The conceptualization model problem—surprise

John Bredehoeft

Abstract The foundation of model analysis is the conceptual model. Surprise is defined as new data that renders the prevailing conceptual model invalid; as defined here it represents a paradigm shift. Limited empirical data indicate that surprises occur in 20–30% of model analyses. These data suggest that groundwater analysts have difficulty selecting the appropriate conceptual model. There is no ready remedy to the conceptual model problem other than (1) to collect as much data as is feasible, using all applicable methods—a complementary data collection methodology can lead to new information that changes the prevailing conceptual model, and (2) for the analyst to remain open to the fact that the conceptual model can change dramatically as more information is collected. In the final analysis, the hydrogeologist makes a subjective decision on the appropriate conceptual model. The conceptualization problem does not render models unusable. The problem introduces an uncertainty that often is not widely recognized. Conceptual model uncertainty is exacerbated in making long-term predictions of system performance.

Résumé C'est le modèle conceptuel qui se trouve à base d'une analyse sur un modèle. On considère comme une surprise lorsque le modèle est invalidé par des données nouvelles; dans les termes définis ici la surprise est équivalente à un change de paradigme. Des données empiriques limitées indiquent que les surprises apparaissent dans 20 à 30% des analyses effectuées sur les modèles. Ces données suggèrent que l'analyse des eaux souterraines présente des difficultés lorsqu'il s'agit de choisir le modèle conceptuel approprié. Il n'existe pas un autre remède au problème du modèle conceptuel que: (1) rassembler autant des données que possible en utilisant toutes les méthodes applicables—la méthode des données complémentaires peut conduire aux nouvelles informations qui vont changer le modèle conceptuel, et (2) l'analyste doit rester ouvert au fait que le modèle conceptuel peut bien changer lorsque des nouvelles informations apparaissent. Dans l'analyse finale le hydrogéologue prend une décision subjective sur le modèle conceptuel approprié. Le problème du le modèle conceptuel ne doit pas rendre le modèle inutilisable. Ce problème introduit une incertitude qui n'est pas toujours reconnue. Les incertitudes du modèle conceptuel deviennent plus importantes dans les cases de prévisions à long terme dans l'analyse de performance.

Resumen La base para hacer un análisis de un modelo es el modelo conceptual. Se define aquí la sorpresa como los datos nuevos que convierten en incoherente al modelo conceptual previamente aceptado; tal como se define aquí esto representa un cambio de paradigma. Los datos empíricos limitados indican que estas sorpresas suceden entre un 20 a un 30% de los análisis de modelos. Esto sugiere que los analistas de modelos de agua subterránea tienen dificultades al seleccionar el modelo conceptual apropiado. No hay otra solución disponible a este problema del modelo conceptual diferente de: (1) Recolectar tanta información como sea posible, mediante la utilización de todos los métodos aplicables, lo cual puede resultar en que esta nueva información ayude a cambiar el modelo conceptual vigente, y (2) Que el analista de modelos se mantenga siempre abierto al hecho de que un modelo conceptual puede cambiar de manera total, en la medida en que se colecte mas información. En el análisis final el hidrogeólogo toma una decisión subjetiva en cuanto al modelo conceptual apropiado. El problema de la conceptualización no produce modelos inútiles. El problema presenta una incertidumbre, la cual a menudo no es tenida en cuenta de manera adecuada. Esta incertidumbre en los modelos conceptuales se aumenta, cuando se hacen predicciones a largo plazo del comportamiento de un sistema dado.

Keywords Numerical modeling · Conceptual models · Groundwater mangement · Data collection and analysis · Mistaken model predictions

Received: 2 April 2004 / Accepted: 13 December 2004

J. Bredehoeft (✉)
The Hydrodynamics Group,
127 Toyon Lane, Sausalito, CA, 94965, USA
e-mail: jdbrede@aol.com
Tel.: +1-415-332-0666
Fax: +1-530-364-8541

DOI 10.1007/s10040-004-0430-5

Introduction

Models today have become a ubiquitous part of science. Uncertainty is a fact with groundwater models. Uncertainty arises in part because there is always an incomplete sample of the subsurface environment in which groundwater flows. A number of methods have been introduced to minimize the uncertainty in the parameter set that the models require. Many of the widely used modeling codes and their associated parameter estimation algorithms, such as MODFLOW-2000 (Hill et al. 2000), are designed to obtain an optimum set of parameters—a set of parameters that minimizes the difference between the model predictions and the observations. Using these methods the analyst can estimate the uncertainty that the parameters introduce into model predictions. However, these methods say very little about the underlying uncertainty associated with the selection of the conceptual model; the focus in this paper is uncertainty associated with the conceptual model.

Many thoughtful people consider conceptualization one of the thorniest problems in modeling (if not science). Every model has as its foundation a *conceptual model*. The conceptual model is the basic idea, or construct, of how the system or process operates; it forms the basic idea for the model (or theory). Zheng and Bennet (1995) provide a good description of the conceptual model:

> A conceptual model contains numerous qualitative and subjective interpretations. The appropriateness of the conceptual model cannot be tested until a numerical model is built and comparisons between field observations and model simulation results are made. Thus one of the most useful things about a numerical model is that it provides a tool to test and improve the conceptual model of a field site. It also provides a guide to future data collection, particularly in those cases where additional data are needed in order to produce a conceptual model consistent with field observations. For this reason, one should not wait until a 'perfect' conceptual model is formulated before starting to assemble the numerical model. Instead, conceptual and numerical modeling should be viewed as an iterative process in which the conceptual model is continuously reformulated and updated.

