
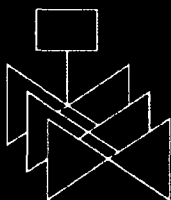
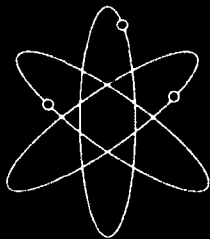
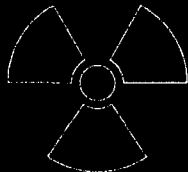
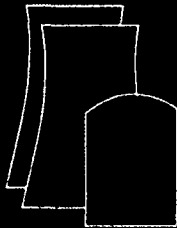


<b>United States Nuclear Regulatory Commission Official Hearing Exhibit</b>	
<b>In the Matter of:</b>	CROW BUTTE RESOURCES, INC. (License Renewal for the In Situ Leach Facility, Crawford, Nebraska)
	<b>ASLBP #:</b> 08-867-02-OLA-BD01
	<b>Docket #:</b> 04008943
	<b>Exhibit #:</b> NRC-094-00-BD01
	<b>Admitted:</b> 9/10/2015
	<b>Rejected:</b>
	<b>Identified:</b> 9/8/2015
	<b>Withdrawn:</b>
	<b>Stricken:</b>
	<b>Other:</b>

**NRC-094**  
**Submitted: 9/8/2015**

**NUREG/CR-6940**  
**PNNL-16396**



# **Combined Estimation of Hydrogeologic Conceptual Model, Parameter, and Scenario Uncertainty with Application to Uranium Transport at the Hanford Site 300 Area**

**Pacific Northwest National Laboratory**

**U.S. Nuclear Regulatory Commission  
Office of Nuclear Regulatory Research  
Washington, DC 20555-0001**

## AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

### NRC Reference Material

As of November 1999, you may electronically access NUREG-series publications and other NRC records at NRC's Public Electronic Reading Room at <http://www.nrc.gov/reading-rm.html>. Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and *Title 10, Energy*, in the Code of *Federal Regulations* may also be purchased from one of these two sources.

1. The Superintendent of Documents  
U.S. Government Printing Office  
Mail Stop SSOP  
Washington, DC 20402-0001  
Internet: [bookstore.gpo.gov](http://bookstore.gpo.gov)  
Telephone: 202-512-1800  
Fax: 202-512-2250
2. The National Technical Information Service  
Springfield, VA 22161-0002  
[www.ntis.gov](http://www.ntis.gov)  
1-800-553-6847 or, locally, 703-605-6000

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

Address: U.S. Nuclear Regulatory Commission  
Office of Administration  
Mail, Distribution and Messenger Team  
Washington, DC 20555-0001

E-mail: [DISTRIBUTION@nrc.gov](mailto:DISTRIBUTION@nrc.gov)  
Facsimile: 301-415-2289

Some publications in the NUREG series that are posted at NRC's Web site address <http://www.nrc.gov/reading-rm/doc-collections/nuregs> are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

### Non-NRC Reference Material

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

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—

The NRC Technical Library  
Two White Flint North  
11545 Rockville Pike  
Rockville, MD 20852-2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

American National Standards Institute  
11 West 42<sup>nd</sup> Street  
New York, NY 10036-8002  
[www.ansi.org](http://www.ansi.org)  
212-642-4900

Legally binding regulatory requirements are stated only in laws; NRC regulations; licenses, including technical specifications; or orders, not in NUREG-series publications. The views expressed in contractor-prepared publications in this series are not necessarily those of the NRC.

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG-XXXX) or agency contractors (NUREG/CR-XXXX), (2) proceedings of conferences (NUREG/CP-XXXX), (3) reports resulting from international agreements (NUREG/IA-XXXX), (4) brochures (NUREG/BR-XXXX), and (5) compilations of legal decisions and orders of the Commission and Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of NRC's regulations (NUREG-0750).

**DISCLAIMER:** This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any employee, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this publication, or represents that its use by such third party would not infringe privately owned rights.

NUREG/CR-6940  
PNNL-16396

# **Combined Estimation of Hydrogeologic Conceptual Model, Parameter, and Scenario Uncertainty with Application to Uranium Transport at the Hanford Site 300 Area**

Manuscript Completed: June 2007  
Date Published: July 2007

Prepared by  
P.D. Meyer, M. Ye (DRI), M.L. Rockhold, S.P. Neuman (UA)  
K.J. Cantrell

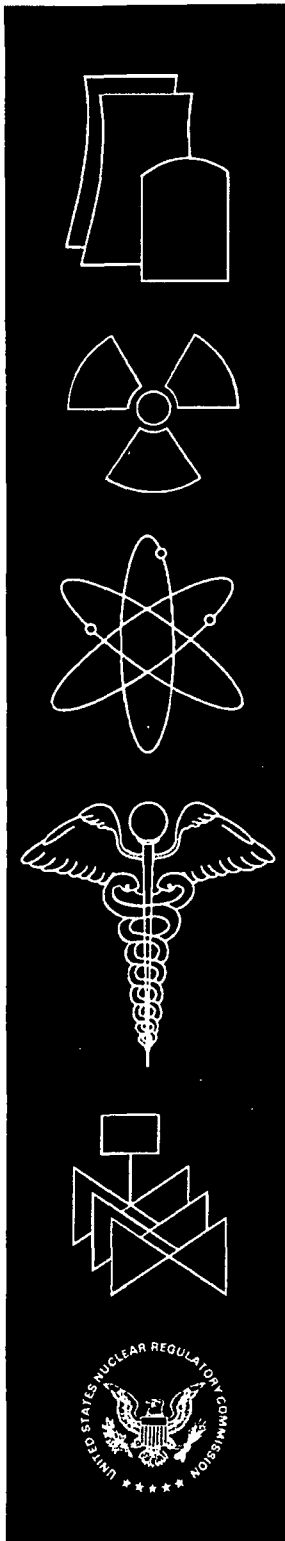
Pacific Northwest National Laboratory  
P.O. Box 999  
Richland, WA 99352

Desert Research Institute  
Las Vegas, NV 89119  
(Currently at Florida State University, Tallahassee, FL 32306)

University of Arizona  
Tucson, AZ 85721

T.J. Nicholson, NRC Project Manager

**Prepared for**  
**Division of Fuel, Engineering and Radiological Research**  
**Office of Nuclear Regulatory Research**  
**U.S. Nuclear Regulatory Commission**  
**Washington, DC 20555-0001**  
**NRC Job Code Y6465**







## ABSTRACT

We describe the development and application of a methodology to systematically and quantitatively assess predictive uncertainty in groundwater flow and transport modeling. The methodology considers the combined impact of hydrogeologic uncertainties associated with the conceptual-mathematical basis of a model, model parameters, and the scenario to which the model is applied. The methodology is based on an extension of a Maximum Likelihood implementation of Bayesian Model Averaging. Model uncertainty is represented by postulating a discrete set of alternative conceptual models for a site with associated prior model probabilities. The prior model probabilities reflect a subjective belief about the relative plausibility of each model based on its apparent consistency with available knowledge and data. Posterior model probabilities are computed and parameter uncertainty is estimated by calibrating each model to observed system behavior. Posterior model probabilities are modifications of the subjective prior values based on an objective evaluation of each model's consistency with available data. Prior parameter estimates are optionally included. Scenario uncertainty is represented as a discrete set of alternative future conditions affecting boundary conditions, source/sink terms, or other aspects of the models. The associated prior scenario probabilities reflect a subjective belief about the rela-

tive plausibility of the alternative scenarios. A joint assessment of uncertainty results from combining model predictions computed under each scenario using as weights the posterior model and prior scenario probabilities. The computed model predictions incorporate parameter uncertainties using, for example, Monte Carlo simulation. The uncertainty methodology was applied to modeling of groundwater flow and uranium transport at the Hanford Site 300 Area. Eight alternative models representing uncertainty in the hydrogeologic and geochemical properties as well as the temporal variability were considered. Two scenarios representing alternative future behavior of the Columbia River adjacent to the site were considered. The scenario alternatives were implemented in the models through the boundary conditions. Alternative models were calibrated using hydraulic head and uranium concentration observations over a seven-year period. Uranium concentrations under each scenario were predicted over a 20-year period. Results demonstrate the feasibility of applying a comprehensive uncertainty assessment to large-scale, detailed groundwater flow and transport modeling. Results also illustrate the ability of the methodology to provide better estimates of predictive uncertainty, quantitative results for use in assessing risk, and an improved understanding of the system behavior and the limitations of the models.

### Paperwork Reduction Act Statement

This NUREG does not contain information collection requirements and, therefore, is not subject to the requirements of the Paperwork Reduction Act of 1995 (44 U.S.C. 3501 et seq.).

### Public Protection Notification

The NRC may not conduct or sponsor, and a person is not required to respond to, a request for information or an information collection requirement unless the requesting document displays a currently valid OMB control number.



## FOREWORD

This report was prepared by Pacific Northwest National Laboratory (PNNL) under a U.S. Department of Energy (DOE) Interagency Work Order (JCN Y6465) with the U.S. Nuclear Regulatory Commission (NRC). This research report describes a systematic, quantitative approach for combining estimations of uncertainties in the data, conceptual models, parameters, and scenarios related to hydrogeology (i.e., the occurrence and distribution of underground water). This is important because the ground-water pathway is a key consideration in assessing the potential for radionuclide transport and possible exposure scenarios to the public and, consequently, it is important to evaluate and understand the related uncertainties. Toward that end, this report provides details on the primary factors that contribute to those uncertainties, and how to quantify them through comparative modeling and analyses of site characterization and monitoring data.

The combined uncertainties approach described in this report uses a statistical method, known as maximum likelihood Bayesian model averaging, to assess predictive uncertainty in ground-water flow and transport modeling. To do so, this approach compares alternative models and assesses their combined predictive uncertainty. This approach is demonstrated by application to ground-water flow and uranium transport modeling at the 300-Area of the DOE Hanford Site. Toward that end, eight alternative models were considered representing uncertainty in hydrogeologic and geochemical properties, as well as the temporal variability. In addition, two scenarios representing alternative future behavior of the Columbia River adjacent to the site were also considered. Results demonstrate the feasibility of applying a comprehensive uncertainty assessment to large-scale, detailed ground-water flow and transport modeling. In addition, the results illustrate the practical benefits of the approach to provide better estimates of predictive uncertainty, quantitative results for use in assessing risk, and an improved understanding of the system behavior and limitations of the models. Although the approach and its applications were designed for reviews of radionuclide transport at complex decommissioning sites, it is also useful for assessing nuclear facility siting, designing ground-water monitoring programs, remediating ground water, and identifying and selecting strategies to preclude offsite migration of abnormal radionuclide releases.

This approach is consistent with the NRC's strategic performance goal of making the agency's activities and decisions more effective, efficient, and realistic by identifying and estimating uncertainties. Toward that end, this report demonstrates, using examples relevant to decommissioning analyses, that sources of uncertainty can be identified, quantified, and integrated using a comparative model analysis approach. This report also illustrates the effectiveness of the combined uncertainty approach to estimate uncertainty in model predictions arising from conceptual, parameter and scenario uncertainties. This information is assisting NRC licensing staff and regional inspectors, Agreement State regulators, and licensees in their decision-making by identifying and quantifying overall uncertainties in performance assessment models.

This report is not a substitute for NRC regulations, and compliance is not required. Consequently, the approaches and methods described in this report are provided for information only, and publication of this report does not necessarily constitute NRC approval or agreement with the information contained herein. Similarly, use of product or trade names is for identification purposes only and does not constitute endorsement by either the NRC or PNNL.

---

Brian W. Sheron, Director  
Office of Nuclear Regulatory Research  
U.S. Nuclear Regulatory Commission



# CONTENTS

Abstract .....	iii
Foreword.....	v
Executive Summary.....	xiii
Acknowledgments .....	xvii
1 Introduction .....	1
2 Quantification of Parameter, Conceptual Model, and Scenario Uncertainties.....	11
2.1 Parameter Uncertainty .....	11
2.1.1 Analysis of Parameter Uncertainty .....	12
2.2 Conceptual Model Uncertainty.....	14
2.2.1 Analysis of Conceptual Model Uncertainty.....	15
2.3 Scenario Uncertainty .....	15
2.3.1 Analysis of Scenario Uncertainty .....	17
3 Joint Estimation of Model, Parameter, and Scenario Uncertainties.....	21
3.1 Bayesian Model Averaging – Combining Conceptual Model and Parameter Uncertainty.....	21
3.1.1 Maximum Likelihood Bayesian Model Averaging (MLBMA).....	22
3.1.1.1 Applicability of MLBMA .....	24
3.2 Incorporation of Scenario Uncertainty .....	24
3.2.1 Bayesian Model Averaging Conditioned on a Specific Scenario .....	25
3.2.2 Scenario Averaging.....	25
3.3 Summary of Uncertainty Assessment Methodology .....	26
4 Example Application .....	29
4.1 Available Site Data and Related Information .....	29
4.1.1 Application Field Site Background.....	29
4.1.2 Hanford 300 Area Data Sources .....	36
4.1.2.1 Surface and Borehole Geophysics .....	36
4.1.2.2 Physical and Hydraulic Properties .....	37
4.1.2.3 Geochemistry, Uranium Sorption, and Transport .....	39
4.1.2.4 Source Terms and Boundary Conditions .....	43

4.1.2.5	Hydraulic Head and Uranium Concentration Data .....	43
4.2	Postulate Alternative Conceptual Models.....	48
4.2.1	Hydraulic Property Heterogeneity .....	52
4.2.2	River Boundary Transience .....	52
4.2.3	Adsorption .....	52
4.2.4	Alternative Conceptual Model Summary .....	53
4.3	Alternative Model Implementations .....	53
4.3.1	Simulation Domain .....	53
4.3.2	Discretization .....	54
4.3.3	Hydrogeologic Zonation and Hydraulic Parameters.....	54
4.3.4	Initial Conditions .....	55
4.3.5	Geochemical Modeling.....	55
4.4	Prior Model Probability .....	55
4.5	Prior Parameter Probability .....	59
4.5.1	Modeling Adsorption in Contaminant Transport Models.....	59
4.5.2	Hanford Site 300 Area Uranium Example .....	60
4.5.2.1	No Site-Specific Data Case.....	60
4.5.2.2	Known Soil Texture and pH Case.....	60
4.5.2.3	Available Regional Batch $K_d$ Values .....	61
4.5.2.4	Available Site-Specific Batch $K_d$ Values .....	61
4.5.2.5	Available Site-Specific, Mechanistically-Based, Adsorption Model Case .....	62
4.5.3	Prior Values of $K_d$ for 300 Area Model Calibration.....	63
4.6	Alternative Hydrologic Scenarios.....	64
4.7	Prior Scenario Probability .....	65
4.8	Model Calibration.....	65
4.8.1	Selection of Calibration Data.....	65
4.8.2	Weights Associated with Calibration Data .....	65
4.8.3	Joint Model Calibration and Calibrated Model Parameters .....	67
4.9	Model Calibration Results.....	69
4.10	Posterior Model Probabilities .....	73

# 1 INTRODUCTION

Regulatory and design applications of hydrogeologic models of flow and contaminant transport often involve using the models to make predictions of future system behavior. For example, the primary U.S. Nuclear Regulatory Commission (NRC) criterion for license termination requires an estimate of maximum dose for a period up to 1000 years from the time of decommissioning. This dose estimate can be made using a simulation model of the transport of residual radionuclide contaminants through the environment. Similarly, the design of a subsurface contaminant remediation system and a network to monitor performance of that system can be facilitated using a model that predicts system behavior under the conditions of the remediation.

A variety of factors conspire to render the predictions of hydrogeologic models uncertain, including

- incomplete knowledge of the system,
- variability in system properties,
- randomness in the system stresses,
- measurement and sampling errors, and
- disparity among sampling, simulation, and actual scales of the system.

Other factors that may affect the degree of uncertainty inherent in these predictions are the potentially long time frame over which predictions may be made (e.g., 1000 years or more), the simulation of transport through multiple exposure pathways and media, and regulatory criteria that may be close to contaminant background concentrations (requiring the prediction of small effects).

Comprehensive and quantitative assessment of the impact of the factors listed above on the uncertainty in predictions of hydrogeologic models may require a significant investment in time for analysis and data collection. Given this, what benefits would justify this investment? Morgan and Henrion (1990) considered this question and concluded that explicit consideration of uncertainty is important for the following reasons.

- When attitudes toward risk are important, e.g., when stakeholders are risk averse (as is commonly the case for people living or working near an environmentally contaminated site), considering uncertainty can improve decision making by quantifying risk.
- When multiple, uncertain sources of information must be combined (a defining characteristic of environmental transport modeling), uncertainty assessment provides a means to systematically

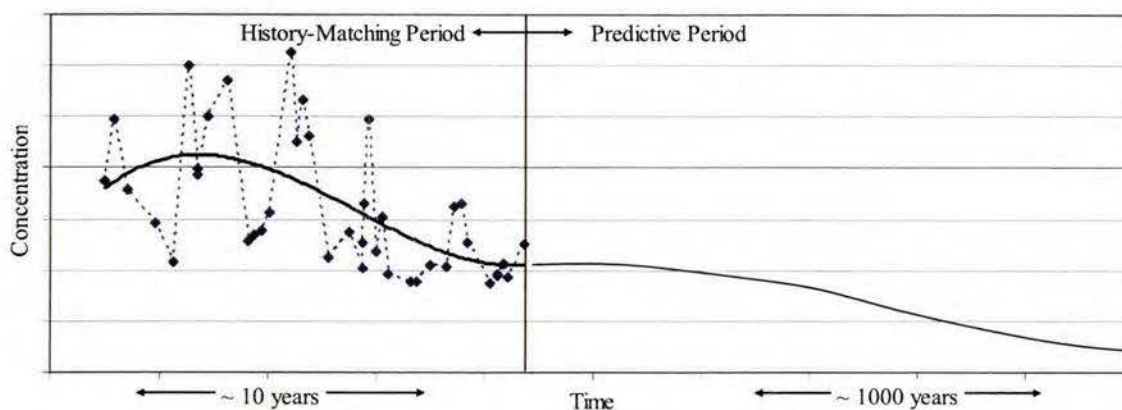
weight and combine the information sources to estimate the precision of the predicted value(s).

- When one of the possible actions is to collect additional data, consideration of uncertainty provides guidance on what data to collect and the potential benefit in reduced uncertainty. This issue is particularly important in hydrogeologic modeling, which is generally constrained by limited characterization data.
- Empirical evidence suggests that “best estimate” answers are regularly biased. Thinking about the uncertainty associated with an analysis can reduce this bias.
- Analysis of uncertainty can help with model development, defining the appropriate level of detail for model components.
- A documented uncertainty analysis helps users of model results evaluate conclusions and limitations, particularly over time as different users arise and project objectives evolve.
- Attempts to characterize and address important uncertainties help analysts fulfill a professional and ethical responsibility to communicate the implications and limitations of their work.

It thus appears that the potential benefits of uncertainty analysis can justify the expense. The main question is how this analysis should proceed. In any case, there are several desirable characteristics that should be cultivated. A methodology for uncertainty assessment should be comprehensive in the sense that all types of uncertainty can be included. The methodology should also be quantitative so that results can be compared to regulatory criteria. Finally, the methodology should be systematic so that it can be applied to a wide range of sites and objectives and to enable the common application of computer codes and methods.

Before presenting the uncertainty assessment methodology adopted in this work, it will be useful to discuss a framework for the application of hydrogeologic models to regulatory decision making. As illustrated in Figure 1-1, the time domain over which model simulations are conducted can be viewed as two distinct periods. The history-matching period consists of the time over which observations of the system are available (as indicated by the diamond symbols in Figure 1-1). Model building and model evaluation take place within the context of the data available during the history-matching period. The predictive period consists of the time over which the behavior of the system is to be predicted for site management purposes. For many





**Figure 1-1. Framework for application of hydrogeologic models to regulatory decision making**

problems, the predictive period will be much longer than the history-matching period.

In general, specific factors resulting in model predictive uncertainty are assessed in the history-matching period and then propagated into the predictive period. Parameter uncertainty in hydrogeologic models is typically addressed in this fashion. Some factors, however, only apply in the predictive period: randomness in future rainfall is an example. As part of the model building and evaluation process, uncertainty in the history-matching period can be reduced by collecting additional data. For example, a pump test could be conducted to reduce uncertainty about the hydraulic conductivity at a specific location. In contrast, uncertainties that apply only to the predictive period cannot be reduced by collecting data. For example, the annual rainfall 10 years in the future is uncertain due to natural variability. Although past and current measurements of annual rainfall can better characterize that variability, the essential randomness of rainfall ten years hence remains.

Although the analysis described here is limited to hydrogeologic uncertainty, it is comprehensive in the sense that the primary uncertainties in most hydrogeologic modeling applications can be included under this framework. Three broad types of uncertainty are considered. Uncertainties are manifested in a hydrogeologic modeling application as uncertainty in model conceptualization, model parameters, and modeling scenarios. The model conceptual basis can be thought of as a hypothesis about the behavior of the system being modeled and the relationships between the components of the system. This conceptualization is typically represented mathematically to render quantitative predictions; thus it is appropriate to talk about a conceptual-mathematical model (sometimes referred to as

model structure). The model parameters are the quantities required to obtain a solution from the model (and thus are model-specific). A scenario is defined here as a future state or condition assumed for a system, with the emphasis on those aspects of a scenario that affect the system hydrology (e.g., future irrigation schemes, ground-water extraction, natural recharge). With reference to the framework of Figure 1-1, conceptual model and parameter uncertainties are assessed in the history-matching period and applied in the predictive period. Scenario uncertainty applies to the predictive period only.

What evidence is there for the relative importance of conceptual model, parameter, and scenario uncertainties in modeling practice? Published results from hydrogeologic model post-audits were reviewed to attribute the primary modeling errors in these applications to conceptual, parameter, or scenario uncertainties. Six additional modeling applications described in Bredehoeft (2005) were included in this review. Results are shown in Table 1-1 and demonstrate the importance of conceptual and scenario uncertainties in contributing to model predictive errors. In 9 of the 16 applications, conceptual model errors were most significant. Model scenario errors were the most significant in 4 of the 16 applications. Parameter errors were most significant in three of the applications. This (limited) review suggests that a comprehensive approach to uncertainty assessment in hydrogeologic modeling should not be limited to parameter uncertainties, but must also consider the potential for significant conceptual model and scenario uncertainties if a realistic estimate of predictive uncertainty is desired.

Given our knowledge of the potential importance of hydrogeologic uncertainties and their categorization as conceptual model, parameter, or scenario uncertainty,



**Table 1-1. Attribution of primary errors in hydrogeologic model applications (after Bredehoeft 2005)**

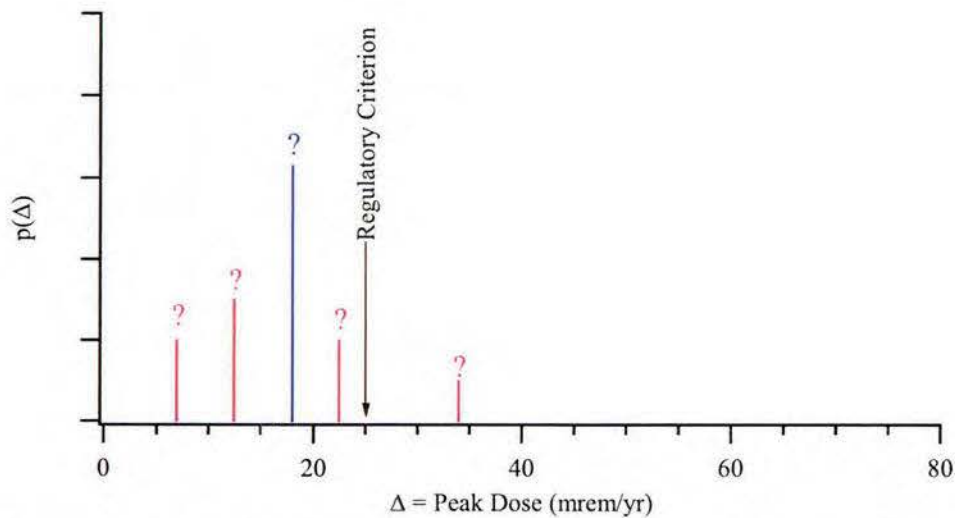
Model Application (Reference)	Comments	Error
Phoenix (Konikow 1986)	Assumed past groundwater pumping would continue in future	Scenario/ Conceptual
Cross Bar Ranch Wellfield (Stewart and Langevin 1999)	Assumed a 75-day, no-recharge scenario would represent long-term maximum drawdown	Scenario/ Conceptual
Arkansas Valley (Konikow and Person 1985)	Needed a longer period of calibration	Scenario/ Parameter
Coachella Valley (Konikow and Swain 1990)	Recharge events unanticipated	Scenario
INEL (Lewis and Goldstein 1982)	Dispersivities poorly estimated	Parameter
Milan Army Plant (Andersen and Lu 2003)	Extrapolated localized pump test results to larger area	Parameter
Blue River (Alley and Emery 1986)	Storativity poorly estimated	Parameter/ Conceptual
Houston (Jorgensen 1981)	Including subsidence in model improved predictions	Conceptual
HYDROCOIN (Konikow et al. 1997)	Boundary condition modeled poorly	Conceptual
Ontario Uranium Tailings (Flavelle et al. 1991)	Inadequate chemical reaction model	Conceptual
Los Alamos (Bredehoeft 2005)	Flow through unsaturated zone not understood	Conceptual
Los Angeles (Bredehoeft 2005)	Flow vectors 90° off in model	Conceptual
Summitville (Bredehoeft 2005)	Seeps on mountain unaccounted for	Conceptual
Santa Barbara (Bredehoeft 2005)	Fault zone flow unaccounted for	Conceptual
WIPP (Bredehoeft 2005)	Assumed salt had no mobile interstitial brine	Conceptual
Fractured Rock Waste Disposal (Bredehoeft 2005)	Preferential flow in unsaturated zone unaccounted for	Conceptual

what are the analysis options? One option is to assume the conceptual model and scenario are known and to address parameter uncertainty only. This has been and remains the most common approach to uncertainty analysis in hydrogeologic modeling. One way to assess the impact of parameter uncertainty is by using a sensitivity approach. An example of this approach is to develop best-estimates of model parameter values using data in the history-matching period and then use these to compute a best-estimate predicted value in the predictive period. Parameter values are then perturbed from their best estimates to determine potential uncertainty in the predicted value. For a decommissioning analysis, the results of this sensitivity approach could be displayed as in Figure 1-2, which compares the predicted best-estimate peak annual dose (in blue) and four sensitivity cases (in red) with the regulatory criterion of 25 mrem/yr. The vertical axis in this figure is the probability of peak dose, representing the degree of plausibility of the model result (Jaynes 2003). The question marks indicate that the actual values of the probabilities are unknown, although statements about the relative values may be possible (e.g., the best-estimate result is presumably more plausible than the sensitivity cases). A variation of this approach is a bounding (conservative) analysis in which the desired predicted value represents the worst plausible behavior

of the system (e.g., the right-most sensitivity case in Figure 1-2).

The problem with a sensitivity approach to assessing the impact of parameter uncertainty is that, as indicated, the probabilities of predicted outcomes are not estimated. There is thus no way to quantitatively estimate the risk involved in a particular decision. In the example of Figure 1-2, the risk corresponds to the probability of failure in meeting the regulatory criterion, that is, the probability that the peak dose is greater than 25 mrem/yr. For a bounding analysis, the significance of the bounding case needs to be assessed to avoid being overly conservative. In Figure 1-2, the significance of the sensitivity case that violates the regulatory criterion rises with its probability, but there is no way to know whether that probability is 1% or 20%.

A quantitative assessment of parameter uncertainty can be completed by computing the probability density function of the desired predicted value. This can be accomplished by assigning a joint probability distribution to model parameters and propagating this through the model, using a Monte Carlo simulation for example. The joint probability distribution is a measure of the degree of plausibility of the values of model parameters and is assigned based on site information and

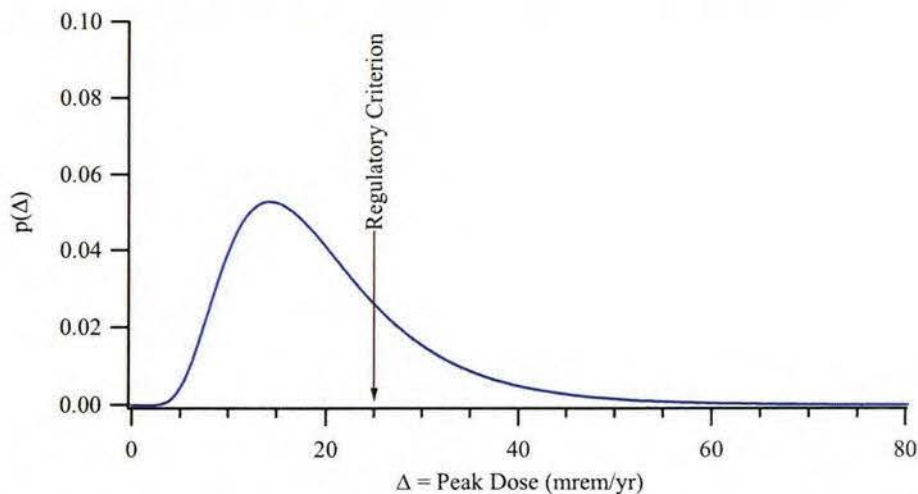


**Figure 1-2. Illustration of possible results of a decommissioning analysis using a sensitivity approach. Best-estimate predicted peak dose is shown in blue. Sensitivity cases are shown in red.**

data available in the history-matching period (using, for example, expert judgment or inverse modeling). Assessing the impact of parameter uncertainty on model predictions in this manner is accepted in policy (EPA 1997; NRC 2003) and is not uncommon in practice. For the decommissioning example, the simulation result could be displayed as shown in Figure 1-3 in which the probability density function of the desired predicted value, peak dose, is compared to the regulatory criterion of 25 mrem/yr. As in Figure 1-2, the vertical axis (probability density) represents the degree of plausibility associated with the predicted value. Unlike the sensitivity approach of Figure 1-2, a quantitative

estimate of risk is available here. The probability of exceeding the regulatory criterion is easily calculated from the density function as the area under that portion of the curve exceeding 25 mrem/yr, as indicated by the hatched area in Figure 1-3.

As stated above, the examples considered so far have assumed that the conceptual model and scenario are known. That is, it is implicit in any assessment that considers only parameter uncertainty that the resulting predictive uncertainty is conditional on the structure of the model. It is generally recognized, however, that a



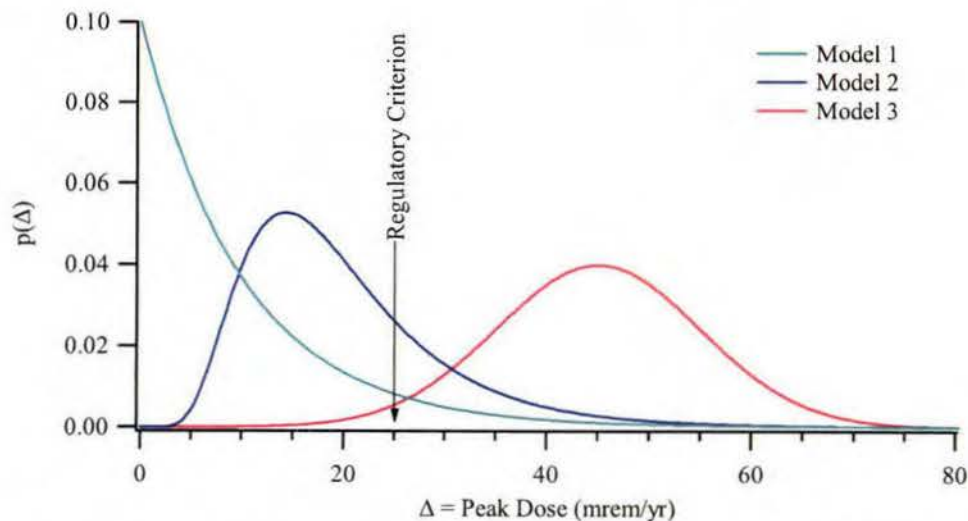
**Figure 1-3. Example results of a decommissioning analysis: the probability density function of peak dose results from the inclusion of parameter uncertainty**

hydrogeologic model of a site is invariably an approximation of the actual system. As a consequence, it may be possible to postulate more than one conceptual model for a site that is consistent with site characterization data and observed system behavior, as evaluated in the history-matching period. Although the potential importance of conceptual model uncertainty has been accepted for some time (Apostolakis 1990; Mosleh et al. 1994), practical methods to assess the impact of model uncertainty on prediction have not found their way into widespread practice.

In a quantitative uncertainty analysis, parameter uncertainty is typically characterized using continuous probability distributions. When characterizing conceptual uncertainty in hydrogeologic modeling, specifying a continuum of conceptual model possibilities is likely to be infeasible. Instead, it is generally more appropriate to postulate a discrete set of alternative conceptual models (Neuman and Wierenga 2003). This suggests a sensitivity approach to conceptual model uncertainty analogous to the sensitivity approach to addressing parameter uncertainty. Namely, each model alternative is used to simulate the desired predicted value, producing a result that might look like the example in Figure 1-4. In this figure there are three alternative conceptual models, each predicting the same quantity, peak dose. Each model result is represented as a probability distribution because the parameters of each model are uncertain. There is no requirement that the models have a common set of parameters or that parameters common to more than one model have the same value (or probability distribution).

A sensitivity approach to assessing model uncertainty has the same drawback as the parameter sensitivity approach discussed with reference to Figure 1-2. Without a quantitative measure of the degree of plausibility of model alternatives, it is impossible to determine the risk of a decision based on the model predictions. Similarly, a conservative approach to model uncertainty relies on an implied belief that the most conservative model has a non-negligible degree of plausibility. A conservative approach in this case would select Model 3 for the comparison with the regulatory criterion; this is easier to justify if the three models are equally plausible than if Models 1 and 2 are significantly more plausible than Model 3. Such a justification requires a quantitative measure of model plausibility.

A quantitative assessment of the combined effects of parameter and conceptual model uncertainty can be achieved by assigning a discrete probability distribution to the model alternatives. The model predictions are then combined using a weighted average with the weight for each model's prediction consisting of that model's probability (e.g., Apostolakis 1990). Analogous to the interpretation of parameter probability, the discrete model probability distribution represents the degree of plausibility of the model alternatives. Shown in Figure 1-5 is the model averaged result for the three-model decommissioning example with model probabilities of 0.5 for Model 1 and 0.25 for the other two models. In this example, Model 1 is thus twice as plausible as Model 2 or 3, with the latter two models being equally plausible. The resulting model-averaged probability density function is properly interpreted as a



**Figure 1-4. Example results of a decommissioning analysis with three conceptual model alternatives: the probability density functions of peak dose result from the inclusion of parameter uncertainty for each model**



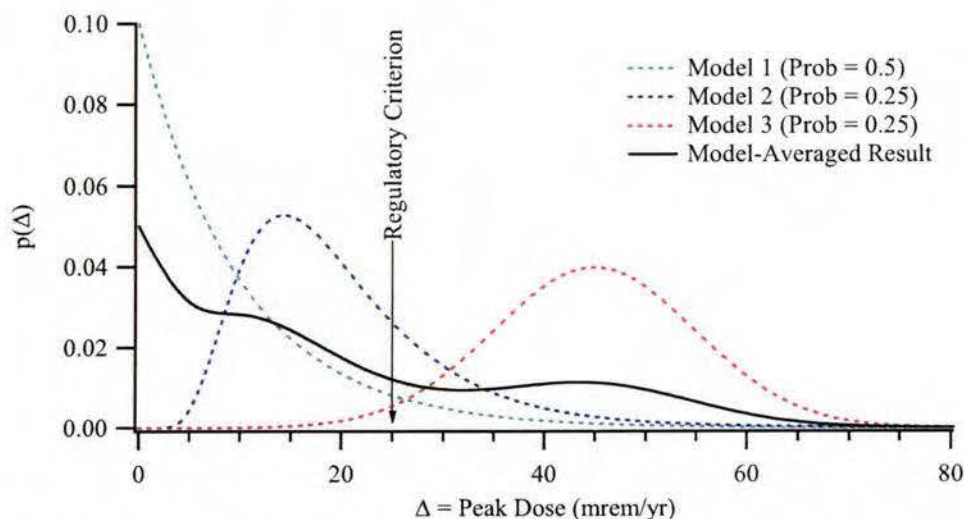
measure of the degree of plausibility of peak dose that takes into consideration the joint effect of parameter and model uncertainties.

The model-averaging approach has been criticized on the basis that there is only one model corresponding to the physical reality and therefore an average over several models has no physical interpretation (e.g., Abramson 1994). Even if one postulates the existence of a model corresponding exactly to reality, the inherent complexity of the hydrogeologic environment makes it unlikely that this model will be contained in any realistic set of alternatives. Rather, although the models are physically based, it is likely that all models considered will be an approximation to physical reality. When appropriately formulated, each model alternative will have some merit in reproducing aspects of the physical system, this merit being quantified by each model's probability. As articulated in Ye et al. (2004), model probability is interpreted as a relative measure with respect to the other model alternatives considered. Thus if an additional model (Model 4) were to be added to the example of Figure 1-5, the individual model probabilities would change to accommodate this, but the probability of Model 1 would still be twice that of Models 2 and 3. A model-averaged consequence has an intuitive and consistent meaning under this interpretation of model probability. Note that we are not averaging the underlying physics, but the predicted consequences of the physics as rendered by the models.

Model averaging has also been criticized for masking information essential to the decision maker (Abramson 1994). This criticism would be valid if the only quan-

tity the decision maker had available was the model-averaged probability density function (the black curve in Figure 1-5) or some measure derived from it, such as the mean peak dose. If, however, the individual model results are presented along with the model-average results and the model probabilities (as in Figure 1-5 and Table 1-2), a fully informed decision can be made. For example, clear differences in the predicted values of models (e.g., Models 1 and 3 in Figure 1-5) suggest that a decision maker might justifiably request additional data/information to better discriminate between these models (i.e., to modify the model probabilities until one model dominates the others). In addition, if some conservatism is to be built into the decision, the model probabilities can provide a basis for selecting the most conservative model (Model 3 in this case) that carries a significant degree of plausibility.

Conservatism could also be included using the model-averaged results, for example, by comparing the regulatory criterion to a high percentile (e.g., the 90<sup>th</sup> percentile) of the model-averaged probability distribution, or by using a probability of exceedance regulatory criterion instead of a deterministic criterion. As given in Table 1-2, the model-averaged mean peak dose is 21.2 mrem/yr (satisfying the deterministic criterion), the 90<sup>th</sup> percentile is 48.5 mrem/yr, and the probability of exceeding 25 mrem/yr is 34%. These factors all suggest that a conservative regulatory action may be preferred in this case, but based on a fully informed consideration of model and parameter uncertainty (i.e., risk), rather than on adoption of the most conservative model.



**Figure 1-5. Model averaged probability density function for the three-model example of Figure 1-4 with model probabilities (see legend) used as weights**

**Table 1-2. Statistics of individual models and model-averaged results shown in Figure 1-5 using model probabilities of 0.5, 0.25, and 0.25 for Models 1, 2, and 3, respectively**

	Mean Dose	Prob (Dose > 25)	90 <sup>th</sup> ile
Model 1	10.0	8.2	23.0
Model 2	20.0	23.9	32.7
Model 3	45.0	97.7	57.8
Model Average	21.2	34.5	48.5

Portraying model results in the manner of Figure 1-5 does not mask essential information, but rather clearly illuminates the effect of model uncertainty and the importance of a consistent and defensible means to estimate model probabilities. Model averaging reduces the risk of relying on an overly conservative model and provides a consistent and quantitative way to address model uncertainty in the context of a regulatory decision.

Parameter and model probabilities, being measures of the degree of plausibility of parameter values and model alternatives, are based on information and data available in the history-matching period. That is, they are estimated as part of the process of model development and evaluation. The characterizations of parameter and model uncertainty can be projected into the predictive period. As mentioned previously, there is an additional source of uncertainty that applies in the predictive period, namely the scenario uncertainty. A scenario is a description of the future conditions under which a model is applied. While scenario development is commonly associated with radioactive waste disposal performance assessment (NEA 2001), the concept applies to any modeling application in which prediction of future system behavior is made. Scenarios are inherently uncertain since they describe conditions in the (uncertain) future.

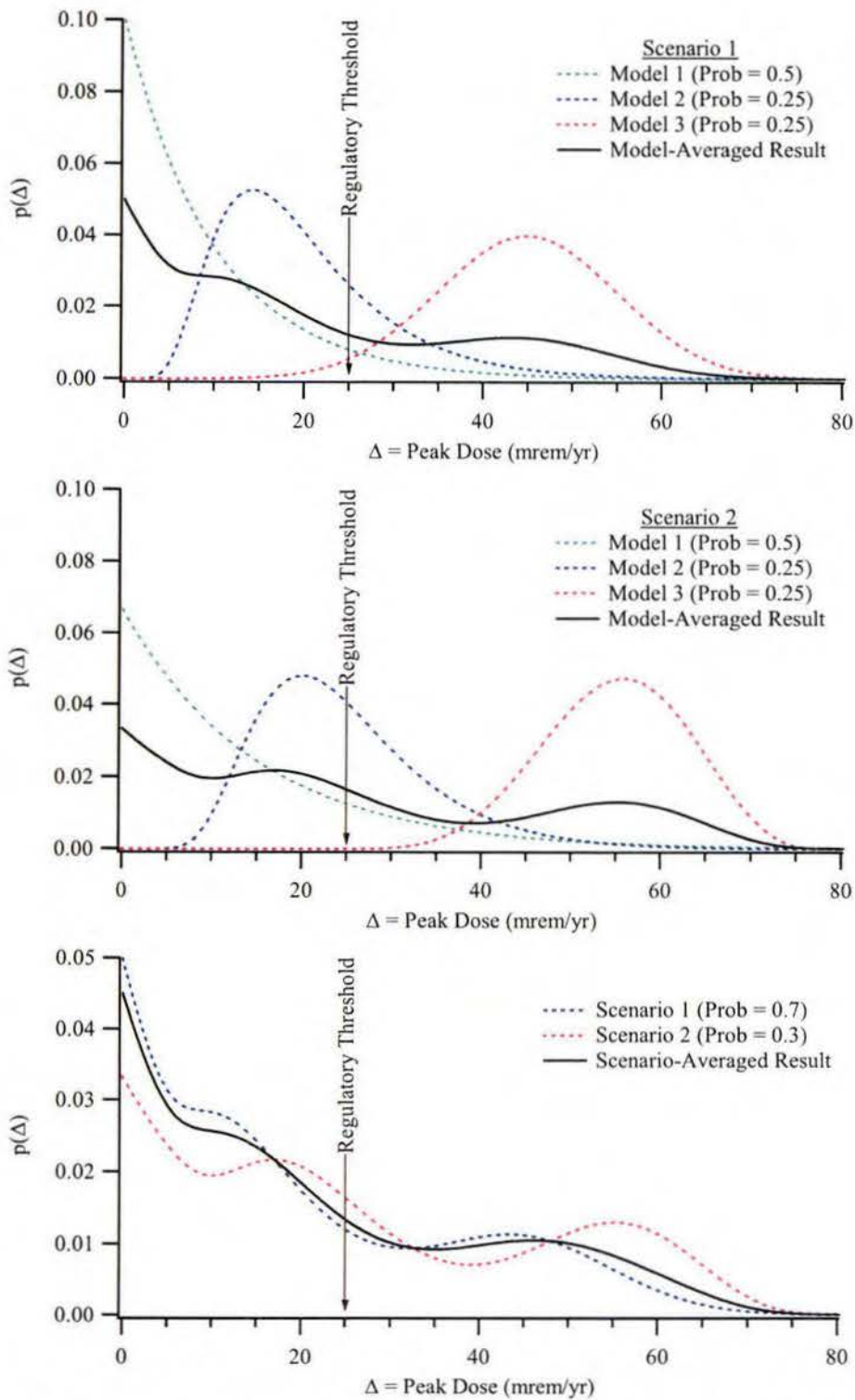
Similar to the representation of conceptual model uncertainty, the uncertainty in future site conditions can be represented as a set of alternative scenarios. In general, only those scenarios that are minimally plausible and have a significant potential impact on the predicted consequence are included in the set of alternatives. In the case where future scenarios are characterized by changes in model inputs such as boundary or source terms (e.g., changes in surface recharge or pumping rates), it may also be useful to represent scenario uncertainties by treating these inputs as random. As in the case of model uncertainty, considering discrete sce-

nario alternatives allows for a systematic specification of scenario uncertainty.

Once the alternative scenarios are defined, the impact of scenario uncertainty can be addressed using a sensitivity approach in which the desired predictions are computed with each model alternative under each alternative scenario. Results for the decommissioning example with two alternative scenarios are shown in Figure 1-6. In this example, Scenario 1 is identical to Figure 1-5. Scenario 2 is an alternative that, in general, increases the expected peak dose by about 5 mrem/yr for each model (the impact on Model 3 is somewhat greater). Although the example is hypothetical, one might think of Scenario 2 as an alternative land use scenario that increases the expected recharge and, as a result, the expected peak dose from residual contamination. For Scenario 2, the model-averaged mean peak dose is 27.3 mrem/yr (violating the deterministic criterion), the 90<sup>th</sup> percentile is 58.5 mrem/yr, and the probability of exceeding 25 mrem/yr is 45%.

For some applications, a sensitivity approach to scenario uncertainty may be the most appropriate point at which to terminate the analysis. The predicted impacts under each scenario can be qualitatively assessed to determine the course of action. Analogous to the characterization of conceptual model uncertainty, it may be possible, however, to assign a discrete probability distribution to the scenario alternatives. In that case, a scenario-averaged probability distribution of the desired predicted value can be computed as a weighted average of the individual scenario (model-averaged) results with the weights being the scenario probabilities. Sample results for the decommissioning example are shown in Figure 1-6 (bottom) using probabilities of 0.7 and 0.3 for the two scenarios. The model-averaged distributions include the impacts of parameter and conceptual model uncertainties under each scenario. The scenario-averaged distribution includes the combined impacts of parameter, conceptual model, and scenario uncertainties. Relevant statistics can be directly determined from the scenario-averaged result. For the example in Figure 1-6, the scenario-averaged mean peak dose is 23.0 mrem/yr, the 90<sup>th</sup> percentile is 52.1 mrem/yr, and the probability of exceeding 25 mrem/yr is 38%. Summary statistics for the individual scenarios and the scenario average are listed in Table 1-3.

As with model probabilities, the scenario probabilities should be interpreted as relative values conditional to the alternatives considered. If an additional scenario alternative were to be added to the two considered in Figure 1-6, the probability of Scenario 1 would still be 2.33 times the probability of Scenario 2, reflecting the judgment that Scenario 1 is more than two times as



**Figure 1-6. Results for two alternative scenarios (top/middle), each of which includes the impact of conceptual model and parameter uncertainties. Model-averaged results for each scenario and scenario-averaged result assuming probabilities of 0.7 and 0.3 for Scenarios 1 and 3, respectively (bottom).**

plausible as Scenario 2. Because the scenario probabilities define a discrete distribution for scenario uncertainty, the sum of all scenario probabilities will be 1.0 for any number of scenario alternatives.

**Table 1-3. Statistics of individual scenarios (model-average results) and the scenario-average results shown in Figure 1-6 (lower plot) using probabilities of 0.7 and 0.3 for Scenarios 1 and 2, respectively**

	Mean Dose	Prob (Dose > 25)	90%ile
Scenario 1 (model-average)	21.2	34.5	48.5
Scenario 2 (model-average)	27.3	44.8	58.5
Scenario Average	23.0	37.6	52.1





## 2 QUANTIFICATION OF PARAMETER, CONCEPTUAL MODEL, AND SCENARIO UNCERTAINTIES

### 2.1 Parameter Uncertainty

One of the primary factors that contribute to hydrogeologic uncertainty is the natural variability of the subsurface. Examples of this variability are illustrated in Figure 2-1, two photos from excavations at the Hanford Site in Washington State. On the left is a trench face from an excavation in the 200 Area. A large variation in soil particle size can be seen, ranging from fine silts to very coarse gravels. The profile shows a layered structure with evidence of cross-bedding; the scale of the structures is on the order of a few centimeters. This variation results in hydraulic and transport properties that may vary over several orders of magnitude on this same small scale. Laboratory and some field measurements are likely to be made on a somewhat larger

scale, perhaps 10 cm or more. Exhaustive sampling to determine the exact nature of the subsurface at this scale will be impossible, thus requiring interpolation between measurements and other indirect methods to estimate properties at unmeasured locations. In addition, the simulation scale for most practical applications (and thus the scale of the parameters) is likely to be significantly larger than the measurement scale, from a few tens of centimeters to many meters.

On the right of Figure 2-1 is a photo of sediments excavated from a trench beneath the South Process Pond in the 300 Area. A large fraction of these sediments is made up of cobble-sized material, which makes it difficult to obtain representative samples for laboratory analysis and to conduct field measurements (e.g., of



Figure 2-1. Photographs of (left) a trench face from an excavation in the 200 Area (photograph by John Selker, Oregon State University) and (right) sediments excavated from a trench beneath the South Process Pond in the 300 Area (from Bjornstad 2003), both on the Hanford Site, Washington



hydraulic conductivity). Also present in these sediments are large clasts composed of semi-consolidated, fine-grained sediments eroded upstream and deposited in the 300 Area during flood events (Bjornstad 2003). These are indicated in Figure 2-1 by the white ellipses. Characterization of these clasts, their preponderance, location, and effect on flow and transport is unlikely to be successful in a deterministic framework.

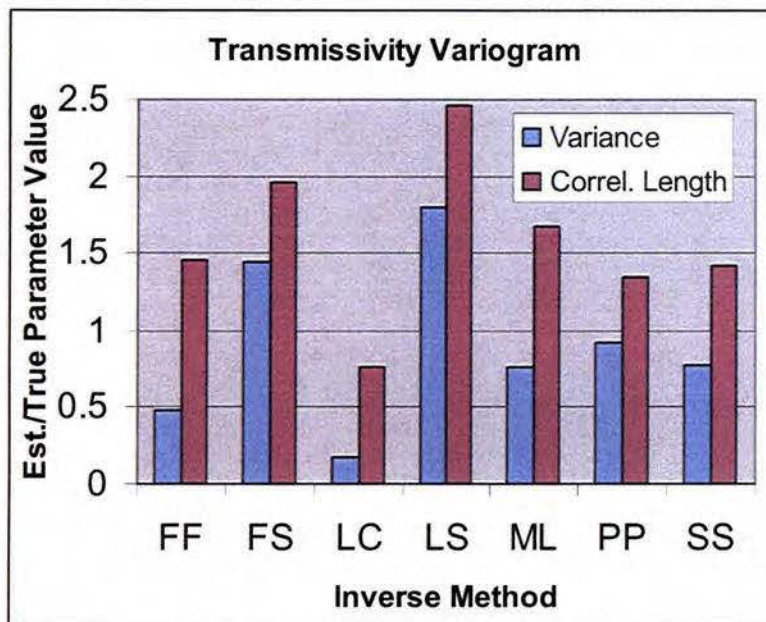
Measurement errors are relatively easy to quantify compared to other sources of uncertainty. While their impact on parameter uncertainty may sometimes be relatively small, Holt et al. (2002) provide evidence that relatively simple measurement errors can introduce significant parameter uncertainties. In their simulations, they also observed that the measurement errors produced spurious parameter correlations, an effect that has likely been poorly appreciated in most applications.

An additional source of parameter uncertainty that has likely not been fully appreciated can be illustrated using results presented in Zimmerman et al. (1998). They compared results from seven models calibrated on the same set of data by different participant groups using different inverse methods. The ratio of estimated to true parameter values for the variance and correlation length of the transmissivity are shown in Figure 2-2 for

each of the inverse methods used. The true transmissivity field was synthetically generated. An exponential model was fit to the average empirical variogram for a set of realizations obtained from each inverse method. The results shown are for Test Problem 1, the simplest transmissivity model used (an isotropic, exponential variogram). Parameter errors reflect two sources of uncertainty: use of different inverse methods and applications by different experts. The latter resulted in differences in model conceptualization for each inverse method, which contributed to the significant differences in parameter estimates. Interestingly, one of the conclusions of this study was that the inverse methods applied did not adequately assess the prediction uncertainty. "The total uncertainty could therefore be better described by the results of the ensemble of several methods, as any one single method in general tends to underestimate the uncertainty." (Zimmerman et al. 1998, pg. 1405) This observation is consistent with the thesis of this report, that including conceptual model uncertainty (in this case, using an ensemble of inverse methods) improves estimates of predictive uncertainty.

