2

Conf-911114--6

PNL-SA--20122

DE92 008339

Received OSTI

MAR 0 2 1992

VALIDATION OF PERFORMANCE ASSESSMENT MODELS

M. P. Bergeron
C. T. Kincaid

November 1991

Presented at the
13th DOE LLW Management Conference
November 19-21, 1991
Atlanta, Georgia

Work supported by
the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Richland, Washington 99352

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

VALIDATION OF PERFORMANCE ASSESSMENT MODELS

Marcel P. Bergeron and Charles T. Kincaid
Pacific Northwest Laboratory[a]
Richland, Washington 99352

INTRODUCTION

The pupose of model validation in a low-level waste site performance assessment is to increase confidence in predictions of the migration and fate of future releases from the wastes. Unlike the process of computer code verfication, model validation is a site-specific process that requires site-specific data. This paper provides an overview of the topic of model validation and describes the general approaches, strategies, and limitations of model validation being considered by various researchers concerned with the subject.

DEFINITIONS

Before we discuss the concept of validation, we would like to distinguish and contrast it with another term commonly used in the area of model validation, code verification. A single, universally accepted definition of code verification exists, while aspects of the definition of validation continue to be debated by scientists throughout the world. A number of international groups have been convened to address the issues of code comparisons (e.g., HYDROCOIN) and validation (e.g., INTRAVAL). The International Atomic Energy Agency (IAEA 1982) glossary of radioactive waste management terminology gives as good a reference as any for both terms:

> "Verification: A computer code is verified when it is confirmed that the conceptual model of the real system is adequately represented by the mathematical solution. Verification can thus be carried out, for example, by comparisons among several similar codes and by comparison of numerical codes with analytical solutions.
>
> "Validation: A conceptual model and the computer code derived from it are validated when it is confirmed that the conceptual model and the derived computer code provide a good representation of the actual processes occurring in the real system. Validation is thus carried out by comparison of calculations with field observations and experimental measurements."

The above rely on the definitions of and, perhaps more important, on an understanding of the relationship among several more commonly used terms, including real system, conceptual model, derived computer code, numerical code, analytical solution, and mathematical solution. The real system is the complete, complex (i.e., numerous physical and chemical processes, nonisothermal effects, heterogeneous soils), three-dimensional, and time-dependent hydrogeologic environment. The conceptual model of a real system is obtained by making simplifying assumptions that reduce the completeness, complexity, dimensionality, and time

---

a Pacific Northwest Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under Contract DE-AC06-76RLO 1830.

dependence of the real system while still retaining what are believed to be salient features and processes of greatest importance. Commonly, our objective is to develop and employ a conceptual model that embodies a minimum of these simplifying assumptions while producing economical and realistic simulations of the real system. Such a conceptual model is then translated into a well-posed mathematical problem. To obtain a solution to the mathematical problem, two approaches are commonly taken. The first approach can be pursued whenever the mathematical problem is sufficiently simple to allow a closed-form analytical solution. The solution is then programmed to create a computer code and obtain the analytical solution. In the second approach, the mathematical problem remains too complex to allow an analytical solution, and a numerical (computer) code using numerical methods (e.g., integrated finite differences, finite elements, etc.) is created and a numerical solution is obtained. Both analytical and numerical solutions are referred to as mathematical solutions.

In the context of these definitions, verification speaks to the consistency between a well-posed mathematical problem and its solution or simulation by the computer code. Verification begins with a conceptual model and only deals with our ability to accurately simulate a site using a particular code. Thus, verification does not deal with the complexity of an actual site. In the simplest terms, verification is confirmation that the computer code is properly encoded and produces solutions consistent with those of other codes and previously obtained analytical solutions that the numerical code is able to simulate.

Benchmarking is another term used in the IAEA (1982) definitions. "Benchmarking is the comparison of solutions of the same problem between computer codes. Benchmarking is used to gain confidence that the code being applied is correctly solving the equations embodied in the code by comparison with an independent calculation or another code. Benchmarking is considered part of the verification step."

Given that one employs a verified code that has been calibrated with data from an actual site to form a numerical model, validation addresses the much broader issue of the ability of a model to represent the real system. A valid model is one confirmed to provide a good representation of the real system. A more recent and subtly different definition of validation is provided by the INTRAVAL Validation Overview and Integration Committee (VOIC). The VOIC defines validation as a three-step process (SKI 1990):

1. understanding the active processes and geologic structure

2. resolving how well one can quantitatively simulate through comparisons to field experiments

3. providing peer review and public scrutiny.

