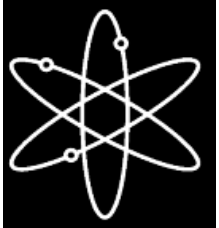
NUREG/CR-6805



A Comprehensive Strategy of Hydrogeologic Modeling and Uncertainty Analysis for Nuclear Facilities and Sites

University of Arizona

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

NUREG/CR-6805

A Comprehensive Strategy of Hydrogeologic Modeling and Uncertainty Analysis for Nuclear Facilities and Sites

Manuscript Completed: December 2002
Date Published: July 2003

Prepared by
S.P. Neuman and P.J. Wierenga

University of Arizona
Tucson, AZ 85721

T.J. Nicholson, NRC Project Manager

Prepared for
Division of Systems Analysis and Regulatory Effectiveness
Office of Nuclear Regulatory Research
U. S. Nuclear Regulatory Commission
Washington, DC 20555-0001
NRC Job Code W6790

ABSTRACT

This report describes a strategy that embodies a systematic and comprehensive approach to hydrogeologic conceptualization, model development and predictive uncertainty analysis. The strategy is comprehensive in that it considers all stages of model building and accounts jointly for uncertainties that arise at each of them. The stages include regional and site characterization, hydrogeologic conceptualization, development of conceptual-mathematical model structure, parameter estimation on the basis of monitored system behavior, and assessment of predictive uncertainty. In addition to parameter uncertainty, the strategy concerns itself with uncertainties arising from incomplete definitions of (a) the conceptual framework that determines model structure, (b) spatial and temporal variations in hydrologic variables that are either not fully captured by the available data or not fully resolved by the model, and (c) the scaling behavior of hydrogeologic variables. The strategy is generic but designed to be of practical use to NRC licensing staff in their review of decommissioning plans, and performance assessment of high-level and low-level radioactive waste disposal sites as well as uranium recovery facilities. An important component of the strategy is a systematic sequence of logical questions, guidelines and criteria with analytical methods appropriate for NRC review and performance evaluation.

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

Appendices

Figures

Figures (continued)

Figures (continued)

Figures (continued)

Tables

EXECUTIVE SUMMARY

This report describes a strategy that embodies a systematic and comprehensive approach to hydrogeologic conceptualization, model development and predictive uncertainty analysis. The strategy is comprehensive in that it considers all stages of model building and accounts jointly for uncertainties that arise at each of them. The stages include regional and site characterization, hydrogeologic conceptualization, development of conceptual-mathematical model structure, parameter estimation on the basis of monitored system behavior, and assessment of predictive uncertainty. In addition to parameter uncertainty, the strategy concerns itself with uncertainties arising from incomplete definitions of (a) the conceptual framework that determines model structure, (b) spatial and temporal variations in hydrologic variables that are either not fully captured by the available data or not fully resolved by the model, and (c) the scaling behavior of hydrogeologic variables.

The strategy is generic but designed to be of practical use to NRC licensing staff in their review of decommissioning plans and performance assessment of high-level and low-level radioactive waste disposal sites as well as uranium recovery facilities. For this purpose, the strategy is cast in the context of a framework that is useful to NRC staff review and performance evaluation needs. The context is defined in terms of corresponding performance measures (as identified by the NRC staff in Appendix A), hydrogeologic analyses that are needed to assess them, the desired reliability of such assessments, and the expenditure (in time, effort and money) that is allowed to achieve it. Examples from various case studies are included to help illustrate some aspects of these analyses. An important component of the strategy is a systematic sequence of logical questions, guidelines and criteria with analytical methods useful for NRC staff review and performance evaluation.

The strategy encourages exploration of varied conceptual frameworks and assumptions at all stages of hydrogeologic model development through a comprehensive evaluation of a broad range of regional and site data, their translation into coherent and internally consistent conceptual-mathematical models, and computation and visualization based on these data and models. Included among these frameworks and assumptions are various model simplification and abstraction schemes. The strategy recognizes that site characterization and monitoring data are expensive and difficult to collect, leading to a ubiquitous scarcity of hard site information. It is therefore critically important to assess the role that such data play in rendering the hydrogeologic performance analysis credible. The strategy stresses the role of characterization and monitoring data in helping one identify and test alternative conceptual models, make rational choices among them, gauge and reduce model bias and uncertainty through proper model selection and calibration, assess the reliability of model predictions, and confirm the assessment through independent peer review as well as at least some degree of direct verification.

The strategy encourages an iterative approach to modeling, whereby a preliminary conceptual-mathematical model is gradually altered and/or refined until one or more likely alternatives have been identified and analyzed. The idea is to start with a conceptual-mathematical model that is as simple as may appear warranted by the

context of the problem and the available data. This model is then tested for qualitative consistency with the available data, made quantitatively compatible with the data through calibration (parameter estimation via an inverse solution), and subjected to an assessment of its predictive reliability (in terms of potential bias and uncertainty). Next, the model structure is altered and/or refined (by modifying its hydrogeologic makeup or features; dimensionality; scale of resolution in space-time; type of driving forces or events including sources, initial and boundary conditions; and governing equations or processes), the model is retested and compared with the former.

The iterative procedure is repeated until there is good reason to conclude that any additional alterations or refinements of the models would lead at best to minor improvements in their predictive capability. Models whose predictive capability is deemed low in comparison to other are discarded and the rest retained for further analysis. The retained models are analyzed jointly to assess their aggregate predictive bias and uncertainty. If the latter are deemed acceptable for the stated purpose of the analysis, the iterative process is halted and the models declared reliable and credible within a clearly stated conditional (on the available information and selected set of models) margin of error. The conditional nature of this error estimate is taken to imply that it constitutes a lower bound, and that the actual predictive error may (and most probably is) larger to an unknown degree.

If the models are deemed unacceptable, a sensitivity analysis is performed to help identify the type and quantity of additional site data that might materially enhance their reliability and credibility. A decision is then made, based on the potential benefit (in terms of bias and uncertainty reduction) and cost (in terms of time, effort and money) of such data, whether or not to collect them and how. If and when new data of significant weight are obtained, the iterative process is repeated till its cost-benefit ratio reaches a value that is not considered worth exceeding.

The strategy recognizes that it is often possible to postulate hydrogeologic conceptual models or hypotheses for a site on the basis of publicly available geologic and geographic information about its surroundings. Additional conceptualization can be done on the basis of generic data about similar regions and the properties of similar materials elsewhere. Several such regional and generic sources of information are identified and discussed in this report. Yet each site is unique and so virtually guaranteed to reveal additional features, properties and behaviors when characterized in some detail locally. Hence the strategy considers local characterization essential for the postulation of acceptably robust conceptual hydrogeologic models for a site.

Regional and site characterization data tend to represent a wide range of measurement scales, not all of which are compatible with the intended scale of hydrogeologic model resolution. The strategy promotes recognition of this important issue and an effort to render the scale of measurement compatible with the scale of model resolution. This can be done by either rescaling the data to fit the scale of model resolution (which often entails averaging or upscaling over computational grid cells) or adapting model resolution to fit the scale of measurement (which often entails adapting the size of grid cells to the size of the data support). Recent advances in scaling theory and practice are briefly reviewed to promote

awareness of fundamental issues associated with the scaling of hydrogeologic variables.

The strategy encourages the analyst to associate model parameters, derived from regional and site characterization data, with statistical measures of their uncertainty. Such prior statistics provide the only way to assess model predictive uncertainty when suitable monitoring data are not available. Methods are discussed and an example given to derive such statistics from generic data sources on the basis of similarity in soil or rock textural features; to update these statistics on the basis of site-specific data using a Bayesian approach; and to interpolate, extrapolate and/or average parameters across the site geostatistically on the basis of measured values at discrete points in space.

While hydrogeologic characterization of a site and its surroundings makes conceptualization possible, it does not provide the means to test conceptual models or compare them with other alternatives. For this, it is necessary to have monitoring data that constitute observations of actual hydrologic behavior at and around the site. Only with such data can one evaluate the ability of models to mimic real system behavior (qualitatively at the conceptual level, quantitatively at the conceptual-mathematical level), improve their ability to do so through calibration against the monitoring data, determine their optimum degree of refinement or complexity, and compare them with each other (qualitatively and quantitatively). The strategy discusses ways to accomplish these tasks in the context of traditional deterministic groundwater flow and transport models.

The strategy encourages the analyst to employ statistical measures of model performance (or fit between simulated and observed hydrologic behaviors) where possible; to determine feasible tradeoffs between model fit and complexity by relying on multiobjective approaches; and to rank models based on (a) statistical indicators of the quality of fit they produce (as obtained from an analysis of residuals), and (b) likelihood-based model discrimination criteria (which penalize models for using an excessive number of free parameters to achieve a given quality of fit). Models that show relatively poor fits and rank low on the list can be subjectively eliminated from further consideration at this stage.

An important byproduct of model calibration is information about parameter uncertainty. When reliable and relevant (in terms of type and scale) prior statistics about the parameters are available from site characterization, the strategy encourages their use as input into the calibration process. Regardless of whether or not suitable prior statistics are available and employed for model calibration, the latter should produce posterior statistics for the optimum parameter estimates. The strategy considers the posterior parameter estimates, and their error statistics, more suitable for the analysis of model predictive uncertainty than the prior estimates and their statistics. This is so because calibration improves the ability of the model to reproduce simulated hydrologic behavior, and a calibrated model is therefore deemed more reliable and credible as a predictive tool than an uncalibrated model. The same holds true for the associated parameter estimation statistics.

The strategy considers two alternative methods to assess the predictive uncertainty of a deterministic groundwater model under assumed future scenarios. The first method relies on Monte Carlo simulation of these scenarios using either prior or posterior

parameter estimates and statistics. The second method establishes approximate error bounds, or confidence limits, for such scenarios by linearization, using the same parameter estimates and statistics. The first approach is more accurate but costly in terms of computer time and effort. In both cases, the strategy favors the use of posterior estimates over that of priors on the understanding that the former are less biased and uncertain than the latter.

To render optimum predictions by means of all calibrated models that have been retained as constituting potentially viable alternatives, and to assess their joint predictive uncertainty, the strategy relies on a newly developed Maximum Likelihood Bayesian Model Averaging approach, MLBMA. The final outcome is conditional on the choice of models and data. It accounts jointly for uncertainties in the conceptual-mathematical model structure and its parameters.

If suitable monitoring data are not available, then there is neither an objective nor a quantitative basis for the comparison of potentially viable conceptual-mathematical models. To quantify the predictive uncertainty of each such model, one can either weigh all models equally or associate them with subjective weights based on the professional judgment of the analyst or a group of experts.

To help evaluate what if any additional data might be worth collecting so as to materially reduce model uncertainty (by further constraining the range of alternative structures and parameters), the strategy suggests conducting a sensitivity analysis to indicate what system behavior appears to be most sensitive to which parameters at what locations. The next step is to consider performing additional site characterization where existing parameter estimates are least certain and the model is relatively insensitive to their values, and monitoring system behavior where it is most sensitive to model parameters while prediction errors appear to be relatively large and consequential.

Calibration of a deterministic groundwater model yields effective parameter estimates, which compensate to some degree for lack of knowledge about variability on space-time scales that are not resolved by the data or the model. This is less true for transport than for flow, because transport processes due to unresolved variations in permeability and porosity cannot be validly represented by traditional effective parameters (such as a constant dispersivity) except in special cases or as a crude approximation. The reliance on effective parameters is also less valid for strongly heterogeneous media with preferential flow paths such as some fractured rocks and soils that contain relatively large and elongated openings created by burrowing animals or plant roots. When the deterministic effective parameter approach fails, the most common and powerful alternative is to employ stochastic concepts and models.

A stochastic approach to groundwater modeling requires that the spatial variability of input parameters, such as permeability and unsaturated soil properties, be characterized with the aid of geostatistical methods. A key component of geostatistical characterization is the assessment of spatial covariance structure (in terms of a covariance function or variogram). The strategy promotes this approach and the report illustrates it by examples. In the absence of site characterization data that are amenable to geostatistical analysis, one can treat the structural parameters (of the variogram or covariance function) as free

parameters to be estimated by model calibration. This too is illustrated by example in the report.

The strategy supports two general approaches to predict stochastic groundwater flow and transport on the basis of geostatistical input: High-resolution numerical Monte Carlo simulation and direct (deterministic) prediction of mean behavior. The latter approach relies on stochastically derived deterministic ensemble mean flow and transport equations in which the dependent variables represent not actual system states (such as head, concentration or flux) but rather their (ensemble) mean values or statistical expectations. These mean or expected values represent optimum unbiased predictors of the unknown, actual system states. Similar values are obtained upon averaging the results of numerous Monte Carlo simulations. Both the Monte Carlo and the mean equation approaches allow the reduction of predictive uncertainty by conditioning the predictions on actual measurements of the input parameters at discrete locations in space.

The variance-covariance of high-resolution (conditional) Monte Carlo simulations serves to quantify predictive uncertainty. An alternative is to compute this variance-covariance directly (deterministically) by means of corresponding moment equations. Both options are supported by the strategy.

Conditioning the stochastic flow and transport equations not only on measured input variables (derived from site characterization) but also on monitored system behavior tends to improve their reliability. This is accomplished through novel inverse procedures, which form part of the proposed strategy.
In virtually all other respects, the strategy is the same for stochastic as for deterministic models of groundwater flow and transport.

FOREWORD

This technical report was prepared by researchers at the University of Arizona (UAZ) under a contract (NRC-04-97-056) with NRC's Office of Nuclear Regulatory Research. The research objective was to develop a methodology with supporting technical bases and guidance for selecting and evaluating appropriate ground-water flow and transport models for use in simulating performance of nuclear waste disposal facilities and possible remediation actions at decommissioning sites. The technical focus was on identifying and quantifying sources of uncertainty in ground-water models as related to specific performance measures defined by NRC regulations and guidance. These performance measures include well-discharge concentrations, expected values or potential maximum values of point concentrations, and dose from drinking water. The focus was not on computer code development, but rather on how to evaluate and compare ground-water models beginning with selection of alternative conceptual models and ending in performance assessment of the nuclear facility and site. Use of this research to support regulatory decisions will contribute to the NRC Strategic Plan's performance goal of making NRC activities and decisions more effective, efficient and realistic.

This report provides a systematic strategy for evaluating site-specific ground-water models and their attendant uncertainties. Specifically the strategy identifies tools and analysis methods for interpreting hydrologic and environmental data important to formulating alternative conceptual models and in estimating their parameter values and modeling assumptions. The report presents approaches for quantitative comparison and ranking of alternative models with field examples. It also addresses fundamental questions related to assessing model predictive uncertainty. Finally, the report considers important questions on model confirmation and assessment of data needs (e.g., when shall data collection end?). Many of these technical questions and the related narratives were developed through discussions between the UAZ researchers and the NRC licensing staff who are seeking ways to reduce the regulatory burden through more realistic modeling.

This report serves as a technology transfer document to the NRC staff for their consultative use in reviewing ground-water models. The implementation of the strategy and its research findings provide a framework for more realistic modeling. Earlier drafts of this report were used in training workshops presented at NRC Headquarters on August 14-15, 2001 and August 1-2, 2002. The report was also circulated to cooperating federal agencies working under a Memorandum of Understanding on research into multi-media environmental models. The report was peer-reviewed by Pacific Northwest National Laboratory investigators who developed a companion hydrologic parameter uncertainty methodology.

Cheryl A. Trottier, Chief
Radiation Protection, Environmental Risk and Waste Management Branch
Division of Systems Analysis and Regulatory Effectiveness
Office of Nuclear Regulatory Research

ACKNOWLEDGMENTS

This work was supported by the U.S. Nuclear Regulatory Commission (NRC) under contract NRC-04-97-056. Special credit is due to Thomas J. Nicholson and Ralph Cady of the NRC for having conceived and championed this research.

The authors acknowledge with thanks contributions by the following individuals: Mr. Tim McCartin of the U.S. NRC for his recent update of Appendix A; Professor Ghislain de Marsily of the University of Paris, France, for having provided data concerning the Fanay-Augères experimental site; Dr. Weihsueh Chiu of the U.S. Environmental Protection Agency for having brought to our attention the extensive work on BMA conducted by the "Seattle school" of statistics at the University of Washington, and for his helpful comments on Appendix E; Drs. Philip D. Meyer and Mark L. Rockhold of Pacific Northwest National Laboratory for their formal review of a draft version of this manuscript.

1 INTRODUCTION

1.1 Background

This report describes a comprehensive strategy for constructing hydrogeologic flow and transport models and assessing their predictive uncertainty. The emphasis is on hydrogeologic conceptualization and its effect on the reliability of predictions relevant to decommissioning reviews and performance assessment (PA) of high-level (HLW) and low-level (LLW) radioactive waste disposal sites, as well as uranium recovery facilities.

The strategy is the outcome of a study commissioned in 1997 by the U.S. Nuclear Regulatory Commission (NRC). Its purpose is to address a concern on the part of the NRC staff that conceptual models of site hydrogeology, and factors (physico-chemical processes, parameters, forcing functions representing sources and initial/boundary conditions) which control subsurface flow and transport, constitute a major source of uncertainty in assessing the expected performance of such sites. The stated objective of the study was to develop a strategy, with supporting technical bases and guidance, for selecting and evaluating appropriate groundwater flow and transport models and assessing uncertainties of related performance measures.

Some of the questions addressed in the report have been the subject of two round-table discussions held at a Workshop on Ground-Water Modeling Related to Dose Assessments on June 23–24, 1999, at NRC Headquarters in Rockville, Maryland.
The first round-table discussion, devoted to Conceptual Model Uncertainties, focused on the following questions: What is a conceptual model? Through what process of information gathering and analysis does one develop and identify a conceptual model? What are typical examples of conceptual models used in groundwater flow and contaminant transport analyses? How important are the assumptions dealing with parameter representation of heterogeneities, dimensionality and scaling factors? What role do conceptual models have in performance assessments? How does one identify alternative conceptual models for a specific site and specified performance measure(s)? How does one quantify differences in the competing conceptual models? Are these differences significant to the performance measure of interest (e.g., well-discharge concentrations, point concentrations of expected values or potential maximum values, drinking-water-related dose? How does one quantify these performance-measure uncertainties for site specific hydrogeologic systems? How important is the issue of model confirmation for conceptual models?

The second round-table discussion dealt with Parameter Uncertainties. It asked: How are conceptual model and parameter uncertainties related? What statistical methods are appropriate for evaluating parameter uncertainty? How can the following topics be handled in a parameter uncertainty strategy: (a) representation heterogeneities; (b) scaling of representative parameter values; (c) assumptions of dimensionality, transient versus steady state flow and transport, and initial and boundary conditions; and (d) model confirmation using specific examples such as the INTRAVAL Project test cases, the decommissioning reviews, and *in situ* leach uranium extraction license applications? What reasonable expectations can be met in quantifying parameter uncertainty in the next 2 years?

The strategy described in this report addresses and answers many of the above questions.

1.2 State-of-the-Art Summary

1.2.1 Preliminary Assessment

Prior to developing the strategy described in this report, a study was conducted to identify information needs and objectives for the evaluation of hydrogeologic flow and transport models relevant to the PA of HLW and LLW radioactive waste disposal sites, decommissioning reviews, and uranium recovery facilities. The study included an assessment of hydrogeologic issues common to performance analyses of nuclear facilities and sites; a review and summary of approaches currently used to identify conceptual, mathematical and computational frameworks for the hydrogeologic flow and transport modeling of such facilities and sites; a discussion of uncertainties that may arise in the context of such framework formulations; a review and summary of approaches currently used to assess these uncertainties; and a preliminary identification of ways to extend and/or modify existing approaches so as to achieve the objectives of the contract. The review included technical literature that focuses on relations among site characterization, monitoring and modeling strategies related to constructing and applying groundwater flow and transport models, and their application to site specific information and databases.

1.2.2 Key Conclusions from Preliminary State-of-the-Art Assessment

The preliminary state-of-the-art assessment led to the following interim conclusions:

- Analyses of groundwater flow and transport typically rely on a single conceptual model of site hydrogeology. This exposes such analyses to Type I model errors, which arise when one rejects (in this case by omission) valid alternative models.
- Critiques of groundwater flow and transport analyses, and legal challenges to such analyses, typically focus on the validity of the underlying conceptual model. By adopting (not rejecting) an invalid conceptual framework, one commits a Type II model error. If exposed to be severe, it may damage one's professional credibility and result in the loss of a legal contest.
- Analyses of model uncertainty based on a single hydrogeologic concept are prone to statistical bias (by committing a Type II error through reliance on an invalid model) and underestimation of uncertainty (by committing a Type I error through under-sampling of the relevant model space).
- The bias and uncertainty that result from reliance on an inadequate conceptual model are typically much larger than those introduced through an inadequate choice of model parameter values. Yet most uncertainty analyses of flow and transport ignore the former and focus exclusively on the latter. This often leads to overconfidence in the predictive capabilities of the model, for which there is no justification.

- Each site is unique and there is no valid way to represent its hydrogeology by means of a generic conceptual model or set of parameters, regardless of purpose.
- Whereas simplicity may at times be desirable and justified, there is no guarantee that simple models of hydrogeology would necessarily yield conservative assessment of system performance; the opposite was found to be true in virtually all cases examined by the report.
- There is no established literature on ways to construct alternative conceptual models of site hydrogeology or methods to assess related uncertainties in groundwater flow and transport analyses.

1.2.3 Uniqueness of Sites versus Non-Uniqueness of Models

The idea that each site is unique and there is no valid way to represent its hydrology by means of a generic conceptual model or set of parameters has found eloquent support in a recent article be Beven (2000). According to him, 350 years of scientific hydrology have established some general hydrologic principles at all scales from single pore to continental river basins. Why then are hydrologic predictions, which are so important in water resources management and pollution control, so uncertain? Why does it appear so difficult to use our understanding, as embodied in the form of theories and models, in practice? The primary reason is that practicing hydrologists cannot deal in generalities but must deal with specific instances of sites that are unique in their characteristics over periods that are unique in their atmospheric and anthropogenic forcing.