### 2.1.1 Analysis of Parameter Uncertainty

The analysis of parameter uncertainty has received much attention in the literature. Helton (1993) and



**Figure 2-2. Ratio of estimated to true parameter values for variance and correlation length of transmissivity for seven different inverse methods. Results from Test Problem 1 of Zimmerman et al. (1998). (FF=Fast Fourier Transform, FS=Fractal Simulation, LC=Linearized Cokriging, LS=Linearized Semianalytical, ML=Maximum Likelihood, PP=Pilot Point, SS=Sequential Self-Calibration)**

McKay (1995) provide discussions of parameter uncertainty that are particularly relevant to dose assessment modeling. See also the recent review of Helton et al. (2006) and the book of Hill and Tiedeman (2007). The primary steps involved in addressing uncertainty in model parameters are

- characterization of parameter uncertainty,
- propagation of parameter uncertainty into model output uncertainty, and
- parameter sensitivity analysis.

To the extent that the parameters considered represent what is unknown about the system and important to predictions, uncertainty analysis based on parameters becomes more useful.

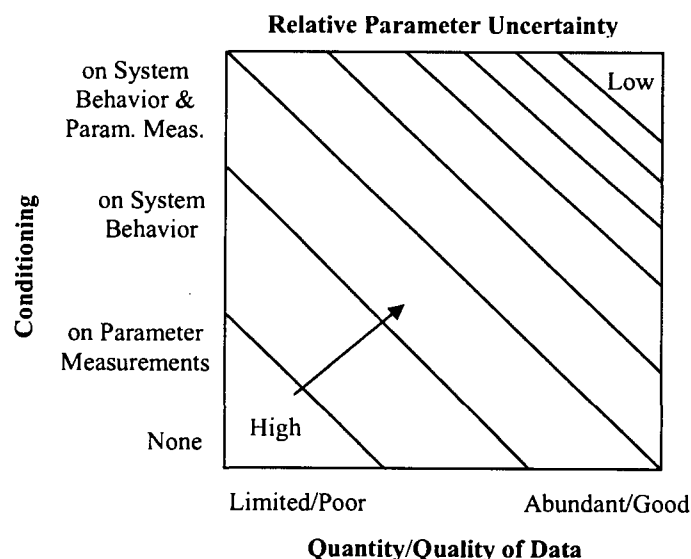
Parameter estimation, including the characterization of parameter uncertainty, is driven by the available data and information. Figure 2-3 is a conceptual representation of the relative parameter uncertainty as a function of the quantity and quality of the data used in the parameter estimation process and the level of conditioning. Conceptually, we expect parameter uncertainty to be reduced as the quantity/quality of data increases and as the level of conditioning increases. No conditioning corresponds to the use of prior parameter estimates. Meyer and Gee (1999) discuss data sources for characterizing prior hydrogeologic parameter uncertainty in the context of dose assessment modeling for license termination decisions. In data-limited applications,

prior parameter probability distributions can also be based on the subjective opinions of one or more experts (Morgan and Henrion 1990).

Meyer et al. (1997) demonstrated the use of information from national-scale databases to specify prior parameter distributions and the subsequent updating of these distributions using site-specific parameter data in a Bayesian approach. This is an example of conditioning on parameter measurements (see Figure 2-3). Another example of this level of conditioning is the application of kriging to interpolate from a set of hydraulic conductivity (or other parameter) measurements.

When observations of state variables (e.g., hydraulic head, radionuclide concentration) are available at a site, formal calibration methods can be used to improve parameter estimates and characterize the uncertainty of these estimates (Carrera and Neuman 1986a; Hill 1998; Hill and Tiedeman 2007). Such inverse modeling is a means of conditioning on system behavior, which can be carried out with or without the inclusion of parameter measurements. Calibrated parameter estimates represent the application of the maximum amount of data/information and yield parameters with the minimum uncertainty. An application to unsaturated flow presented in Wang et al. (2003) illustrates the relationships between the data used in parameter estimation and the resulting prediction uncertainty.

Note that parameter estimates obtained without conditioning or that are conditioned only on parameter



**Figure 2-3. Conceptual response of parameter uncertainty to the quantity and quality of data and the level of conditioning**

measurements may be independent of a model. As discussed in Meyer and Gee (1999), however, there must be a correspondence between the estimates and the parameters assigned those estimates, e.g., a model that has a single value of a parameter representing a site must be assigned a value that represents a mean. Similarly, the uncertainty in that parameter value must represent uncertainty in the mean. Because they rely on an inverse model, calibrated parameter estimates are model-dependent and can be expected to change if the underlying flow or transport model is modified. In fact, most calibration methods assume the model is correct. Errors thus represent the uncertainty in parameters given that the model is correct. This will underestimate parameter uncertainty. Hill (1998) and Hill and Tiedeman (2007) discuss these and related issues.

Zimmerman et al. (1998) evaluated a variety of calibration methods using a set of hypothetical (generated) data based on the Waste Isolation Pilot Plant site. Transmissivity fields for two-dimensional groundwater flow models were calibrated on four test problems. One of their conclusions was that the calibrated models consistently underestimated the “true” variability in transport. The maximum likelihood (Carrera and Neuman 1986a) and sequential self-calibration (Gomez-Hernandez et al. 1997) methods were consistently ranked higher than the other methods. The sequential self-calibration method offers the advantage of producing spatially variable transmissivity fields that honor the spatial statistics of the transmissivity field. A calibrated, stochastic groundwater simulation can be carried out using a set of these fields in a Monte Carlo simulation. Calibration must be carried out for each realization, however. The maximum likelihood method is more general and can be applied to the calibration of a wide variety of parameters, including statistical parameters. Other alternatives include direct calibration of stochastic moment equations, recently demonstrated by Hernandez et al. (2006), and the gradual deformation method of Hu et al. (2001).

Computer codes that can be adapted to the calibration of any simulation model are available (Poeter et al. 2005; Doherty 2004). A method for calibrating geostatistically-simulated parameter fields (similar to the sequential self-calibration method) has recently been demonstrated using PEST (Doherty 2003).

A variety of methods for propagating parameter uncertainty are available, including Monte Carlo simulation, the first-order, second-moment method (Kunstmann et al. 2002; Vecchia and Cooley 1987), the stochastic response surface method (Isukapalli et al. 1998), and stochastic moment methods (Dagan and Neuman 1997; Zhang 2001). Monte Carlo simulation is the most gen-

erally applicable method. The stochastic moment methods are appealing because of their potential computational advantage over Monte Carlo simulation. Recent progress in handling conditions that introduce nonstationarities (Zhang 2001) have made these methods more generally applicable.

Uncertainties must be defined on a site-specific basis and the importance of individual sources may vary site by site or even with different objectives at the same site. Determination of the parameters that are most important to the prediction uncertainty is the final element of an assessment of parameter uncertainty. This is generally carried out through the implementation of sensitivity analysis (Saltelli et al. 2000a, 2004; Helton 1993). Sensitivity measures may also be obtained during the calibration procedure (Hill 1998; Tiedeman et al. 2003; Hill and Tiedeman 2007). Global sensitivity methods (Borgonovo et al. 2003; Saltelli et al. 2000b, 2004; McKay 1995; Hill and Tiedeman 2007) partition the total prediction variance according to the contribution of each parameter and also determine the contribution to prediction variance due to interactions between parameters.

## 2.2 Conceptual Model Uncertainty

As discussed in the previous chapter, hydrogeologic uncertainty may result in valid alternative model structures or conceptualizations. When multiple model conceptualizations are consistent with the available data, it may not be justifiable to rely on a single model structure. Relying on a single conceptual representation of a system has two potential pitfalls: the rejection by omission of valid alternatives, and reliance on an invalid representation by failing to adequately test it. The potential consequences are underestimation of uncertainty by under-sampling model space and biased results by relying on an invalid model.

As mentioned previously, conceptual model uncertainty refers here to uncertainty in both the conceptualization of the system and the mathematical implementation of that conceptualization in a model. Conceptual model alternatives are based on the available site data and other relevant information; each represents a distinct conceptualization of system characterization or behavior. For example, alternative conceptual models might be represented by the presence and absence of leakage from an underlying aquifer; or the presence and absence of matrix-fracture interaction in a fractured rock. Each conceptualization may be implemented in more than one way: for example, a fractured rock may be represented as an equivalent porous medium or as a discrete network of fractures. The process of conceptual-mathematical model development may



be iterative as additional site data becomes available and conceptual models are updated.

### 2.2.1 Analysis of Conceptual Model Uncertainty

Methods for the quantification of conceptual model uncertainty are much less well established than those addressing parameter uncertainty. Mosleh et al. (1994) provide a good introduction to the issues involved. Neuman and Wierenga (2003) discuss a wide variety of issues related to hydrogeologic conceptual model uncertainty, including many instances of its practical importance.

While it is generally possible to specify a reasonable probability distribution representing the complete set of possibilities for the value of a parameter, it is not generally possible to specify the complete set of possible conceptual model alternatives. As a result, conceptual model uncertainty has generally been represented as a discrete distribution, with a small number of model alternatives taken as the complete set of possibilities.

Any approach based on evaluation of a discrete set of alternative models will only be as good as the set of alternatives. That is, if the set of alternatives does not represent the full range of possibilities, conceptual model uncertainty will be underestimated. In Neuman and Wierenga's (2003) extensive discussion of conceptual model uncertainty, they provide some advice on the generation of alternatives, summarized as follows.

- From the assembled database of site-specific data and other relevant information, consider alternative representations of space-time scales, number and type of hydrogeologic units, flow and transport property characterization, system boundaries, initial conditions, fast flow paths, controlling transport phenomena, etc.
- Each conceptual model alternative should be supported by key data.
- Minimize inconsistencies, anomalies, and ambiguities.
- Apply the principle of Occam's window (Jefferys and Berger 1992; Madigan and Raftery 1994) according to which one considers only a relatively small set of the most parsimonious models among those which, *a priori*, appear to be hydrologically most plausible in light of all knowledge and data relevant to the purpose of the model and, *a posteriori*, explain the data in an acceptable manner.
- Maximize the number of experts involved in the generation of alternative conceptualizations.
- Articulate uncertainties associated with each alternative conceptualization.

Because the set of alternative conceptual models is unlikely to represent the full range of possibilities, evaluations of model uncertainty should be viewed as relative comparisons. That is, they may be used to conclude that one model is better than another for the intended purpose, but they cannot necessarily be used to conclude that any model is a good model. In addition, as stated above, the contribution of model uncertainty to overall prediction uncertainty will be underestimated.

Having defined the set of alternatives, the options for addressing conceptual model uncertainty include the following.

- Evaluate each alternative and select the best model. This may be carried out through an informal comparison (James and Oldenburg 1997; Cole et al. 2001b) or through evaluation of a formal model selection criterion (Burnham and Anderson 2002). As mentioned, selection of a single model may not always be justifiable.
- Evaluate each alternative and combine the results using some weighting scheme, such as the likelihood-based weighting of Beven and Freer (2001), the multimodel ensemble approach of Krishnamurti et al. (2000), the model likelihood weighting of Burnham and Anderson (2002), and the model probability weighting of Draper (1995).

Neuman (2003) and Ye et al. (2004) reviewed these and other approaches that have been used to address conceptual model uncertainty. As described in the previous chapter and presented in detail in the following chapter, the method used here is based on model probability weighting.

### 2.3 Scenario Uncertainty

The concept of a scenario has been much discussed in the literature related to performance assessment of nuclear waste disposal facilities. Definitions provided by practitioners working in this area emphasize the notion of scenarios as a set of alternative future conditions for assessing facility performance. Scenario development generally takes place within an iterative process of modeling a system (NEA 2001). The system in this case is the waste disposal facility, the surrounding natural environment, and any external factors acting on these.

In the nuclear waste disposal arena, scenario development is intimately linked with the assessment of features, events, and processes (FEPs) for a system. These terms generally refer to characteristics of the system (features), factors acting on the system (events), and

phenomena governing the behavior of the system (processes). NEA (2000) provides a database of generic FEPs from which site-specific scenarios can be developed through a process of screening and adjustment based on site-specific information. Scenario development and the use of the FEP process are not limited to the assessment of proposed waste disposal facilities. An application of this process for a site-wide assessment of impacts from (numerous) currently contaminated areas and future waste disposal operations at the Hanford Site is given by Solar et al. (2001). In this application, generic FEPs were excluded based on a low probability of occurrence and/or a lack of consequences. Thus only those FEPs that have an impact on the performance criteria were included in the scenarios. Consideration of FEPs was deferred when there was insufficient information to make a decision about inclusion in or exclusion from the analysis. A similar procedure was followed by Moschandreas and Karuchit (2002).

In the context of dose assessment modeling for decommissioning, the U.S. Nuclear Regulatory Commission (NRC) refers to "exposure scenarios," which it defines as "*reasonable sets of human activities related to the future use of the site. Therefore, scenarios provide a description of future land uses, human activities, and behavior of the natural system*" (NRC 2000, p. C26). A number of exposure scenarios are defined in NRC (2000).

As indicated by the above discussion, the term scenario is used to indicate a general statement about possible future conditions. For example, a climate change scenario is a general statement describing a possible change in climate. A residential farmer exposure scenario is a general description of the pathways leading to possible exposure. To evaluate the effect of a scenario, however, the specific characteristics of the scenario must be specified. Climate change, for example, may involve changes to the average and extreme values of temperature and to the timing, spatial distribution, and total amount of precipitation and recharge. A residential farmer scenario involves specification of the extent and rate of irrigation and the location and rate of possible groundwater extraction. In this report we are primarily concerned with the specific characteristics of scenarios. In addition, our interest is limited to the effect of a scenario on the hydrologic aspects of a system. Scenario elements that may affect the hydrologic characteristics of a site include: geological events (e.g., earthquake, landslide), climatic events (e.g., flood, change in precipitation or temperature), changes in engineered components (e.g., surface or subsurface barriers whose hydrologic properties change over time), and human activities (e.g., excavation, well drill-

ing, land use changes that affect the rate and chemistry of recharge). Hydrologic effects of a scenario could include slow changes over time (e.g., a gradual increase in precipitation), or sudden events (e.g., a flood).

We refer to the future hydrologic state of a system as a hydrologic scenario (which we will also refer to simply as a scenario). Regardless of how large or small a system of interest may be, it is always part of a larger system (which makes it open) and made up of smaller systems (which render it complex). To render any reasonable hydrologic predictions for an open, complex system, we must be able to describe its present and future features, hydrologic processes operating within it, and events that drive them. The system's features include the geometric boundaries, geology, hydrogeologic properties, nature and distribution of permeating fluids, topography, physiography, and climate. Hydrologic processes include fluid flow, advective and diffusive chemical and energy transport, phase transitions such as evaporation and condensation, geochemical and biological alterations, and radioactive decay. Events include driving forces acting on the system's external boundaries (e.g., precipitation, evaporation, transpiration, infiltration, chemical spills) and acting internally (e.g., pumping and injection through wells, contamination from nonaqueous liquid sources at depth, contamination through wells or other openings).

No matter what anthropogenic or natural changes occur in a system in the predictable (though uncertain) future, such actions will at most modify one or more of the above three system elements (features, events, processes) in a predictable (though generally uncertain) manner. To postulate hydrologic scenarios for a system thus translates to evaluating the possible future features, events, and processes of the system. To determine the effect of hydrologic scenarios on exposure it is necessary that the features, events, and processes be amenable to mathematical descriptions, i.e., a mathematical model. Typically the features will be expressed in terms of a parameterized description of the system's geometry and its properties, the processes in terms of governing equations that include these parameters, and the events in terms of initial, boundary, and source terms constraining or entering into the governing equations.

If a system is defined (arbitrarily) such that the soil surface is one of its boundaries, then a scenario with conditions acting at or near this surface (climatic conditions, precipitation, evapotranspiration, surface or near surface contamination) would manifest itself in the form of boundary conditions. If a river is modeled (arbitrarily) as a boundary, a scenario defining conditions in the river (stage, contaminant levels) would also

be reflected in the boundary conditions. (It may be better to make the river part of the system being modeled, in which case conditions upstream would form the boundary). In general, the water table would be part of the system and so computed from the system model. If it is treated (arbitrarily) as a boundary, however, then a scenario affecting conditions at the water table would enter as part of the boundary conditions. Similarly, water pH would generally be computed using the system model, but if it is prescribed then a scenario that affects water pH would enter the analysis as a modification to the system features. If a scenario specifies that water is pumped or contamination takes place through a well, this is part of an internal source/sink term. If a scenario specifies that a farmer irrigates a field with contaminated water, this is a boundary condition. If the system is subject to tectonic events such as faulting (often associated with earthquakes), the occurrence of such a scenario would change the system features and thus its parametric description. If a dyke or volcano intrudes into the system, this will again change the system features and its parametric description. If heat generated at a repository causes water to boil near the repository and to condense at some distance from it, this would be reflected in the governing equations.

In summary, a hydrologic scenario can be reduced to a set of conditions described by the three elements of any simulation model: geometry and parameterization (features), structure of governing equations (processes), and driving forces (events). Since we presently know (imperfectly) how to deal with parameter uncertainty, uncertainty in the model structure, and uncertainty in the forcing terms, we know mathematically how to deal with scenario uncertainty. Most of the time a scenario will impact mainly forcing terms, which are often easier to deal with.

### 2.3.1 Analysis of Scenario Uncertainty

Uncertainty in scenarios can be considered on several levels. One approach is to admit the possibility that the future exposure scenario is unknown. Given that the future occurrence of a particular exposure scenario depends entirely on human behavior, however, specifying probabilities of occurrence for exposure scenarios is a speculative proposition and averaging the consequence over multiple exposure scenarios appears unjustifiable. A conservative approach may be more appropriate in this case.

Given a particular exposure scenario, however, there may still be many uncertain elements associated with that exposure scenario. For example, the future infiltration rate depends on future precipitation rates and pat-

terns, whether or not the land is irrigated, and the rate of irrigation if it is used. Future uncertainty in the infiltration rate may affect exposures through pathways involving groundwater migration. These uncertainties in the future hydrologic conditions are more amenable to representation using probability theory and are the aspects of scenario uncertainty to which the methods discussed here are applicable.

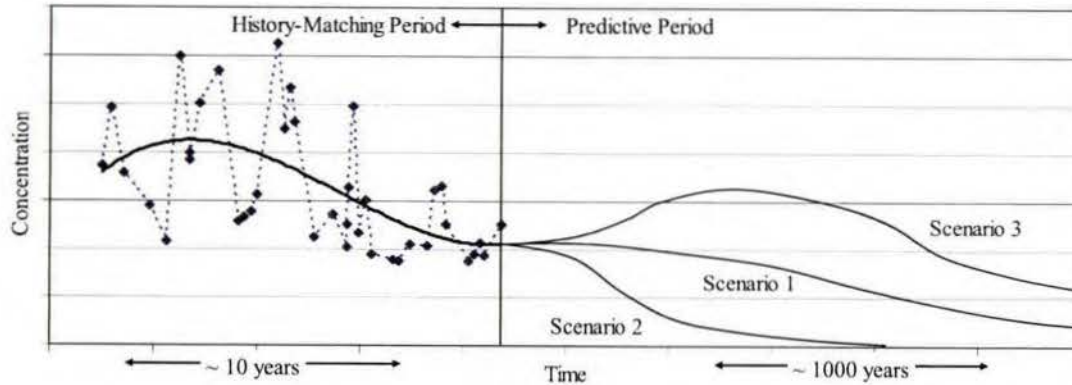
The primary considerations in an analysis of hydrologic scenario uncertainty are (1) determining those future hydrologic conditions that are most important to include in the analysis, (2) characterizing the uncertainty of those conditions, and (3) evaluating hydrologic scenario uncertainty jointly with model and parameter uncertainty.

Identification of future hydrologic conditions that have potentially significant impacts on exposure are best undertaken as part of a model building process that includes exposure scenario development. Established methodologies, such as that described in NEA (2001), can provide guidance on the process. In addition to the use of the International FEP database, NRC (2000) suggests the use of Kennedy and Streng (1992), Shippers (1989), and Shippers and Harlan (1989) for identifying appropriate exposure pathways for decommissioning sites.

Upon identifying the exposure pathways, hydrologic conditions affecting contaminant transport along the pathways can be specified. Consideration of conceptual model, parameter, and model forcing uncertainties during the model building process (i.e., in the history matching period of Figure 1-1) will help to identify uncertainties in the predictive period of model application. Principles articulated in Neuman and Wierenga (2003) for generating alternative conceptual models may be useful in postulating alternative hydrologic scenarios. Among these are the principle of Occam's window according to which one would consider only a relatively small set of the most parsimonious scenarios among those which, *a priori*, appear to be hydrologically most plausible in light of all knowledge and data relevant to the purpose of the model. Figure 2-4 illustrates the modeling framework with scenario uncertainty represented as a set of three discrete scenarios. An additional principle valuable in postulating alternative scenarios is to maximize the number of experts involved in their generation.

Since scenarios describe future conditions, the probability of occurrence must be assigned to each scenario based on a subjective understanding and judgment of the plausibility of the scenario. This does not mean that scenario probabilities need be speculative. It does





**Figure 2-4. Framework for hydrogeologic modeling with scenario uncertainty represented by three alternative scenarios**

mean, however, that there is no opportunity to condition one's belief about the probability of occurrence of a given scenario on observations. In Bayesian terms, scenario uncertainty can be characterized by prior probabilities only. As discussed in the previous chapter, prior probabilities are appropriately interpreted as a subjective measure of the degree of plausibility based on current knowledge, experience, and judgment. Thus the evidence available to establish the probability of a scenario will depend on the specific scenario. For example, established methods are available to estimate the probability of a specific flood occurring within a 1000-year period. No such methods are available for estimating the probability that irrigated agriculture will occur at a site. However, the probability of irrigated agriculture need not be complete speculation. Evaluation of the fraction of nearby land with similar soil and climatic attributes that is under irrigation would be a reasonable basis for establishing the probability of irrigated agriculture occurring at a site, assuming that other factors affecting the occurrence of irrigated agriculture (e.g., climatic, economic, and demographic conditions) remain relatively unchanged.

As discussed in the following chapter, using the methodology presented in this report requires that the alternative scenarios must be mutually exclusive and collectively exhaustive. Although it is likely impossible to prove that a set of scenarios is collectively exhaustive, a relatively small set of scenarios may adequately represent the primary sources of uncertainty in future hydrologic conditions, particularly if the scenarios can be expressed in a general way. An example is a climate change scenario, which may have several impacts on the models. By specifying the scenario general way we avoid having to consider each of the individual impacts separately. Because we require that the set of alternative scenarios is collectively exhaustive, scenario prob-

abilities should be interpreted as relative probabilities (i.e., relative to the other scenarios in the set).

Alternative scenarios are often likely to be characterized as discrete events. Climate change, floods, and introduction of irrigated agriculture are all examples of discrete events affecting the hydrologic conditions at a site. Such events are often not mutually exclusive (e.g., the occurrence of irrigated agriculture does not preclude the occurrence of climate change). By defining scenarios as possible combinations of alternative events, the scenarios can be made mutually exclusive. An example for three events is shown in Table 2-1. A "1" in the table signifies the occurrence of the event in a scenario and a "0" indicates the absence of that event. Scenario 1 in Table 2-1 has none of the events occurring and might be referred to as a reference scenario, perhaps characterized by the continuation of current hydrologic conditions into the future. For  $n$  events, this procedure will result in  $2^n$  scenarios; some of these scenarios may be discarded because of an insignificant probability or because they are not of regulatory concern.

If the scenarios are enumerated from a set of events such as in Table 2-1, the scenario probabilities can be determined from estimates of the marginal and conditional probabilities of the events. Procedures to ensure that the results are consistent with probability theory are available (e.g., De Kluyver and Moskowitz 1984; Brauers and Weber 1988). If the events are independent, the scenario probabilities can be easily computed from the marginal probabilities of the events, as illustrated in Table 2-1. Note that the marginal probabilities for the events characterizing the scenarios may sum to more than 1.0, but the scenario probabilities must total 1.0.



Although this method has been described using a characterization of scenario uncertainty as a set of discrete alternatives, it is possible to consider additional parameter uncertainty specific to a scenario. Consider for example, a scenario that involves a change in land use to irrigated agriculture with a consequent change in the water flux to the soil surface, which represents the boundary of the system. The scenario could specify a

constant irrigation rate. Alternatively, the irrigation rate in this scenario could be modeled as a random variable and included, for example, as a sampled parameter in a Monte Carlo simulation along with any uncertain parameters evaluated in the history matching period.

The following chapter presents a quantitative methodology for assessing the combined impact of parameter, conceptual model, and scenario uncertainties.

**Table 2-1. Example formulation of mutually exclusive scenarios from three scenario-characterizing events. Marginal probabilities for the three events and resulting scenario probabilities assuming independence between events are given.**

		Events Characterizing Scenarios			
		Climate Change (p=0.3)	Flood (p=0.2)	Irrigated Agriculture (p=0.6)	
Scenarios	1	0	0	0	0.224
	2	1	0	0	0.096
	3	0	1	0	0.056
	4	1	1	0	0.024
	5	0	0	1	0.336
	6	1	0	1	0.144
	7	0	1	1	0.084
	8	1	1	1	0.036



### 3 JOINT ESTIMATION OF MODEL, PARAMETER, AND SCENARIO UNCERTAINTIES

A method is presented here to provide an optimal way of combining the predictions of several alternative models and assessing their joint predictive uncertainty, with consideration of parameter, conceptual model, and scenario uncertainties. This method relies on the specification of a set of alternative models and scenarios and weights the alternative model results by a measure of the model probabilities and the alternative scenarios by a measure of the scenario probabilities. The method closely follows the presentation of Draper (1995).

#### 3.1 Bayesian Model Averaging – Combining Conceptual Model and Parameter Uncertainty

A formal method of evaluating prediction uncertainty with consideration of model and parameter uncertainty is Bayesian Model Averaging (BMA) (Draper 1995; Hoeting et al. 1999). Using the notation of Hoeting et al. (1999), if  $\Delta$  is the predicted quantity, its posterior distribution given a set of data  $\mathbf{D}$  is

$$p(\Delta|\mathbf{D}) = \sum_{k=1}^K p(\Delta|M_k, \mathbf{D}) p(M_k|\mathbf{D}) \quad (1)$$

where  $\mathcal{M} = (M_1, \dots, M_K)$  is the set of all models considered,  $p(\Delta|M_k, \mathbf{D})$  is the posterior distribution of  $\Delta$  for model  $M_k$ , and  $p(M_k|\mathbf{D})$  is the posterior model probability for model  $M_k$ . Referring back to Figure 1-5,  $p(\Delta|\mathbf{D})$  is the solid black curve (the model-averaged result of interest) while the individual model results,  $p(\Delta|M_k, \mathbf{D})$ , are shown as the dashed curves. Model probabilities,  $p(M_k|\mathbf{D})$ , are given in the legend of Figure 1-5.

Model uncertainty is represented in (1) by the discrete set of models  $\mathcal{M}$ . Parameter uncertainty enters (1) as the random contribution to

$$p(\Delta|M_k, \mathbf{D}) = \int p(\Delta|M_k, \mathbf{D}, \theta_k) p(\theta_k|M_k, \mathbf{D}) d\theta_k \quad (2)$$

where  $\theta_k$  is the vector of parameters associated with model  $M_k$  and  $p(\theta_k|M_k, \mathbf{D})$  is the posterior prob-

ability density of  $\theta_k$  given  $M_k$  and  $\mathbf{D}$ . Given the joint parameter probability distribution,  $p(\theta_k|M_k, \mathbf{D})$ , (2) could be solved using, for example, Monte Carlo simulation.

Posterior model probability is given by Bayes' theorem,

$$p(M_k|\mathbf{D}) = \frac{p(\mathbf{D}|M_k) p(M_k)}{p(\mathbf{D})} = \frac{p(\mathbf{D}|M_k) p(M_k)}{\sum_{l=1}^K p(\mathbf{D}|M_l) p(M_l)} \quad (3)$$

where  $p(\mathbf{D}|M_k)$  is the likelihood of model  $M_k$  and  $p(M_k)$  is the prior probability of model  $M_k$ . The model likelihood can be expressed as

$$p(\mathbf{D}|M_k) = \int p(\mathbf{D}|\theta_k, M_k) p(\theta_k|M_k) d\theta_k \quad (4)$$

where  $p(\theta_k|M_k)$  is the prior probability density of  $\theta_k$  under model  $M_k$ , and  $p(\mathbf{D}|\theta_k, M_k)$  is the joint likelihood of model  $M_k$  and its parameters  $\theta_k$ .

The model-averaged values of the posterior mean and variance of  $\Delta$  are (Draper 1995)

$$E[\Delta|\mathbf{D}] = \sum_{k=1}^K E[\Delta|\mathbf{D}, M_k] p(M_k|\mathbf{D}) \quad (5)$$

$$Var[\Delta|\mathbf{D}] = \sum_{k=1}^K Var[\Delta|\mathbf{D}, M_k] p(M_k|\mathbf{D}) + \sum_{k=1}^K (E[\Delta|\mathbf{D}, M_k] - E[\Delta|\mathbf{D}])^2 p(M_k|\mathbf{D}) \quad (6)$$

In (6), the first term on the right-hand side represents within-model variance; the second term represents between-model variance. Note that the predictive probabilities (1) and leading moments (5) and (6) are weighted by the posterior probabilities of the individual models.

To apply BMA, one formally requires that the prior model probabilities sum up to one,

$$\sum_{k=1}^K p(M_k) = 1. \quad (7)$$

This implies that all possible models of relevance are included in  $\mathcal{M}$  (i.e., the set is collectively exhaustive), and that all models in  $\mathcal{M}$  differ from each other sufficiently to be considered mutually exclusive (i.e., the joint probability of any two models is zero). In practice, it may be impossible to demonstrate that the set of models is collectively exhaustive. In this case, model uncertainty may be underestimated and model probability must be interpreted as relative to the other models in  $\mathcal{M}$ , as implied by the fact that all probabilities computed using BMA are conditional on  $\mathcal{M}$ .

As discussed in Ye et al. (2004) and Meyer et al. (2004), basing the analysis on a set of model alternatives that do not encompass all possibilities implies a relative comparison between models. We thus interpret prior model probabilities to be subjective values reflecting a belief about the relative plausibility of each model based on its apparent (qualitative, *a priori*) consistency with available knowledge and data. Ye et al. (2005) discuss the use of maximum entropy and expert judgment to determine prior model probabilities.

Whereas prior model probabilities are subjective, the posterior model probabilities are modifications of these subjective values based on an objective evaluation of each model's consistency with available data. Hence, the posterior probabilities are valid only in a comparative, not in an absolute, sense. They are conditional on the choice of models (in addition to being conditional on the data) and may be sensitive to the choice of prior model probabilities (see Ye et al. 2004, 2005 for additional discussion of sensitivity to prior model probabilities). This sensitivity is expected to diminish with increased level of conditioning on data.

### 3.1.1 Maximum Likelihood Bayesian Model Averaging (MLBMA)

Computational difficulties in the BMA approach include the calculation of  $p(\Delta|M_k, \mathbf{D})$  in (1) and

$p(\mathbf{D}|M_k)$  in (3), which may require exhaustive Monte

Carlo simulations of the prior parameter space  $\theta_k$  for each model. This may be computationally and hydrologically very demanding. MLBMA (Neuman 2003; Ye et al. 2004; Meyer et al. 2004) uses two approximations to simplify the application of BMA in hydrogeological modeling. First, approximating  $p(\Delta|M_k, \mathbf{D})$  by  $p(\Delta|M_k, \hat{\theta}_k, \mathbf{D})$ , where  $\hat{\theta}_k$  is the

maximum likelihood (ML) estimate of  $\theta_k$  based on the likelihood  $p(\mathbf{D}|\theta_k, M_k)$ , was suggested by Taplin

(1993) and was shown to be useful in the BMA context by Draper (1995), Raftery et al. (1996) and Volinsky et al. (1997). Second, Neuman (2003) proposed evaluating the posterior model probability using an expression for  $p(M_k|\mathbf{D})$  derived by expanding the terms in the

integrand of (4) in a Taylor series about  $\hat{\theta}_k$  (Kashyap 1982). A related approach based on Laplace approximations has been used in the BMA context by Draper (1995) and Kass and Raftery (1995). Kashyap's expression can be written (Ye et al. 2004) as

$$p(M_k|\mathbf{D}) = \frac{\exp\left(-\frac{1}{2}\Delta KIC_k\right) p(M_k)}{\sum_{l=1}^K \exp\left(-\frac{1}{2}\Delta KIC_l\right) p(M_l)} \quad (8)$$

where

$$\Delta KIC_k = KIC_k - KIC_{min}, \quad (9)$$

$$KIC_k = NLL_k + N_k \ln\left(\frac{N}{2\pi}\right) + \ln\left|\bar{\mathbf{F}}_k(\mathbf{D}|\hat{\theta}_k, M_k)\right| \quad (10)$$

$KIC_k$  is the so-called Kashyap information criterion for model  $M_k$ ,  $KIC_{min}$  is its minimum value over all candidate models, and

$$NLL_k = -2 \ln p(\mathbf{D}|\hat{\theta}_k, M_k) - 2 \ln p(\hat{\theta}_k|M_k) \quad (11)$$

the negative log likelihood of  $M_k$  evaluated at  $\hat{\theta}_k$ .

Here  $N_k$  is the dimension of  $\theta_k$  (number of parameters associated with model  $M_k$ ),  $N$  is the dimension of  $\mathbf{D}$  (number of discrete data points), and  $\bar{\mathbf{F}}_k$  is the normalized (by  $N$ ) observed (as opposed to ensemble mean) Fisher information matrix having components

$$\bar{\mathbf{F}}_{k,ij} = -\frac{1}{N} \frac{\partial^2 \ln p(\mathbf{D}|\theta_k, M_k)}{\partial \theta_i \partial \theta_j} \bigg|_{\theta_k = \hat{\theta}_k} \quad (12)$$

In the absence of prior information about the parameters, one simply drops the term  $-2 \ln p(\hat{\theta}_k|M_k)$  from  $NLL_k$ . This reflects common practice in model calibration.

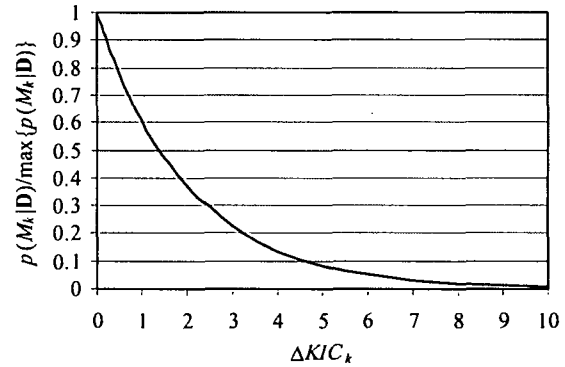
Increasing the number of parameters  $N_k$  allows  $-\ln p(\mathbf{D}|\hat{\theta}_k, M_k)$  to decrease and  $N_k \ln N$  to increase. When  $N_k$  is large, the rate of decrease does not compensate for the rate of increase and  $KIC_k$  grows while  $p(M_k|\mathbf{D})$  diminishes. This means that a more parsimonious model with fewer parameters is ranked higher and assigned a higher posterior probability. On the other hand,  $-\ln p(\mathbf{D}|\hat{\theta}_k, M_k)$  diminishes with  $N$  at a rate higher than linear so that as  $N$  grows, there may be an advantage to a more complex model with larger  $N_k$ .

The last term in (10) reflects the information content of the available data. Among models having an equal number of parameters, which fit a given set of observations equally well, the Fisher information term causes  $KIC$  to favor models with relatively small information content per unit sample or, equivalently, a correspondingly large parameter estimation variance. Looking at this from a different but related angle, one expects a model having a large specific (per unit sample) information content (and small estimation variance) to exhibit improved performance (better fit for a given complexity) and vice versa. If increasing the specific information content of a model fails to improve its performance relative to another model, then according to  $KIC$  the former model has a lesser probability of being correct than does the model with lesser specific information content; selecting a model with greater information content would not be justified. Among models having different numbers of parameters that fit a given set of observations equally well,  $KIC$  may favor more complex models if their specific information content per unit sample is comparatively small.

As shown in Ye et al. (2004), alternative models can have different types and numbers of parameters, but the latter must be estimated and the models compared considering a single data set  $\mathbf{D}$ . As additional data become available, they can be included in  $\mathbf{D}$  and the analysis updated accordingly. For a comparison of two- and three-dimensional models, data distributed in three-dimensional space may need to be projected onto a two-dimensional plane as done by Ando et al. (2003) or averaged in the third dimension as suggested by Neuman and Wierenga (2003, Appendix B).

Previously,  $KIC_k$  has been used (e.g., Carrera and Neuman 1986a,b; Samper and Neuman 1989a,b) as an optimum decision rule for the ranking of competing models. The highest-ranking model is that corresponding to  $KIC_{\min}$ . Note that  $KIC$  has no intrinsic meaning; it is only the differences between  $KIC$  values that have meaning. Thus the use of  $\Delta KIC$  in (8) reflects the in-

terpretation of  $p(M_k|\mathbf{D})$  as a relative probability suitable for comparing the models within the set  $\mathcal{M}$ . The effect of  $\Delta KIC$  on the posterior model probability is shown in Figure 3-1. Model probability is normalized



**Figure 3-1. The effect of  $\Delta KIC$  on posterior model probability given as a fraction of the probability for the best model**

in this figure by the maximum model probability (i.e., the probability of the best model). Relatively small  $KIC$  differences result in large differences in model probabilities due to the exponential in (8). Model probability is less than one percent for  $\Delta KIC$  values greater than 10.

A nonasymptotic version of the familiar Akaike (1974) information criterion ( $AIC$ ) has also been advocated for use in computing relative model weights (Burnham and Anderson 2002). This version of  $AIC$ , derived by Hurvich (1989) can be written as

$$AICc_k = AIC + \frac{2N_k(N_k + 1)}{N - N_k - 1} = NLL_k + 2N_k + \frac{2N_k(N_k + 1)}{N - N_k - 1} \quad (13)$$

Model probabilities can be estimated by computing  $\Delta AICc_k$  values as in (9) and substituting these in (8).

Poeter and Anderson (2005) discuss the use of  $AICc$  in groundwater applications. Appendix C of this report contains an application of MLBMA to recharge model uncertainty that includes a comparison of model probabilities computed using  $KIC$  and  $AIC$ . Ye et al. (2007)<sup>1</sup> discuss alternative criteria for computing model probabilities and present an analysis that sug-

<sup>1</sup> Ye, M., P.D. Meyer, and S.P. Neuman, On model selection criteria in multimodel analysis, *Water Resour. Res.*, in review, 2007.

gests  $KIC$  may be a better model discrimination criterion than  $AIC$  or  $AICc$ .

### 3.1.1.1 Applicability of MLBMA

Using the maximum likelihood method in BMA has several advantages. It can be applied to both complex and simplified models. It can be applied to deterministic models as described by Carrera and Neuman (1986a,b) and Carrera et al. (1997) and also to stochastic models based on moment equations as demonstrated by Hernandez et al. (2002, 2003, 2006). Application of maximum likelihood also yields parameter sensitivity information.

Including prior information in the maximum likelihood calibration is an option that allows one to condition the parameter estimates not only on site monitoring (observational) data but also on site characterization data, from which prior parameter estimates are usually derived. When both sets of data are considered to be statistically meaningful, the posterior parameter estimates are compatible with a wider array of measurements than they would be otherwise and are therefore better constrained (potentially rendering the model a better predictor).

Maximum likelihood yields a negative log likelihood criterion  $NLL_i$  (Eq. 11) that includes two weighted square residual terms: a generalized sum of squared differences between simulated and observed state variables arising from  $-2 \ln p(\mathbf{D}|\hat{\theta}_i, M_i)$ , and a generalized sum of squared differences between posterior and prior parameter estimates arising from  $-2 \ln p(\hat{\theta}_i | M_i)$ . The first is weighted by a matrix proportional to the inverse covariance matrix of state observation errors. The second is weighted by a matrix proportional to the inverse covariance matrix of prior parameter estimation errors. Maximum likelihood allows the statistical parameters of the errors to be estimated. When these statistical parameters are known (i.e., not estimated), maximum likelihood reduces to generalized least squares estimation. In this case, available codes such as PEST (Doherty 2004) and UCODE (Poeter et al. 2005) can be applied.

Maximum likelihood estimation yields an approximate covariance matrix for the estimation errors of  $\hat{\theta}_i$ .

Upon considering the parameter estimation errors of a calibrated deterministic model  $M_i$  to be Gaussian or log Gaussian, one easily determines  $p(\Delta|M_i, \hat{\theta}_i, \mathbf{D})$  in (2) by Monte Carlo simulation of  $\Delta$  through random

perturbation of the parameters. The simulation also yields corresponding approximations  $E[\Delta|M_i, \hat{\theta}_i, \mathbf{D}]$

of  $E[\Delta|M_i, \mathbf{D}]$ , and  $Var[\Delta|M_i, \hat{\theta}_i, \mathbf{D}]$  of

$Var[\Delta|M_i, \mathbf{D}]$ , in (4) and (5). If  $M_i$  is a geostatistical model as in the example of Ye et al. 2004 or a stochastic moment model of the kind considered by Hernandez et al. (2002, 2003, 2006), it yields  $E[\Delta|M_i, \hat{\theta}_i, \mathbf{D}]$  and  $Var[\Delta|M_i, \hat{\theta}_i, \mathbf{D}]$  directly without Monte Carlo simulation.

One final point regarding the applicability of MLBMA. In the most data-limited application, one in which there are no system observations with which to calibrate a model and the only available parameter information is that available from generic databases, (1) reduces to

$$p(\Delta) = \sum_{k=1}^K p(\Delta|M_k) p(M_k) \quad (14)$$

That is, model predictions can still be made using prior parameter estimates and model averaging can still be carried out, but only with prior model probabilities. Since the predictions and model probabilities are not conditioned on state variable observations, however, the results are expected to be more uncertain and potentially more biased.

## 3.2 Incorporation of Scenario Uncertainty

MLBMA predictions are computed using the following procedure. Each model in  $\mathcal{M}$  is calibrated in the history-matching period using the dataset  $\mathbf{D}$ . The calibrated models are then used to simulate the system behavior in the predictive period with each model's result weighted by its posterior model probability. For calibration, the models must reflect the system makeup, processes, and forcing of the history matching period and must be capable of producing the quantities in  $\mathbf{D}$  (typically head and concentration measurements). For prediction, the models must reflect the future scenario(s) and must be able to produce the quantities required to evaluate site safety/performance. This will, in general, require that changes be made to the models between the history-matching and prediction periods. For example, a climate change scenario may require modification of the upper boundary condition representing precipitation or recharge. A residential farmer scenario may require the inclusion in the models of a sink term representing a pumped well. For the purposes of the analysis presented in this section, we assume that all the models in  $\mathcal{M}$  retained for prediction (i.e., those

models with non-negligible posterior model probabilities) were constructed such that they can be easily modified to simulate any scenario considered.

### 3.2.1 Bayesian Model Averaging Conditioned on a Specific Scenario

Formally, scenario uncertainty can be quantitatively assessed jointly with model and parameter uncertainties following the methodology described by Draper (1995) and applied in a nuclear waste disposal context (albeit without the inclusion of model uncertainty) by Draper et al. (1999). Consider an uncertain scenario in which the uncertainty is modeled discretely as a set of alternative scenarios,  $\mathcal{S} = (S_1, \dots, S_I)$ . For a given scenario,  $S_i$ , the posterior distribution of a predicted quantity can thus be interpreted as conditional on that scenario and equation (1) becomes

$$p(\Delta | \mathbf{D}, S_i) = \sum_{k=1}^K p(\Delta | M_k, \mathbf{D}, S_i) p(M_k | \mathbf{D}, S_i) \quad (15)$$

Posterior model probability conditional on a given scenario can be expressed similarly by modifying equation (3).

$$\begin{aligned} p(M_k | \mathbf{D}, S_i) &= \frac{p(\mathbf{D} | M_k, S_i) p(M_k | S_i)}{p(\mathbf{D} | S_i)} \\ &= \frac{p(\mathbf{D} | M_k) p(M_k | S_i)}{\sum_{l=1}^K p(\mathbf{D} | M_l) p(M_l | S_i)} \end{aligned} \quad (16)$$

The simplifications made in the rightmost equality of equation (16) are based on the assumption that the dataset,  $\mathbf{D}$ , is independent of the scenario. That is, the occurrence of any particular scenario in the future does not affect the probability of observing the data,  $\mathbf{D}$ , in the past. As a result, the model likelihoods,  $p(\mathbf{D} | M_k)$ , are not a function of the scenario and do not need to be recomputed under each scenario. In contrast, prior model probability,  $p(M_k | S_i)$ , is potentially a function of the scenarios. That is, the occurrence of specific future hydrologic conditions may have an impact on the relative plausibility of the various models. Thus posterior model probability is a function of the scenario only through the possible dependence of prior model probabilities on the scenario. As in equation (7), prior model probabilities under a given scenario must sum to one.

$$\sum_{k=1}^K p(M_k | S_i) = 1 \quad (17)$$

Posterior mean and variance of  $\Delta$  become (see Appendix A)

$$E(\Delta | \mathbf{D}, S_i) = \sum_{k=1}^K E(\Delta | M_k, \mathbf{D}, S_i) p(M_k | \mathbf{D}, S_i) \quad (18)$$

$$\begin{aligned} Var(\Delta | \mathbf{D}, S_i) &= \sum_{k=1}^K [Var(\Delta | M_k, \mathbf{D}, S_i) \\ &+ [E(\Delta | M_k, \mathbf{D}, S_i)]^2] p(M_k | \mathbf{D}, S_i) \\ &- [E(\Delta | \mathbf{D}, S_i)]^2 \end{aligned} \quad (19)$$

Equation (19) can be rewritten as

$$\begin{aligned} Var(\Delta | \mathbf{D}, S_i) &= \sum_{k=1}^K Var(\Delta | M_k, \mathbf{D}, S_i) p(M_k | \mathbf{D}, S_i) \\ &+ \sum_{k=1}^K [E(\Delta | M_k, \mathbf{D}, S_i) - E(\Delta | \mathbf{D}, S_i)]^2 p(M_k | \mathbf{D}, S_i) \end{aligned} \quad (20)$$

where the two terms on the right hand side represent within-model and between-model variance for a given scenario.

### 3.2.2 Scenario Averaging

Averaging equation (15) over all scenarios using scenario probabilities  $p(S_i) = p(S_i | \mathbf{D})$  as weights gives

$$\begin{aligned} p(\Delta | \mathbf{D}) &= \sum_{i=1}^I p(\Delta | \mathbf{D}, S_i) p(S_i) \\ &= \sum_{i=1}^I \sum_{k=1}^K p(\Delta | M_k, \mathbf{D}, S_i) p(M_k | \mathbf{D}, S_i) p(S_i) \end{aligned} \quad (21)$$

where  $p(\Delta | \mathbf{D})$  is implicitly conditional on all scenarios and model structures. Probabilities  $p(\Delta | S_i, M_k, \mathbf{D})$  and  $p(M_k | S_i, \mathbf{D})$  can be obtained by Monte Carlo simulation and (16), respectively. For the averaging in (21) we require that the scenarios given in  $\mathcal{S} = (S_1, \dots, S_I)$  are mutually exclusive and collectively exhaustive. That is,



$$\sum_{i=1}^I p(S_i) = 1. \quad (22)$$

The posterior mean of  $\Delta$ , including the effects of scenario uncertainty, is (see Appendix B)

$$\begin{aligned} E(\Delta|\mathbf{D}) &= \sum_{i=1}^I E(\Delta|\mathbf{D}, S_i) p(S_i) \\ &= \sum_{i=1}^I \sum_{k=1}^K E(\Delta|M_k, \mathbf{D}, S_i) p(M_k|\mathbf{D}, S_i) p(S_i) \end{aligned} \quad (23)$$

where  $E(\Delta|\mathbf{D}, S_i)$  is evaluated by (18). The posterior variance of  $\Delta$  is (see Appendix B)

$$\begin{aligned} Var(\Delta|\mathbf{D}) &= \sum_{i=1}^I \sum_{k=1}^K [Var(\Delta|M_k, \mathbf{D}, S_i) \\ &+ [E(\Delta|M_k, \mathbf{D}, S_i) - E(\Delta|\mathbf{D})]^2] p(M_k|\mathbf{D}, S_i) p(S_i) \\ &- [E(\Delta|\mathbf{D})]^2. \end{aligned} \quad (24)$$

Equation (24) can be rewritten as (see Appendix B)

$$\begin{aligned} Var(\Delta|\mathbf{D}) &= \sum_{i=1}^I Var(\Delta|\mathbf{D}, S_i) p(S_i) \\ &+ \sum_{i=1}^I [E(\Delta|\mathbf{D}, S_i) - E(\Delta|\mathbf{D})]^2 p(S_i) \end{aligned} \quad (25)$$

where  $E(\Delta|\mathbf{D}, S_i)$  and  $Var(\Delta|\mathbf{D}, S_i)$  can be estimated by equations (18) and (19) or (20). The first term on the right hand side of (25) is the variance within scenarios; the second term is the variance between scenarios.

By substituting equation (20) into (25), the posterior variance can be rewritten in the manner of Draper (1995) as

$$\begin{aligned} Var(\Delta|\mathbf{D}) &= \sum_{i=1}^I \sum_{k=1}^K Var(\Delta|M_k, \mathbf{D}, S_i) p(M_k|\mathbf{D}, S_i) p(S_i|\mathbf{D}) \\ &+ \sum_{i=1}^I \sum_{k=1}^K [E(\Delta|M_k, \mathbf{D}, S_i) - E(\Delta|\mathbf{D})]^2 \\ &\quad \cdot p(M_k|\mathbf{D}, S_i) p(S_i|\mathbf{D}) \\ &+ \sum_{i=1}^I [E(\Delta|\mathbf{D}, S_i) - E(\Delta|\mathbf{D})]^2 p(S_i|\mathbf{D}). \end{aligned} \quad (26)$$

This expression consists of three terms:

(1) the variance within models and scenarios,

$$\sum_{i=1}^I \sum_{k=1}^K Var(\Delta|M_k, \mathbf{D}, S_i) p(M_k|\mathbf{D}, S_i) p(S_i|\mathbf{D}) \quad (27)$$

(2) the variance between models within scenarios,

$$\sum_{i=1}^I \sum_{k=1}^K [E(\Delta|M_k, \mathbf{D}, S_i) - E(\Delta|\mathbf{D}, S_i)]^2 p(M_k|\mathbf{D}, S_i) p(S_i|\mathbf{D}) \quad (28)$$

(3) and the variance between scenarios,

$$\sum_{i=1}^I [E(\Delta|\mathbf{D}, S_i) - E(\Delta|\mathbf{D})]^2 p(S_i|\mathbf{D}). \quad (29)$$

The equations provided above can be applied to estimate the individual and collective contribution to model predictive uncertainty of parameter, conceptual model, and scenario uncertainties. Parameter and conceptual model uncertainty are considered using maximum likelihood Bayesian model averaging (MLBMA) in the history-matching period (the period for which system state data exist). To incorporate scenario uncertainty, the MLBMA results are repeatedly applied in the predictive period under a set of alternative scenarios. Because the scenarios describe future conditions, the scenario probabilities represent prior estimates and cannot be updated using the (past) system state data. Incorporation of scenario uncertainty using the method described here thus does not require any additional calibration (beyond that conducted in the MLBMA analysis), but does require additional probabilistic calculations. For example, solution of equation (21) could be accomplished using a Monte Carlo simulation of each model within each scenario. This is straightforward, albeit computationally expensive for large or complex numerical models.

As time progresses and additional data are collected, these data can be incorporated in an expanded data set,  $\mathbf{D}$ , and the Bayesian analysis updated accordingly. If of sufficient duration, the additional data may cause one to modify the prior probabilities assigned to the alternative scenarios.