Note, this process provides a "good representation of the real system" as called for in the IAEA definition; however, the VOIC has provided a more explicit statement of the process necessary to ensure a good representation. Laboratory experiments and simulation of these experiments can serve to identify and validate relevant physical processes. The VOIC has also made explicit reference to the important role that an understanding of the geologic structure plays in the validation process.

CATEGORIES OF PREDICTIVE ERRORS

Predictive errors resulting from modeling could generally be categorized into the following areas:

- conceptual model errors
- errors in estimation of source-term release
- errors in estimation of hydraulic and transport parameters
- unforeseen changes in hydrologic or other conditions assumed in predictions.

To illustrate these errors in the context of model confidence, we would like to discuss two modeling cases. In these modeling cases, ground-water models were developed and calibrated to site-specific conditions. The calibrated models were then used to predict conditions within the aquifer for some period into the future. In both cases, further data became available or were collected to represent the future conditions within the modeled area, and comparisons were made between the predictions and actual behavior. The comparisons presented interesting insights into the area of model confidence and what limitations we face with model validation. One case concerns modeling studies conducted within the Snake River Plain aquifer in the vicinity of the Idaho National Engineering Laboratory (INEL) in southeast Idaho. The other case concerns modeling studies in an irrigated area within the Arkansas River Valley in southeast Colorado. The following discussions present summaries of these cases.

Snake River Plain Aquifer in Southeast Idaho

The first case revisits a modeling study at the INEL documented in a report by Robertson (1974). The following is a summary of information presented in the Robertson report along with extracts from two later reports by Lewis and Goldstein (1982) and Konikow and Patten (1985).

The original modeling by Robertson (1974), considered to be one of the most significant efforts in terms of its approach and success in predicting observed data and information, focused on simulating the effect of two liquid waste discharge facilities, the Test Reactor Area (TRA) and the Idaho Chemical Processing Plant (ICPP) (Figure 1). Liquid wastes had been discharged at these two facilities continuously since 1952. The modeling involved development of a solute transport model that was used to simulate the movement of discharged contaminants into the underlying basalt aquifer, referred to as the Snake River Plain aquifer.

The model, based on a method of characteristics model very similar to that developed by Konikow and Bredehoeft (1978), was calibrated in stages. A flow model was first calibrated to a 2600-mi^2 area surrounding the TRA and ICPP areas. A steady-state model was calibrated to measurements made in May and June 1965; the model was then calibrated in the transient mode to head changes measured between December 1964 and December 1965. This flow model provided the hydrologic basis for a solute transport model that was calibrated for a smaller area wherein the contamination had primarily occurred. The transport model was calibrated to the

