ground water

Issue Paper/

From Models to Performance Assessment: The Conceptualization Problem

by John D. Bredehoeft[1]

Abstract

Today, models are ubiquitous tools for ground water analyses. The intent of this paper is to explore philosophically the role of the conceptual model in analysis. Selection of the appropriate conceptual model is an a priori decision by the analyst. Calibration is an integral part of the modeling process. Unfortunately a wrong or incomplete conceptual model can often be adequately calibrated; good calibration of a model does not ensure a correct conceptual model. Petroleum engineers have another term for calibration; they refer to it as *history matching*. A caveat to the idea of history matching is that we can make a prediction with some confidence equal to the period of the history match. In other words, if we have matched a 10-year history, we can predict for 10 years with reasonable confidence; beyond 10 years the confidence in the prediction diminishes rapidly. The same rule of thumb applies to ground water model analyses. Nuclear waste disposal poses a difficult problem because the time horizon, 1000 years or longer, is well beyond the possibility of the history match (or period of calibration) in the traditional analysis. Nonetheless, numerical models appear to be the tool of choice for analyzing the safety of waste facilities. Models have a well-recognized inherent uncertainty. Performance assessment, the technique for assessing the safety of nuclear waste facilities, involves an ensemble of cascading models. Performance assessment with its ensemble of models multiplies the inherent uncertainty of the single model. The closer we can approach the idea of a long history with which to match the models, even models of nuclear waste facilities, the more confidence we will have in the analysis (and the models, including performance assessment). This thesis argues for prolonged periods of observation (perhaps as long as 300 to 1000 years) before a nuclear waste facility is finally closed.

Introduction—Models

Models play a key role in the analysis of many, if not most, ground water problems. They are especially important in predicting the behavior of nuclear waste facilities far into the future. The Waste Isolation Pilot Plant (WIPP, a geologic repository for transuranic wastes in New Mexico) was recently opened and is receiving nuclear weapons waste. Yucca Mountain (the proposed high-level nuclear waste repository in Nevada) is near the licensing stage. Hydrogeological models play a key role in assessing the safety of these facilities.

The purpose of this paper is to discuss philosophically the use of models in making predictions. Many of these ideas have been expressed elsewhere, yet they seem worth restating. In particular, I want to examine the role that the conceptual model plays in analysis. I take a historical perspective in developing these ideas.

In the 19th century, various laws that describe the movement of heat, electricity, and ground water through a continuum were derived. Of special concern to those of us that investigate ground water is Darcy's law. By applying the principle of conservation of mass and incorporating Darcy's law as a constitutive relationship, we can derive a partial differential equation that describes the hydraulic head throughout a porous medium. Once the head is determined, we can apply Darcy's law to derive the ground water flow vectors throughout the system. These principles form the basis for all ground water flow and transport models.

For many ground water problems with simple geometry, simple parameter distributions, and simple boundary conditions analytical solutions to the mathematical problem can be derived. Because there is an analogy between the flow of ground water and the flow of both electricity and heat, we often can find the mathematical solution for the appropriate boundary value problem in the literature on

---

[1] The Hydrodynamics Group, 127 Toyon Lane, Sausalito, CA 94965; jdbrede@aol.com

heat flow and/or electrical flow. The analogy between heat flow and ground water flow was enriched in the 1930s and 1940s when Theis (1935) suggested that transient ground water flow was analogous to unsteady heat flow, and Jacob (1940) derived the transient ground water flow equation from first principles. Ground water in the 1940s and 1950s went through a period when various boundary value problems were solved for pumping wells; numerous pumping test procedures were developed. Many of the pumping test solutions could be found in the classical literature on heat flow. Carslaw and Jaeger (1959) was on the shelf of most serious ground water hydrologists. Many of the classical solutions involved numerical integration of mathematical functions; the digital computer enhanced the capability to numerically integrate these functions.