A number of ideas flow from this description:

1. The conceptual model is based on the subjective judgment of the analyst. Often one loses sight of this fact.
2. A numerical model provides a tool by which to test the appropriateness of the prevailing concept.
3. One can expect the conceptual model to be continuously updated as new information is acquired.

Conventional wisdom (or conventional dogma) dictates one's conceptual model—more on this below. The point is that the conceptual model results from decisions made by the analyst, either actively or passively.

In a *Ground Water* issue paper, entitled *From Models to Performance Assessment—The Conceptualization Problem* this author drew the following conclusions (Bredehoeft 2003):

1. Modelers tend to regard their conceptual models as immutable.
2. Time and again errors in prediction revolve around a poor choice of the conceptual model.
3. More often than not, data will fit more than one conceptual model equally well.
4. Good calibration of a model does not ensure a correct conceptual model.
5. Probabilistic sampling of the parameter sets does not compensate for uncertainties in what are the appropriate conceptual models, or for wrong or incomplete models.

Several of these points are illustrated in the discussion of specific models below.

This paper stemmed from a discussion with Shlomo Neuman concerning the *Ground Water* paper referred to above (Bredehoeft 2003); the particular concern in this paper is the role of surprise in conceptualization. Unfortunately, there is very little empirical data to bolster this author's argument for surprise. For the most part the argument relies upon data from the author's experience as both a scientist and a groundwater consultant. That experience involves a limited data set—usually too limited. What little empirical data there are from others tend to reinforce the author's experience.

In this paper the author argues that *true scientific surprise happens sufficiently frequently that it makes model predictions subject to additional significant errors*. First one needs a definition of surprise before continuing the argument.

Surprise

In the context of this paper *surprise* is defined as the collection of new information that renders one's original conceptual model invalid. The above quotation from Zheng and Bennett (1995) established the idea that the conceptual model is expected to change as new information is obtained. Changes in conceptual models range across a spectrum from modest changes in parameter values at one end of the spectrum to a complete paradigm shift in what was the prevailing conceptual model at the opposite extreme. Surprise, as this author defines it, entails a complete paradigm shift; the original concept needs to be rejected, or completely revised and a new model introduced to explain the new information. Since there is a spectrum of possible changes it may be difficult to define whether the change in a conceptual model is within the usual expectation or whether a surprise (a complete paradigm shift) occurred.

DOI 10.1007/s10040-004-0430-5

There are numerous examples of surprise in the history of science. One can illustrate by discussing the surprise that revolutionized geology in the later part of the 20th century—plate tectonics.

Moving continents was a conceptual idea introduced by the geophysicist Alfred Weggener in 1912. Most earth scientists, especially in United States, were collectively adamant that this was a far-fetched notion that could not be true. Something happened in the late 1950s and 1960s that caused earth scientists to take another look at continental movement—the earth science community acquired new data.

A group of geophysicists collected magnetometer data from the ocean floor first in the North Pacific off Cape Mendocino on the west coast of North America. Remnant magnetic strips were discovered in the bedrock beneath the ocean floor (Mason 1958). These were explained as a result of sea floor spreading (Vine and Mathews 1963). Quickly the geologic community had to revisit continental drift—plate tectonics was proposed as a conceptual model. Today, with Global Positioning Satellites (GPS) technology, we can now measure the relative motion of the plates in real time—plate motions are real.

Every geologist knows this story—scientists tend to think of these surprises as great exceptions. Plate tectonics is the surprise that revolutionized the study of the Earth in the 20th century. In groundwater hydrology the Theis equation revolutionized the science in the 1930s.

Prevailing paradigms

What one chooses as a conceptual model is a function of the status of knowledge in science. As a young geologist one of the author's mentors, N.W. Bass, remarked—"*a geologic report is always a progress report*"—this author continues to reflect on this remark. For example, plate tectonics changed geology and changed our conceptual model of tectonics. Theis (1935) and Jacob (1940) changed groundwater hydrology by introducing the transient theory of groundwater flow.

One can take for example the Theis/Jacob equation. Before Theis/Jacob the prevailing idea regarding groundwater flow was that the flow was incompressible as reflected by the Thiem equation. Following Theis/Jacob hydrogeologists analyze groundwater flow as compressible. One can go on with these examples; the point is that the prevailing science conditions the ideas of scientists—and their conceptual models.

As the prevailing paradigms change so will the conceptual models. The fact that science changes over time probably makes very little difference if one restricts their model predictions to the near term. However, as scientists begin to predict for long periods into the future the fact that science itself is dynamic, and changes over time, lends further uncertainty to long-term predictions.

Plate tectonics and the Theis equation are big issues that changed the conceptual models that geologists and groundwater professionals use to view the earth. The relevant question is—*are there similar surprises, that are not so global, yet render one's initial conceptual model invalid?* In the author's experience there are.

One can distinguish two kinds of paradigm shifts that lead to surprise:

1. Surprise that flows from a revision in scientific theory.
2. Surprise that flows from new information about a particular site.

The following examples illustrate surprise that flows from new site information.