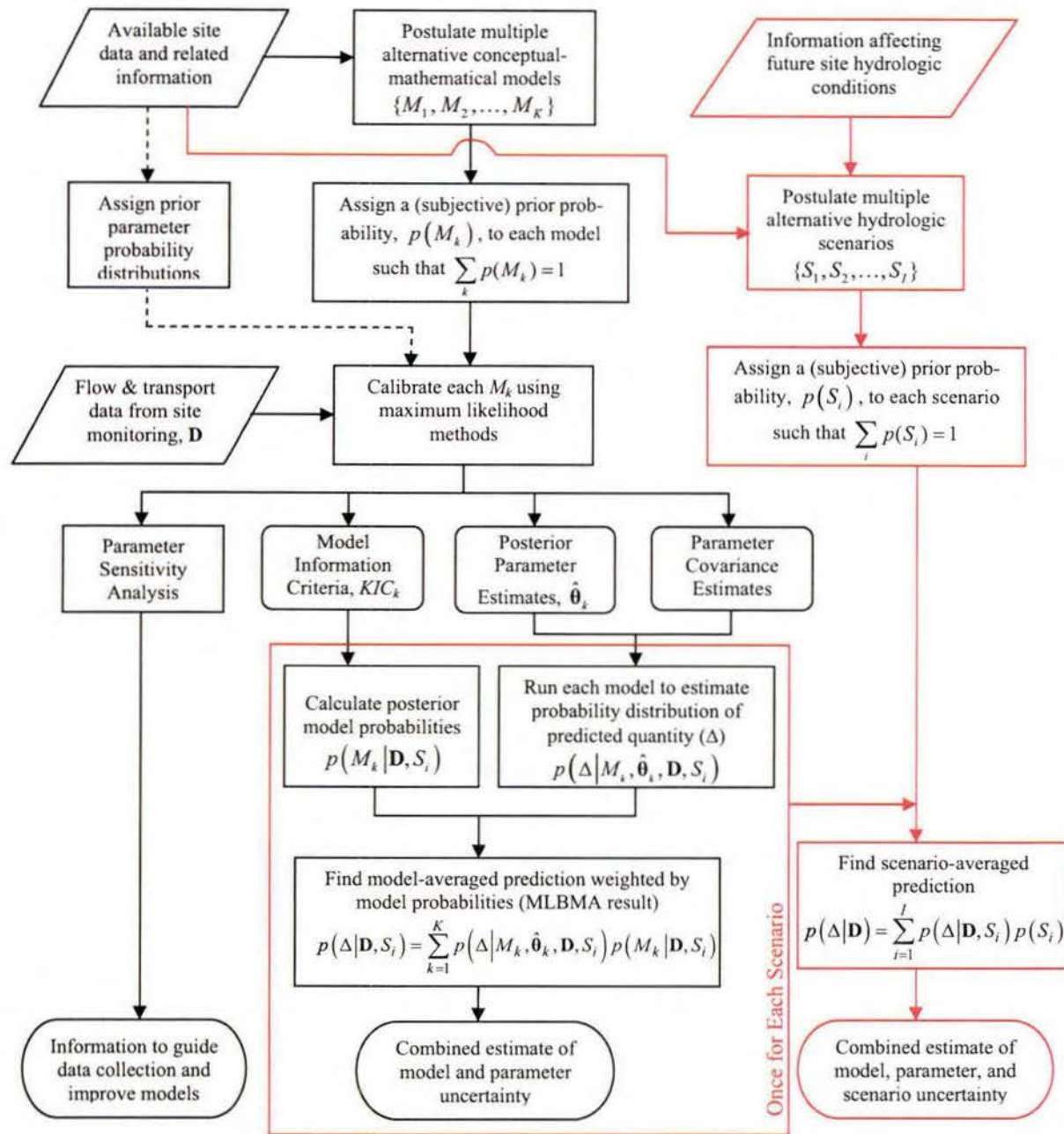
### 3.3 Summary of Uncertainty Assessment Methodology

To implement MLBMA the following steps are followed.



1. Postulate alternative conceptual-mathematical models for a site. Guidance provided in Neuman and Wierenga (2003) may be useful in this step.
2. Assign a prior probability to each model using, for example, the method of Ye et al. (2005).
3. Optionally assign prior probabilities to the parameters of each model, using, for example, guidance provided in Meyer and Gee (1999).
4. Postulate alternative scenarios affecting the future hydrologic conditions at the site.
5. Assign prior probabilities to each scenario using, for example, expert elicitation.
6. Obtain posterior maximum likelihood parameter estimates, and estimation covariance, for each model by inversion (model calibration). In many cases, available codes such as PEST (Doherty, 2004) and UCODE (Poeter et al. 2005) can be applied to this step.
7. Calculate a posterior probability for each model using the model calibration results and the prior model probabilities.
8. Predict quantities of interest using each model.
9. Assess prediction uncertainty (distribution, variance) for each model using Monte Carlo or stochastic moment methods.
10. Weight predictions and uncertainties by the corresponding posterior model probabilities and sum the results over all models.
11. Repeat steps 7-10 for each alternative scenario. Prior model probabilities may be modified for each scenario.
12. Weight the results of step 10 for each scenario by the appropriate scenario probability and sum the results over all scenarios.

A flowchart illustrating the MLBMA approach to combined estimation of conceptual model and parameter uncertainty is shown in Figure 3-2. Elements of the flowchart related to the analysis of hydrologic scenarios are shown in red. As mentioned above, the methodology can be conducted in an iterative fashion; as additional data become available, they can be included in the data set, **D**, and steps 6-12 repeated. In some cases, earlier steps in the methodology may be repeated, such as when additional data indicate that other models or scenarios should be considered.



**Figure 3-2. Flowchart for the combined estimation of model, parameter, and scenario uncertainties using Maximum Likelihood Bayesian Model Averaging. Data boxes contain duplicated data/information and are shown separately to clarify the application of the data/information.**

## 4 EXAMPLE APPLICATION

This chapter describes an application of the methods presented in the previous chapter to assess the impact of conceptual model, parameter, and scenario uncertainty in hydrogeological modeling for dose assessment. The MLBMA methodology for the combined estimation of hydrogeological conceptual model and parameter uncertainty was applied to the modeling of spatial variability in unsaturated, fractured rock by Ye et al. (2004). Seven alternative variogram models of air permeability were evaluated in that application, with the variogram and drift parameters of the models estimated using data from single-hole injection tests. The application illustrated the superiority of using the model-averaged result from MLBMA over any individual variogram model. The application described here advances the uncertainty assessment methodology by including the impact of hydrologic scenario uncertainty and by applying the methods to a problem more directly relevant to dose assessment: namely, transient groundwater flow and transport of a radionuclide from an uncontrolled disposal. The presentation in this chapter generally follows the flowchart of Figure 3-2.

### 4.1 Available Site Data and Related Information

In selecting a site for the application of the uncertainty methodology a number of factors were considered. To avoid any potential conflict-of-interest issues and limitations on the use of data, NRC-licensed sites were not considered. The application site needed to have existing contamination and sufficient historical data to characterize changes in flow and contaminant transport over time, to calibrate the models to be developed for the site, and to evaluate the methodology. It was desired that the contaminant at the site be a radionuclide commonly occurring at NRC decommissioning sites and therefore of concern to the NRC staff. In addition, the hydrologic characteristics of the site should reflect conditions not uncommon at decommissioning sites.

#### 4.1.1 Application Field Site Background

The site chosen for the example application was the 300 Area in the southeastern part of the U.S. Dept. of Energy (DOE) Hanford Site in Richland, Washington (see Figure 4-1). The 300 Area is an industrial area at which uranium fuel was manufactured for use in the reactors located in the northern part of the Hanford Site. In addition to the former nuclear fuel fabrication facilities, the site contains fuels research laboratories, several solid waste burial grounds, and liquid effluent

disposal sites (e.g., process trenches, process ponds) to which uranium and other contaminants were discharged during the production process. The primary discharges of uranium-laden liquid wastes (in terms of liquid volume and uranium activity) occurred at the 316-2 North and 316-1 South Process Ponds (from 1944-1975) and at the 316-5 Process Trenches (from 1975 to 1984). The locations of the 300 Area boundary and the waste sites are shown in Figure 4-2. Also indicated on this figure is the boundary of the 300-FF-5 Operable Unit, a region encompassing potential groundwater contamination and defined for regulatory purposes. In this report, the term “300 Area” will refer to a region that includes that portion of the 300-FF-5 Operable Unit immediately surrounding the 300 Area boundary.

The Hanford Site is located in the semiarid Pasco Basin of the Columbia Plateau in southeastern Washington State, within the rain shadow of the Cascade Mountain Range. The Hanford Site is characterized as a shrub-steppe ecosystem that is adapted to the region’s mid-latitude, semiarid climate (Neitzel 1998). Such ecosystems are typically dominated by a shrub overstory with a grass understory. Livestock grazing and agricultural production prior to government control of the Hanford Site contributed to colonization by non-native vegetation species that currently dominate portions of the landscape. In addition, summer range fires have tended to eliminate fire-intolerant species and have allowed more opportunistic and fire-resistant species a chance to become established. The dominant non-native species on the site is cheatgrass.

Average daily maximum temperature varies from about 4°C in late December and early January to 33°C in late July. On average, there are 48 days during the summer months with a maximum temperature greater than or equal to 32°C. From late-November through early March, minimum temperatures average less than or equal to 0°C. The recorded maximum temperature is 45°C; the recorded minimum is -30°C.

Precipitation at Richland, Washington (National Climatic Data Center Cooperative Station number 457015) has averaged 18.1 cm/yr since 1948, with 52 percent of the annual precipitation occurring from November through February. Days with more than 1.3 cm of precipitation occur on average less than once each year. Seasonal average snowfall is 22.4 cm with a maximum recorded monthly snowfall of 52.1 cm; the maximum recorded seasonal snowfall is 99.8 cm.

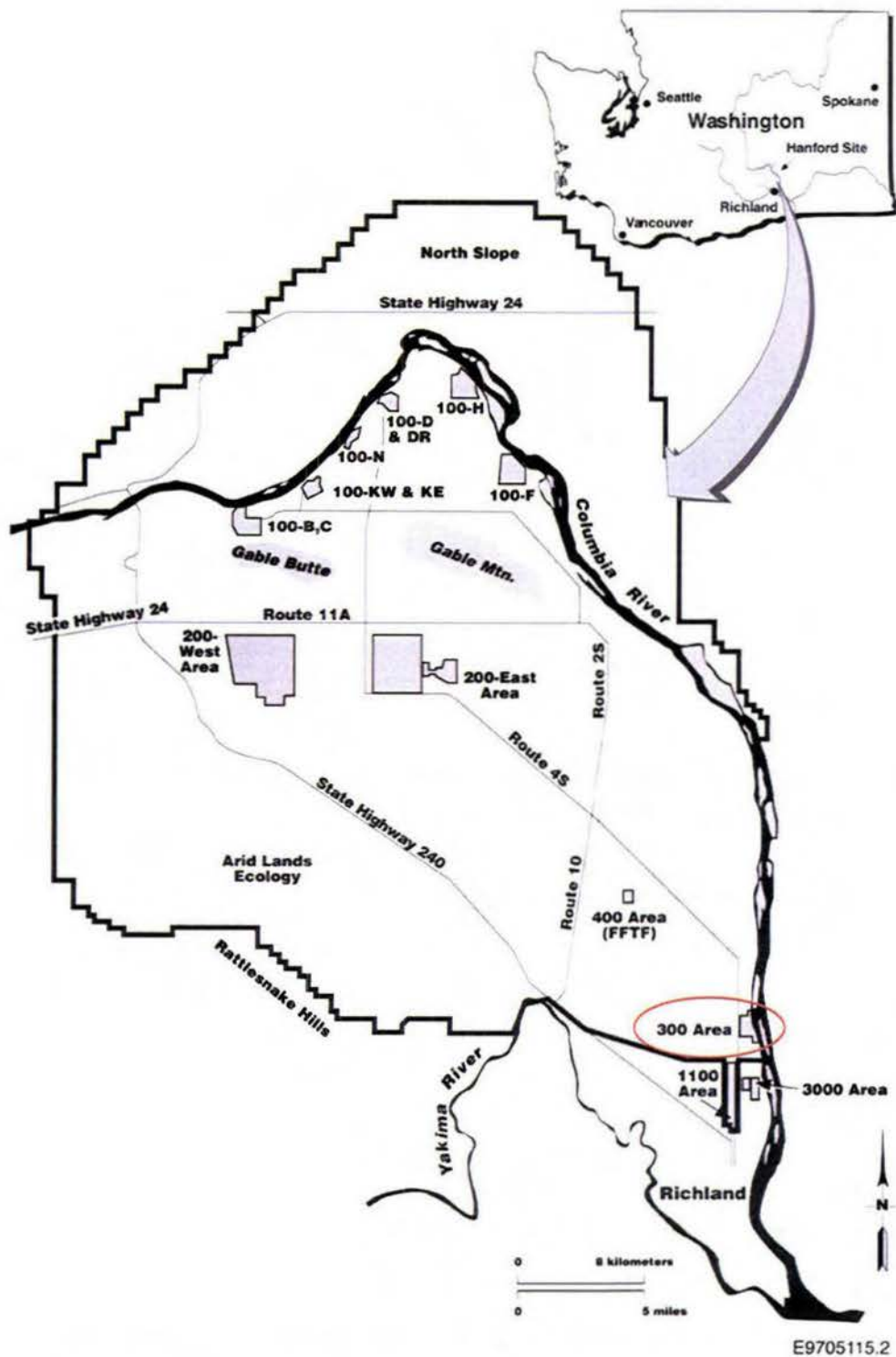


Figure 4-1. The Hanford Site and its location within Washington State. The 300 Area is in the southeastern part of the site adjacent to the Columbia River.



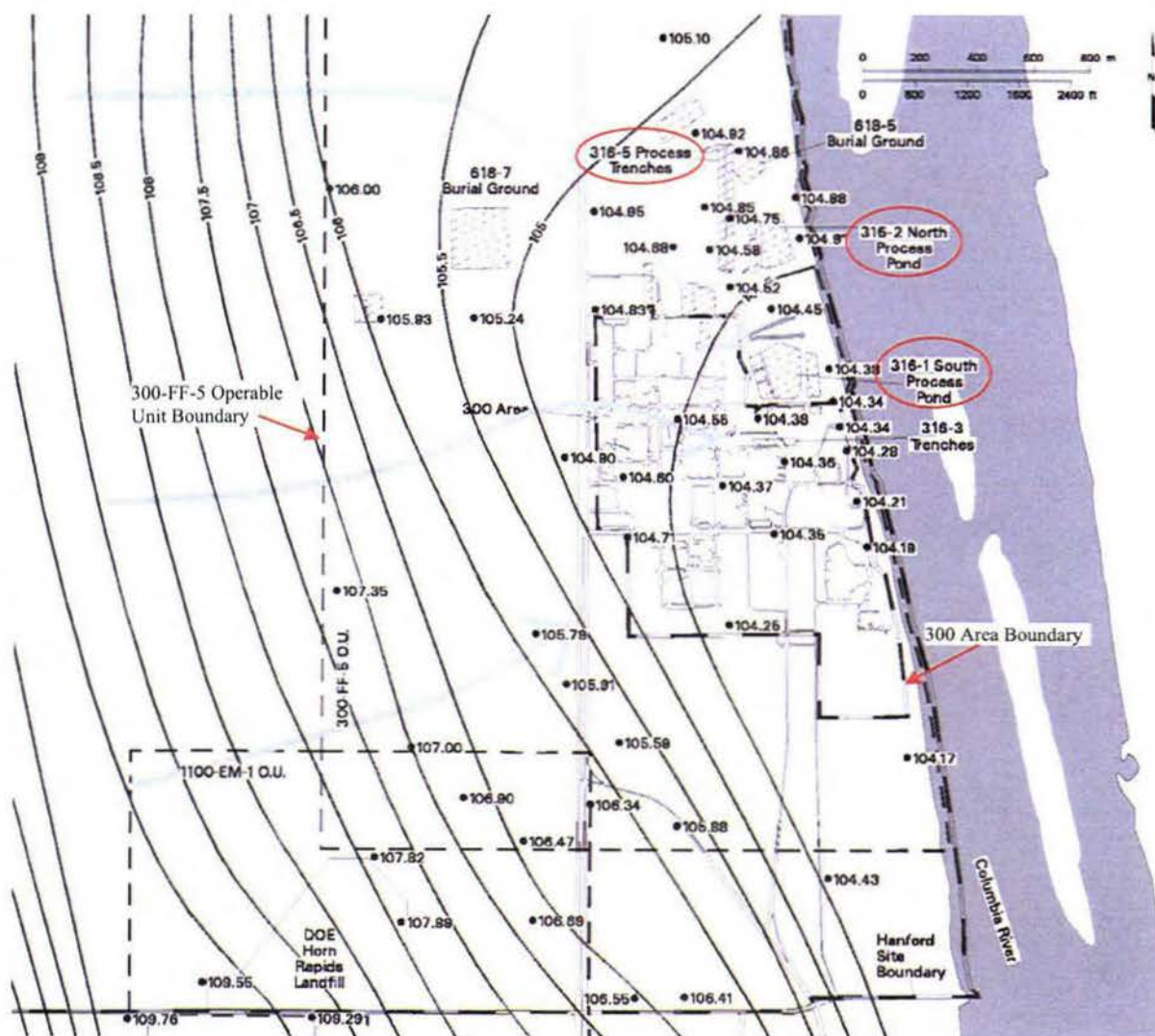


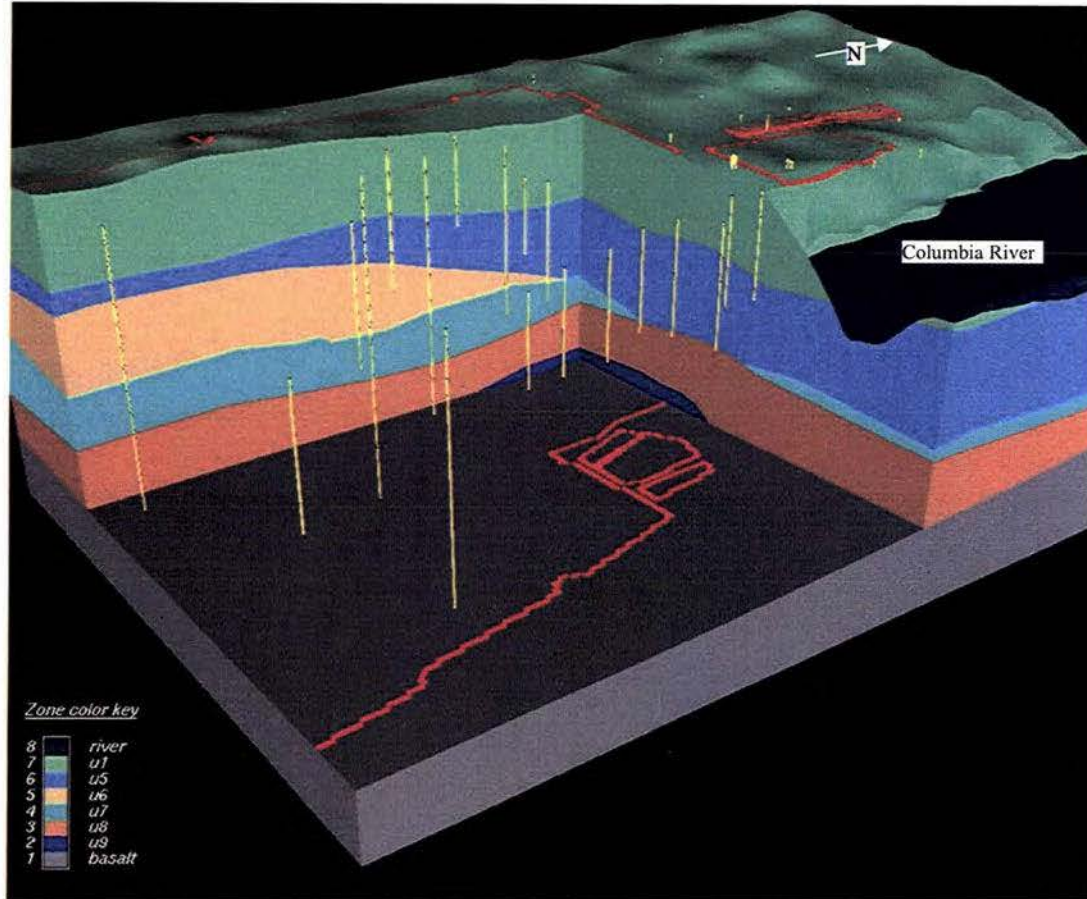
Figure 4-2. Plan view of the 300 Area primary liquid waste discharge sites, boundaries of the 300 Area proper and the 300-FF-5 Operable Unit, and the water table elevation in March 2003 (NAVD88, m) with inferred groundwater flow directions (from Hartman et al. 2004)

The semiarid climate results in fairly low rates of groundwater recharge. Natural recharge rates across the Hanford Site are estimated to range from 0 to more than 10 cm/yr depending on surface soils, vegetation, and topography (Fayer and Walters 1995). Minimal recharge rates occur in fine-textured soils where deep-rooted plants prevail. Larger recharge rates are likely to occur in areas with coarse, gravelly surface sediments and little or no vegetation.

Data from a number of boreholes located in the 300 Area have been used to develop a geologic conceptual

model of the site. The location of these boreholes (vertical yellow lines) and the interpreted stratigraphy at the 300 Area are shown in Figure 4-3. Units shown in Figure 4-3 are

- u1: Hanford Formation flood deposits consisting of high-permeability sand- and gravel-dominated facies,
- u5, u7: Ringold Formation fluvial gravel facies of generally lower permeability than the Hanford Formation deposits,



**Figure 4-3. Stratigraphy of the 300 Area. Red lines indicate the boundary of the 300 Area, the North and South Process Ponds, and the 316-5 Process Trenches, vertical yellow lines indicated the location of boreholes, u1 indicates the Hanford Formation, and u5-u9 indicate units of the Ringold Formation (vertical exaggeration unknown).**

- u6: Ringold Formation overbank deposits consisting of silts and silty sands,
- u8: Ringold Formation lacustrine deposit consisting of a low-permeability clay,
- u9: Ringold Formation fluvial deposit of silty sand and gravel.

The low permeability clay unit (u8) is continuous across the 300 Area, occurs at an average depth of about 30 m, and acts as a confining layer. As occurs across the Hanford Site, extensive basalt units underlie the 300 Area. East-west cross-sections through this geologic model are shown in Figure 4-4; the locations of the western 300 Area boundary, the edge of the Columbia River, and the approximate water table in 2001 are also shown.

Approximately 30 years of groundwater elevation and uranium concentration data exist for wells located in the 300 Area. Average depth to groundwater in the 300 Area is about 12 m with the water table typically occurring in the lower Hanford Formation near the interface with the Ringold Formation. The 300 Area is bounded by the Columbia River on the east with groundwater flow in the unconfined aquifer generally flowing from west to east toward the river, as indicated in Figure 4-2. There is some convergence of flow, however, with hydraulic head measurements indicating a southwesterly flow in the northern portion of the 300 Area and a northwesterly flow in the southern portion. Changes in river stage have been observed to affect heads in the unconfined aquifer as much as several hundred meters away from the river (Campbell et al.



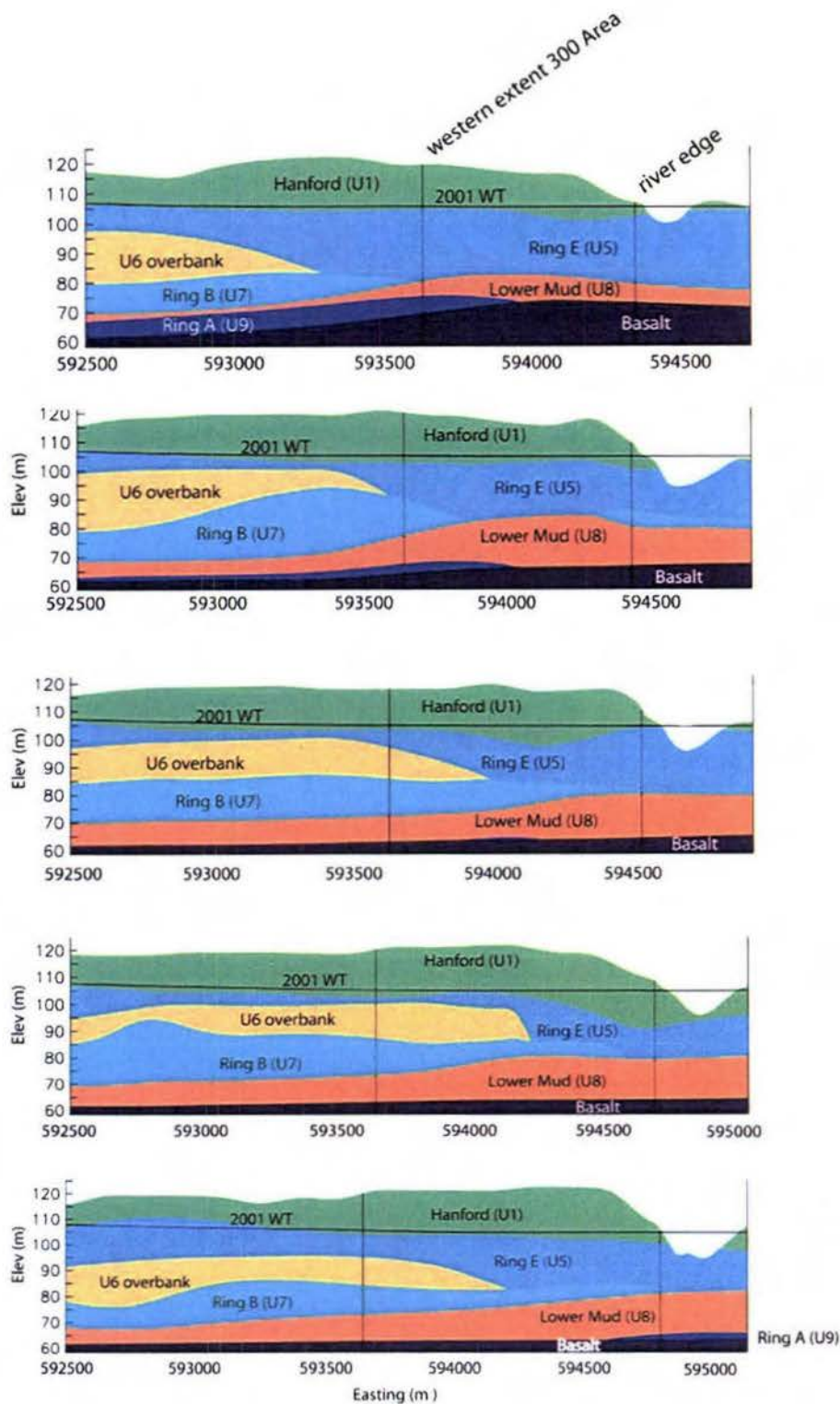


Figure 4-4. East-west cross-sections through the 300 Area at (top to bottom) Northing coordinates 117000, 116500, 116000, 115500, and 115000 m (refer to Figure 4-17 for coordinate locations)





1993; Lindberg and Chou 2001). High river stage can also impact uranium concentrations a similar distance inland from the river (Hartman et al. 2004). Differences in observed uranium concentrations during low and high river stages also have been observed, as shown in Figure 4-5.

An extensive characterization of the 300 Area was carried out beginning in the late 1980s as part of a Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) investigation. Data obtained during this investigation illustrate the complex interaction between site geochemistry, hydrology, and the waste site discharges. After removing contaminated soil from the 316-5 Process Trenches (beginning in 1991), uranium groundwater concentration near the southern end of the trenches was significantly reduced. Uranium concentration in well 399-1-17A is shown in Figure 4-6. This well, whose location is indicated in Figure 4-5, is screened near the top of the aquifer. The concentration remained low as long as uranium-free water continued to be discharged to the trenches. The Remedial Investigation/Feasibility Study report completed during this period (DOE 1995) concluded that

the uranium source was sorption-controlled and that the concentration of uranium in the groundwater would fall below 20  $\mu\text{g/l}$  throughout the 300 Area by the year 2003 as a result of dilution and transport/dilution to the Columbia River. When all discharges to the process trenches ceased (in December 1994), however, the uranium concentration in well 399-1-17A increased rapidly and remains above the 30  $\mu\text{g/l}$  drinking water standard.

Groundwater uranium concentrations continue to exceed the drinking water standard in the 300 Area over an area of about 0.4  $\text{km}^2$  (Hartman et al. 2004), indicating that there is a long-term, slow-release source of uranium present. Lindberg and Chou's (2001) conceptual model attributed this source to uranium remaining in the vadose zone. During high river stages, river water infiltrates the river banks, raising heads in the aquifer and mobilizing uranium present in the vadose zone (see Figure 4-7). Recent geochemical studies of sediments sampled beneath the north and south process ponds indicate that uranium desorption in the vadose zone is a nonequilibrium process (Zachara et al. 2005).

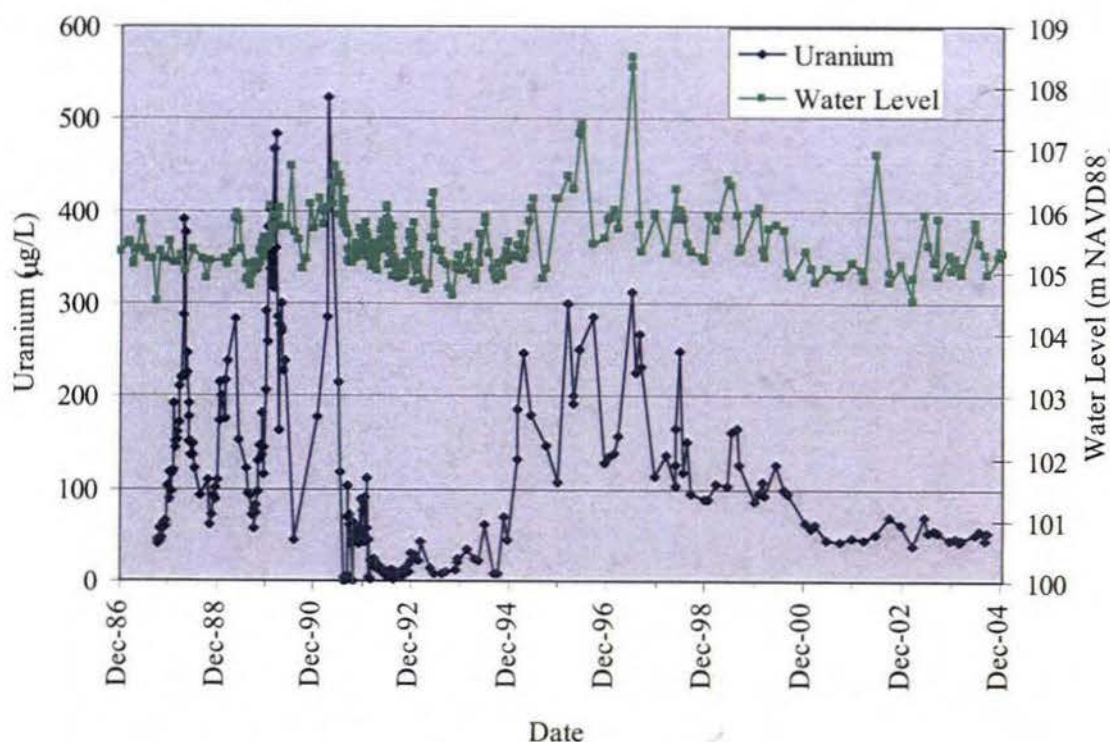


Figure 4-6. Water level and uranium concentration (top of aquifer) in well 399-1-17A near the southern end of the 316-5 Process Trenches. Contaminated soil was removed starting in 1991. Discharge of uranium-free water continued until December 1994.

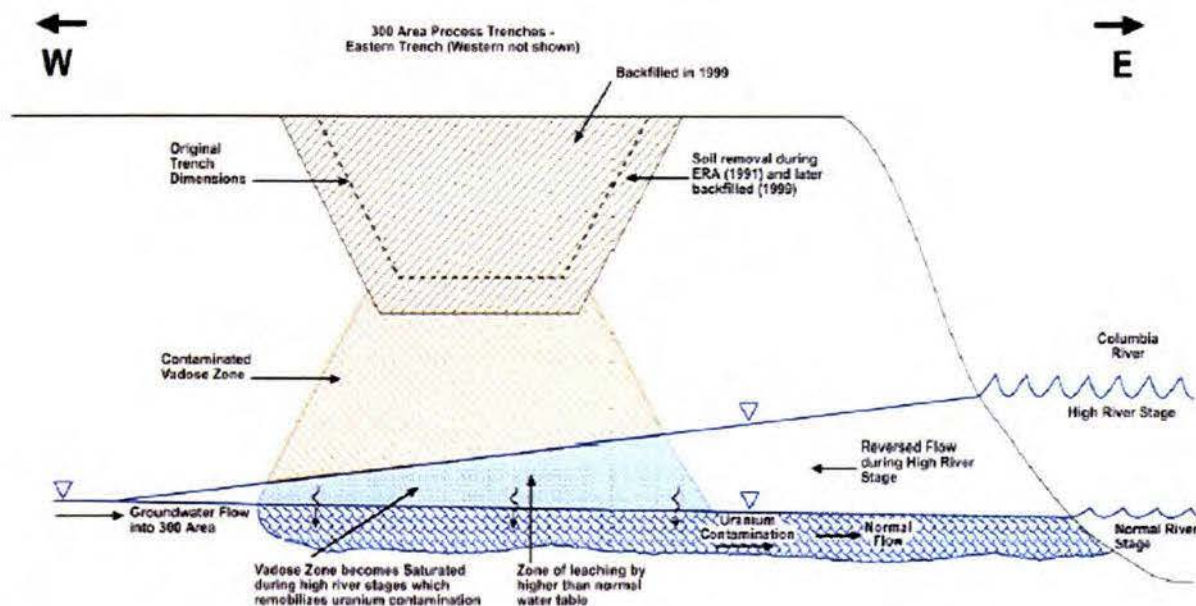


Figure 4-7. Conceptual model of uranium remobilization in the 300 Area during high river stages (from Lindberg and Chou 2001)

#### 4.1.2 Hanford 300 Area Data Sources

This section provides a summary of sources of data that are available to support modeling of groundwater flow and uranium transport in the 300 Area of the Hanford Site. Five categories of data are distinguished:

- surface and borehole geophysics,
- physical and hydraulic properties,
- geochemistry, uranium sorption, and transport,
- source terms and boundary conditions, and
- hydraulic head and uranium concentration data.

Each of these data categories is discussed in the following sections. In addition, ongoing research activities that are being conducted by other groups are briefly described. Data and models developed by these other groups could possibly be used to provide additional information in support of this modeling effort.

Most of the available physical and hydraulic property and borehole geophysics data for sediments within the boundaries of the 300-FF-5 operable unit are contained in two reports and/or associated data archives (Schalla et al. 1988; Swanson et al. 1992). The report by Schalla et al. (1988) is a compendium of technical information on the 300 Area, including sediment and groundwater chemistry data, and hydrogeologic characterization data based on installation of 18 groundwater monitoring wells. Sediment chemistry data from the 300 Area

Process Trenches and North and South Process Ponds are also reported by Schalla et al. (1988; Tables 7.3 and 7.4). The report by Swanson et al. (1992) summarizes characterization data obtained from the installation of 19 additional groundwater monitoring wells that were added to augment the existing 60-well network in the 300-FF-5 operable unit (which includes the wells installed by Schalla et al. 1998). A general summary of these and other available site characterization data is provided in the Phase I Remedial Investigation Report for the 300-FF-5 Operable Unit (DOE 1994).

More recent sampling has been conducted of near-surface sediments, and sediments underlying the North and South Process Ponds, to characterize geochemical conditions and uranium transport behavior in these sediments (Serne et al. 2002; Zachara et al. 2005). Some of these studies are described in more detail later in this section.

##### 4.1.2.1 Surface and Borehole Geophysics

Upon completion of each of the 18 groundwater monitoring wells described by Schalla et al. (1988), geophysical logs were taken using natural gamma, neutron, and density probes. Plots of these data are provided in Appendix A of Schalla et al. (1988), and the original logs are maintained in PNNL data archives.



Spectral gamma surveys were conducted in 8 of the monitoring wells installed by Swanson et al. (1992; Table 4). No artificial radionuclides were detected by the spectral gamma probe in the wells that were surveyed. Gross gamma surveys were also conducted in 11 of the monitoring wells installed by Swanson et al. (1992). The gross gamma data were said to be marginally useful for correlating thick, fine-grained sequences such as the lower Ringold mud unit (Swanson et al. 1992).

The results of surface geophysical surveys that were conducted in the 300 Area are described by Kunk et al. (1993) and Kunk (1993). The objectives of the geophysical surveys were to

- Evaluate the reflective properties of the Hanford and Ringold formations, the water table, and the top of the basalt.
- Determine the existence of a proposed N-S trending paleochannel located near the eastern boundary of the 300-FF-5 operable unit, parallel to the present-day Columbia River.
- Define the lateral extent of the Ringold lower mud unit below the operable unit.

Four surface geophysical techniques were used:

- shallow high-resolution seismic reflection,
- seismic refraction,
- electromagnetic induction (EMI), and
- ground-penetrating radar (GPR).

The shallow high-resolution seismic reflection surveys were the first attempt to use this technique at Hanford. Kunk et al. (1993) indicated that the contacts between Holocene-age eolian soils and Missoula flood deposits, the Hanford/Ringold contact, and the lower mud unit of the Ringold formation were generally able to be mapped using this technique. However, the source frequency and the proximity of the water table to the Hanford/Ringold contact generally made these two features appear as one reflector. This was also found to be the case for the lower mud unit, which is generally within 9 m (30 ft) of the basalt bedrock. Kunk et al. (1993) suggested that a higher-frequency source might be able to distinguish between the water table and the Hanford/Ringold contact, but cautioned that a higher-frequency source would also have lower energy, making it less able to penetrate and resolve features at deeper depths. The seismic data did not show the N-S trending paleochannel that was proposed by Lindberg and Bond (1979).

The EMI and GPR surveys showed a pervasive reflector between depths of 0 to 4 m (14 ft). This feature was

interpreted to be a Holocene-age soil horizon that separates the eolian silts and sands from the underlying Hanford formation (Kunk et al. 1993). This feature was inversely related to the elevation of the ground surface. When topographic corrections were made to the data, this feature became a relatively flat surface (Kunk et al. 1993). This interpreted Holocene-age soil horizon was absent in several areas, which Kunk et al. (1993) suggested could be due to erosion, human disturbance, or that the horizon was never or poorly developed in these areas.

Several locations also showed trough-shaped GPR anomalies which coincided with a missing section of the interpreted Holocene soil horizon. Kunk et al. (1993) suggested that these features could be relatively recent paleochannels, but noted that none of these features were of the size, depth, or magnitude suggested by Lindberg and Bond (1979). They also noted that most of the interpreted paleochannels appear to post-date the Holocene soil horizon. A second reflecting horizon was also detected on several profiles, usually 1.2 to 1.5 m (4 to 5 ft) below the interpreted Holocene soil horizon. This horizon was also interpreted to be a buried paleosol (Kunk et al. 1993).

We speculate that the seeps along the banks of the Columbia River adjacent to the 300 Area (Hulstrom 1993) occur where one or more of these paleosols intersects the shoreline. The capillary barrier effect resulting from these finer-grained paleosols overlying coarser sand and gravel units should promote lateral flow through the finer-grained paleosols. It is also likely that a significant fraction of the residual uranium contamination that exists in the vadose zone in the 300 Area may reside in these finer-grained sediments, due to their greater specific surface area.

#### 4.1.2.2 *Physical and Hydraulic Properties*

All 18 of the wells discussed by Schalla et al. (1988) were drilled by the cable-tool method. Sieve data (sand fraction only;  $0.05 \text{ mm} < \text{particle diameter} \leq 2 \text{ mm}$ ) were collected for 147 samples taken at approximately 1.5-m-depth (5 ft.) intervals from five of the wells, at depths ranging from 1.2 to 55 m below ground surface. Additional particle-size data were collected for selected fine-grained sub-samples from some of these wells using the hydrometer method (Gee and Bauder 1986). Laboratory measurements of vertical saturated hydraulic conductivities were also made on some of these sub-samples. All of the physical and hydraulic property data reported by Schalla et al. (1988) were available only in hard copy. Therefore the sieve data for the 147 samples (from five wells) reported by Schalla et al. (1988) were entered into an Excel® spreadsheet to fa-

cilitate later use in estimating hydraulic and transport parameters and their spatial distribution. These data show gravel (>2 mm particle size) contents ranging from 0 to 84%, with an average of 33%.

Aquifer test analyses were conducted for 13 of the wells installed by Schalla et al. (1988). Calculated values of horizontal hydraulic conductivity for wells screened in the Hanford formation are up to approximately 15,000 m/d, which reflects the very coarse nature of these sediments. Calculated horizontal hydraulic conductivities for the Ringold formation were typically lower, ranging from a few to several hundred meters per day (Schalla et al. 1988; Table D.2). The pump test results were critically evaluated by Spane (1991, unpublished letter report to Craig Swanson, WHC), who determined that the results from about half of the pump tests were invalid due to the methods used for data analysis (semi-log straight-line solutions and log-log type curve matching procedures) and recharge boundary effects resulting from the proximity of some of the wells to the Columbia River. Some of these pump test data have since been re-analyzed.<sup>2</sup>

The report by Swanson et al. (1992) summarizes data obtained during drilling of 19 groundwater monitoring wells. Sixteen of the wells were drilled by the cable-tool method and 3 were drilled by the sonic-drill method. A total of 227 sediment samples were collected during installation of 11 of these wells, at 1.5-m-depth (5 ft.) intervals, using a 10-cm-diameter (4 in.), 60-cm-long (2 ft.), split-spoon sampler with stainless steel or lexan liners. According to DOE (1994, p.2-8), the following physical property tests were performed on the samples collected by Swanson et al. (1992):

- sieve analyses, to determine particle-size distributions for the sand to gravel-sized fractions of the sediments,
- hydrometer analyses, to determine particle-size distributions for the silt and clay-sized fractions of the sediments,
- permeameter tests, to determine vertical saturated hydraulic conductivities, and
- moisture content.

The DOE (1994) report also states that the results of these tests are provided in Swanson et al. (1992). Unfortunately, this is not the case. The report by Swanson et al. (1992) contains the following data:

- $\text{CaCO}_3$  (%),
- moisture content (%),
- porosity (%),
- specific gravity, and
- bulk density.

However, no sieve, hydrometer, or permeameter test data are contained in the report. Laboratory-measured particle-size distribution data for samples from three of Swanson's wells were obtained from the ROCSAN database (<http://vlprod.rl.gov/vlib/app/>). Further inquiries with former staff from the Westinghouse Hanford Company Geotechnical Engineering Laboratory (GEL), which no longer exists, and subsequent searching of archived data records, led to sieve data for 53 samples from 10 of Swanson's wells, and 23 samples from various other locations in the 300 Area. All of the samples in the GEL records were from depths shallower than 10 m (33 ft.) below ground surface.

The other missing data that were supposed to be in the Swanson et al. (1992) report (hydrometer and permeameter test data) have not been located and may never have been reported or published. It should be noted, however, that measurements of vertical hydraulic conductivities on core samples of gravel- and cobble-dominated sediments, such as those found in the 300 Area, are generally considered to be unreliable since the measurements would be biased low due to the very large particle sizes relative to the size of the core barrel. Therefore, even if the permeameter data were located, they would probably not be useful.

The pump test results reported by Schalla et al (1988), Swanson (1992), and others were reviewed. The majority of the pump tests were single well tests; storage coefficients were not estimated for these tests. Swanson (1992) reports specific yield values of 0.37 and 0.016 for two multi-well tests conducted near the western 300-FF-5 boundary. Wurstner et al. (1995) and Thorne and Newcomer (1992) estimated the specific yield of the Hanford Formation to be in the range of 0.1 to 0.3 and that of the Ringold U5 unit to be in the range of 0.05 to 0.2.

Current estimates of average hydraulic conductivities for the different hydrogeologic units underlying the 300 Area are given in Table 4-1 (column 2). Also shown in Table 4-1 are site-wide estimated ranges for the hydraulic conductivities of these same units based on interpretation of measurements (column 3) and inverse modeling (columns 4 and 5). Hydraulic conductivity values for the 300 Area presented in Table 4-1 are representative of the horizontal conductivity. Vertical hydraulic conductivities have generally been as-

---

<sup>2</sup> Personal communication, Paul Thorne, PNNL, March 2005.



sumed to be 0.01 to 0.1 times the horizontal conductivities in Hanford Site groundwater modeling.

#### 4.1.2.3 Geochemistry, Uranium Sorption, and Transport

Cantrell et al. (2003) describe a database of adsorption data measured on Hanford Site sediment samples for contaminants occurring at the site. Measured values for Uranium (VI)  $K_d$  range from approximately 0 to 1000 mL/g. For natural groundwater conditions at the Hanford Site, Cantrell et al. (2003) concluded that uranium adsorption is relatively low; a range of  $K_d$  values from 0.2 to 4 mL/g being appropriate for these conditions.

Krupka and Serne (2002) reviewed the geochemical factors affecting the behavior of uranium in vadose zone sediments. Relevant to the 300 Area uranium plume is their conclusion that "near sources of uranium release, solubility processes are particularly important for those sediments that become partially saturated with water or completely dry between periods of recharge, such as the surface soils and vadose zone sediments. Under these conditions, the concentration of uranium in the residue pore fluids may exceed the solubility limits for U(VI)-containing minerals and/or co-precipitates with other minerals, such as iron oxides. Characterization studies at DOE sites, such as the Hanford, Fernald, Oak Ridge, and Savannah River sites, suggest that sediments and soils contaminated from

disposal or spills of uranium-containing liquid wastes at these sites can contain uranium-containing minerals or co-precipitates" (Krupka and Serne 2002, pg. 6.6). Such conditions apply to the disposal of uranium-containing liquid wastes in the 300 Area and may have contributed to the apparent long-term source of uranium in the vadose zone.

Uranium sorption and transport data from six samples of 300 Area sediments, collected in December 2000 and February 2001, are described by Serne et al. (2002). One of the samples, considered to be representative of uncontaminated background sediment, was collected from the face of an excavated pit located west of the 300 Area. Two samples were collected from the North Process Pond; one from an excavated trench along the southern border of the pond, and the other from the northeast corner of the wall of the pond (Figure 4-8). Three additional samples were collected near a building in the northern portion of the 300 Area proper (the 303-K building). All of the samples were sieved in the field to remove gravel particles greater than 6.35 mm (1/4 in.). Several types of tests were conducted on these samples: column leach tests, batch adsorption and leach tests, and column adsorption/desorption tests (Serne et al. 2002). The column leach tests were performed on all six samples using deionized water to simulate leaching of residual contaminants by percolating rainwater. The batch leach tests were conducted on five sub-samples of the sediments

**Table 4-1. Estimated hydraulic conductivities (m/d) for the hydrogeologic units underlying the Hanford 300 Area.**

Hydrogeologic Unit	Current Average Estimates for 300 Area	Estimated Range for Hanford Site Based on Pump Tests, Slug Tests, and Some Lab Tests <sup>1</sup>	Inverse Model Estimates for Hanford Site <sup>2</sup>	Inverse Model Estimates for Hanford Site <sup>3</sup>
U1 – Hanford	1500	1 – 1e6	2 – 30,000	192 – 37,100
U5 – Ringold sand/gravel	150	0.1 – 200	0.1 – 4,000	3
U6 – Ringold overbank	0.01	0.0003 – 0.09	0.01 – 0.1	
U7 – Ringold sand/gravel	43	0.1 – 200	0.008 – 90	
U8 – Ringold lacustrine	5e-5	0.0003 – 0.09		0.0002
Basalt	5e-5			

<sup>1</sup> Wurstner et al. (1995); Thorne and Newcomer (1992)

<sup>2</sup> Cole et al. (2001a)

<sup>3</sup> Vermeul et al. (2003)

used for the column leach tests, using de-ionized water, actual groundwater from the 300 Area, and simulated vadose zone pore water (10x ionic strength). Batch adsorption tests were performed on one sample using uranium-spiked pore-water simulant, and a uranium-spiked pore-water simulant with 10x ionic strength. Flow-through column adsorption and desorption tests were conducted on one sample using non-spiked and uranium-spiked pore-water simulants, and 10x ionic strength pore-water simulants. Serne et al. (2002) also

measured the particle-size distributions and bulk mineralogical characteristics of their samples and estimated  $K_d$  values from their column and batch data.

Uranium leach tests conducted on the 300 Area samples by Serne et al. (2002) using a simulated pore water solution resulted in equivalent  $K_d$  values between 70 and 300 mL/g. Based on these results and characterization studies, Serne et al. concluded that the release of uranium from the near-surface sediments is dominated



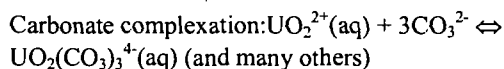
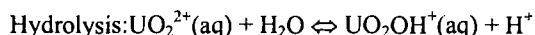
**Figure 4-8. Photos of sediments from the location of samples obtained by Serne et al. (2002) for geochemical analyses at the south side (top) and the northeast corner (bottom) of the North Process Pond**

by dissolution of discrete uranium minerals or uranium-containing co-precipitates, not by desorption of adsorbed uranium. Adsorption tests on uncontaminated samples obtained from the 300 Area resulted in  $K_d$  values that ranged from 0 mL/g to more than 100 mL/g depending on the solution chemistry. Inorganic carbon solution concentration had the greatest impact on adsorption. Solution pH was also important, but is expected to be relatively constant in the field due to buffering by the soil. With solution chemistry held constant, a linear adsorption model appears appropriate over the range of uranium concentrations observed in the 300 Area groundwater. A nonlinear model may be appropriate at the higher uranium concentrations expected in the vadose zone. In response to varying solution chemistry (primarily alkalinity), the  $K_d$  values are expected to vary spatially. Serne et al. (2002) suggested  $K_d$  values of 0 to 1 mL/g in the near-surface vadose zone, 2 to 4 mL/g in the deeper vadose zone and unconfined aquifer, and values at least as high as 7 mL/g in sediments with solution diluted by Columbia River water.

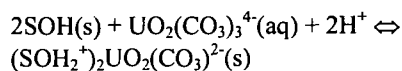
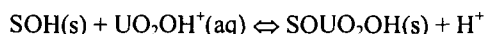
Additional experimental work on uranium sorption and transport in 300 Area sediments is currently being conducted at PNNL and at the U.S. Geological Survey (Zachara et al. 2005). The samples used in these experiments were collected from excavations at four locations: two in the North Process Pond, and two in the South Process Pond. At each location, samples were obtained every 2 ft. to a depth of 4 ft. below the water table. (Note that the Ponds had been excavated to remove potentially contaminated materials; as a result, the excavations proceeded from locations below grade. The water table occurred at 10, 16, 18, and 21 feet below the top of the four excavations.) Figure 4-9 shows the face of one of the excavated pits and sediments at depths where samples were obtained. Geochemical analyses and transport experiments were conducted with sub-samples of the bulk sediments consisting of only the less than 2 mm particle size fraction.

Zachara (2004) suggested that the following geochemical reactions should be considered for U(VI) sorption and transport in 300 Area sediments:

#### Aqueous complexation

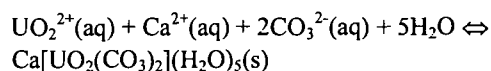
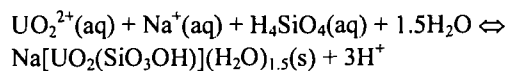


#### Adsorption



where SOH = SiOH, FeOH, AlOH

#### Precipitation



The significance of U(VI) aqueous complexation reactions is that the net charge of the complexed uranyl ion becomes more negative, which could potentially lead to enhanced transport (e.g., via anion exclusion effects). However, the complexed aqueous species may also undergo surface complexation (or adsorption) reactions, thus rendering the uranyl ion less mobile. Various mineral precipitation reactions are also possible. All of these reactions are strongly dependent on water chemistry, which is potentially a very important consideration for the 300 Area due to the differing river and groundwater chemistries.

The experimental results of Zachara et al. (2005) indicate that the degree of U(VI) sorption in 300 Area sediments is dependent on a number of other factors, including dissolved inorganic carbon (DIC), silt and clay content, extractable Fe(III), and chlinochlore (chlorite) content. The calculated  $K_d$  values ranged from 0.27 to 38.4 mL/g, with higher values generally being found for finer-grained samples. In an earlier study, Serne et al. (1992) found that U(VI) activity in 300 Area sediments increased with decreasing particle size, presumably due to more reactive surface area, but that most of the activity was actually associated with the coarser size fractions, simply because the bulk of the material in these sediments is very coarse.

The sorption and desorption data of Zachara et al. (2005) reached a common equilibrium point, but significant times were required. Equilibration of samples with artificial groundwaters showed rapid desorption for the first 24 hours of reaction followed by a slower, steady release of uranium for approximately 1 week. It is posited that the latter component is a result of the slow dissolution of U(VI) in precipitated carbonate minerals (Zachara et al. 2005). As a less likely alternative, Zachara et al. (2005) suggest that the slow release may also be diffusion of U(VI) from immobile water within sediment micro-porosity.