The Conceptualization Problem

The various pumping test solutions involve different conceptual models of the well and its geologic environment. For example, the response of a well pumping at a constant rate from an extensive confined aquifer can be analyzed as if (1) the overlying and underlying beds are impermeable (the Theis solution), (2) the overlying and underlying beds are leaky without storage (the leaky aquifer solution), or (3) the overlying and underlying beds are leaky with storage (the modified leaky aquifer solution, Hantush 1964). A pumping test in the Dakota Sandstone at Wall, South Dakota, illustrates the point (Bredehoeft et al. 1983). Different investigators fit the Wall data to both the Theis solution and the modified leaky aquifer solution; the data fit either solution equally well. We obtain a different answer depending on the conceptual model chosen; the predictions of long-term future well performance will be different depending on which model is selected.

Usually the conceptual model chosen for analysis is an a priori decision of the analyst. Sometimes the fit of the data to the analytical solution will suggest that the conceptual model is inappropriate, but more often than not the data will fit more than one conceptual model equally well, as was the case at Wall, South Dakota. My point is that we can choose the wrong conceptual model, fit the data, and get a wrong answer. In the 1940s and 1950s, hydrogeologists did not call these solutions to pumping test model analyses, but we might today.

We cannot overemphasize the role of the choice of the conceptual model in any analysis of a ground water system. A wrong conceptual model invariably leads to poor predictions, no matter how well the model is fit to the data. Time and again, the errors in prediction revolve around a poor choice of the conceptual model (Konikow and Bredehoeft 1992; Oreskes and Belitz 2001). Modeling invariably involves simplifying the real system into a conceptual model that can be analyzed; that conceptual model must capture the essence of the problem. Choosing the appropriate conceptual model is usually a matter of professional judgment. It is how we conceptualize the problem that separates good analysts from poor ones, especially today when anyone can run codes such as MODFLOW.

We tend to regard our conceptual model as immutable. Yet I remember one of my mentors as a young geologist, N.W. Bass, said to me: "A geologic report is always a progress report." I continue to reflect on this remark. What we choose as a conceptual model is a function of the status of knowledge in science. For example, plate tectonics changed geology and changed our conceptual model of tectonics. Theis (1935) and Jacob (1940) changed ground water hydrology by introducing the transient theory of ground water flow. Finding chlorine–36 at depth in Yucca Mountain has caused the community to rethink transport through a fractured unsaturated zone. As yet, there is no consensus on the appropriate conceptual model for transport within Yucca Mountain (National Research Council 2001). The point is that our conceptual model changes with advances in the science; concepts are by no means static.

Models

At the time well tests were the standard tool for analysis for hydrogeologists, it was apparent to some individuals that it would be of great value to find a procedure to solve the more global problem of flow through a porous medium with varying parameter distributions and complex boundary conditions. In other words, to treat flow in an entire aquifer. A group at the U.S. Geological Survey, led by Herb Skibitski, developed the resistor-capacitor electrical network as an *analog model* for ground water flow. This introduced into ground water the idea of a *model* of an aquifer system.

A parallel effort was under way in the petroleum industry where reservoir engineers simulated flow in realistic hydrocarbon reservoirs. A petroleum reservoir is often more complex than saturated ground water systems because of the presence of multiphases—oil, gas, and water. The petroleum engineering effort to model a reservoir used the digital computer. Some of the best applied mathematicians of the 1950s and 1960s worked on developing numerical methods to solve the equations of flow in porous media. The petroleum industry referred to both the computer codes and the models of specific fields as *reservoir simulators.* The term used to describe the analyses was *reservoir simulation* rather than modeling.

As digital computers grew in power, the analog methods of the 1950s and 1960s used in ground water were replaced by digital computer methods in the 1970s. As digital computers became more powerful and less costly, modeling became widely used. With the power of today's PCs, models of ground water systems are now commonplace.

The digital computer codes had the added benefit that solute transport also could be modeled. In a general way, sets of partial differential equations could be solved simultaneously. The analog models dealt only with the solution of the ground water flow equation. The digital computer added new dimensions to modeling. A whole industry has grown up that produces models of ground water flow and transport that are easily implemented. There are a number of pre- and post-processors for MODFLOW and MT3D, the most common of the ground water flow and transport codes. The pre- and post-processors make modeling relatively easy, and enable very large problems, involving big grids, to be modeled. Without the pre- and post-processors, large grids are exceedingly difficult to implement—for all practical purposes, they become intractable.