Examples of hydrogeologic surprise

The author began modeling in the 1960s; much of his research career at the U.S. Geological Survey (USGS) was spent in exploring conceptual models that could explain fluid phenomena observed in the subsurface. He left the USGS at the end of 1994; since then he has been consulting—going on 10 years. The author's consulting experience is most relevant here. While consulting he worked on 21 major investigations that involved models. These were a varied spectrum of studies—water supply, contaminant transport, Superfund Sites (designated by the U.S. Environmental Protection Agency—EPA), nuclear waste, etc. In these 21 studies there have been surprises four to six times—some instances are not so clear. Examples of the surprises encountered by the author are discussed below.

Before proceeding it is of interest to further clarify the definition of surprise. A surprise is new information that causes a rethinking of the conceptual model. It is not simply a revision in a parameter set, or a new calibration of the model.

WIPP (The Waste Isolation Pilot Plant), New Mexico, USA

The original concept for the Waste Isolation Pilot Plant (WIPP) was that the mine for radioactive waste disposal in the Permian, Salado Salt would be dry. This was the view of a National Academy of Science Panel that convened in the 1950s and suggested salt as the preferred medium for a nuclear repository—the panel included such distinguished hydrogeologists as C.V. Theis and King Hubbert. At the WIPP site it was originally recognized that the salt contained 0.5% brine in vesicles within the salt crystals, but it was thought that the brine within the vesicles was unlikely to migrate. The original concept was that a mine in salt would be dry. Once an exploratory mine was opened at WIPP, it was found that the salt contained 1–3% brine in interstices between the salt crystals, and this brine would migrate into the mine. Measurements made during heater experiments, before the heaters were turned on, demonstrated conclusively that brine was flowing to the mine. Later a sealed room experiment proved even more conclusively that brine was moving into the mine. Under normal operations the ventilation system removed the moisture.

Thus the salt mine that was originally conceived as dry now became damp. Moisture generates gas within the repository. The original concept had to be totally revised.

The change in concept did not mean WIPP was unviable as a repository; however, it did require a totally revised conceptual model and a revised analysis.

Yucca Mountain, Nevada, USA

Yucca Mountain had a similar surprise when data became available from the exploratory underground drift. Chlorine 36 was found that was generated from nuclear bomb testing. This suggests that there is a so-called *fast path* for moisture movement through the mountain that is unaccounted for by the accepted theory for water transport in an unsaturated media. The theory for flow through the unsaturated zone has been modified to uncouple flow through major fractures and faults from the matrix. The jury is still out on Yucca Mountain; finding chlorine 36 at depth does not by itself make Yucca Mountain a bad site for a nuclear repository. However, it caused a reevaluation of what is the appropriate conceptual model to describe moisture movement in the mountain.

Both the surprises at WIPP and Yucca Mountain came when scientists/engineers mined into the subsurface—in both cases mining provided better geologic access and new information. The two nuclear repositories in the United States had, what the author would term, surprises that rendered the original conceptual models unviable—two for two. In both instances there were large scientific teams that initially proposed conceptual models that later surprises showed were invalid.

LANL (Los Alamos National Laboratory), New Mexico, USA

Radioactive wastes are disposed of at Los Alamos National Laboratory (LANL). Most of the wastes were placed in excavations on the top of the dry mesas. The moisture movement beneath the mesas has been extensively studied. The moisture movement, as reflected by the moisture profile data, is enigmatic, at best. The uncertainty in the unsaturated zone flow is captured by this quotation from the performance assessment document (Los Alamos National Laboratory 1997):

> ...It is critical to distinguish these possibilities and determine if the apparent source term is physical or not, because the region may or may not provide a moisture source driving flux at greater depths throughout the unsaturated zone beneath the mesa. This controls the overall unsaturated transit time to the ground water and thus the impact on the site Performance Assessment depends critically upon resolving the hydrologic flows in this region (the unsaturated zone). ...

The canyons rather than the mesas are where most of the groundwater recharge occurs at Los Alamos. Even though the flow through the unsaturated zone on the mesas is not understood, it may not cause a significant problem for waste disposal at Los Alamos.

Los Angeles, California, USA

The author engaged in a consulting project in the Los Angeles area where the concern was the migration of contaminants. A flow model was created, and calibrated to the available head data in the usual manner; unfortunately the head data were not uniformly distributed throughout the area of interest. In the investigation flow vectors were mapped that were 90° different from those predicted by the accepted model for the area. These vectors could not be explained by a simple “tweaking” of the existing model; the vectors were reconciled with a different conceptual model.

In this instance the flow model was very large, involving approximately ten layers and more than several hundred thousand nodes. Even with a preprocessor it was time consuming to extensively revise the model. There was reluctance on the part of the initial modeling team to recast the model with a new conceptual design. The initial model had gained a momentum of its own; even though it was shown to be seriously flawed.

Santa Barbara, California, USA

The author investigated a contaminated site in Santa Barbara. Several models were created of the site. There was ambiguity on how a recent fault west of the site impacted groundwater flow. The author’s model analysis indicated that the fault was permeable, contrary to others that conceived the fault to be impermeable. Making the fault permeable changed the pattern of groundwater flow.