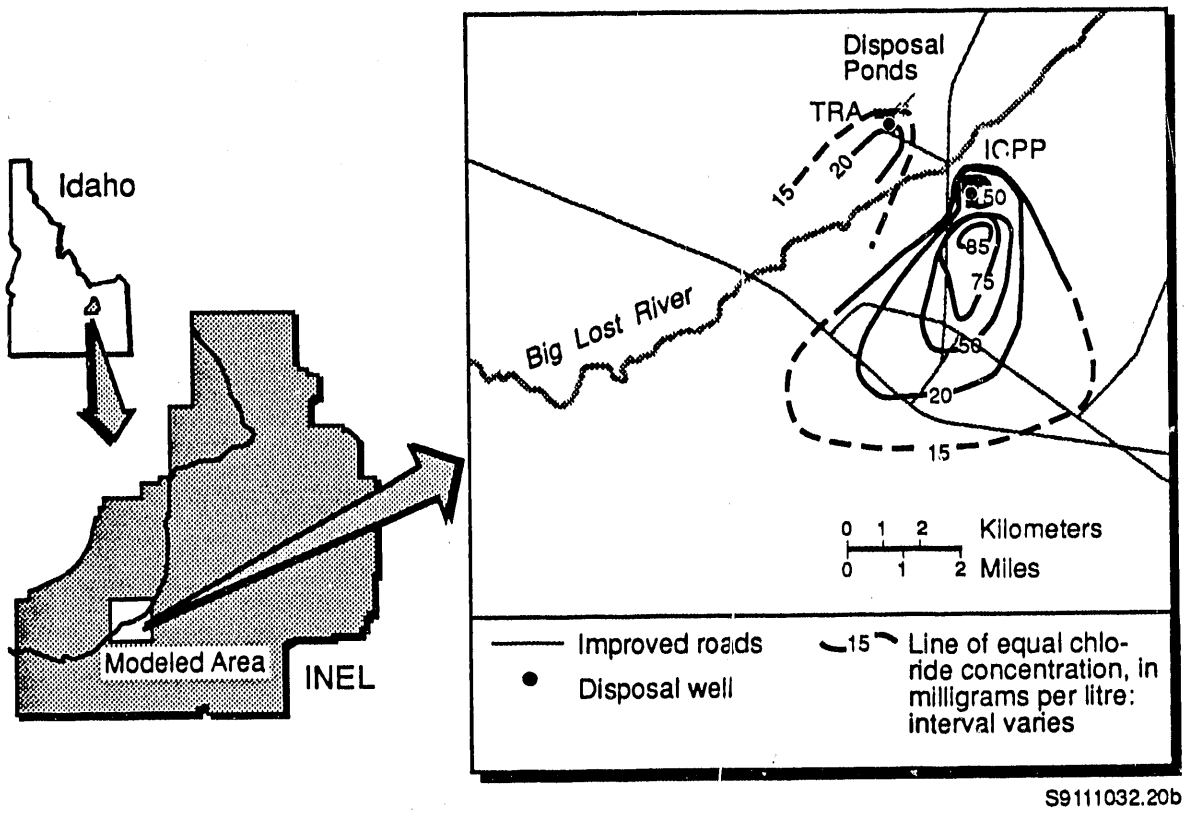

Figure 1. Map of the Test Reactor Area (TRA)-Idaho Chemical Processing Plant (ICPP) Vicinity Showing Observed Distribution of Waste Chloride in the Snake River Plain Aquifer in 1972.

20-year history of tritium and chloride contamination that was observed in 45 wells near and downgradient of known sources of contamination. The distribution of waste chloride as it was observed in November 1972 is also depicted in Figure 1. The calibrated solute transport model was used by Robertson to predict future concentrations of chloride, tritium, and strontium-90 for the years 1980 and 2000 for alternative future conditions. An important assumption was that disposal at the TRA and ICPP facilities would continue at 1973 disposal rates.

Lewis and Goldstein (1982) collected additional data downgradient of the TRA and ICPP facilities within areas where the Robertson model had predicted waste chloride and tritium transport. A distribution of waste chloride measured in October 1980 and the predicted distribution for 1980 by Robertson (1974) are compared in Figure 2. A comparison of these results indicates that the observed and predicted contaminant plumes are in general agreement in terms of direction, extent, and magnitude. Certainly, given the purpose of the original modeling, the modeling effort was considered very successful. However, Lewis and Goldstein indicated that while the overall match was quite good, as one would expect when the details were examined, there are a number of differences: the observed plume is broader and has more lateral spreading than predicted, the leading edge of the plume had not spread as far south as predicted, and a secondary plume of contamination originating from the TRA was not detected in the 1980 measurements to the extent predicted.