Calibration

An integral part of the modeling procedure is calibration. Calibration involves fitting the model output to a set of observations. Hopefully, at some point in the process, the model parameters are adjusted so that an *adequate* fit to the observations is achieved. Originally, calibration was a trial-and-error procedure. In recent years, the process of adjusting the parameters to achieve an adequate fit has been automated.

Numerical measures of the goodness-of-fit between the observations and the model predictions have been devised. The numerical measures provide the appearance that judging the adequacy of fit during calibration is no longer subjective. However, in the end, what constitutes an adequate fit is a subjective decision. Statistical measures of the goodness-of-fit can be calculated, but the question of whether a model is calibrated is a decision left to the analyst.

There are problems in the calibration procedure. As suggested by the previous discussion of pumping test analysis, the calibration commonly does not test our conceptual model. *In other words, a model involving a wrong or incomplete conceptual model can be adequately calibrated.* It is generally conceded that a model, even if it is well calibrated, is nonunique; another parameter set might result in an equally good calibration (Bethke 1992).

Post Audits

Since models have now been around for several decades, it is possible in certain limited instances to evaluate their performance. Predictions were made that can now be compared to what happened to a particular system. Many audits do not really test the adequacy of the model because what took place with the real system was not a scenario that was analyzed initially. Typically, pumping followed a different pattern than anticipated.

There are a limited number of post audits of model predictions; they are not reassuring. Many models did not provide good predictions (Anderson and Woessner 1992; Konikow and Bredehoeft 1992). Many models suffered from a conceptual omission: an important process was overlooked. In other cases, the range of parameters was much larger than was included in the model analysis. Models are known to have provided poor predictions, even models that were thought to have been well calibrated.

Validation

Validation is a term promoted by the nuclear waste community. Different people variously define validation; there is no consensus on what it means. Furthermore, in most cases, the goal of calibration and validation are the same: In both cases, we seek to create the best possible representation of the system. We as a community have formulated restrictive, and rather special, definitions of what it means to validate a code.

Recognizing that the traditional history match was impossible, the nuclear waste community set out to test different codes in situations where shorter histories of performance were available. They called this test of the models *validation*. This is only one of many specialized definitions of validation. This test of the codes was no different than the calibration procedure models normally undergo. If the model of a specified system could be adequately calibrated, the code was deemed validated. In many instances, we can substitute the words *well calibrated* for *validated* without changing significantly the author's meaning.

There are both pragmatic and philosophical grounds to avoid the idea of validation. The idea of validation (or invalidation) is deeply rooted in the philosophy of science. On philosophical grounds, Popper (1968) argued that scientific theory can be invalidated—not validated. Of course, Popper is not the only philosopher of science. Others, notably the pragmatists, of which John Dewey is perhaps the best known, argued that we learn from experience, observations, and mistakes (Menand 2001). The pragmatists argued we never find real truth, but we do get closer to understanding. Kuhn (1970) suggested that scientists try to make existing theory work until finally the evidence indicates that it does not; then they embrace a new theory. None of these philosophers argued that one could validate.

It is unfortunate that we have allowed the term *validation* to become a part of the model lexicon. Oreskes and Belitz (2001) summarize the status of validation:

> "The inherent uncertainties of models have been widely recognized, and it is commonly acknowledged that the term 'validation' is an unfortunate one, because its root—valid—implies a legitimacy that we are not justified in asserting. . . . But old habits die hard and the term persists. In formal documents of major national and international agencies that sponsor modeling efforts, and in the work of many modelers, 'validation' is still widely used in ways that assert or imply assurance that the model accurately reflects the underlying natural processes, and therefore provides a reliable basis for decision-making. This usage is misleading and should be changed. Models cannot be validated. The reasons why have been outlined in detail elsewhere (Konikow and Bredehoeft 1992; Oreskes et al. 1994)."