Summitville, Colorado, USA

Summitville is a Superfund site, a mine in Colorado where gold was recovered in a heap leach operation. The U.S. Environmental Protection Agency and the State of Colorado reclaimed the site after the mining company went bankrupt. The reclamation was based on a model analysis, done by the mining company before going bankrupt, that indicated that by plugging an adit, that drained the mine workings, the water table would rise to an elevation that flooded several open pits in the area of the mining. The hypothesis of the reclamation plan was that tailings placed in the open pits would become saturated with groundwater that in turn would greatly retard the formation of acid mine drainage from the tailings. Geological mapping of the area had clearly identified a host of natural seeps in the vicinity of the mines. The mining company modeling totally overlooked the natural seeps that were drained by the adit, but would be reactivated once the adit was plugged. The natural seeps would drain groundwater keeping water levels below the open pit.

Based on the modeling approximately $100 million was spent to move tailings back into the open pits. The adit was plugged, but the groundwater levels did not rise to fill the pits; the old seeps were reactivated all over the mountain. Without groundwater saturating the tailings acid mine waters are generated more or less unimpeded.

At some point society must take action even with incomplete information. Summitville was an instance where the agencies responsible for reclaiming the site were anxious to act; they wanted to demonstrate that they were proactive. There was abundant information to suggest that the remedy selected would not alleviate the acid mine drainage. In the rush to action, the cautionary information, as well as carefully designed experiments to test the conceptual design were bypassed. The Superfund program enables the reclaiming agencies to recover their costs; however, the U.S. Federal Courts have ruled that where the remedy implemented is not effective the party causing the contamination is not obligated for the expense. In this case the government recovered only a token amount of the cost of reclamation. Acid mine drainage continues from the site.

Other investigations

In the author's consulting he has been involved with 15 other model investigations. There is no indication in the other investigations that the conceptual models were sufficiently flawed that a surprise occurred.

Post audits

The author's limited set of investigations represents a small sample of the world of groundwater models. It is of interest to increase the sample size. Konikow (1986) and Anderson and Woessner (1992) performed post-audits of model predictions. They attempted to assess the question—*did the model do a good job of predicting the future response of the groundwater system?* The focus here is different; the author is asking were there surprises in these model studies. Even so, their data is of interest. In some instances what they suggest as a weakness in the conceptual model seems to be an incomplete parameter distribution.

Groundwater hydrologists have been modeling using analog models since the 1960s and digital groundwater models since the 1970s. There should now be a number of models that could be evaluated by a post-audit. Unfortunately, while many model predictions were made, in many instances the development scenarios analyzed were very different from the actual development that followed the model analysis. Society acted differently than the scenarios that were analyzed.

One can read two messages from the fact that society's behavior was different than the scenarios analyzed. On the one hand, the results of the model scenarios may have induced a more enlightened behavior. For example, the community would reduce its reliance on groundwater as a result of the analysis; this was done at El Paso, Texas, USA. On the other hand, society may be viewed as much more fickle, behaving in a different manner than what was anticipated in the model studies. In either case the number of investigations in which the model scenarios anticipated actual development is small; as a consequence the number of post-audits of models is also small.

Arkansas River Valley, Colorado, USA

Konikow and Bredehoeft (1974) used a flow and transport model to analyze the build up of salt in the alluvial aquifer associated with the Arkansas River. Their model predicted a long-term salt accumulation in the aquifer; the early model results indicated that salt would continue to accumulate in the aquifer. Konikow and Person (1985) used 1982 data to evaluate the predictions made in 1974. The patterns of salt concentration were similar to those predicted; however, there was no observed long-term accumulation of salt. The salt concentration was stable over time with only annual fluctuations. The 1974 model was based upon only one year of data for calibration. Konikow and Person (1985) concluded that one year was too short a period of calibration to resolve the issue of a long-term trend in the salt accumulation.

Blue River Basin, Nebraska, USA

Alley and Emery (1986) did a post audit of an analog model that was built in 1965 for the Blue River Basin, Nebraska. The model predicted declines in both groundwater levels and streamflow as a result of pumping groundwater for irrigation. The post audit showed that the model over-predicted the decline in groundwater levels and under-predicted the decline in streamflow. It appears that the parameters of the model were poorly estimated; the storativity was too low. More water was induced to flow from the stream than the model predicted. Whether the interconnection with the stream was poorly defined conceptually is unclear.

Coachella Valley, California, USA

Konikow and Swain (1990) reviewed a model analysis that Swain (1978) created of both groundwater flow and transport for the Coachella Valley, California. The model was originally calibrated to 40 years of data. The model was used to predict water levels seven years into the future. The post audit showed large errors in the predicted water levels. There were very significant recharge events from tributary creeks in the area that were totally unanticipated in the initial conceptualization, even though there was 40 years of prior data with which to calibrate. The wet year recharge events were unexpected; the unanticipated wet year recharge events were a surprise.

Houston, Texas, USA

The modeling at Houston is somewhat different; Jorgensen (1981) reviewed the history of modeling the area where a series of three models was constructed to model groundwater conditions. The U.S. Geological Survey (USGS) created an electric analog model in the early 1960s. This model predicted drawdowns in the Houston area, but did not include in the conceptual model the land subsidence caused by the groundwater pumping. In 1975 a second analog model was constructed by the USGS that included the compression of the clay confining layers that creates land subsidence. In the late 1970s the U.S. Geological Survey created a digital aquifer model that described the regional area that included Houston. The conceptual model for the digital model was much the same as the 1975 analog model. The Houston modeling is an instance of iterative modeling in which each new model builds on the previous effort: each model is an improvement. The later models were improved conceptually by including the land subsidence.