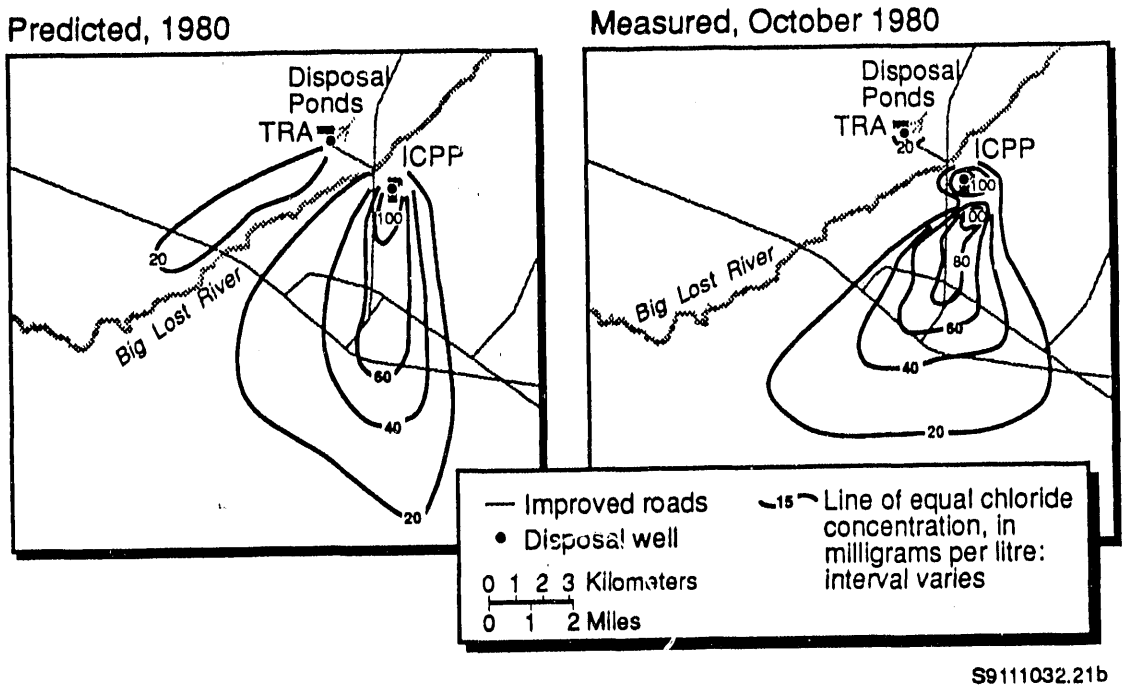

Figure 2. Distribution of Waste Chloride in the Snake River Plain Aquifer.

Lewis and Goldstein (1982) suggested a number of factors that would contribute to these discrepancies. These factors were 1) less dilution from recharge occurred between 1977 and 1980 because of below-normal river flow, 2) chloride disposal at the ICPP increased for several years prior to 1980, 3) the model grid may have been too coarse, 4) inaccurate hydraulic and transport parameters were selected in model calibration, 5) vertical components of flow and transport not considered in the two-dimensional analysis may have been significant, 6) the quality and quantity of wells may not have been sufficient to map the observed plumes, and 7) the numerical methods may have introduced some error.

Arkansas River Valley in Southeast Colorado

Konikow and Person (1985) revisited a modeling effort documented in Konikow and Bredehoeft (1974) for an irrigated area within the Arkansas River Valley in southeast Colorado (Figure 3). In that study, a solute transport model was applied to an 11-mile reach of the valley to predict changes in salinity in response to spatially and temporally varying irrigation stresses. The model was calibrated in 1973 based on measurements made in the area in 1971 and 1972 (Figure 4). These data indicated a trend of increasing salinity of about 2% to 3% per year if then-current irrigation practices continued. This basic trend provided the basis for the use of the calibrated model in predicting the long-term changes in ground-water salinity.