Reservoir Engineering: A Pragmatic Approach

The ground water community could take a lesson from petroleum reservoir engineering. The usual practice is to history match the reservoir simulator output to some temporal history of production. This is calibration in the ground water lexicon. Based on the match, a prediction of future performance is made, but one is cautious in extending that prediction much beyond a period equal to the production history. In other words, the rule of thumb is that, if we make a 10-year history match, we might be reasonably confident in predicting the next 10 years of performance; however, beyond 10 years, the confidence in a prediction greatly diminishes.

The reservoir engineering community makes no claims about the validity of the model. They simply imply: (1) we have a model that we think incorporates the appropriate physics and chemistry, including the appropriate parameter set, that matches an observed temporal history of reservoir performance; and (2) we will use that model to predict future reservoir performance. Furthermore, continued monitoring of the system is used to refine and improve the model.

Reservoir simulation is important in the petroleum industry. A small improvement in the fraction of petroleum recovered from a reservoir can amount to many millions of dollars. It is worth looking to reservoir engineering practice as a guide to modeling ground water systems, especially systems involving high risk to society, such as nuclear waste repositories.

Many of the same techniques are used with most normal ground water models. In many water supply models, we have a history of the response of a ground water system to stress. One makes a model that reproduces the history (is calibrated), and then makes predictions of future performance. A well-known caveat is that if the system reaches a new state, the past history may be a poor analog for future performance. Perhaps an example is worth mentioning.

Ground water is being mined from the Denver Basin aquifers in the area just to the south of Metropolitan Denver. Water levels over much of the Denver Basin are declining at rates of 20 to 30 feet per year. So far, the aquifers are still artesian over much of the basin. The question arises as to what will happen as the artesian head is removed and the system becomes water table. Theory suggests that the rate of water level decline will slow as the aquifers pass to water table conditions. However, there are a number of complicating factors. The aquifers are composed of multiple lenticular sand bodies that are not continuous either vertically or across the basin. The layered nature of the sand bodies that make up the permeable portion of the aquifer restrict its drainage; a highly layered aquifer may drain quite differently than a massive thick sand. In addition, there are extreme drawdowns during summer pumping periods that cause the sand bodies to become unsaturated during the summer and then saturate again during the winter; this cyclic drawdown tends to trap air in the sands. There is a debate raging among concerned professionals over what the impact of the complications will be on the water table drainage of the Denver Basin aquifers. We can only speculate until there is some history of how the system responds under water table conditions.

Modeling as an Iterative Process

Good modeling is an iterative process. As new data are acquired, the model is revisited and adjusted (or recalibrated) so that the model predictions are consistent with all the data, including the new data. The model becomes a living tool for analysis. With this paradigm, the modeling strategy changes; it requires continued monitoring and model updating.

We see this strategy at work in many ground water problems. Many problems, especially where there is major concern over the water supply, have been modeled numerous times. I was recently in the Tampa Bay area where there have been three models of seawater intrusion built during the last decade; as digital computers increased in power, each model was more complex. The models progressed from two-dimensional cross sections to fully three-dimensional representations of the system. Each improved the representation of the system. A new flow model is also under construction for the area. The modeling continues because it provides new insights and increased confidence in understanding.

The iterative process is important in addressing the adequacy of the conceptual model. A mismatch between the model prediction and the observed data should raise the issue of conceptualization: Is the mismatch a result of parameter misadjustments or does it suggest that we rethink the conceptual model?

Nuclear Waste

That brings me to the licensing of nuclear waste facilities. The time horizons for these facilities are long—1000 to 10,000 years or longer. The usual model practice of matching a temporal history of system performance and then predicting for a more or less equal period is out of the question. The models are used to predict system performance well beyond an observable history. As modelers, we have to hope that we have (1) included all of the relevant processes in our conceptual model of the system, (2) described the appropriate boundary conditions that are operable through the time horizon of our prediction, and (3) captured the parameters, and their uncertainty, in our representation of the system. This is a tall order.