According to Beven, the latter has not been considered as a problem in the past. The attitude has been that if the modeling concepts were sufficiently accurate then it would be possible to represent the uniqueness of individual sites by specific values of model parameters. In principle this might be possible; in practice it proves to be untenable because of the impossibility of fully describing the system, and particularly the heterogeneity of its subsurface characteristics.

In Beven's view, experience with predictive models in a wide variety of hydrological contexts suggests that, given the data available in even intensively studied locations, there is generally a very wide range of models and parameter sets within models that will yield acceptable simulations. Beven and Freer (2001) have attributed to Hornberger and Speer (1981) the notion that this is not simply a problem of identifying a correct or optimal model given limited data. Beven (1993) has called this the *equifinality* problem to stress that this is a generic problem in modeling complex hydrologic systems, and not simply a problem of identifying the "true" model structure or parameter set. According to Beven (2000), the equifinality problem arises from the limitations of current model structures in representing heterogeneous surface and subsurface flow systems, from limitations of measurement techniques and scales in defining system characteristics and the initial and boundary conditions for a model, and from the uniqueness of individual sites.

Beven (2000) opines that in recent years there has been a move towards more and more detailed studies of flow in rivers, soils and aquifers, using more and more sophisticated instrumentation. The result has

generally been to demonstrate complexity of flow pathways, due to heterogeneity on different scales and interactions between the geometry of the flow domain and the prevailing hydraulic gradients and flow pathways (including both "dead zones" and "preferential" flows). It has proven difficult to use such information in applications of predictive models which are generally required to be used at scales larger than the detailed measurements.

The problem is much less one of understanding than of uniqueness of place. It is not so much that the complexity of the measurements is not accessible to adequate interpretation (or even in some cases to being adequately modeled). The problem is that the measurements are specific to the study site with little or no guarantee that they can be considered representative even of sites that are hydrologically similar (in climate, physiography, ecology and geology). But to do detailed measurements throughout an area of interest is both impractical and unfeasibly expensive. The unique characteristics of a site of interest are therefore inherently unknowable. All that can be done is to constrain the model representations of the site to those that are acceptably realistic, usually in the sense of being consistent with the data.

Given limited measurements on a hydrologic system, a uniquely correct model would not be identifiable even if one had a perfect model available. The current generation of model structures, while useful, may not be sufficiently accurate representations of the complexities of the flow pathways that are easy to perceive and understand but difficult to represent in mathematical form without making grossly simplifying assumptions. This implies that there may also then be many different model structures that might be consistent with the available observational data. Likewise, it will always be impossible to have sufficient information to identify a uniquely correct set of model parameters for any given model structure. We should therefore ***expect*** equifinality of model structures and parameter sets in the representation of unique places and hydrologic systems.

In Beven's view, existing models are useful only because of a form of circular reasoning called model calibration. Hydrologists are prepared to accept that a model is adequate if it gives reasonable predictions of the available observations. Because of uniqueness of place, one does not expect that a model will provide reasonable predictions with only *a priori* estimates of parameter values (why should it?). Some form of calibration is therefore necessary. Such models may be more useful in allowing a decision to be made than in providing accurate predictions.

1.2.4 Desired Level of Model Complexity

The degree of complexity that a groundwater model should incorporate was the subject of a special session, "Groundwater models: How much complexity is warranted?" at the 1998 Spring Meeting of the American Geophysical Union (Hunt and Zheng, 1999). The central questions addressed at the session were: We know that the natural world is characterized by complex and interrelated processes. Yet do we need to explicitly incorporate these intricacies and to perform the tasks we are charged with? In this era of expanding computer power and development of sophisticated preprocessors and postprocessors, are bigger and better machines making better models?

A common concern expressed at the special AGU session was that the extensive data required to construct a complex groundwater model may result in less time spent understanding the system and more time spent constructing and managing data input and output. It was suggested that careful quality assurance and quality control are required when the input data are large, and that peer review of modeling and data acquisition activities may be important in the regulatory context. On one hand, the opinion was voiced that contaminant transport models do not necessarily enjoy greater success as predictive tools with the inclusion of increasing levels of complexity; one should take models for what they are, namely, powerful heuristic tools with limited predictive capabilities. On the other hand, it was recognized that to capture certain flow and transport phenomena, models of appropriate complexity may be required. The degree of complexity should thus be in part a function of its intended use. Fractal scaling and stochastic methods were recognized as capable of capturing flow and transport behaviors in complex hydrogeologic systems by means of relatively simple models.

1.2.5 Traditional Approach to Model Uncertainty Analysis

The traditional approach to hydrologic model uncertainty analysis has been to postulate a deterministic model structure and treat its parameters as being imperfectly known. To quantify this imperfect knowledge, one must postulate a *prior* parameter uncertainty model.

In cases where a statistically significant set of measured, site specific parameter values is available, one could postulate a Type A probabilistic model of prior parameter uncertainty based on statistics derived from these data. Ideally, the data would include measurements of the parameters proper on spatial and temporal scales compatible with the scale of resolution to be adopted in the intended model. The data would also include probabilistic information about errors of measurement and test interpretation that suffer from a known amount (ideally zero) of statistical bias.

In cases where no such data are available in statistically significant quantities, one has the option of postulating a Type B model of prior parameter uncertainty on the basis of subjective probabilities. Such a model should always be suspected of suffering from an unknown amount of statistical and personal bias. Statistical bias is introduced due to lack of site specific information about mean values of the parameters in question. A personal bias tends to manifest itself in the form of assigned uncertainty measures (most importantly bias and error variance) that are either too small or too large. The first is a manifestation of over-confidence in the model parameters, the second of unduly low confidence in their values.

Intermediate between Type A and Type B parameter uncertainty models is the case where indirect information about the parameters is available, from which relevant prior statistics can be derived formally. Such information may include (a) off-site measurements of the parameters proper (quite often on scales other than those corresponding to the intended scale of model resolution) and/or (b) surrogate measurements on site that are known to correlate with the parameters of interest (for example, porosities or geophysical signatures that correlate in known ways with permeabilities, water contents or fracture densities). Statistics derived from off-site

data must be considered potentially biased (due to a lack of site-specific information about mean parameter values and incompatibility of geology and scale). The associated variance may be too small or too large, depending on the quantity and quality of such data. Statistics derived from surrogate data may suffer from poorly defined correlations and incompatibility of scale.

The traditional approach to reduce parameter bias and uncertainty has been to calibrate the model against monitored observations of hydrologic system behavior. The process is also known as parameter identification or estimation, inverse solution or history matching. Its goal is to render the model compatible with available monitoring data and thereby increase the chance that it would provide reliable predictions of hydrologic behavior in the future.

The last thirty years have seen major advances in the development of theories and algorithms for the estimation of deterministic model parameters. Many (though not all) of these theories and algorithms are "statistical" in that they include analyses of parameter estimation uncertainty. Such analyses typically accept, but do not necessarily require, information about prior parameter statistics as input. The output includes *posterior* statistics of parameter estimation errors, which are generally less biased and smaller than the prior estimation errors. A recent summary and comparison of various statistical inverse methods for groundwater flow models has been published by Zimmerman *et al.* (1998). Public domain or commercially available inverse codes include those developed by Poeter and Hill (1998), Doherty *et al.* (1994), Finsterle (1999a-c) and Simunek *et al.* (1999). A detailed set of guidelines for the effective calibration of deterministic groundwater flow models has been prepared by Hill (1998).

Bias and uncertainty in model parameters translate into bias and uncertainty in hydrologic model predictions. To assess the latter *a priori*, it is necessary to "propagate" the former through the model under assumed future scenarios. This is in contrast to *a posteriori* assessment on the basis of so-called post audits, in which the model predictions are compared with monitored future behavior of the hydrologic system. If the *post facto* comparison is deemed successful, the model is considered "confirmed" or "validated" with reference to the available monitoring data.

The most common way to propagate input errors through an otherwise deterministic model is by means of Monte Carlo simulations. This is done by generating multiple, equally likely sets of randomized inputs; computing deterministically a set of corresponding model outputs for each; and analyzing the resultant multiple, equally likely random output sets statistically. This approach is common in the performance assessment of nuclear facilities and sites, in which prior input statistics of Type B are often used due to insufficiency of site specific parameter and monitoring data. Another approach is to associate the model predictions with approximate error bounds, or confidence limits, computed on the basis of linear regression theory applied to the (typically nonlinear) groundwater inverse model (Poeter and Hill, 1998; Hill, 1998).

1.2.6 Accounting for Conceptual Model Uncertainty

Carrera and Neuman (1986a-b) and Samper and Neuman (1989a-b) have noted that an inadequate model structure

(conceptualization) is far more detrimental to its predictive ability than is a suboptimal set of model parameters. This helps explain why the National Research Council (2000), in a recent study of Research Needs in Subsurface Science for the Environmental Management Program of the U.S. Department of Energy, has listed as second among four recommended research emphases the development of tools and methodologies for conceptual modeling with emphasis on heterogeneity, scale and uncertainty bounds on the basis of field experimental data.

Recently, a panel was convened by the National Research Council (2001) to describe the process through which conceptual models of flow and transport in the fractured vadose zone are developed, tested, refined and reviewed. The panel concluded that development of the conceptual model is the most important part of the modeling process. The conceptual model is the foundation of the quantitative mathematical representation of the field site (i.e., the mathematical model), which in turn is the basis for the computer code used for simulation. Reasonable alternative conceptualizations and hypotheses should be developed and evaluated. In some cases, the early part of a study might involve multiple conceptual models until alternatives are eliminated by field results.

According to the panel, it is important to recognize that model predictions require assumptions about future events or scenarios, and are subject to uncertainty. Meaningful quantification of uncertainty should be considered an integral part of any modeling endeavor, as it establishes confidence bands on predictions given the current state of knowledge about the system. A suite of predictions for a range of different assumptions and future scenarios is more useful than a single prediction.

In fact, we have seen earlier that hydrologic models are prone to equifinality (Beven, 2000) in both their structures and parameter sets. This means that there is uncertainty not only about the parameter values that should enter into a given model (as characterized by its structure), but also about the very structure (conceptual and mathematical) of the model that should represent the hydrologic system of interest. The traditional approach to model uncertainty analysis, which considers only a single deterministic model structure, fails to adequately sample the complete space of plausible hydrologic models. As such, it is prone to modeling bias and underestimation of model uncertainty.

An example of how one could account quantitatively for structural model uncertainties was given by James and Oldenburg (1997). They investigated the uncertainty of simulated TCE concentrations, at the point of potential human exposure, due to uncertainty in the parameters (permeability, porosity, diffusivity, solubility, adsorption) and variations in the conceptual-mathematical model (injection rate of TCE source; initial TCE source saturation; regional groundwater flow; heterogeneity of permeability). The authors used the three-dimensional code T2VOC to simulate three-phase (gas, aqueous, NAPL), three-component (air, water, VOC) nonisothermal flow based on an actual site with a 25 m thick vadose zone and a saturated zone. To assess parameter uncertainty associated with a given model, they used the inverse code ITOUGH2. Their final step was to assess the range of outcomes that one obtains with the entire set of alternative conceptual-

mathematical models. James and Oldenburg found that uncertainties in their model outcomes span orders of magnitude, and that both parameter and model uncertainty contribute significantly to this wide range of outcomes. They concluded that "risk assessment and remediation selection ... is meaningful only if analysis includes quantitative estimates of ... uncertainty" in both the parameters and the conceptual-mathematical models.

A similar approach has been advocated more recently by Samper and Molinero (2000). The authors consider the main uncertainties in predicting groundwater flow and transport to be those associated with the selection of future scenarios, choice of model structure and assignment of model parameters. The authors consider parameter uncertainty to be minor in comparison to structural (i.e. conceptual) model errors. They suggest to evaluate model predictive uncertainty by calibrating a number of conceptual-mathematical models against available monitoring data, to retain those calibrated models that can adequately reproduce past observations, to assess the predictive uncertainty of each model due to the uncertainty of its parameters, to treat the predictive uncertainty of each model as being equally likely, and to produce a single combined range of predictive uncertainties.

Rather than relying on model calibration and treating the outcomes of different structural models as being equally likely, Beven and Binley (1992) have proposed a strategy to which they refer as GLUE (Generalized Likelihood Uncertainty Estimation). The strategy calls for the identification of several alternative structural models and the postulation of a prior probabilistic model of parameter uncertainty for each. Each structural model, coupled with its corresponding parameter uncertainty model, is used to generate Monte Carlo realizations of past hydrologic behaviors and to compare the results with monitored system behavior during the same period. Likelihood measures are defined to gauge the degree of correspondence between each simulated and observed record of system behavior. If a likelihood measure falls below a subjectively defined "rejection criterion," the corresponding combination of model structure and parameter set are discarded. Those combinations which pass this test are retained to provide predictions of system behavior under selected future scenarios. Each prediction is weighted by a corresponding normalized likelihood measure (so as to render the sum of all likelihood measures equal to one), to produce a likelihood-weighted cumulative distribution of all available predictions.

As of 1992, the GLUE strategy has been applied to a variety of hydrologic problems including rainfall-runoff modeling, solute transport in rivers and groundwater, flood frequency and flood inundation analyses, critical load modeling, nitrogen budget calculations and the analysis of distributed land surface to atmosphere vapor and heat fluxes. For recent discussions of GLUE and its applications the reader is referred to Beven (2000) and Beven and Freer (2001).

A Bayesian approach to the quantification of errors in a single groundwater model was recently proposed by Gaganis and Smith (2001). Like GLUE, it relies on Monte Carlo simulations without model calibration and on subjective criteria of “model correctness.”

It must be understood that the set of predictions one produces with any given choice of alternative structural models and

parameter sets, by whatever method, is conditional on the choice of models and the data used to support them. As such, these predictions do not represent all possibilities but only a limited range of such possibilities, associated with these models and data. Any change in the latter would generally lead to a different assessment of predictive model uncertainty. There thus appears to be no way to assess the uncertainty of hydrologic predictions in an absolute sense, only in a conditional or relative sense.

1.2.7 Multiobjective Tradeoffs between Model Fit and Complexity

The National Research Council (2001) panel on conceptual models of flow and transport in the fractured vadose zone noted that a conceptual model is by necessity a simplification of the real system. However, the degree of simplification must be commensurate with the problem being addressed. Yet neither the panel nor the broader literature offer any systematic guidelines about how to determine the optimum degree of model simplification for a given purpose.

One important measure of model complexity is the number of free parameters (degrees of freedom) that it contains. This is the number of parameters that are free to be adjusted during calibration so as to achieve an optimum model fit with monitored system behavior. A model with a given set of free parameters, which describe a spatially varying quantity such as permeability or porosity, might be more or less complex depending on their mode and degree of variation in space. Several such measures of complexity have been introduced by Emsellem and de Marsily (1971), Neuman (1973), Distefano and Rath (1975) and Neuman and De Marsily (1976).

The approach taken by these authors was to calibrate a series of hydrologic models with gradually increasing degrees of model complexity. Each calibration yields a measure of model fit that corresponds to a particular measure of model complexity. By plotting the measure of fit versus the measure of complexity, one usually finds the former to improve (decrease) as the model becomes more complex. The improvement is rapid when the model is relatively simple and slow when the model is relatively complex. Clearly, the standard hydrologic objectives of achieving a good fit between simulated and observed system behavior, and keeping the model simple, are in conflict. Instead of one optimum solution to this multiobjective or Pareto optimization problem, there is a range of feasible (or Pareto) solutions among which one must choose the optimum tradeoff subjectively (Neuman, 1973). The optimum tradeoff is achieved at a point beyond which any further increase in model complexity would bring about only a relatively insignificant improvement in model fit. A similar idea has been espoused more recently by Sun and Yeh (1985) and summarized by Sun (1994) in a book on Inverse Problems in Groundwater Modeling.

It is possible to postulate multiple, non-commensurate model fit criteria (Gupta *et al.*, 1998) and model complexity measures (Neuman, 1973) within a multiobjective framework. Typically, models lack the ability to simulate all relevant aspects of hydrologic behavior with an equal degree of fidelity. For example, a subsurface transport model may not be able to simulate accurately observed contaminant concentration versus time records at several monitoring locations simultaneously without becoming overly complex and cumbersome. Or the model may be unable to reproduce

equally well the peak arrival time, magnitude, rate of ascent and rate of recession of a contaminant breakthrough record. One then has the option of defining multiple criteria of model fit, each representing a different aspect of desired model performance. Multiobjective programming techniques are available (*e.g.*, Duckstein *et al.*, 1991) that should then make it possible for the modeler to investigate the range of feasible solutions to the problem and solve it by determining (subjectively) an acceptable tradeoff between all of its conflicting and incommensurate objectives. A computationally efficient and mathematically rigorous method to generate binary graphs of tradeoff between each pair of conflicting objectives, and to determine an optimum tradeoff between all of them interactively, was developed by Neuman and Krzystofowicz (1977).

1.2.8 Combined Measures of Model Fit and Complexity

Several combined measures of model fit and complexity have been developed in the context of time series analysis by Akaike (1974, 1977), Hannan (1980) and Kashyap (1982). In recent years, such measures have been used with increasing frequency to rank and discriminate between alternative geostatistical (Samper and Neuman, 1989a-b; Chen *et al.*, 2000) and dynamic (Carrera and Neuman, 1986a-b; Finsterle and Faybishenko, 1999) groundwater models. The model discrimination criteria combine a log-likelihood function, which gauges the quality of model fit to monitoring data, with a term proportional to the number of free model parameters. By ranking the models in ascending order of their associated discrimination criteria, one favors the least complex among those that perform equally well in terms of model fit. Among models of equal complexity, those that fit the monitoring data better are favored.

The model discrimination criterion of Kashyap contains yet another term that gauges the information content of the available data. It thus allows considering models of growing complexity as the data base improves in quantity and quality. Stated otherwise, the criterion recognizes that when the data base is limited and/or of poor quality, one has little justification for selecting an elaborate model with numerous parameters. Instead, one should then prefer a simpler model with fewer parameters, which nevertheless reflects adequately the underlying hydrogeologic structure of the system, and the corresponding flow and transport regime. Kashyap's criterion favors that model which, among all alternatives considered, is least probable (or likely, in an average sense) to be incorrect. Stated otherwise, the criterion minimizes the average probability of selecting the wrong model among a set of alternatives.

1.2.9 Effects of Scale on Hydrologic Predictability

In September 2000, a workshop entitled Predictability and Limits to Prediction in Hydrologic Systems was held at the National Center for Atmospheric Research in Boulder, Colorado, under the auspices of the National Research Council. The workshop concluded that (Entekhabi, 2000) hydrologic systems contain heterogeneous geological, topographic, and ecological features that vary on multiple scales. This multiplicity of scales has a pervasive effect on hydrologic characterization and prediction. In general, the effective response of systems at larger scales is not completely determined by scaling local processes.

In hydrologic science, heterogeneity is a rule and it cannot necessarily be fully captured by randomization of parameters (essentially an after-thought). Interactions among micro-scale features often lead to effects that are not completely represented in macro-scale predictions based on effective parameters for micro-scale models.
Examples include enhanced surface flux due to land-breeze circulations over heterogeneous patches, regional recharge and discharge patterns over complex terrain. There are processes and conditions when the effective parameter approach to scaling may be feasible. In the remaining circumstances, the macro-scale and micro-scale predictive relations for hydrologic processes may have different functional forms and dependencies. Furthermore, there may be organizing principles at work that result in simple procedures for statistically relating variables across a wide range of scales in the hydrologic system. We add that such principles have been noted and procedures developed (with applications to the permeability and dispersivity of geologic media) by Neuman (1990, 1994), Di Federico and Neuman (1997; see also Neuman and Di Federico, 1998) and Di Federico *et al.* (1999).

According to Entekhabi (2000), it has been traditional to assume that spatially and temporally averaged hydrologic quantities are less prone to predictive uncertainty than quantities that are localized in space-time. While this may be true for systems that are statistically homogeneous and stationary, such systems are the exception rather than the rule. Spatial and temporal averages are not necessarily more predictable as traditionally believed if the averaging covers a scale that contains a strong transition or change in behavior (analogous to bifurcation in nonlinear dynamic systems).

1.2.10 Prediction Uncertainty Due to Random Fluctuations in Hydrologic Variables

Hydrogeologic environments consist of natural soils and rocks that exhibit both systematic and random spatial variations in hydraulic and transport properties on a multiplicity of scales. Groundwater flow and solute transport take place under the action of forces whose exact nature, magnitude and space-time distribution are generally uncertain. Traditional deterministic methods of analyses capture at best the larger-scale, systematic components of these variations and distributions. They however fail to resolve smaller scale variations or account for their uncertain nature. Yet capturing hydrologic complexity on a wide range of scales, and quantifying the associated uncertainty, is proving to be of paramount importance for a wide array of applications such as the characterization and remediation of contaminated sites, risk analysis and monitoring design for sites subject to potential future contamination, and performance assessment of nuclear facilities and sites.

A recent study of groundwater and soil cleanup by the National Research Council (1999) recognizes that the subsurface is usually highly heterogeneous and characterizing this variability is extremely difficult. This heterogeneity and difficulty in characterization complicate the design of subsurface cleanup systems because predicting system performance under such uncertain conditions is difficult. An accurate knowledge of geological heterogeneities is vital for evaluating the hydrogeologic limits on subsurface contaminant remediation.

In a recent editorial titled "It's the Heterogeneity!", the Editor of the most widely read groundwater journal (Wood, 2000) opines that the natural system is so complex that it will be many years before one can effectively deal with heterogeneity on societally important scales.