## South Process Pond - Pit#2

Prepared by:  
Bruce Bjornstad  
Applied Geology  
and Geochemistry  
PNNL

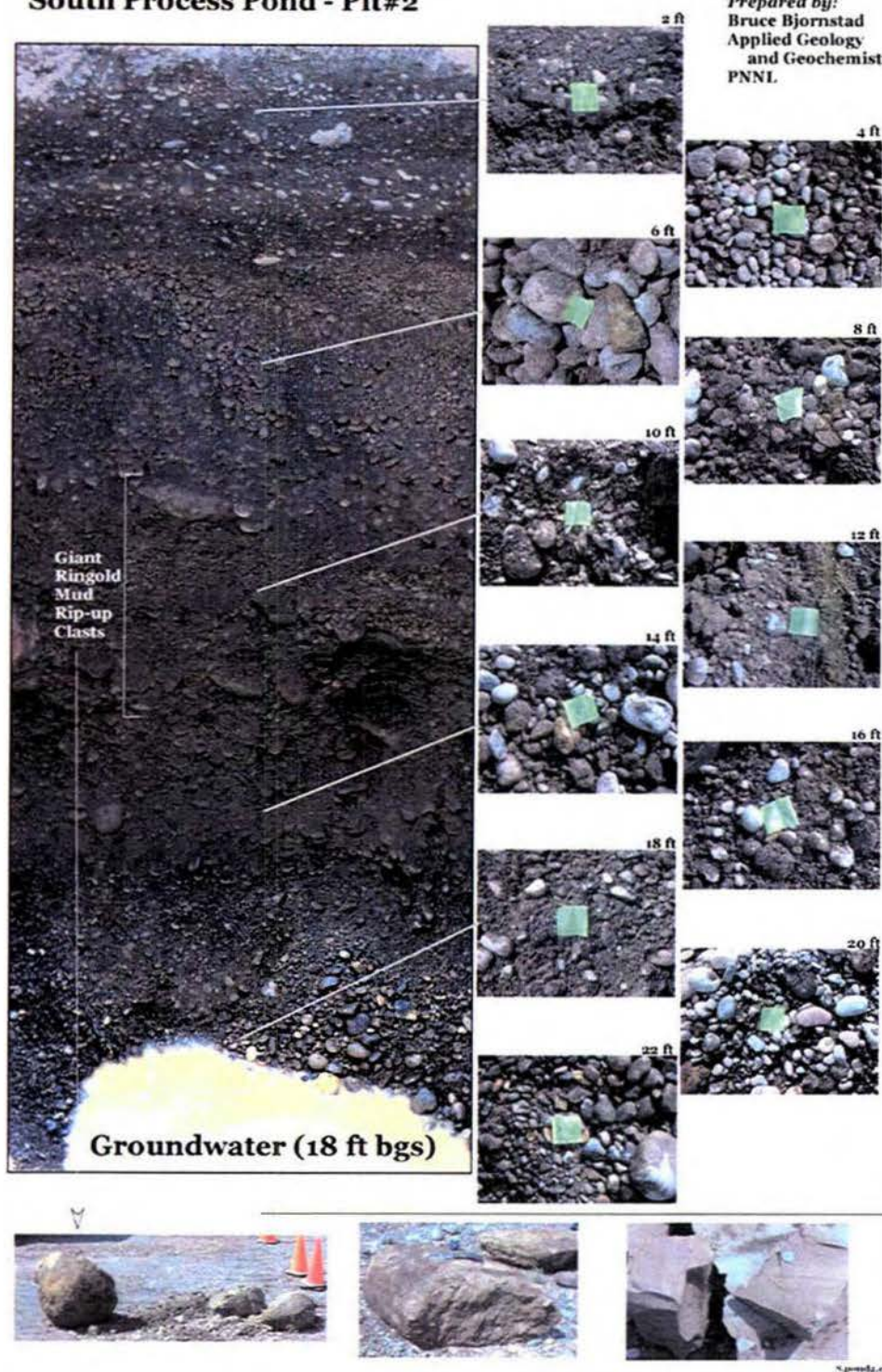
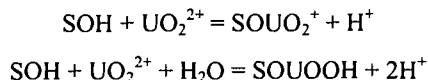


Figure 4-9. Photograph of a pit excavated in the base of the South Process Pond at the Hanford Site 300 Area and close-up photos of sediment at discrete depths where samples were taken (from Bjornstad 2003)

Two models have been developed to represent the sorption/desorption data described by Zachara et al. (2005). One is a distributed-rate, nonequilibrium reactive transport model (Zachara et al. 2005; see also Liu et al. 2004). The other is a semi-empirical surface complexation model using the following surface complexation reactions for U(VI) in 300 Area sediments.



where S refers to surface complexation sites. A total site density of 3.84  $\mu\text{moles}/\text{m}^2$  of BET surface area was assumed; the FITEQL model (Westall 1982) was used to determine optimal formation constants.

Davis et al. (2004) also presented “isotherm” data for several samples. These data exhibited apparent threshold or maximum adsorption capacities, signifying Langmuir-type isotherm behavior. The maximum adsorption capacities appeared to be proportional to measured BET surface areas.

#### 4.1.2.4 Source Terms and Boundary Conditions

The historical uranium releases and associated artificial recharge of water to the 300 Area process ponds and trenches are not well documented. Thus uranium transport from the 300 Area has been evaluated within a site-wide probabilistic modeling framework (Bryce et al. 2003). Records of historical uranium releases and artificial recharge of wastewater to the process ponds and trenches in the 300 Area have been estimated using the Soil Inventory Model (SIM; Simpson et al. 2001). Estimated liquid volume and uranium discharges to the major facilities at the 300 Area are shown in Figure 4-10. There are some discrepancies, however, between the results from SIM and water discharge estimates to the 316-5 facility (North Process Trench) reported by Lindberg and Chou (2001). Such differences are attributable to the uncertainties in the actual waste disposal history of the 300 Area.

Because the 300 Area is adjacent to the Columbia River, groundwater heads near the river are strongly influenced by the river stage, which fluctuates in response to seasonal changes in river flows and operation of the hydropower facilities in the Columbia River basin. An hourly record of river stage data at the 300 Area is available beginning in 1991. A much longer record of streamflow data is available from a gage located below Priest Rapids Dam, the nearest upstream dam (USGS Gage No. 12472800). Monthly average streamflow below Priest Rapids Dam (Figure 4-11) illustrates the historical variability of the Columbia River flow. The apparent change in variance of the

discharge in the early 1970's corresponds with the completion of the three Canadian dams constructed as part of the Columbia River Treaty. The last of these dams, and the largest (Mica Dam), was completed in 1973.

Monthly averages underestimate the actual variability of the river stage as illustrated in Figure 4-12, which shows the monthly and daily averages juxtaposed on the hourly data for the river stage at the 300 Area during 1996. Using monthly average river stage as a model boundary condition results in an error of greater than five feet during peak river stage in 1996 (late February). Daily averages appear to be a much better approximation, although it is evident that there are significant fluctuations on the hourly time scale that would not be represented with a boundary condition based on daily average river stage. The effect of the river stage variability on the groundwater table can be seen in Figure 4-13, which shows the water table under average (March 2000) and high (June 1997) river stage conditions.

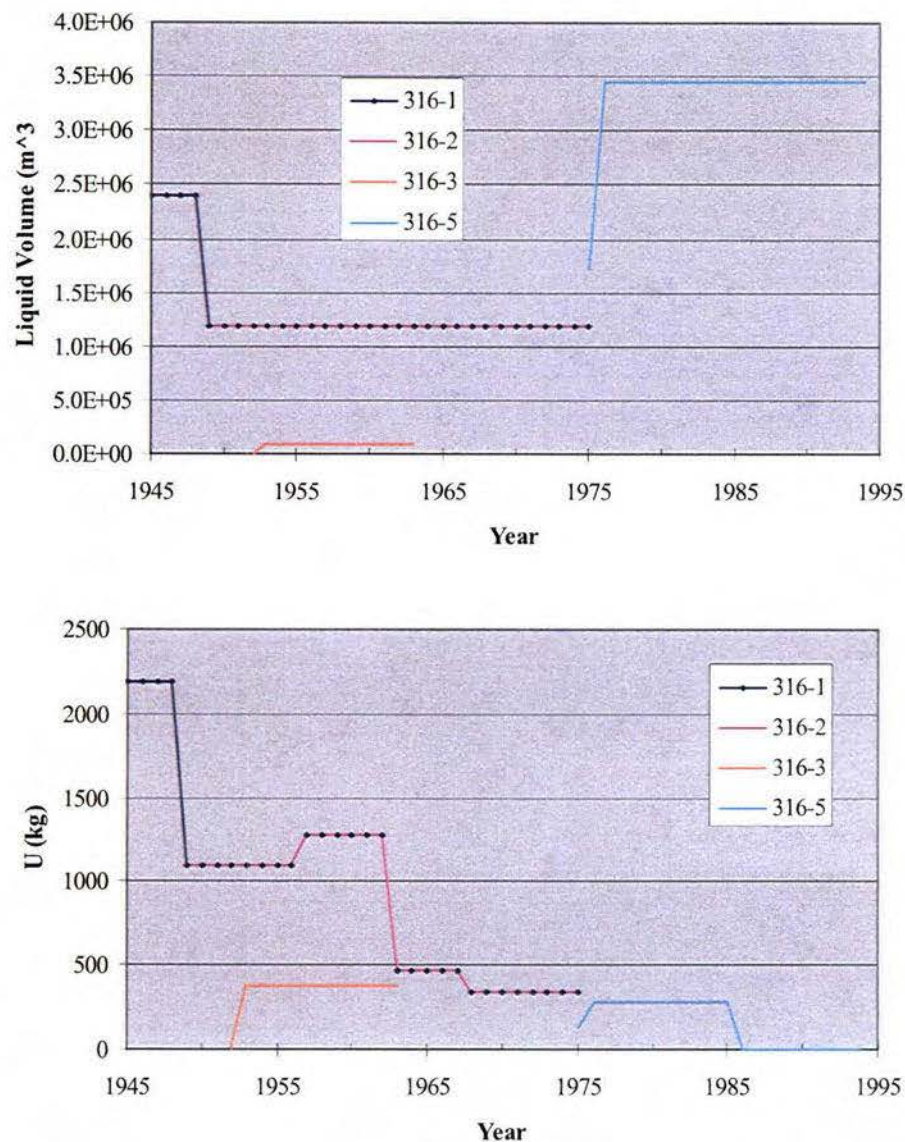
Groundwater velocities shift direction near the river during the high stage condition as discharge from the aquifer to the river is reduced and recharge to the aquifer from the river increases. Note that the high river stage will saturate unsaturated sediments and produce three-dimensional flow near the river. Simulations conducted by Waichler and Yabusaki (2005) indicate that the impact of the river boundary on a tracer extended much farther inland when using hourly river data than when using the monthly data. Computational requirements of modeling may increase significantly as the period of the boundary condition is reduced.

Waichler et al. (2005) used Columbia River discharge data and river stage measurements to calibrate a one-dimensional hydrodynamic model of the Hanford Reach of the Columbia River for the period of 1940-2004. This model can be used to generate hourly river elevations along the boundary of the 300 Area.

#### 4.1.2.5 Hydraulic Head and Uranium Concentration Data

Hydraulic head data measured in groundwater wells in the 300 Area are available from 1950 to the present (see Figure 4-14). The locations of measured heads and the measurement intervals have varied over that time with significant periods of no data in most wells. Because of the high frequency variation in river levels adjacent to the 300 Area and the rapid response of groundwater heads to the river level (particularly near the river), the frequency of head measurements is often



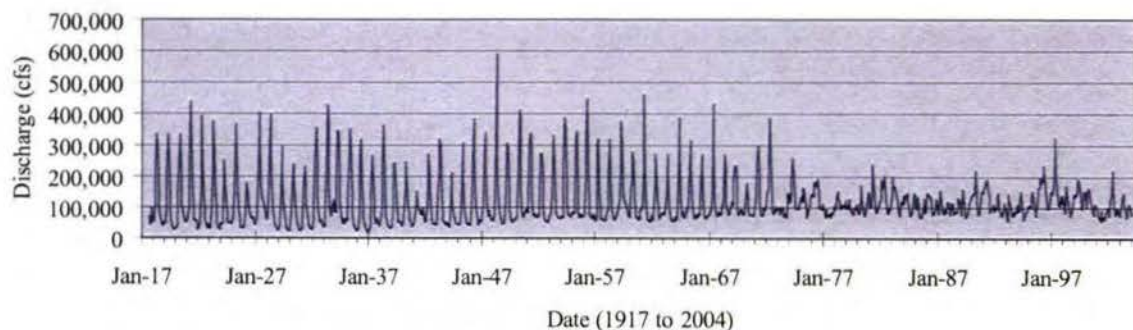


**Figure 4-10. Estimates of liquid volume (top) and U-238 discharged to the 316-1 South and 316-2 North Process Ponds, the 316-5 Process Trenches, and the 316-3 trenches**

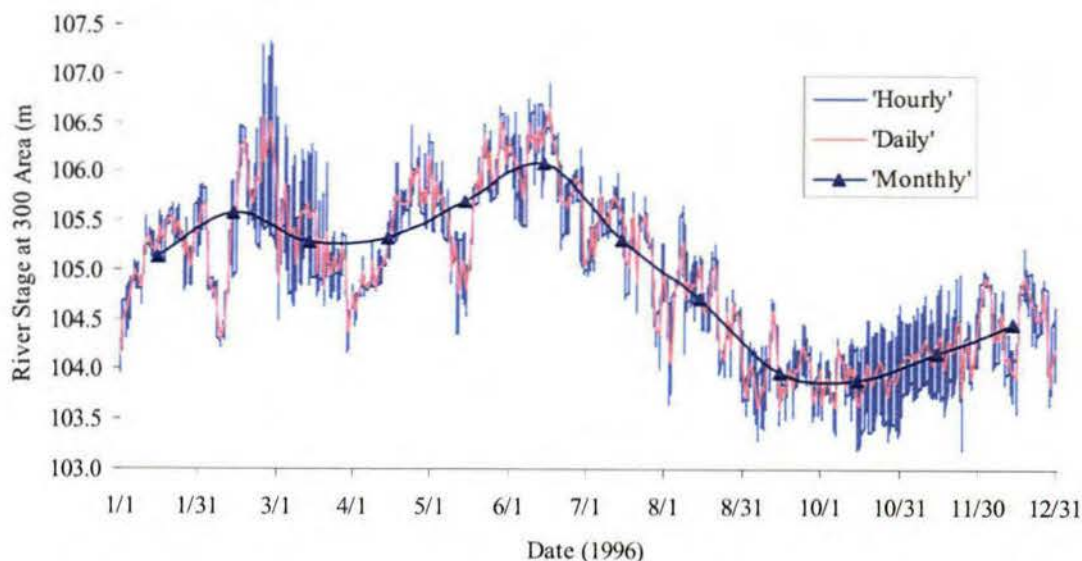
insufficient to accurately represent actual head variation in time.

Campbell et al. (1993) instrumented a network of wells in the 300 Area with pressure transducers and data loggers. During parts of 1991-1993, hourly head data were collected from this well network and were automatically transferred from the field to the lab by radio-telemetry. Hourly head measurements have also recently been resumed in a network of wells. These data

are being used in conjunction with the pump test results from Schalla et al. (1988) and Swanson (1992) to develop and calibrate a three-dimensional flow model in support of a review of the Record of Decision for the 300 Area groundwater (300-FF-5 operable unit). This model is similar in spatial extent to the models used here, but has differences due to incorporation of the most recently obtained data and the specific requirements of the model to support the Record of Decision.



**Figure 4-11. Monthly discharge as measured below Priest Rapids Dam, the nearest dam upriver from the Hanford Site 300 Area**



**Figure 4-12. Hourly, daily average, and monthly average river stage at the 300 Area in 1996**

Groundwater samples are collected quarterly or semi-annually from selected monitoring wells in the 300 Area for chemical analyses, including uranium. There are scattered measurements of uranium concentrations in groundwater in 1959, 1967-68, and the early 1970s, but more regular measurements did not begin until the late 1970s (see Figure 4-14). Figure 4-15 shows the extent of uranium contamination of the groundwater in 1959 during discharge to the North and South Process Ponds. Figure 4-16 shows a time series of uranium concentration in groundwater from 1977 to 2004.

Uranium groundwater concentration data from the 300 Area have recently been analyzed using geostatistical methods to generate estimates of the total mass of uranium in 300 Area groundwater at selected times and the uncertainty of these estimates (Murray et al. 2004).



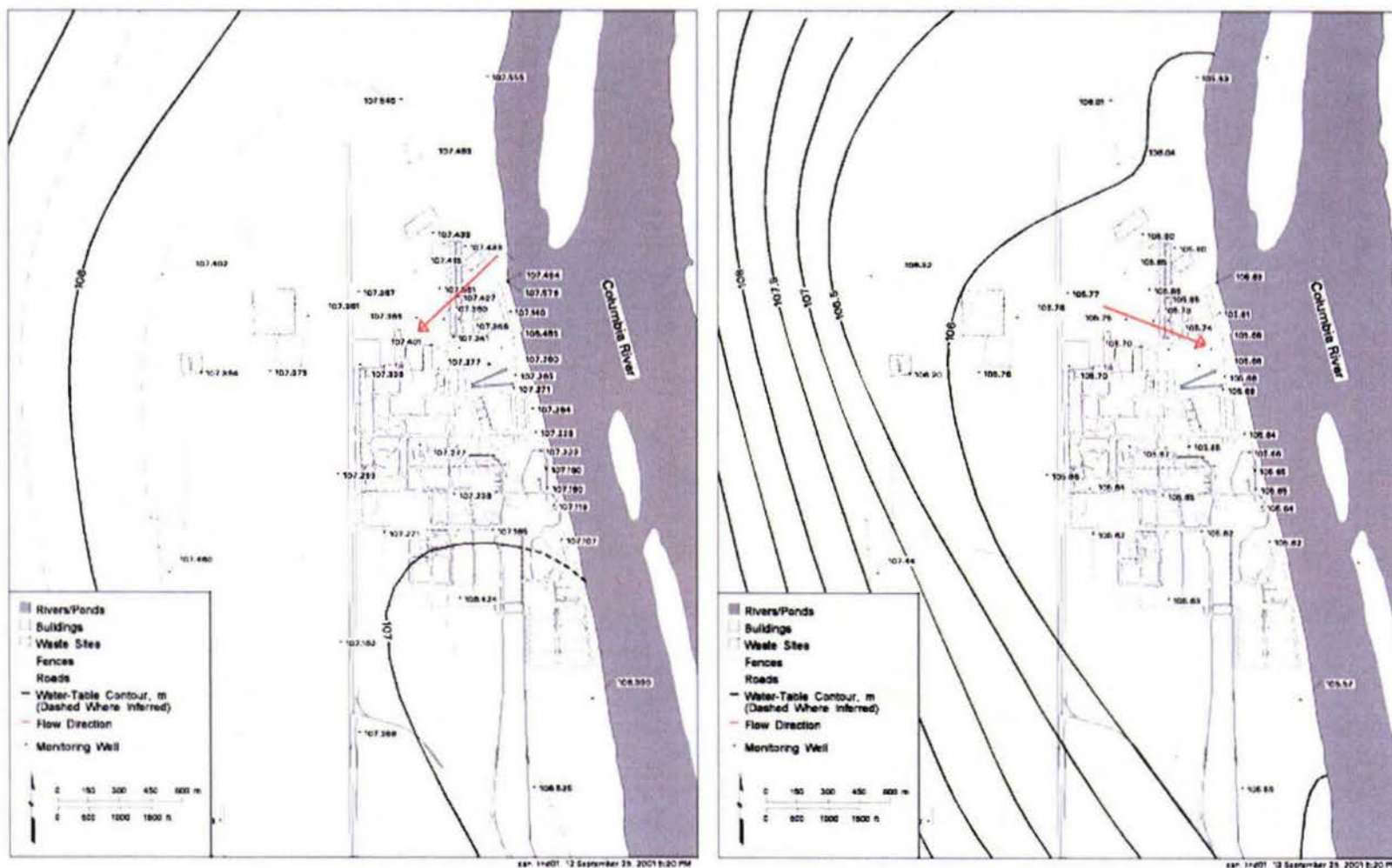
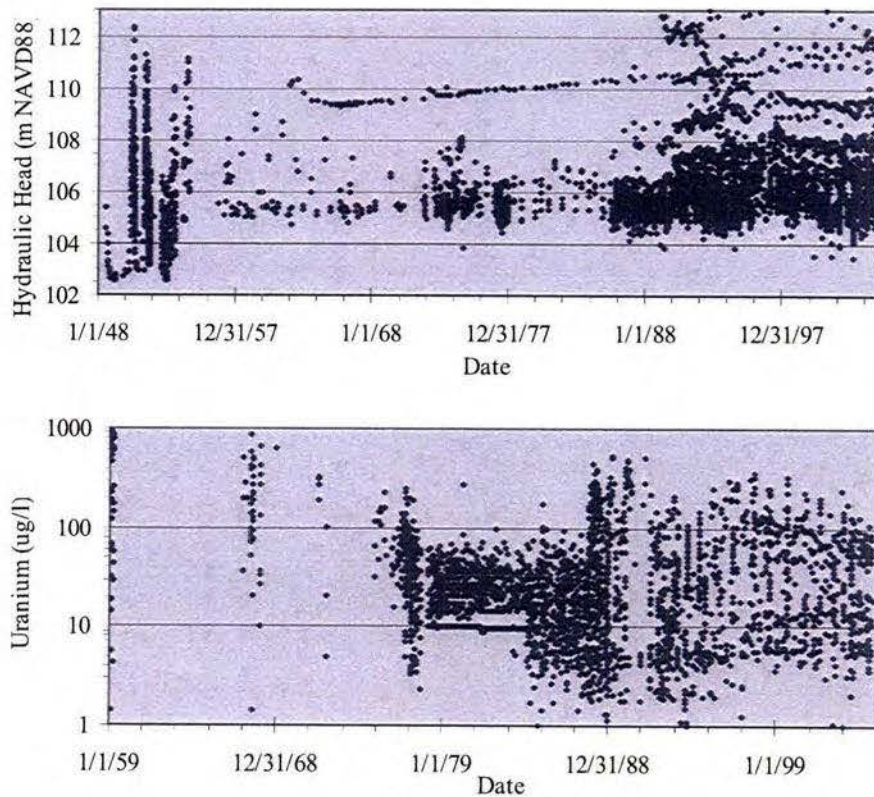


Figure 4-13. Inferred water table in the 300 Area during (left) average river stage and (right) high river stage. Groundwater movement in the uranium source region is indicated by the red arrow. (from Lindberg and Chou 2001).

Chemical analyses were conducted on sediment samples from some of the wells installed by Swanson et al. (1992). Sediment concentration data for selected radionuclides (including uranium), inorganic metals, cyanide, and volatile organic compounds for 51 samples from eight of these wells are reported by DOE (1994; Appendix B). These analyses were conducted on four to eight samples from each well, with sample depths ranging from 8.2 to 54 m (27 to 177 ft) bgs. The U-238 data from these samples all have concentrations less than 3.5  $\mu\text{g/g}$ . The concentration of U-238 in an “uncontaminated background sample” collected by Serne et al. (2002; sample B11493) was about 5  $\mu\text{g/g}$ . Therefore the sediments around the process trenches and ponds appear to be at or below background concentrations. Serne et al. (2002; p. iv) state, however, that “less than 4% of the existing uranium in the contaminated near-surface sediments readily leaches into simulated rainwater over a period of 6 months.” This statement suggests that it may be difficult to accurately determine the amount of ura-

nium that still resides in the unsaturated zone in the 300 Area. Ongoing data collection efforts are directed at better defining the inventory of contaminants in the unsaturated zone.

Time-dependent head and uranium concentration data were used in model calibrations. There are no time-dependent moisture content or pressure head measurements in the 300 Area vadose zone. The only uranium concentration data available in the vadose zone are from samples collected for geochemical analysis, which represent a single point in time. Total uranium concentrations for the sediment samples collected from the 300 Area by Serne et al. (2002) ranged from 5 to 989  $\mu\text{g/g}$ , and for the samples collected from the excavations in the North and South Process Ponds by Zachara et al. (2005) from 5 to 238  $\mu\text{g/g}$ . Zachara et al. (2005) estimated that the fraction of uranium available to desorb or dissolve was 4% to 8% for the upper samples and 8% to 67% for the deeper vadose zone samples.



**Figure 4-14. Available data from the 300 Area: (top) hydraulic head and (bottom) uranium concentrations in groundwater**



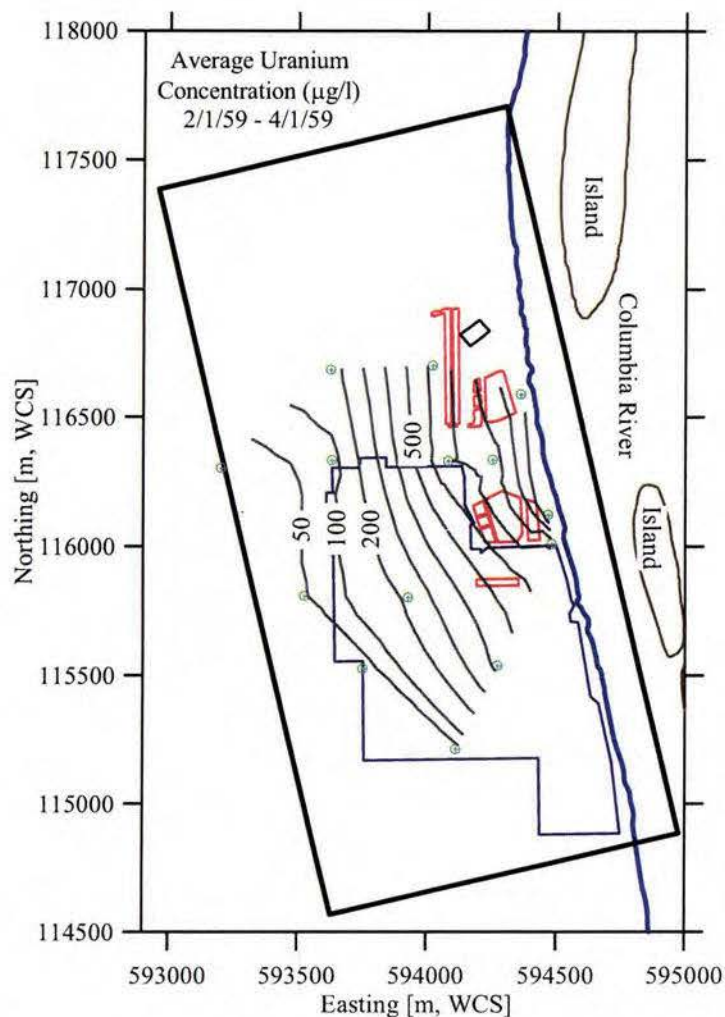


Figure 4-15. Uranium concentration ( $\mu\text{g/l}$ ) contours interpreted from groundwater measurements in 1959 made at the wells indicated in green

#### 4.2 Postulate Alternative Conceptual Models

As described in Chapter 3, application of the Maximum Likelihood Bayesian Model Averaging method for jointly assessing conceptual model and parameter uncertainties requires the formulation of alternative conceptual models. Based on the characteristics of the site and the available data, a number of conceptual model alternatives were considered to potentially contribute significant uncertainty. The conceptual model elements for which alternative representations were considered included the following.

- *Configuration of the geologic units.* Units contacts have been interpreted based on available

borehole data. Uncertainties arise because the location of these contacts is not always obvious and must be interpolated between boreholes. In addition, differences between units u5 and u7 are small and there is significant overlap in the ranges of observed properties for units u1 and u5.

- *Heterogeneity within the geologic units.* Although it is clear that there is significant spatial variability of properties within a geologic unit, data is limited to the location of boreholes. As a result, it is difficult to specify hydraulic property heterogeneity within geologic units.
- *Spatial and temporal variability of recharge.* Recharge at the Hanford Site is sensitive to the surface material (e.g., building, asphalt, soil),



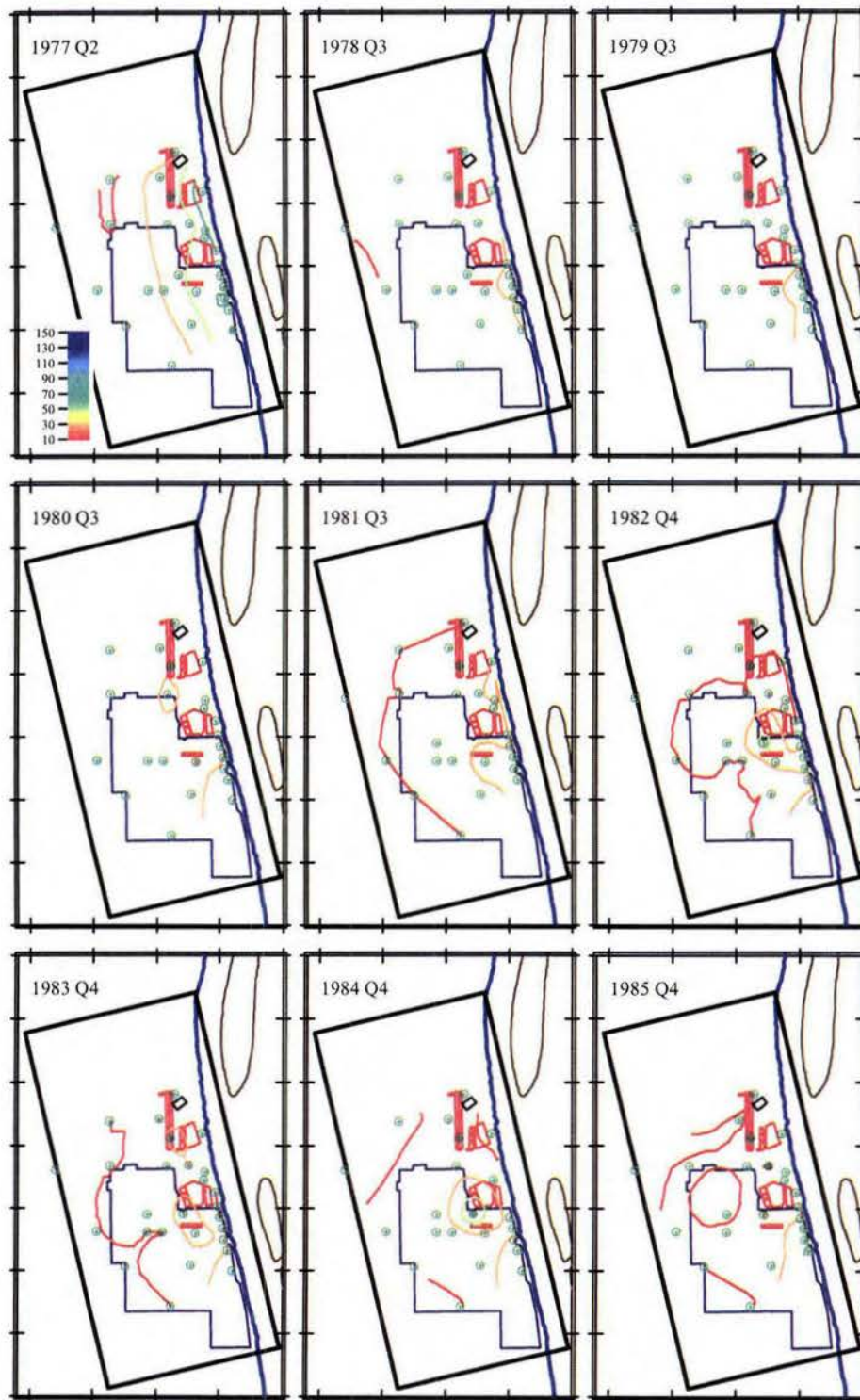


Figure 4-16. Uranium concentration contours derived from measurements in the year and quarter indicated; measurement locations shown in green, contours at 10, 30, 50...150  $\mu\text{g/l}$ . (A) 1977-1985

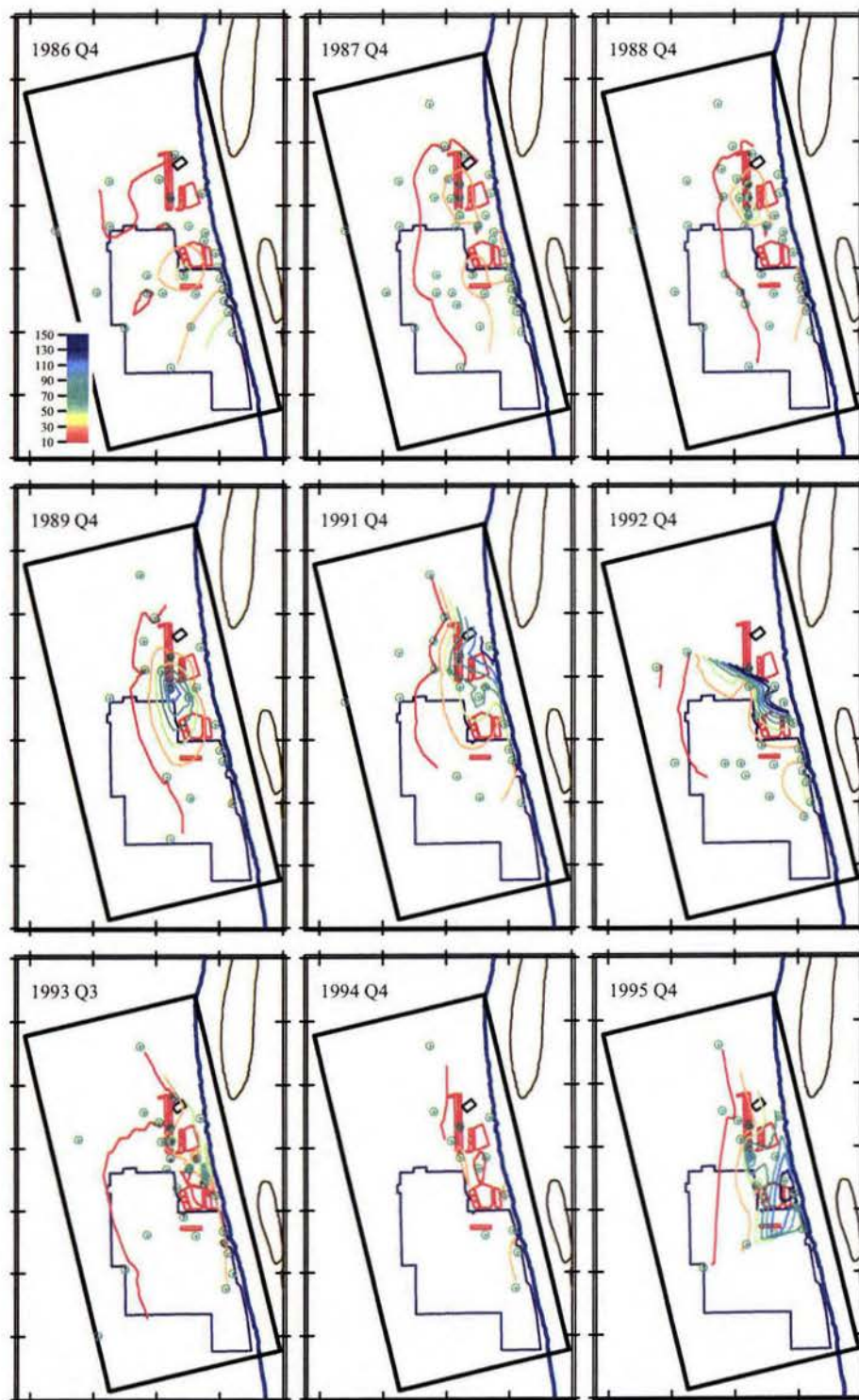


Figure 4-16 cont. (B) 1986-1995



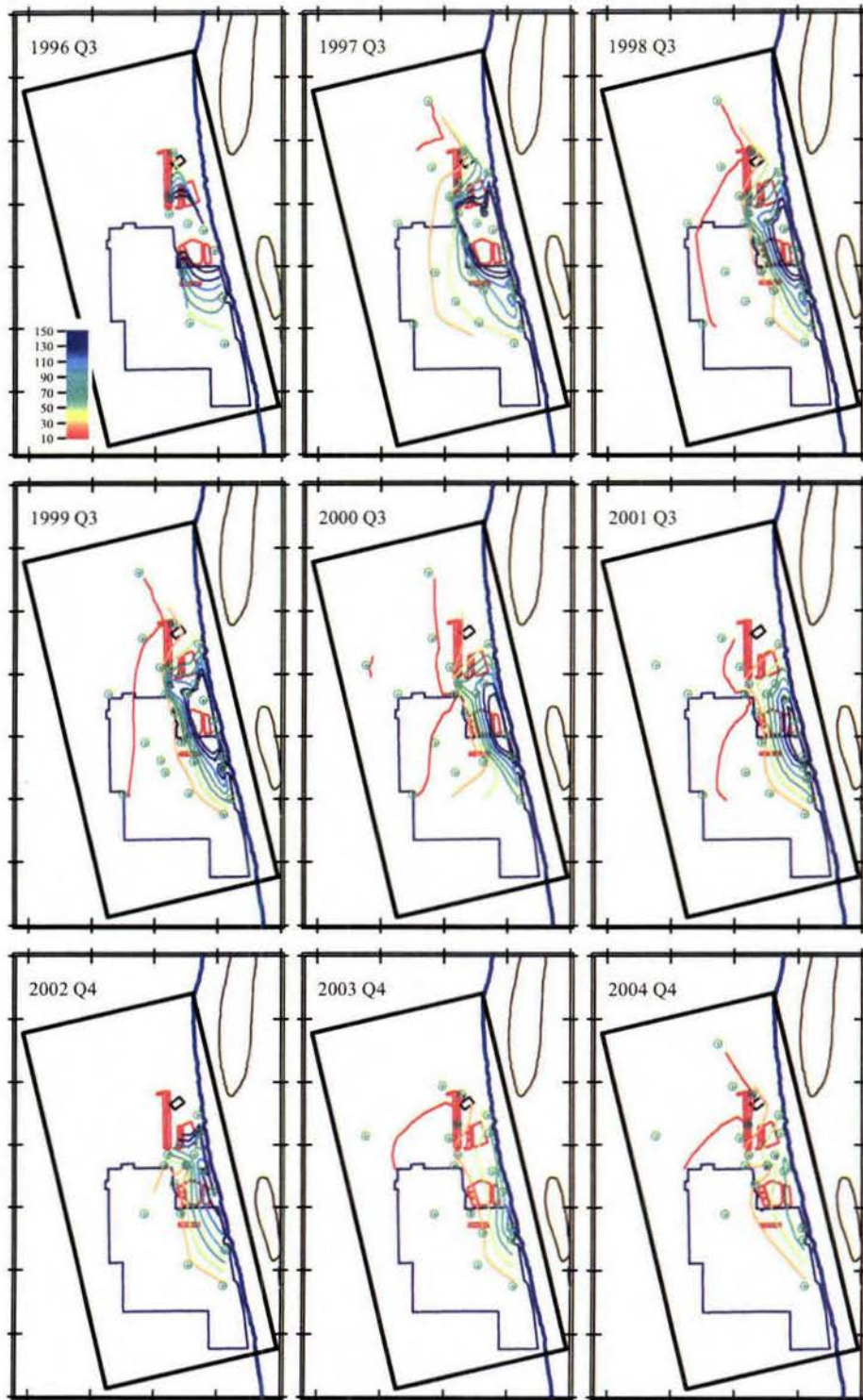


Figure 4-16 cont. (C) 1996-2004

the soil texture, and vegetation (if soil is the surface material). Recharge is likely to vary seasonally and from year to year. Nonetheless, the natural recharge flux is small relative to the groundwater flux and the impact on groundwater from temporal fluctuations in river stage is likely to be much greater than that arising from temporal variability in recharge.

- *River boundary transience.* River stage is known to vary significantly on an hourly basis. The impact on transport of the scale at which this variability is modeled is of potential importance.
- *Uranium release rates.* Disposals to the ponds have been estimated from available records, but significant uncertainties remain. Fluxes of water and contaminants to the unsaturated zone and the fluxes from the unsaturated zone to groundwater are even more uncertain. *Spatial and temporal variability of uranium adsorption.* Reactive processes in the unsaturated and saturated zones have been quantified recently, but significant uncertainty remains in the appropriate models to use. The suitability of applying lab-scale data to the field-scale is also uncertain. Geochemical modeling uncertainty was not the focus of this application, however, so the alternative models considered for adsorption were limited.
- *Heads along the lateral boundaries.* Boundary heads are based on a limited number of boreholes, particularly in the north and northwest part of the site. A well was recently added in this area.
- *Representation of unsaturated flow and transport.* The unsaturated zone at the 300 Area is a significant portion of the transport path. There are, however, only limited data on the unsaturated hydraulic properties and no in situ record of soil water content/pressure data for the 300 Area unsaturated zone. In addition, there are limited data on uranium concentrations in the unsaturated zone. Some site data indicate that transport in the unsaturated zone was primarily vertical. At the same time, the extent of the groundwater plumes in the past may have been increased by lateral transport in the unsaturated zone. In any case, it is computationally easier to model only saturated flow and transport, thus motivating a simplified representation of unsaturated zone flow and transport.
- *Darcy's Law alternative.* As a result of the rapid fluctuations in river stage, an alternative form of Darcy's Law that adds a term in the time-derivative of flux was considered. This would require modification of groundwater flow codes and would likely only be applicable very close to the river and only during high-frequency river

stage fluctuations. Waichler and Yabusaki (2005) determined that Darcy's Law was valid at the 300 Area.

To illustrate the application of the uncertainty assessment methodology, the alternative conceptual models represented uncertainties in three of these elements: the characterization of heterogeneity in hydraulic properties, temporal variability in the Columbia River stage, and spatial variability in adsorption.

#### 4.2.1 Hydraulic Property Heterogeneity

Two alternatives were postulated that represent a degree of uncertainty in the hydraulic property heterogeneity. One of the alternatives assumed the hydrogeologic units were located as given in the geologic model discussed above and that hydraulic properties were uniform within each unit. The other alternative assumed a homogeneous characterization of hydraulic properties over the entire model domain. Justifications for this approach include observations that Unit 6 is absent over much of the model domain near the river (see Figure 4-4), Units 5 and 7 are hydraulically similar, a relatively small portion of the Hanford Formation is saturated, and many of the data are from wells screened over a length of 4.5 m (15 ft) or more.

#### 4.2.2 River Boundary Transience

As discussed above, the Columbia River stage at the 300 Area fluctuates significantly, sometimes within a single day. Representing the transient river behavior has an impact on the computational requirements of the modeling. As a result, it is not uncommon to average river fluctuations over some period of time or to model a transient river boundary as a steady-state boundary. For this application, two alternatives were considered: a steady-state and a transient river boundary condition. A monthly averaging period for the transient river alternative was used, based on practical concerns in limiting the computational requirements.

#### 4.2.3 Adsorption

Two alternative representations of adsorption were included in this application. Both alternatives assumed a linear equilibrium adsorption model. One alternative assumed a spatially homogeneous  $K_d$  throughout the model domain. The other alternative used two  $K_d$  zones, one near the river and one for the rest of the model domain. This allowed the model to represent, to a limited extent, the differences in adsorption due to water chemistry differences in the groundwater and river water.



#### 4.2.4 Alternative Conceptual Model Summary

The alternative conceptualizations discussed above were combined to generate a set of model alternatives used in the uncertainty assessment application. The alternative models for the 300 Area uranium plume were combinations of (1) homogeneous vs. zoned hydraulic conductivity, (2) steady-state vs. transient river boundary condition, and (3) homogeneous vs. zoned  $K_d$ . Eight combinations of these three factors were modeled and are listed in Table 4-2. Note that these models covered the saturated zone only.

Preliminary analyses demonstrated that the models with a steady-state river boundary (odd numbered models in Table 4-2) provided a poor representation of observed heads and uranium concentrations and that these models would be unable to compete with the transient models in terms of posterior model probability. As a result, the steady-state models were dropped from the analysis and are not included in the remaining discussion of this application.

#### 4.3 Alternative Model Implementations

The alternative conceptual models described in the previous section were implemented using MODFLOW (Harbaugh et al. 2000) and MT3DMS (Zheng and Wang 1999). This section describes the elements of the alternative models.

The initial models used a simulation period starting in 1945 and extending through the period of uranium dis-

and ponds was based on the liquid fluxes used for the source terms in the Systems Assessment Capability simulations (Bryce et al. 2003). Uranium mass loadings for the liquid waste disposal sites were also based on information from the Systems Assessment Capability project. The calibrated models for the simulations beginning in 1945 were unable to represent the large-scale extent of uranium contamination in groundwater observed prior to 1976 (e.g., Figure 4-15) and the large-scale introduction of significant uranium to the groundwater in 1995-97 (Figure 4-16 B and C) either through geochemical changes in the aquifer beginning in 1995 (with the cessation of discharges to the 316-5 trench) or the presence of a long-term source of uranium in the vadose zone activated during the flood of 1997.

To reduce the uncertainty associated with the uranium sources and to simplify the model calibrations, the period of September 1997 to December 2004 was selected as the simulation period for calibration. It was assumed that there were no significant sources of uranium to the groundwater during this period.

##### 4.3.1 Simulation Domain

The simulation domain for the 300 Area models was selected to cover the observed (past and current) and anticipated future extent of the uranium plume. A plan view of the three-dimensional simulation domain is shown in Figure 4-17, including locations of the Columbia River boundary, the boundary of the 300 Area, the primary disposal area boundaries, and existing

**Table 4-2. Model alternatives considered in the 300 Area application of uncertainty assessment**

charges to the waste disposal sites. The recharge representing liquid waste disposal at the process trenches

Model	Hydraulic Property Heterogeneity	River Boundary	$K_d$ Spatial Variability
1	Homogeneous $K_s$	Steady-State	Homogeneous $K_d$
2	Homogeneous $K_s$	Transient	Homogeneous $K_d$
3	5 $K_s$ Zones	Steady-State	Homogeneous $K_d$
4	5 $K_s$ Zones	Transient	Homogeneous $K_d$
5	Homogeneous $K_s$	Steady-State	2 $K_d$ Zones
6	Homogeneous $K_s$	Transient	2 $K_d$ Zones
7	5 $K_s$ Zones	Steady-State	2 $K_d$ Zones
8	5 $K_s$ Zones	Transient	2 $K_d$ Zones

monitoring well locations. The three-dimensional simulation domain was 1386.8 m (dimension of x) by 2893.1 m (dimension of y) by 47.3 m deep. The rotation angle was 13.6 degrees counterclockwise from the east-west direction.

#### 4.3.2 Discretization

The simulation domain was discretized into a MODFLOW/MT3DMS grid with 15 layers, 54 rows, and 32 columns (approximately 26,000 grid cells). Grid cell size varied from 34.67 m to 69.34 m in the x direction (perpendicular to the river), from 36.16 m to 144.66 m in the y direction, and 1.0 m to 5.75 m in the z direction. Figure 4-18 illustrates the grid discretization. The grid was dense near the contamination sources (the ponds and trenches) with finer discretization also applied to the lateral boundaries. The vertical discretization was finer within the depth of the Columbia River than below the river. However, the first layer was coarse to incorporate fluctuation of the river stage, since the river was treated as a constant head boundary.

The simulation period for model calibration was from September 1997 through December 2004. Because the transient river boundary was modeled using monthly average river stage, each month constituted a stress period for which river stage was constant. Natural recharge was assumed constant in space and time for the entire simulation period.

#### 4.3.3 Hydrogeologic Zonation and Hydraulic Parameters

As discussed earlier, geologic analysis from borehole data reveals that there are five major hydrogeologic units in the 300 Area among six total units located above the underlying basalt. Using unit contact information from the borehole data, a three-dimensional model of the spatial distribution of the hydrogeologic units was developed (see Figure 4-3 and Figure 4-4). This model was used to project the spatial distribution of the hydrogeologic units onto the MODFLOW grid as illustrated in Figure 4-18 using the MODFLOW Hydrogeologic-Unit Flow (HUF) package. Zonation of the distribution coefficient was different than the hydrogeologic zonation and will be discussed in Section 4.3.5.

Estimated representative (average) values for hydraulic conductivity of the 300 Area sediments were given in Table 4-1. The parameter estimates were used as initial value and their ranges were used to constrain parameter adjustment in inverse modeling described in Section 4.8.

The upper boundary of the 300 Area flow models was a constant flux boundary with an applied flux of 55.4 mm/yr representing long-term average recharge based on the analysis of Fayer and Walters (1995). The bottom boundary of the models was a zero-flux boundary representing the top of the confining Lower Mud unit (u8). Although flow from the underlying confined basalt aquifer to the unconfined aquifer may occur at locations where the lower mud unit (u8) is incomplete, this unit appears to be contiguous throughout the model domain. Upward flux from the basalt aquifer was not anticipated to affect uranium concentration and was not considered in the models.

The eastern boundary of the model domain was the Columbia River. Groundwater is generally discharged to the river. During high river stage, there is infiltration of river water through the river bank resulting in mixing within the near-river groundwater and a potential reversal of groundwater flow (see the discussion in Section 4.1.2.4). In the geologic model the river penetrates through the Hanford formation and into the upper Ringold (u5) except for a portion of the boundary around Northing coordinate 115500 m (see Figure 4-4). The shape and depth of the river are irregular, which was reflected in both the geologic model and the 300 Area simulation model grid (Figure 4-18). In the flow models, the river was treated as a prescribed head boundary with temporally varying head. As discussed in Section 4.1.2.4, the river stage fluctuates significantly on an hourly time-scale. For computational reasons, however, the river boundary varied no more frequently than monthly for the simulation models. The monthly average river stage at the 300 Area as simulated by Waichler et al. (2005) was taken as the eastern head boundary. Figure 4-19 shows the simulated river stage over the period of site operation.

The western boundary of the flow models was treated as a prescribed head boundary with the head value along the boundary varying in space. The assigned head values were based on groundwater heads observed in wells located near the boundary. Observed heads supported the assignment of zero flux to the southern boundary of the nominal model. The northern boundary was also treated as a zero-flux boundary, although this assumption has less support from observations since there has only recently been a monitoring well in the northwest portion of the model domain.

The boundaries of the transport model coincided with those of the flow model. The upper boundary along with the bottom, northern, western, and southern boundaries were treated as zero flux transport boundaries. The eastern boundary along the Columbia River was treated as an outflow boundary. That is, mass

transport across the boundary occurred by advection only and solely in the direction from the groundwater to the river.

#### 4.3.4 Initial Conditions

The initial condition for the flow models was based on average heads across the simulation domain. The initial uranium concentration in groundwater was based on observed concentrations in August 1997 (Figure 4-20, left). Artificial observations were used to provide a smooth interpolation of concentration over layer 1 of the simulation domain (Figure 4-20, right). These interpolated concentrations were also applied to layers 2-5 of the model. Fifty percent of the layer 1 concentration was applied to layer 6 and 25% to layer 7 for the initial condition. Layers 8-15 had an initial concentration of zero. As discussed in the introduction to Section 4.3, although groundwater contamination occurred much earlier, the simulations for this application were begun in September 1997 to reduce the uncertainty associated with the uranium sources and to simplify the model calibrations.

#### 4.3.5 Geochemical Modeling

Desorption/adsorption experiments conducted on 300 Area soil samples indicate that uranium chemistry at the 300 Area may be quite complex (Serne et al. 2002; Zachara et al. 2005). Reproducing observed uranium transport may thus require a reactive transport model, such as the non-equilibrium adsorption model or the surface complexation model described in Zachara et al. (2005). Application of a reactive transport model was beyond the scope of this project, however.

The alternative models implemented here assumed that adsorption at the site can be represented as a time-invariant, equilibrium process. Serne et al. (2002) indicated that a linear adsorption model is appropriate for concentrations currently observed in the groundwater, but that a non-linear model may be applicable for higher concentrations potentially present in the vadose zone. A range of distribution coefficient values ( $0 \leq K_d \leq 100$ ) have been estimated from laboratory experiments on the <2 mm particle size fraction (Serne et al. 2002; Zachara et al. 2005). It is generally expected that these experimental  $K_d$  values will be reduced in the field due to the presence of coarse (>2 mm) sediments (Kaplan et al. 2000). Serne et al. (2002) concluded that likely ranges for field values of  $K_d$  are from 0 to 1 mL/g in the near surface vadose zone influenced by evapotranspiration, 2 to 4 mL/g in the unconfined aquifer sediments not influenced by dilution of river water, and 7 mL/g or higher when groundwater is diluted by river water. Their proposed spatial variabil-

ity in  $K_d$  values was primarily due to differences in dissolved inorganic carbon.

The effect of river water on uranium adsorption was approximated in the model alternatives implemented here by using a spatially variable, zoned  $K_d$  with different underlying values for the groundwater within a mixing zone adjacent to the Columbia River. For the alternative models with heterogeneous  $K_d$ , the model domain was separated into two parts by the black lines shown in Figure 4-18. This zonation was chosen based on an approximate zone of mixing as reported by Waichler and Yabusaki (2005) using observed nitrate concentrations. Prior values for  $K_d$  are discussed in Section 4.5.

#### 4.4 Prior Model Probability

For the 300 Area application, equal prior model probabilities were assumed to simplify the analysis. This reflects the *a priori* belief that the alternative models are all equally plausible. For many applications, non-uniform prior model probabilities may be appropriate. In these cases, expert judgment combined with the method of Ye et al. (2005) can be used to assign prior model probabilities. Appendix C contains an application of MLBMA to alternative recharge models in which a panel of experts was used to assign prior model probabilities.