HYDROCOIN—Salt Dome (hypothetical site)
The HYDROCOIN project was designed to test several codes on well-defined test cases. Unfortunately, most of the test cases did not have analytical solutions so that there was no well-defined correct answer; one could only see if the various models produced similar answers. A 'salt dome' test case was set up to simulate variable density groundwater flow over the top of a salt dome in which salt was dissolving into the groundwater. The boundary condition specified at the top of the salt was a constant concentration boundary. Six groups analyzed the problem using six different models. Konikow et al. (1997) showed that all six of these groups had not created the constant concentration boundary condition specified in the problem—all six had gotten it wrong.

INEL (Idaho National Engineering Laboratory), Idaho, USA
Robertson (1974) created a flow and transport model for the INEL site. Using his model he predicted the movement of radioactive constituents; tritium was the most mobile of the constituents. In 1980 eight wells were drilled along the southern boundary of INEL in an effort to detect the tritium plume. Lewis and Goldstein (1982) reported on the data from the eight new wells, and evaluated Robertson's earlier prediction. Robertson predicted a longer plume with less transverse dispersion than what was observed in 1980. The problem in Robertson's prediction relates to values used for transverse and longitudinal dispersivity. The transverse dispersion at the INEL site is usually large; it as large as the longitudinal dispersion.

Phoenix, Arizona, USA
An electric analog flow model was built of the Phoenix area by the USGS in 1968 (Anderson 1968). The model was calibrated with 40 years of data—1923–1964. Predictions were made using the model. Konikow (1986) did a post audit of the Phoenix model. The predictions of future groundwater levels proved to be much too pessimistic. Groundwater pumping was greatly reduced by the Central Arizona Project that brought surface water from the Colorado River to Phoenix. The post audit indicated that the Phoenix model contained bias in the parameter distribution. Konikow (1986) was able to demonstrate that this bias was not related to errors in the projected pumping. The technology is improved since the 1960s; there are new tools to estimate parameter distributions—MODFLOWP, UCODE, PEST. These methods should facilitate better identification of the parameters

Uranium Tailings, Ontario, Canada
Flavelle et al. (1991) used a flow and transport model to predict the movement of hydrogen ions (H^+) out of a uranium tailings pile. The model was calibrated to data taken in 1983 and 1984. The model was then used to predict concentrations for 1989. Data collected in 1989 showed that the inner portion of the plume was predicted with reasonable accuracy; however, the outer part of the plume was poorly predicted. This indicated that the single distribution coefficient used in the model was an inadequate representation of the ongoing chemical interactions associated with the plume—the conceptual model was inadequate.

Summary

Table 1 summarizes both the author's consulting experience and the post-audit results reported by others.

Table 1 A summary of surprises in modeling results

Prototype	Modeler	Model type	Surprise	Comments
Arkansas Valley	Konikow	Transport	No	Need longer period of calibration
Blue River	Emery	Flow	No	Need better parameters
Coachella Valley	Swain	Flow	Yes	Recharge events unanticipated
Houston	Jorgensen	Flow/ subsidence	?	Iterative modeling
HYDROCOIN	Konikow	Transport	Yes	Boundary condition modeled poorly
INEL	Robertson	Transport	No	Need better parameters
Los Alamos	Los Alamos	Unsaturated flow	?	Flow through unsaturated zone not understood
Los Angeles area	Bredehoeft	Flow	Yes	Flow vectors 90° off in model
Ontario U tailings	Flavelle	Transport	Yes	Need more than one distribution coefficient
Phoenix	Anderson	Flow	No	Need better parameters
Summitville	Bredehoeft	Flow	Yes	Seeps on mountain unaccounted for
Santa Barbara	Bredehoeft	Transport	?	Fault zone flow unaccounted for
WIPP	WIPP project	Flow	Yes	Salt had 1–3% interstitial brine
Yucca Mountain	YM project	Unsaturated flow	Yes	Chlorine 36 indicates fast flow path
Other models	15	Flow/ transport	No	Bredehoeft's consulting—no conceptual problems
Total	29		7 yes (3 ?)	

Discussion

This review points out that surprises occur in the model investigations reviewed with reasonable frequency—20–30% of the time. This indicates that it is not an easy matter to conceive of the appropriate conceptual model. The relevant question then is—*how often is the chosen conceptual model inadequate, and yet new information is not uncovered to invalidate the prevailing conceptual model?* Can one extrapolate from the data to suggest that this situation occurs 20–30% of the time, even though a surprise is not discovered? The available post-audit data suggests that groundwater analysts make mistakes in constructing their conceptual models reasonably frequently.

Can hydrogeologists solve the conceptualization problem?

A number of authors have considered the conceptual model problem. Most investigators propose a strategy for constructing alternative conceptual-mathematical models of subsurface flow and transport, and then selecting the best among them, or alternatively rejecting those models that are deemed to provide an inadequate fit to the observations. Among the studies are: Carrera and Neuman (1986), Sun et al. (1995), Neuman (2003), Neuman and Wierenga (2003), and Tsai et al. (2003a, b). The selection strategy is based upon the analyst proposing a set of alternative conceptual models. Neuman and Wierenga are careful not to suggest that the set of conceptual models selected is a complete set of all possible models. Considering a set of possible conceptual models will undoubtedly improve model selection.

