Advances in Water Resources **15** (1992) 167–173

The role of the postaudit in model validation

Mary P. Anderson

Department of Geology and Geophysics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

&

William W. Woessner

Department of Geology, University of Montana, Missoula, Montana 59812, USA

Most researchers agree that validation is a demonstration that a model is capable of making accurate predictions at a site-specific field setting. A successful demonstration of validation requires completion of a series of steps that form a modeling protocol. These steps include model design and calibration, and verification of the governing equation, the computer code, and the model itself. The strictest form of validation is to demonstrate that the model can accurately predict the future. This type of validation test has been called a postaudit. Results of five postaudits suggest that it will be difficult and probably impossible to validate groundwater models by means of a postaudit because it is impossible to characterize the field setting in sufficient detail. Attention should instead be focused on good modeling protocol including providing a complete description of model design, a thorough assessment of model calibration, and an uncertainty analysis.

Key words: model validation, postaudit, modeling protocol, uncertainty analysis.

INTRODUCTION

A groundwater model is generally recognized to be the preferred tool for synthesizing the many factors involved in analyzing complex groundwater problems. But at the same time, results of groundwater models are often viewed with skepticism. In the last 5 years or so an 'enormous amount of skepticism appears to have developed, with a resulting attitude of 'Prove it!' having replaced the more passive and accepting faith of earlier years'.[1] Skepticism arises from an increasing concern over the validity of using models to make long-term predictions about the configuration of a flow system or concentrations of contaminants in groundwater. The absence of proof that models can make accurate long-term predictions has led to demands for what is usually termed model validation, where validation is usually understood to be a demonstration of accuracy.

A model consists of a governing equation and a set of boundary and initial conditions specific to a given field problem. The model is also understood to have associated ranges of site-specific parameter values. The code is a generic computer program that contains an algorithm capable of solving the mathematical model numerically. The modeling process consists of using the code to solve a site-specific field problem. Model validation takes place during the final steps in the modeling process when the accuracy of the model is tested. As such, the success of model validation depends on satisfactory completion of all the other steps in the modeling process. This sequence of steps forms a protocol for modeling. There have been numerous applications of groundwater models to field problems, but as yet there is no standard protocol to provide guidance during modeling or when reporting modeling results. Efforts to develop such a protocol are currently underway.[2,3] The modeling protocol advocated by Anderson and Woessner[26] is shown in Figure 1.

The concept of validation is sometimes confused with verification. Verification, like validation, refers to establishing accuracy. It can be used in reference to the governing equation, the code, or the model. All three types of verification are part of a modeling protocol. Verification of the governing equation consists of demonstrating that the equation used in the model accurately describes the processes of flow and/or transport in porous media, i.e. that the governing