Performance Assessment

A nuclear waste facility is judged safe if the predicted exposure to radioactivity to an individual located near the boundary of the facility is below a set standard. To make the dose calculation, transport of radioactive components of the wastes is investigated along various exposure pathways. Transport of radioactive isotopes of concern, often by moving subsurface fluids, is predicted within various components of the repository system along the pathways of concern. Various models of the transport processes are linked to perform a performance assessment. The performance assessment is run stochastically so that a probabilistic prediction of the radioactive dose to the hypothetical individual of concern is computed. Ewing et al. (1999) review the use of performance assessment.

Performance assessment sounds obscure in the abstract; perhaps an example will illustrate the procedure. The Waste Isolation Pilot Plant (WIPP) is a salt mine, 2200 feet deep, in the Permian Basin of New Mexico, near Carlsbad, where nuclear waste created by the U.S. weapons programs is being buried. The original concept was that the Salado Salt in which the mine is built would deform plastically and encapsulate the buried nuclear waste within a period of several decades. There are problems with this concept. Once an exploratory mine was constructed, the salt was found to contain 1% to 3% interstitial brine—brine between the salt crystals. This brine migrates into the mine. During mining, the mine ventilation removes the moisture; however, once the mine is closed, the brine accumulates in the closed rooms. Under humid or partially wet conditions, steel drums containing much of the waste will react with the brine producing hydrogen gas, and cellulose in the waste will biodegrade, producing additional gas. Under these conditions, the repository becomes a pressurized, sealed mine in which the pore fluids (brine and gas) resist the plastic collapse of the salt. Finding 1% to 3% brine within the salt required a revised conceptual model for WIPP.

Further complicating WIPP are commercial grade potash deposits that overlie the mine, and oil and gas fields in the surrounding area. The oil and gas fields are believed to extend beneath the repository. In evaluating the safety of the repository, the U.S. Environmental Protection Agency (EPA) insisted that the scenario of drilling into the repository be assessed. EPA directed that, for the assessment, the current rate of drilling in the area, using the current drilling technology, be extended throughout the 10,000-year time horizon of analysis. The attorney general of New Mexico challenged in court EPA's idea of extending current technology and current drilling frequency into the future. The U.S. Court of Appeals agreed with EPA that it was reasonable to use the current technology and frequency over the entire time horizon as a surrogate measure of the risk from an unknown future in which both drilling frequency and technology will undoubtedly change.

Most investigators thought that the Achilles' heel of WIPP was the human intrusion scenario. Extending the current drilling rate for the 10,000-year planning horizon means statistically that WIPP will be drilled into several times with a probability of 1.0. Using EPA's imposed conditions, there will be drilling hits into the repository.

Performance Assessment: A Cascade of Models

A number of models of the mine and its environment were linked into a single system: the *performance assessment model.* At the base of the pyramid of performance assessment models was a model of the near field; the actual mine, and the reservoir formed by the fluid-filled nuclear wastes. The basic fluid model of the mine describes the multiphase pressure environment within the mine (1) once the mine is sealed, (2) the salt deforms around the waste, and (3) the moist waste and steel drums produce gases. A submodel predicts the temporal concentration of radioactive chemical species of concern in the fluids contained within the waste. An additional submodel predicts the rock mechanics of the salt deformation in response to the fluid pressure in the repository.

The near-field model was embedded in a far-field model that represents the geologic setting that contains the mine. As explained previously, human intrusion through subsequent drilling into the facility is a major concern. Additional submodels of the performance assessment ensemble describe the exhumation of nuclear waste by subsequent drilling. There are models of how drilling through the repository waste brings waste to the surface.

The performance assessment model is operated in a stochastic mode so that a probabilistic prediction is generated. Performance assessment recognizes that the parameters of the models are incompletely known. Using a Latin Hypercube sampling procedure, the parameters of the various submodels are sampled from their expected distribution, although in many instances the assumed parameter distributions are highly uncertain. The idea is that, by running the performance assessment model with repeated sampling of the parameters, we can calculate a statistical distribution of the probable radioactive dose to the hypothetical individual of concern. It is possible through sensitivity analysis to identify the parameters that most control the predicted dose of radioactivity. It is, however, difficult to determine how errors are propagated through the suite of interconnected models (Konikow and Ewing 1999).