The complex nature of the subsurface and of the conditions that control groundwater flow and transport within it throw into question the reliability of traditional deterministic methods of hydrogeologic analysis. The emphasis is therefore shifting from deterministic to probabilistic methods that are better suited for these needs. The trend has become to describe the spatial variability and scaling of hydrogeologic medium properties geostatistically, and to analyze subsurface fluid flow and solute transport stochastically. This trend has been documented in a number of recent books including those by Dagan and Neuman (1997) and Zhang (2001).

The most common method of stochastic analysis is high-resolution computational Monte Carlo simulation that produces a large number of equally likely results. These nonunique results are summarized in terms of statistically averaged quantities, their variance-covariance, and perhaps higher moments of the corresponding sample probability distributions. Results that honor measured values of medium properties are said to be conditioned on these data. Upon conditioning the simulations on measured values of parameters in space, one obtains (among others) conditional mean flow and transport variables that constitute optimum unbiased predictors of these unknown random quantities. One also obtains conditional second moments (variance-covariance) that provide a measure of the associated prediction errors. To condition the predictions on system monitoring data, one must either discard random simulations that do not reproduce the observations, or employ an inverse procedure of the kind developed for this purpose by Sahuiquillo *et al.* (1992), Gómez-Hernández *et al.* (1997; 2000) and Capilla *et al.* (1997).

Monte Carlo analysis requires knowing the multivariate probability distribution of relevant hydrogeologic properties, which is difficult to infer from commonly available data. To achieve a high space-time resolution of relevant stochastic phenomena, it requires the use of large space-time grids with very small discretization intervals. To yield sample statistics that converge to their theoretical (ensemble) counterparts requires numerous repetitions (realizations). The net result is a large amount of computational time and storage, which are considered uneconomical for many practical applications.

This has given impetus to the development of alternative stochastic methods that allow one to compute the conditional mean, variance and covariance of groundwater flow and transport variables directly, without Monte Carlo simulation. This is done on the basis of conditional moment equations that may be nonlocal (integro-differential) and contain parameters that depend on more than one point in space and/or time. Conditioning additionally renders the parameters of such equations dependent not only on medium properties but also on the information that one has about these properties (scale, location, quantity and quality of data). Hence the parameters of conditional moment equations are nonunique. Exact conditional moment equations have been developed for steady state (Neuman and Orr, 1993; Neuman *et al.*, 1996) and transient (Tartakovsky and

Neuman, 1998, 1999) groundwater flow in saturated porous media, steady state flow in a certain class of unsaturated soils (Tartakovsky *et al.*, 1999; Lu *et al.*, 2002), and advective (Neuman, 1993) or advective-dispersive (Zhang and Neuman, 1996) transport of a nonreactive solute.

Though the underlying stochastic flow and transport regimes are taken to obey Darcy's law and Fick's analogy, these relations are generally not obeyed by the conditional mean flow and transport regimes except in special cases or as localized approximations. Such localized approximations yield familiar-looking differential equations which, however, acquire a non-traditional meaning in that their parameters (hydraulic conductivity, seepage velocity, dispersivity) and state variables (hydraulic head, concentration, fluid and solute fluxes) are information-dependent and thus nonunique (Guadagnini and Neuman, 1999a-b, 2001; Ye *et al.*, 2002). Whereas nonlocal moment equations contain information about predictive uncertainty, localized (differential) equations generally do not.

To render conditional moment equations workable, one must adopt some form of closure approximation, such as perturbation (Guadagnini and Neuman, 1999a-b, 2001; Ye *et al.*, 2002; Lu *et al.*, 2002) or an assumption of Gaussianity (Neuman *et al.*, 1999; Amir and Neuman, 2001; Wang *et al.*, 2002). Though perturbative closure approximations are not guaranteed to work for strongly heterogeneous media, they have proven to work remarkably well in two-dimensional finite element analyses of superimposed mean uniform and convergent steady state (Guadagnini and Neuman, 1999a-b) and transient (Ye *et al.*, 2002) flows in saturated porous media and steady-state flow in unsaturated media (Lu *et al.*, 2002).

In the aforementioned moment analyses, conditioning was limited to prior measurements of parameters (mainly permeability) at discrete points in space. To condition the predictions on state variables such as hydraulic head and concentration, it is necessary to develop corresponding inverse algorithms of the kind recently described by Hernandez *et al.* (2002).

1.3 Outline of Proposed Strategy

1.3.1 Comprehensive Approach to Modeling and Uncertainty Analysis

This report describes a strategy that embodies a systematic and comprehensive approach to hydrogeologic conceptualization, model development and predictive uncertainty analysis. The strategy is comprehensive in that considers all stages of model building and accounts jointly for uncertainties that arise at each of them. The stages include regional and site characterization, hydrogeologic conceptualization, development of conceptual-mathematical model structure, parameter estimation on the basis of monitored system behavior, and assessment of predictive uncertainty. In addition to parameter uncertainty, the strategy concerns itself with uncertainties arising from incomplete definitions of (a) the conceptual framework that determines model structure, (b) spatial and temporal variations in hydrologic variables that are either not fully captured by the available data or not fully resolved by the model, and (c) the scaling behavior of hydrogeologic variables.

The strategy addresses these issues within a unified and coherent theoretical framework. It does so by relying on both traditional and nontraditional ideas and techniques. Some nontraditional ideas and techniques are

taken from the recent literature and some are entirely new. Most are amenable to immediate implementation with the aid of existing software. Though there is at present no publicly available software for the solution of stochastic flow and transport problems by means of the moment equation method (which is part of the strategy), such software is expected to become available within a few years. In the meantime, such problems can be tackled effectively with existing software by means of the Monte Carlo option provided in the strategy.

1.3.2 NRC Licensing Context

The strategy is generic but designed to be of practical use to NRC staff in their review of decommissioning plans, and performance assessment of high-level and low-level radioactive waste disposal sites as well as uranium recovery facilities. For this purpose, the strategy is cast in the context of a framework that is useful to NRC staff review and performance evaluation needs. The regulatory context is defined in terms of corresponding performance measures (as identified by the NRC staff in Appendix A), hydrogeologic analyses that are needed to assess them, the desired reliability of such assessments, and the expenditure (in time, effort and money) that is allowed to achieve it. Examples from various case studies are included to help illustrate some aspects of these analyses.

An important component of the strategy is a systematic sequence of logical questions, guidelines and criteria with analytical methods that may be useful to the NRC staff in their review and performance evaluation. Though the contract does not call for the development or demonstration of practical data processing and modeling tools or software, such tools and software are identified where possible. To the extent feasible, the strategy is supported by field experimental data that cover a range of technical problems facing the NRC staff.

1.3.3 Hydrogeologic Conceptualization and Data

The strategy encourages exploration of varied conceptual frameworks and assumptions at all stages of hydrogeologic model development through a comprehensive evaluation of a broad range of regional and site data, their translation into coherent and internally consistent conceptual-mathematical models, and computation and visualization based on these data and models. Included among these frameworks and assumptions are various model simplification and abstraction schemes. The strategy recognizes that site characterization and monitoring data are expensive and difficult to collect, leading to a ubiquitous scarcity of hard site information. It is therefore critically important to assess the role that such data play in rendering the hydrogeologic performance analysis credible. The strategy stresses the role of characterization and monitoring data in helping one identify and test alternative conceptual models, make rational choices among them, gauge and reduce model bias and uncertainty through proper model selection and calibration, render and assess the reliability of model predictions, and confirm the assessment through independent peer review as well as at least some degree of direct verification.

1.3.4 Iterative Approach

The strategy encourages an iterative approach to modeling, whereby a preliminary conceptual-mathematical model is gradually altered and refined until one or

more likely alternatives have been identified and analyzed. The idea is to start with a conceptual-mathematical model that is as simple as may appear warranted by the context of the problem and the available data. This model is then tested for qualitative consistency with the available data, made quantitatively compatible with the data through calibration, and subjected to an assessment of its predictive reliability (in terms of potential bias and uncertainty). Next, the model structure is altered and/or refined (by modifying its hydrogeologic makeup or features; dimensionality; scale of resolution in space-time; type of driving forces or events including sources, initial and boundary conditions; and governing equations or processes), the model is retested and compared with the former. The iterative procedure is repeated till there is good reason to conclude that any additional alterations or refinements of the models would lead at best to minor improvements in their predictive capability. Models whose predictive capability is deemed low in comparison to other are discarded and the rest retained for further analysis. The retained models are used to render a joint prediction. The corresponding prediction error is quantified by means of a joint variance. If the joint prediction and associated error variance are deemed acceptable for the stated purpose of the analysis, the iterative process is halted and the models declared reliable and credible within a clearly stated conditional (on the available information and selected set of models) margin of error. The conditional nature of this error estimate is taken to imply that it constitutes a lower bound, and that the actual predictive error may (and most probably is) larger to an unknown degree.

If the models are deemed unacceptable, a sensitivity analysis is performed to help identify the type and quantity of additional site data that might materially enhance their reliability and credibility. A decision is then made, based on the potential benefit (in terms of bias and uncertainty reduction) and cost (in terms of time, effort and money) of such data, whether or not to collect them and how. If and when new data of significant weight are obtained, the iterative process is repeated till its cost-benefit ratio reaches a value that is not considered worth exceeding.

1.3.5 When to Stop Collecting Data and Refining the Model?

It is worth repeating that the questions (1) what degree of model refinement is enough for a given regulatory purpose, and (2) how much and what kind of data are enough for the task, are recast here in terms of economic and policy questions, which are the responsibility of managers and decision-makers to answer. The strategy is designed to help the analyst address the questions by how much would further model refinements, and/or additional site characterization and monitoring, improve the prediction of performance measures derived from the hydrogeologic analysis. In other words, the strategy addresses the worth of model refinement and data collection in terms of their contributions to the potential enhancement of model reliability and credibility, not in terms of their marginal cost-benefit. Its aim is to allow managers and decision-makers to make informed decisions about the time, manpower and budget that they deem worth allocating to these activities in light of such enhancement under existing administrative, budgetary and policy constraints.

1.3.6 Assimilating Regional and Site Characterization Data

The strategy recognizes that it is often possible to postulate hydrogeologic conceptual models or hypotheses for a site on the basis of publicly available geologic and geographic information about its surroundings. Additional conceptualization can be done on the basis of generic data about similar regions and the properties of similar materials elsewhere. Several such regional and generic sources of information are identified and discussed in this report. Yet each site is unique and so virtually guaranteed to reveal additional features, properties and behaviors when characterized in some detail locally. Hence the strategy considers local characterization essential for the postulation of acceptably robust conceptual hydrogeologic models for a site.

Regional and site characterization data tend to represent a wide range of measurement scales, not all of which are compatible with the intended scale of hydrogeologic model resolution. The strategy promotes recognition of this important issue and an effort to render the scale of measurement compatible with the scale of model resolution. This can be done by either rescaling the data to fit the scale of model resolution (which often entails averaging or upscaling over computational grid cells) or adapting model resolution to fit the scale of measurement (which often entails adapting the size of grid cells to the size of the data support). Recent advances in scaling theory and practice are briefly reviewed to promote awareness of fundamental issues associated with the scaling of hydrogeologic variables.

The strategy encourages the analyst to associate model parameters, derived from regional and site characterization data, with statistical measures of their uncertainty. Such prior statistics provide the only way to assess model predictive uncertainty when suitable monitoring data are not available. Methods are discussed and an example given to derive such statistics from generic data sources on the basis of similarity in soil or rock textural features; to update these statistics on the basis of site-specific data using a Bayesian approach; and to interpolate, extrapolate and/or average parameters across the site geostatistically on the basis of measured values at discrete points in space.

1.3.7 Testing and Calibration against Monitoring Data

While hydrogeologic characterization of a site and its surroundings makes conceptualization possible, it does not provide the means to test conceptual models or compare them with other alternatives. For this, it is necessary to have monitoring data that constitute observations of actual hydrologic behavior at and around the site. Only with such data can one evaluate the ability of models to mimic real system behavior (qualitatively at the conceptual level, quantitatively at the conceptual-mathematical level), improve their ability to do so through calibration against the monitoring data, determine their optimum degree of refinement or complexity, and compare them with each other (qualitatively and quantitatively). The proposed strategy introduces ways to accomplish these tasks in the context of traditional deterministic groundwater flow and transport models.

The strategy encourages the analyst to employ statistical measures of model performance (or fit between simulated and observed hydrologic behaviors) where possible; to determine feasible tradeoffs

between model fit and complexity by relying on multiobjective approaches; and to rank models based on (a) statistical indicators of the quality of fit they produce (as obtained from an analysis of residuals), and (b) likelihood-based model discrimination criteria (which penalize models for using an excessive number of free parameters to achieve a given quality of fit). Models that show relatively poor fits and rank low on the list can be subjectively eliminated from further consideration at this stage.

An important byproduct of model calibration is information about parameter uncertainty. When reliable and relevant (in terms of type and scale) prior statistics about the parameters are available from site characterization, the strategy encourages their use as input into the calibration process. Regardless of whether or not suitable prior statistics are available and employed for model calibration, the latter should produce posterior statistics for the optimum parameter estimates. The strategy considers the posterior parameter estimates, and their error statistics, more suitable for the analysis of model predictive uncertainty than the prior estimates and their statistics. This is so because calibration generally improves the ability of the model to reproduce simulated hydrologic behavior, and a calibrated model is therefore deemed more reliable and credible as a predictive tool than an uncalibrated model. The same holds true for the associated parameter estimation statistics.

1.3.8 Predictive Uncertainty of Deterministic Models

The strategy considers two alternative methods to assess the predictive uncertainty of a deterministic groundwater model under assumed future scenarios. The first method relies on Monte Carlo simulation of these scenarios using prior or posterior parameter estimates and statistics. The second method establishes approximate error bounds, or confidence limits, for such scenarios by linearization, using the same parameter estimates and statistics. The first approach is more accurate but costly in terms of computer time and effort. In both cases, the strategy favors the use of posterior estimates over that of priors on the understanding that the former are less biased and uncertain than the latter.

To render a joint prediction and assess the corresponding joint uncertainty based on all calibrated models that have been retained as constituting potentially viable alternatives, the strategy introduces a new approach. The approach is based on the idea of Maximum Likelihood Bayesian Model Averaging (MLBMA) recently proposed by Neuman (2002). MLBMA relies on probabilistic maximum likelihood (ML) concepts to (a) calibrate each model against observed space-time variations in system states (pressure, water content, concentration), considering prior information about relevant soil properties; (b) eliminate models that cannot be so calibrated with acceptable fidelity; (c) predict future system behavior or performance measures (travel times, concentrations, mass rates), and assess corresponding predictive uncertainty, using each model; and (d) average the results using posterior model probabilities as weights. MLBMA supports the principle of parsimony in that among models having similar predictive capabilities, it favors those having fewer parameters and are thus simpler. It renders an earlier Bayesian Model Averaging (BMA) approach (Hoeting *et al.*, 1999) computationally feasible by basing it on a maximum likelihood approximation due to Kashyap (1982) and

the maximum likelihood parameter estimation method of Carrera and Neuman (1986a). Its final outcome is a conditional assessment of predictive uncertainty that accounts jointly for uncertainties in the conceptual-mathematical model structure and its parameters.

MLBA differs in important ways from earlier approaches. It differs fundamentally from the GLUE approach of Beven and Binley (1992; see also Beven, 2000, and Beven and Freer, 2001), as well as from the Bayesian approach of Gaganis and Smith (2001), in that it (a) relies on posterior (post-calibration) rather than prior statistics, (b) avoids the need for time-consuming Monte Carlo simulations, and (c) obviates the need for subjective model acceptance thresholds or criteria. MLBA differs from the approach of James and Oldenburg (1997) and Samper and Molinero (2000) in that it does not generally treat calibrated models as being equally likely.

It is important to reemphasize that the set of predictions one produces with any given choice of alternative structural models and parameter sets, by whatever method, is conditional on the choice of models and the data used to support them. As such, these predictions do not represent all possibilities but only a limited range of such possibilities, associated with these models and data. Any change in the latter would generally lead to a different assessment of predictive model uncertainty. There thus appears to be no way to assess the uncertainty of hydrologic predictions in an absolute sense, only in a conditional or relative sense.

If suitable monitoring data are not available, then there is neither an objective nor a quantitative basis for the comparison of potentially viable conceptual-mathematical models. Likewise, there are no posterior but only prior parameter estimates and statistics to deal with. To quantify the predictive uncertainty of each such model, one again has the option of relying on Monte Carlo or a linearized analysis based on the available prior statistics. To aggregate these results, one can either weigh all models equally or associate them with subjective weights based on the professional judgment of the analyst or a group of experts.

1.3.9 Assessing the Need for More Data

To help evaluate what if any additional data might be worth collecting so as to materially reduce model uncertainty (by further constraining the range of alternative structures and parameters), the strategy suggests conducting a sensitivity analysis to indicate what system behavior appears to be most sensitive to which parameters at what locations. The next step is to consider performing additional site characterization where existing parameter estimates are least certain and the model is relatively insensitive to their values, and monitoring system behavior where it is most sensitive to model parameters while prediction errors appear to be relatively large and consequential.

1.3.10 Accounting for Unresolved Variations in Space-Time

Calibration of a deterministic groundwater model yields effective parameter estimates, which compensate to some degree for lack of knowledge about variability on space-time scales that are not resolved by the data or the model. This is less true for transport than for flow, because transport processes due to unresolved variations in permeability and porosity cannot be validly represented by traditional effective parameters (such as a

constant dispersivity) except in special cases or as a crude approximation. The reliance on effective parameters is also less valid for strongly heterogeneous media with preferential flow paths such as some fractured rocks and soils that contain relatively large and elongated openings created by burrowing animals or plant roots. When the deterministic effective parameter approach fails, the most common and powerful alternative is to employ stochastic concepts and models.

A stochastic approach to groundwater modeling requires that the spatial variability of input parameters, such as permeability and unsaturated soil properties, be characterized with the aid of geostatistical methods. A key component of geostatistical characterization is the assessment of spatial covariance structure (in terms of a covariance or variogram function). The strategy promotes this approach and the report illustrates it by examples. In the absence of site characterization data that are amenable to geostatistical analysis, one can treat the structural parameters (of the covariance or variogram function) as free parameters to be estimated by model calibration. This too is illustrated by example in the report.

1.3.11 Stochastic Modeling

The strategy supports two general approaches to predict stochastic groundwater flow and transport on the basis of geostatistical input: High-resolution numerical Monte Carlo simulation and direct (deterministic) prediction of mean behavior. The latter approach relies on stochastically derived deterministic ensemble mean flow and transport equations in which the dependent variables represent not actual system states (such as head, concentration or flux) but rather their (ensemble) mean values or statistical expectations. These mean or expected values represent optimum unbiased predictors of the unknown, actual system states. Similar values are obtained upon averaging the results of numerous Monte Carlo simulations. Both the Monte Carlo and the mean equation approaches allow the reduction of predictive uncertainty by conditioning the predictions on actual measurements of the input parameters at discrete locations in space.
The variance-covariance of high-resolution (conditional) Monte Carlo simulations serves to quantify predictive uncertainty. An alternative is to compute this variance-covariance directly (deterministically) by means of corresponding moment equations. Both options are supported by the strategy.

Conditioning the stochastic flow and transport equations not only on measured input variables (derived from site characterization) but also on monitored system behavior tends to improve their reliability. This is accomplished through novel inverse procedures, which form part of the proposed strategy.

In virtually all other respects, the strategy is similar in principle (though not in all details) for stochastic and deterministic models of groundwater flow and transport.

2 NATURE OF HYDROGEOLOGIC SITE MODELS

Prior to developing a model, one must define what one means by the term.

2.1 Site-Specific Nature of Hydrogeologic Models

Hydrogeologic models are by nature site-specific. Though there is an established (and evolving) set of general hydrogeologic principles that apply to many sites on many scales, they are insufficient to either describe (conceptualize) or quantify (model) the hydrogeology of a particular site over a given range of scales. Because each site is unique, general principles must always be supplemented by regional and site-specific data to be useful for conceptualizing and modeling subsurface flow and transport at a site, regardless of purpose.

General hydrogeologic principles may serve to develop generic methods and tools of hydrogeologic analysis that are applicable, in principle, to many sites on varied scales. Among numerous examples one may cite software tools for reconstruction and visualization of site hydrostratigraphy and structure, statistical and geostatistical analysis of site data, interpretation of pumping tests and simulation of flow and transport processes. Whereas numerical simulation codes are often referred to as "models," we avoid such usage of the term on the understanding that they are tools rather than models. To transform simulation codes into models, one must apply them to particular hydrogeologic circumstances that represent either a hypothetical or a real setting. Hydrogeologic models of hypothetical settings are research and demonstration tools that may be useful for the exploration and exposition of generic principles and hypotheses. They have also been incorporated in multi-media codes sometimes used for dose and risk assessment. Though they may serve a useful purpose, hydrogeologic models of hypothetical settings must not be mistaken for valid models of particular sites. To render them such, one must ground these models in real data which represent specific geographic locations, hydrogeologic conditions and natural or anthropogenic forcing that prevails at these locations.

The strategy described in this report deals with site-specific rather than hypothetical models of subsurface flow and transport within a regulatory context that is relevant to NRC review and performance evaluation needs.

2.2 What is a Hydrogeologic Model?

For purposes of the strategy developed in this report, a hydrogeologic model is a conceptual and mathematical construct that serves to analyze, qualitatively and quantitatively, subsurface flow and transport at a site in a way that is useful for NRC review and performance evaluation. It consists of a description of the site and circumstances being modeled, a set of regional and site data to support this description, a mathematical framework that incorporates both the description and the data, and a mathematical (analytical or numerical) evaluation of system behavior and performance. The evaluation typically consists of simulating (reconstructing or predicting) and analyzing space-time variations in quantities such as hydraulic head or pressure, solute concentration, fluid and solute flux and velocity, solute travel time and associated performance measures. A hydrogeologic model thus consists of a

conceptual and a mathematical component.