Advances in Water Resources 0309-1708/92/$05.00

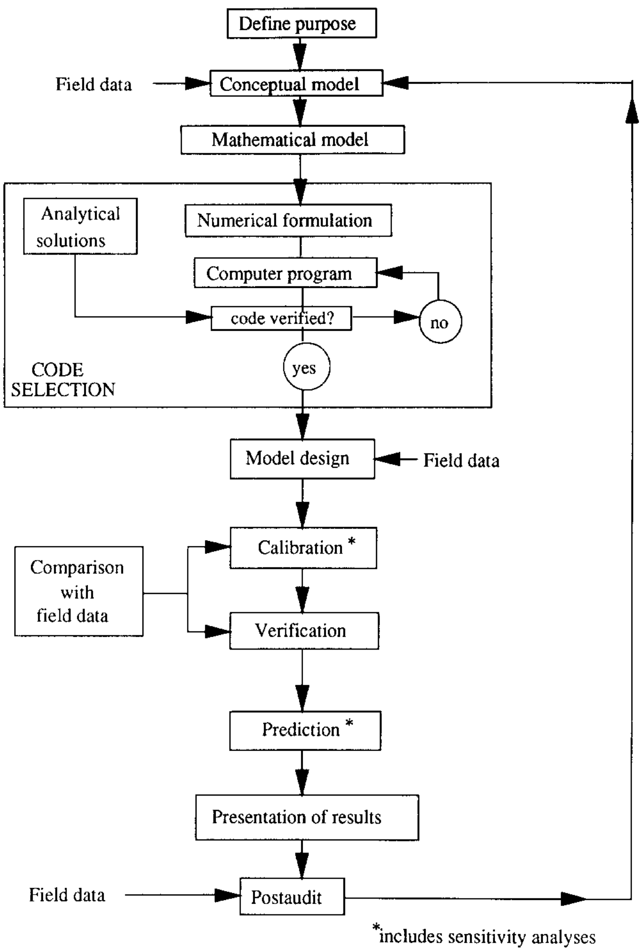

Fig. 1. A modeling protocol.[26]

equation is appropriate for the processes of interest. Verification of a governing equation may consist of a demonstration that a model based on the equation can reproduce and predict short-term results from laboratory and field experiments. Groundwater flow models are accepted as verified in this sense; it is commonly believed that Darcy's law and conservation of mass accurately describe groundwater flow at a macroscopic scale and that average or 'effective' parameters may be defined to characterize the porous medium at the scale of a representative elementary volume. Mass transport models, on the other hand, have not passed this type of generic verification test. Laboratory studies and calibration of models to field data demonstrate that the advection–dispersion equation, currently used in most solute transport models, does not always reproduce system behavior. Consequently, there is much debate over the appropriate way to quantify terms for dispersion and certain chemical reactions. If the governing equation used in a model has not passed a verification test, it must be recognized that modeling results may be incorrect.

Code verification involves comparison of the numerical solution generated by the model with one or more analytical solutions or with other numerical solutions. Verification of the code ensures that the computer program accurately solves the equations that constitute the mathematical model.

Model verification has been used synonymously with model validation, but we prefer to distinguish between these two concepts. We also treat model calibration separately from verification and validation, although all three are tests of model accuracy. During calibration a set of values for aquifer parameters and stresses is found that approximately reproduces field-measured heads and flows and/or concentrations. Calibration is done by trial-and-error adjustment of parameters or by using an automated parameter estimation code. The purpose of model verification is to establish greater confidence in the model by using the set of calibrated parameter values and stresses to reproduce a second set of field data. According to Konikow,[4] a model is verified 'if its accuracy and predictive capability have been proven to lie within acceptable limits or error by tests independent of the calibration data.' This step is also sometimes called historical data validation.[5] In a typical verification exercise, values of parameters and hydrologic stresses determined during calibration are used to simulate a transient response that has been measured in the field. Unfortunately it is often impossible to verify a model because only one set of field data is available. That data set, of course, is needed for calibration. If this is the case, the model cannot be verified. A calibrated but unverified model can still be used to make predictions as long as careful sensitivity analyses of both the calibrated model and the predictive model are performed and evaluated. Predictions resulting from calibrated but unverified models generally will be more uncertain than predictions derived from verified models.

Whereas model calibration and verification demonstrate that the model can mimic past behavior, model validation, as defined here, tests whether the model can predict the future. This type of validation test has been called a predictive validation[5] or a postaudit. In a postaudit, new field data are collected several years after the modeling study was completed to determine whether the prediction came true. If the model's prediction was accurate the model may be considered valid for that particular site for the conditions simulated. Several authors[1,6–10] stressed that claims of validation require qualifiers as to the conditions for which the model has been validated and those for which it should not be used. For example, Tsang[8] observed that it may not be reasonable to require the current generation of transport models to predict concentrations at a point in space at a given moment in time. However, it may be reasonable to expect the model to simulate average transport behavior within the problem domain. Hence, it is necessary to state the performance measures used as one of the qualifiers of the validation.

In this paper the role of the postaudit as a form of validation is considered. The postaudit should occur long enough after the prediction was made to ensure that there has been adequate time for significant changes to occur. A postaudit performed too soon after the initial calibration may lead to the conclusion that the prediction came close to estimating the observed values, when in fact not enough time elapsed to allow the system to move sufficiently far from the calibrated solution. In the literature, validation has been used mainly in reference to models for assessing the potential of contaminant movement from high level radioactive waste repositories.[11] This type of modeling requires prediction on the order of 10 000 or 100 000 years, whereas in most other engineering applications of groundwater models the time frame is of the order of tens of years. Because of the long-term nature of predictions required to assess the potential for contaminant movement from high level radioactive waste repositories, model validation by means of a postaudit will have limited utility in this context.