WIPP was judged safe largely on the basis of performance assessment. The WIPP performance assessment will form the template for future safety analyses of nuclear waste repositories in the United States.

Bredehoeft (1997, 1998) argued in the case of WIPP that certain human intrusion scenarios were inadequately examined. One of these scenarios was drilling with air. Drilling with air makes penetrating a highly pressurized repository much more hazardous. The weight of the drilling mud compensates for part or all of the high pressure in the repository; when drilling with air, there is no mud column to compensate for the pressure in the repository. Given high pressure in the repository, drilling with air exhumes more waste. A second scenario of concern was a leak in a reinjection brine well that created an extended hydraulic fracture. Such hydraulic fractures had occurred, associated with water flooding for oil recovery within the New Mexico portion of the Delaware Basin. A hydraulic fracture into the repository could create a fluid short circuit and potentially move large volumes of brine through the repository, leaching and transporting hazardous radionuclides. The U.S. Department of Energy and EPA viewed these scenarios as low probability events.

Many of the individuals concerned with the safety of WIPP believed that the human intrusion scenario dictated by EPA was unlikely. Because human intrusion is the most probable foreseeable failure scenario, these individuals felt the repository was inherently safe.

Linking a cascade of models compounds the calibration problems associated with each component model. Many of the models used for performance assessment at WIPP were theoretical and poorly calibrated. Extending the time horizon to 10,000 years further compounds the difficulties. The hope is that the statistical sampling of the important parameters in the performance assessment model will provide a probabilistic range of future outcomes. If 95% or 99% of these outcomes are within acceptable limits, the repository is judged to be safe. This approach does not address the problem that we may have overlooked something important in our conceptual model of the system. *Repeated sampling of a large parameter set may compensate for the uncertainty in the parameter values for the models used in performance assessment, but it does not compensate for wrong or incomplete conceptual models.*

Probabilistic performance assessment raises the issue of precision versus accuracy. The probabilistic approach may give the illusion that the modeler has quantified the error associated with the model. However, if darts are thrown at the wrong target, the spread of darts does not provide an assessment of whether the right target was chosen.

History Matching in Nuclear Waste Facilities

As indicated previously, the time horizon for formally predicting doses from nuclear waste facilities is 1000 to 10,000 years, or longer. Some of the longer lasting radioactive isotopes will persist well beyond 10,000 years. Given that the time horizon of the predictions is 10,000 years or longer, there is no opportunity for the traditional history match followed by a more or less equal period of prediction. Given the current strategy, there is, at best, a set of experiments of limited duration to which the models can be calibrated.

Yucca Mountain

One of the principal tools for evaluating the suitability of Yucca Mountain as a repository will be performance assessment—performance assessment similar to that used at WIPP. There are additional complications at Yucca Mountain. The waste to be emplaced at Yucca Mountain will generate heat. At issue is whether the loading will be relatively dense, producing high temperatures within the host rock, temperatures above the boiling point of water, or whether the spent nuclear fuel will be distributed more widely, keeping the host rock temperatures below the boiling point of water. Many investigators have concluded that the higher temperatures greatly increase the uncertainty of how the ground water system within the mountain will respond, and are to be avoided.

At both WIPP and Yucca Mountain, there were surprises once mining allowed scientists/engineers to actually visit the underground. At WIPP, the salt observed in the underground contained 1% to 3% interstitial brine. The original concept was that the only brine in the salt was in vesicles contained within the salt crystals—about 0.5%. Finding interstitial brine meant that the facility would be moist, a fact that was not included in the original conceptual model. The brine at WIPP did not preclude using the facility as a repository; it greatly complicated the analysis of the safety of the facility. It increased the uncertainty of the prediction of performance. In the end, WIPP was still judged to be safe by EPA, as well as by much of the scientific community (National Research Council 1996).

At Yucca Mountain, water containing chlorine–36 that was derived from atmospheric testing of nuclear weapons was found in the underground drift. The chlorine–36 indicates a *fast-path* for moisture movement in the mountain, a path that is unpredicted by the conventional theory of transport in the unsaturated zone, even a fractured unsaturated zone. The task at Yucca Mountain is to predict transport in a fractured, unsaturated media, subjected to a heat load for a prolonged period—10,000 years.