Whereas much has been written about the mathematical component of hydrogeologic models, relatively little attention has been devoted to the conceptual component. In most mathematical models of subsurface flow and transport, the conceptual framework is tacitly assumed to be given, accurate and unique. All three premises are challenged by the strategy in this report, which focuses in a major way on issues associated not only with mathematical but also with conceptual aspects of hydrogeologic model building and evaluation. For this, we need to define what is meant by conceptual and mathematical modeling, and how do the two interrelate? This is done next.

2.3 What is a Conceptual Model of Site Hydrogeology?

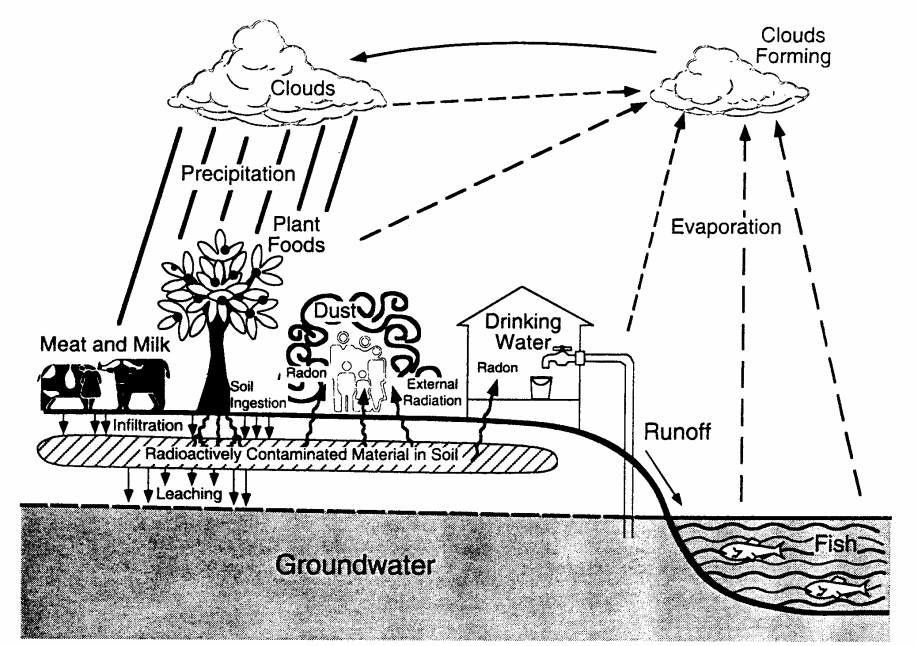
We stated that a hydrogeologic model consists of a description of the site and circumstances being modeled, a set of regional and site data to support this description, a mathematical framework that incorporates both the description and the data, and a mathematical (analytical or numerical) evaluation of system behavior and performance. A conceptual model embodies the descriptive component of the model, which may be both qualitative and quantitative.

According to an exchange of letters between Paul Roman, Alfred A. Brooks, and Lorenzo de la Torre in Physics Today (1998), physical (and hydrogeologic) reality exists objectively but is not fully or directly accessible to us. Instead, it reveals itself to us via sensory impressions and experiences (through observations and experiments) in signatures or bits. To assimilate these separate pieces of information we describe, contemplate, correlate, connect, systematize, interpret and integrate them into a body of knowledge. On the basis of this knowledge we form mental constructs, or contextual structures, which we call conceptual models of reality.

Conceptual descriptions of hydrogeologic systems may take many forms. According to Lein (1997, p. 85), conceptual models take the forms of charts, pictures and diagrams depicting system arrangements and flows. Examples relevant to NRC staff performance assessment use include a conceptual diagram and a flow chart of exposure pathways considered in the RESRAD (Yu, 1993) dose assessment model (Figure 2-1), a similar flow chart (Figure 2-2) from a technical guide to groundwater model selection issued jointly by three agencies (EPA, DOE, NRC, 1994); a flow chart of input data and processes considered in the performance assessment of geosphere transport (Figure 2-3) by Gautschi (1995), and a flow chart of processes that may impact the performance of an engineered barriers system at a potential HLW repository in Nevada (Figure 2-4).

A definition more directly relevant to hydrogeologic modeling has been given by Anderson and Woessner (1992). According to them (*ibid*, p. 28), a conceptual model is a pictorial representation of the groundwater flow system, frequently in the form of a block diagram or a cross section (Figure 2-5). Its development includes (*ibid*, p. 6) identification of hydrostratigraphic units and system boundaries; assembly of field data including information on the water balance and data needed to assign values to aquifer parameters and hydrologic stresses; and a field visit to help keep the modeler tied into

reality and exert a positive influence on his subjective modeling decisions. The purpose of a conceptual model is (*ibid*, p. 28) to simplify the field problem and organize the associated field data so that the system can be analyzed more readily by means of a numerical model; in their view, the nature of the conceptual model determines the dimensions of this model and the design of the grid.



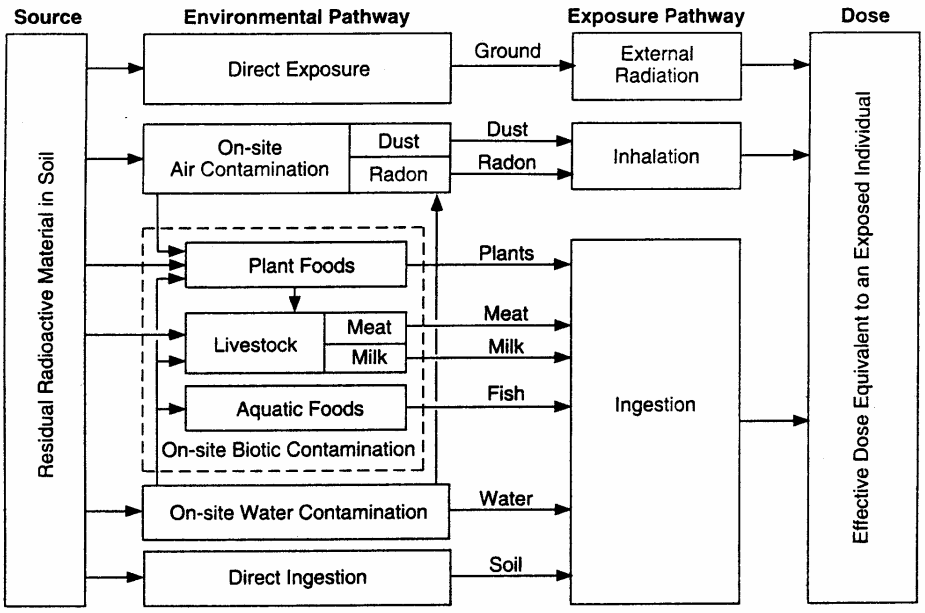


Figure 2-1. Conceptual diagram and flow chart of exposure pathways in RESRAD (after Yu, 1993).

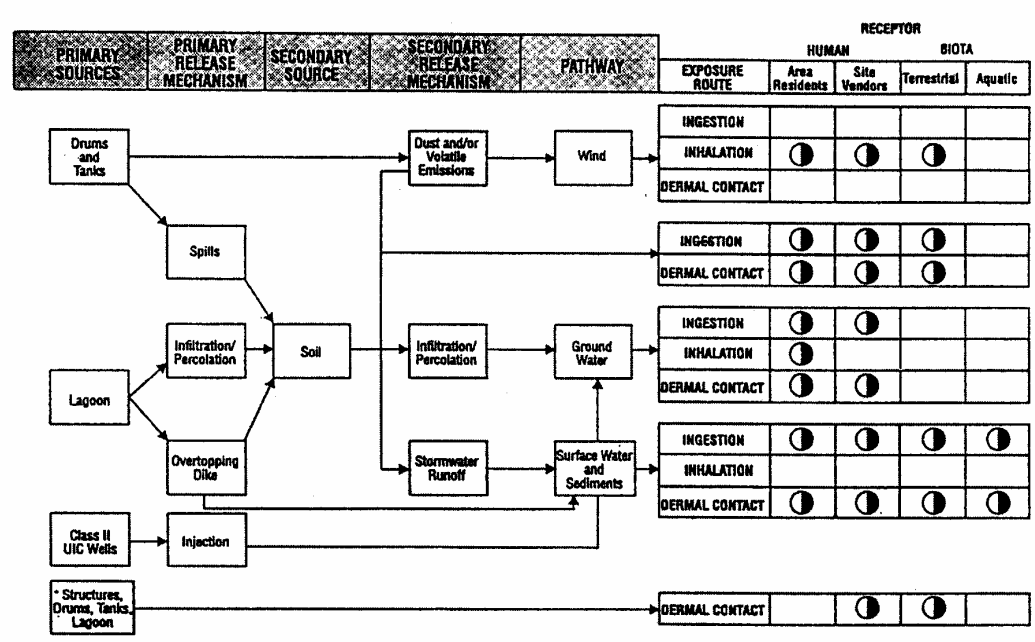


Figure 2-2. Conceptual model
as depicted by EPA, DOE and NRC (1994) guide to model selection.

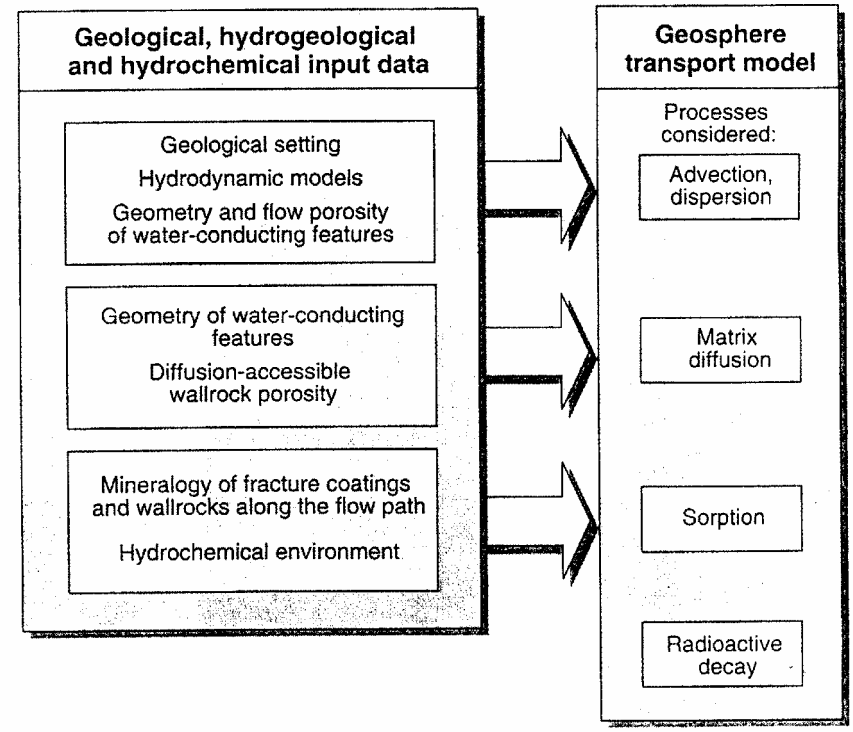


Figure 2-3. Input data and processes considered in performance assessment of geosphere transport by Gautschi (1995). With permission, NAGRA.

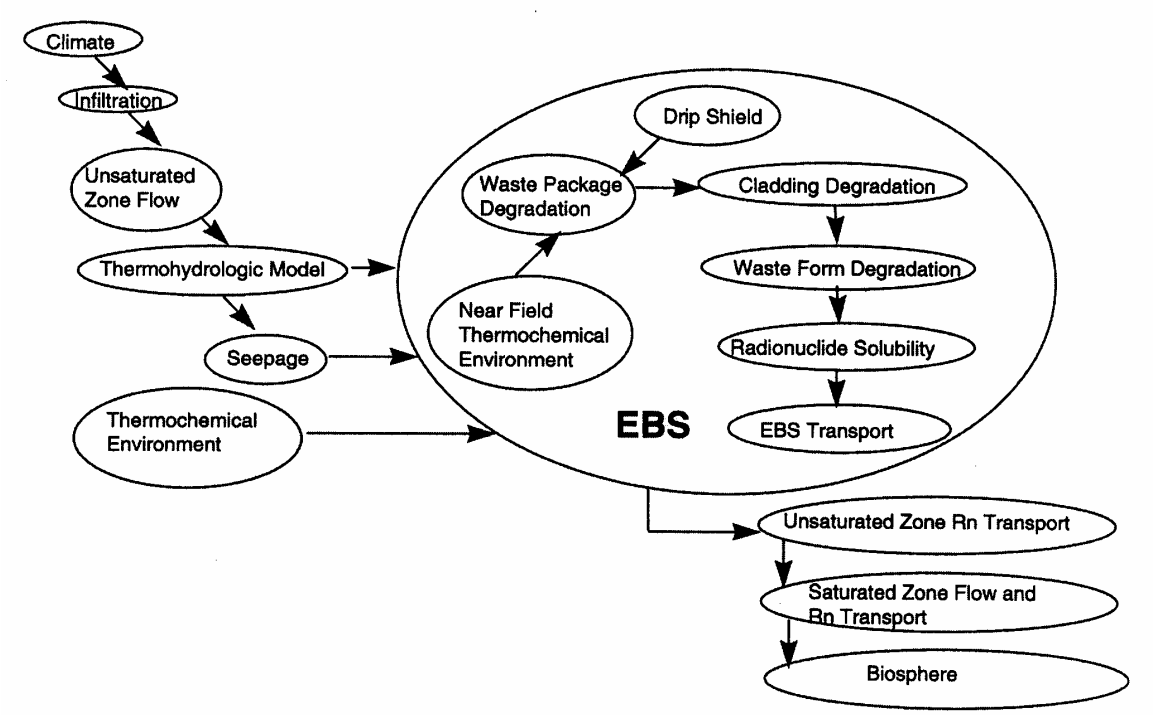


Figure 2-4. Processes affecting performance of an engineered barriers system (from USGS presentation to Expert Elicitation Panel, 1997).

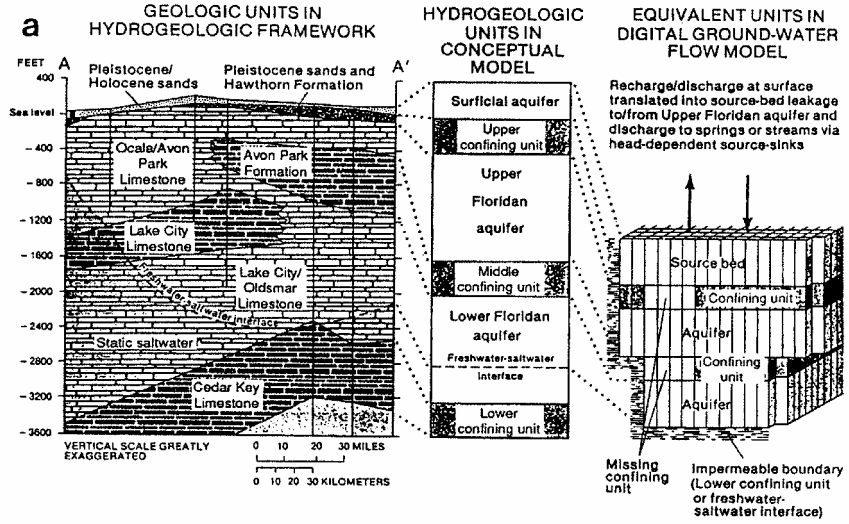


Figure 2-5. Translation of geologic information into conceptual model for numerical modeling: (a) Floridan aquifer system, W to E cross section for central Florida (after Anderson and Woessner, 1992). With permission, Academic Press.

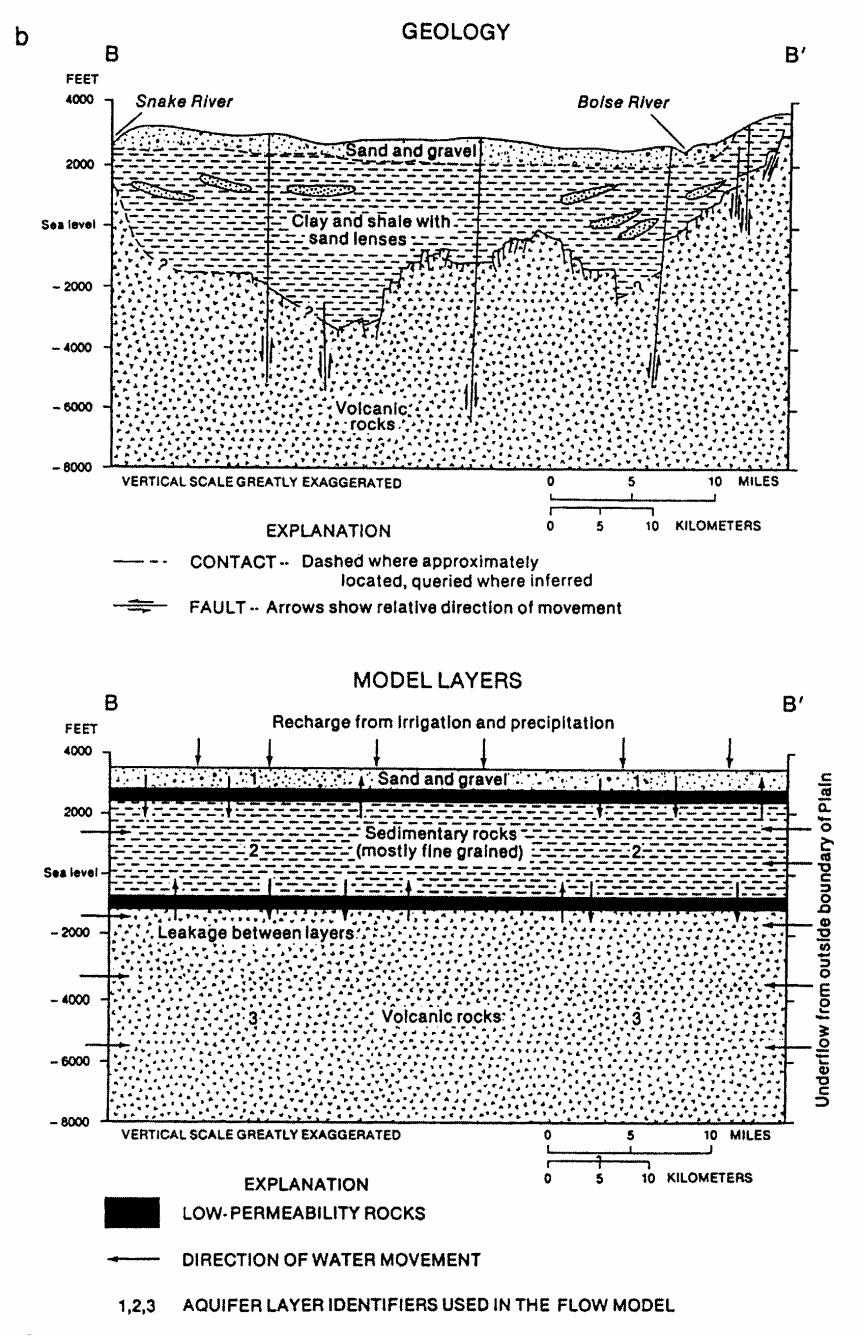


Figure 2-5 (Continued). Translation of geologic information into conceptual model for numerical modeling: (b) Snake River Plain aquifer system, southwest to northeast cross section; (after Anderson and Woessner, 1992). With permission, Academic Press.

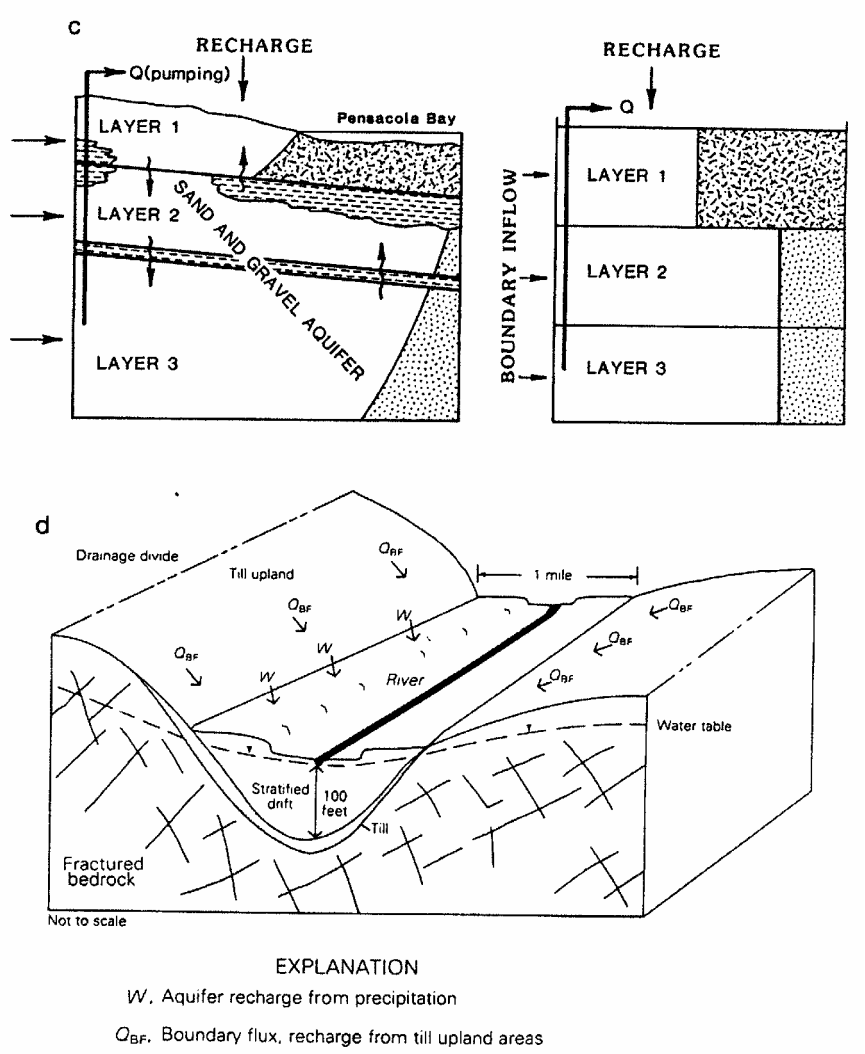


Figure 2-5 (Continued). Translation of geologic information into conceptual model for numerical modeling: (c) Sand and gravel aquifer in Pensacola, Florida; (d) Glacial-drift river-valley aquifer in Rhode Island (after Anderson and Woessner, 1992). With permission, Academic Press.