Postaudits have not been considered a routine part of the modeling process and in fact they may not be necessary if the purpose of the model is to analyze current steady-state behavior or to make short-term predictions. However, when the model is used to make predictions on the order of tens or hundreds of years, a postaudit is an important step in building the case that a model produces meaningful results.

The modeling process, including analysis of the postaudit, requires subjective judgements about the magnitude of acceptable error.[12] Errors include measurement error in the field data and modeling error represented by the differences between field and simulated values. While the magnitude and distribution of errors can be analyzed quantitatively, a subjective judgement is always required in deciding whether the errors are tolerable. Such judgements must be tied to the purpose of the modeling effort and based on hydrogeologic expertise and evidence.

LESSONS FROM POSTAUDITS

To date, five postaudits of modeling studies have been reported in the literature.[13–17] Three of these include postaudits of solute transport models. In all five cases the models did not accurately predict the future. Two conclusions emerge:

(1) Inaccurate predictions were partly caused by errors in the conceptual model of the hydrogeological system.
(2) Inaccurate predictions resulted from a failure to use appropriate values for assumed future stresses such as recharge, pumping and contaminant loading rates.

The more serious problem from the perspective of using a postaudit as a proof of validation is the first of these. It is unfortunate but true that there will always be errors in the conceptual model. Even a detailed site characterization can never eliminate uncertainties about parameter values, processes, and conditions at a site. While it is true that uncertainties involved in estimating future stresses are often large, these are less serious because the model could be rerun using accurate values for the stresses once they become known. Then if the conceptual model and the calibrated parameter values accurately represent system behavior, the validation would be successful.

Uncertainty about future stresses should be built into predictive simulations by means of a sensitivity analysis of the prediction (Fig. 1) in which several different scenarios are simulated using different assumed trends in the applied stresses in order to define a range in the predicted values. In the modeling work analyzed in the postaudits described below, and for most modeling done in the 1960s and 1970s, this step was not performed.

Summaries of postaudits

Postaudit results are reviewed by Anderson and Woessner[26] and are briefly summarized below.

(1) Konikow[13] performed a postaudit of a two-dimensional electric analog model of the Salt River and Lower Santa Cruz River Basins, Arizona, that was calibrated against 40 years of record (1923–64) and then used to predict water-level changes during the following 10 years (1965–74). During the postaudit, analysis of observed water-level changes in 77 wells during 1965–74 showed that the model consistently predicted lower water levels than actually occurred.

The errors in the prediction can be accounted for, in part, by the failure to use accurate future pumping stresses in the simulation. The modeler assumed that future pumping would continue at the 1964 rate when in fact pumping declined after 1965. During the postaudit, examination of the distribution of pumping and the predicted errors in water levels suggested that incorrect assumed pumping rates were partly responsible for the erroneous prediction but that other sources of error were also present. Because the analog model has been disassembled, it was not possible to run the model again to isolate the sources of error. However, it is likely that improvements in the conceptual model, e.g. including land subsidence and a three-dimensional representation of the system, would improve the model's predictive ability by better representing changes in aquifer storage and transmissivity.

(2) Alley and Emery[14] examined predictions of 1982 water-level declines and streamflow depletions for the Blue River Basin, Nebraska, made in 1965 with an electric analog model. Declines in water levels and streamflow were predicted to occur as a result of increases in pumpage for irrigation. The postaudit showed that the model overestimated the decline in

groundwater levels and underestimated the amount of streamflow depletion.

Net groundwater withdrawals in agricultural areas are difficult to estimate because it is usually necessary to infer net withdrawals from estimates of irrigated acreage, groundwater recharge, and consumptive use of irrigation water (irrigation efficiency). Analyses performed by Alley and Emery[14] suggested that net groundwater withdrawals used in the analog simulation were too low. The model overestimated groundwater level declines because it assumed that all of the net groundwater withdrawals would come from storage in the aquifer, when in fact some water comes from induced recharge from the stream. Furthermore, Alley and Emery[14] speculated that storage coefficients used in the model were too low. They concluded that: 'Considerable uncertainty about the basic conceptualization of the hydrology of the Blue River basin greatly limits the reliability of groundwater models developed for the basin.'