Ye et al. (2005) proposed to maximize a measure of entropy composed of prior model probabilities subject to constraints representing prior expert knowledge of the relative model probabilities. Specifically, they proposed to assign prior model probabilities by solving the following nonlinear, constrained optimization problem.

$$\max_{p_k} H = -\sum_{k=1}^K p_k \log p_k \quad (30)$$

subject to

$$\begin{aligned} g_i &\leq 0 \quad i = 1, \dots, I \\ h_j &= 0 \quad j = 1, \dots, J \\ \sum_{k=1}^K p_k &= 1 \end{aligned} \quad (31)$$

where  $H$  is the entropy measure,  $p_k \equiv p(M_k)$ , the prior model probability, and  $g_i$  and  $h_j$  are specified relationships between the various  $p_k$  values. For example, the constraint  $2p_2 - p_1 \leq 0$  expresses the prior judgment that Model 1 is at least twice as plausible as

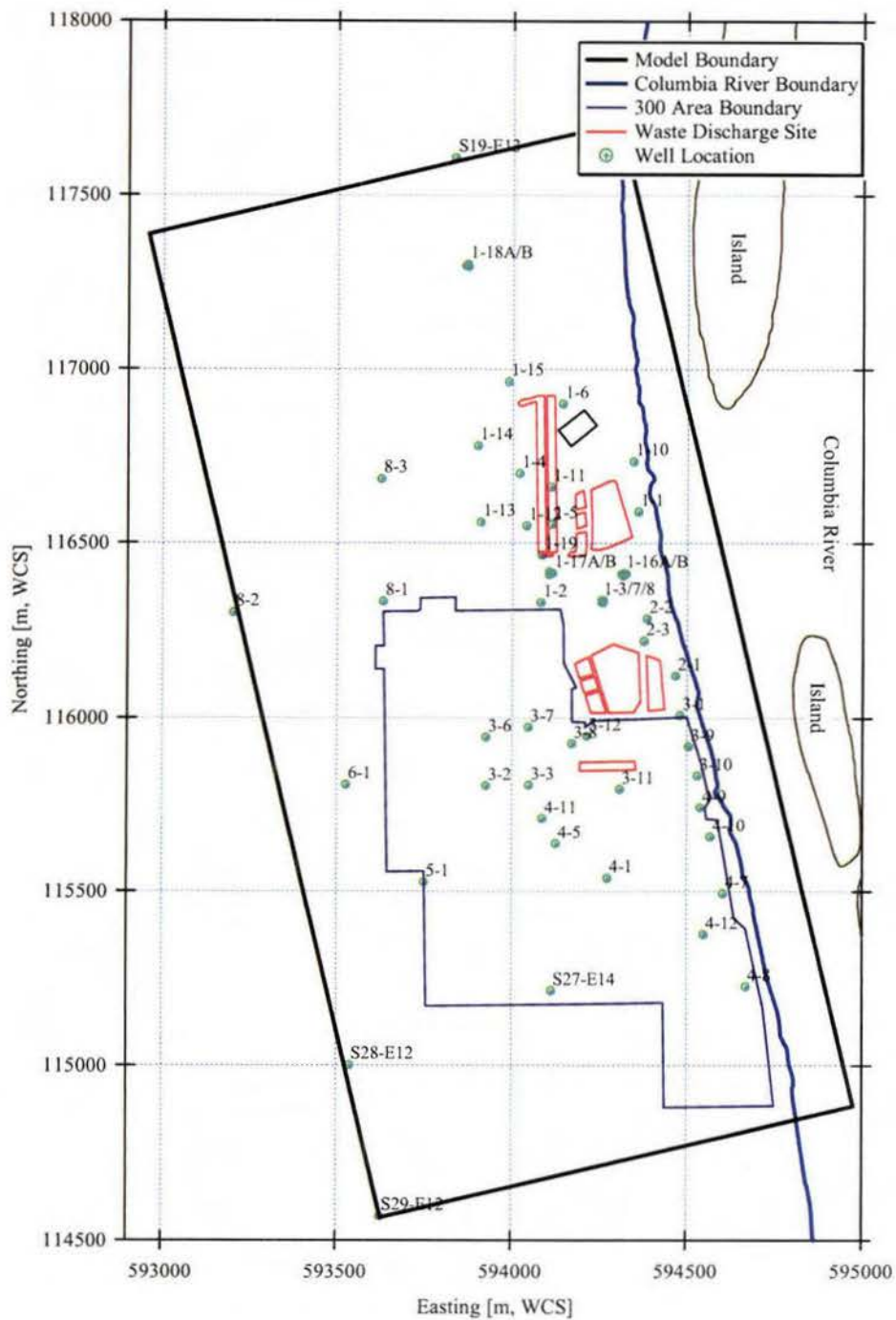


Figure 4-17. Simulation domain and locations of major features of the 300 Area simulation model



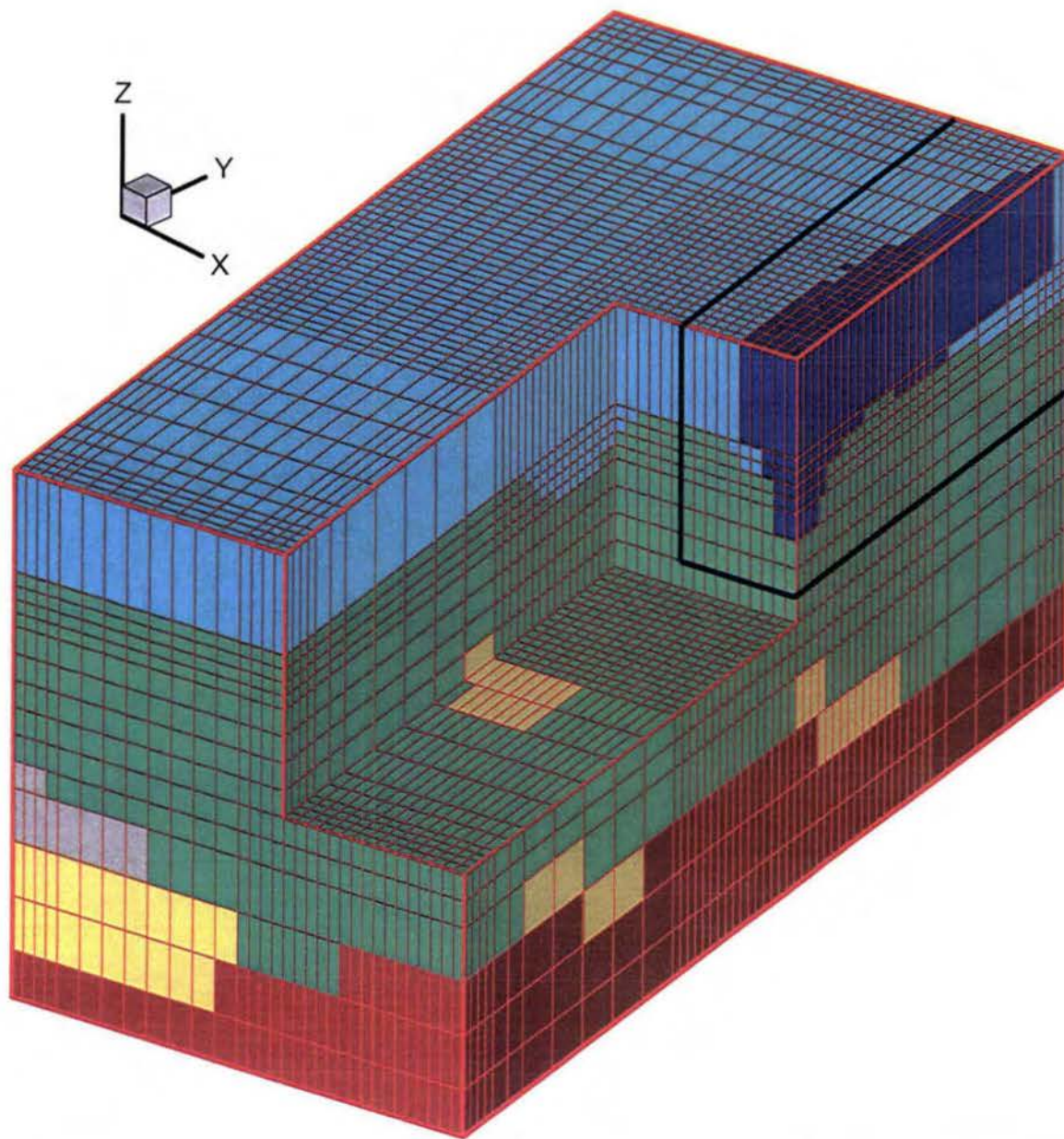
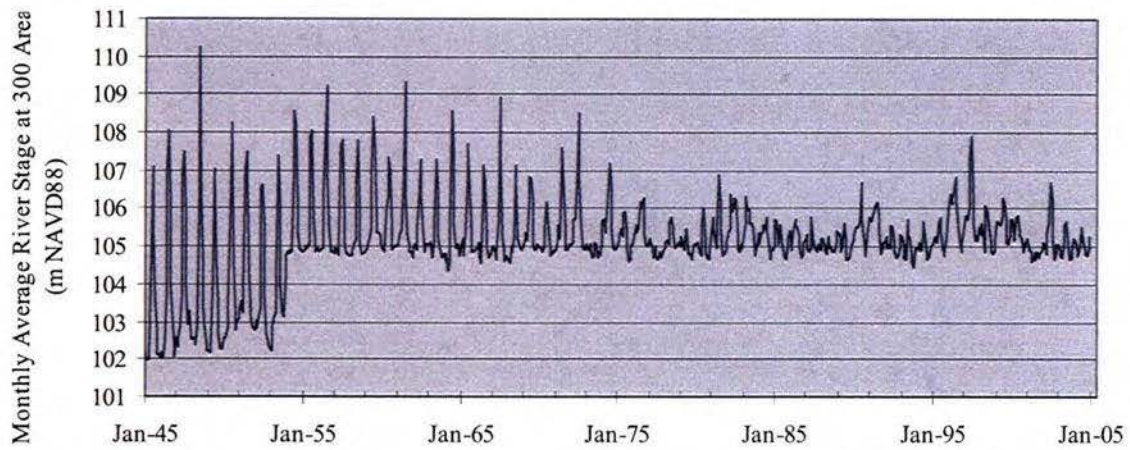
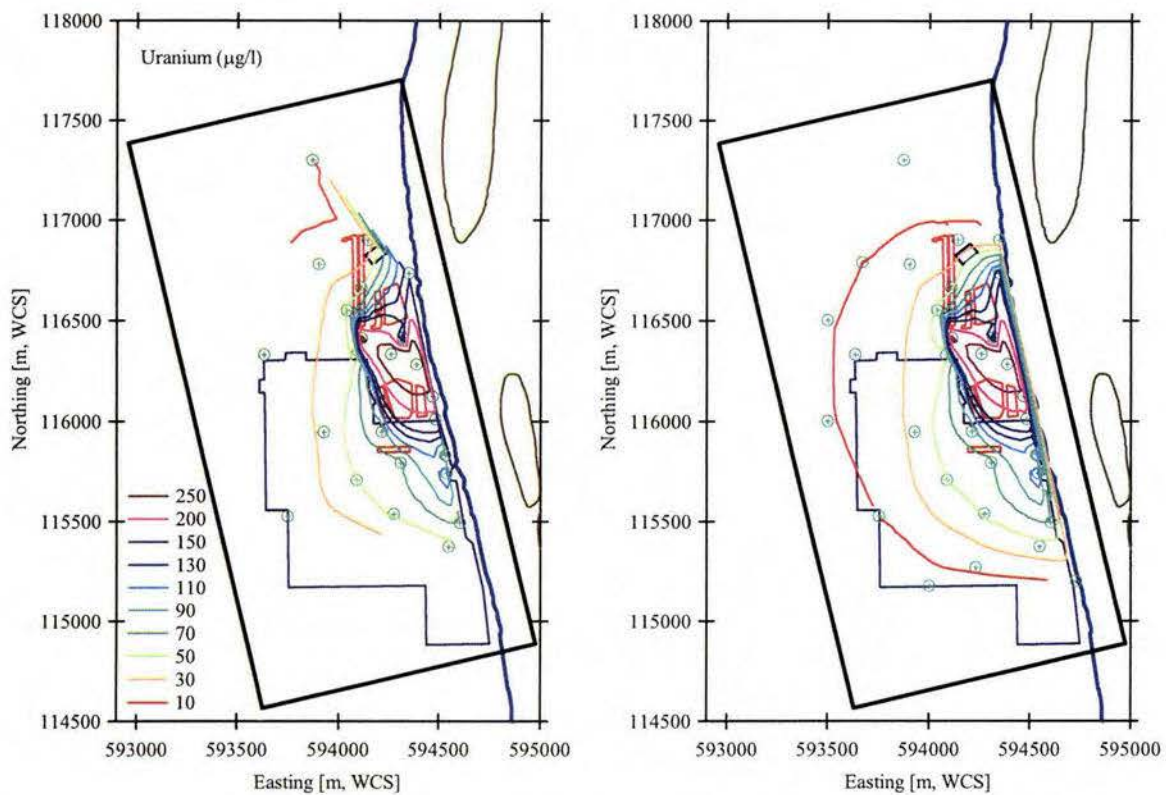


Figure 4-18. Discretization of three-dimensional simulation domain. Cells are inactive within the Columbia River (shown in blue). Hydrogeologic units are also shown in the figure as follows: u1 (teal), u5 (green), u6 (grey), u7 (yellow), u8 (red). The black lines indicate an adsorption zone near the river.



**Figure 4-19. Monthly average river stage at the 300 Area simulated by Waichler et al. (2005) and used as a boundary condition in the groundwater flow models**



**Figure 4-20. Contours of (left) observed uranium concentration in groundwater in August 1997 and (right) modified values used as the initial condition in layer 1 of the simulation models**



Model 2. If the constraints are logical in that they conform to Ockham's razor and to behavior expected on theoretical and/or empirical grounds (i.e., if one excludes arbitrary solutions that are not based on the principle of parsimony coupled with sound expert knowledge), then the corresponding informed solution is preferred over the non-informative neutral choice (i.e.,  $p_k = 1/K$ ).

If an expert, or a group of experts, is unable to select one set of constraints among several alternative sets, Ye et al. (2005) recommended choosing among alternative expert opinions *a posteriori*, that is, on the basis of posterior measures of model quality. If sufficient data are available to conduct a cross-validation, Ye et al. (2005) recommended choosing the set of prior probabilities that yields optimum predictive performance as measured by an objective criterion (such as mean squared prediction error). If there are not enough data to conduct a meaningful cross-validation Ye et al. (2005) suggested selecting the set of prior probabilities that maximizes the likelihood of the set of model alternatives in light of the data. That is, they suggested choosing the set of prior probabilities that maximizes

$$p(\mathbf{D}|\mathcal{M}) = \sum_{k=1}^K p(\mathbf{D}|M_k) p(M_k) \quad (32)$$

$$\approx \sum_{k=1}^K \exp\left(-\frac{1}{2} KIC_k\right) p(M_k)$$

Once  $KIC_k$  values, which are independent of prior model probabilities, have been computed, then Eq. (32) can be easily calculated for any admissible (i.e., arrived at via entropy maximization on the basis of prior expert judgment) prior probability set.

#### 4.5 Prior Parameter Probability

The methodology described in Chapter 3 allows for the application of prior parameter probabilities as an optional component of the maximum likelihood objective function used in the parameter estimation (Eq. 11 and Step 3 of Section 3.3) and as an uncalibrated, but uncertain parameter input to the computation of prediction uncertainty in Step 9 of Section 3.3. The assessment of prior parameter uncertainty was discussed in general terms in Section 2.1. This section describes the use of generic and site-specific data to estimate uncertainty in the average linear adsorption coefficient at the Hanford Site 300 Area and application of this data to specify prior values of  $K_d$  in the model calibrations.

##### 4.5.1 Modeling Adsorption in Contaminant Transport Models

Adsorption is one of the most important geochemical mechanisms for slowing or preventing the migration of contaminants through the vadose zone and in groundwater systems. The most common method used to describe contaminant adsorption on complicated matrices such as soils and sediments is the distribution coefficient or  $K_d$  model:

$$K_d = S/C_{aq} \quad (33)$$

where  $S$  is the concentration of the contaminant on the solid and  $C_{aq}$  is the concentration in the aqueous phase. Implicit assumptions typically made when using this model in groundwater transport codes are that the  $K_d$  is constant (conditions that affect the  $K_d$  do not change), adsorption reaches equilibrium quickly relative to the groundwater flow rate, and the adsorption is reversible. These are not necessarily realistic assumptions. The popularity of the distribution coefficient approach among modelers stems from its simplicity and the fact that it can be applied to complex matrices and solutions for which it would be difficult or impossible to obtain all the required thermodynamic and mechanistic surface adsorption data. This advantage is also the primary drawback of this approach. Because the model is empirical, it should be strictly applied only to the same geochemical conditions under which the  $K_d$  was measured. This condition can be relaxed if a variable is known to have no or minimal influence on adsorption of the contaminant of interest. When geochemical conditions that affect adsorption vary significantly within the system being modeled, then the constant  $K_d$  model can produce erroneous results. This problem can be avoided by varying the  $K_d$  spatially and temporally as needed due to changing geochemical conditions. In addition, if mechanisms other than adsorption are important for immobilizing contaminants of interest (such as precipitation), the  $K_d$  model may not be appropriate.

Surface complexation modeling is a mechanistic-based approach for describing adsorption that can account for variable geochemical conditions. Most commonly used groundwater transport codes do not have the capability to model sophisticated geochemical interactions. Reactive transport codes that integrate thermodynamics to model complex geochemical interactions are computationally expensive and typically have intensive input data requirements. As a result, these more sophisticated approaches to contaminant transport modeling are not generally applied, but are applied at sites where resources permit more in-depth research.

An alternative to using the more geochemically comprehensive reactive transport codes, is to use surface complexation data and geochemical equilibrium codes to calculate  $K_d$  values that can be used in groundwater transport codes that rely on  $K_d$  values to describe adsorption. In this approach  $K_d$  values that vary spatially and/or temporally can be used as input data to the groundwater transport code.

Keeping in mind the caveats discussed above, adopting the use of Eq. (33) requires estimating an appropriate value of  $K_d$  and the uncertainty in that value. A variety of data sources are available to provide a basis for this estimation. Table 4-3 lists a possible hierarchy of this data, ranging from generic literature values to detailed, site-specific geochemical measurements. The availability of data determines the appropriate methods used to estimate  $K_d$  values. It is assumed in Table 4-3 that as the quantity and quality of site-specific characterization data become available, the quality of adsorption data that can be determined or measured will increase and the uncertainty will decrease. The challenge is to converge on an appropriate balance between minimizing  $K_d$  uncertainty and minimizing the resources needed to reduce that uncertainty.

#### 4.5.2 Hanford Site 300 Area Uranium Example

Application of the concepts embodied in Table 4-3 are illustrated here using as an example uranium adsorption at the Hanford Site 300 Area.


##### 4.5.2.1 No Site-Specific Data Case

A number of literature compilations of  $K_d$  values are available. Sheppard and Thibault (1990) provide a compilation of  $K_d$  values for four major soil types (sand, loam, clay, and organic). Across all soil types, the overall range for the uranium  $K_d$  is from 0.03 to  $4.0 \times 10^5$  L/kg (mL/g). A look-up table for estimating  $K_d$  values as a function of pH (for pH values ranging from 3 to 10) is provided in EPA (1999). The overall range of  $K_d$  values across the entire pH range of 3 to 10 was  $< 1$  to  $1.0 \times 10^6$  L/kg. It is clear from these data sources, that the range of  $K_d$  values for a wide range of possible conditions is very large for uranium. From a transport modeling perspective, this indicates that uranium mobility can range from being essentially completely mobile (moves at the same rate as water) to essentially completely immobile. One could define a  $K_d$  probability distribution to reflect this large range of possible values, but the resulting transport uncertainty would likely be so large as to be unworkable.

##### 4.5.2.2 Known Soil Texture and pH Case

When the soil texture and pH are known and can be used to represent average site conditions, this information can be used to narrow the range of  $K_d$  values used for transport modeling and hydrologic uncertainty assessments. As indicated previously, Sheppard and Thibault (1990) provide a compilation of  $K_d$  values for four major soil types (sand, loam, clay, and organic). In this compilation, mineral soils were categorized by texture into sand, clay, or loam. Soils that contained  $\geq 70\%$  sand-sized particles were classified as sand soils.

**Table 4-3. Hierarchical application of data to determine best estimates and uncertainty in  $K_d$  values**

Available Site-Specific Input Data	Methodology Used to Determine Best Estimates and Range of $K_d$ Values	Resulting Uncertainty in Adsorption Modeling Results
No Site-Specific Data	Available Literature Compilations of Generic $K_d$ Values	Decreasing Uncertainty, Increasing Resource Requirements 
Soil Texture and pH	Refinement of $K_d$ Values Using Available Literature Compilations of Generic $K_d$ Values Based on Limited Site-Specific Data	
Site-Specific Batch $K_d$ Measurements	Better Defined $K_d$ Values and Ranges Based on Site-Specific Empirical Measurements	
Detailed Geochemical Information, Site-Specific Adsorption Measurements and Model Development	Use of More Scientifically Defensible Adsorption Models and Site-Specific Data to Calculate $K_d$ Values for Variable Conditions	



Mineral soils that contained  $\geq 35\%$  clay-sized particles were classified as clay soils. Loam soils had an even distribution of sand-, clay-, and silt-sized particles or consisted of up to 80% silt-sized particles. Soils containing  $> 30\%$  organic matter were classified as organic soils.

Previous work conducted on 300 Area sediments can be used as a source for soil texture and pH data (Zachara et al. 2005). Table 4-4 contains major size fraction data for a sediment sample collected at 16 feet below ground surface (bgs) from the 300 Area North Process Pond (NPP) along with the pH value measured on a 1:2 water extract. Included are percentages for the whole sample and the sample fraction less than 2 mm.

Because the 300 Area sediments are composed of predominantly cobbles, they don't fit into the classification scheme of Sheppard and Thibault (1990). To handle this problem, the Sheppard and Thibault method for estimating  $K_d$  values will be used on the less than 2 mm size fraction and a gravel correction will be handled separately. For the NPP-1 (16 ft bgs) sample, the less than 2 mm size fraction contains 70% sand. As a result, this sediment is classified as a sand using the Sheppard and Thibault (1990) scheme. From Table 3 in Sheppard and Thibault (1990), the mean uranium  $K_d$  value for sand is 35 L/kg, with a range of 0.03 to 2200 L/kg.

A methodology for adjusting  $K_d$  values to account for the gravel content of soils and sediments is provided in Cantrell et al. (2003). This report indicates that for high  $K_d$  contaminants (Cs, Sr, and Pu), the following equation is recommended (see Appendix A, Kaplan and Serne 2000).

$$Kd_{gc} = (1-f) Kd_{<2mm} + (f)0.23 Kd_{<2mm} \quad (34)$$

where  $Kd_{gc}$  is the gravel-corrected  $K_d$  value,  $f$  is the weight fraction gravel, and  $Kd_{<2mm}$  is the  $K_d$  value determined using the less than 2 mm material. For low  $K_d$  contaminants, the following equation is recommended.

$$Kd_{gc} = (1-f) Kd_{<2mm} \quad (35)$$

**Table 4-4. Textural characteristics and pH for sediment sample NPP-1 obtained 16 ft. below ground surface from the North Process Pond (Zachara et al. 2005)**

Sediment ID	Entire Sample	Fraction <2mm
Cobbles (%)	91.7	-
Sand (%)	6.5	70
Silt (%)	0.77	13
Clay (%)	1.01	17
pH	7.98	7.98

The potential range of uranium  $K_d$  spans values generally considered to be both low and high. Because the mean  $K_d$  value for uranium is 35 L/kg, Eq. (34) was used for the gravel corrections.

Using a value of 0.917 for  $f$  in Eq. (34), the gravel-corrected mean uranium  $K_d$  value for 300 Area sediments was calculated to be 10 L/kg with a range of 0.009-650 L/kg, based on Table 3 in Sheppard and Thibault (1990).

This  $K_d$  value estimation methodology has reduced the uncertainty for the  $K_d$  values by over two orders of magnitude and has provided a mean estimate as well. A probability distribution consistent with the mean and range could be assigned to represent  $K_d$  uncertainty. This approach still produces a relatively large uncertainty for uranium  $K_d$  values.

#### 4.5.2.3 Available Regional Batch $K_d$ Values

$K_d$  values measured on Hanford Site sediments and Hanford groundwater range from 0.16 to 104 L/kg (on  $< 2$  mm size materials) (Cantrell et al. 2003). None of these measurements were on 300 Area sediments. As part of the composite analysis being conducted for the Hanford Site, best estimate  $K_d$  values were selected for various waste chemistry/source categories and impact categories (Last 2004). Estimates of minimum and maximum values were also provided. These values were determined based upon a thorough review of the data compiled in Cantrell et al. (2003). For natural Hanford groundwater conditions, the best estimate  $K_d$  value for uranium was taken to be 0.8 L/kg, with a minimum of 0.2 L/kg and a maximum of 4 L/kg. Applying the gravel correction appropriate for the 300 Area (Eq. 35 with  $f = 0.917$ ), a best estimate of 0.07 L/kg was determined, and a minimum and maximum of 0.02 and 0.33 L/kg were determined. Even though these Hanford Site values could be considered site specific, they are not specific to the 300 Area.

#### 4.5.2.4 Available Site-Specific Batch $K_d$ Values

Leaching and adsorption characteristics of uranium in near-surface sediments collected from the 300 Area of the Hanford Site were studied by Serne et al. (2002). Adsorption  $K_d$  values were measured over a range of solution variables. These included two artificial groundwater compositions; one referred to as high ionic strength solution and one referred to as low ionic strength. The low ionic strength solution was a 10:1 dilution of the high ionic strength solution.  $K_d$  measurements with these solutions were conducted over a range of uranium concentrations that varied from 50 to 5000 ppb U. Other experiments were conducted at

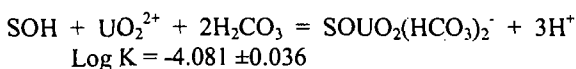
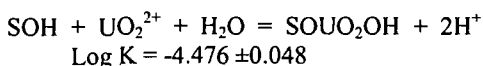
variable pH values and variable bicarbonate concentrations. The  $K_d$  values measured under conditions that most closely resemble those of 300 Area groundwater were those measured using the low ionic strength solution. Only these  $K_d$  values will be reviewed here.

Ten  $K_d$  measurements were made using the low ionic strength solution. As indicated, the U concentration was varied from 50 to 5000 ppb. Results from both the low ionic strength and high ionic strength solutions indicated that the adsorption was linear over this range of U concentration (i.e., the  $K_d$  model was appropriate). The average and standard deviation for the ten  $K_d$  values measured at low ionic strength were 3.77 and 0.71 L/kg, respectively, with a minimum of 2.99 and a maximum of 5.04. Applying the same gravel correction used previously, a corrected average and standard deviation of 0.31 and 0.06 L/kg, respectively, were calculated. The geometric mean was 0.31 L/kg also. Assuming  $K_d$  follows a lognormal distribution, the mean plus and minus three standard deviations yields an approximate range of values from 0.18 to 0.53 L/kg.

#### 4.5.2.5 Available Site-Specific, Mechanistically-Based, Adsorption Model Case

To account for the variation in the observed  $K_d$  values, a new geochemical conceptual model was developed to describe U(VI) adsorption on 300 Area sediments (Zachara et al. 2005). This conceptual model included a surface complexation model (SCM) based on the generalized composite approach (Davis et al. 2004, 2002; Davis and Curtis 2003). A preliminary SCM based on this work was used here (Davis, J.A., personal communication, 2005). U(VI) adsorption to 300 Area sediments was measured over a range of uranium concentrations and alkalinity that was significantly greater than those observed at the 300 Area. This ensured that the SCM was applicable to all groundwater conditions that could be expected at the 300 Area.

Two equations were found to provide the best fit to all the experimental adsorption data. The two equations are given below, including the log of the equilibrium constant and its standard deviation:



These equations were used along with major ion composition data available from the HEIS (1994) database for 300 Area wells to calculate  $K_d$  values. The equilib-

rium distribution between the solid and aqueous phase was calculated using the SpecE8 module of Geochemists Workbench® (Bethke 2005). The thermodynamic database thermo.com.v8.r6+.dat was modified to include the same formation constants for U(VI) solution species used in the preliminary SCM.

$K_d$  values were calculated for groundwater compositions determined from 19 wells in the 300 Area. Only wells that had a complete analysis set were used. This included pH, major cations and anions, and alkalinity. Wells that were near the river were avoided because these wells are seasonally influenced by the Columbia River. Only one analysis was used for each well to calculate the  $K_d$  values. In general, the latest available analysis that included the required suite of analytes was used. Table 4-5 shows the groundwater compositions used to calculate the  $K_d$  values, along with the calculated  $K_d$  values. A gravel fraction of 0.917 in Eq. (35) was used to correct the derived  $K_d$  values.

The average and standard deviation for the nineteen  $K_d$  values that were calculated was 3.18 and 2.64 L/kg, respectively, with a geometric mean of 2.14 L/kg. The minimum and maximum values calculated were 0.25 and 9.96 L/kg. Assuming  $K_d$  follows a lognormal distribution, the mean plus and minus three standard deviations yield an approximate range of values from 0.10 to 44 L/kg. This large variation in calculated  $K_d$  values illustrates that relatively minor changes in solution chemistry can have relatively large impacts on how strongly U(VI) is adsorbed to Hanford sediments.

The variation in  $K_d$  values calculated using the SCM approach was significantly greater than that determined from both the 300 Area site-specific  $K_d$  measurements and  $K_d$  measurements made using sediments collected throughout the Hanford Site. In the case, of the 300 Area site-specific measurements, the limited variation of the  $K_d$  measurements was due to the fact that the background solution composition was identical for each measurement. For the Hanford-wide  $K_d$  measurements, natural Hanford groundwater samples (uninfluenced by waste site releases) were used as the background solution. In the case of the  $K_d$  values calculated for the 300 Area using the SCM, variable solution compositions measured from wells throughout the 300 Area were used. In some cases, relatively low pH values were observed in wells near the process trenches and ponds. Natural Hanford groundwater pH values typically range from about 7.8 to 8.2. At the 300 Area some pH values as low as 7.35 were used to calculate the  $K_d$  values. This explains the reason for the relatively large variation in  $K_d$  values calculated using the SCM.

**Table 4-5. 300 Area groundwater compositions ( $\mu\text{g/L}$ ) used to calculate  $K_d$  ( $\text{L/kg}$ ) values using the SCM, along with the calculated  $K_d$  values. Units for alkalinity are in  $\mu\text{g/L}$  as  $\text{CaCO}_3$ . Included at the bottom are the average, standard deviation, and coefficient of variation (COV) for each of the chemical constituents and  $K_d$  values for all the wells.**

Well ID	Sample Date	Alkalinity	Ca	Cl	Mg	Nitrate	pH	K	Na	Sulfate	$K_d$
399-1-12	6/6/1996	120000	44000	16300	9500	22150	7.47	2600	20000	51300	6.52
399-1-13A	7/17/1992	120000	48500	16014	9900	25650	7.86	4500	22000	40720	1.26
399-1-14A	6/4/1996	140000	47000	16500	11000	23300	7.46	5500	22000	53200	4.84
399-1-17A	12/16/2004	140000	56800	18400	12800	23000	7.46	5040	27700	53800	3.46
399-1-21A	4/30/1992	121000	47200	20200	9505	17300	7.40	4710	18250	46000	7.42
399-1-5	5/20/1992	111000	47950	18500	9710	21200	7.35	2550	17250	40720	9.96
399-3-11	6/11/1992	123000	46150	17500	9255	15100	7.55	5195	20250	39000	4.05
399-3-12	4/23/1992	129000	45550	18200	9350	15100	7.80	5105	20650	41000	1.51
399-3-2	12/9/1991	116000	41650	16014	8705	19000	7.98	5105	16700	27000	1.15
399-3-3	5/13/1992	98000	31550	8900	6420	7970	7.79	3660	12350	23000	4.78
399-4-1	12/10/1991	120000	41000	16014	8050	25650	7.82	4800	18000	40720	1.95
399-4-11	12/11/1991	120000	42000	16014	8100	25650	7.61	4600	18000	40720	3.99
399-5-1	4/21/1992	158000	63850	17300	12800	66400	7.84	6605	24250	54000	0.54
399-6-1	4/21/1992	134000	50700	20700	11150	31400	8.30	5985	22150	40000	0.25
399-8-1	7/15/1992	110000	43000	14200	8750	15300	8.16	5250	14000	36200	0.73
399-8-4	1/8/1992	106000	40750	16014	8865	15900	7.87	3420	17550	25000	2.04
399-8-5A	4/16/1996	110000	43000	14000	8900	24000	7.80	5200	14000	32000	2.21
699-S27-E14	11/29/1990	151000	51200	11600	10600	24680	7.47	6250	21000	35300	3.28
699-S28-E12	4/21/1992	146000	64450	11900	13400	68600	7.90	6835	24700	54000	0.48
Average		124895	47174	16014	9829	25650	7.73	4890	19516	40720	3.18
Standard Deviation		16141	7922	2931	1778	15702	0.26	1184	3910	9713	2.64
Coeff. of Variation (%)		12.9	16.8	18.3	18.1	61.2	3.4	24.2	20.0	23.9	83.0

In summary, a hierarchy of data was applied to specifying the uncertainty in  $K_d$  values at the Hanford Site 300 Area. Uncertainty in this case was represented as a range of possible values, which was generally expected to be reduced as more site-specific data was applied. Figure 4-21 displays the resulting ranges of  $K_d$  values derived for the five levels of data. If an average value was available that is shown also. (As noted above, when the  $K_d$  estimates are supported by limited data, the resulting uncertainty may be unworkably large. When batch  $K_d$  measurements are the only source of data, they may underestimate the uncertainty by failing to account for potential variation in water chemistry. The use of site-specific adsorption measurements and an SCM, when applied to observed water chemistry data resulted in a realistic range of  $K_d$  values. Since these values were computed using a common gravel fraction, the derived  $K_d$  uncertainty would be somewhat larger if the variation in sediment particle size distribution were included.

#### 4.5.3 Prior Values of $K_d$ for 300 Area Model Calibration

The distribution of  $K_d$  based on the 300 Area surface complexation model (Section 4.5.2.5) could be used in model calibration by specifying a prior estimate for the value of  $K_d$  as the mean of the distribution and deriving a weight for the prior estimate from the variance of the distribution. Because two of the models (Models 6 and 8) have two distinct  $K_d$  zones (and thus two  $K_d$  parameters for which prior information could be supplied), the data described in Section 4.5.2.5 were supplemented with a set of seven  $K_d$  estimates derived from the surface complexation model using water chemistry data from seven wells near the river (Table 4-6). The mean  $K_d$  for the near-river wells was 20.1  $\text{L/kg}$  with a standard deviation of 17.5  $\text{L/kg}$ . For the entire set of 26  $K_d$  estimates, the mean was 7.75  $\text{L/kg}$  with a standard deviation of 11.7  $\text{L/kg}$ .

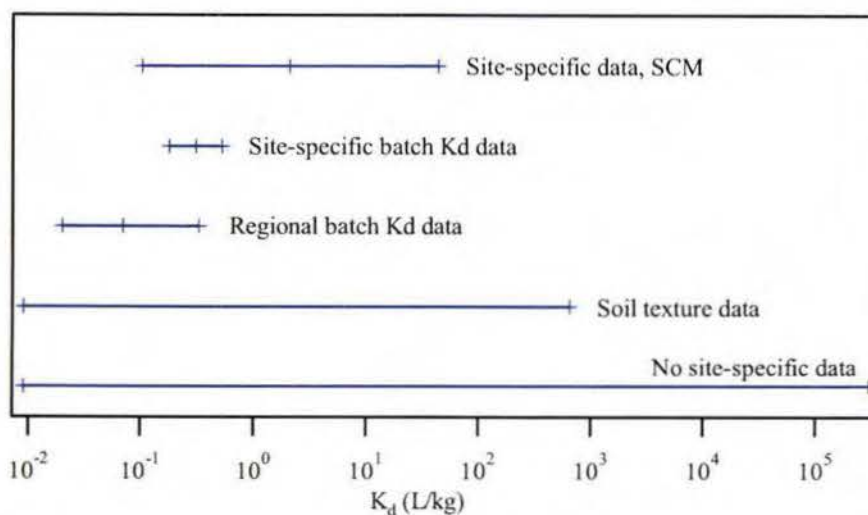
The mean values for the inland and near-river  $K_d$  estimates were used as prior estimates for the two  $K_d$  parameters in Models 6 and 8. For Models 2 and 4 in which the  $K_d$  was uniform, the mean value for the en-



tire set of  $K_d$  estimates was used as the prior estimate. For all models, the weight applied to the prior estimates was taken as the inverse of the standard deviation of the entire set of  $K_d$  estimates multiplied by the number of estimates (26 in this case). Because the log of  $K_d$  was estimated in the calibration, the prior values and the weight were computed on the log  $K_d$  estimates. Prior estimates and the weight for the log( $K_d$ ) parameters are given in Table 4-7 for each of the model alternatives.

**Table 4-6.  $K_d$  values estimated for near-river wells using a surface complexation model**

Well ID	$K_d$ (L/kg)
399-1-10A	7.3
399-1-16A	5.5
399-3-10	31.7
399-3-9	51.1
399-4-10	18.9
399-4-7	1.5
399-4-9	25.0



**Figure 4-21. Ranges of  $K_d$  values for the Hanford Site 300 Area derived using a hierarchy of data**

**Table 4-7. Prior estimates and calibration weight for the log of distribution coefficient parameters**

Parameter	Model Applicability	Prior Estimate	Weight
Kd	2 and 4, full domain	-2.46443	45.51
Kd1	6 and 8, inland	-2.67047	45.51
Kd2	6 and 8, near-river	-1.905192	45.51

#### 4.6 Alternative Hydrologic Scenarios

The baseline hydrologic scenario for the uncertainty assessment application to uranium transport at the 300 Area was based on a continuation of current conditions

at the site. That is, it was assumed that the current hydrologic conditions at the site will continue in the future. This was implemented by simulating a 20-year period from January 2005 through December 2024 with the river boundary condition taken to be the river stage during the period 1975 to 1994. This boundary condition was selected for convenience. In a more sophisticated analysis, the river stage could be modeled as a random variable (correlated in time) sampled from a distribution consistent with the statistics of the observed river stage.

When considering alternative hydrologic scenarios at the site, several changes to the future hydrologic conditions can be reasonably proposed. (1) Future land uses in the surrounding area may change as restoration of the Hanford Site progresses. One such change considered for a performance assessment of waste disposal at the Hanford Site is agricultural development of land after completion of Site restoration (anticipated to be mid-21<sup>st</sup> century). Evans et al. (2000) examined the potential for agricultural development at the Hanford

Site and estimated recharge rates resulting for the anticipated irrigation. Mixed use development is also being considered for the site. (2) The operation of dams on the Columbia River may change as a result of changes to dam structures, dam aging, and water use priorities. These changes will have an impact on the magnitude and variability of river stage at the 300 Area. (3) Climate change is anticipated to impact the hydrology within the Columbia River basin in a number of ways (Leung et al. 2004). Changes in precipitation and temperature may directly impact the recharge rate at the 300 Area and may indirectly impact the river stage through resulting changes in dam operations.

One alternative hydrologic scenario was considered in this application. This scenario assumed that changes to dam operation results in Columbia River flows that are closer to its free-flowing state. This was implemented by simulating a 20-year period from January 2005 through December 2024 with the river boundary condition taken to be the river stage during the period 1955 to 1974. The river stage during this time period reflects the much higher peak spring/summer flows that existed prior to the completion of the Canadian dams on the Columbia River (Figure 4-19). While this scenario is not plausible in the near-term, it was used here to demonstrate the uncertainty methodology in this application using a scenario that has a significant impact on flow and transport.

## 4.7 Prior Scenario Probability

As discussed in Section 2.3.1, since scenarios describe future conditions, prior scenario probabilities must be assigned to each scenario based on a subjective understanding and judgment of the plausibility of the scenario. While formal methods could be used to estimate scenario probabilities [e.g., expert judgment and the approach of Ye et al. (2005) could be applied], for this application, arbitrary prior scenario probabilities were used to illustrate the uncertainty assessment methodology. A prior probability of 0.7 was assigned to the baseline hydrologic scenario and a probability of 0.3 was assigned to the alternative scenario.

## 4.8 Model Calibration

Model calibration (inverse modeling) was conducted to obtain maximum likelihood parameter estimates so that the *NLL* (Eq. 11) and Fisher information matrix (Eq. 12) could be calculated to compute *KIC* (Eq. 10) and posterior model probability (Eq. 8). The calibration was implemented using PEST, whose utilities for MODFLOW and MT3DMS parameter estimation were used to link MODFLOW/MT3DMS/PEST. The utilities contain a series of subroutines to extract simulated

heads and concentrations from MODFLOW and MT3DMS output files corresponding to measured heads and concentrations selected for model calibration. Residuals between the simulated and measured heads and concentrations form an objective function, which was minimized iteratively to find optimum model parameters. In this application, prior information on the  $K_d$  parameter(s) was used and included in the objective function.

### 4.8.1 Selection of Calibration Data

The available head and concentration observations in the 300 Area for the period of September 1997 to December 2004 were reviewed for use in the calibration of the alternative models. Observations from 21 wells were selected to provide spatially and temporally representative coverage (Figure 4-22). Three wells were screened in the lower portion of the aquifer (well numbers are indicated in Figure 4-22) while the remaining 18 wells were screened near the top of the aquifer. A subset of data from each of the 21 wells was used in the calibration with the general goal of two observations per year at each well if available. The distribution in time of the selected calibration data is shown in Figure 4-23. In total there were 430 observations used in the calibration, of which 222 were head measurements and 208 were concentration measurements.

### 4.8.2 Weights Associated with Calibration Data

The weights for head measurements were proportional to the inverse of estimated observation errors and included components related to well-altitude error, well-location error, nonsimulated transient error, and measurement error. More discussion of these individual error components can be found in Belcher et al. (2004). At the Hanford Site, the well-location error was small enough to be negligible (Vermeul et al. 2003). The well-altitude error was less than 0.01 m, given a maximum head gradient of about 0.25%. The measurement error was about 0.03 m corresponding to a depth of observation of 30 m. To incorporate nonsimulated transient error, final head observation error was considered as 0.06 m.

According to Hill (1998), given a measurement accuracy,  $a_h$ , so that  $h - a_h \leq h^* \leq h + a_h$  ( $h^*$  and  $h$  being "true" and measured head values), the standard deviation of observation error is  $\sigma_h = a_h / 1.96$ . This assumes that head observation error follows a normal distribution with zero mean, and 1.96 is determined from the 95% confidence interval. Therefore, the

weights associated with head observations can be estimated via

$$w_h = \frac{1}{\sigma_h} = \frac{1}{a_h / 1.96} = \frac{1.96}{a_h} \quad h - a_h \leq h^* \leq h + a_h \quad (36)$$

With a head observation error of  $a_h = 0.06$  m, this gives a uniform weight of 32.667 for each head observation used in the calibration.

The weights associated with concentration observations were initially estimated using lab-reported analytical errors for uranium concentration measurements at well 399-1-17A, a well near the southern boundary of the 316-5 Process Trenches with a long history of detectable uranium concentrations. The total analytical error reported is shown in Figure 4-24 and can be seen to be an approximately constant proportion of the reported uranium concentration. Assuming normally distributed measurement errors and a 95% probability that the true uranium concentration is within the range defined by

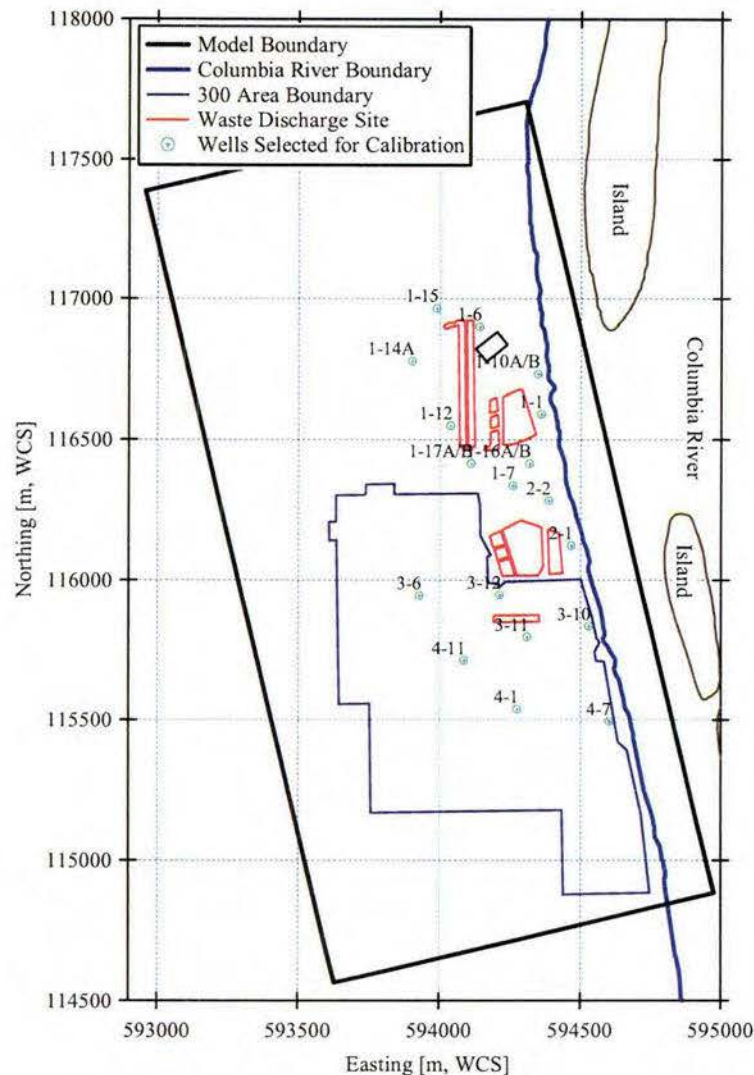
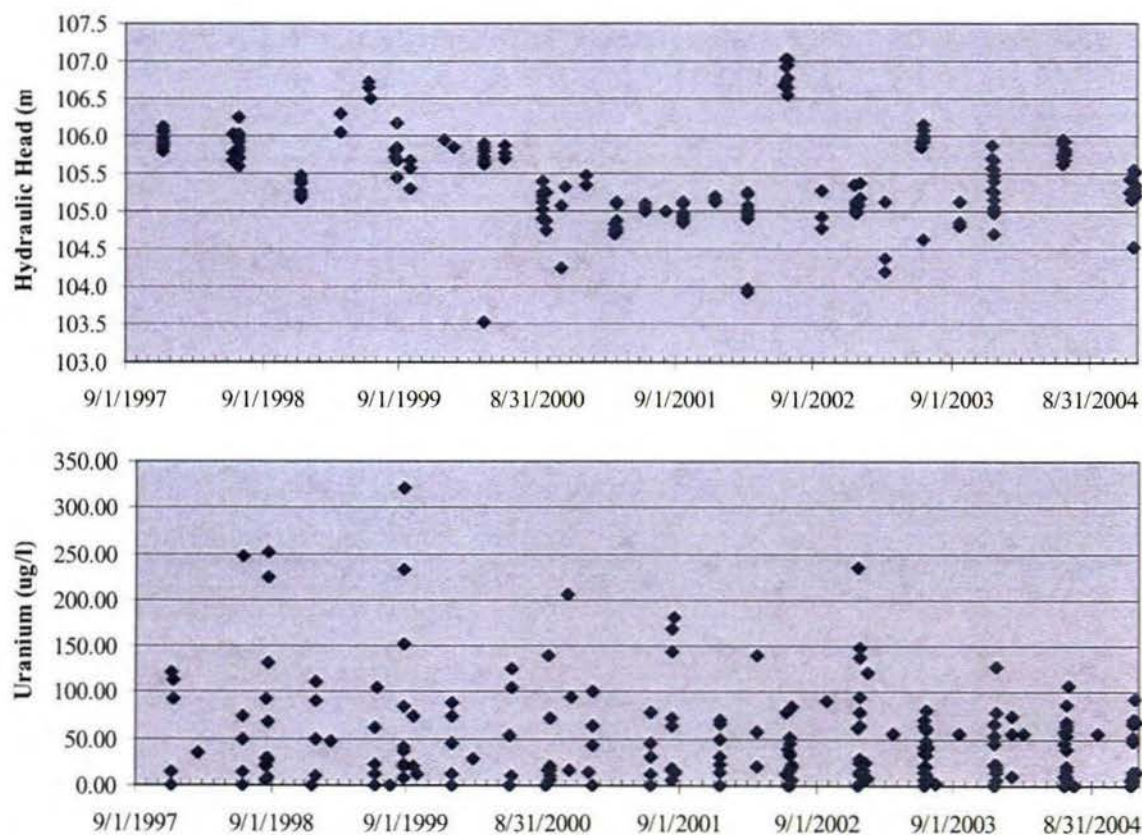


Figure 4-22. Distribution of observation wells from which calibration data were selected





**Figure 4-23. Distribution of calibration data in time for (top) head and (bottom) concentration observations**

the measured concentration plus or minus the reported analytical error, the standard deviation of the measurement errors was estimated as the reported analytical error divided by 1.96. The resulting concentration weights decrease as the observed concentrations increase. The calibration thus emphasized those observations where the concentrations were relatively small. The resulting weighted concentration residuals were not independent of the weighted simulated concentrations as shown in the upper plot of Figure 4-25. Observations with very low values of concentration (such as those from the deeper wells) also contributed an inordinate amount to the calibration objective function.

To improve the calibration results, equal weights were applied to the concentration observations. The weight selected, 0.349, was the median of the weights computed as inversely proportional to the observed concentrations. This corresponds to a standard deviation of observation error of  $2.87 \mu\text{g/l}$ , or a uniform measurement error of  $5.62 \mu\text{g/l}$ . The weighted concentration residuals from a calibration with equal concentration

observation weights were independent of the weighted simulated concentration as shown in the lower plot of Figure 4-25. Weighted concentration residuals increased with simulated concentrations, suggesting that the weights on the lowest concentration observations may be too small.

#### **4.8.3 Joint Model Calibration and Calibrated Model Parameters**

Joint calibration of the flow and transport models was necessary, since the Fisher information matrix (Eq. 12) of MLBMA must be obtained simultaneously by jointly calibrating the flow and transport processes. This results in joint calibration of maximum likelihood estimates for flow and transport parameters. In addition, Sun and Yeh (1990) have shown that using concentration observations in a joint calibration can improve the calibration of flow model parameters.