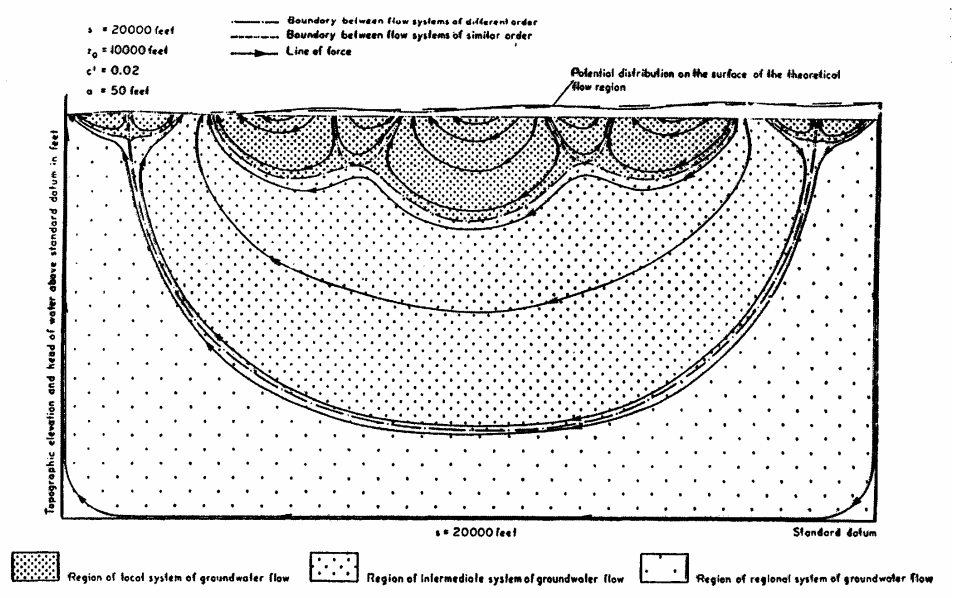


Figure 2-6. Theoretical flow pattern and flow systems according to Tóth (1963). With permission, AGU.

One of the earliest and best known pictorial hydrogeologic concept representations is the classic depiction, in Figure 2-6, of streamlines defining hydrogeologic systems and subsystems due to Tóth (1963). Other examples of pictorial concept representations include a hydrogeologic block diagram of arid zone hydrogeology in the southwestern United States (Figure 2-7), a map of groundwater flow and basin divides in the Death Valley area (Figure 2-8), and a conceptual model of crystalline basement with features relevant to flow, and its exploration, in Switzerland (Figure 2-9).

Meyer and Gee (1999) consider a conceptual model to be a hypothesis that describes the main features of site geology, hydrology, geochemistry, and relationships between geologic structure and fluid flow and contaminant transport. In their view, mathematical modeling is a process of testing this hypothesis. A conceptual model is a pictorial, qualitative description of the groundwater system in terms of hydrogeologic units, system boundaries including time-varying inputs and outputs, and hydraulic as well as transport properties including their spatial variability. A mathematical model constitutes a quantitative representation of such a conceptual model.

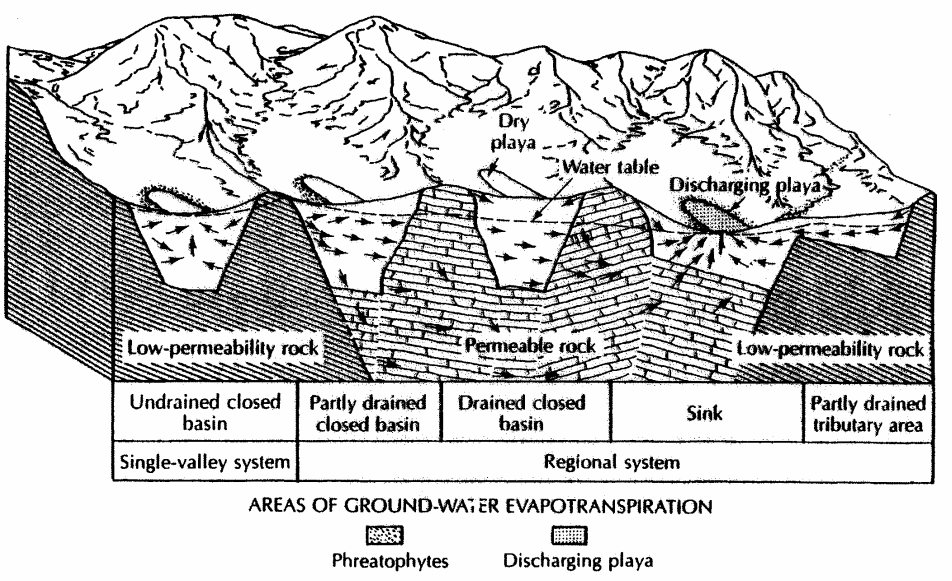


Figure 2-7. Hydrogeologic block diagram of arid zone hydrogeology (after Eakin *et al.*, 1976).

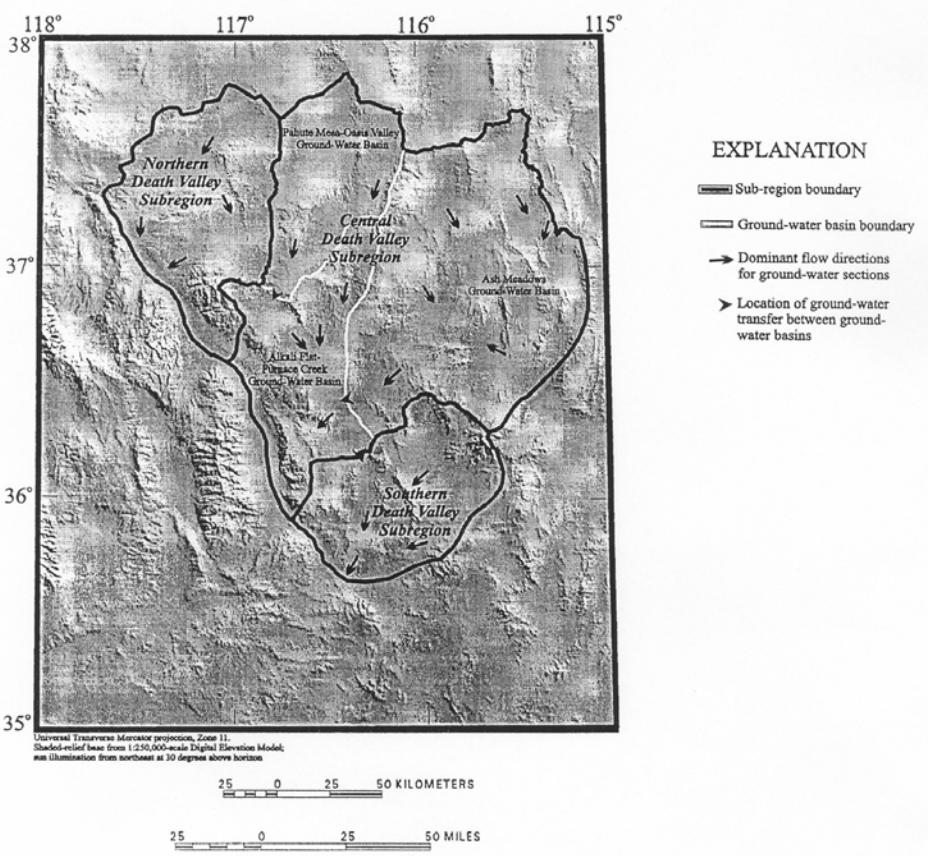


Figure 2-8. Map of groundwater flow and basin divides in the Death Valley area (after D'Agnese *et al.*, 1996).

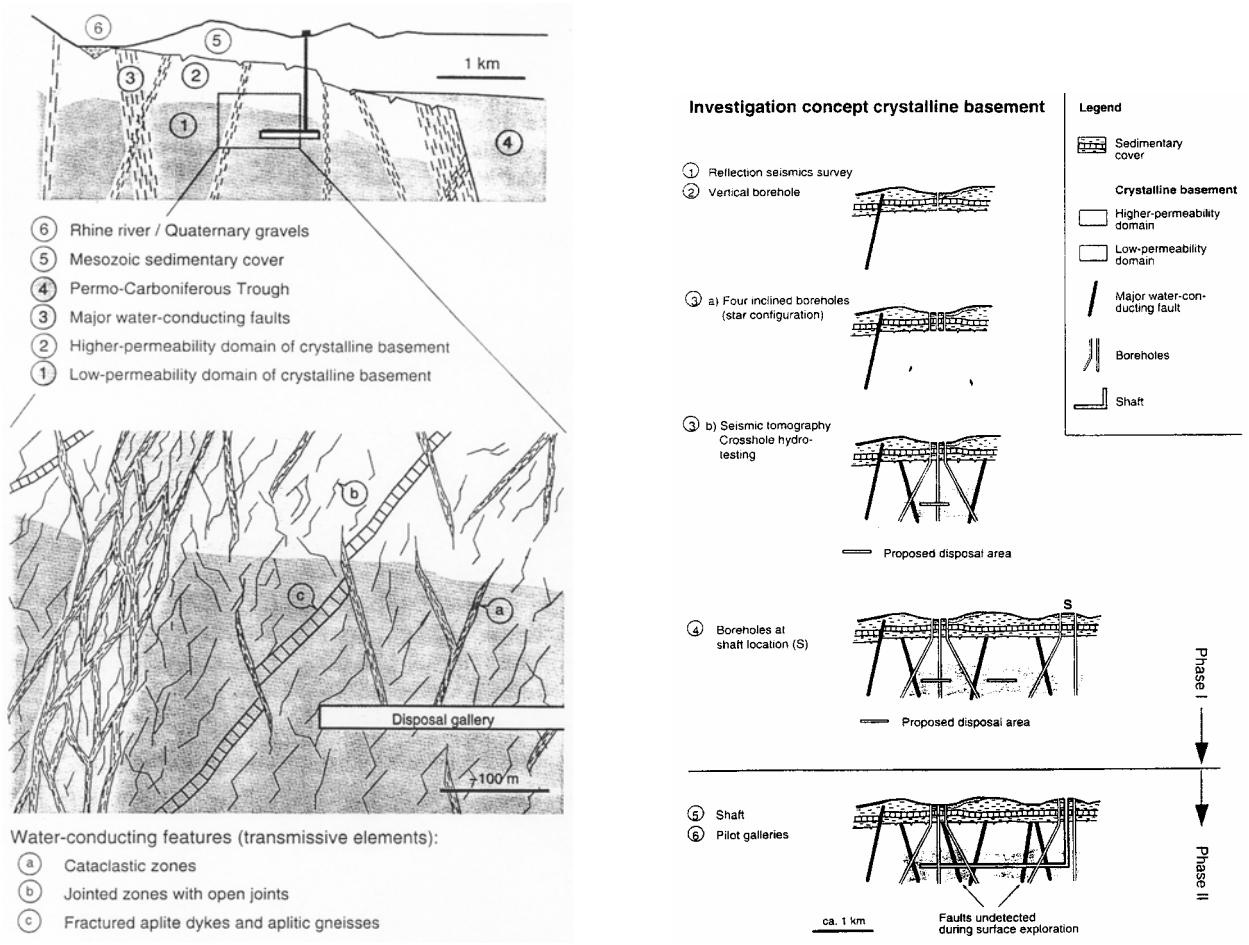


Figure 2-9. Conceptual model of crystalline basement in Switzerland with features relevant to flow, and corresponding concept of its characterization (after Gautschi, 1995). With permission, NAGRA.

Kolm and van der Heijde (1996) describe an integrated, stepwise method for the qualitative conceptualization and quantitative characterization of natural and anthropogenic subsurface envirochemical systems. They define the latter as hydrogeologic and hydrochemical systems that contain chemical species of concern to environmental management. According to the authors, a conceptual model of an envirochemical system includes a qualitative assessment of how chemicals enter, move through or are retained in and leave the system. It also includes a qualitative description of the source, transport, fate and distribution of each targeted chemical or, in the case of unknown sources, a hypothesis concerning source locations and strengths based on conceptualized transport and fate processes and known distribution of chemicals. The conceptual model is described and visualized using cross-sections and plan view illustrations. The authors do not present examples of conceptual models but instead illustrate, by means of flow charts (Figure 2-10), how their suggested procedure for the joint conceptualization and characterization of envirochemical and groundwater systems would operate.

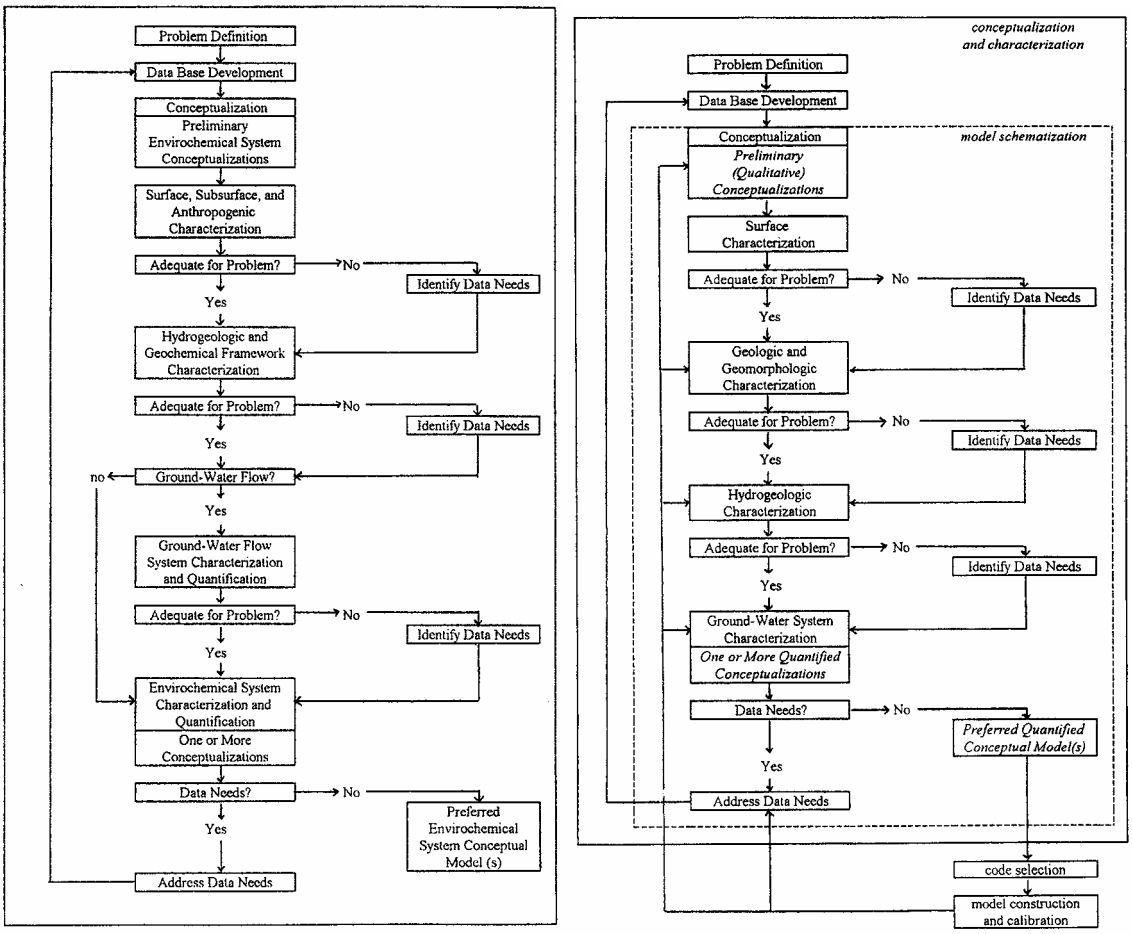


Figure 2-10. Procedure for joint conceptualization and characterization of envirochemical and groundwater systems according to Kolm and van der Heijde (1996). With permission, IAHS press.

It is of interest to note that conceptual models do not appear in the otherwise extensive index of a book by the National Research Council (1990) devoted to scientific and regulatory applications of groundwater models. However, they are the focus of a more recent study by a National Academy panel (NRC, 2001) of conceptual flow and transport models in the fractured vadose zone. The panel defines a conceptual model as an evolving hypothesis identifying the important features, processes and events controlling fluid flow and contaminant transport of consequence at a specific field site in the context of a recognized problem.

A conceptual model is qualitative and expressed by ideas, words and figures. It is a hypothesis because it must be tested for internal consistency and for its ability to represent the real system in a meaningful way. The hypothesis evolves (is revised and refined) during testing and as new information is gathered. Another study by a National Academy panel (NRC, 2000) considers conceptual modeling to include the representation of hydrogeologic heterogeneity and scale on the basis of field experimental data.

2.4 What is a Conceptual-Mathematical Model of Site Hydrogeology?

We stated that a hydrogeologic model consists of a description of the site and circumstances being modeled, a set of regional and site data to support this description, a mathematical framework that incorporates both the description and the data, and a mathematical (analytical or numerical) evaluation of system behavior and performance. A conceptual-mathematical model embodies the descriptive component of the model, cast in the form of mathematical equations suitable for system evaluation. The equations represent the symbolic framework, or structure, of the conceptual-mathematical model. Ideally, they provide a mathematical description of all physico-chemical (and other) processes that are considered relevant to flow and transport at the site on a given range of space-time scales subject to well-defined forcing. Defining the scales of the system includes specifying its spatial and temporal dimensions. Forcing includes sources, initial and boundary conditions. Suitability for system evaluation means that the symbolic conceptual-mathematical framework, or structure, is cast in a form that lends itself to computation and subsequent presentation (numerical, graphical) and analysis (deterministic, statistical) of the results.

Mathematical hydrogeologic models are described succinctly in Figure 2-11. Their essence is the formulation of conceptual models of site hydrogeology and flow/transport dynamics in mathematical language. This includes mathematical definition and description of relevant space-time frames, dimensions and scales, system topology and geometry, interactions (called processes) between kinematic (flux, velocity) and dynamic (force, stress, energy) quantities, forcing terms (initial and boundary conditions, sources), state equations (functional relationships between state variables), and corresponding parameters (their functional representations). It leads to a system of mathematical statements, relationships and equations (algebraic, differential, integro-differential, Boolean, deterministic, statistical, stochastic) that ultimately allow one to interpret and explain existing observations, and predict future conditions, events and observations related to site hydrogeology and corresponding flow/transport dynamics. These statements, relationships and equations constitute a mathematical conceptualization of the system, not merely a calculational tool. They complement the qualitative conceptual picture and, together with it, form a conceptual-mathematical framework or model for the hydrogeologic system. Whether such a model is analytical or numerical (written in the language of calculus or algebra) is merely a technical, not a fundamental, distinction.

For us this implies that one cannot legitimately disassociate hydrogeologic interpretations, or conceptual models, from corresponding mathematical and computational models for purposes of quantitative environmental impact and performance assessments of a given nuclear facility or site; these models are intimately linked and define a single conceptual-mathematical framework for quantitative site analyses. An example of how a given qualitative conceptual model may be associated with different conceptual-mathematical models, depending on scale, is given by Figure 2-12 for the crystalline basement of Northern Switzerland.