(3) Konikow and Bredehoeft[18] used an early version of the USGS Method of Characteristics model[19] to predict concentrations of dissolved solids in the aquifer adjacent to a portion of the Arkansas River in southeastern Colorado. Salinity is a problem in this area, owing to recycling of irrigation water. Konikow and Bredehoeft[18] calibrated the flow model to a transient flow field determined from data collected during the 1971–72 study. The solute transport model was also calibrated to data obtained during 1971–72. The model predicted that dissolved solids concentrations would increase steadily through 1982. Statistical evaluation of the historical data set including data collected in 1982, showed that dissolved solids had not increased in the aquifer above 1971–72 levels. This suggested that the aquifer is in dynamic equilibrium with respect to salinity.[15] If true, this would mean that current irrigation practices could be continued indefinitely without causing further groundwater salinity degradation.

Konikow and Person[15] showed that the error in the original prediction was due to calibration during a period of decreasing river discharge. During the 1971–72 calibration period, river discharge was declining after a record high in 1966. Concentration of dissolved solids in the river is inversely proportional to discharge; during 1971–72, river water recharging the aquifer was increasing in salinity. The model propagated this trend into the future. Statistical tests showed that this short-term trend, although statistically significant, was not representative of the long-term salinity trend. The postaudit showed that the flow portion of the model was adequately calibrated.

Person and Konikow[20] recalibrated the model using an improved regression equation to relate salinity to measured specific conductance. Data from 1971, 1972, and 1982, were used in calculating the new regression equation. Calibration of the model is sensitive to this relationship because about half of the irrigation water is diverted from the river. They also improved the conceptual model of the system by incorporating a lag time for solutes to travel through the unsaturated zone. The recalibrated model successfully simulated the observed long-term trend in salinity from 1971 to 1982. Finally, they used the recalibrated model to demonstrate that the system is in dynamic equilibrium with current irrigation practices.

Although the recalibrated model could accurately simulate the observed long-term trend in salinity, it should be noted that the recalibration used field data collected in 1982 in order to simulate 1982 conditions. Only 3 years of detailed data, including 1982, were available for the 1971–82 simulation. It is therefore relevant to ask whether the model could have predicted the long-term trend in the absence of the 1982 data. Through statistical analysis of temporal salinity trends using a 32-year record of estimated stream salinities, Person and Konikow[20] demonstrated that a 4-year sampling period was needed to calibrate the solute transport model to within 10% of the observed mean salinity, while 1 year's worth of data was sufficient to calibrate the flow model. The implication is that the long-term trend could have been predicted without the 1982 data, as long as a 4-year record of salinity trends was available for calibration. This finding suggests that evaluation of a conceptual model should include not only the spatial properties of the system, but also the hydrologic response time and temporal trends that characterize the system.

(4) Robertson[21] used a two-dimensional groundwater flow model to simulate flow in a basalt aquifer beneath the Idaho National Engineering Laboratory (INEL). Robertson[21] calibrated the model to an assumed steady-state flow field. He coupled the flow model to a solute transport model calibrated to the observed concentrations of chloride in groundwater in 1958 and 1969. Simulation of tritium and strontium-90 plumes were also simulated and compared with field data.

Robertson[21] then used the calibrated model to predict chloride and tritium concentrations in 1980. Lewis and Goldstein[16] performed a postaudit of those predictions and concluded that the contaminant plumes predicted by the simulation extended farther downgradient than the actual plumes because of conservative worst-case assumptions in the model input and inaccurate approximations of subsequent waste discharge and aquifer recharge conditions. The model assumed that waste disposal through a disposal well south of the river would continue at 1973 rates when disposal rates actually increased. The model also assumed that the Big Lost River would recharge the aquifer in odd-numbered years, when in fact there were high flows from 1974 to 1976, followed by 4 years of low flows when no recharge occurred.