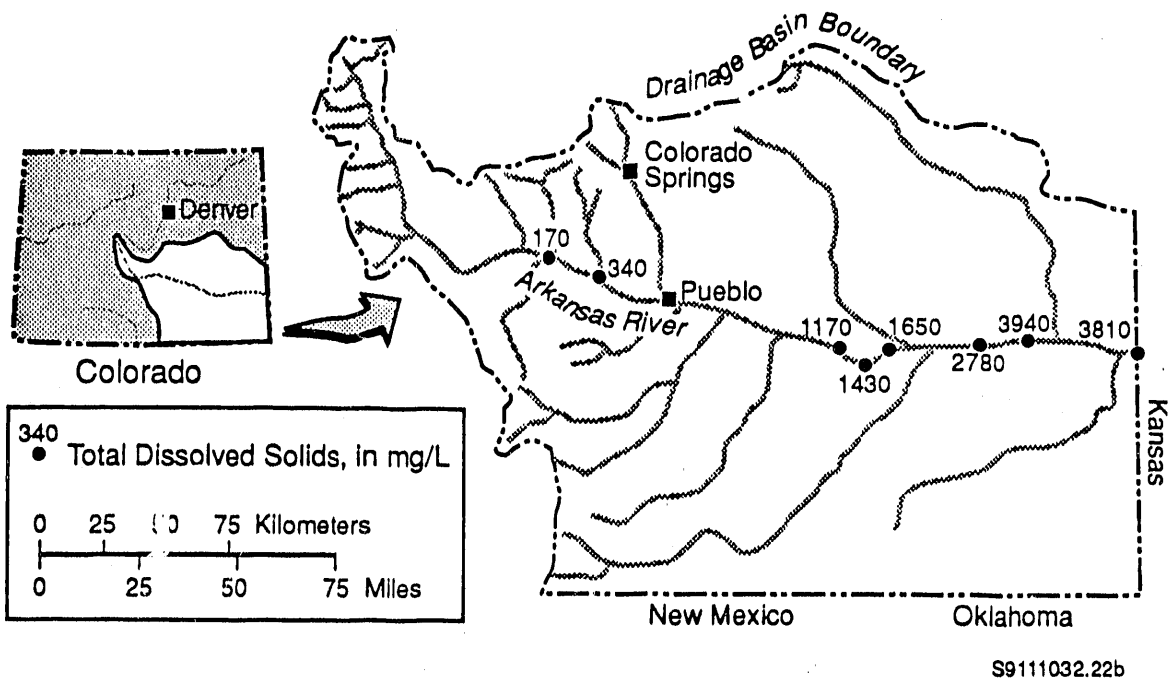

Figure 3. Downvalley Increase in Salinity in Arkansas River Valley in Southeast Colorado.

Konikow and Person (1985) reexamined the model results in light of new data that were collected in the area in the winter of 1982 (see Figure 4). Nonparametric and parametric statistical tests, used in the evaluation, indicated that while there was a statistically significant increase in salinity between 1971 and 1972, data from 1972 and 1982 showed no significant net change. The difference between 1972 and 1982 data and the results from original predictions and new predictions made with a recalibrated model are shown in Figure 5.

Analysis of historical data (1895, 1923, 1959-1961, and 1964) not considered in the original study supported a hypothesis that ground-water salinity in the study area had reached a long-term dynamic equilibrium condition. Konikow and Person (1985) concluded that the original model predictions were invalid because the model was calibrated to a period that was not representative of the long-term trends in ground-water salinity changes.

LIMITS OF MODEL VALIDATION

Confidence in most modeling efforts involves a two-step process that first builds confidence in the computer code (i.e., code verification) and then builds confidence in the site-specific model (i.e., model validation). The code verification process involves 1) the comparison of the code with relevant analytical solutions and 2) the comparison of code results with others on