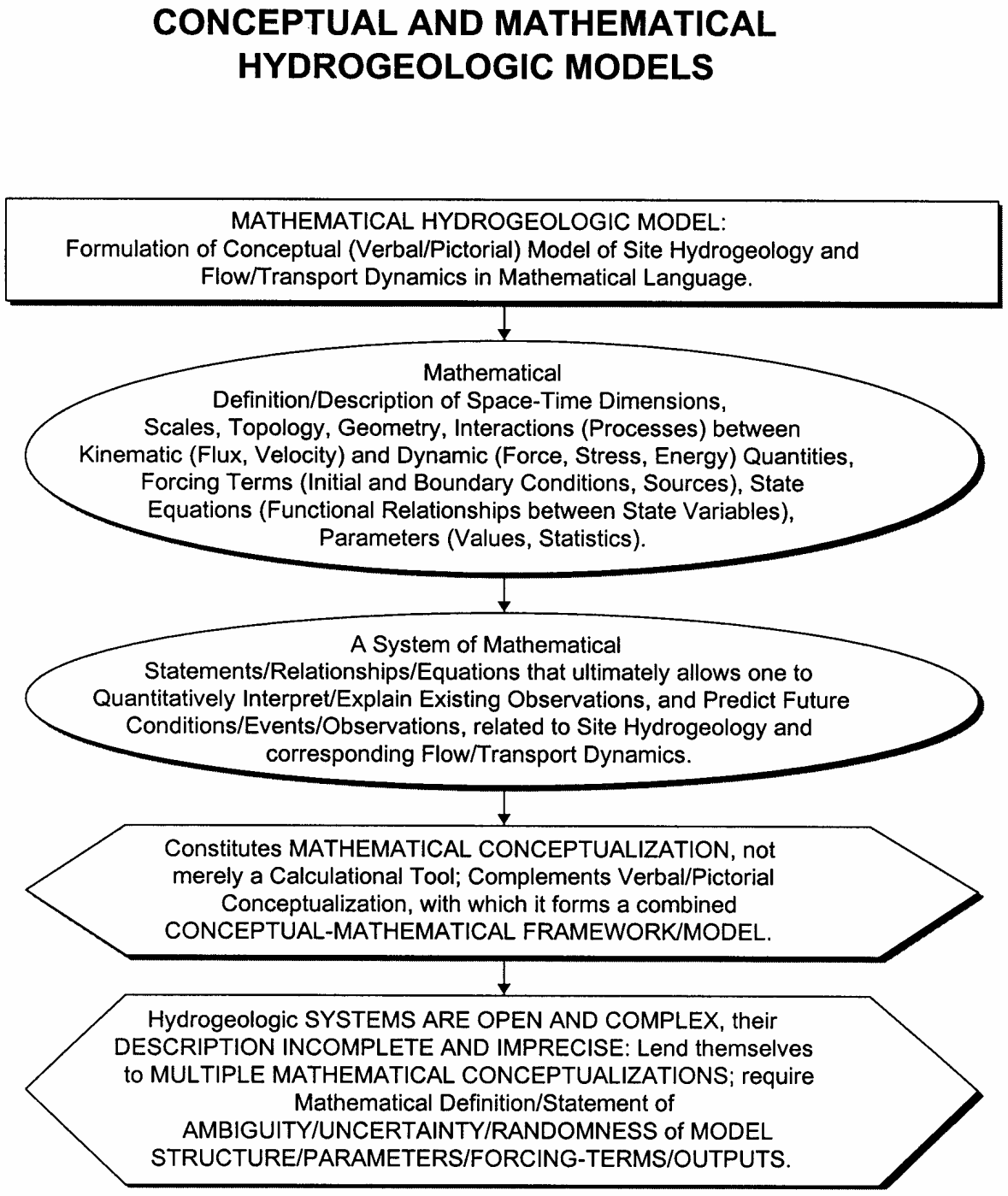


Figure 2-11. Mathematical formulation of hypotheses.

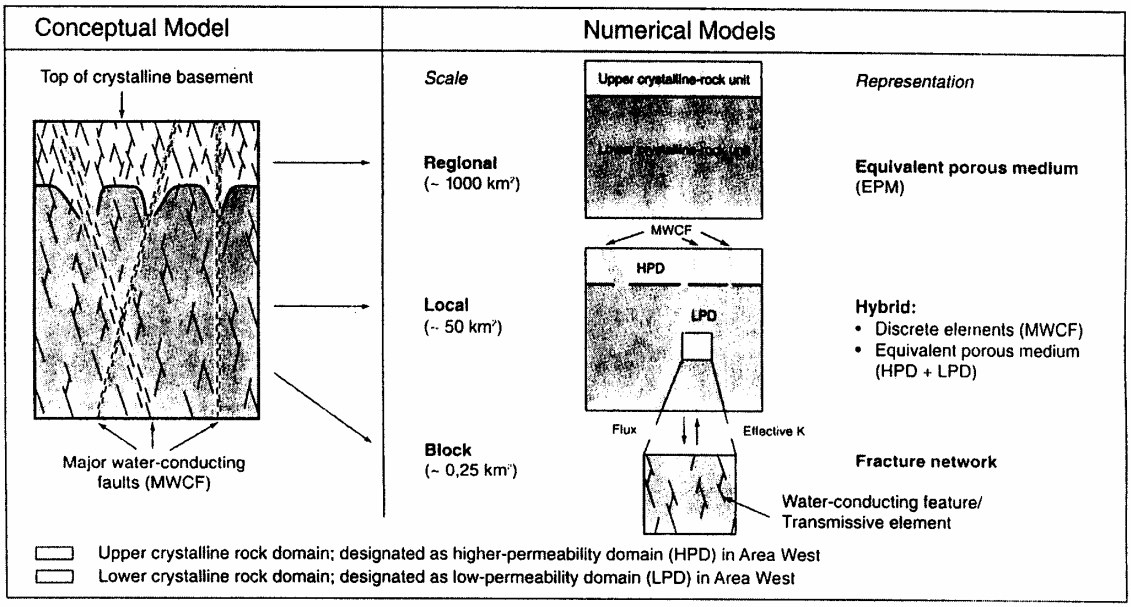


Figure 2-12. Different conceptual-mathematical models associated with a given qualitative conceptual model of crystalline basement in Northern Switzerland, depending on scale (after Gautschi, 1995). With permission, NAGRA.

We mentioned that a groundwater code is a tool and not a model. However, selecting a particular code implies identifying specific processes that may govern flow and transport at a site, their symbolic mathematical representation, and their numerical approximation. Selecting the space-time dimensions and size of a computational grid determines the dimensions and scale of the system being modeled. Choosing the sizes of space-time discretization intervals defines the scales at which flow and transport processes are resolved. Specifying the location and type of sources, initial and boundary conditions identifies the forcing. Both material properties and forcing terms are associated with parameters that must eventually be assigned numerical values. Choosing the modes of their representation (parameterization) defines the resolution scales of these parameters (and thus of material properties and forcing terms) in space-time. Once this has been accomplished in a way that is supported by all relevant regional and site data, the code has been transformed from a mere tool to a *bona fide* conceptual-mathematical model of site hydrogeology.

2.5 What is a Hydrogeologic Model?

A conceptual-mathematical model describes subsurface flow and transport in symbolic mathematical and algorithmic language. As such, it constitutes the mathematical framework, or structure, of a hydrogeologic model for a site. The framework contains input parameters that represent material properties and forcing terms (sources, initial and boundary conditions). Assigning numerical values to these parameters transforms the symbolic framework into a working mathematical model of the hydrogeologic system. This hydrogeologic model can then be run to simulate

(reconstruct or predict) and analyze space-time variations in quantities such as hydraulic head or pressure, solute concentration, fluid and solute flux and velocity, solute travel time and associated performance measures.

A hydrogeologic model thus consists of a conceptual-mathematical structure (which embodies a qualitative mathematical description of the system) and a set of parameter values (which render the model quantitative).

2.6 Non-Uniqueness of Hydrogeologic Models

As pointed out by Roman, Brooks and de la Torre (Physics Today, 1998), physical reality exists objectively but is not fully or directly accessible to us. Instead, it reveals itself to us via sensory impressions and experiences (through observations and experiments) in signatures or bits. To assimilate these separate pieces of information we describe, contemplate, correlate, connect, systematize, interpret and integrate them into a body of knowledge. On the basis of this knowledge we form mental constructs, or contextual structures, which we call conceptual models of reality. It has been recognized by Einstein that such constructs, or structures, cannot be extracted from experience but must be freely invented. It therefore follows that observed reality may lend itself not just to one but to multiple conceptualizations.

The same applies to hydrogeologic systems. The formulation of a conceptual model is inherently subjective in that it relies on limited available site observations and data, as well as experience and insights developed through work on similar sites and/or related problems; hydrogeologic conceptualization is susceptible to biases arising from the disciplinary background and experience of the analyst, and/or by different perceptions of the problem as influenced by external social and political factors (NRC, 2001).

Contrary to engineering systems which are generally closed, relatively simple and well-defined, hydrogeologic systems are open, complex and only partially defined. The open nature of hydrogeologic systems means that they are not amenable to fully controlled experimentation; as they are additionally complex, their description must remain forever incomplete and imprecise. Corresponding conceptual-mathematical models are therefore fallible scientific constructs, not credible engineering tools; they undergo a never ending process of modification, rejection and/or replacement as new scientific evidence and/or reasoning emerge. In other words, the conceptual-mathematical structure of a hydrogeologic site model is nonunique.

The latter has not been considered as a problem in the past (Beven, 2000). The attitude has been that if the modeling concepts were sufficiently accurate then it would be possible to represent the site uniquely by specific values of model parameters. In principle this might be possible; in practice it proves to be untenable because of the impossibility of fully describing the system, and particularly the heterogeneity of its subsurface characteristics.

Experience with models in a wide variety of hydrologic settings and contexts indicates that, regardless of the amount and quality of regional and site data available at a given site, it is generally neither possible nor justifiable to describe all relevant aspects of site behavior and performance by a unique

model. Instead, there is generally a wide range of model structures and parameter values that appear to "work" for a given site and purpose. In other words, neither the structure nor the parameter values of a hydrologic site model are unique. This "equifinality" problem (Beven, 1993, 2000) is generic and not simply one of identifying a system's "true" model structure or parameter values. In fact, a "true" model for a hydrologic system does not exist. It is therefore best (NRC, 2001) to consider a broad range of reasonable alternative hypotheses and base the model on a variety of different types of data.

2.7 Accounting for Hydrogeologic Complexity

2.7.1 Effects of Heterogeneity and Scale

Hydrologic systems are inherently heterogeneous on a multiplicity of scales (Entekhabi, 2000). The nonuniqueness of hydrologic models, applied to unique sites, arises from the limitations of current model structures in representing heterogeneous surface and subsurface flow systems, from limitations of measurement techniques and scales in defining system characteristics and the initial and boundary conditions for a model, and from the uniqueness of individual sites (Beven, 2000).

In recent years there has been a move towards more and more detailed studies of flow in rivers, soils and aquifers, using more and more sophisticated instrumentation (Beven, 2000). The result has generally been to demonstrate complexity of flow pathways, due to heterogeneity on different scales and interactions between the geometry of the flow domain and the prevailing hydraulic gradients and flow pathways (including both "dead zones" and "preferential" flows). It has proven difficult to use such information in applications of predictive models, which are generally required to be used at scales larger than the detailed measurements.

The problem is not so much that the complexity of the measurements is not accessible to adequate interpretation or even, in some cases, to being adequately modeled (Beven, 2000). The problem is that detailed characterization of a site is technically impractical and unfeasibly expensive. Even if this was practical and feasible, it would not necessarily lead to an improved representation of the system on larger scales. This is so because the effective behavior of a hydrologic system on larger scales is not completely determined by the scale-up of local processes; interactions among smaller-scale features and processes often lead to effects that are not completely captured by effective parameters in larger-scale models (Entekhabi, 2000).

There are processes and conditions when the effective parameter approach to scaling may be feasible. In the remaining circumstances, the macro-scale and micro-scale predictive relations for hydrologic processes may have different functional forms and dependencies (Entekhabi, 2000).

2.7.2 "Simple" Models of Complex Systems

The unique heterogeneous and scale-dependent characteristics of a site are inherently unknowable and not given to accurate or unique representation by a model. All one can do is to represent system complexities in simplified form and constrain the model so it provides an acceptably realistic representation of the system, as manifested by its consistency with available site data.

Demonstrating consistency with site data is important because a model is a valid reflection of reality only in the sense, and to the extent, that it yields qualitative and quantitative explanations and predictions of real experiences (Physics Today, 1998). Only if a hydrogeologic model is comparable with site characterization and monitoring data would it reflect actual and anticipated conditions at a site. Models may legitimately consider conditions and/or data that lie outside the observed range, but they must nevertheless remain compatible with available site data when applied within their range. Models which cannot be so applied may constitute appealing logical constructs which are however speculative and not subject to verification, confirmation or validation.

The level of hydrogeologic complexity that one incorporates in a model should be commensurate with its purpose (NRC, 2001). Some flow and transport phenomena may not be modeled with desired fidelity without accounting for system complexities that control them. Yet the details of these complexities may be deemed less relevant to the problem than their effect on flow and transport, and to characterize and model them may be considered too difficult and costly. One should then seek a model that captures the effects of these complexities implicitly, without reproducing them in detail explicitly.

A widely practiced approach has been to simplify complex hydrogeologic system for modeling purposes in an *ad hoc* and subjective manner (as illustrated for a popular dose assessment code in Figure 2-13). It is generally not clear that this approach captures adequately all aspects of site complexity that have a significant impact on the problem being tackled. It is also not clear that such models have the space-time resolution, which may be required to render them comparable and compatible with site data. Without this, the models remain interesting logical constructs which, however, may not be valid representations of actual site conditions and may thus yield unreliable predictions.

To insure that relevant aspects of hydrogeologic complexity are reflected in modeled system behavior and performance, and that the model is compatible with site data, it is important that the process of simplification be done systematically and objectively. This can be done by filtering out undesirable details through formal averaging of the governing equations in space-time or in probability space, in a way which retains and renders their influence on the model implicit.

Averaging three- or two-dimensional equations across one spatial dimension renders them two- or one-dimensional, respectively. Averaging transient equations over time may (but need not) render them representative of a steady state. Averaging can also be done over subdomains of the site being modeled, and over multiple time intervals. In each case, the averaging results in governing flow and transport equations that contain upscaled quantities. If the space-time scales of these quantities differ from those of the available site data, then either the model or the data must be rescaled to render them compatible and comparable with each other.

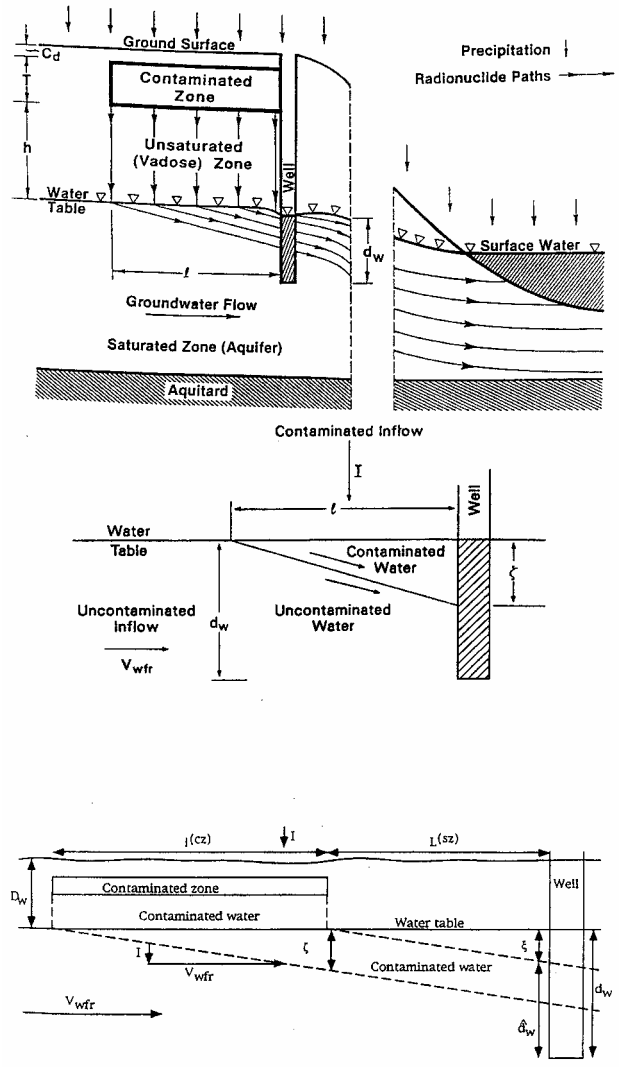


Figure 2-13. Simplified water pathways and well contamination processes in the RESRAD model (after Yu, 1993).

It has been traditional to assume that spatially and temporally averaged hydrologic quantities are less prone to predictive uncertainty than quantities that are localized in space-time. While this may be true for systems that are statistically homogeneous and stationary, such systems are the exception rather than the rule. Spatial and temporal averages are not necessarily more predictable as traditionally believed if the averaging covers a scale that contains a strong transition or change in behavior (Entekhabi, 2000).

This and other scale-related problems are generally avoided if the averaging is done in probability space. Such "ensemble" averaging leads to stochastic equations that contain statistical moments of hydrogeologic variables (considered random), most commonly the mean and variance-covariance. The mean is a predictor of

system behavior or performance, and the variance-covariance is a measure of predictive error. Both are smoother (vary more slowly in space-time) than their random counterparts and, in this sense, render the model relatively "simple." Despite their smoothness, both moments are defined on the same space-time scales as are the random hydrogeologic variables on which they are based. Stochastic models thus achieve smoothness and simplicity without any need to average or upscale in space-time. As they are typically conditioned on site measurements (*i.e.*, they honor the data), stochastic models are compatible with these measurements both in scale and magnitude. Yet another advantage of the stochastic method over space-time averaging is that it yields measures of predictive uncertainty. Stochastic approaches are increasingly recognized as offering a way to deal with complex, scale-dependent heterogeneous systems by means of relatively simple models (Hunt and Zheng, 1999).

3 CONTEXTUAL FRAMEWORK

The first step in model development is to define its context and purpose.

3.1 Purpose of Hydrogeologic Models Within NRC Regulatory Context

Within a NRC regulatory context, the purpose of hydrogeologic site models is to help one analyze, qualitatively and quantitatively, subsurface flow and transport at a site in a way that is useful for NRC staff review of decommissioning plans and performance of high-level and low-level radioactive waste disposal sites and uranium recovery facilities. For this purpose, the strategy developed in this report is cast in the context of a framework that is useful to NRC staff review and performance evaluation needs. The framework is defined in terms of performance measures identified by the NRC licensing staff (see Appendix A).

3.2 Questions that Hydrogeologic Models Must Address

The contextual framework defines key questions to which groundwater flow and transport models are expected to provide answers; it helps narrow down the problem as well as the range and type of potential model applications. The questions are stated in Figure 3-1 in the following order:

- For what ***purposes*** will the model be used? The answer includes defining the nature and magnitude of an existing and/or potential problem, location of the problem area, its causes, potential short- and long-term solutions and/or remedies, their anticipated consequences and costs-benefits, issues needing resolution, and criteria by which a solution and/or remedy will be selected.
- To what ***hydrogeologic system*** will the models be applied? The answer includes defining the hydrogeologic environment that is or may potentially be affected by the problem and/or its solution or remedy, the corresponding site and surrounding environs of (potential) concern, and the regional hydrogeologic setting.
- Under what ***circumstances and scenarios*** will the modeled hydrogeologic system operate? This requires defining predevelopment, current and potential undisturbed and disturbed site and regional conditions, and corresponding natural as well as anthropogenic influences, under which flow and transport would be modeled.
- What ***measures*** will be adopted to assess ***performance*** of the hydrogeologic system? This requires specifying measures and/or criteria related to site and regional hydrogeology that would be used to identify issues, potential solutions and/or remedies, their cost-efficiency and ability to meet regulatory requirements.
- What aspects of site and regional ***hydrogeology*** and ***flow/transport dynamics*** are expected to ***impact*** these ***performance measures***, and how? The answer involves defining key elements of the hydrogeologic system, key flow and transport mechanisms, key natural and anthropogenic influences, and space-time scales that may potentially impact these measures, as well as the manners in which such impacts might occur.
- With what ***reliability***, ***certainty*** and ***accuracy*** does one need to predict ***performance measures***? ***How important*** a role will such measures play in the

decision process? What is the ***worth*** (in time/effort/resources) ***of assessing performance*** measures for the site ***at specified levels of reliability, certainty, and accuracy***? These are **the most basic questions** that must be addressed before any site investigations and modeling efforts are initiated, as the answer would have a major impact on the time, effort and resources that could validly be expended in pursuit of performance assessment for the site.

3.3 Performance Measures

The strategy in this report focuses on groundwater flow and transport primarily in the sense, and to the extent, that they impact performance measures. Such measures are typically articulated within the broader context of a system in which groundwater is only one among several components. However, our strategy focuses solely on groundwater aspects of these broader criteria. These have been identified by the NRC staff in Appendix A as a "Note on Performance Measures and Hydrology Issues Pertaining to Groundwater Flow and Transport Models of Licensed Nuclear Sites." Appendix A provides information about performance measures based on regulatory standards, pertinent hydrology and groundwater issues that are related to performance, and uncertainty commonly encountered in groundwater flow and transport modeling. The NRC staff note focuses on low-level waste, decommissioning, high-level waste and uranium recovery sites. It emphasizes that the end point of the analysis is the performance measure for the licensed facility obtained through a performance assessment. As outlined in Appendix A, the hydrologic measures and issues need only be explored and dealt with to the extent required by performance assessment, taking into account the significant uncertainties and the hazard involved, and that some uncertainties may be satisfactorily addressed through bounding analyses.

Our interpretation of Appendix A is that the following groundwater-related performance measures need to be assessed by means of hydrogeologic flow and transport models.

Performance Measures for Low-Level Waste Disposal Sites:

- Concentrations of radioactive materials at receptor locations contributing to annual dose to an average member of a critical group.
- Concentrations in well water at site boundary resulting in the highest composite dose.
- Cumulative releases to the environment to be As Low As Reasonably Achievable (ALARA).

Performance Measures for Decommissioning Sites:

- Residual radioactivity resulting in Total Effective Dose Equivalent (TEDE) to an average member of a critical group at receptor locations, or human access points, including drinking water.
- Radionuclide concentrations that contribute toward the dose received by an individual who uses well water on site, or by an individual off-site who uses water from a well at the site boundary.
- Cumulative releases to the environment ALARA.