Lewis and Goldstein[16] also pointed out that the conceptual model used by Robertson[21] was highly simplified. It was not unusual in the 1970s to use simplistic conceptual models in contaminant transport modeling, assuming two-dimensional, steady-state flow and a homogeneous and isotropic aquifer. We now know that these assumptions are usually inappropriate for simulations of complex contaminant plumes such as those at the INEL. However, recent modeling of this system by Goode and Konikow[22] demonstrated that the inclusion of transient effects in the model does not explain the anomalies in the Robertson simulation. Another possibility would be to use a fracture flow model to simulate the basalt aquifer. The conceptual model used by Robertson[21] viewed the aquifer as a continuous porous medium. It is likely that flow in this aquifer would be better approximated using a dual porosity model that included flow through the fractures as well as matrix diffusion.

(5) Flavelle *et al.*[17] simulated the release of hydrogen ions (H^+) from a tailings pile situated in glaciofluvial deposits in Ontario, Canada. The flow model was calibrated to heads observed in 1989 in the inner part of the plume where pH was less than 4·8. The solute transport model was calibrated by varying the distribution coefficient so that the velocity of contaminants in the inner part of the plume matched the observed positions of the plume in 1983 and 1984. The calibrated model was then used to predict the plume configuration in 1989.

Field measurements collected in 1989 showed that the model predicted pH in the inner portion of the plume reasonably well but not at the outer edges of the plume because the simulated velocities were too low. The values of distribution coefficient calibrated to the inner portion of the plume poorly represented conditions at the outer edge of the plume. It is not surprising that a single value for the distribution coefficient did not simulate the complex geochemistry occurring within the plume. The investigators concluded that even though their site is one of the most thoroughly studied uranium tailings sites in Canada, the data were not complete enough for a successful model validation.

Discussion

All of the postaudits indicate that errors in the predictions can be attributed at least partly to errors in the conceptual model. Model validation, therefore, requires a good conceptual model. Herein lies a major difficulty because a good conceptual model requires accurate and complete field characterization. Field characterization is always incomplete, thereby introducing uncertainty into the conceptual model. Continual improvement of the conceptual model requires periodic collection of field data and a trial and error process of model improvement over many years. It is rare to find such a large commitment of time and money to a modeling effort.

It is likely that there have been hundreds of predictive modeling studies performed since the 1960s. The fact that only five postaudits are reported in the literature suggests that at least in the USA, models are often used in a crisis mode rather than a management mode. In other words, a model is constructed to answer some pressing question so that a management decision can be made. After the model has served this purpose, it is 'shelved' and forgotten or discarded. Most models constructed in the USA are not used for management of the groundwater system on a day-to-day, month-to-month, or even year-to-year basis.

Ideally, models should be archived so that the model can be revived years later when a new modeling objective is defined or new field data become available. For example, Jorgensen[23] described a succession of three increasingly more sophisticated models used to predict drawdowns in the aquifer system underlying Houston, Texas, and the surrounding area. The first model was an electric analog model constructed in the early 1960s. The model accurately simulated observed water-level declines in and adjacent to the City of Houston, but did not reliably simulate drawdowns in outlying areas. The failure of the model was attributed to the lack of sufficient field data to formulate an adequate conceptual model of the system. Between 1965 and 1975, the conceptual model of the system was improved following the acquisition of new field data. A second analog model was constructed in 1975 using a four-layer representation of the system and including the effects of vertical leakage across clay units and the release of water from storage owing to compaction of clays. The simulated clay compaction was used to assess the ability of the model to predict land subsidence. Although the model accurately simulated drawdown, except near the boundaries, it did not accurately simulate the distribution of observed land subsidence. In the late 1970s, as part of a Regional Aquifer System Assessment (RASA) study, a five-layer finite difference model was used to simulate the Houston area once again. This model simulated a larger area than the 1975 analog model and thereby eliminated boundary effects that had been a problem with the analog simulation. The finite difference model used essentially the same conceptual model as the 1975 analog model but incorporated a revised distribution of clay layers and time-dependent storage coefficients for the clays. This model accurately simulated both drawdowns and land subsidence.