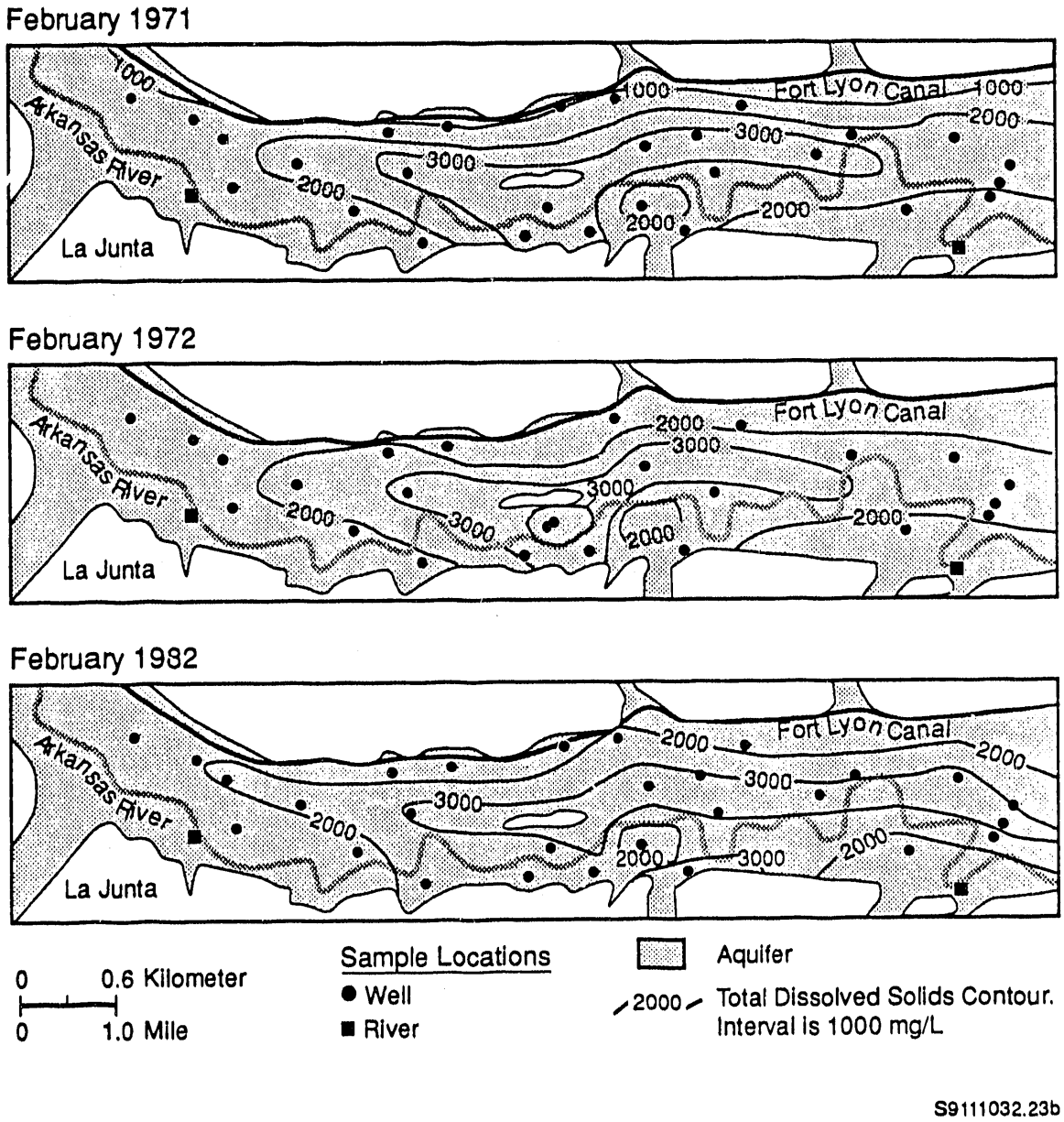

Figure 4. Dissolved Solids Concentration in Arkansas River Valley in Southeast Colorado

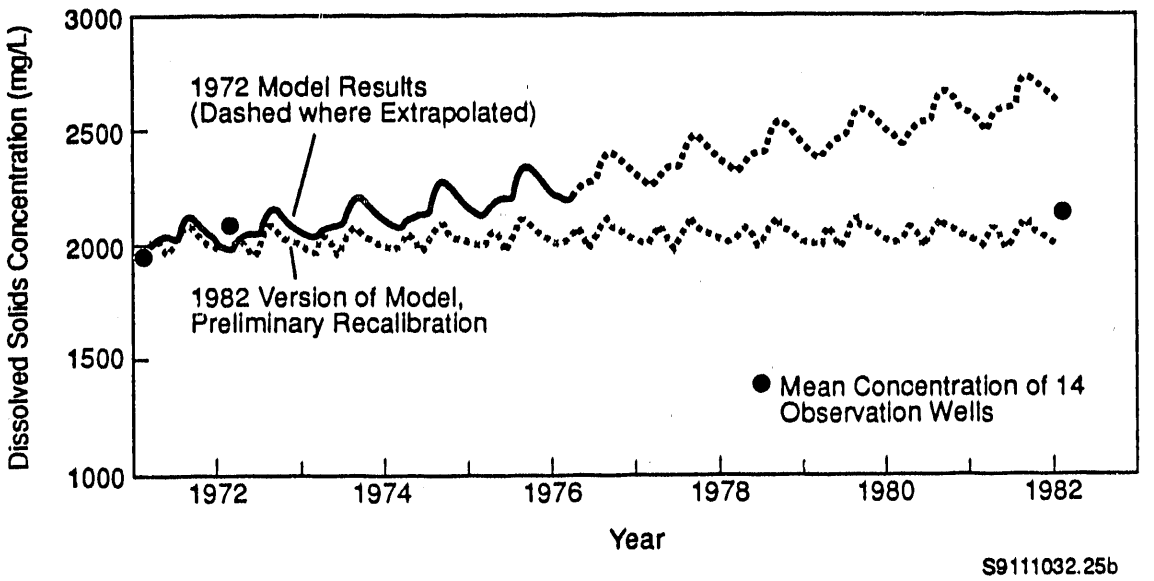

Figure 5. Comparision of Predicted Model Results with Mean Dissolved Solids Concentration.

relevant problems (i.e., benchmarking). Confidence in the site model (i.e, the conceptual model and computer code) is achieved by calibration of the model to one set of data and validation of the model to independent set(s) of data/information. These data could be from other site-specific information or from field experiments performed at sites with similar settings.