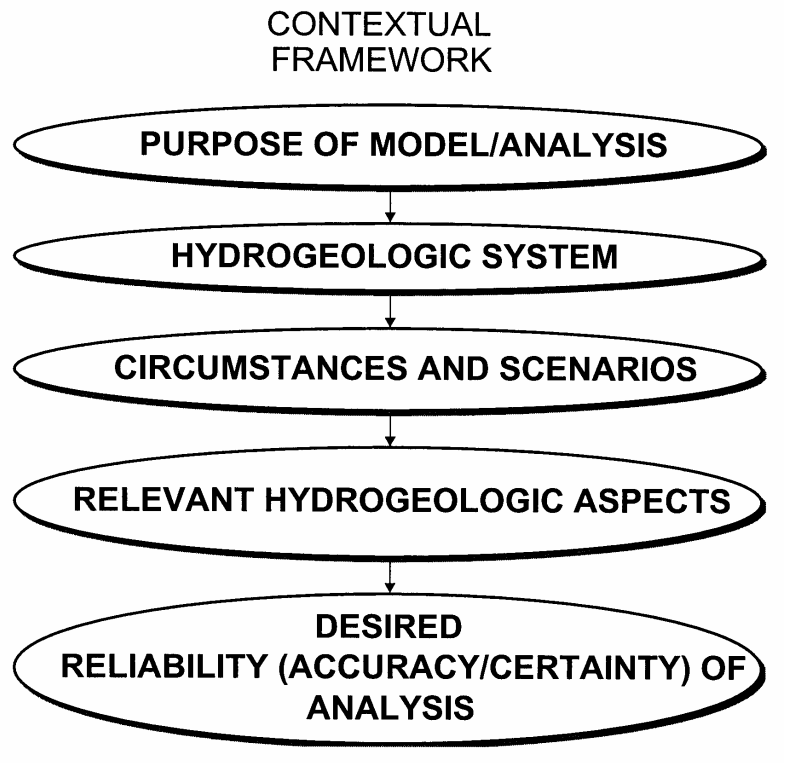


Figure 3-1. Contextual framework of modeling.

Performance Measure for High-Level Disposal at Yucca Mountain:

- Radionuclide concentrations that contribute to the expected dose or TEDE received by an average member of a critical group 20 km down gradient of the proposed repository.

Performance Measures for Uranium Recovery and Tailing Sites:

- Concentrations of specified chemicals and radionuclides.

At DOE mill tailing sites under Title I

- The model is to provide a technical basis for selecting a restoration strategy and for determining risk to humans and the environment.

At non-DOE mill tailing sites under Title II

- Maximum concentrations of selected constituents and chemical parameters.
- The model is aid in guiding and supporting
  - ✓ Detailed groundwater characterization,
  - ✓ Specific criteria for monitoring and restoration.

At In-Situ Leach (ISL) uranium extraction sites

- Monitoring to assure restoration of ore zone and affected aquifers, based on background levels and EPA standards.
- Restoration goals for each constituent on a well-by-well or well-field average basis.

- Primary restoration goal of returning the ore zone and aquifers to their preoperational (baseline) levels, which could rely on statistical measures such as average parameters, range of baseline concentrations, and 99% confidence intervals.
- Secondary restoration standards of returning the ore zone and aquifers to a pre-ISL class of use such as drinking water, livestock, agricultural or limited use.
- If a parameter cannot be restored to a secondary standard, demonstrate that it would not
  - ✓ threaten public health and safety or the environment,
  - ✓ significantly degrade water use.

3.4 Hydrogeologic Context of NRC Performance Measures

3.4.1 Hydrogeologic Aspects of Performance Criteria

The above performance criteria establish the contextual framework within which the strategy in this report is developed.

From a hydrogeologic point of view, the main differences between the various performance measures relate to the spatial and temporal scales of the problem; the type of environment (whether saturated or unsaturated, confined or unconfined, porous or fractured); the driving mechanisms (whether infiltration or pumping, recharge or discharge); the space-time resolution needed to accurately assess performance measures; the practicality and cost of attaining such a resolution; the potential effect of a mismatch between required and attained resolution on the accuracy of the computed performance measures; and the level of inaccuracy and uncertainty that performance assessment can tolerate.

Common to all the above performance measures is that they involve complex, three-dimensional, heterogeneous hydrogeologic environments; three-dimensional, generally transient groundwater flow regimes; open flow systems with indeterminate driving (forcing) terms (initial conditions, recharge and discharge, other source terms and boundary conditions); severely limited access to, and possibility to explore, the interior of the groundwater flow system; the uncertain nature and distribution of contaminants and their sources both at and below the soil surface, on site and off site; the lack of adequate science and technology to render reliable long-term predictions of pollutant migration and dilution with groundwater; and the difficulty in verifying such predictions by direct measurement.

3.4.2 Implications vis-à-vis Hydrogeologic Conceptualization and Modeling

The commonalties just identified among the various performance criteria imply that all of them require developing a good understanding, and reliable models, of subsurface flow and transport. One cannot do so without recognizing and considering the full complexity of a site before attempting to represent it by means of a simplified conceptual-mathematical model. This is true regardless of the specific performance criteria one needs to address.

The process of model simplification or abstraction must be based on a prior decision by the hydrogeologist and performance assessment analyst on the degree of reliability, certainty and accuracy with which given performance measures need to be predicted, their importance in the decision process, and the amount of time, effort and resources that assessing them at

the specified levels of reliability, certainty and accuracy would justify in each specific case. These questions are of a regulatory nature and cannot be addresses by hydrologists without a case-by-case dialogue with regulators. As such, they are not addressed by the strategy in this report. Instead, the strategy stipulates that these fundamental questions must be addressed and resolved by the NRC licensing staff before any site investigation and modeling are initiated. This is so because the answers impact in a major way the degree to which one justifies the use of a simplified conceptual-mathematical model for what is ubiquitously a complex hydrogeologic system with three-dimensional, transient groundwater flow and contaminant transport.

Therefore, it is very important that the hydrogeologist first articulated these processes in their naturally complex setting as best understood in light of available site data and the state of prevailing hydrogeologic knowledge prior to postulating groundwater flow and transport models for performance assessment of a given site. Even if performance assessment is ultimately conducted with the aid of highly simplified conceptual-mathematical models of groundwater flow and transport, the strategy deems it essential for the credibility of the assessment that these models derive objectively (and if possible formally) from a more complete description of site hydrogeology, based on clearly reasoned and properly defended arguments.

Rather than taking the attitude that a limited performance assessment goal justifies a limited view of hydrogeology, the strategy adopts the attitude that a comprehensive description of hydrogeology is required to properly adapt a hydrogeologic model to such a limited goal. This manifests itself as a key criterion and guiding principle for the construction of conceptual-mathematical models in the strategy.

4 ASSEMBLY AND ORGANIZATION OF HYDROGEOLOGIC KNOWLEDGE BASE

Having determined the context and purpose of a model, one must next assemble and organize a corresponding knowledge base.

4.1 What Constitutes a Hydrogeologic Knowledge Base?

As illustrated diagrammatically in Figure 4-1, a hydrogeologic knowledge base includes all relevant (to the contextual framework) qualitative and quantitative data, observations, concepts, theories and models which pertain to past, current and anticipated groundwater flow and transport conditions and system behavior at and around a site, of both a generic and a site-specific nature. It includes information concerning the accessibility and reliability of data and knowledge. Key data categories include site and regional physiography, topography, climate, meteorology, soils, vegetation, land use, geomorphology, geology, geophysics, surface and subsurface hydrology, inorganic and organic hydrochemistry, radiochemistry, natural and anthropogenic isotopes, remotely sensed data, *etc*. Of special relevance to hydrogeologic model development are regional and site data that allow one to define the distribution of hydrostratigraphic units on a variety of scales; their geologic structure; rock and soil types; their textural, physical, flow and transport properties; fluid types; their state of saturation, pressure, temperature and density; chemical constituents and isotopes; major contaminants in soil, rock and groundwater; and their sources.

Key knowledge categories include available concepts, methods and tools to integrate, manage, access, visualize, analyze, interpret and process such data in a manner suitable for flow and transport analyses under uncertainty, to perform such analyses, and to assess the corresponding ambiguities and uncertainties at various levels of hydrogeologic, conceptual and mathematical complexity in various contextually relevant hydrogeologic environments.

4.2 Why Does One Need Site Characterization and Monitoring Data?

Hydrogeologic modeling must not take place in a vacuum. All sites are unique and so cannot be validly represented by a generic model. Instead, hydrogeologic models must be solidly grounded in a broad array of regional and site data to validly represent particular locales.

Site characterization data form the foundation on which one postulates one or more conceptual-mathematical models for an area and assigns initial values to their input parameters (*i.e.*, material properties and forcing terms such as sources, initial and boundary conditions). To test and compare these models among themselves qualitatively and/or quantitatively, one also needs monitoring data that constitute observations of actual hydrologic behavior at and around the site. Only with such data can one evaluate the ability of models to mimic real system behavior (qualitatively at the conceptual level, quantitatively at the conceptual-mathematical level), improve their ability to do so through calibration against the monitoring data, determine their optimum degree of refinement or complexity, rank them and assess their cumulative impact on predictive uncertainty, and compare them with each other.

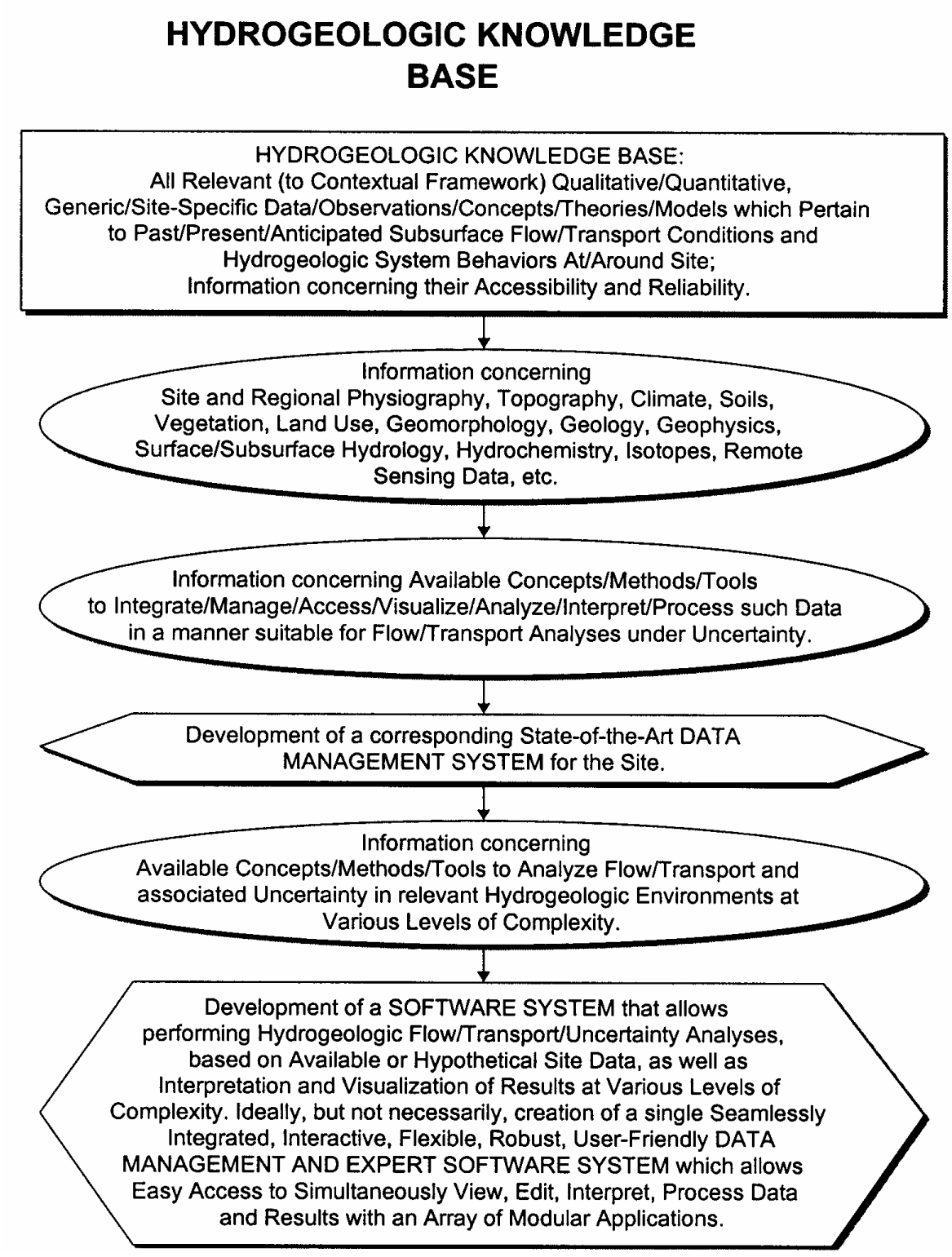


Figure 4-1. Content of knowledge base.

The strategy in this report recognizes that site characterization and monitoring data are expensive and difficult to collect, leading to a ubiquitous scarcity of hard site information. It is therefore critically important to assess the role that such data play in rendering the hydrogeologic performance analysis credible. The strategy stresses the role of characterization and monitoring data in helping one identify and test alternative conceptual models, make rational choices among them, gauge and reduce model bias and uncertainty through proper model selection and calibration, assess the reliability of model predictions, and confirm the assessment through independent peer review as well as at least some degree of direct verification.

4.3 Sources of Hydrogeologic Information

Sources of site-specific information include technical reports, maps, charts and data files from local, state and federal agencies and institutions, including universities and their research arms; corresponding sites on the Internet; and visits to the site and its surroundings. Sources of generic information include textbooks, monographs, reference books and compendia on hydrogeology; soil physics; the dynamics of flow and transport in porous and fractured media; testing, sampling and monitoring of soils and aquifers; groundwater modeling; isotope hydrology; and hydrogeochemistry. Additional sources of generic information include research and review papers and reports pertaining to specific aspects of the above topics.

4.3.1 Example: The Arizona Water Information Directory

An excellent example of how one can identify sources of information relevant to hydrogeologic analysis and modeling is provided by the Arizona Water Information Directory (Tellman, 2001). This guide to locating agencies, organizations, and university specialists with information about water in Arizona is based in part on a 1998 publication titled "Where to Get Free (or Almost Free) Information About Water in Arizona" by Tellman. It documents an extensive and broad range of available resources for water information in Arizona, many of which (primarily federal agencies) may also serve as sources of information for other parts of the United States. The publication is free upon request from the University of Arizona Water Resources Research Center, whose web site (www.ag.arizona.edu/azwater/) also supports complex search queries and is updated periodically as new information becomes available.

Examples of data sources related to geology include U.S. Geological Survey (USGS) (www.usgs.gov), Arizona Geological Survey (www.azgs.state.az.us) and University of Arizona Institute for the Study of Planet Earth (www.ispe.arizona.edu). Data concerning aquifers can be obtained from the USGS, Arizona Department of Water Resources (www.water.az.gov), U.S. Environmental Protection Agency (U.S. EPA) (www.epa.gov), and U.S. Water Conservation Laboratory (www.uswcl.ars.ag.gov). Well data are available from the Arizona Department of Water Resources, Arizona Small Utilities Association (www.asua.org), and Salt River Project (www.srpnet.com). Information about soils is available from the Arizona Department of Agriculture (www.agriculture.state.az.us) and Cooperative Extension (www.ag.arizona.edu/extension). Data related to climate and weather can be obtained from the U.S. Natural Resources

Conservation Service (www.az.nrcs.usda.gov), U.S. Water Conservation Laboratory, Southwest Watershed Research Center (www.tucson.arg.ag.gov), Arizona State University Office of Climatology (www.geography.asu.edu/climatolog), and University of Arizona Institute for the Study of Planet Earth. Evaporation and evapotranspiration data are obtainable from the Arizona Meteorological Network (AZMET) (www.ag.arizona.edu/azmet) and Southwest Watershed Research Center. Streamflow data are available from the USGS, U.S. Forest Service (www.fs.fed.us) and Salt River Project. Numerous additional sources of information are listed by various categories.

4.3.2 Example: Maricopa Agricultural Center

In the process of developing this strategy, we have investigated experimentally groundwater flow and transport in unsaturated and saturated soils at the Maricopa Agricultural Center near Phoenix, Arizona. To support this investigation, we relied on several publicly available sources of information. These included reference maps downloaded from the University of Arizona library (www.srnr.arizona.edu/nbs/gap/nbiidata.html), meteorological data from the AZMET Network, distribution of soils and irrigated land from the Arizona Land Resource Information System (ALRIS) (www.land.state.az.us/alris/htmls/data2.html), as well as irrigation rates, distribution of wells, well logs, well construction data, pumping rates, and groundwater levels from archived records of the Arizona Department of Water Resources and/or the U.S. Geological Survey.

4.3.3 Example: EDR Environmental Data Resources, Inc.

EDR Environmental Data Resources, Inc. (www.edrnet.com), is a good commercial source of environmental information. EDR offers current and historical environmental risk management information, industry publications and market research, newsletters, a daily business news service, training workshops, and state-of-the-art online services including interactive mapping.

The service allows one to search a variety of government records for information about specified target areas within a given radius. Examples include topographic maps and gradients from the U.S. Geological Survey; flood zone data from the Federal Emergency Management Agency; information about wetlands from the U.S. Fish and Wildlife Service; groundwater flow direction, velocity and depth to water table from published sources; geologic age and rock stratigraphic units from the USGS; soil maps and soil survey data from the U.S. Department of Agriculture Soil Conservation Service; locations of water wells and springs from the USGS, U.S. EPA and state agencies; as well as other environmental data.

4.4 Data Management and Expert Software Systems

Ideally, the knowledge base would be contained in an integrated, interactive, flexible, robust and user-friendly data management and expert software system which allows easy access to each member of a multidisciplinary team to simultaneously view, edit, and interpret available and/or hypothetical site data and results through all phases of a project; with applications organized by discipline, each with a variety

of modules, which work together and share a common data base; and with seamless access to data as well as to flow and transport analytic and interpretive tools of varied types and levels of complexity.

4.4.1 Example: GMS

The Groundwater Modeling System (GMS) (www.scisoft-gms.com) is a sophisticated package for groundwater modeling, marketed by the Scientific Software Group (www.scisoftware.com) and other commercial entities. It provides tools for various phases of the modeling process including site characterization, model development, post-processing, calibration, and visualization. GMS supports triangulated irregular networks (TINs), solids, borehole data, two- and three-dimensional geostatistics, and both finite element and finite difference models in two and three spatial dimensions. Currently supported groundwater flow and transport codes include MODFLOW, MODPATH, MT3D, RT3D, FEMWATER, SEEP2D, SEAM3D, PEST, UCODE and UTCHEM. Due to its modular nature, GMS can be customized to include selected modules and interfaces.

GMS comes in identical PC and UNIX versions. It allows one to directly import data from, or export data to, ARC/INFO, ArcCAD, and ArcView. GMS makes it relatively easy for the user to quickly develop a conceptual model and a corresponding numerical model for the area being studied. For example, a TIFF or JPEG image of an aerial photo/scanned-in map, or an AutoCAD or MicroStation DXF drawing of the site can be displayed as background images allowing the user to define points, polylines, and polygons to represent spatially associated modeling data. Boundary conditions and parameter values can be directly assigned to these graphical entities. Points can define well pumping data or point sources for contaminants; polylines can define rivers, drains, or model boundaries; and polygons can define areal data such as lakes, differing recharge zones or hydraulic conductivities. Once the conceptual model has been defined, GMS will construct a grid, automatically refined around wells with the cells outside the model boundary already deactivated. The defined modeling data is then superimposed onto the grid with the appropriate parameters. For example, conductances assigned to polylines such as drains and rivers are automatically computed according to the length of the polyline segment within each cell. At this stage, the model is completely defined and no cell editing is required. If the user decides to change the conceptual model (move a boundary, add additional wells, etc.), these changes can be made quickly. Drawing tools are also provided, which allow the user to draw text, lines, polylines, arrows, rectangles, *etc.*, in order to add annotation to the graphical representation of the model.

GMS allows the user to construct TINs and solid models, and display borehole data. TINs are formed by connecting a set of x-y-z points (either scattered, gridded, or from boreholes) with edges to form a network of triangles. TINs can be used to represent the surface of a geologic stratum and can be displayed in oblique view with hidden surfaces removed. Three-dimensional models of stratigraphy can be constructed using solids and cut to create cross sections anywhere on the grid. The solid model can be shaded to generate realistic images.

GMS allows one to manage borehole data for site characterization. A borehole can contain either stratigraphy data or sample data or both. Stratigraphy data are used to

represent soil layers that are encountered in a soil boring. The soil layers are represented using contacts and segments. A segment represents a soil layer, and a contact is the interface between two segments. Contacts and segments can be used to construct TINs, solids and 3D finite-element meshes. Sample data represent data obtained by continuous sampling along the length of the hole. Cone penetrometer data and down-hole geophysical data are examples of sample data. Sample data are stored in data sets that can be manipulated as other data sets in GMS. For example, sample data from a cone penetrometer test may include data sets for tip resistance, sleeve resistance, and friction ratio. Sample data can be converted to scatter points, which can be interpolated to a three-dimensional grid, or mesh from which isosurfaces and color-shaded contours can be generated. Sample data can also be used to infer soil stratigraphy.

With GMS, one can interpolate from groups of scattered data in two or three dimensions to other objects (meshes, grids, TINs). Two-dimensional interpolation is used (for example) to generate transmissivities for a layer of a three-dimensional grid. Three-dimensional interpolation is used (for example) to assign initial conditions (head and/or concentrations) to each node in a three-dimensional grid. Interpolation schemes supported by GMS include linear (2D), inverse distance weighting (2-3D), Clough-Tocher piece-wise cubic patch based on finite-elements (2D), natural neighbor based on Thiessen polygons (2D) or polyhedrals (3D), ordinary and universal kriging using the GSLIB software package with graphical (anisotropic) variogram editing, and log interpolation. One can also interpolate from groups of 3D scatter points to any of the other data types (meshes, grids, TINs).