Performance assessment is dependent on having a correct conceptual model of transport within the mountain. At Yucca Mountain, the appropriate conceptual model for simulating unsaturated transport in the fractured tuffs at the site is unclear. A recent study by the National Research Council (2001) concluded that there is no consensus within the hydrogeologic community of what the appropriate conceptual model is to describe transport in a fractured, unsaturated zone. The generally accepted theory does not predict the chlorine–36 movement. This lack of a clear conceptual model greatly increases the uncertainty associated with performance assessment at Yucca Mountain. Without a consensus on the appropriate conceptual model, predictions of future system performance become highly questionable, at best.

Where Are We as a Community of Modelers?

Models are useful in integrating and synthesizing our knowledge about hydrogeologic systems in a way that allows us to make predictions about the future performance of the system. Most of us regard models as our best tools for the task. However, anyone engaged in this processes recognizes its inherent uncertainty. Modelers also recognize a pervasive element of professional judgment in creating models and judging their effectiveness. To some extent, these ideas are embedded in what we generally refer to as model calibration. Unfortunately, model calibration may or may not adequately test our conceptual model. Too often, an incomplete conceptual model can pass the test of being calibrated. Too often, the models have proven to be incomplete or wrong. As hydrogeologists, we make mistakes.

Oreskes et al. (1994) summarize the uncertainty in modeling; they state, " . . . the establishment that a model accurately represents the 'actual processes occurring in a real system' is not even a theoretical possibility."

Probabilistic performance assessment does not overcome the inherent uncertainty in modeling. Performance assessment is conducted in a probabilistic mode to compensate for the uncertainties in the parameters (and perhaps the boundary conditions). As suggested previously, uncertainties in what are the appropriate conceptual models are not compensated for by probabilistic sampling of the parameter sets of wrong or incomplete conceptual models.

Oreskes and Belitz (2001) regard the conceptual model as the most difficult problem in modeling; they state:

> "Conceptualization is probably the most thorny issue in modeling. It is the foundation of any model, and everyone knows that a faulty foundation will produce a faulty structure. . . . Yet what to do about it remains a problem. Much attention in model assessment has focused on quantification of error, but how do we quantify the error in a mistaken idea? . . . It is uncertainty rooted in the foundations of our knowledge, a function of our limited access to and understanding of the natural world. Almost by definition, conceptual error cannot be quantified. We don't know what we don't know, and we can't measure errors that we don't know we've made."

Iterative modeling in which we continue to monitor and revise the models to fit new data provides the best opportunity to avoid errors, including errors of conceptualization. However, iterative modeling, while it improves our odds for success, is not foolproof; models still have an inherent uncertainty.

Given the inherent uncertainty associated with models, Oreskes and Belitz (2001) ask the relevant question: Are predictions necessary for policy decisions? Uncertainty associated with model predictions may make alternative strategies or complementary courses of action more reasonable for society. We should examine the alternatives.

A Return to History Matching

The closer we can approach the idea embedded in the reservoir engineering concept of history matching, the more confidence we have in predictions. We would like the period of the history match to approach as nearly as possible the length of the prediction—our rule of thumb for confidence in prediction discussed previously.

Yucca Mountain could be a case in point. At Yucca Mountain, it seems that nuclear wastes could be emplaced in a retrievable mode within the repository for a long period. Our uncertainty in the models of the basic processes at Yucca Mountain argues strongly for a long period of observation.

The concept of Yucca Mountain could be changed to one of monitored retrievable storage for an indefinite period, perhaps 300 to 1000 years. A long period of monitoring of the facility could provide a history of performance to which the models could be repeatedly matched and improved. At the time that the models are demonstrated to reproduce the performance of the repository for a greatly extended period, society will be in a much stronger position to judge the suitability of the site as a permanent repository. I would urge that we rethink the nuclear repository at Yucca Mountain with the idea of keeping the repository open for observation for a prolonged and, for now, indefinite period.