While the idea of formally considering more than one conceptual model is good, this author has never seen the idea put into practice. In his experience, analysts have selected what they judge to be an appropriate, single conceptual model. The parameters of the model are often changed as new data is acquired; however, there is a reluctance to abandon the conceptual model unless new data shows that the original concept can no longer fit the data. At WIPP, for example, there was a reluctance to consider that brine could exist between the salt crystals and would migrate into the mine until the evidence was irrefutable.

Clifford Voss (USGS) reflects the view of other analysts who would like to go further with the idea of a set of possible conceptual models; Voss stated (personal communication):

> In the end, we may not know which is correct, or which is 'best', but as imaginative hydrogeologists, ... we can imagine a full range of possibilities.

The Voss strategy presupposes that one can envision the entire set of possibilities. If anything, the history of science teaches humility; science is replete with true surprises. For example, no one conceived that the Salado Salt at WIPP contained 3% brine in the interstices between the salt crystals before the exploratory mine was opened. Similarly, no one conceived of a fast path for flow at Yucca Mountain before Chlorine 36 was found in the exploratory mine drift. These were surprises, totally unanticipated by large scientific teams.

Surprise is a part of science. The data indicate that in many cases hydrogeologists were not sufficiently informed to imagine what is the entire set of possible conceptual models. Furthermore, the scientific community is constrained in its selection of conceptual models by the prevailing scientific dogma—dogma that changes with time. Surprise is here to stay; it adds to the uncertainty of the scientific endeavor. It adds to the uncertainty of model predictions; it is an additional uncertainty that the hydrogeologic community lives with.

Shlomo Neuman also recognizes that surprises occur; in correspondence with the author he states:

> Yet no matter how large the supporting database may be, there always is a possibility that new observations and experimental data become available which the existing theory (or model) can neither reproduce nor explain ...

Given the fact that such surprises occur, it is presumptuous to imagine that one can envision the entire suite of possible conceptual models in complex hydrogeologic settings. Given surprise, the quotation from Oreskes and Belitz (2001) is even more appropriate:

> We don't know what we don't know, and we can't measure errors that we don't know we've made.

As suggested above, considering a set of conceptual models will aid in solving the problem; however, as long as one admits that surprises can occur the conceptual model problem is not solved—the problem is here to stay.

Do hydrogeologists continue to model?

Of course—models are useful in integrating and synthesizing our knowledge about hydrogeologic systems in a way that allows us to both (1) gain insights into how the systems function, and (2) make predictions about future performance. Most hydrogeologists, including this author, regard models as our best tools for the task. However, anyone engaged in this process must recognize its inherent uncertainty. The conceptual model problem makes model predictions inherently more uncertain. The fact that surprise occurs should give the community more pause—pause that reinforces the notion of uncertainty.

One can identify three broad categories of model use: (1) conceptual model synthesis, (2) short-term predictions for management, and (3) long-term predictions.

Conceptual model synthesis

One uses a model to synthesize mathematically the conceptual model of a site, and to test the concept against the data. This is extremely useful and provides a means to screen alternative concepts. This use of the model is often most useful in rejecting concepts that are infeasible, or most unlikely. Zheng and Bennett (1995) in their definition of a conceptual model, quoted above, expressed this view of model use.

In many instances data are lacking on certain parameters, the model provides a means to estimate the numerical value of these parameters. One can then ask questions—are the parameters values reasonable, or do we need a new conceptual model? For example, often one does not have hydraulic conductivity values for confining layers because of the difficulties associated with acquiring such data. The model can be used to estimate confining layer conductivities.

Analyzing more than a single process at a given site provides different information; often this information provides new insights. For example, analyzing both flow and transport at a site provides more information than flow alone. Transport introduces quantitative estimates of porosity and dispersivity that can provide better insights into the actual mechanism of groundwater flow at the site.

Using models for conceptual model synthesis is most appropriate; they provide the analyst with improved professional judgment. In the end, this may be the most important use of models, more import than future predictions.

What if—model predictions in the near future

Models are useful in making predictions on how a groundwater system will behave if one takes certain actions. For example, how will the system respond if we put a new well field at a particular location? Models can be used to analyze management options.

Petroleum engineers have perfected the art of short-term reservoir predictions. They look at the models from a pragmatic perspective. They create the mathematical model from their best understanding of the prevailing theory. They then apply the model to a particular petroleum reservoir. They adjust the parameters to match an observed history of reservoir performance—they call this match a "history match" (rather than model calibration). They then use the history-matched model to make a prediction of reservoir performance. However, they have caveats regarding their predictions. The rule of thumb is not to rely on the predictions much beyond a period equal to the period of history match. In other words, if one matches a 10-year reservoir history the engineer has some confidence in making a 10-year prediction. Beyond the 10-year prediction the engineer questions his confidence in the prediction.

Petroleum reservoir engineers avoid making claims that they have the correct conceptual model. They say simply we did the best we could to create what we think is an appropriate model of the reservoir. We will use this model to make a prediction of performance in which we have confidence, for a period equal to our history match. These rules of thumb could well be applied to groundwater analyses.