When one examines this approach to confidence building, several limitations to the validation of any model obviously exist. First, and foremost, it is important to understand that there is no such thing as absolute or everlasting validation; validation is always relative to the existing body of empirical knowledge. This knowledge includes the extent and limits of our scientific understanding and the body of site-specific data that exists at the time. The fact that validation is relative to this knowledge is not bad because the fit between existing data and current theory is nearly always imperfect. As our understanding of the environment improves and as more site-specific data are collected, the validation of a model must be revisited and confirmed again.

The difficulty of confidence building (i.e., model validation) increases significantly with increasing spatial and temporal scale of the modeling effort. Economics typically dictate that data density decreases as the scale of investigation increases. The results from INTRAVAL indicate that individual processes may be investigated with laboratory-scale experiments and

the resulting observations/measurements may be simulated very closely with the appropriate model. However, there is increasing uncertainty as the spatial and temporal scales increase to the field or site scale with desired analysis times of tens, hundreds, or thousands of years.

Another limiting aspect of validation is its site-specific character; a model shown to be valid for one site or hydrogeologic profile may not be valid at another site a relatively short distance away. The ability to claim and demonstrate the broad-based validity of a model (i.e., to extend its valid application to another nearby site) will depend on the degree of similarity in the active physical, geochemical, and microbiological processes and in the similarity of hydrogeologic structure.

An aspect of validation of great significance is the data-intensive effort that is required for validation. The common situation is that the data in the quantity or the quality required are available at a site of interest. Validation requires that one test a predetermined model using independent data. To first create the "predetermined" model, one commonly calibrates the model to a site-specific data set. Then, a second, independent data set that captures the character of the site under some altered state is needed for validation of the model. Thus, two independent data sets are required; one for calibration and another for validation activities.

The quality of a validation effort is also impacted by the availability of field measurements of model performance measures rather than field measurements of readily measured parameters. For example, it is relatively easy to measure the water-table elevation of an aquifer; however, one needs to know the hydraulic and transport properties and ground-water velocity to validate flow and contaminant transport models of the aquifer. These performance measures are more difficult to measure in the field. In the unsaturated zone, it is relatively easy to sample or measure moisture content of the soil; however, models of water flow predict water matrix potential or pressure, and are used to estimate water, hence contaminant, velocity. Thus, the difficult to obtain field measurements of matrix potential or soil water velocity are preferable for a validation study.

Validation efforts at any site are impacted by each of the above limitations. Data sets for subsurface flow and transport model validation are, in general, limited. Records of contaminant sources, migration, and fate in the unsaturated zone at most sites are at best incomplete and do not at this time present an opportunity for site-specific validation. Therefore, validation of unsaturated zone flow and transport models may have to rely on offsite data sets. Pending the development of independent calibration and validation data sets for sites, models will be applied using as much site-specific data as possible.

OTHER MODEL VALIDATION METHODS

Given the limited nature of model validation, an approach based on the underlying philosophy of "As Validated As Reasonably Achievable" (AVARA) is proposed. This approach recognizes that complete validation is not achievable and infers that validation efforts should acknowledge that resources are limited. Validation should be limited to important components of the performance assessment model using available data and information. This approach also suggests that other methods to validate steps in the modeling process include 1) data review and evaluation; 2)

conceptual model development; and 3) selection, development, and implementation of numerical models.

A recent paper on model validation by Tsang (1991) provides a list of other methods for validation that might be considered. These methods are summarized below:

Event Validation -- Qualitative comparison of events of simulation model with those in real system.

Face Validation -- Peer review by technical experts.

Trace Validation -- Analysis of behavior of different elements or entities of model are traced to determine if logic is correct.

Historical Methods -- 1) Evaluate modeling assumptions; 2) validate each assumption, when possible, with empirical tests; and 3) compare input-output relationship of model with field behavior.

Internal Validation -- Sensitivity analysis of important model parameters. Particularly important in validation of stochastic models. High degree of variability in model results may cause model results to be questionable.

Historical Data Validation -- Part of site data are used for model calibration. Model predictions are made to check against remaining data.

Predictive Validation -- 1) Model is used to provide predictions for a given field or laboratory experiment and 2) further measurements are made to check predictions.