The arguments for early closure of Yucca Mountain do not seem scientific, but rather political. Political considerations can often be changed by persuasive scientific arguments. Uncertainty associated with the predictions of the system behavior is a good reason not to be in a hurry to close the repository. Early closing of the repository may well be premature.

Acknowledgments

I wish to thank my colleagues Bill Alley, Mary Anderson, Ken Belitz, Lenny Konikow, and Naomi Oreskes for their insights in reviewing this paper.

References

Anderson, M.P., and W.W. Woessner. 1992. The role of the postaudit in model validation. *Advances in Water Resources* 15, 167–173.

Bethke, C. 1992. The question of uniqueness in geochemical modeling. *Geochimica et Cosmochimica Acta* 56, 4315–4320.

Bredehoeft, J.D., C.E. Neuzil, and P.C.D. Milly. 1983. Regional flow in the Dakota Aquifer: A study in the role of confining layers. U.S. Geological Survey Water-Supply Paper 2237.

Bredehoeft, J.D. 1997. *Air Drilling into WIPP.* Submitted to U.S. EPA Compliance Docket No. A–93–02—WIPP (prepared for New Mexico attorney general).

Bredehoeft, J.D. 1998. *Drilling with Mud and Air into WIPP—Revisited.* Submitted to U.S. EPA Compliance Docket No. A–93–02—WIPP (prepared for New Mexico attorney general).

Carslaw, H.S., and J.C. Jaeger. 1959. *Heat Conduction in Solids,* 2nd edition. Oxford, U.K.: Oxford University Press.

Ewing, R.C., M.S. Tierney, L.F. Konikow, and R.P. Rechard. 1999. Performance assessments of nuclear repositories: A dialogue on their value and limitations. *Risk Analysis* 19, 933–958.

Hantush, M.S. 1964. Hydraulics of wells. In *Advances in Hydroscience,* vol. 1, ed. V.T. Chow, 281–432. New York: Academic Press.

Jacob, C.E. 1940. On the flow of water in an elastic artesian aquifer. *Transactions of the American Geophysical Union* 2, 585–586.

Konikow, L.F., and J.D. Bredehoeft. 1992. Groundwater models cannot be validated. *Advances in Water Resources* 15, 75–83.

Konikow, L.F., and R.C. Ewing. 1999. Is a probabilistic performance assessment enough? *Ground Water* 37, no. 4: 481–482.

Kuhn, T.S. 1970. *The Structure of Scientific Revolution,* 2nd edition. Chicago: University of Chicago Press.

Menand, L. 2001. *The Metaphysical Club: A Story of Ideas in America.* New York: Farrar, Straus, and Giroux.

National Research Council. 1996. *The Waste Isolation Pilot Plant: A Potential Solution for the Disposal of Transuranic Waste.* Committee on the Waste Isolation Pilot Plant. Washington, D.C.: National Academy Press.

National Research Council. 2001. *Conceptual Models of Flow and Transport in the Fractured Vadose Zone.* Washington, D.C.: National Academy Press.

Oreskes, N., K. Shrader-Frechette, and K. Belitz. 1994. Verification, validation, and confirmation of numerical models in earth sciences. *Science* 263, 641–646.

Oreskes, N., and K. Belitz. 2001. Philosophical issues in model assessment. In *Model Validation: Perspectives in Hydrological Sciences,* ed. M.G. Anderson and P.D. Bates, 23–41. New York: John Wiley and Sons.

Popper, K.R. 1968. *The Logic of Scientific Discovery.* New York: Harper and Row.

Theis, C.V. 1935. The relation between lowering the piezometric surface and the rate and duration of discharge of a well using groundwater storage. In *Transactions of the American Geophysical Union,* 16th annual meeting, part 2; 519–524. Washington, D.C.: AGU.

Editor's Note: We invited Dr. Bredehoeft to contribute an issue paper on a topic of his choice to mark our 40th anniversary celebration this year of the publication of the first issue of *Ground Water.* Dr. Bredehoeft is a former editor-in-chief of the journal.