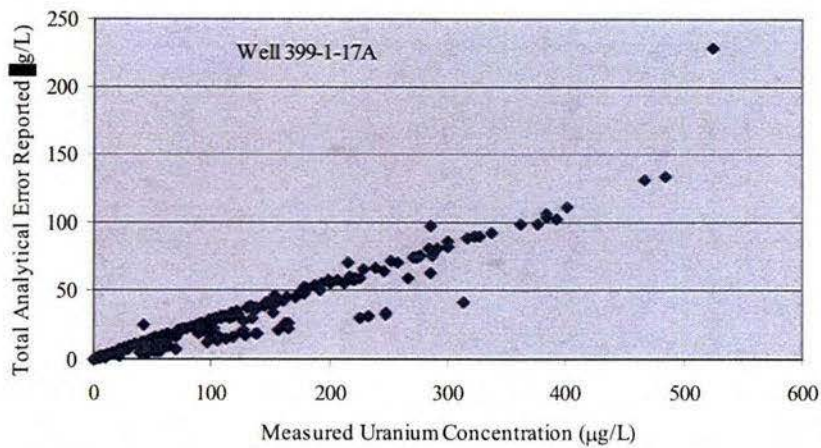


Figure 4-24. Plot of total analytical error with observation of uranium concentration at well 399-1-17A

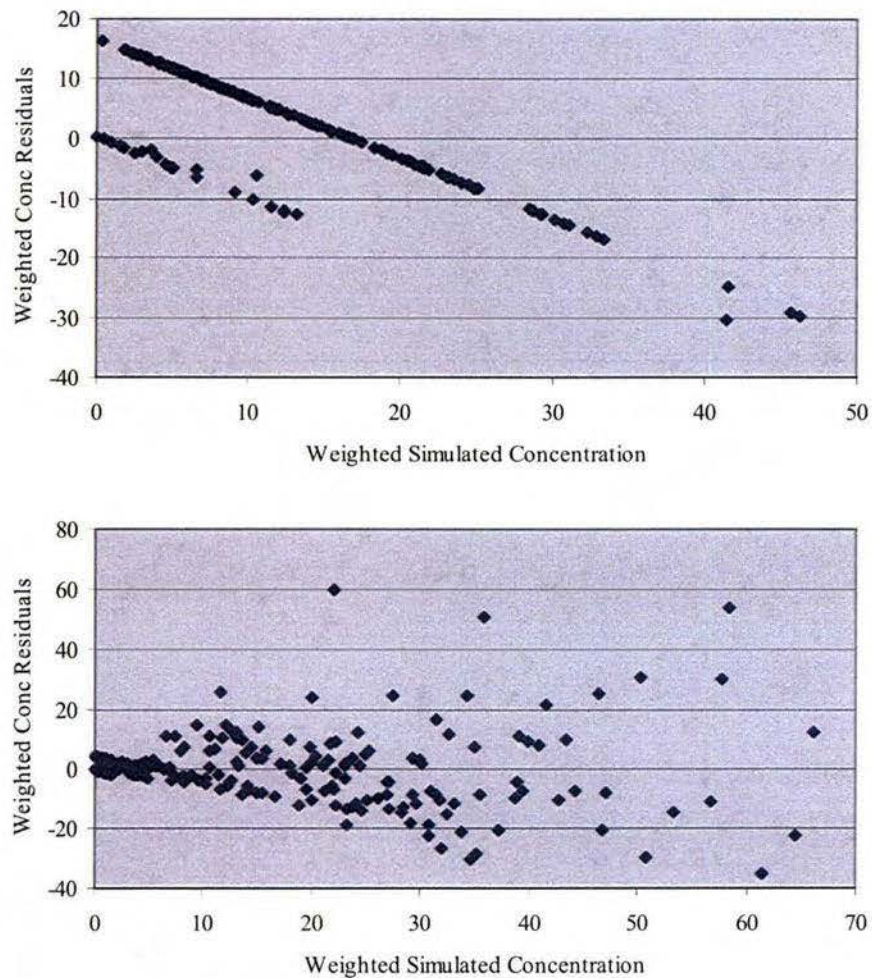


Figure 4-25. Weighted concentration residuals versus weighted simulated concentrations for Model 4 using (top) concentration weights inversely proportional to observed concentrations and (bottom) equal concentration weights



Following Medina and Carrera (1996), *NLL* (Eq. 11) of MLBMA was defined as

$$NLL = N \ln(2\pi) + \ln|C| + (\mathbf{d} - \mathbf{d}^*)^T C^{-1} (\mathbf{d} - \mathbf{d}^*) \quad (37)$$

where  $N$  are the number of head and concentration observations plus the number of prior parameter estimates,  $\mathbf{d}^*$  is a vector of observations and prior parameter estimates (length  $N$ ),  $\mathbf{d}$  is a vector of computed head, concentration, and parameter values (length  $N$ ), and  $C$  is the combined covariance matrix of the observations and prior parameter estimates. The covariance matrix was taken to be diagonal with elements equal to the inverse of the weights determined in Section 4.8.2 for the head and concentration observations and Section 0 for the prior parameter estimates. As the first two terms on the right hand side of (Eq. 38) were constant, the PEST implementation used the objective function

$$S = (\mathbf{d} - \mathbf{d}^*)^T C^{-1} (\mathbf{d} - \mathbf{d}^*) \quad (38)$$

which was minimized iteratively using the Gauss-Levenberg-Marquardt method (Doherty 2004).

Preliminary parameter sensitivity analysis examined the impact on the flow and transport model results of horizontal hydraulic conductivity (e.g.,  $Kh\_1$  for unit 1), the ratio of horizontal to vertical hydraulic conductivity ( $VANI$ ), the specific yield of unit 1 ( $Sy\_1$ ), porosity, longitudinal dispersivity ( $\alpha_{pL}$ ), the ratio of vertical to longitudinal dispersivity ( $R\text{-}\alpha_{pV}$ ), and the linear adsorption distribution coefficient ( $K_d$ ). The method of Morris (1991) was used to estimate the distributions of elementary effects for these parameters. The mean of each distribution represents the first-order effect of each parameter. The standard deviation of each distribution represents the effect of parameter interactions (without information about which parameters are interacting). Results for Model 4 are presented in Figure 4-26. The distribution of values represented by the box plots in this figure results from computing the sensitivities for each of the model outputs used in the calibration (all head and concentration observations). The most important parameters to include in the calibration, as suggested by the sensitivity analysis, were the horizontal hydraulic conductivity of units 1 and 5, the linear adsorption distribution coefficient, longitudinal dispersivity, and the ratio of vertical to longitudinal dispersivity. Model results were relatively insensitive to the remainder of the considered parameters.

Preliminary calibrations confirmed the sensitivity results. Calibration results were more stable when the

insensitive parameters were not included in the estimation process. The calibration results were also improved by fitting the log of the hydraulic conductivity and distribution coefficient parameters. In addition, because of a high correlation between  $\alpha_{pL}$  and  $R\text{-}\alpha_{pV}$ , the final calibrations did not fit  $R\text{-}\alpha_{pV}$ . The values of all non-fitted parameters were fixed during calibration at estimates based on literature values and professional judgment.

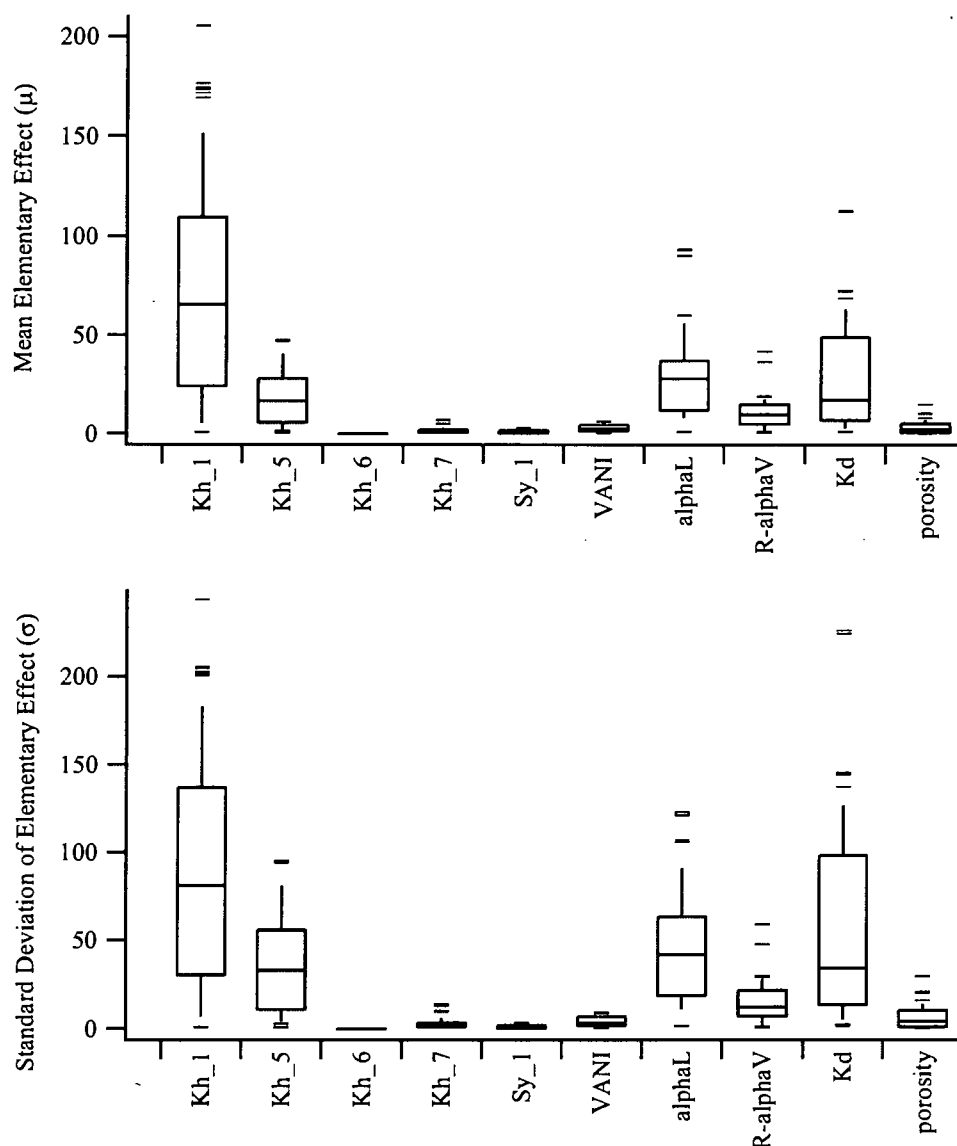
Table 4-8 lists initial values of the calibrated parameters and the ranges used to constrain parameter variation within PEST for the four alternative models. Parameter ranges were set large enough to ensure that the optimal parameter values fell between the limits. In this table, the homogeneous horizontal hydraulic conductivity of models 2 and 6 was denoted by  $Kh$ . The heterogeneous hydraulic conductivity of models 4 and 8 was denoted by  $Kh\_1$  and  $Kh\_5$  for units 1 and 5. The homogeneous distribution coefficient of models 2 and 4 was denoted as  $K_d$  while the heterogeneous  $K_d$  for zones 1 and 2 of models 6 and 8 were denoted as  $K_{d1}$  and  $K_{d2}$ , respectively. The initial parameter values were based on preliminary calibration results for  $Kh\_1$ , Table 4-1 for  $Kh\_5$ , mean values from the surface complexation model for the  $K_d$  parameters, and judgment for  $\alpha_{pL}$ . Parameters that were fixed in the calibrations are listed in Table 4-9. In most cases, values of the fixed parameters were based on previous model applications at the Hanford Site or the literature.

## 4.9 Model Calibration Results

Each of the four alternative models was calibrated over the period of September 1997 to December 2004 using the same set of head and uranium concentration observations. As stated previously, the flow and transport models were calibrated jointly. Final calibrated parameter values for each model are listed in Table 4-10 and presented graphically in Figure 4-27. Of the parameters estimated, sensitivities to  $Kh\_5$  and  $\alpha_{pL}$  were the lowest. This is reflected in the large differences between models for the calibrated values of these parameters and in the relatively large confidence limits for  $Kh\_5$ .

Calibrated results for the hydraulic conductivity parameters differ from the current average estimates for the 300 Area given in Table 4-1, but are consistent with the ranges of values based on the field tests and previous inverse modeling for the Hanford Site (Table 4-1). Calibrated hydraulic conductivity for the homogeneous models was significantly less than the unit 1 hydraulic conductivities from the heterogeneous models.





**Figure 4-26. Estimated means (top) and standard deviations (bottom) of the elementary effects of Model 4 computed using the method of Morris (1991). Boxplots represent the distribution over all head and concentration measurements used in the calibration. Median, 10, 25, 75, and 90 percentiles shown; outliers are outside 5 and 95 percentiles.**

Calibrated values of the dispersivity were small for the homogeneous models – much smaller than the grid cell size in the primary flow direction (35 m) and very close to the initial value. For the heterogeneous models, the calibrated values of the dispersivity were much larger, on the order of the grid cell size.

Calibrated values for the  $K_d$  parameters were greater than the mean values computed using the surface com-

plexation model results (Sections 4.5.2.5 and 0) but were within the estimated limits derived from the geochemical modeling. Calibration of Model 6 resulted in nearly equal values for Kd1 and Kd2 while the Model 8 calibration resulted in a value of Kd1 greater than Kd2. Each of these results is contrary to the postulated geochemical conceptualization and indicates that a more sophisticated geochemical model than used here may be necessary.

**Table 4-8. Initial values and upper and lower limits (in brackets) for parameters estimated in the calibration. Empty spaces indicate that the parameter was not present in a model. Log-transformed parameters were calibrated for all parameters except alphaL.**

	<b>Kh</b> (m/d)	<b>Kh_1</b> (m/d)	<b>Kh_5</b> (m/d)	<b>alphaL</b> (m)	<b>Kd</b> (L/kg)	<b>Kd1</b> (L/kg)	<b>Kd2</b> (L/kg)
<b>Model 2</b>	15000 [1,2x10 <sup>5</sup> ]			3.0 [0.1,100]	8.0 [0.1,500]		
<b>Model 4</b>		15000 [1,2x10 <sup>5</sup> ]	150 [0.1,5000]	3.0 [0.1,100]	8.0 [0.1,500]		
<b>Model 6</b>	15000 [1,2x10 <sup>5</sup> ]			3.0 [0.1,100]		3.0 [0.1,500]	20.0 [0.1,500]
<b>Model 8</b>		15000 [1,2x10 <sup>5</sup> ]	150 [0.1,5000]	3.0 [0.1,100]		3.0 [0.1,500]	20.0 [0.1,500]

**Table 4-9. Values of parameters that were fixed in the calibrations**

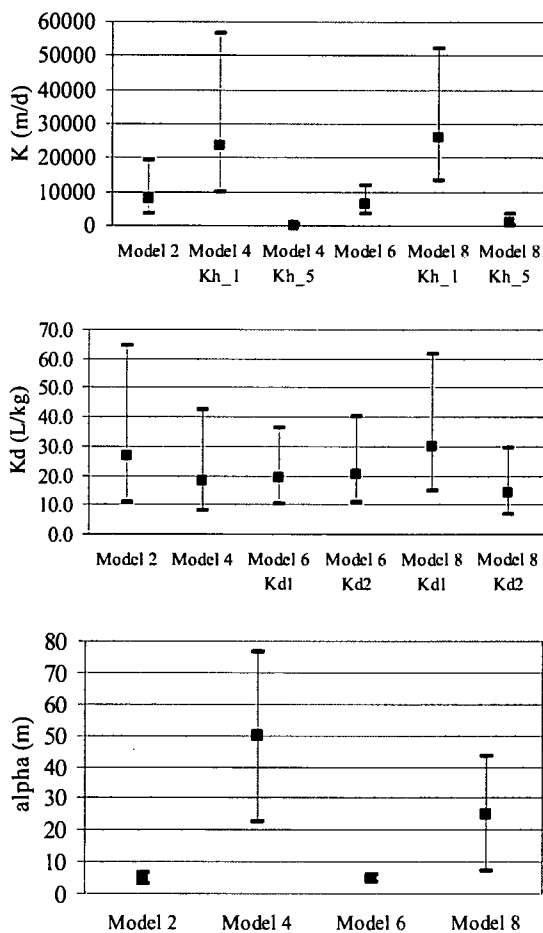
<b>Parameter</b>	<b>Value</b>
Kh_6	0.01 m/d
Kh_7	43 m/d
Kh_8	5.0x10 <sup>-5</sup> m/d
Sy	0.2
VANI	10
Ratio of transverse to longitudinal dispersivity	1.0
R-alphaV	0.1
Bulk Density	2100 kg/m <sup>3</sup>
Molecular Diffusion Coefficient	8.64x10 <sup>-6</sup> m <sup>2</sup> /d

The initial and final calibration objective function components are listed in Table 4-11 for each alternative model. The total sum of squared weighted residuals (SSWR) is the sum of the SSWR due to the head and concentration observations as well as the prior estimates of the  $K_d$  parameters (Eq. 39). The total objective function favors Model 4 over Model 8 by a small margin and favors the models with heterogeneous hydraulic conductivity (Models 4 and 8) by a wide margin over the homogeneous models (Models 2 and 6).

The better fit of Models 4 and 8 is due primarily to the improved fit to the head observations and primarily to the improved head fit at wells 3-6, 4-1, and 4-11 (see Figure 4-22 for the locations of these wells). The SSWR for the head observations actually increased during the calibration for Models 2 and 6. This was apparently necessary to decrease the SSWR for the concentration observations from their large initial values. The initial concentration SSWR for Models 2 and 6 was significantly larger than for Models 4 and 8.

**Table 4-10. Calibrated parameter values and 95% upper and lower linear confidence limits (in brackets)**

	<b>Kh</b> (m/d)	<b>Kh_1</b> (m/d)	<b>Kh_5</b> (m/d)	<b>alphaL</b> (m)	<b>Kd</b> (L/kg)	<b>Kd1</b> (L/kg)	<b>Kd2</b> (L/kg)
<b>Model 2</b>	8097 [3385,19371]			4.9 [3.2,6.6]	26.3 [10.8,64.2]		
<b>Model 4</b>		23,508 [9733,56775]	11 [0.4,263]	49.6 [22.8,76.3]	18.0 [7.6,42.2]		
<b>Model 6</b>	6237 [3261,11926]			4.6 [3.3,5.9]		19.0 [9.9,36.2]	20.5 [10.5,40.1]
<b>Model 8</b>		26,093 [13079,52055]	780 [179,3410]	25.0 [6.9,43.1]		30.2 [14.8,61.4]	14.3 [7.0,29.1]



**Figure 4-27. Comparison of final calibrated parameter values: (top) hydraulic conductivity parameters, (middle) adsorption parameters, and (bottom) dispersivity**

There were relatively small differences between models in the final fit to the concentration observations. Model 8 with its two  $K_d$  zones fits both the head and concentration observations slightly better than Model 4, but the contribution to the objective function from the  $K_d$  prior component is significantly larger for Model 8.

Plots of simulated versus observed heads and uranium concentrations are shown in Figure 4-28 for Model 4. Results for the other models appear similar and are not shown here. Observed heads are underestimated for the highest heads and overestimated for the lowest heads, primarily as a result of the fixed head on the western boundary of the model. The pattern is similar for the uranium concentrations with the largest concentrations being somewhat severely underestimated. These results

suggest that there is room for improvement in the models.

A comparison between observed and simulated values at selected wells is shown in Figure 4-29 for two wells located between the major uranium sources. Well 399-2-2 is close to the river while well 399-1-17A is farther inland. (See Figure 4-22 for well locations). Calibration data indicate observations that were included in the calibration. Measured data are observations that were not included in the calibration. There are only minor differences between heads simulated by the four model alternatives at these wells. For concentrations, Models 2 and 6 produce similar results, but there are differences in the results of the other two models. In general, the calibrated models appear to fit the observed heads better than the observed concentrations. Figure 4-30 presents calibration results for two wells located south of the major uranium sources. Well 399-3-10 is located near the river while well 399-4-11 is much farther inland. Models 2 and 6 produce very similar results at these wells for both heads and concentrations. Unlike the other wells, at well 399-4-11 there are noticeable differences between the heads simulated by the models.

Simulated uranium concentrations at the end of December 2004 are compared in Figure 4-31 to the observed values from the fourth quarter of 2004. Results are shown for Model 4. Simulated values are presented as banded color contours where the color interfaces represent simulated concentration values of 10, 30, 50, 70, 90, and 110  $\mu\text{g/l}$ . The observed values are represented by labeled contours at 10, 30, 50, and 70  $\mu\text{g/l}$ . In general, simulated concentrations appear to be somewhat lower than the observed concentrations except near the river where there are two regions of elevated concentration in the simulated results. Overall, the simulated results reflect the general character of the observed concentrations at the end of the calibration period.

**Table 4-11. Calibration objective function components for each alternative model**

Model:		2	4	6	8
SSWR_head	Initial	32405	29949	32405	29949
	Final	32824	28595	33104	28170
SSWR_conc	Initial	93077	44495	96113	48351
	Final	30219	30404	30474	29869
SSWR_prior	Initial	280	280	133	133
	Final	1619	1070	1961	2747
SSWR	Initial	125761	74724	128651	78433
	Final	64662	60069	65539	60786



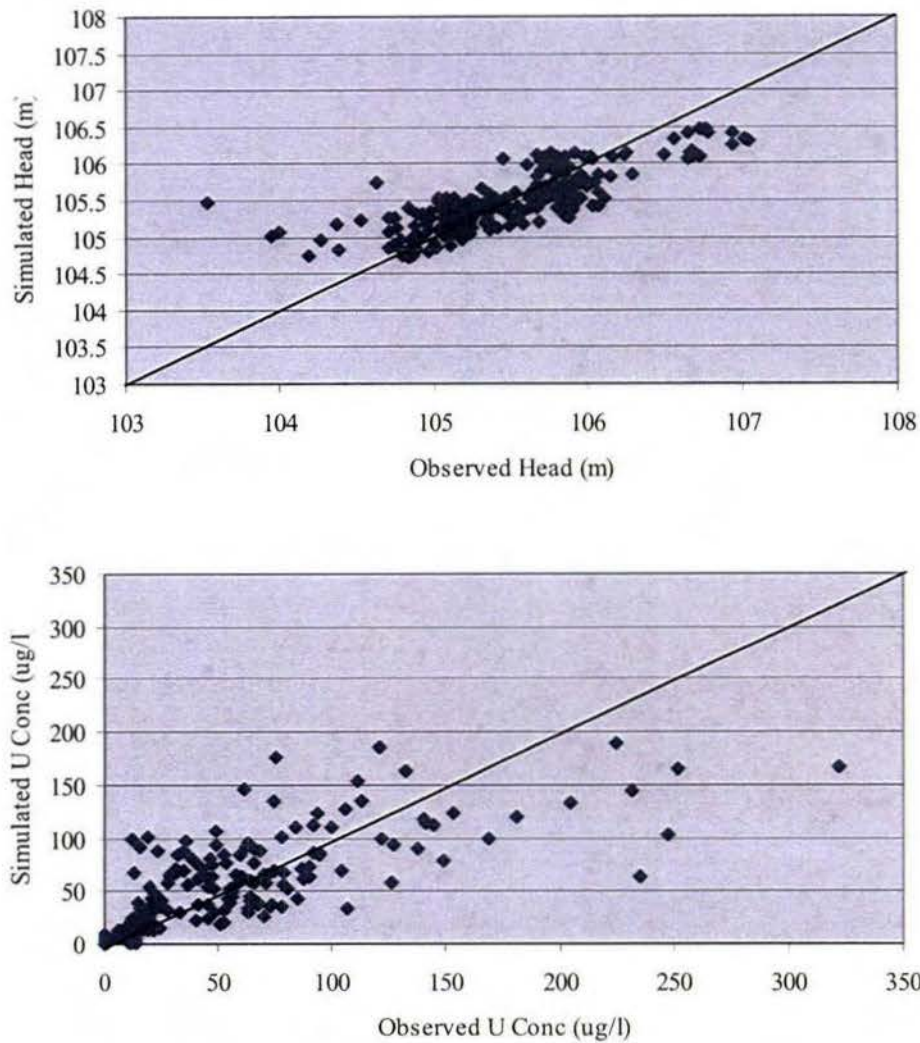


Figure 4-28. Simulated versus observed (top) heads and (bottom) uranium concentrations for Model 4

#### 4.10 Posterior Model Probabilities

The results of the model calibrations were used to compute posterior model probabilities as described in Section 3.1.1. Equation (8) provides the expression for posterior model probabilities and contains two needed quantities: the prior model probabilities, which were assumed to be an equal 25% for the four models of this application, and the  $\Delta KIC$  value for each model. The expression for  $KIC$  (Eq. 10) contains three terms. The negative log likelihood term was computed here as

$$NLL = N \ln(\sigma^2) \quad (39)$$

Where  $\sigma^2$  is the maximum likelihood estimate of the error variance

$$\sigma^2 = \frac{SSWR}{N} \quad (40)$$

$NLL$  represents the fit to the observations used in the calibration. The smaller this value, the better the fit. As can be seen in Table 4-12, Model 4 has the smallest value of  $NLL$  followed by Model 8, Model 2, and Model 6.

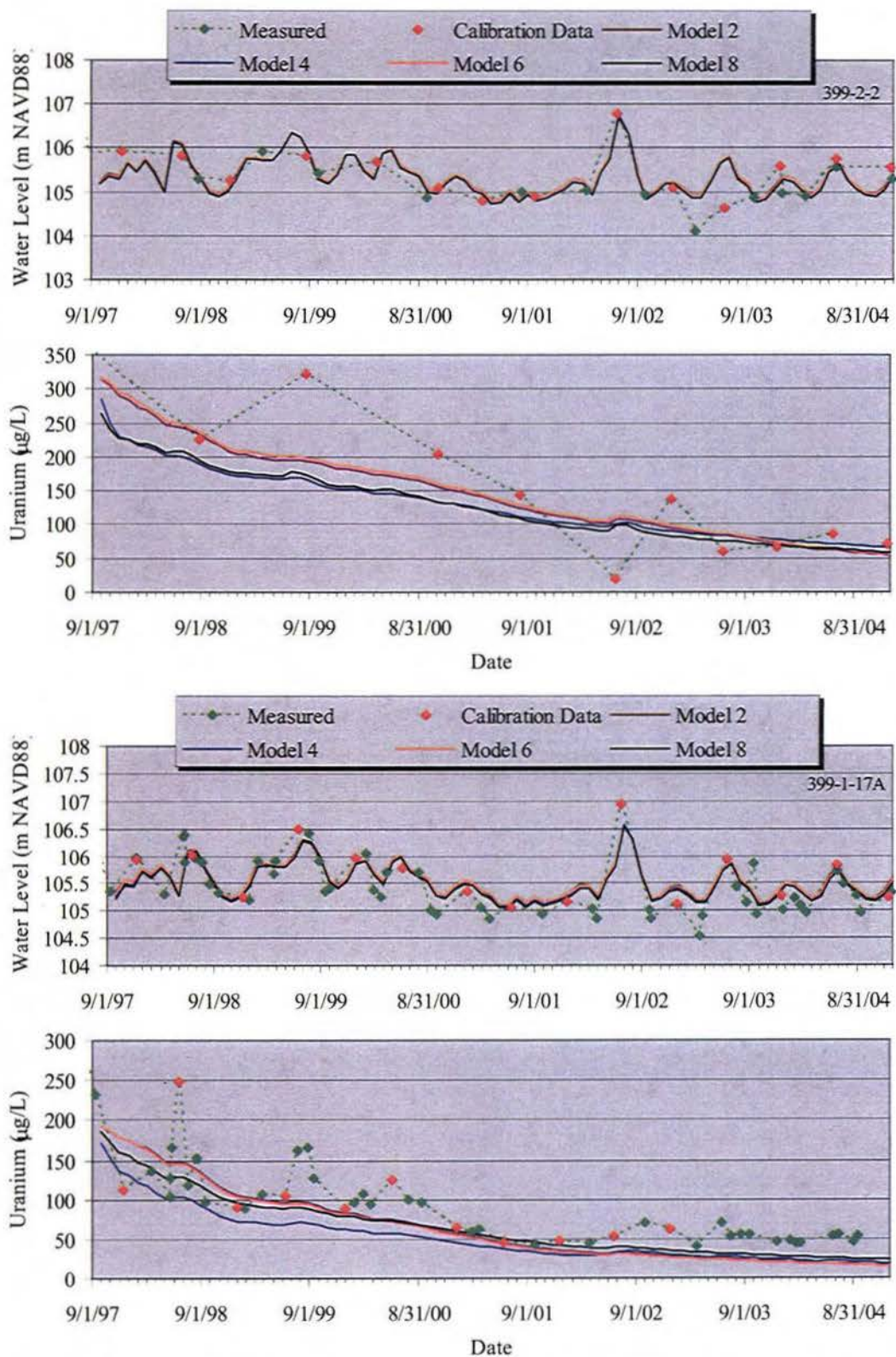


Figure 4-29. Calibration results for (top two plots) well 399-2-2 and (bottom two plots) well 399-1-17A



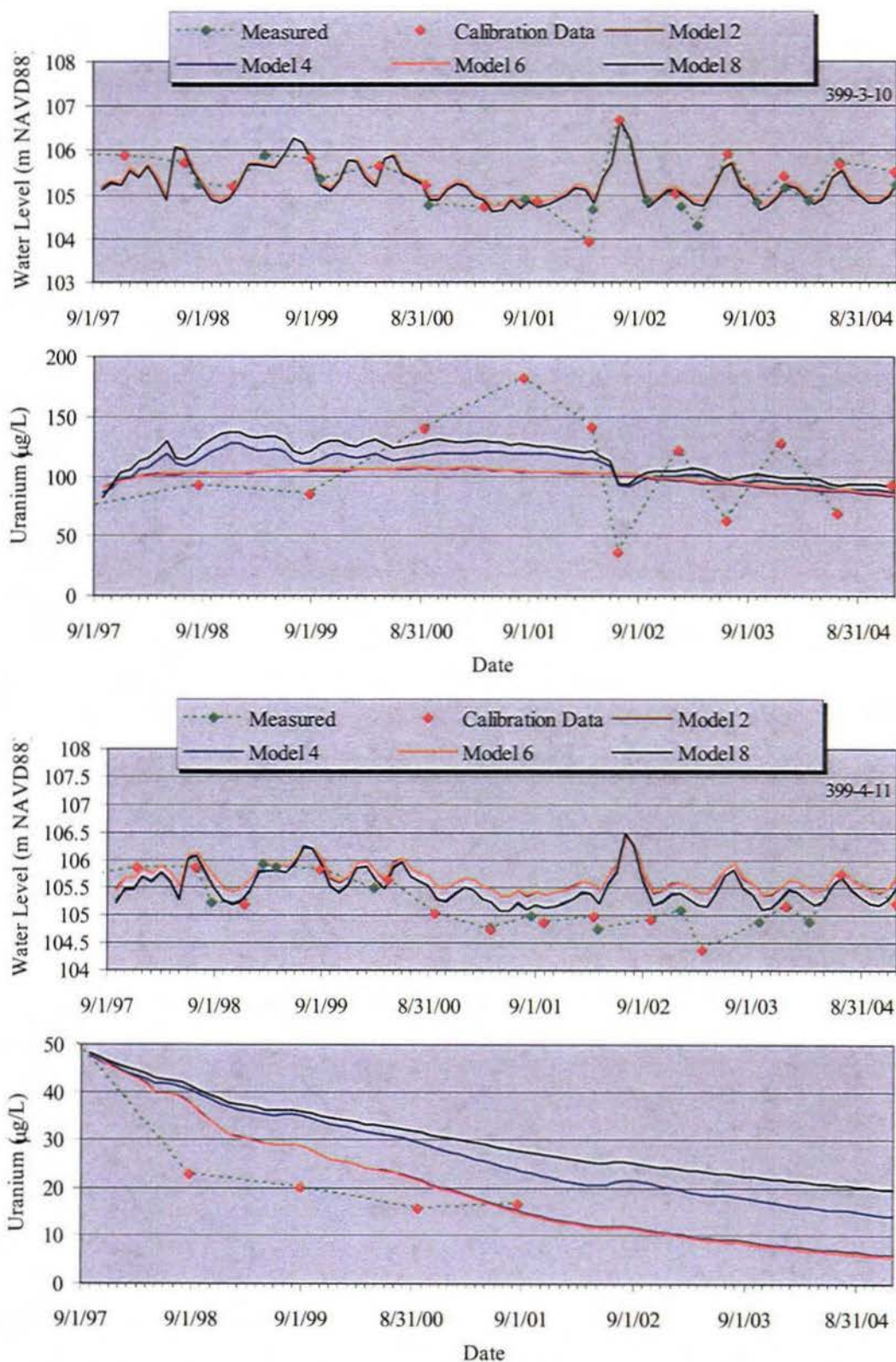
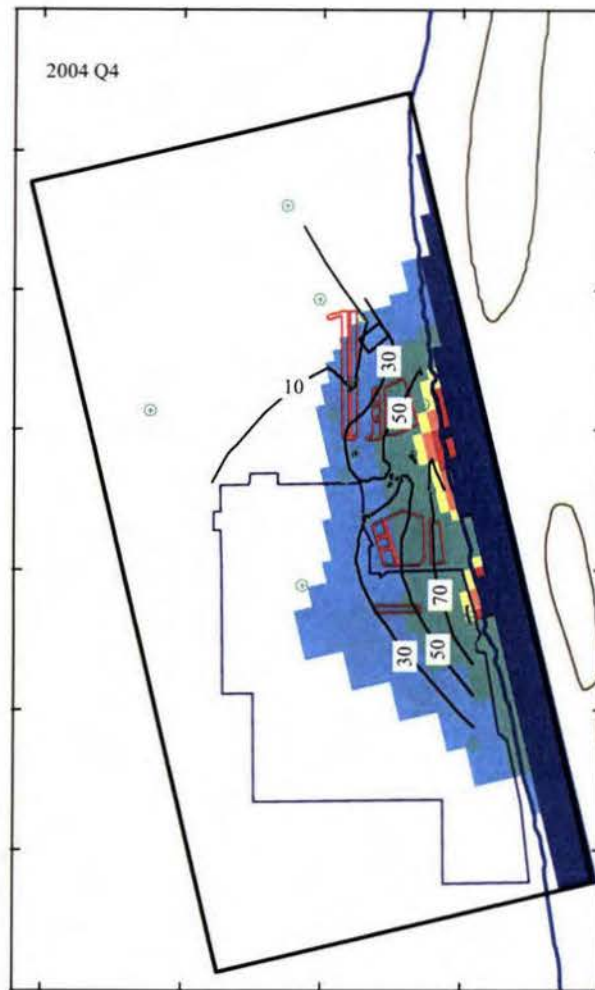


Figure 4-30. Calibration results for (top two plots) well 399-3-10 and (bottom two plots) well 399-4-11





**Figure 4-31. Model 4 simulated uranium concentrations in December 2004 (color) overlain by contours based on observed uranium concentration ( $\mu\text{g/l}$ ) in the fourth quarter of 2004. Color levels indicate the 10, 30, 50, 70, 90, and 110  $\mu\text{g/l}$  locations. Dark blue indicates the Columbia River.**

The second term in  $KIC$ ,  $N_k \ln(N/2\pi)$ , is a function of the number of calibrated parameters,  $N_k$ , which varies between models, and the number of observations, which is constant for all models. This term is sometimes interpreted as a representation of model complexity. In this sense,  $KIC$  favors models that are less complex. Values of this term are given in Table 4-12 and can be seen to favor Model 2 over all other models and Models 4 and 6 over Model 8.

The third term in the computation of  $KIC$  is the normalized observed Fisher information matrix (Eq. 12). This term can be expressed as

$$\ln|\bar{\mathbf{F}}| = \ln\left|\frac{1}{N}\mathbf{I}\right| = \ln(N^{-N_k}|\mathbf{I}|) = -N_k \ln N + \ln|\mathbf{I}| \quad (41)$$

where  $\mathbf{I}$  is the observed Fisher information matrix, computed here as

$$\ln|\mathbf{I}| = \ln|\Sigma^{-1}| = -\ln|\Sigma| \quad (42)$$

and  $\Sigma$  is the maximum likelihood estimate of the parameter covariance matrix. If the parameter covariance matrix is reported by the parameter estimation software, the determinant can be computed directly with care taken to observe two requirements.

**Table 4-12. Calibration results and computation of model probabilities using *KIC***

Model:	2	4	6	8
SSWR	64662	60069	65539	60786
$N_k$	3	4	4	5
$\sigma^2 = \text{SSWR}/N$	150.38	139.70	152.42	141.36
$NLL = M\ln(\sigma^2)$	2155.65	2123.97	2161.44	2129.07
$N_k \ln(N/2\pi)$	12.68	16.90	16.90	21.13
$\ln  \bar{\mathbf{F}} $	-20.20	-38.69	-13.48	-39.71
<i>KIC</i>	2148.13	2102.18	2164.87	2110.49
Rank	3	1	4	2
$\Delta KIC$	45.95	0.00	62.69	8.32
$p(M_k \mathbf{D})$	0.00%	98.46%	0.00%	1.54%

For the first requirement, if the parameter covariance is reported for transformed parameters, as is done in PEST, the covariance must be recomputed for the untransformed parameters. For PEST this can be accomplished by setting the parameter NOPTMAX=-1 in the PEST control file and rerunning PEST without parameter transformations and with initial parameter values set to the optimal values.

For the second requirement, if the parameter estimation software uses least squares (as does PEST) instead of maximum likelihood, the parameter covariance reported is the least squares estimate and must be converted to the maximum likelihood estimate as follows.

$$\Sigma = \frac{N - N_k}{N} \Sigma_{LS} \quad (43)$$

where  $\Sigma_{LS}$  is the least squares estimate of the parameter covariance matrix.

PEST reports the eigenvalues of the (least squares) parameter covariance matrix, which were used to evaluate Eq. (42),

$$\begin{aligned} \ln |\Sigma| &= \ln \left| \frac{N - N_k}{N} \Sigma_{LS} \right| = \ln \left( \frac{N - N_k}{N} \right)^{N_k} + \ln |\Sigma_{LS}| \\ &= N_k \ln \left( \frac{N - N_k}{N} \right) + \sum_{i=1}^{N_k} \ln \lambda_i \end{aligned} \quad (44)$$

where  $\lambda_i$  are the eigenvalues. The third term of *KIC* was thus computed using the PEST output as

$$\ln |\bar{\mathbf{F}}| = -N_k \ln N - N_k \ln \left( \frac{N - N_k}{N} \right) - \sum_{i=1}^{N_k} \ln \lambda_i \quad (45)$$

Eigenvalues were evaluated by running PEST with the control parameter NOPTMAX=-1 and with all parameters untransformed and set at their optimal values.

As discussed in Section 3.1.1, the Fisher information term causes *KIC* to favor models with relatively small information content per unit sample or, equivalently, a correspondingly large parameter estimation variance. Values of  $\ln |\bar{\mathbf{F}}|$  for the four alternative models are given in Table 4-12. It can be seen that this term favors Model 8 over Model 4 (slightly) with Model 2 and Model 6 having much larger values.

Final *KIC* values are given in Table 4-12 with the result that Model 4 is preferred over Models 8, 2, and 6 in that order. The  $\Delta KIC$  value for Model 8 is 8.32 with Models 2 and 6 having much higher values. With equal prior probabilities, the posterior probabilities were computed using Eq. (8) and are also given in Table 4-12. The result is that Model 4 is preferred to the near exclusion of the other models. Referring to Figure 3-1, it can be seen that the sensitivity of model probability to  $\Delta KIC$  means that the small differences in the three terms of *KIC* can produce a large difference in posterior model probabilities.

#### 4.11 Prediction

The fact that Model 4 dominates the other alternatives in terms of posterior model probabilities means that only Model 4 would need to be applied in the predictive period. To better illustrate the full application of the uncertainty methodology, however, we used all four models for prediction under the baseline scenario. For the alternative scenario, only Model 4 was applied.

There is an independent reason to apply even low-probability models in the predictive period. If the conditions of the predictive period are significantly different than the calibration period, differences between models that were not apparent in the calibration may appear during prediction. Applying even low-probability models in the predictive period may bring these differences to light. At that point, the conditions of the predictive and calibration periods and the data used in the calibration can all be reviewed to determine whether the model probabilities should be re-evaluated.

##### 4.11.1 Predictive Simulation Description

The predictive period for the 300 Area application was from January 2005 to January 2025. The simulation domain and model discretization were unchanged for the predictive simulations. Groundwater heads and uranium concentrations from the end of the calibration period were used as the initial conditions for the pre-

dictive modeling. Boundary conditions used in the predictive simulations were unchanged except for the head boundary representing the Columbia River stage. For the baseline scenario, the river stage simulated by Waichler et al. (2005) for the period 1975 to 1994 was used in the predictive period (see Figure 4-19). For the alternative scenario, the river stage simulated by Waichler et al. (2005) for the period 1955 to 1974 was used. In both cases, monthly average river stage was used with one-month stress periods in the flow models.

#### 4.11.1.1 Modeled Parameter Uncertainty

The calibration results were used to simulate parameter uncertainty in the predictive period by assuming that the calibrated parameters (or the log of the parameters in the case of the hydraulic conductivity and distribution coefficient parameters) followed a multivariate Gaussian distribution with mean values given by the optimal parameter values and a covariance given by the maximum likelihood estimated parameter covariance (Eq. 43).

The Latin hypercube sampling code of Iman and Shortencarier (1984) was used to generate 200 realizations of the calibrated parameters for each of the four model alternatives. The resulting parameter distributions and correlations accurately reflected the input distributions determined from the calibration results. Sample distributions are shown in Figure 4-32 for  $K_h$  and  $\alpha_L$  of Model 4.

As mentioned in Section 2.3.1, characteristics of the scenarios could have been modeled as random variables. Specifically, the future river stage is clearly uncertain – modeling the river stage as a temporally correlated stochastic process would have been justified. The simulated results of Waichler et al. (2005) could be used to derive appropriate statistics of the stochastic process for the two scenarios. Each Monte Carlo realization would then include not only the randomly sampled parameters described above, but also a randomly sampled river stage time series. Scenario uncertainty would thus include not only the effect of the two scenarios (baseline and the alternative), but also the random variation in the river stage time series.

#### 4.11.2 Baseline Scenario

The parameter realizations discussed in the previous section were used in a Monte Carlo simulation to compute flow and uranium transport over the predictive period. The resulting head and concentration distributions included the impact of parameter uncertainty. A Monte Carlo simulation was carried out for each of the

four model alternatives to incorporate the impact of model uncertainty.

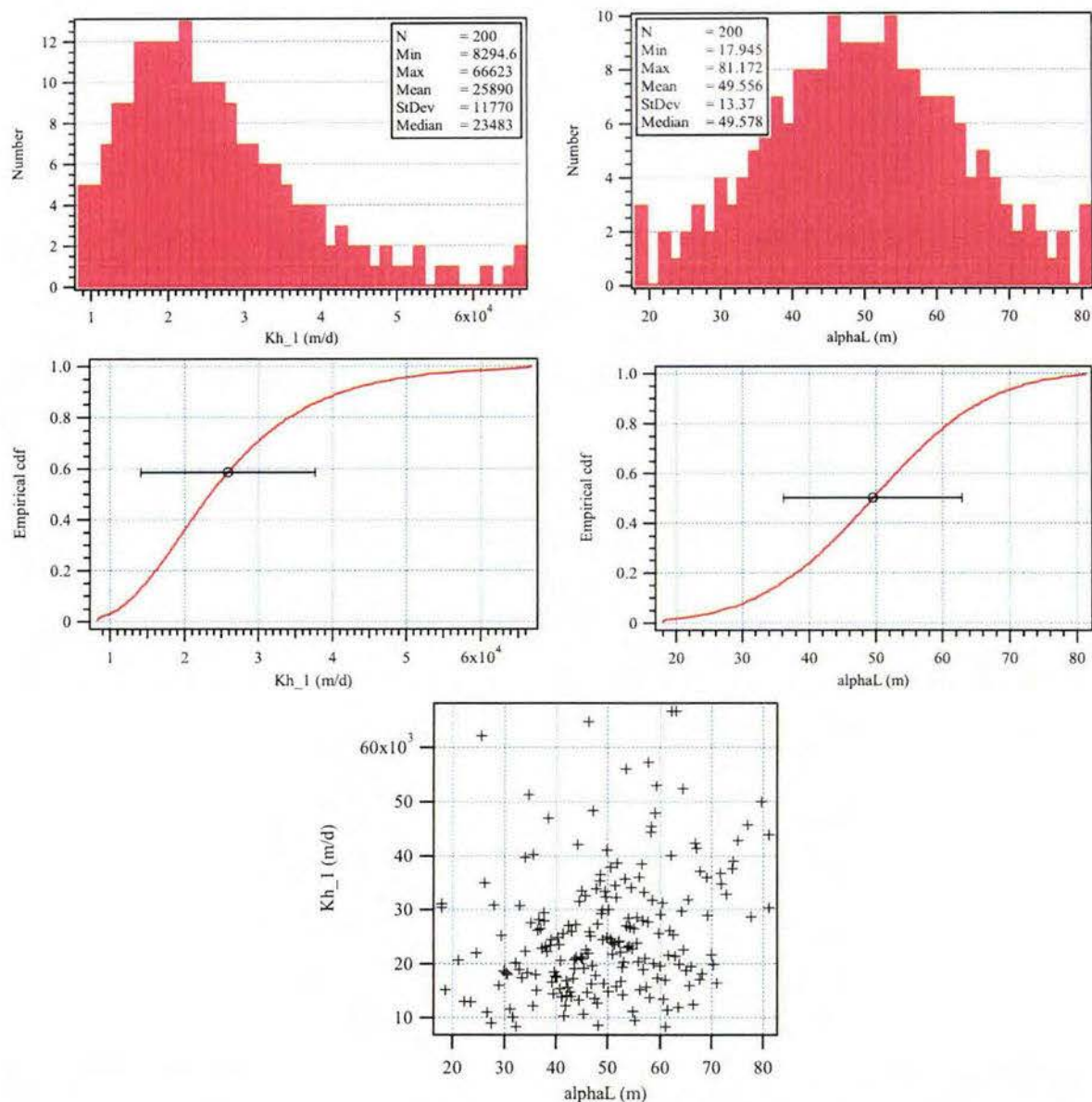
There are various types of prediction results that may be of interest. For example, the mean and variance of concentration can be viewed as a function of time to examine the evolution of the plume over the predictive period and where the greatest concentration uncertainty exists (that portion due to parameter uncertainty). Figure 4-33 shows the concentration mean and variance in January 2010. This is a view looking northwest with the river grid cells removed. The variance is generally largest along the river and in the areas of high mean concentration.

It may also be of interest to examine the predicted output at a specific location. Figure 4-34 shows the uranium concentration at well 399-1-1 near the river as predicted by Model 4. Concentrations computed for all 200 realizations are shown as a function of time along with the average concentration. The concentration variance grows over time at this location. In addition, the concentration at specific times appears to be somewhat skewed.

For regulatory or management applications, the predicted value of interest is likely to be a single quantity that can be derived from the spatial and temporal distributions of concentration. Examples include the maximum concentration along a compliance boundary during the predictive period, the length of time for the maximum concentration to fall below a limit, and the peak dose over the predictive period as computed from predicted concentrations and a specified exposure scenario. A probability distribution for each of these quantities could be estimated from the Monte Carlo simulation results. These distributions, one for each alternative model, are the  $p(\Delta | M_k, \mathbf{D})$  of Eq. (1). To illustrate such results for the 300 Area application, the distributions of predicted uranium concentration at well 399-1-1 on 1/1/2025 are shown in Figure 4-35 for each of the four model alternatives and the model average using the posterior model probabilities from Table 4-12. Empirical probability density functions based on the 200 realizations are shown in the upper plot while the lower plot shows the empirical cumulative distribution functions. As determined by the posterior model probabilities, the model-average distribution closely follows the distribution from Model 4.

The distributions of concentration for Models 2 and 6 were nearly identical at this time and location - the heterogeneity of  $K_d$  (Model 6) had little impact on the predicted concentration when the hydraulic conductivity was homogenous throughout the domain. This was not unexpected as the calibrated values of  $K_d$  and





**Figure 4-32. Sample histograms, cumulative distribution functions (mean  $\pm$  one standard deviation shown), and scatterplot for Model 4 parameters  $Kh_1$  and  $\alpha L$  resulting from 200 Latin hypercube realizations (correlation = 0.27)**

$K_d$  for Model 6 were nearly the same (Table 4-10) with a correlation coefficient of 0.86. In addition, the calibrated value of the homogeneous  $K_d$  for Model 2 was similar to the calibrated values of  $K_d$  and  $K_d$  for Model 6. Models 4 and 8 with heterogeneous hydraulic conductivity had much larger mean uranium concentrations at well 399-1-1 on 1/1/2025 than the homogeneous

models. The variance of predicted concentration was also much larger for the heterogeneous models. The heterogeneity of  $K_d$  in Model 8 had a noticeable effect on the predicted uranium concentration distribution, producing a smaller mean and somewhat smaller variance than Model 4 with homogeneous  $K_d$ .

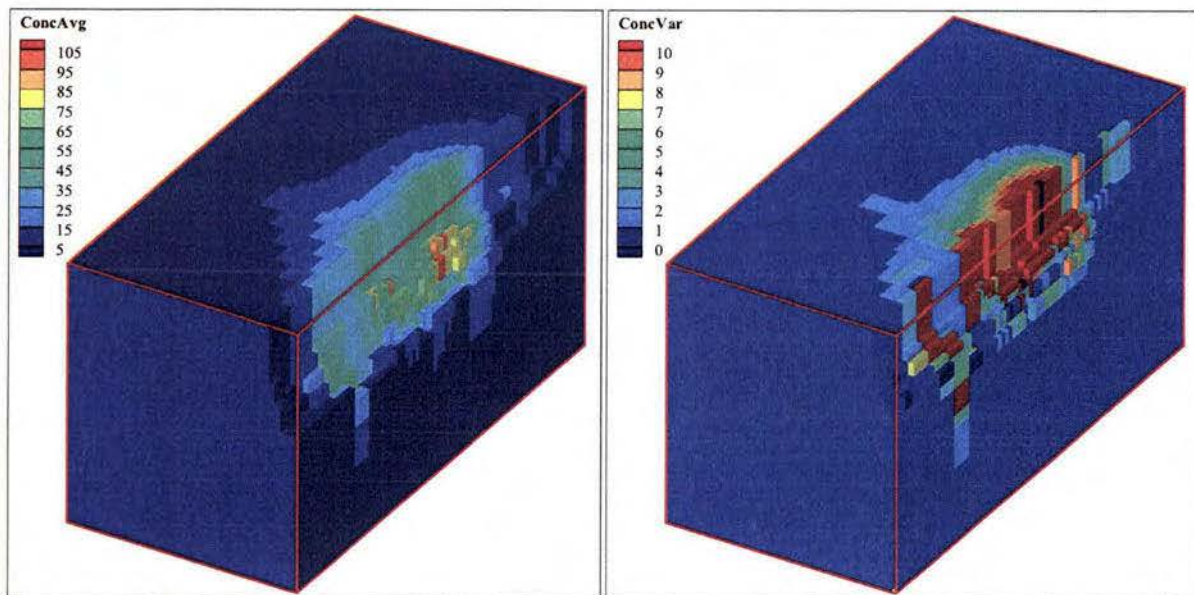


Figure 4-33. Monte Carlo simulation results for Model 4 in January 2010: (left) concentration mean ( $\mu\text{g/l}$ ) and (right) concentration variance

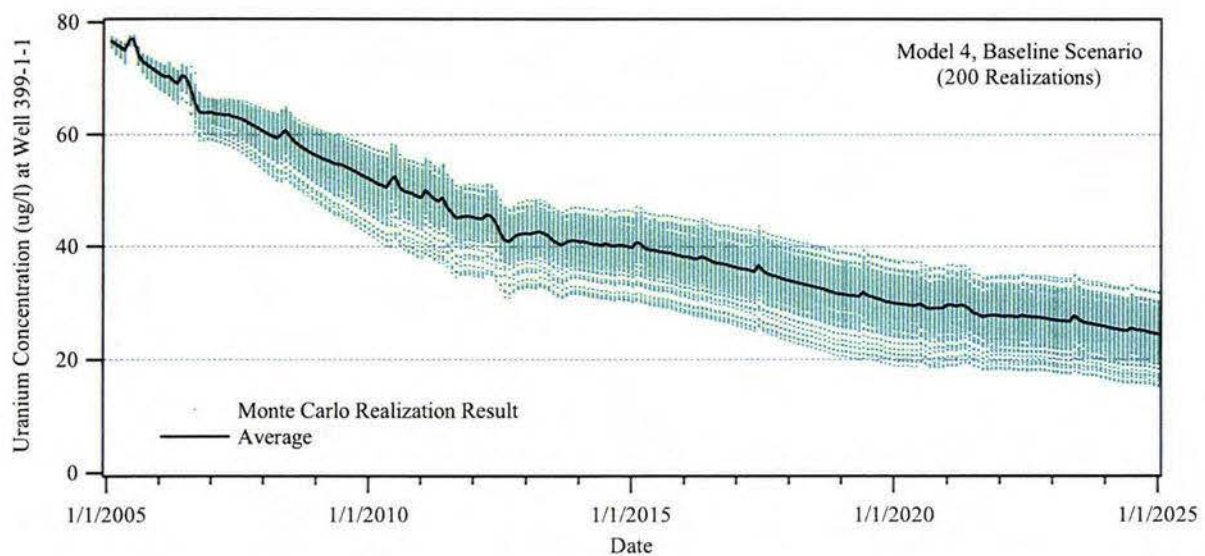


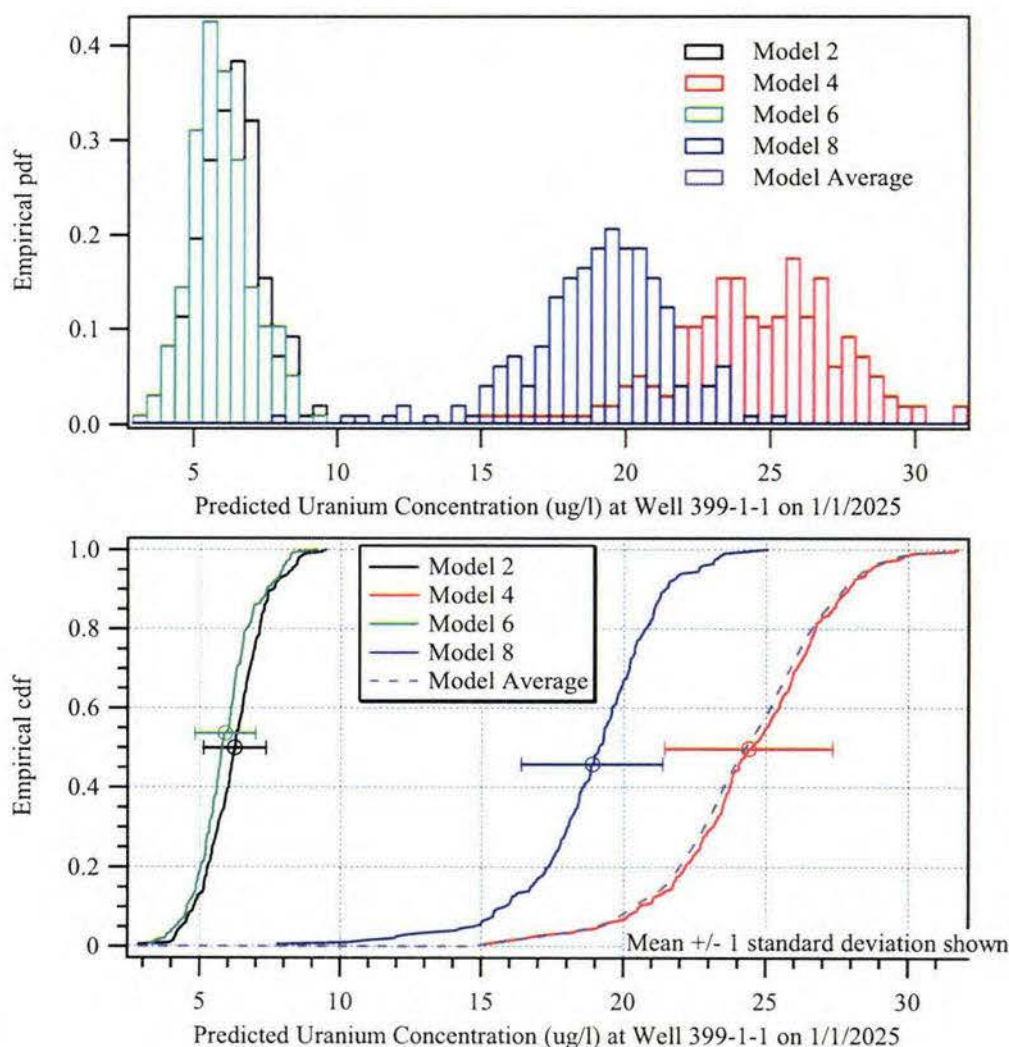
Figure 4-34. Monte Carlo simulation results for Model 4 under the baseline scenario: uranium concentration ( $\mu\text{g/l}$ ) at well 399-1-1 for 200 realizations and average concentration



The model-averaged concentration mean and variance can be computed from the individual model results using Eqs. (18) and (19) or (20). Because of the dominant probability for Model 4 in this application, the model-averaged mean and variance were essentially the same as the mean and variance of Model 4. The model-averaged distribution shown in Figure 4-35 can be used to estimate risk-related quantities. For example, the probability that the uranium concentration at well 399-1-1 exceeds 30  $\mu\text{g/l}$  on 1/1/2025 is estimated to be about 1%.