Groundwater models are especially useful in assessing the sustainability of a groundwater reservoir. Using the model one can estimate whether the system, given a particular development, will be able to sustain the stress indefinitely into the future, or will there be unwanted impacts. For example, can pumping from a particular aquifer be sustained indefinitely? This author argued that a model analysis is the best tool to answer this question (Bredehoeft 2002).

There are many other examples of management questions for which the model is most useful. For example, will pumping from an aquifer near the seacoast induce seawater intrusion? A corollary question—is there a better location for the pumping that will minimize, or control the intrusion? Another example, how best does one cleanup a contaminated aquifer?

Groundwater systems that are large and involve the water table are slow to respond to stress. Often it takes several hundred years for such systems to reach a new equilibrium state where there is no additional change in groundwater storage. Even so, the author includes these in this class of analysis.

The short-term predictive model is useful for making enlightened management decisions. The list of examples where models were used to address management questions is very extensive.

Long-term management decisions—long-term predictions

Hydrogeologists are now being asked to make long-term predictions of groundwater system performance, especially in association with the site selection of nuclear waste facilities. Groundwater models are being the basis for "Performance Assessment (PA)" in the site evaluation of nuclear waste facilities. Predictions of performance are being made to 1,000 and 10,000 years—sometimes longer. It is in these instances that the conceptual model problem becomes most daunting. There is no history for the system that comes anywhere close to the period that is being predicted—the petroleum engineer's rule of thumb cannot be applied.

Performance Assessment treats the uncertainty in the model parameters by running the model iteratively with parameters sampled from a probable range of possible values. The model predictions are examined statistically. If a large majority of the predictions fall within a range considered safe, then at least one criterion for a safe repository is satisfied. Performance Assessment does not test the adequacy of the conceptual model. The conceptual model may be all-important in making good long-term predictions of performance.

When predictions extend to 1,000 years, or longer, one can expect science itself to change. For example, the current transport theory, with its changing dispersivity with distance, is thought by many to be inadequate. One might expect a different transport theory to emerge in the next 1,000 years. This could change long-term predictions of transport.

Long-term model predictions are subject to the greatest error. One can expect great uncertainty in these predictions. Conceptual model problems play a large role in the uncertainty of these analyses. As suggested above, analyzing a set of potential conceptual models, while it may help, does not solve the problem.

Summary and conclusions

There are a host of applications where a groundwater model, used prudently, is quite useful. Perhaps foremost among these applications is the use of the model in synthesizing information, testing alternative conceptual models mathematically, and in providing the analyst an understanding of how the system functions—i.e., formulating professional judgment. Models are also useful in making management decisions. Usually, these entail predictions in the relatively short-term. Groundwater models involve much greater uncertainty when used to make long-term predictions. It is in the long-term predictions where conceptual model uncertainty adds the largest element of 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) summarized 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 above, 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 one continues 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 another 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 in the light of uncertainty associated with model predictions in an effort to find a truly robust solution. This is especially true where society is taking actions that have consequences far in the future.

At some point society finds it necessary to take action, action in the case of groundwater based upon some degree of incomplete information. (Hydrogeologists never have a complete description of the subsurface.) For societal actions that have far reaching consequences, it is prudent to seek truly robust solutions to the problem. For example, the U.S. Department of Energy is actively engaged in preparing a license application for a facility at Yucca Mountain to serve as a high-level nuclear waste repository for the country. Deep geologic disposal, as envisioned for Yucca Mountain, was proposed to sequester nuclear waste within the earth where it would be removed from the human environment. As originally conceived society, once the facility was closed, one could abandon the site without causing a large risk to humankind. The question arises—when does one 'close' the facility?

As currently conceived the facility at Yucca Mountain will dispose of nuclear fuel rods in a density that creates a high thermal load for the repository host rocks. Models of how the host rocks and their entrained moisture will behave under this high thermal loading push the envelope of the available models.

This author argues that because of the uncertainties discussed above, especially the conceptual model problem uncertainties, society would be best served by treating the early history of waste disposal at Yucca Mountain more as a scientific experiment. Extensive data could be collected during and following the emplacement of wastes in the repository. These data could be used to compare against the model predictions of performance. The models could be revised in the iterative manner described above, and new predictions made. Should problems arise, modifications in the repository operations might be called for. In an extreme case, some, or all the waste might need to be retrieved. Society may find that the waste itself is more valuable in the future, and should be retrieved and reprocessed. Under this scenario one would only close the facility once one was convinced that the model was adequately calibrated for a prolonged period of observation, and was satisfied with the future predictions of performance. There seems to be no strong scientific reason to rush to close the facility—there are good reasons not to rush.

Viewing long-term nuclear waste disposal as a scientific experiment requires rethinking, and a different mindset. It is the kind of rethinking that Oreskes and Belitz (2001) suggested when they asked the question:

...are predictions necessary for policy decisions?