The example described by Jorgensen[23] is not a postaudit because a long-term prediction of the model was not evaluated. Rather, successive improvements in the conceptual model were made in an effort to achieve a better calibration to observed conditions. Jorgensen's example also illustrates the iterative way in which a model may be improved as new information is obtained.

Ideally, this is the way all models should develop. Improvements in the conceptual model will result in improved predictions.

SUMMARY AND DISCUSSION

Model validation carries the implication that site-specific models can make accurate predictions. If we require that a model accurately reproduce existing conditions and make accurate short-term predictions, it may be sufficient to follow the steps in the modeling protocol shown in Fig. 1 up to the postaudit.

If we require that a model make accurate long-term predictions in order to be considered valid, a postaudit is recommended. However, it is necessary to wait several years after a prediction is made before a postaudit can be performed. In applications to high level radioactive waste disposal, for example, it may not be practical to wait the length of time necessary before the prediction can be tested under the conditions for which validation is required.

Another difficulty is that a successful postaudit requires an accurate conceptual model of the site and accurate estimates of the magnitude and timing of future stresses. Defining an accurate conceptual model is an iterative process of continually up-dating and improving the field data base. Even with a large commitment of time and money, it is likely that the complete 'truth' about a site will never be known. Estimation of future stresses requires that the modeler foresee the future. Such forecasting necessarily introduces another large source of uncertainty into the model, as the results of the postaudits reviewed above demonstrate. For this reason, it is important to perform a series of predictive simulations to establish a range of probable outcomes. (Each of the postaudits reviewed above assessed the accuracy of only one predictive simulation.) Alternatively, the calibrated model could be rerun after the postaudit when the future stresses are known. If the calibrated model formulated prior to the postaudit accurately predicts field measurements when accurate stresses are input, the model can be considered validated for the conditions simulated.

The issue of validation is mainly a regulatory one, not a scientific one.[6] Tsang[5] pointed out that validation is not possible without a thorough understanding of the relevant physical and chemical processes and the system structure. Because our understanding of a system will always be incomplete a model can never be proven valid from a scientific standpoint. Hence, regulators must be content with some degree of partial validation, which requires detailed qualification of the conditions of validation. Such limited validation may be less than satisfying.

Given the difficulties of carrying out a successful validation and the low probability of success, it seems wise to seek an alternative to validation as a regulatory objective. Model validation is not a fruitful exercise because uncertainties in the conceptual model will always exist. Hence, uncertainty analysis should be built into the modeling strategy from the onset. For example, a modeling strategy involving uncertainty analysis coupled with probability and risk assessment was described by Freeze *et al.*[24] According to NRC[1] (p. 232): 'Such information is ultimately both more useful and more realistic than a certification that a model is or is not validated.' The regulatory focus should shift from demands for validation to demands for good modeling protocol, including providing a complete description of model design, a thorough assessment of model calibration, and an uncertainty analysis. Existing protocols for validation, e.g. the protocol proposed by the US Department of Energy (Voss[25]), should be replaced by protocols for performing and documenting the entire modeling process.

REFERENCES

1. NRC (National Research Council), *Ground Water Models: Scientific and Regulatory Applications*. National Academy Press, Washington, DC, 1990.
2. GRMR (Ground Water Monitoring Review), An update on ASTM Subcommittee D-18.21 on ground water and vadose zone investigations, *Ground Water Monitoring Review*, **X** (1990) 74–7.
3. State of California, Department of Health Services, Scientific and Technical Standards of Hazardous Waste Sites (Draft), Vol. 2: Exposure Assessment, 1990.
4. Konikow, L.F., Calibration of ground-water models. In *Verification of Mathematical and Physical Models in Hydraulic Engineering*, American Society of Civil Engineers, New York, 1978, pp. 87–93.
5. Tsang, C.-F., The modeling process and model validation. *Ground Water*, **29** (1991) 825–31.
6. Davis, P.A. & Goodrich, M.T., A proposed strategy for the validation of ground-water flow and solute transport models. In *GEOVAL-1990: Symposium on Validation of Geosphere Flow and Transport Models*. Organisation for Economic Co-Operation and Development (OECD), Nuclear Energy Agency, Paris, France, 1991, pp. 580–88.
7. Tsang, C.-F., Comments on model validation. *Transport in Porous Media*, **2** (1987) 623–9.
8. Tsang, C.-F., Tracer travel time and model validation. *Radioactive Waste Management and the Nuclear Fuel Cycle*, **13** (1989) 311–323.
9. Heijde, P.K.M. van der, El-Kadi, A.I. & Williams, S.A., Groundwater Modeling: An Overview and Status Report, EPA/600/2-89/028, US EPA, 1988.
10. Heijde, P.K.M. van der, Quality assurance in the development and application of groundwater models. In *Calibration and Reliability in Groundwater Modeling*, ed. K. Kovar, IAHS Publication No. 195, 1990, pp. 271–8.
11. *GEOVAL-1990: Symposium on Validation of Geosphere Flow and Transport Models*. Organisation for Economic Co-Operation and Development (OECD) Nuclear Energy Agency, Paris, France, 1991.
12. Woessner, W.W. & Anderson, M.P., Selecting calibration values and formulating calibration targets for ground