Tsang (1991) suggested another method that could be important for model validation would be the use of multiple assessment groups. This method would involve making independent assessments of the same hydrogeologic system by two or more technical groups. Differences in assumptions, approaches, and results will likely 1) cross verify underlying assumptions/approaches, 2) provide alternate assumptions and approaches, and 3) stimulate new consideration and understanding of the modeled system. This approach would, in general, lead to improving the overall model validation process by helping to eliminate gross errors and to increasing credibility with the scientific community and the public.

OPPORTUNITIES FOR FURTHER VALIDATION

Monitoring of site behavior and performance conducted during the operational period offers other important opportunities for collecting data and information to support model validation. Model validation using these data can help improve confidence in model predictions prior to the

site closure period. Surrogate experiments could be designed and conducted during the operational period to validate components or specific assumptions related to performance assessment. Such experiments could include:

- lysimeters used to evaluate infiltration through protective barriers
- field leaching designed to evaluate waste form release
- tracer tests to examine near-field transport in vadose zone
- tracer tests to evaluate solute transport in saturated zone.

SUMMARY AND CONCLUSIONS

Model validation in low-level waste site performance assessment is needed to increase confidence in predictions of the migration and fate of future releases from these wastes. Unlike code verification, where the focus is on the appropriateness and accuracy of the analytical or numerical implementation of a computer code, model validation is a site-specific process. A site model is considered validated when it is confirmed that the conceptual model and the implemented computer code provide a good representation of the actual processes occurring at the site. The current practice of developing and fine tuning a site-specific model (i.e., calibration) requires site-specific data; validation requires independent calibration and validation data sets.

While validation and confidence in predictions are the goals we strive for, there are limits to model validation. These limits exist because validation is always relative to our existing knowledge of the geologic, hydrologic, and geochemical processes operable at the site. A number of post-audits of model predictions documented in the literature demonstrate the difficulties one can encounter in making predictions with limited information. Significant discrepancies between predictions and the actual results occur in response to errors in data interpretation used to formulate the site conceptual model, errors in the estimates of source-term release and of hydraulic and transport parameters used in the model predictions, and unforeseen changes in assumed hydrologic conditions (e.g., increased gradient or ground-water usage). These discrepancies suggest the accuracy of predictions is heavily dependent on both the representativeness of available data and the quality of the interpretation of those data.

These limitations force us to conduct and accept qualified validation of performance assessment models and codes based on often limited site information. An AVARA approach, using multiple validation methods, should be the guiding principle. Model predictions need to be reviewed and updated as more information is collected during operational monitoring or is made available through the use of carefully selected surrogate experiments conducted, monitored, and analyzed prior to the post-closure period.

REFERENCES

IAEA (International Atomic Energy Agency). 1982. Radioactive Waste Management Glossary. IAEA-TECDOC-264, Vienna, Austria.

Konikow, L. F., and Bredehoeft, J. D. 1974. "Modeling Flow and Chemical Quality Changes in an Irrigated Stream-Aquifer System." Water Resources Research 10(3):546-562.

Konikow, L. F., and Bredehoeft, J. D. 1978. "Computer Model of Two-Dimensional Solute Transport and Dispersion in Groundwater." Techniques of Water-Resources Investigations, Book 7, Chapter C2, U.S. Geological Survey, Reston, Virginia.

Konikow, L. F., and E. P. Patten, Jr. 1985. Groundwater Forecasting, in Hydrological Forecasting. Editors, M. G. Anderson and T. P. Burt. John Wiley & Sons, Limited, New York, New York.

Konikow, L. F., and Person, M. 1985. "Assessment of Long-Term Salinity Changes in an Irrigated Stream-Aquifer System." Water Resources Research 21(11):1611-1624.

Lewis, B. D., and Goldstein, F. J. 1982. Evaluation of a Predictive Ground-Water Solute-Transport Model at the Idaho National Engineering Laboratory, Idaho. Water-Resources Investigation 82-25, U. S. Geological Survey. Idaho Falls, Idaho.

Robertson, J. B. 1974. Digital Modeling of Radioactive and Chemical Waste Transport in the Snake River Plain Aquifer at the National Reactor Testing Station, Idaho. Open-File Report IDO-22054, U. S. Geological Survey, Idaho Falls, Idaho.

SKI (Swedish Nuclear Power Inspectorate). 1990. The International Project: Background and Results 1990. Nuclear Energy Agency, Organization for Economic Co-Operation and Development, Paris, France.

Tsang, C. F. 1991. "The Modeling Process and Model Validation." Groundwater 29(6):825-831.

END

DATE
FILMED
4/01/92

I