### 4.11.3 Alternative Scenario

As discussed previously, the alternative scenario considered in this application assumed variability in Columbia River stage at the 300 Area reflective of conditions prior to 1975. This was modeled by simulating uranium transport from January 2005 through December 2024 using a head boundary condition along the river equal to the estimated river stage from 1955 to 1974. Compared to the baseline scenario, this boundary condition was more variable with significantly higher late spring/early summer peaks (Figure 4-19). For the alternative scenario, the random realizations of parameter values used in the Monte Carlo simulation were



**Figure 4-35. Probability distributions for the predicted uranium concentration at well 399-1-1 on 1/1/2025: (top) empirical probability density functions for the four model alternatives and the model average and (bottom) empirical cumulative distribution functions**



identical to the values used in the baseline scenario. 200 realizations of groundwater flow and uranium transport were simulated.

Uranium concentration at well 399-1-1 over the entire predictive period is shown for all 200 realizations in Figure 4-36. These results can be compared to the baseline scenario results shown in Figure 4-34. For the alternative scenario, the mean concentration is much more variable and falls from its peak value more quickly, particularly in the first few years of the simulation. The concentration variance appears to grow more rapidly in the early part of the simulation and appears to be greater throughout the predictive period than under the baseline scenario.

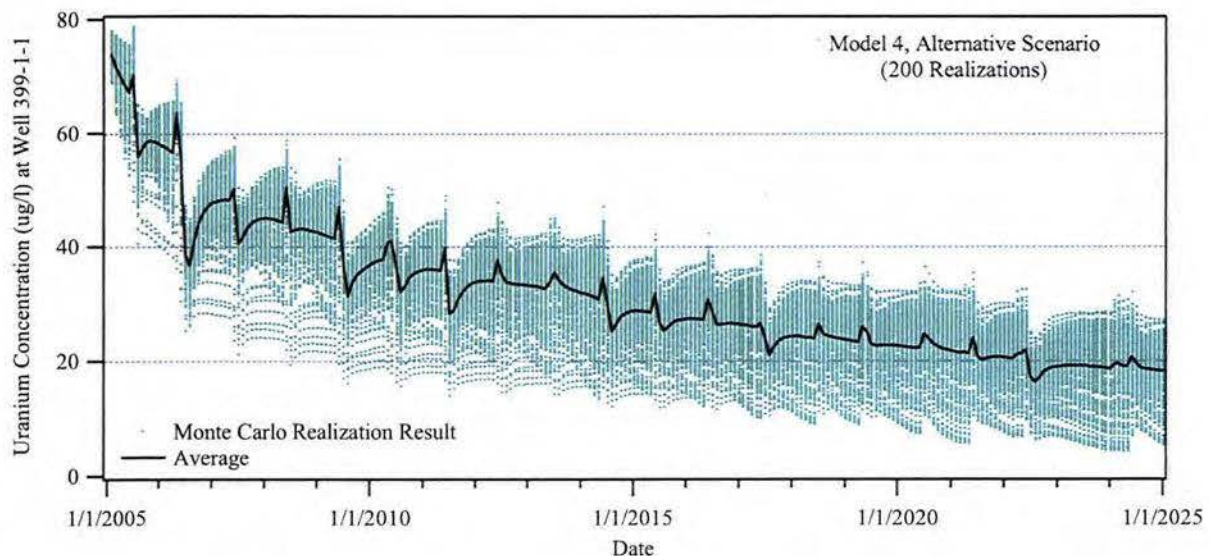
Model-averaged results could be computed for the alternative scenario in a manner identical to that discussed above for the baseline scenario and shown in Figure 4-35 for the predicted concentration at well 399-1-1 on 1/1/2025. Corollary results for the alternative scenario are not shown here because the model-averaged results are nearly identical to the results from Model 4 (discussed in the following section) due to the dominant probability of Model 4.

#### 4.11.3.1 Scenario-Averaged Prediction

Probability distributions for predicted quantities that include parameter, model, and scenario uncertainty can

be computed using Eq. (21). The model-averaged probability distributions for each scenario were averaged over the alternative scenarios using the scenario probabilities as weights. For this application, the baseline scenario had a probability of 0.7 and the alternative scenario had a probability of 0.3. Model-averaged probability density functions for the uranium concentration at well 399-1-1 on 1/1/2025 are shown in the top plot of Figure 4-37 for the two scenarios. The scenario-averaged density function was computed as the weighted average of these two densities. The corresponding cumulative distribution functions are shown in the lower plot of Figure 4-37. The mean concentration is smaller for the alternative scenario while the concentration variance is larger. The distribution for the alternative scenario also appears to be more highly skewed.

Risk-related quantities can be estimated directly from the scenario-averaged distribution. For example, the probability that the uranium concentration at well 399-1-1 exceeds 25  $\mu\text{g/l}$  on 1/1/2025 is estimated to be about 30 percent. This estimate includes the impact of parameter uncertainty, conceptual model uncertainty, and scenario uncertainty. At the same time, the contribution to the total uncertainty from each of these components is made clear in the figures (Figure 4-35 and Figure 4-37). These contributions can also be computed using the equations presented in Section 3. The mean and variance (Eqs. 23 and 24) of the predicted



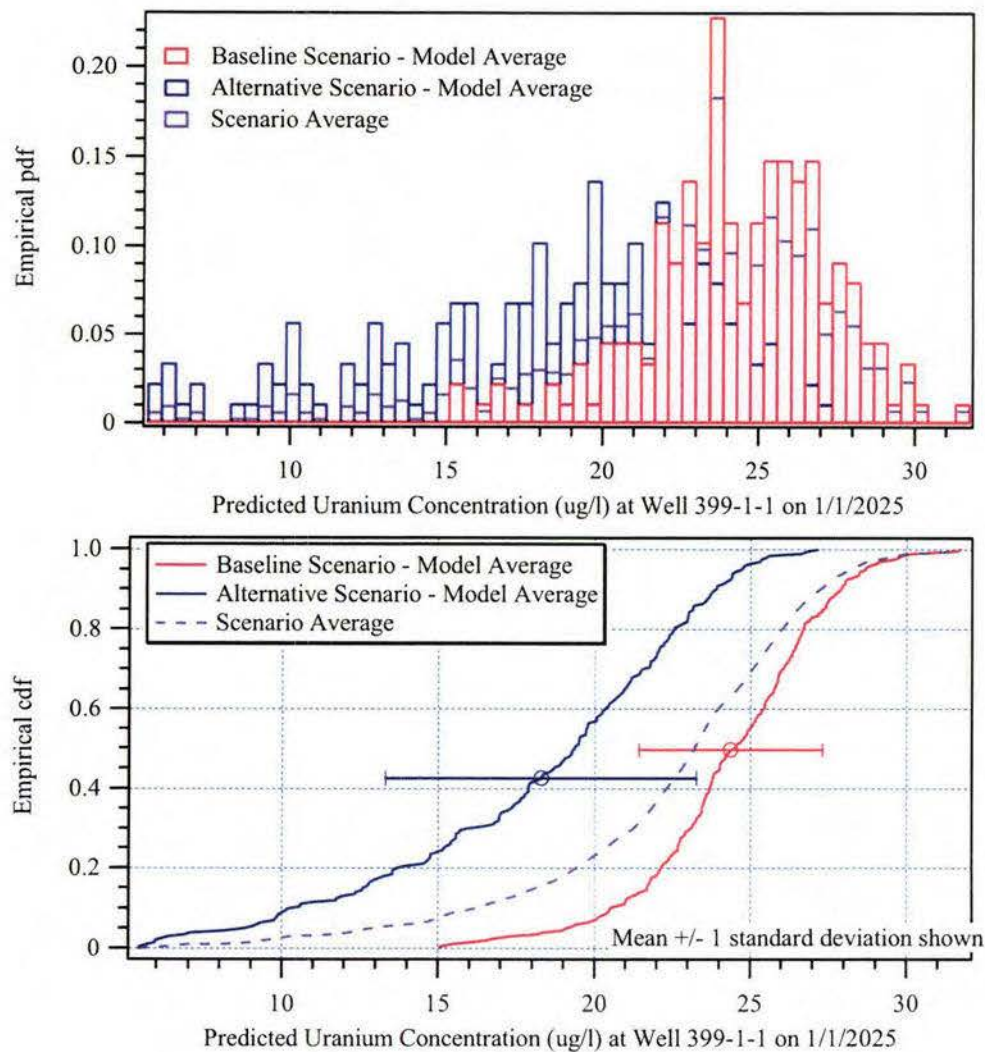
**Figure 4-36. Monte Carlo simulation results for Model 4 under the alternative scenario: uranium concentration ( $\mu\text{g/l}$ ) at well 399-1-1 for 200 realizations and average concentration**

concentration at well 399-1-1 on 1/1/2025 are given in Table 4-13. The variance is broken down into the within-scenario and between-scenario components as computed from Eq. (25).

The impact of changes in the model or scenario probabilities can be readily computed and viewed graphically. In addition, differences in the distributions of predicted values due to the model or scenario alternatives are readily apparent. Thus the impact on risk-related quantities can be easily seen. For example, the probability that the uranium concentration at well 399-1-1 exceeds 25 µg/l on 1/1/2025 is estimated to be about 45 percent under the baseline scenario and just 5 percent under the alternative scenario. Such information may be valuable in justifying the use of specific scenarios and supporting decisions based in part on the simulation results.

**Table 4-13. Mean, variance, and standard deviation of individual scenarios and the scenario average for the predicted concentration at well 399-1-1 on 1/1/2025**

	Baseline Scenario	Alternative Scenario	Scenario Average
Probability	0.7	0.3	
Mean	24.40	18.30	22.57
Variance	8.70	24.77	21.34
Std Dev	2.95	4.98	4.62
Within-Scenario Variance			13.52
Between-Scenario Variance			7.82



**Figure 4-37. Probability distributions for the predicted uranium concentration at well 399-1-1 on 1/1/2025: (top) empirical probability density functions for Model 4 under the two scenarios and the scenario average and (bottom) empirical cumulative distribution functions**



## 5 CONCLUSIONS

This report has described a methodology to systematically and quantitatively assess predictive uncertainty in groundwater flow and transport modeling with consideration of the combined impact of hydrogeologic uncertainties associated with the conceptual-mathematical basis of a model, model parameters, and the scenarios to which the model is applied. The application of this methodology to realistic, field-scale groundwater flow and transport modeling has been demonstrated using uranium contamination at the 300 Area of the U.S. Department of Energy Hanford Site as an example.

A review of published results from hydrogeologic model post-audits and other model applications demonstrated that the primary errors in groundwater modeling predictions can often not be assigned to errors in parameter values. In 9 of the 16 applications reviewed, conceptual model errors were most significant. Model scenario errors were the most significant in 4 of the 16 applications. Parameter errors were most significant in three of the applications. This (limited) review suggests that a comprehensive approach to uncertainty assessment in hydrogeologic modeling should not be limited to parameter uncertainties, but must also consider the potential for significant conceptual model and scenario uncertainties if a realistic estimate of predictive uncertainty is desired.

In a quantitative uncertainty analysis, parameter uncertainty is typically characterized using continuous probability distributions. When characterizing conceptual uncertainty in hydrogeologic modeling, specifying a continuum of conceptual model possibilities is likely to be infeasible. Instead, it is generally more appropriate to postulate a discrete set of alternative conceptual models. Given a set of alternatives, one might suggest that model uncertainty be resolved by simply comparing predictions from the alternative models. This approach has a serious drawback, however. Without a quantitative measure of the degree of plausibility of model alternatives, it is impossible to determine the risk of a decision based on the model predictions. Similarly, justifying the adoption of the most conservative model for use in prediction will be easier if the plausibility of the conservative model can be compared to the alternatives.

Maximum Likelihood Bayesian Model Averaging was used here to estimate model probabilities, which are a measure of the relative degree of plausibility of a set of alternative models. MLBMA is an optimal and computationally feasible way of combining the predictions of several competing models and assessing their joint

predictive uncertainty. It is a general method, applying to deterministic and stochastic models, and to complex and simplified models.

The model-averaging approach has been criticized on the basis that there is only one model corresponding to the physical reality and therefore an average over several models has no physical interpretation. Although the models considered here are physically-based, the inherent complexity of the hydrogeologic environment means that all models considered will be an approximation to physical reality. When appropriately formulated, each model alternative will have some merit in reproducing aspects of the physical system, this merit being quantified by each model's probability. Since we interpret model probability as a relative measure with respect to the other model alternatives considered, a model-averaged consequence has an intuitive and consistent meaning. Note that in an application of MLBMA the predicted consequences of the physics as rendered by the models are averaged, not the underlying physics.

Model averaging has also been criticized for masking information essential to the decision maker. It is demonstrated in this report, both in the abstract and in the concrete application to uranium transport at the 300 Area, that the MLBMA results can be presented so as to fully inform a decision, clearly illuminating the effect of model uncertainty and the importance of a consistent and defensible means to estimate model probabilities. Model averaging reduces the risk of relying on an overly conservative model and provides a consistent and quantitative way to address model uncertainty in the context of a regulatory decision.

In the framework adopted in this report, estimated parameter values and model probabilities are based on information and data available for the model calibration, referred to here as the history-matching period. That is, parameter values and model probabilities are estimated as part of the process of model development and evaluation. The characterizations of parameter and model uncertainty resulting from this process can be projected into the period over which model predictions are needed, referred to here as the predictive period. As defined here, a scenario is a description of the future conditions under which a model is applied, conditions that are inherently uncertain and apply only in the predictive period. A comprehensive assessment of predictive uncertainty thus requires that scenario uncertainty be explicitly considered. However, in the limited definition used here, a scenario can be reduced to a set of

conditions described by the three elements of any simulation model: geometry and parameterization, structure of governing equations, and forcing. Since we presently know, albeit imperfectly, how to deal with parameter uncertainty, uncertainty in the model structure, and uncertainty in the forcing terms, we know mathematically how to deal with scenario uncertainty. Most of the time a scenario will impact mainly forcing terms, which are often easier to deal with.

MLBMA was extended to include scenario uncertainty in a joint estimation of predictive uncertainty. This requires postulating a set of alternative scenarios and the probability of their occurrence. As with model probabilities, the scenario probabilities are interpreted as relative measures of the degree of plausibility of the scenarios. Unlike model probabilities, at the time of prediction there is no opportunity to condition the scenario probabilities on observation. Scenario probabilities are thus prior probabilities and are appropriately interpreted as subjective measures based on current knowledge, experience, and judgment. The MLBMA methodology requires that alternative scenarios be mutually exclusive, a condition that can be enforced by defining the scenarios as possible combinations of alternative events that need not be mutually exclusive. Scenario probabilities can be determined from estimates of the marginal and conditional probabilities of these events.

In deriving MLBMA with scenario averaging, the assumption was made that the dataset used for model calibration is independent of the scenario. That is, the occurrence of any particular scenario in the future does not affect the probability of observing the data in the past. As a result, the model calibrations are not a function of the scenario and do not need to be recomputed under each scenario. This significantly reduces the computational requirements of the methodology. To incorporate scenario uncertainty, predictive simulations are conducted with the alternative models under each alternative scenario. While this is computationally expensive, it is straightforward and would be carried out in any case unless scenario uncertainty is completely ignored. Although not considered in the application presented here, it is possible to include additional parameter uncertainty specific to a scenario (i.e., not present in the history-matching period) in the predictive simulations.

A previous application of MLBMA illustrated the superiority of using the model-averaged result from MLBMA over any individual model. The application to the Hanford Site 300 Area uranium plume advances the uncertainty assessment methodology by including the impact of hydrologic scenario uncertainty and by

applying the methods to a more complex problem and relevant system: transient groundwater flow and transport of a radionuclide from an uncontrolled disposal. The application should be viewed as an example, however, since significant uncertainties were ignored (primarily uncertainty in uranium sources) and models were simplified in critical ways (e.g., the geochemistry, saturated flow only). The application illustrates the very real tradeoffs that must be made when modeling complex systems between incorporating complexity in the model and the computational demands of model calibration and uncertainty assessment. Application of the methodology to a similarly-scaled, three-dimensional, variably-saturated model incorporating reactive geochemistry is currently likely to be computationally infeasible. In spite of these limitations, the application illustrates all aspects of the methodology and can serve as a template for future applications.

As described in the report, there exists at the 300 Area a significant amount of data regarding aquifer heads, uranium concentrations, and site geochemistry. Despite this, there are significant uncertainties that remain, including the hydraulic properties of the unsaturated zone, the extent and magnitude of the current uranium contamination in the unsaturated zone, and the historical flux of uranium to groundwater. As is the case at the 300 Area, the historical data is not always optimal for use in model calibration. Where models are to be used in site management, data collection needs to be targeted for use in model development and evaluation, including calibration and uncertainty assessment.

Data continue to be collected at the 300 Area site and the regulatory models continue to be refined. Four large-diameter boreholes were completed recently with continuous split-spoon samples collected to better define the distribution of uranium. Physical and geochemical properties have been measured on more than 100 of these samples. Slug tests and pump tests were performed on the completed wells to provide hydraulic parameter estimates. In addition, a large-scale tracer test was recently conducted using one of the wells as the injection well. These recent data were not used in the application presented here. Due to the iterative nature of the methodology, however, these data could be incorporated in an updated analysis in a very straightforward manner.

For this application, prior parameter information was valuable in producing reasonable estimates of the uranium adsorption coefficients. Without the prior information, adsorption coefficients were unreasonably large. Since the prior information exists, including it in the calibration was justified. The poor calibration results obtained without the prior information suggest,

however, that the geochemical conceptualization of the models needs to be improved. This is also suggested by the failure of the calibrations to produce the expected relationship between adsorption coefficients in the two zones used in Models 6 and 8, and by the inability to reproduce large observed concentrations.

Selection of weights for the calibration data had a significant effect on the calibration results. Actual laboratory measurement error for the uranium concentration data is likely to be proportional to the concentration. Using weights based on estimated laboratory measurement errors produced inferior calibrations to those obtained using equal weights for the concentration observations. The use of equal weights implies that a concentration-independent sampling error dominates the uranium concentration measurement error.

Model probabilities are very sensitive to the magnitudes of the *KIC* differences. Thus changes in model assumptions and inputs, such as fixed parameter values and uranium sources, could alter the model probabilities. Changes in calibration inputs, such as the selection of data, may have the same effect. The sensitivity of model probability to factors influencing *KIC* differences (as well as other model discrimination criteria such as *AIC*) warrants further investigation.

Although the methodology is formulated using maximum likelihood parameter estimation, results from a least squares parameter estimation can be used as well. The 300 Area application used the least squares code PEST to complete the model calibrations. Computation of the model probabilities from the PEST results were completed in a spreadsheet.

For the 300 Area application, one model dominated the other alternatives. This result implies that only the dominant model needed to be used in predictive simulations. For illustrative purposes, however, all four models were used. In some cases, it may be reasonable to apply low-probability models to the prediction. If the conditions of the predictive period are significantly different than the calibration period, differences between models that were not apparent in the calibration

may appear during prediction. Applying even low-probability models in the predictive period may bring these differences to light. At that point, the conditions of the predictive and calibration periods and the data used in the calibration can all be reviewed to determine whether the model probabilities should be reevaluated.

Predictive simulations were carried out for the 300 Area application over the period 2005-2025 using all four models and under two alternative scenarios. Monte Carlo simulation was used to simulate parameter uncertainty. Selected uranium concentration results were shown to illustrate the model and scenario averaging. These figures and the accompanying tables illustrate the individual contributions to predictive uncertainty of the model, parameter, and scenario uncertainties. In this case, model uncertainty was negligible due to the dominance of a single model. Parameter and scenario uncertainties were significant.

In a regulatory application of the methodology, probability distributions of concentration would also likely be computed. The regulatory quantity of interest may be a quantity derived from predicted concentration, however, such as peak dose over a given time period or a mortality risk to a specific population. In these cases, the execution of the methodology would be identical except that the desired regulatory quantity would be computed for each Monte Carlo realization and its distribution would be averaged over the alternative models and scenarios. Figures illustrating the results would be for the regulatory quantity instead of for contaminant concentration.

Application of the uncertainty methodology requires the calibration and probabilistic simulation of multiple models and is therefore computationally demanding. For the 300 Area application, however, computations were carried out on inexpensive desktop computers. In addition, the application was completed using primarily public-domain software tools. The 300 Area application thus demonstrates the practicability of applying a comprehensive uncertainty assessment to large-scale, detailed groundwater flow and transport modeling.





## 6 REFERENCES

- Abramson, L.R., "Model Uncertainty from a Regulatory Point of View," in *Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis, Model Uncertainty: Its Characterization and Quantification*, NUREG/CP-0138, pp. 119-123, U.S. Nuclear Regulatory Commission, Washington, DC, 1994.
- Akaike, H., A new look at statistical model identification, *IEEE Trans. Automat. Contr.*, AC-19, 716-722, 1974.
- Alley W.M. and P.A. Emery, Groundwater model of the Blue River basin, Nebraska—twenty years later, *J. Hydrol.*, 85:225-249, 1986.
- Andersen, P.F. and S. Lu, A post audit of a model-designed ground water extraction system, *Ground Water*, 41(2):212-218, 2003.
- Ando, K., Kostner, A., and S.P. Neuman, Stochastic continuum modeling of flow and transport in a crystalline rock mass: Fanay-Augères, France, revisited, *Hydrogeology Journal*, 11(5), 521-535, 2003.
- Apostolakis, G., The concept of probability in safety assessments of technological systems, *Science*, 250:1359, 1990.
- Belcher, W.R. (ed), *Death Valley Regional Ground-Water Flow System, Nevada and California – Hydrogeologic Framework and Transient Ground-Water Flow Model*, U.S. Geological Survey Scientific Investigation Report 2004-5205, 2004.
- Bethke, C.M., The Geochemist's Workbench® Release 6.0 GWB Reaction Modeling Guide; RockWare, Inc., Golden, CO, 2005.
- Beven K.J. and J. Freer, Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrology*, 249:11-29, 2001.
- Bjornstad, B., Sampling and Hydrogeology of the Vadose Zone Beneath the 300 Area Process Ponds (unpublished report), 2003.
- Borgonovo, E., G.E. Apostolakis, S. Tarantola and A. Saltelli, Comparison of global sensitivity analysis techniques and importance measures in PSA, *Reliability Engineering and System Safety*, 79:175-185, 2003.
- Brauers, J. and M. Weber, A new method of scenario analysis for strategic planning, *J. Forecasting*, 7(1):31-47, 1988.
- Bredehoeft, J., The conceptualization model problem – surprise, *Hydrogeology J.*, 13:37-46, 2005.
- Bryce R.W., C.T. Kincaid, P.W. Eslinger, and L.F. Morasch (eds), *An Initial Assessment of Hanford Impact Performed with the System Assessment Capability*, PNNL-14027, Pacific Northwest National Laboratory, Richland, Washington, 2003.
- Burnham, K.P. and A.R. Anderson, *Model selection and multiple model inference: a practical information-theoretical approach*, 2nd edition, New York, Springer, 2002.
- Campbell, M.D., W.J. McMahon, and K.R. Simpson, *Water Level Measurements for Modeling Hydraulic Properties in the 300-FF-5 and 100 Aggregate Area Operable Units*, PNL-8580, Pacific Northwest Laboratory, Richland, Washington, 1993.
- Cantrell, K.J., R.J. Serne, and G.V. Last, *Hanford Contaminant Distribution Coefficient Database and Users Guide*, PNNL-13895, Rev. 1, Pacific Northwest National Laboratory, Richland, Washington, 2003.
- Carrera J. and S.P. Neuman, Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, *Water Resour. Res.*, 22(2):199-210, 1986a.
- Carrera, J. and S.P. Neuman, Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data. *Water Resour. Res.* 22(2), 228-242, 1986b.
- Carrera, J., A. Medina, C. Axness, and T. Zimmerman, Formulations and computational issues of the inversion of random fields. In: *Subsurface Flow and Transport: A Stochastic Approach* (ed. by G. Dagan and S. P. Neuman), 62-79. Cambridge University Press, Cambridge, United Kingdom, 1997.
- Cole C.R., M.P. Bergeron, S.K. Wurster, P.D. Thorne, S. Orr, and M.I. McKinley, *Transient Inverse Calibration of Hanford Site-Wide Groundwater Model to Hanford Operational Impacts—1943 to 1996*, PNNL-13447, Pacific Northwest National Laboratory, Richland, Washington, 2001a.

- Cole, C.R., M.P. Bergeron, C.J. Murray, P.D. Thorne, S.K. Wurstner, and P.M. Rogers, *Uncertainty Analysis Framework – Hanford Site-Wide Groundwater Flow and Transport Model*, PNNL-13641, Pacific Northwest National Laboratory, Richland, Washington, 2001b.
- Dagan G. and S.P. Neuman (eds.), *Subsurface Flow and Transport: A Stochastic Approach*, Cambridge University Press, Cambridge, United Kingdom, 1997.
- Davis J.A., D.E. Meece, M. Kohler, and G.P. Curtis, Approaches to surface complexation modeling of uranium(VI) adsorption on aquifer sediments, *Geochim. Cosmochim. Acta*, 68, 3621-3641, 2004.
- Davis J.A., and G.P. Curtis, *Application of Surface Complexation Modeling to Describe Uranium(VI) Adsorption and Retardation at the Uranium Mill Tailings Site at Naurita, Colorado*, NUREG/CR-6820, U.S. Nuclear Regulatory Commission, Rockville, Maryland, 2003.
- Davis, J.A., D.L. Bond, and P. Fox, "Labile U(VI) and Progress Toward a Generalized Surface Complexation Model," presented at a workshop on Conceptual Model Development and Reactive Transport Modeling for the 300 Area Uranium Plume in 300-FF-5, held May 10-11, 2004, Richland, Washington, 2004, <http://www.hanford.gov/cp/gpp/public/workshops.cfm>.
- De Kluyver, C.A. and H. Moskowitz, Assessing scenario probabilities via interactive goal programming, *Management Science*, 30(3):273-278, 1984.
- DOE, *Phase 1 Remedial Investigation Report for the 300-FF-5 Operable Unit*, DOE/RL-93-21, Rev. 0, Volumes 1 and 2, U.S. Dept. of Energy, Richland, Washington, 1994.
- DOE, *Remedial Investigation/Feasibility Study Report for the 300-FF-5 Operable Unit*, DOE/RL-94-85, Rev. 0, U.S. Dept. of Energy, Richland, Washington, 1995.
- Doherty, J., *PEST*, Model-Independent Parameter Estimation, User's Manual: 5th Edition, Watermark Numerical Computing, Australia, 2004. <http://www.sspa.com/pest/>
- Doherty, J., Ground water model calibration using pilot points and regularization, *Ground Water*, 41(2):170-177, 2003.
- Draper, D., Assessment and propagation of model uncertainty, *J. Roy. Statist. Soc. Ser. B*, 57(1):45-97, 1995.
- Draper, D., A. Pereira, P. Prado, A. Saltelli, R. Cheal, S. Eguilior, B. Mendes, S. Tarantola, Scenario and parametric uncertainty in GESAMAC: A methodological study in nuclear waste disposal risk assessment, *Computational Physics Communications*, 117, 142-155, 1999.
- EPA, *Guiding Principles for Monte Carlo Analysis*, EPA /630/R-97/001, U.S. Environmental Protection Agency, Washington, DC, 1997.
- EPA, *Understanding Variation in Partition Coefficient,  $K_d$  Values, Volume II: Review of Geochemistry and Available  $K_d$  Values for Cadmium, Cesium, Chromium, Lead, Plutonium, Radon, Strontium, Thorium, Tritium ( $^3\text{H}$ ), and Uranium*, EPA 402-R-99-004B, U.S. Environmental Protection Agency, Washington DC, 1999. <http://www.epa.gov/radiation/technology/partition.htm>
- Evans, R.G., M.J. Hattendorf, and C.T. Kincaid, *Evaluation of the Potential for Agricultural Development at the Hanford Site*, PNNL-13125, Pacific Northwest National Laboratory, Richland, Washington, 2000.
- Fayer, M.J. and T.B. Walters, *Estimated Recharge Rates at the Hanford Site*, PNL-10285, Pacific Northwest Laboratory, Richland, Washington, 1995.
- Flavelle P., S. Nguyen S, and W. Napier, Lessons learned from model validation - a regulatory perspective, in *GEOVAL-1990: Symposium on Validation of Geosphere Flow and Transport Models*, Organization for Economic Co-operation and Development (OECD), Nuclear Energy Agency, Paris, France, pp. 441-448, 1991.
- Gee G.W. and J.W. Bauder, Particle-size analysis, pp. 383-423 In A Klute (ed.) *Methods of soil analysis*, Part 1, 2nd ed., Agron. Monogr. 9, Am. Soc. Agron., and Soil Sci. Soc. Am., Madison, Wisconsin, 1986.
- Gomez-Hernandez, J.J., A. Sahuquillo, and J.E. Capilla, Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, I. Theory, *J. Hydrology*, 203:162-174, 1997.
- Harbaugh, A.W., E.R. Banta, M.C. Hill, and M.G. McDonald, *MODFLOW-2000, the U.S. Geological Survey modular ground-water model -- User guide to modularization concepts and the Ground-Water Flow Process*, U.S. Geological Survey Open-File Report 00-92, 121 pp., 2000. <http://water.usgs.gov/nrp/gwsoftware/modflow2000/modflow2000.html>



- Hartman, M.J., L.F. Morasch, and W.D. Webber (eds.), *Hanford Site Groundwater Monitoring for Fiscal Year 2003*, PNNL-14548, Pacific Northwest National Laboratory, Richland, Washington, 2004.
- HEIS, *Hanford Environmental Information System*, Environmental Information Systems Department, Fluor Hanford, Inc., Richland, Washington, 1994.
- Helton, J.C., Uncertainty and sensitivity techniques for use in performance assessment for radioactive waste disposal, *Reliability Engineering and System Safety*, 42:327-367, 1993.
- Helton, J.C., J.D. Johnson, C.J. Sallaberry, and C.B. Storlie, Survey of sampling-based methods for uncertainty and sensitivity analysis, *Reliability Engineering and System Safety*, 91:1175-1209, 2006.
- Hernandez, A.F., S.P. Neuman, A. Guadagnini, and J. Carrera-Ramirez, Conditioning steady state mean stochastic flow equations on head and hydraulic conductivity measurements, 158-162, *Proc. 4<sup>th</sup> Intern. Conf. on Calibration and Reliability in Groundwater Modeling (ModelCARE 2002)*, edited by K. Kovar and Z. Hrkal, Charles University, Prague, Czech Republic, 2002.
- Hernandez, A. F., S. P. Neuman, A. Guadagnini, and J. Carrera, Conditioning mean steady state flow on hydraulic head and conductivity through geostatistical inversion, *Stochastic Environmental Research and Risk Assessment*, 17, DOI10.1007/s00477-003-0154-4, 2003.
- Hernandez, A.F., S.P. Neuman, A. Guadagnini, and J. Carrera, Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media, *Water Resour. Res.*, 42, W05425, doi:10.1029/2005WR004449, 2006.
- Hill, M.C., *Methods and Guidelines for Effective Model Calibration*, U.S. Geological Survey Water-Resources Investigations Report 98-4005, U.S. Geological Survey, Denver, Colorado, 1998.
- Hill, M.C. and C.R. Tiedeman, *Effective Groundwater Model Calibration*, Wiley, New York, 455 pp., 2007.
- Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky, Bayesian model averaging: A tutorial, *Statist. Sci.*, 14(4):382-417, 1999.
- Holt, R.M., J.L. Wilson, and R.J. Glass, Spatial bias in field-estimated unsaturated hydraulic properties, *Water Resour. Res.*, 38(12), 1311, doi:10.1029/2002WR001336, 2002.
- Hu, L.Y., G. Blanc, and B. Noetinger, Gradual deformation and iterative calibration of sequential stochastic simulations, *Math. Geol.*, 33(4):475-489, DOI10.1023/A:1011088913233, 2001.
- Hulstrom L.C., *Sampling and Analysis of the 300F-5 Operable Unit Springs and Near Shore Sediments and River Water*, WHC-SD-EN-TI-125, Westinghouse Hanford Company, Richland, Washington, 1993.
- Hurvich, C.M. and C.-L. Tsai, Regression and time series model selection in small samples, *Biometrika*, 76(2):297-307, 1989.
- Iman, R.L., and Shortencarier, M.J., *A Fortran 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models*, NUREG/CR-3624, Technical Report SAND83-2365, Sandia National Laboratories, Albuquerque, NM, 1984.
- Isukapalli, S.S., A. Roy and P.G. Georgopoulos, Stochastic response surface methods (SRSMs) for uncertainty propagation: application to environmental and biological systems, *Risk Analysis*, Vol. 18, No. 3, 1998.
- James A.L. and C.M. Oldenburg, Linear and Monte Carlo uncertainty analysis for subsurface contaminant transport simulation, *Water Resour. Res.*, 33(11):2495-2508, 1997.
- Jaynes, E.T. *Probability Theory, The Logic of Science*, Cambridge University Press, Cambridge, UK, 727 pp., 2003.
- Jefferys, W.H. and J.O. Berger, Ockham's razor and Bayesian analysis, *American Scientist*, 80(1):64-72, 1992.
- Jorgensen, D.G., Geohydrologic models of the Houston District, Texas, *Ground Water*, 28:418-428, 1991.
- Kaplan, D.I. and R.J. Serne, *Geochemical Data Package for the Hanford Immobilized Low-Activity Tank Waste Performance Assessment (ILAW PA)*, PNNL-13037, Rev. 1, Pacific Northwest National Laboratory, Richland, WA, 2000.
- Kaplan, D.I., I.V. Kutnyakov, A.P. Gamberdinger, R.J. Serne, and K.E. Parker, Gravel-corrected  $K_d$  values, *Ground Water*, 38(6):851-857, 2000.
- Kashyap, R.L., Optimal choice of AR and MA parts in autoregressive moving average models. *IEEE Trans. Pattern Anal. Mach. Intel. PAMI* 4(2): 99-104, 1982.

- Kass, R.E. and A.E. Raftery, Bayes factors, *J. Amer. Statist. Assoc.*, 90(430):773-795, 1995.
- Kennedy, W.E. Jr. and Streng, D.L., *Residual Radioactive Contamination From Decommissioning: Technical Basis for Translating Contamination Levels to Annual Total Effective Dose Equivalent*, NUREG/CR-5512, Volume 1, U.S. Nuclear Regulatory Commission, Washington, D.C., October 1992.
- Konikow L.F., Predictive accuracy of a ground-water model-lessons from a postaudit, *Ground Water*, 24:173-184, 1986.
- Konikow L.F. and M. Person, Assessment of long-term salinity changes in an irrigated stream-aquifer system, *Water Resour. Res.*, 21:1611-1624, 1985.
- Konikow L.F. and L.A. Swain, Assessment of predictive accuracy of a model of artificial recharge effects in the Coachella Valley, California. In: Simpson E.S., Sharp J.M., Jr. (eds.) *Selected Papers on Hydrogeology from the 28th Geological Congress*, Verlag Heinz Heise, Washington, DC, USA, pp. 433-449, 1990.
- Konikow L.F., W.E. Sanford, and P.J. Campbell, Constant concentration boundary condition: lessons from the HYDRCOIN variable-density groundwater benchmark problem, *Water Resour. Res.*, 33:2253-2261, 1997.
- Krishnamurti, T.N., C.M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, Multimodel ensemble forecasts for weather and seasonal climate, *J. Climate*, 13(23):4196-4216, 2000.
- Krupka K.M., and R.J. Serne, *Geochemical Factors Affecting the Behavior of Antimony, Cobalt, Europium, Technetium, and Uranium in Vadose Sediments*, PNNL-14126, Pacific Northwest National Laboratory, Richland, Washington, 2002.
- Kunk, J.R., *Geophysical Studies: 300-FF-5 Operable Unit*. WHC-SD-EN-DP-059, Westinghouse Hanford Company, Richland, Washington, 1993.
- Kunk J.R., S.M. Narbutovskih, K.A. Bergstrom, and T.M. Mitchell, *Phase I Summary of Surface Geophysical Studies in the 300-FF-5 Operable Unit*, WHC-SD-EN-TI-069, Westinghouse Hanford Company, Richland, Washington, 1993.
- Kunstmann, H., W. Kinzelbach, and T. Siegfried, Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modeling, *Water Resour. Res.*, 38(4):10.1029/2000WR000022, 2002.
- Last G.V., G.W. Gee, E.J. Freeman, W.E. Nichols, K.J. Cantrell, B.N. Bjornstad, M.J. Fayer, D.G. Horton, *Vadose Zone Hydrogeology Data Package for the 2004 Composite Analysis*, PNNL-14702, Rev. 0, Pacific Northwest National Laboratory, Richland, WA, 2004.
- Leung L.R., Y. Qian, X. Bian, W.M. Washington, J. Han, and J.O. Roads, Mid-century ensemble regional climate change scenarios for the western United States, *Climatic Change*, 62(1-3):75-113, 2004.
- Lewis B.D. and F.J. Goldstein, *Evaluation of a Predictive Groundwater Solute-Transport Model at the Idaho National Engineering Laboratory Testing Station, Idaho*, U.S. Geological Survey Water Resources Investigation No. 82-25, 1982.
- Lindberg, J.W., and F.W. Bond, *Geohydrology and Ground-Water Quality Beneath the 300 Area, Hanford Site, Washington*, PNL-2949, Pacific Northwest Laboratory, Richland, Washington, 1979.
- Lindberg J.W., and C.J. Chou, *300 Area Process Trenches Groundwater Monitoring Plan*, PNNL-13645, Pacific Northwest National Laboratory, Richland, Washington, 2001.
- Liu, C., J. Zachara, and N. Qafoku, "Modeling of Uranium Sorption/Desorption with Advection," presented at a workshop on *Conceptual Model Development and Reactive Transport Modeling for the 300 Area Uranium Plume in 300-FF-5*, held May 10-11, 2004, Richland, Washington, 2004. (<http://www.hanford.gov/cp/gpp/public/workshops.cfm>)
- Madigan D. and A.E. Raftery, Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Amer. Statist. Assoc.*, 89(428):1535-1546, 1994.
- McKay, M.D., *Evaluating Prediction Uncertainty*, NUREG/CR-6311, U.S. Nuclear Regulatory Commission, Washington, D.C., 1995.
- Medina, A. and J. Carrera, Coupled estimation of flow and solute transport parameters, *Water Resour. Res.*, 32(10), 3063-3076, 10.1029/96WR00754, 1996.
- Meyer, P.D., M.L. Rockhold, and G.W. Gee, *Uncertainty Analyses of Infiltration and Subsurface Flow and Transport for SDMP Sites*, NUREG/CR-6565, U.S. Nuclear Regulatory Commission, Washington, D.C., 1997. (<http://nrc-hydro-uncert.pnl.gov/>)

- Meyer, P.D. and G.W. Gee, *Information on Hydrologic Conceptual Models, Parameters, Uncertainty Analysis, and Data Sources for Dose Assessments at Decommissioning Sites*, NUREG/CR-6656, U.S. Nuclear Regulatory Commission, Washington, D.C., 1999. (<http://nrc-hydro-uncert.pnl.gov/>)
- Meyer, P.D., M. Ye, S.P. Neuman, and K.J. Cantrell, *Combined Estimation of Hydrogeologic Conceptual Model and Parameter Uncertainty*, NUREG/CR-6843 (PNNL-14534), U.S. Nuclear Regulatory Commission, Washington, DC., 2004.
- Morgan, M.G., and M. Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, Cambridge, United Kingdom, 1990.
- Morris, M.D., Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33(2):161-174, 1991.
- Moschandreas, D.J. and S. Karuchit, Scenario-model-parameter: a new method of cumulative risk uncertainty analysis, *Environment International*, 28, 247-261, 2002.
- Mosleh, A., N. Siu, C. Smidts, and C. Lui (eds.), *Model Uncertainty: Its Characterization and Quantification, Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis*, NUREG/CP-0138, U.S. Nuclear Regulatory Commission, Washington, D.C., 1994.
- Murray C.J., Y. Chien, and P.D. Thorne, *A Geostatistical Analysis of Historical Field Data on Tritium, Technetium-99, Iodine-129, and Uranium*. PNNL-14618, Rev. 0, Pacific Northwest National Laboratory, Richland, Washington, 2004.
- NEA, *Features, Events, and Processes (FEPs) for Geologic Disposal of Radioactive Waste, An International Database, 2000 Edition*, Nuclear Energy Agency-Organization For Economic Co-operation and Development, France, 2000.
- NEA, *Scenario Development Methods and Practice*, Nuclear Energy Agency-Organization For Economic Co-operation and Development, France, 2001.
- Neitzel, D.A. (ed.), *Hanford Site National Environmental Policy Act (NEPA) Characterization*, PNNL-6415, Rev. 10, Pacific Northwest National Laboratory, Richland, Washington, 1998.
- Neuman, S.P., Maximum likelihood Bayesian averaging of alternative conceptual-mathematical models, *Stochastic Environmental Research and Risk Assessment*, 17, DOI10.1007/s00477-003-0151-7, 2003.
- Neuman, S.P. and P.J. Wierenga, *A Comprehensive Strategy of Hydrogeologic Modeling and Uncertainty Analysis for Nuclear Facilities and Sites*, NUREG/CR-6805, U.S. Nuclear Regulatory Commission, Washington, D.C., 2003.
- NRC, *NMSS Decommissioning Standard Review Plan*, NUREG-1727, U.S. Nuclear Regulatory Commission, September, 2000.
- NRC, *Consolidated NMSS Decommissioning Guidance, Decommissioning Process for Materials Licensees*, NUREG-1757 (Vol. 2, Appendix I), U.S. Nuclear Regulatory Commission, Washington, DC, 2003.
- Poeter E.P., M.C. Hill, E.R. Banta, S.W. Mehl, and S. Christensen, *UCODE\_2005 and Six Other Computer Codes for Universal Sensitivity Analysis, Inverse Modeling, and Uncertainty Evaluation*, U.S. Geological Survey Techniques and Methods Report TM 6-A11, U.S. Geological Survey, Denver, Colorado, 2005.
- Poeter, E.P. and D. Anderson, Multimodel ranking and inference in ground water modeling, *Ground Water*, 43(4):597-605, 2005.
- Raftery A.E., D. Madigan, C.T. Volinsky, Accounting for model uncertainty in survival analysis improves predictive performance, in: *Bayesian Statistics*, J. Bernardo, J. Berger, A. Dawid, A. Smith (eds.), Oxford Univ. Press, pp. 323-349, 1996.
- Saltelli, A., K. Chan, and E.M. Scott (eds.), *Sensitivity Analysis*, John Wiley & Sons LTD, Chichester, England, 475 pp., 2000a.
- Saltelli, A., S. Tarantola and F. Campolongo, Sensitivity analysis as an ingredient of modeling, *Statistical Science*, Vol. 15, 4:377-395, 2000b.
- Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, Wiley, New York, 232 pp., 2004.
- Samper, F.J. and S.P. Neuman, Estimation of spatial covariance structures by adjoint state maximum likelihood cross-validation: 1. Theory, *Water Resour. Res.*, 25(3), 351-362, 1989a.
- Samper, F.J. and S.P. Neuman, Estimation of spatial covariance structures by adjoint state maximum likelihood cross-validation: 1. Synthetic experiments, *Water Resour. Res.*, 25(3), 363-371, 1989b.



- Schalla R., R.W. Wallace, R.L. Aaberg, S.P. Airhart, D.J. Bates, J.V.M. Carlile, C.S. Cline, D.I. Dennison, M.D. Freshley, P.R. Heller, E.J. Jensen, K.B. Olsen, R.G. Parkhurst, J.T. Rieger, and E.J. Westergard, *Interim Characterization Report for the 300 Area Process Trenches*, PNL-6716, Pacific Northwest Laboratory, Richland, Washington, 1988.
- Serne R.J., C.W. Lindenmeier, P.K. Bhatia, and V.L. LeGore, *Contaminant Concentration vs. Particle Size for 300 Area North Process Pond Samples*, WHC-SD-EN-TI-049 (PNL-8401), Westinghouse Hanford Company, 1992.
- Serne R.J., C.F. Brown, H.T. Schaef, E.M. Pierce, M.J. Lindberg, Z. Wang, P. Gassman, and J. Catalano, *300 Area Uranium Leach and Adsorption Project*, PNNL-14022, Pacific Northwest National Laboratory, Richland, Washington, 2002.
- Sheppard, M.L., and D.H. Thibault, Default soil solid/liquid partition coefficients, K<sub>ds</sub>, for four major soil types: A compendium, *Health Physics*, 59(4):471-478, 1990.
- Shippers, L.R., *Background Information for the Development of a Low-Level Waste Performance Assessment Methodology: Identification of Potential Exposure Pathways*, NUREG/CR-5453, Volume 1, Sandia National Laboratories, 1989.
- Shippers, L.R., and C.P. Harlan, *Background Information for the Development of a Low-Level Waste Performance Assessment Methodology: Assessment of Relative Significance of Migration and Exposure Pathways*, NUREG/CR-5453, Volume 2, Sandia National Laboratories, 1989.
- Simpson, B., R.A. Corbin, and S.F. Agnew, *Hanford Soil Inventory Model*, BHI-01496, Rev. 0, Bechtel Hanford, Inc., Richland, Washington, 2001.
- Solar, L.G., G.V. Last, B.A. Napier, V.J. Rohay, and F.J. Schelling, *The Application of Feature, Event, and Process Methodology at the Hanford Site*, BHI-01573, Rev. 0, Bechtel Hanford, Inc., Richland, Washington, 2001.
- Stewart, M. and C. Langevin, Post Audit of a numerical prediction of wellfield drawdown in a semiconfined aquifer system, *Ground Water*, 37(2):245-252, 1999.
- Sun, N-Z and W.W-G Yeh, Coupled Inverse Problems in Groundwater Modeling, I. Sensitivity Analysis and Parameter Identification, *Water Resour. Res.*, 26 (10), 2507-2525, 1990.
- Swanson L.C., G.G. Kelty, K.A. Lindsey, K.R. Simpson, R.K. Price, and S.D. Consort, *Phase 1 Hydrogeologic Summary of the 300-FF-5 Operable Unit, 300 Area*, WHC-SD-EN-TI-052, Rev. 0, Westinghouse Hanford Company, Richland, Washington, 1992.
- Taplin, R.H., Robust likelihood calculation for time series, *J. Roy. Statist. Soc. Ser. B*, 55:829-836, 1993.
- Thorne P.D. and D.R. Newcomer, *Summary and Evaluation of Available Hydraulic Property Data for the Hanford Site Unconfined Aquifer System*, PNL-8337, Pacific Northwest Laboratory, Richland, Washington, 1992.
- Tiedeman, C.R., M.C. Hill, F.A. D'Agnese, C.C. Faunt, Methods for using groundwater model predictions to guide hydrogeologic data collection, with application to the Death Valley regional groundwater flow system, *Water Resour. Res.*, 39(1), 1010, doi:10.1029/2001WR001255, 2003
- Vecchia, A.V. and R.L. Cooley, Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model, *Water Resour. Res.*, 22(2):95-108, 1987.
- Vermeul V.R., M.P. Bergeron, C.R. Cole, C.J. Murray, W.E. Nichols, T.D. Scheibe, P.D. Thorne, S.R. Waichler, and Y. Xie, *Transient Inverse Calibration of the Site-Wide Groundwater Flow Model (ACM-2): FY03 Progress Report*, PNNL-14398, Pacific Northwest National Laboratory, Richland, Washington, 2003.
- Volinsky C.T., D. Madigan, A.E. Raftery, R.A. Kronmal, Bayesian model averaging in proportional hazard models: assessing the risk of a stroke. *J. Roy. Statist. Soc. Ser. C* 46, 433-448, 1997.
- Waichler, S.R. and S.B. Yabusaki, *Flow and Transport in the Hanford 300 Area Vadose Zone-Aquifer-River System*, PNNL-15125, Pacific Northwest National Laboratory, Richland, Washington, 2005.
- Waichler, S.R., W.A. Perkins, and M.C. Richmond, *Hydrodynamic Simulation of the Columbia River, Hanford Reach, 1940-2004*, PNNL-15226, Pacific Northwest National Laboratory, Richland, Washington, 2005.
- Wang, W., S.P. Neuman, T. Yao, and P.J. Wierenga, Simulation of large-scale field infiltration experiments using a hierarchy of models based on public, generic, and site data, *Vadose Zone Journal*, 2:297-312, 2003.

- Westall J.C., *FITEQL – A Program for the Determination of Chemical Equilibrium Constants from Experimental Data; User's Guide*, Version 2.0. Chemistry Department, Oregon State University, Corvallis, Oregon, 1982.
- Wurstner S.K., P.D. Thorne, M.A. Chamness, M.D. Freshley, and M.D. Williams, *Development of a Three-Dimensional Groundwater Model of the Hanford Site Unconfined Aquifer System: FY 1995 Status Report*, PNL-10886, Pacific Northwest Laboratory, Richland, Washington, 1995.
- Ye, M., S.P. Neuman, and P.D. Meyer, Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resour. Res.*, 40, W05113, doi:10.1029/2003WR002557, 2004.
- Ye, M., S.P. Neuman, P.D. Meyer, and K. Pohlmann, Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff, *Water Resour. Res.*, 41, W12429, doi:10.1029/2005WR004260, 2005.
- Zachara, J.M., "Batch Measurements of Desorption/Adsorption from Vadose Zone and Aquifer Sediments," presented at a workshop on *Conceptual Model Development and Reactive Transport Modeling for the 300 Area Uranium Plume in 300-FF-5*, held May 10-11, 2004, Richland, Washington, 2004.  
<http://www.hanford.gov/cp/gpp/public/workshops.cfm>.
- Zachara, J.M., J.A. Davis, C. Liu, J.P. McKinley, N. Qafoku, D.M. Wellman, and S.B. Yabusaki, *Uranium Geochemistry in Vadose Zone and Aquifer Sediments from the 300 Area Uranium Plume*, PNNL-15121, Pacific Northwest National Laboratory, Richland, Washington, 2005.
- Zhang, D., *Stochastic Methods for Flow in Porous Media*, Academic Press, 2001.
- Zheng, C. and P.P. Wang, *MT3DMS, A modular three-dimensional multi-species transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems; documentation and user's guide*, U.S. Army Engineer Research and Development Center Contract Report SERDP-99-1, Vicksburg, MS, 202 pp., 1999.  
<http://hydro.geo.ua.edu/mt3d/>
- Zimmerman, D.A., G. de Marsily, C.A. Gotway, M.G. Marietta, C.L. Axness, R.L. Beauheim, R.L. Bras, J. Carrera, G. Dagan, P.B. Davies, D.P. Gallegos, A. Galli, J. Gomez-Hernandez, P. Grindrod, A. L. Gutjahr, P.K. Kitanidis, A.M. Lavenue, D. McLaughlin, S.P. Neuman, B.S. RamaRao, C. Ravenne, and Y. Rubin, A comparison of seven geostatistically-based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resour. Res.*, 34(6):1373-1413, 1998.