water flow simulations. In *Proceedings of: Solving Ground Water Problems with Models.* Association of Ground Water Scientists and Engineers, National Ground Water Association, Dallas, TX, 1992, pp. 199–212.
13. Konikow, L.F., Predictive accuracy of a ground-water model — lessons from a postaudit. *Ground Water*, **24** (1986) 173–84.
14. Alley, W.M. & Emery, P.A., Groundwater model of the Blue River Basin, Nebraska — twenty years later. *Journal of Hydrology*, **85** (1986) 225–50.
15. Konikow, L.F. & Person, M. Assessment of long-term salinity changes in an irrigated stream-aquifer system. *Water Resources Research*, **21** (1985) 1611–24.
16. Lewis, B.D. & Goldstein, F.S., Evaluation of a predictive ground-water solute-transport model at the Idaho National Engineering Laboratory, Idaho. Water Resources Investigations Report 82–85, US Geological Survey, 1982.
17. Flavelle, P., Nguyen, S. & Napier, W., Lessons learned from model validation — a regulatory perspective. In *GEOVAL-1990: Symposium on Validation of Geosphere Flow and Transport Models.* Organisation for Economic Co-Operation and Development (OECD), Nuclear Energy Agency, Paris, France, 1991, pp. 441–8.
18. Konikow, L.F. & Bredehoeft, J.D., Modeling flow and chemical quality changes in an irrigated stream-aquifer system. *Water Resources Research*, **10** (1974) 546–562.
19. Konikow, L.F. & Bredehoeft, J.D., Computer model of two-dimensional solute transport and dispersion in ground water. In *Techniques of Water-Resources Investigation*, Book 7, Chapter C2, US Geological Survey, 1978.
20. Person, M. & Konikow, L.F., Recalibration and predictive reliability of a solute transport model of an irrigated stream-aquifer system. *Journal of Hydrology*, **87** (1986) 145–65.
21. Robertson, J.B., Digital modeling of radioactive and chemical waste transport in the Snake River Plain aquifer at the National Reactor Testing Station, Idaho. Open File Report IDO-22054, US Geological Survey, 1974.
22. Goode, D.J. & Konikow, L.F., Reevaluation of large-scale dispersivities for a waste chloride plume: effects of transient flow. In *Calibration and Reliability in Groundwater Modeling*, ed. K. Kovar. IAHS Publ. 195, D. Reidel, Dorchrecht, 1990, pp. 417–26.
23. Jorgensen, D.G., Geohydrologic models of the Houston District, Texas. *Ground Water*, **19** (1981) 418–28.
24. Freeze, R.A., Massmann, J., Smith, L., Sperling, T. & James, B., Hydrogeological decision analysis: 1. A framework. *Ground Water*, **28** (1990) 738–66.
25. Voss, C.I., A proposed methodology for validating performance assessment models for the DOE Office of Civilian Radioactive Waste Management Program. In *High Level Radioactive Waste Management*, Vol. 1. American Society of Civil Engineers, New York, 1990, pp. 359-63.
26. Anderson, M.P. & Woessner, W.W., *Applied Groundwater Modeling: Simulation of Flow and Advective Transport.* Academic Press, New York, 1992.