Acknowledgements I wish to thank several of my colleagues for their thoughtful comments and review. In particular my thanks go to Jesus Carrera, Leonard Konikow, Shlomo Neuman, Naomi Oreskes, Robert Schneider, Clifford Voss, and Chunmiao Zheng, and for their suggestions and constructive criticism

References

Alley WM, Emery PA (1986) Groundwater model of the Blue River basin, Nebraska—twenty years later. J Hydrol 85:225–249

Anderson MP, Woessner WW (1992) The role of postaudit in model validation. Adv Water Resour 15:167–173

Anderson TW (1968) Electric analog analysis of ground-water depletion in central Arizona. U.S. Geological Survey Water-Supply Paper 1860, 21 pp

Bredehoeft JD (2002) The water budget myth revisited: why hydrogeologists model. Ground Water 40:340–345

Bredehoeft JD (2003) From models to performance assessment—the conceptualization problem. Ground Water 41:571–577

Carrera J, Neuman SP (1986) Estimation of aquifer parameters under steady-state and transient conditions. Water Resour Res 22:199–210

Flavelle P, Nguyen S, Napier W (1991) Lessons learned from model validation—a regulatory perspective: in GEOVAL-1990: symposium on validation of gepsphere flow and transport models. Organization for Economic Co-Operation (OECD), Nuclear Energy Agency, Paris, France, pp 441–448

Hill MC, Banta ER, Harbaugh AW, Anderman ER (2000) Modflow-2000, the U.S. Geological Survey modular ground-water model: user guide to observation, sensitivity, and parameter estimation processes, and three post-processing programs: U.S. Geological Survey Open-File Report 2000-184, 209 pp

Jacob CE (1940) On the flow of water in an elastic artesian aquifer. Transactions of American Geophysical Union, part 2, pp 585–586

Jorgensen DG (1981) Geohydrologic models of the Houston District, Texas. Ground Water 28:418–428

Konikow LF (1986) Predictive accuracy of a ground-water model—lessons from a postaudit. Ground Water 24:173–184

Konikow LF (1995) The value of postaudits in groundwater model applications. In: El-Kadi AI (ed) Groundwater models for resources analysis and management. Lewis Publishers—CRC, Boca Raton, FL, pp 59–78

Konikow LF, Bredehoeft JD (1974) Modeling flow and chemical-quality changes in an irrigated stream-aquifer system. Water Resour Res 10:546–562

Konikow LF, Person M (1985) Assessment of long-term salinity changes in an irrigated stream-aquifer system. Water Resour Res 21:1611–1624

Konikow LF, Sanford WE, Campbell PJ (1997) Constant concentration boundary condition: lessons from the HYDRCOIN variable-density groundwater benchmark problem. Water Resour Res 33:2253–2261

Konikow LF, Swain LA (1990) Assessment of predictive accuracy of a model of artificial recharge effects in the Coachella Valley, California. In: Simpson ES, Sharp JM Jr (eds) Selected papers on hydrogeology from the 28th Geological Congress. Verlag Heinz Heise, Washington, DC, USA, pp 433–449

Lewis BD, Goldstein FJ (1982) Evaluation of a predictive groundwater solute-transport model at the Idaho National Engineering Laboratory testing station, Idaho. U.S. Geological Survey Water Resources Investigation 82–85:71

Los Alamos National Laboratory (LANL) (1997) Performance Assessment and Composite Analysis for the Los Alamos Material Disposal Area G: Report-54G-013.R.2 (Appendix 2c—analysis of liquid phase transport in the unsaturated zone at a mesa top disposal facility, p 28)

Mason RG (1958) A magnetic survey off the west coast of North America between latitudes 36 and 38° N and 121 and 128° W. Geophys J 1:320-329

Neuman SP (2003) Maximum likelihood Bayesian averaging of alternative conceptual mathematical models. Stochastic Environ Res Risk Assess 17(5):291–305, doi: 10.1007/s00477-003-0151-7

Neuman SP, Wierenga PJ (2003) A comprehensive strategy of hydrogeologic modeling and uncertainty analysis for nuclear facilities and sites. NUREG/CR-6805, U.S. Nuclear Regulatory Commission, Washington, DC

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

Oreskes N, Belitz K (2001) Philosophical issues in model assessment. Chapter 3. In: Anderson MG, Bates PD (eds) Model validation: perspectives in hydrological sciences. Wiley, New York, pp 23–41

Robertson JB (1974) Digital modeling of radioactive and chemical waste transport in the Snake River Plain aquifer at the National reactor Testing Station, Idaho. U.S. Geological Survey Open-File Report. IDO—22054, 41 pp

Sun MZ, Jeng MC, Yeh WWG (1995) A proposed geological parameterization method of parameter identification in three dimensional groundwater modeling. Water Resour Res 31:89–102

Swain LA (1978) Predicted water-level and water-quality effects of artificial recharge in the upper Coachella Valley, California, using a finite-element digital model. U.S. Geological Survey Water-Resources Investigation 77–29:54

Theis CV (1935) The relation between lowering the piezometric surface and the rate and duration of discharge of a well using groundwater storage. Transactions American Geophysical Union, 16th Annual Meeting, part 2, pp 519–524

Tsai FTC, Sun NZ, Yeh WWG (2003a) Global-local optimization for parameter structure identification in three-dimensional groundwater modeling. Water Resour Res 39:1043

Tsai FTC, Sun NZ, Yeh WWG (2003b) A combinatorial optimization scheme for parameter structure identification in groundwater modeling. Ground Water 41:156–169

Vine FJ, Mathews DH (1963) Magnetic anomalies over oceanic ridges. Nature 199:947–949

Zheng C, Bennett G (1995) Applied contaminant transport modeling. Wiley, NewYork

DOI 10.1007/s10040-004-0430-5