## Appendix A

### Derivation of Posterior Mean and Variance Considering Model Structure and Parameter Uncertainty for a Given Scenario

Posterior mean  $E(\Delta|S_i, \mathbf{D})$  can be evaluated via

$$E(\Delta|S_i, \mathbf{D}) = \int \Delta p(\Delta|S_i, \mathbf{D}) d\Delta \quad (\text{A1})$$

Substituting the expression of  $p(\Delta|S_i, \mathbf{D})$  (8) into (A1) gives

$$E(\Delta|S_i, \mathbf{D}) = \int \Delta \sum_{k=1}^K p(\Delta|M_k, S_i, \mathbf{D}) p(M_k|S_i, \mathbf{D}) d\Delta \quad (\text{A2})$$

which can be rewritten as

$$E(\Delta|S_i, \mathbf{D}) = \sum_{k=1}^K \int \Delta p(\Delta|M_k, S_i, \mathbf{D}) d\Delta p(M_k|S_i, \mathbf{D}) \quad (\text{A3})$$

Considering that  $E(\Delta|M_k, S_i, \mathbf{D}) = \int \Delta p(\Delta|M_k, S_i, \mathbf{D}) d\Delta$ , (A3) leads to (18) directly.

Posterior variance  $Var(\Delta|S_i, \mathbf{D})$  can be evaluated via

$$\begin{aligned} Var(\Delta|S_i, \mathbf{D}) &= E\left(\left[\Delta - E(\Delta|S_i, \mathbf{D})\right]^2 \middle| \mathbf{D}\right) = \int \left[\Delta - E(\Delta|S_i, \mathbf{D})\right]^2 p(\Delta|S_i, \mathbf{D}) d\Delta \\ &= \int \Delta^2 p(\Delta|S_i, \mathbf{D}) d\Delta - 2E(\Delta|S_i, \mathbf{D}) \int \Delta p(\Delta|S_i, \mathbf{D}) d\Delta + \left[E(\Delta|S_i, \mathbf{D})\right]^2 \int p(\Delta|S_i, \mathbf{D}) d\Delta \end{aligned} \quad (\text{A4})$$

Recalling (A1) and  $\int p(\Delta|S_i, \mathbf{D}) d\Delta = 1$ , the last two terms of (A4) are

$$-2E(\Delta|S_i, \mathbf{D}) \int \Delta p(\Delta|S_i, \mathbf{D}) d\Delta + \left[E(\Delta|S_i, \mathbf{D})\right]^2 \int p(\Delta|S_i, \mathbf{D}) d\Delta = -\left[E(\Delta|S_i, \mathbf{D})\right]^2 \quad (\text{A5})$$

and (A4) thus becomes

$$Var(\Delta|S_i, \mathbf{D}) = \int \Delta^2 p(\Delta|S_i, \mathbf{D}) d\Delta - \left[E(\Delta|S_i, \mathbf{D})\right]^2 \quad (\text{A6})$$

Substituting the expression of  $p(\Delta|S_i, \mathbf{D})$  (15) into the first term at the right hand side of (A6) gives

$$\begin{aligned}
\int \Delta^2 p(\Delta|S_i, \mathbf{D}) d\Delta &= \int \Delta^2 \sum_{k=1}^K p(\Delta|M_k, S_i, \mathbf{D}) p(M_k|S_i, \mathbf{D}) d\Delta \\
&= \sum_{k=1}^K \int \Delta^2 p(\Delta|M_k, S_i, \mathbf{D}) d\Delta p(M_k|S_i, \mathbf{D}) \\
&= \sum_{k=1}^K E(\Delta^2|M_k, S_i, \mathbf{D}) p(M_k|S_i, \mathbf{D})
\end{aligned} \tag{A7}$$

Substituting the expression of variance

$$E(\Delta^2|M_k, S_i, \mathbf{D}) = \text{Var}(\Delta|M_k, S_i, \mathbf{D}) + [E(\Delta|M_k, S_i, \mathbf{D})]^2 \tag{A8}$$

into (A7) gives

$$\int \Delta^2 p(\Delta|S_i, \mathbf{D}) d\Delta = \sum_{k=1}^K \left[ \text{Var}(\Delta|M_k, S_i, \mathbf{D}) + [E(\Delta|M_k, S_i, \mathbf{D})]^2 \right] p(M_k|S_i, \mathbf{D}) \tag{A9}$$

Substituting (A9) into (A6) yields (19) directly.

Equation (19) can be written as

$$\begin{aligned}
\text{Var}(\Delta|S_i, \mathbf{D}) &= \\
&\sum_{k=1}^K \text{Var}(\Delta|M_k, S_i, \mathbf{D}) p(M_k|S_i, \mathbf{D}) + \sum_{k=1}^K [E(\Delta|M_k, S_i, \mathbf{D})]^2 p(M_k|S_i, \mathbf{D}) - [E(\Delta|S_i, \mathbf{D})]^2 \tag{A10}
\end{aligned}$$

Consider the expression

$$\begin{aligned}
&\sum_{k=1}^K [E(\Delta|M_k, S_i, \mathbf{D}) - E(\Delta|S_i, \mathbf{D})]^2 p(M_k|S_i, \mathbf{D}) = \sum_{k=1}^K [E(\Delta|M_k, S_i, \mathbf{D})]^2 p(M_k|S_i, \mathbf{D}) \\
&- 2E(\Delta|S_i, \mathbf{D}) \sum_{k=1}^K E(\Delta|M_k, S_i, \mathbf{D}) p(M_k|S_i, \mathbf{D}) + [E(\Delta|S_i, \mathbf{D})]^2 \sum_{k=1}^K p(M_k|S_i, \mathbf{D})
\end{aligned} \tag{A11}$$

Given (17) and (18), (A11) becomes

$$\begin{aligned}
&\sum_{k=1}^K [E(\Delta|M_k, S_i, \mathbf{D}) - E(\Delta|S_i, \mathbf{D})]^2 p(M_k|S_i, \mathbf{D}) = \\
&\sum_{k=1}^K [E(\Delta|M_k, S_i, \mathbf{D})]^2 p(M_k|S_i, \mathbf{D}) - [E(\Delta|S_i, \mathbf{D})]^2
\end{aligned} \tag{A12}$$

Substituting (A12) into (A10) leads to (20) directly.

## Appendix B

### Derivation of Posterior Mean and Variance Considering Model Structure, Parameter, and Scenario Uncertainties

Posterior mean  $E(\Delta|\mathbf{D})$  can be evaluated via

$$E(\Delta|\mathbf{D}) = \int \Delta p(\Delta|\mathbf{D}) d\Delta \quad (\text{B1})$$

Substituting the expression of  $p(\Delta|\mathbf{D})$  (21) into (B1) gives

$$E(\Delta|\mathbf{D}) = \int \Delta \sum_{i=1}^I \sum_{k=1}^K p(\Delta|M_k, S_i, \mathbf{D}) p(M_k|S_i, \mathbf{D}) p(S_i) d\Delta \quad (\text{B2})$$

which can be rewritten as

$$E(\Delta|\mathbf{D}) = \sum_{i=1}^I \sum_{k=1}^K \int \Delta p(\Delta|M_k, S_i, \mathbf{D}) d\Delta p(M_k|S_i, \mathbf{D}) p(S_i) \quad (\text{B3})$$

Considering that  $E(\Delta|M_k, S_i, \mathbf{D}) = \int \Delta p(\Delta|M_k, S_i, \mathbf{D}) d\Delta$ , (B3) leads to (23) directly.

Posterior variance  $Var(\Delta|\mathbf{D})$  can be evaluated via

$$\begin{aligned} Var(\Delta|\mathbf{D}) &= E\left(\left[\Delta - E(\Delta|\mathbf{D})\right]^2|\mathbf{D}\right) = \int \left[\Delta - E(\Delta|\mathbf{D})\right]^2 p(\Delta|\mathbf{D}) d\Delta \\ &= \int \Delta^2 p(\Delta|\mathbf{D}) d\Delta - 2E(\Delta|\mathbf{D}) \int \Delta p(\Delta|\mathbf{D}) d\Delta + \left[E(\Delta|\mathbf{D})\right]^2 \int p(\Delta|\mathbf{D}) d\Delta \end{aligned} \quad (\text{B4})$$

Recalling (B1) and  $\int p(\Delta|\mathbf{D}) d\Delta = 1$ , the last two terms of (B4) are

$$-2E(\Delta|\mathbf{D}) \int \Delta p(\Delta|\mathbf{D}) d\Delta + \left[E(\Delta|\mathbf{D})\right]^2 \int p(\Delta|\mathbf{D}) d\Delta = -\left[E(\Delta|\mathbf{D})\right]^2 \quad (\text{B5})$$

and (B4) thus becomes

$$Var(\Delta|\mathbf{D}) = \int \Delta^2 p(\Delta|\mathbf{D}) d\Delta - \left[E(\Delta|\mathbf{D})\right]^2 \quad (\text{B6})$$

Substituting the expression of  $p(\Delta|\mathbf{D})$  (21) into the first term on the right hand side of (B6) gives

$$\begin{aligned}
\int \Delta^2 p(\Delta|\mathbf{D}) d\Delta &= \int \Delta^2 \sum_{i=1}^I \sum_{k=1}^K p(\Delta|M_k, S_i, \mathbf{D}) p(M_k|S_i, \mathbf{D}) p(S_i) d\Delta \\
&= \sum_{i=1}^I \sum_{k=1}^K \int \Delta^2 p(\Delta|M_k, S_i, \mathbf{D}) d\Delta p(M_k|S_i, \mathbf{D}) p(S_i) \\
&= \sum_{i=1}^I \sum_{k=1}^K E(\Delta^2|M_k, S_i, \mathbf{D}) p(M_k|S_i, \mathbf{D}) p(S_i)
\end{aligned} \tag{B7}$$

Substituting the expression of variance

$$E(\Delta^2|M_k, S_i, \mathbf{D}) = \text{Var}(\Delta|M_k, S_i, \mathbf{D}) + [E(\Delta|M_k, S_i, \mathbf{D})]^2 \tag{B8}$$

into (B7) gives

$$\begin{aligned}
\int \Delta^2 p(\Delta|\mathbf{D}) d\Delta &= \\
&\sum_{i=1}^I \sum_{k=1}^K \left[ \text{Var}(\Delta|M_k, S_i, \mathbf{D}) + [E(\Delta|M_k, S_i, \mathbf{D})]^2 \right] p(M_k|S_i, \mathbf{D}) p(S_i)
\end{aligned} \tag{B9}$$

Substituting (B9) into (B6) yields (24) directly.

Considering (19), equation (24) can be written as

$$\begin{aligned}
\text{Var}(\Delta|\mathbf{D}) &= \sum_{i=1}^I \left( \text{Var}(\Delta|S_i, \mathbf{D}) + [E(\Delta|S_i, \mathbf{D})]^2 \right) p(S_i) - [E(\Delta|\mathbf{D})]^2 \\
&= \sum_{i=1}^I \text{Var}(\Delta|S_i, \mathbf{D}) p(S_i) + \sum_{i=1}^I [E(\Delta|S_i, \mathbf{D})]^2 p(S_i) - [E(\Delta|\mathbf{D})]^2
\end{aligned} \tag{B10}$$

Consider the expression below

$$\begin{aligned}
&\sum_{i=1}^I [E(\Delta|S_i, \mathbf{D}) - E(\Delta|\mathbf{D})]^2 p(S_i) \\
&= \sum_{i=1}^I [E(\Delta|S_i, \mathbf{D})]^2 p(S_i) - 2E(\Delta|\mathbf{D}) \sum_{i=1}^I E(\Delta|S_i, \mathbf{D}) p(S_i) + [E(\Delta|\mathbf{D})]^2 \sum_{i=1}^I p(S_i)
\end{aligned} \tag{B11}$$

which, using (22), becomes

$$\begin{aligned}
&\sum_{i=1}^I [E(\Delta|S_i, \mathbf{D}) - E(\Delta|\mathbf{D})]^2 p(S_i) \\
&= \sum_{i=1}^I [E(\Delta|S_i, \mathbf{D})]^2 p(S_i) - 2E(\Delta|\mathbf{D}) \sum_{i=1}^I E(\Delta|S_i, \mathbf{D}) p(S_i) + [E(\Delta|\mathbf{D})]^2
\end{aligned} \tag{B12}$$

Using (19), (B12) can be written as



$$\begin{aligned}
& \sum_{i=1}^I \left[ E(\Delta|S_i, \mathbf{D}) - E(\Delta|\mathbf{D}) \right]^2 p(S_i) \\
&= \sum_{i=1}^I \left[ E(\Delta|S_i, \mathbf{D}) \right]^2 p(S_i) - 2 \left[ E(\Delta|\mathbf{D}) \right]^2 + \left[ E(\Delta|\mathbf{D}) \right]^2 \\
&= \sum_{i=1}^I \left[ E(\Delta|S_i, \mathbf{D}) \right]^2 p(S_i) - \left[ E(\Delta|\mathbf{D}) \right]^2
\end{aligned} \tag{B13}$$

Considering (22), (B13) leads to

$$\sum_{i=1}^I \left[ E(\Delta|S_i, \mathbf{D}) \right]^2 p(S_i) - \left[ E(\Delta|\mathbf{D}) \right]^2 = \sum_{i=1}^I \left[ E(\Delta|S_i, \mathbf{D}) - E(\Delta|\mathbf{D}) \right]^2 p(S_i) \tag{B14}$$

Substituting (B14) into (B10) gives (25) directly.



## Appendix C

### On Evaluation of Recharge Model Uncertainty: A Priori and A Posteriori

M. Ye, K. Pohlmann, J. Chapman, and D. Shafer

Previously published in Proceedings of the International High-Level Radioactive Waste Management  
Conference  
April 30 – May 4, 2006, Las Vegas, Nevada

#### Abstract

Hydrologic environments are open and complex, rendering them prone to multiple interpretations and mathematical descriptions. Hydrologic analyses typically rely on a single conceptual-mathematical model, which ignores conceptual model uncertainty and may result in bias in predictions and under-estimation of predictive uncertainty. This study is to assess conceptual model uncertainty residing in five recharge models developed to date by different researchers based on different theories for Nevada and Death Valley area, CA. A recently developed statistical method, Maximum Likelihood Bayesian Model Averaging (MLBMA), is utilized for this analysis. In a Bayesian framework, the recharge model uncertainty is assessed, a priori, using expert judgments collected through an expert elicitation in the form of prior probabilities of the models. The uncertainty is then evaluated, a posteriori, by updating the prior probabilities to estimate posterior model probability. The updating is conducted through maximum likelihood inverse modeling by calibrating the Death Valley Regional Flow System (DVRFS) model corresponding to each recharge model against observations of head and flow. Calibration results of DVRFS for the five recharge models are used to estimate three information criteria (AIC, BIC, and KIC) used to rank and discriminate these models. Posterior probabilities of the five recharge models, evaluated using KIC, are used as weights to average head predictions, which gives posterior mean and variance. The posterior quantities incorporate both parametric and conceptual model uncertainties.

#### 1 INTRODUCTION

Hydrologic analyses are commonly based on a single conceptual-mathematical model. Yet hydrologic environments are open and complex, rendering them prone to multiple interpretations and mathematical descriptions. This is true regardless of the quantity and quality of available hydrologic information and data. Focusing on only one conceptual-mathematical model may lead to a Type I model error, which arises when one rejects (by omission) valid alternative models. It may also result in a Type II model error, which arises when one adopts (fails to reject) an invalid conceptual-mathematical model. Indeed, critiques of hydrologic analyses, and legal challenges to them, typically focus on the validity of the underlying conceptual (and by implication mathematical) model. If a proposed model is found to be severely deficient, hydrologic analysis based on the single model may damage professional credibility of the work;

result in the loss of a legal contest; and lead to adverse environmental, economic and political impacts ([1-2]).

The need to properly assess conceptual model uncertainty has motivated the recent development of a Maximum Likelihood Bayesian Model Averaging method (MLBMA) [3-4]. MLBMA is being applied in our study to assess conceptual model uncertainty in the Death Valley Regional Flow System (DVRFS) model, developed by the U.S. Geological Survey [5] to simulate the regional flow system in southwest Nevada and southeast California. This area includes the U.S. Department of Energy proposed Yucca Mountain nuclear repository, the nation's first long-term permanent geologic repository of spent nuclear fuel and high-level radioactive waste.

Our study is focused on assessing conceptual model uncertainty due to five alternative recharge models listed in Table C-1: (1) the Maxey-Eakin (ME) model [6], (2) two distributed parameter watershed (DPW) models, one with and one without a runoff-runoff component [7], and (3) two chloride mass balance (CMB) models, each with different zero-recharge masks, one for alluvium and one for both alluvium and elevation [8]. These five models are based on different methodologies for estimating recharge and have different levels of complexity, and they all have been used for groundwater modeling in Nevada. Recharge estimates of the five models are plotted in Figure C-1, and they are significantly different. A large amount of conceptual model uncertainty exists in the recharge models [9]. Since recharge significantly affects modeled groundwater flow paths and travel times, it is important to evaluate the recharge model uncertainty and quantify its propagation through the groundwater modeling process.

Using MLBMA, we assess the recharge model uncertainty *a priori* and *a posteriori*. The terms "*a priori*" and "*a posteriori*" refer primarily to how or on what basis our assessment is conducted. An assessment is conducted *a priori* if it is based on prior information without calibrating the regional flow model (of which the recharge model is a component) against site observations (e.g., hydraulic head and groundwater flux). The prior information includes assessment of model uncertainty from a similar site and/or expert judgments based on one's professional experience. An assessment is conducted *a posteriori* when the regional flow model (of which the recharge model is a component) is calibrated against site observations. In MLBMA, assessments *a priori* and *a posteriori* are quantified by prior and posterior model probabilities, as discussed below.



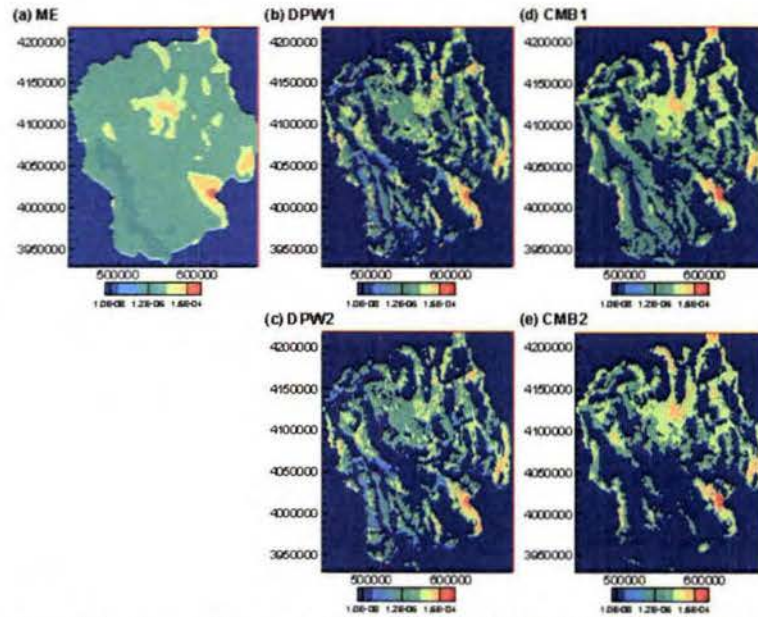


Figure C-1. Illustration of the five recharge models: (a) ME, (b) DPW1, (c) DPW2, (d) CMB1, and (e) CMB2.

Table C-1. Abbreviation and Description of the Five Recharge Models.

Models	Model Description
ME	Maxey-Eakin model
DPW1	Distributed parameter watershed model with runoff-runoff component
DPW2	Distributed parameter watershed model without runoff-runoff component
CMB1	Chloride mass balance model with fluvial mask
CMB2	Chloride mass balance model with fluvial and elevation masks

## 2 MAXIMUM LIKELIHOOD BAYESIAN MODEL AVERAGING (MLBMA)

To render our paper complete and self-contained, we start with a brief description of MLBMA; for additional details the reader is referred to [3-4]. If  $\Delta$  is the desired predicted quantity given a set of  $K$  alternative models, then its posterior distribution, given a discrete set  $\mathbf{D}$  of site data, is

$$p(\Delta|\mathbf{D}) = \sum_{k=1}^K p(\Delta|M_k, \mathbf{D}) p(M_k|\mathbf{D}) \quad (C1)$$

where  $p(\Delta|M_k, \mathbf{D})$  is the posterior distribution of  $\Delta$  under model  $M_k$  and  $p(M_k|\mathbf{D})$  is posterior probability of  $M_k$ . With consideration of parametric and conceptual model uncertainty, mean and variance of  $\Delta$  are

$$E[\Delta|\mathbf{D}] = \sum_{k=1}^K E[\Delta|\mathbf{D}, M_k] p(M_k|\mathbf{D}) \quad (C2)$$

$$Var[\Delta|\mathbf{D}] = \sum_{k=1}^K Var[\Delta|\mathbf{D}, M_k] p(M_k|\mathbf{D}) + \sum_{k=1}^K (E[\Delta|\mathbf{D}, M_k] - E[\Delta|\mathbf{D}])^2 p(M_k|\mathbf{D}) \quad (C3)$$

where  $E[\Delta|\mathbf{D}, M_k]$  and  $Var[\Delta|\mathbf{D}, M_k]$  are mean and variance of  $\Delta$  under model  $M_k$  due to uncertainty of parameters associated with  $M_k$ . The weight  $p(M_k|\mathbf{D})$  used to average model predictions and corresponding predictive variance is posterior model probability of model  $M_k$ , evaluated using Bayes' rule

$$p(M_k|\mathbf{D}) = \frac{p(\mathbf{D}|M_k)p(M_k)}{\sum_{l=1}^K p(\mathbf{D}|M_l)p(M_l)} \quad (C4)$$

where  $p(\mathbf{D}|M_k)$  is likelihood of model  $M_k$  (a measure of consistency between model predictions and site observations  $\mathbf{D}$ ) and  $p(M_k)$  is prior probability of  $M_k$ . Estimating prior model probability will be discussed in detail in Section III. The posterior model probability is conditioned on site observations explicitly and prior information implicitly. According to [4], equation (C4) can be approximated as

$$p(M_k|\mathbf{D}) \approx \frac{\exp\left(-\frac{1}{2}KIC_k\right)p(M_k)}{\sum_{l=1}^K \exp\left(-\frac{1}{2}KIC_l\right)p(M_l)} \quad (C5)$$

where  $KIC$  is Kashyap information criterion defined as [10]

$$KIC_k = (N - N_k) \ln \sigma_k^2 - N_k \ln 2\pi + \ln |\mathbf{X}_k^T \boldsymbol{\omega} \mathbf{X}_k| \quad (C6)$$

where  $N$  is number of calibration data  $\mathbf{D}$ ,  $N_k$  is number of parameters  $\boldsymbol{\theta}_k$  associated with model  $M_k$ ,  $e$  is the natural number,  $\boldsymbol{\omega}$  (the same for all models) is weight matrix associated with calibration data  $\mathbf{D}$ , and  $\mathbf{X}$  is sensitivity matrix with element  $X_{k,ij} = \partial D_{k,i} / \partial \theta_{k,j}$  evaluated at maximum likelihood parameter estimates,  $\hat{\boldsymbol{\theta}}_{k,j}$  ( $D_{k,i}$  being predictions at locations of  $D_i$  by model  $M_k$ ).  $\hat{\boldsymbol{\theta}}_{k,j}$  can be estimated using maximum likelihood (or, equivalently, generalized least square) methods, which also gives the calculated error variance,  $\sigma_k^2$ ,

$$\sigma_k^2 = \frac{\mathbf{e}_k^T \boldsymbol{\omega} \mathbf{e}_k}{N} = \frac{WSSR_k}{N} \quad (C7)$$

where  $\mathbf{e} = \mathbf{D} - \mathbf{D}'$  is residual and  $WSSR_k$  is weighted sum of squared residual of model  $M_k$ . All the quantities above can be estimated based on results of model calibration using common software such as MODFLOW2000 [11].

### 3. EVALUATE RECHARGE MODEL UNCERTAINTY: A PRIORI

Conceptual uncertainty of the five recharge models is first evaluated, *a priori*, using prior probabilities of the models. Prior model probability is interpreted by [4] as subjective values reflecting the analyst's (or a group of analysts') belief about the relative plausibility of each model (or a group of models) based on its apparent (qualitative, *a priori*) consistency with available knowledge and data. The analyst's perception, degree of reasonable belief [12], or confidence [13] in a model is ideally based on expert judgment, which is considered by [14] as the basis of conceptual model development. Hence we view integrating expert judgment in MLBMA (by specifying subjective prior probabilities) to be a strength rather than a weakness. According to this view, the models included in the model set must be those (and only those) that experts consider being of potential relevance to the problem at hand. Given a set of alternative models, their prior probabilities sum up to one,

$$\sum_{k=1}^K p(M_k) = 1 \quad (C8)$$

This implies that all possible models of relevance are included in the model set (collective exhaustiveness), and that all models in the set differ from each other sufficiently to be considered mutually exclusive (the joint probability of two or more models being zero).

Following the process suggested by [15], an expert elicitation is conducted to elicit professional judgments from seven experts on uncertainty of the five recharge models. Elicited prior probabilities of the five models are plotted in Figure C-2, which shows that the maximum and minimum prior probabilities are 45% and 5%, respectively. Models ME, DPW1, and CMB2 are the three most plausible models, and do not receive the minimum prior probability from any expert. Although the experts evaluate the models from various aspects (e.g., model assumptions and sensitivity of model predictions to model parameters), no experts place more than 50% prior probability on any model.

The prior model probabilities are aggregated using simple averaging, i.e.,

$$P_k = \frac{1}{NE} \sum_{i=1}^{NE} P_{ik} \quad (C9)$$

where  $NE=7$  is number of experts and  $P_{ik}$  is the prior probability expert  $i$  assigns to alternative model  $M_k$ . The aggregated prior model probabilities are plotted in Figure C-3. Models DPW1 and DPW2 have the largest and smallest probability, respectively, since the experts regard that including the runoff-runoff component is more realistic. Probability of the model CMB2 is larger than that of CMB1, since experts regard that elevation mask is an important feature in recharge estimation. Model ME is ranked as the second most plausible model and has prior model probability of 25%, although this model is the simplest one and its recharge estimation is significantly different from that of other four models. Although prior probabilities given by each expert are significantly different (Figure C-2), the aggregated probabilities are more or less uniform, considering that the equally likely prior probability is 20%. The largest deviation from the equally likely prior probability is only 10% for model DPW1. This manifests the inherent uncertainty in the recharge models, since they are developed independently based on solid physical principles and assumptions, calibrated with site measurements, and have all been applied to water resource management in Nevada. Since none of the models dominates over other models and all models have prior model probabilities larger than 5%, there is no justification to select one model and discard others, *a priori*.

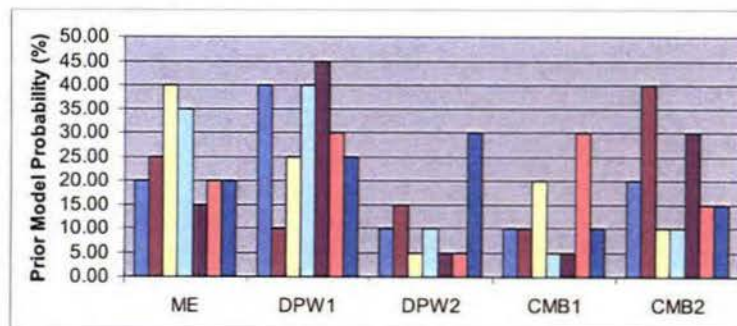




Figure C-2. Column chart of prior probabilities of the five models given by seven experts. Columns of each model represent elicited prior model probability from one expert. Model names are explained in Table C-1.

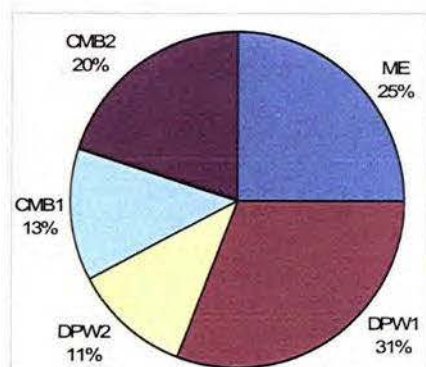


Figure C-3. Prior probabilities of the five recharge models obtained through an expert elicitation. Model names are explained in Table C-1.

#### 4. EVALUATE RECHARGE MODEL UNCERTAINTY: A POSTERIORI

Recharge model uncertainty is assessed, *a posteriori*, by maximum likelihood model calibration against site observations. Results of model calibration are used to estimate model likelihood  $p(\mathbf{D}|M_k)$ , which, in turn, is used to evaluate posterior model probability  $p(M_k|\mathbf{D})$  in (C4). Whereas prior model probabilities must in our view remain subjective, the posterior model probabilities are modifications of these subjective values based on an objective evaluation of each model's consistency with available data.

##### 4.1 Model Calibration Using MODFLOW2000

Plausibility and uncertainty of each of the five recharge models is evaluated by calibrating the Death Valley Regional Flow System (DVRFS) model, of which the recharge model is a component. DVRFS was modeled by [5] using MODFLOW2000, and a three-dimensional hydrogeologic framework based on characterization of regional geology, hydrology, and hydrogeology. The recharge model used in DVRFS is DPW1 developed by [7]. Our study is to assess recharge model uncertainty in the modeling framework of DVRFS, without modifying its other components. DVRFS was calibrated using MODFLOW2000 against a total of 4,963 observations of head (2,227), head change (2,672), discharge (49), and constant-head flow (15). These observations are also used in our calibration.

Our calibration process, however, is different from that of DVRFS, which calibrated 55 model parameters, 23 in the steady-state model and 32 in the transient model. Our model calibration is based on the transient model only, since there is insufficient information to identify how the 23 parameters are calibrated in the steady-state model. In addition, only some of the 55 parameters are calibrated in our study, due to different purposes of our study. Specifically, 32 of the 55 parameters are calibrated for DPW1 and DPW2. The two models estimate precipitation (not recharge), which is converted to recharge within DVRFS by dividing the top model layer into five recharge zones. Recharge coefficients in two zones are calibrated against site observations. Since the other three models estimate recharge directly,



recharge coefficients are not used and therefore only 30 parameters are calibrated. All other calibration parameters are the same as those used in DVRFS. Although MLBMA allows different models having different numbers of calibrated parameters, we intend to calibrate the same model parameters for all the recharge models so that model ranking and uncertainty analysis are on the same basis. In the same line, model calibration is conducted in the same manner for all the recharge models. Specifically, all the model calibrations use identical initial parameter values, convergence criterion, and other calibration variables such as parameter log transform and damping factors.

Model calibration results corresponding to the five recharge models are summarized in Table C-2 and Figure C-4. Table C-2 lists WSSR (weighted sum of squared residuals) of the four kinds of observations, respectively, and total WSSR. WSSR of DVRFS is also listed for comparison. The table shows that, except for recharge model ME, the values of WSSR of the models are close and are lower than that of DVRFS. This is not surprising since our calibration is based on model calibration of DVRFS to a certain extent, and can be regarded as further calibration of DVRFS. The largest relative differences of WSSR occur for the observations of discharge and constant-head flow. Figure C-4 plots some of the calibrated parameters whose values are noticeably different between the recharge models (values of other calibrated parameter are close). The values of calibrated parameters in DVRFS are also plotted for comparison. Although WSSR corresponding to the recharge models are similar, some parameter values are different, indicating different responses of the regional flow system simulation to the recharge models, which provides a basis for model discrimination. The largest difference of parameter values occurs for hydraulic conductivity of volcanic rock units (K3) such as K3BRU123 and K3CTM. While most of the parameters are within the parameter ranges given in [5], several parameter values exceed the ranges. This, however, is not surprising, since the ranges are based on limited information of site measurements. All values of the calibrated parameters are considered reasonable.

TABLE C-2. Weighted Sum of Squared Residuals (WSSR) for all Kinds of Observations Corresponding to the Five Recharge Models and DVRFS.

Type of observation	Observation Number	DVRFS	ME	DPW1	DPW2	CMB1	CMB2
Hydraulic head	2227	23083.22	26321.55	20030.92	20296.37	20215.87	19803.57
Head changes	2672	13348.08	11805.63	12599.57	12752.66	12372.07	12057.11
Discharge	49	637.64	2078.36	674.43	611.12	1001.34	1062.06
Constant-head flow	15	438.15	1520.28	296.94	350.56	863.24	641.49
Total	4963	37507.10	41725.82	33601.86	34010.71	34452.52	33564.23

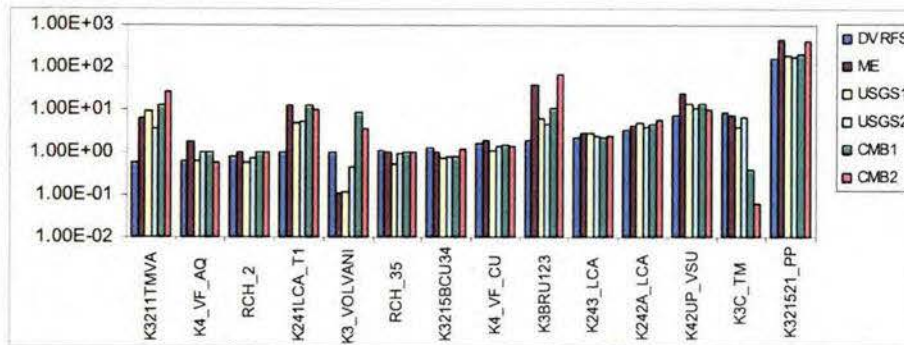


Figure C-4. Comparison of values of some calibrated parameters corresponding to the five recharge models.

#### 4.2 Posterior Model Probability

Posterior model probabilities of the recharge models are calculated using equation (C6) based on the model calibration results listed in Table C-3, which also lists three information criteria (AIC, BIC, and KIC) commonly used to rank alternative models. AIC and BIC are evaluated via

$$AIC_k = -2 \ln p(\mathbf{D} | M_k) + 2N_k = N \ln \sigma_k^2 + N_k \quad (C10)$$

$$BIC_k = -2 \ln p(\mathbf{D} | M_k) + N_k \ln N = N \ln \sigma_k^2 + N_k \ln N \quad (C11)$$

These information criteria rank alternative models not only based on their goodness-of-fit (as measured by WSSR) but also on the principle of parsimony, which states that a simple model (with lower number of parameters) is considered more plausible than a complex model if their predictions fit observations equally well. The three information criteria rank the five recharge models at almost the same order, with models ME, DPW2, and CMB1 ranked as the least plausible. In addition, the information criteria show that DPW1 and CMB2 are more plausible than DPW2 and CMB1, respectively, which is consistent with the results of the expert elicitation (Figure C-3). Nevertheless, model ME is ranked as least plausible after model calibration. AIC and BIC rank CMB2 as the best model, while KIC ranks DPW1 as the best one. Inconsistency of model ranking given by different information criteria is not uncommon. Among the three criteria, KIC is favored since it incorporates quality of data used for model calibration [3] and can yield more reliable model rankings in various circumstances [e.g., 4].

TABLE C-3. Quality Criteria, Ranking, And Prior/Posterior Probabilities Associated With The Five Recharge Models.

	ME	DPW1	DPW2	CMB1	CMB2
$N_k$	30	32	32	30	30
$WSSR$	41726	33602	34011	34453	33564
$\ln F $	360	346	344	349	346
AIC	10627	9556	9616	9676	9547
Rank	5	2	3	4	1
BIC	10822	9764	9824	9871	9742
Rank	5	2	3	4	1

KIC	10808	9718	9775	9852	9720
Rank	5	1	3	4	2
$p(M_k)$	25%	30%	11%	13%	20%
$p(M_k \mathbf{D})$	0	83.16%	0	0	16.84%
$p(M_k)$	20%	20%	20%	20%	20%
$p(M_k \mathbf{D})$	0	76.70%	0	0	23.30%

Posterior model probabilities are evaluated using (C5) for two sets of prior model probabilities. One set has informative priors obtained from the expert elicitation and the other one treats the five models equally likely. Regardless of prior probabilities, posterior probabilities of models ME, DPW2, and CMB2 are zero, indicating that they are implausible given the calibration data. This is so even though model ME received a relatively large prior probability from the experts. The effects of prior on posterior model probability is observed for models DPW1 and CMB2. Posterior probability of DPW1 decreases 6.43% when its prior probability decreases 10%. This results in a concomitant increase of 6.43% in the posterior probability of model CMB2, even though its prior probability does not change. Although it is expected that sensitivity of posterior to prior model probability diminishes as the amount conditioning (calibration) data increases, this study shows that, even with 4,963 observations, sensitivity to prior probability does not disappear. In this case, using informative prior model probability (obtained from expert elicitation in this study) may increase accuracy of model uncertainty assessment, as suggested in [16]. Note that just like prior probabilities, posterior probabilities are valid only in a comparative, not in an absolute, sense. They are conditional on the choice of models, calibration data, and prior information used to estimate prior model probabilities.

## 5. BAYESIAN MODEL AVERAGING

Based on equations (C2) and (C3), Bayesian model averaging is used to yield the posterior mean and variance to incorporate both parametric and conceptual model uncertainty. The posterior mean represents the optimum prediction and the posterior variance measures the associated predictive uncertainty. Monte Carlo simulation is used to assess parametric uncertainty and estimate  $E[\Delta|\mathbf{D}, M_k]$  and  $Var[\Delta|\mathbf{D}, M_k]$  for model  $M_k$ . Multivariate normal distributions are used to generate 200 parameter realizations of the calibrated parameters. The mean of the normal distribution is the maximum likelihood parameter estimate  $\hat{\theta}_k$  of model  $M_k$  and the covariance matrix is  $\sigma^2(\mathbf{X}_k^T \omega \mathbf{X}_k)^{-1}$  [3, 11]. Estimation of posterior mean and variance using equations (C2) and (C3) is straightforward. Figure C-5 plots mean head predictions corresponding to the five recharge models (Figures C-5b – C-5f) and the MLBMA (Figure C-5a) posterior mean head in the first (top) model layer at stress period 87 (1998, the last year of the transient model). The MLBMA mean head (Figure C-5a) is an average of the mean heads for DPW1 and CMB2, since the other three models have zero posterior probabilities. Mean head contours of DPW1 and DPW2 are similar to each other, as are the contours of CMB1 and CMB2, owing to the similarity of the two pairs of recharge models. Since these four recharge models are different from ME, the contour of ME is different from the four contours in Figures C-5c – C-5f. Figure C-6 plots the cumulative distribution function (CDF) of the mean head and standard deviation of head predicted by models DPW1 and CMB2 and MLBMA for the entire simulation domain at stress period 87. For mean head predictions, the CDFs of DPW1, CMB2, and MLBMA are almost identical, due to the similarity of mean head predictions of DPW1 and CMB2 (shown in Figures C-5b and C-5f). Nevertheless, the standard deviation of head

prediction of MLBMA is larger than that of models DPW1 and CMB2, since MLBMA considers both parametric and conceptual model uncertainty, while the two single models address only parametric uncertainty.

## 6. CONCLUSIONS

This study assesses conceptual model uncertainty of five recharge models within the modeling framework of DVRFS, of which each recharge model is a component. Conceptual model uncertainty is first assessed, *a priori*, using expert judgment gathered from an expert elicitation. The experts placed higher probabilities on DPW1 and CMB2 than DPW2 and CMB1, respectively. However, since the recharge models are developed by different researchers based on different theories, prior model probabilities elicited from the experts are around the average value of 20%. This indicates that one cannot select one model for predictions and discard all others *a priori*. Since prior information cannot fully assess conceptual model uncertainty, model calibration is needed to assess conceptual model uncertainty *a posteriori* based on observations of head and flow. DVRFS is used as the framework for numerical modeling and only its recharge component varies for different recharge models (other components remain the same). Based on model calibration results using MODFLOW2000, three information criteria (AIC, BIC, and KIC) are evaluated to rank the models. Model ranking of AIC and BIC are the same, but different from that of KIC. Consistent with results of expert elicitation, DPW1 and CMB2 are ranked more plausible than DPW2 and CMB1, respectively. However, as opposed to the results of expert elicitation, model ME is ranked as least plausible. This suggests the importance of uncertainty assessment *a posteriori*. Posterior model probabilities are evaluated using KIC, which is considered superior to AIC and BIC. Models ME, DPW2, and CMB1 have zero posterior probabilities. Sensitivity of posterior to prior probabilities for models DPW1 and CMB2 does not disappear, although 4,963 observations are used for model calibration. Note that posterior probabilities are valid only in a comparative, not in an absolute, sense. They are conditional on the choice of models, calibration data, and prior information used to estimate prior probabilities. Bayesian model averaging is conducted to estimate posterior mean and variance of head and flux. Posterior variance of MLBMA is larger than the variance of any single model, since conceptual model uncertainty is also addressed. Our research results can be extended to incorporate conceptual model uncertainty in flow path delineation, which can in turn be used to design networks for detection and monitoring of potential radionuclide transport in the saturated zone of the Death Valley Regional Flow System, where Yucca Mountain is located.



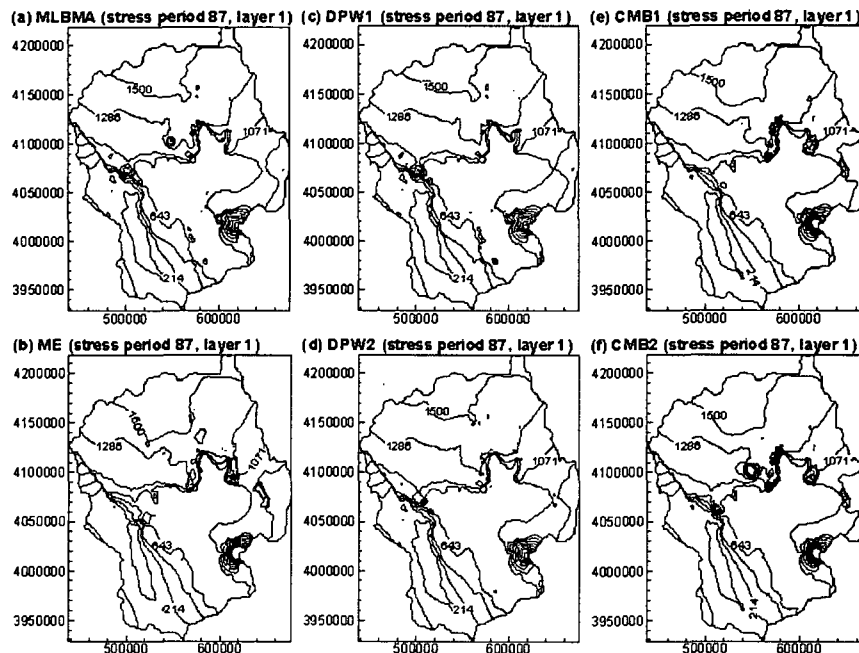


Figure C-5. Mean head predicted by (a) MLBMA, (b) ME, (c) DPW1, (d) DPW2, (e) CMB1, and (f) CMB2 in the first (top) layer at stress period 87 (1998).

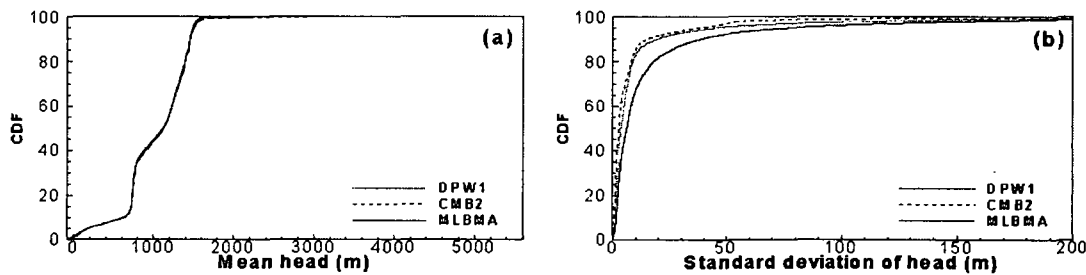


Figure C-6. Cumulative distribution function (CDF) of mean head and standard deviation of head prediction over the whole simulation domain at stress period 87 (1998).

## ACKNOWLEDGMENTS

Research is funded through contract DE-AC52-00NV13609 with the U.S. Department of Energy, National Nuclear Security Administration Nevada Site Office (NNSA/NSO). We appreciate Dr. Glendon Gee, Dr. Greg Pohll, Mr. Chuck Russell, Dr. Joe Hevesi, and Mr. Randy Lacznia for their participation in the expert elicitation.

## REFERENCES

- [1] NATIONAL RESEARCH COUNCIL, "Conceptual Models of Flow and Transport in the Fractured Vadose Zone", National Academy Press, Washington, DC (2001).

- [2] S.P. NEUMAN and P.J. WIERENGA, "A comprehensive strategy of hydrogeologic modeling and uncertainty analysis for nuclear facilities and sites", NUREG/CR-6805, U.S. Nuclear Regulatory Commission, Washington, D.C. (2003).
- [3] S.P. NEUMAN, "Maximum likelihood Bayesian averaging of alternative conceptual-mathematical models, Stochastic Environmental Research and Risk Assessment", 17(5), 291-305, DOI: 10.1007/s00477-003-0151-7 (2003).
- [4] M. YE, S.P. NEUMAN, and P.D. MEYER, "Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff", Water Resour. Res., 40, W05113, doi:10.1029/2003WR002557 (2004).
- [5] W.R. BELCHER (ed). "Death Valley regional ground-water flow system, Nevada and California – Hydrogeologic framework and transient ground-water flow model", U.S. Geological Survey Scientific Investigation Report 2004-5205 (2004).
- [6] G.B. MAXEY and T.E. EAKIN, "Ground water in White River Valley, White Pine, Nye, and Lincoln Counties, Nevada". No. 8, State of Nevada Office of the State Engineer prepared in cooperation with the United State Department of the Interior Geological Survey, Carson City, Nevada (1949).
- [7] J.A. HEVESI, A.L. FLINT, and L.E. FLINT, "Simulation of Net Infiltration and Potential Recharge Using a Distributed Parameter Watershed Model for the Death Valley Region, Nevada and California". Water-Resources, Investigations Report 03-4090. Sacramento, CA: U.S. Geological Survey (2003).
- [8] C.E. RUSSELL, and T. MINOR, "Reconnaissance Estimates of Recharge Based on an Elevation-dependent Chloride Mass-balance Approach", DOE/NV/11508-37. Prepared for the U.S. Department of Energy, National Nuclear Security Administration Nevada Operations Office. Las Vegas, NV: Desert Research Institute (2002).
- [9] K. REHFELDT (ed), "Hydrologic Data For The Groundwater Flow And Contaminant Transport Model Of Corrective Action Units 101 And 102: Central And Western Pahute Mesa, Nye County, Nevada", Stoller-Navarro Joint Venture (2004).
- [10] R.L KASHYAP., "Optimal choice of AR and MA parts in autoregressive moving average models", IEEE Trans. Pattern Anal. Mach. Intel. PAMI, 4(2), 99-104 (1982).
- [11] M.C. HILL, E.R. BANTA, A.W. HARBAUGH, and E.R. ANDERMAN, MODFLOW-2000, "The U.S. Geological Survey modular groundwater model – User guide to the observation, sensitivity, and parameter estimation processes and three post-processing programs", USGS Open-File Report 00-184, Reston, Virginia: USGS (2000).
- [12] H. JEFFREYS, "*Scientific Inference*" (2<sup>nd</sup> ed.), Cambridge, U.K., Cambridge University Press, (1957).
- [13] E. ZIO, and G.E. APOSTOLAKIS, "Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive waste repositories", *Reliability Engineering and System Safety*, 54, 225-241 (1996).
- [14] J. BREDEHOEFT, "The conceptualization model problem-surprise", *Hydrogeol. J.*, 13, 37-46 (2005).
- [15] R.L. KEENEY AND D. VON WINTERFELDT, "Eliciting probabilities from experts in complex technical problems", *IEEE Transactions On Engineering Management*, 38 (3): 191-201 (1991).
- [16] M. YE, S. P. NEUMAN, P. D. MEYER, and K. POHLMANN (2005), "Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff", *Water Resour. Res.*, 41, W12429, doi:10.1029/2005WR004260.

<b>NRC FORM 335</b> (9-2004) NRCMD 3.7		<b>U.S. NUCLEAR REGULATORY COMMISSION</b>		<b>1. REPORT NUMBER</b> (Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)  NUREG/CR-6940 PNNL-16396										
<b>BIBLIOGRAPHIC DATA SHEET</b> (See instructions on the reverse)														
<b>2. TITLE AND SUBTITLE</b>  Combined Estimation of Hydrogeologic Conceptual Model, Parameter, and Scenario Uncertainty with Application to Uranium Transport at the Hanford Site 300 Area				<b>3. DATE REPORT PUBLISHED</b>										
				MONTH  July	YEAR  2007									
				<b>4. FIN OR GRANT NUMBER</b>  Y6465										
<b>5. AUTHOR(S)</b>  P.D. Meyer (PNNL) , M.Ye (DRI), M.L. Rockhold (PNNL), S.P. Neuman (UA), and K.J. Cantrell (PNNL) (PNNL - Pacific Northwest National Laboratory, DRI - Desert Research Institute, UA - University of Arizona)				<b>6. TYPE OF REPORT</b>  Technical										
				<b>7. PERIOD COVERED</b> (Inclusive Dates)  August 2002 - February 2007										
<b>8. PERFORMING ORGANIZATION - NAME AND ADDRESS</b> (If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)  <table border="0" style="width: 100%;"> <tr> <td>Pacific Northwest National Laboratory</td> <td>Desert Research Institute</td> <td>University of Arizona</td> </tr> <tr> <td>P.O. Box 999</td> <td>Las Vegas, NV 89119</td> <td>Tucson, AZ 85721</td> </tr> <tr> <td>Richland, WA 99352</td> <td></td> <td></td> </tr> </table>						Pacific Northwest National Laboratory	Desert Research Institute	University of Arizona	P.O. Box 999	Las Vegas, NV 89119	Tucson, AZ 85721	Richland, WA 99352		
Pacific Northwest National Laboratory	Desert Research Institute	University of Arizona												
P.O. Box 999	Las Vegas, NV 89119	Tucson, AZ 85721												
Richland, WA 99352														
<b>9. SPONSORING ORGANIZATION - NAME AND ADDRESS</b> (If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)  Division of Fuel, Engineering and Radiological Research Office of Nuclear Regulatory Research U.S. Nuclear Regulatory Commission Washington, DC 20555-0001														
<b>10. SUPPLEMENTARY NOTES</b>  T.J. Nicholson, NRC Project Manager														
<b>11. ABSTRACT</b> (200 words or less)  A methodology was developed and applied to systematically and quantitatively assess predictive uncertainty in ground-water flow and transport modeling. The methodology considers the combined impact of hydrogeologic uncertainties associated with the conceptual-mathematical basis of a model, model parameters, and the scenario to which the model is applied. The methodology is based on an extension of a Maximum Likelihood implementation of Bayesian Model Averaging. Model uncertainty is represented by postulating a discrete set of alternative conceptual models for a site with associated prior model probabilities. The prior model probabilities reflect a subjective belief about the relative plausibility of each model based on its apparent consistency with available knowledge and data. Posterior model probabilities are computed and parameter uncertainty is estimated by calibrating each model to observed system behavior. Scenario uncertainty is represented as a discrete set of alternative future conditions affecting boundary conditions, source/sink terms, or other aspects of the models. A joint assessment of uncertainty results from combining model predictions computed under each scenario using as weights, the posterior model and prior scenario probabilities. The uncertainty methodology was applied to modeling of ground-water flow and uranium transport at the Hanford Site 300-Area. Results demonstrate the feasibility of applying a comprehensive uncertainty assessment to large-scale, detailed ground-water flow and transport modeling. Results also illustrate the ability of the methodology to provide better estimates of predictive uncertainty, quantitative results for use in assessing risk, and an improved understanding of the system behavior and the limitations of the models.														
<b>12. KEY WORDS/DESCRIPTORS</b> (List words or phrases that will assist researchers in locating the report.)  Bayesian Model Averaging dose assessment ground-water modeling hydrogeologic conceptual model model calibration model uncertainty parameter uncertainty scenario uncertainty uncertainty				<b>13. AVAILABILITY STATEMENT</b>  unlimited										
				<b>14. SECURITY CLASSIFICATION</b> (This Page)  unclassified										
				(This Report)  unclassified										
				<b>15. NUMBER OF PAGES</b>										
				<b>16. PRICE</b>										



Federal Recycling Program