The GMS user interface is divided into ten separate modules. A module is provided for each of the basic data types supported by the system. These include a Triangulated Irregular Network (TIN) Module, Borehole Module, Solid Module, 2D Mesh Module, 2D Grid Module, 2D Scatter Point Module, 3D Mesh Module, 3D Grid Module, 3D Scatter Point Module, and Map Module. Switching from one module to another can be done instantaneously to facilitate the simultaneous use of several data types when necessary. To provide a consistent interface for all the modules, GMS associates generic scalar or vector data set values with any object. Each data set can be either steady state or transient. TINs, meshes, grids, and scatter point sets all have an associated list of scalar data sets and a list of vector data sets. Boreholes have a list of scalar data sets. Each set has a single vector or scalar value for each node, cell, borehole sample point, or scatter point. Data sets can be used to represent various quantities such as total heads computed by a groundwater model or starting heads used as initial conditions for input to a transient groundwater model. Data sets can be imported from a file or created by interpolation from a group of scattered points. The data can be altered through various mathematical operations and contoured or displayed as isosurfaces.

A variety of tools are available in GMS for panning, zooming, and rotating 3D objects. Editing and model interaction can take place in any view. Visualization options range from 2D contour plots to 3D isosurfaces. A groundwater model can be displayed in plan view or 3D oblique view and rotated interactively. Cross sections and fence diagrams may be cut arbitrarily anywhere in the model. Hidden surface removal and color and light source shading can be used to generate high-quality images. Contours and color fringes can be used to display the

variation of input data or computed results. Cross sections and isosurfaces can be interactively generated from 3D meshes, grids and solids, allowing the user to quickly visualize the 3D model. Both steady state and transient solutions can be displayed in animation using vector, isosurface, color fringe or contour animation. For example, animation of a transient solution allows the user to observe how head, drawdown, velocity and contaminant concentration vary with time. In addition, GMS can also sweep an isosurface through the 3D model. The minimum and maximum isosurface values are determined from the model and the program will then linearly interpolate and display multiple isosurfaces in rapid succession. This allows the user to quickly understand the spatial variation of a contaminant plume. Film loops are saved in Microsoft Video for Windows (*.avi) format and can be played back outside of GMS using almost any multimedia player or presentation package such as Microsoft PowerPoint.

4.4.2 Example: FRAMES

A Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES) is being developed by Pacific Northwest National Laboratory (www.pnl.gov/hightechcomm/license/programs/frames.pdf). FRAMES is an open-architecture, object-oriented system that provides an environmental database. It aids decision-makers and environmental users in constructing a Conceptual Site Model that is real-world based. FRAMES provides a platform to link the users' preferred codes with others that are required to perform environmental assessment, analyze the results and display them graphically.

FRAMES contains sockets for a collection of computer models that simulate elements of a source, fate and transport, exposure, and risk-assessment system. Currently available sockets include

- Contaminant Source and Release to Environment
- Overland
- Vadose Zone
- Saturated Zone
- Atmospheric
- Surface Water
- Exposure
- Dose
- Human Health Impacts
- Sensitivity/Uncertainty

Future planned sockets will include

- Ecological Impacts
- GIS
- Remediation Technology
- Cost Analysis
- Data Quality Objective
- Life-Cycle Management
- Conceptual Site Module

FRAMES provides file specifications that describe how all site information is stored within the framework and passed between modules. These file specifications are not associated with the model-specific information, only with the transfer of information to the outer framework or another modules.

With a Windows-based plug-and-play user interface, the user builds a Conceptual Site Model using media icons to represent the flow of contamination through the environment. Then, the user selects the simulation models to be used for the analysis. FRAMES allows the user to produce multiple unique analyses with one software tool.

FRAMES' modularization produces several types of time varying outputs including

- Contaminant mass remaining at the source
- Contaminant fluxes from the source
- Atmospheric concentrations and soil deposition
- Intake or dose
- Hazard quotient or risk

The platform allows one to

- Evaluate protectiveness for hazardous waste sites (U.S. EPA)
- Support D&D license termination rule (U.S. NRC)
- Link models related to both human health and ecological impacts from past and current practices at government installations (U.S. Department of Defense)
- Provide flexible environmental modeling tools for assessing human health and environmental compounds for a waste site, program, installation, and complex wide scale (U.S. DOE and its contractors)
- Assess environmental issues and set policy (international companies and government agencies)

4.4.3 Example: SEDSS

The U.S. DOE, Office of Science and Technology, through the Mixed Waste Landfill Integrated Demonstration program, has funded the development of the Sandia Environmental Decision Support System (SEDSS) (www.sandia.gov/Subsurface/factshts/ert/sedss.pdf). The latter is a decision support strategy and automated decision support tool for aiding environmental decision support makers in selecting appropriate characterization and remediation schemes. It is a concise, widely applicable process or framework for formulating, addressing and solving environmental problems. SEDSS is designed to extend the application of risk-based performance assessment methodologies to environmental restoration activities. It explicitly accommodates uncertainty while integrating risk analysis with data collection.

SEDSS provides access to site data through a GIS; a set of analysis tools necessary to qualitatively and quantitatively evaluate a site based on available information; and a platform for graphically displaying and documenting the results of these analyses. To address uncertainty, SEDSS uses probabilistic methods such as Monte Carlo simulation and geostatistical analysis. SEDSS includes a Conceptual Model Manager (CMM) that assists users in defining and documenting the environmental problem. A series of interactive windows allows the user to step through the problem to identify assumptions about the site environment as well as sources of contamination, transport pathways, and human and environmental exposure routes. The user's response to the CMM queries provides the framework for the quantitative evaluation of risk.

Some examples of decisions that SEDSS may help answer are the following: How should resources (cost and time) be prioritized based on estimates of risk? Is a site safe? If the site is unsafe, what remedial action or containment is necessary and optimal? When is remediation complete and how can it be defended as adequate? Is a potential new waste-facility-design safe (*i.e.*, can a permit application be defended)? Is a monitoring network adequate? Eventually, a user of SEDSS will be:

- Guided in establishing performance objectives for a given site;

- Provided with immediate access to both data stored in numerical or graphical forms;
- Guided in the development of a conceptual model of site conditions, including the types and condition of the waste, the pathways through which the waste migrates, the mechanisms by which human receptors can be exposed to the waste, and the human health risks associated with that exposure;
- Assisted in setting up and executing numerical analysis to evaluate the user's concept of contaminant movement, exposure, and risks that accommodate the users existing uncertainty in site characteristics;
- Provided with the ability to determine which additional data would be important in altering or substantiating a current decision on risk or remediation approach, and tools to evaluate the cost/benefit of acquiring that data; and
- Guided in on-site, real-time collection of additional data.

A prototype of SEDSS exists for a Sun workstation or server running the Unix operating system. Its capabilities include the first four of the above steps, providing the ability to determine risk from radioactive or hazardous contaminants moving via the ground-water pathway. The prototype software will optimize the location of a monitoring well network for detection of hazardous contaminants under U.S. EPA RCRA regulations, and perform risk analysis for the groundwater pathway.

4.4.4 Example: FIELDS

U.S. EPA Region 5 is developing a Fully Integrated Environmental Location Decision Support (FIELDS) System (www.epa.gov/region5fields/static/pages/mission.htm) to help identify, assess, communicate and solve priority environmental problems in specific geographic areas. FIELDS began as an effort to more effectively solve contaminated sediment problems in and around the Great Lakes. The FIELDS team now supports and has applied their technology tools to numerous sediment, soil and groundwater sites in U.S. EPA Region 5 and beyond. The system can be applied to environmental problem on various scales involving soils, surface water, ground water, air and the ecology at diverse site facilities, watersheds and urban areas. FIELDS employs Geographic Information Systems (GIS), Global Positioning System (GPS), database, analysis and imaging software combined in a modular fashion. It supports the EQuIS database by Earthsoft (www.earthsoft.com) which holds chemistry data from wells, bore holes, groundwater, surface water and soil samples. It also has a geology database, which can contain information on well development.

4.4.5 Example: SADA

The University of Tennessee at Knoxville is collaborating with Oak Ridge National Laboratory in the development of a Spatial Analysis and Decision Assistance (SADA) System (www.sis.utk.edu/cis/sada/) under funding by the U.S. EPA and Department of Energy (U.S. DOE). SADA is an evolving freeware system that incorporates tools from environmental assessment fields into a unified problem solving environment. The tools include integrated modules for visualization, geo-spatial analysis, statistical analysis, human health risk assessment, cost-benefit analysis, sampling design, and decision analysis. The capabilities of SADA can be used independently or collectively to address site specific concerns when characterizing a contaminated site, assessing risk, determining the location of future

samples, and designing remedial action. The U.S. EPA FIELDS team has funded the development of an ecological risk module for SADA.

SADA provides a number of methods for the exploration of environmental data that is categorized by depth during remedial investigations (generally soil and groundwater). Data exploration tools include two- and three- dimensional data visualization options. Three-dimensional information is presented as multiple slices (layers) or by volume. The volume approach allows visualization of all depths at once. SADA accepts map layers from GIS and allows one to select a sub-region of the site for geo-spatial and risk analyses. Geo-spatial analysis tools include methods for assessing spatial correlation among data, modeling spatial correlation, and producing concentration, risk, probability, variance, and cleanup maps. Spatial data can be interpolated via ordinary kriging, indicator kriging, inverse distance or nearest neighbor methods.

SADA can produce site-specific cost-benefit curves that demonstrate the specific relationship between a given remedial cleanup goal and the corresponding cost. This cleanup goal can be a concentration value or a particular human health risk scenario. It also provides different strategies to determine future sample locations, depending on the choice of geo-spatial interpolator. The estimate rank approach identifies unsampled locations that are modeled to have high concentration levels relative to the existing data. It can be useful for verifying the extent of hotspot regions and is available for any of the interpolation schemes. It does not account, however, for data variability. Consequently, it may place sampling points at locations that are high in concentration values but are relatively well characterized. The variance rank approach fills new samples into unsampled locations that have high estimation variances. Since it gives no weight to the magnitude of concentrations, samples may appear where data are sparse but where corresponding concentrations are very low relative to the decision rule. This approach is available only with ordinary kriging. The percentile rank approach considers both the magnitude and variance so as to avoid sampling well-characterized hot spots or sparse areas with very low detected or non-detected values. The uncertainty rank approach places new samples in areas where there is the greatest uncertainty about exceeding a cleanup goal. It helps delineate the boundaries of an area of concern. Finally, the secondary constraint approach allows the user to specify a minimum distance between any new sample locations and any previously sampled data.

4.4.6 Example: LANL EES Integrated GIS and Model-Based Data Management System

The Earth and Environmental Sciences (EES) Division at Los Alamos National Laboratory (LANL) is integrating numerical vadose zone modeling results with GIS into a single platform to store, retrieve, query and display data (EES Division, 2001). The GIS system is comprised of a Raid Storage device, an Oracle database, a SDE (Spatial Data Engine) for querying, and Arc/View GIS to view the data. Traditional GIS data such as contours, historical building footprints, and known source locations are stored as points, lines and polygons with attributes. Vadose zone flow and transport modeling results obtained with the three-dimensional FEHM code are stored as points with attributes, such as temperature or pressure at a given node. The system is able to store, query, visualize and compare both numerical model results and geographic data

in an efficient manner. This has proven to be helpful in creating both conceptual and numerical models of subsurface flow and transport at the site.

4.4.7 Example: LANL EES Geologic Model of LANL Site and Española Basin

The EES Division at LANL is developing (personal communication, 2001) geologic models for the Environmental Restoration Project and Groundwater Protection Program on two scales: Laboratory (138 square miles) and Española Basin (2500 square miles). The models, assembled from geologic maps, borehole data, high-resolution total station mapping, and interpretative cross-sections, are managed using Oracle. GIS Arc/Info is used to assemble the various data sources and to generate gridded surfaces representing geologic unit boundaries. These surfaces are assembled into three-dimensional geologic models using Stratamodel.

At present, the models are built into 30 geologic units, from approximately 50,000 data records. The geology ranges from Precambrian basement to recent ash-flow tuffs to localized basalt flows. The models embed all of the spatial relations of the actual geology and can be used to generate geologic maps, cross-sections, etc. They will be the framework for numerical models of hydrologic and contaminant transport processes in the area. The Laboratory scale model will serve as a framework for detailed studies of the fate of contaminants such as in the Los Alamos Canyon region. At the Española Basin scale, the model will help interpret transport pathways through the regional groundwater system.

5 QUALITATIVE CONCEPTUALIZATION AND SCREENING OF HYPOTHESES

Having assembled and organized a hydrogeologic knowledge base for a site, one should be ready to describe in qualitative terms the hydrogeologic makeup of the site, the prevailing flow and transport regimes, and the way in which these regimes are expected to evolve under future scenarios of interest. As each site is defined by a geographic area and a depth dimension, its qualitative hydrogeologic description should capture all three of these spatial dimensions. Several alternative descriptions should be considered, examined in light of the available information and eliminated if in conflict with key data.

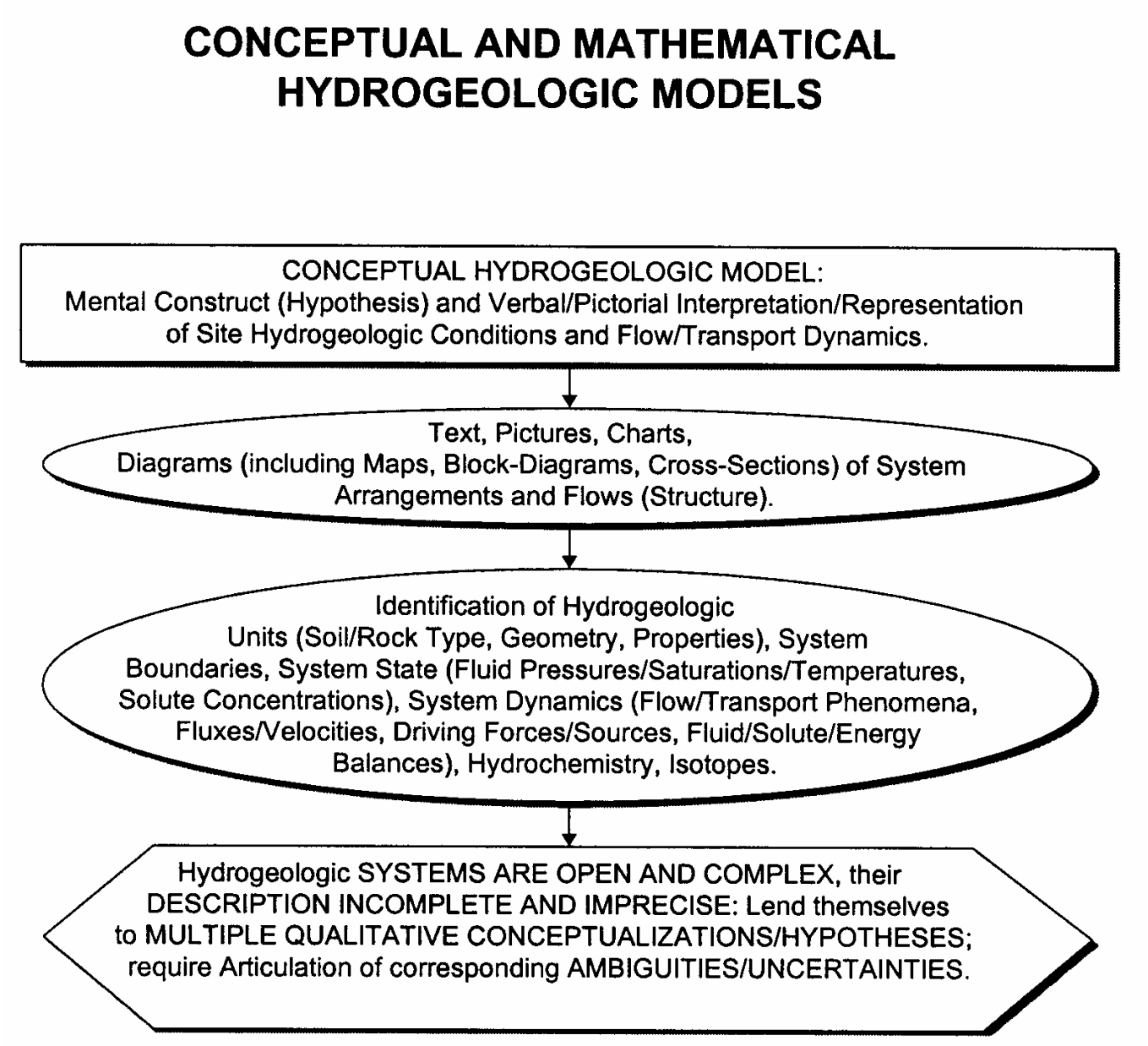


Figure 5-1. Qualitative conceptual-mathematical modeling.

5.1 What is a Qualitative Conceptual Model of Site Hydrogeology?

The qualitative aspect of conceptual hydrogeologic modeling is encapsulated in Figure 5-1. A conceptual hydrogeologic model is a mental construct or hypothesis accompanied by verbal, pictorial, diagrammatic and/or tabular interpretations and representations of site hydrogeologic conditions as well as corresponding flow/transport dynamics. It is customarily presented in the form of written text accompanied by pictures, charts and diagrams (including maps, block-diagrams, cross-sections, panel diagrams, vertical/horizontal profiles) of system components, arrangements and relationships (flows) known collectively as structure. A conceptual model identifies relevant hydrogeologic units (soils, aquifers, aquitards, aquicludes, bedrock) and features (faults, intrusions, fractures), their makeup (mineralogy, petrography, texture, cementation, porosity, permeability, related petrophysical and geophysical properties), geometry (horizontal and vertical dimensions, boundary and internal configurations), system states under undisturbed and disturbed conditions (types of fluids that permeate the soils and rocks, their pressure, saturation, density and temperature, dissolved solutes and their concentrations), flow and transport dynamics and kinematics (active flow and transport processes; their driving mechanisms; fluid and solute fluxes and velocities; fluid, solute and energy balances), hydrochemistry and isotopes.

5.2 Why Interpret Site Hydrogeology in Three Dimensions?

A hydrogeologic knowledge base typically includes information obtained from boreholes, soil and geologic outcrops, underground openings and geophysical surveys. This adds a depth dimension to geographic information about the site and its surroundings which is largely two-dimensional. A hydrogeologic knowledge base is thus inherently three-dimensional, and should allow one to describe the system in three-dimensions. This is important because hydrogeology is inherently three-dimensional, complex, and manifests itself on a multiplicity of scales. As such, it cannot be meaningfully captured in a single two-dimensional map or cross-section, nor can it be described adequately by a one-dimensional profile such as a stratigraphic column or a lithologic log, that show only selected details on a given scale. Instead, hydrogeologic data must be interpreted in three-dimensions and in some detail on a range of scales (from regional down to thin sections). There virtually always is at least some regional and site information that should allow one to do so; seldom can lack of data be validly quoted as a reason for doing less. This is true regardless of how complex or simple a set of models one ultimately employs for the assessment of site performance and the analysis of corresponding uncertainty.

Not only the hydrogeologic makeup of a site but also the corresponding flow and transport regimes can and must be explored and described qualitatively in three-dimensions on a range of scales. Seldom can an inability or lack of need to do so be validly quoted as justification for doing less. This is true regardless of how narrow is the contextual framework, how limited is the available knowledge (data) base, or how restricted are the available (financial, technological, computational) resources. Doing otherwise may result in mischaracterization of the system and misrepresentation of flow and transport in the mathematical model.

5.3 How to Formulate Hydrogeologic Hypotheses?

A systematic approach to the qualitative conceptualization and screening of hydrogeologic hypotheses is depicted in Figure 5-2. It requires as a prerequisite the availability or acquisition of expertise in the qualitative and quantitative interpretation of hydrogeologic field data. With such expertise, and a healthy dose of hydrogeologic insight, one may proceed to conceptualize hydrogeologic units and features on the regional, site and subsite scales in terms of their hydrostratigraphy, lithology, vertical and horizontal boundaries, structural features such as folds, faults, offsets and intrusions, textural features such as grain size, cementation, microstructure and fracturing, flow and transport properties such as porosity, permeability, dispersivity and sorption coefficients, related pedological and petrophysical properties (spatial, directional and textural) as well as distributions and variations (heterogeneity and anisotropy) on a range of scales, both between and within hydrogeologic units and relevant structural and textural features.

One must assemble or develop descriptions of geologic outcrops, well logs, air and satellite images, maps and cross-sections, panel diagrams, quantitative records of measured variables, and qualitative descriptions of observed phenomena. It is useful, but not necessary, to embed these in tools for computer management of comprehensive hydrogeologic data of the kind discussed in Chapter 4.

The next step is to describe the space-time distributions of fluid types (water, air, nonaqueous phase liquids) and corresponding states such as saturation, pressure, temperature and density, followed by significant solutes and their concentrations, as well as the delineation of saturated, vadose and perched zones, all on a range of scales between and within hydrogeologic, structural and textural units and features. This in turn allows one to describe active and anticipated flow and transport phenomena such as advection (of solutes or particulates), convection (of heat), diffusion, dispersion and sorption, their modes (discrete features; single, dual or multiple continua) and scales of manifestation (regional, site, subsite, long-term, short-term) boundary and internal mechanisms, forces and sources that drive them (infiltration, evapotranspiration, recharge, discharge, pumping), as well as their relative intensities and significance.

The next step is to delineate in a qualitative but internally consistent and coherent manner contours of equal hydraulic potential (head), pressure, saturation, density, temperature and solute concentration together with corresponding flowlines (streamlines, pathlines, streaklines), vectors of fluid and solute flux and velocity, and isochrones (of groundwater residence time and solute travel time). The potential for the development of fast flow paths, on scales smaller than is represented by these contours and vectors (focused and episodic infiltration, preferential wetting, high-permeability channels, instability of fluid fronts, fingering), must be articulated at this stage. One should also assess the overall balance of fluids, solutes and energy within the system. The task can be facilitated by means of descriptions, maps, cross-sections and panel diagrams concerning site geology and hydrogeology; records of precipitation, irrigation, evapotranspiration, runoff, river stage and discharge, wetlands; well and spring hydrographs; hand-drawn horizontal and vertical flow nets; simple water balance formulae; and published case studies of hydrologically similar sites.

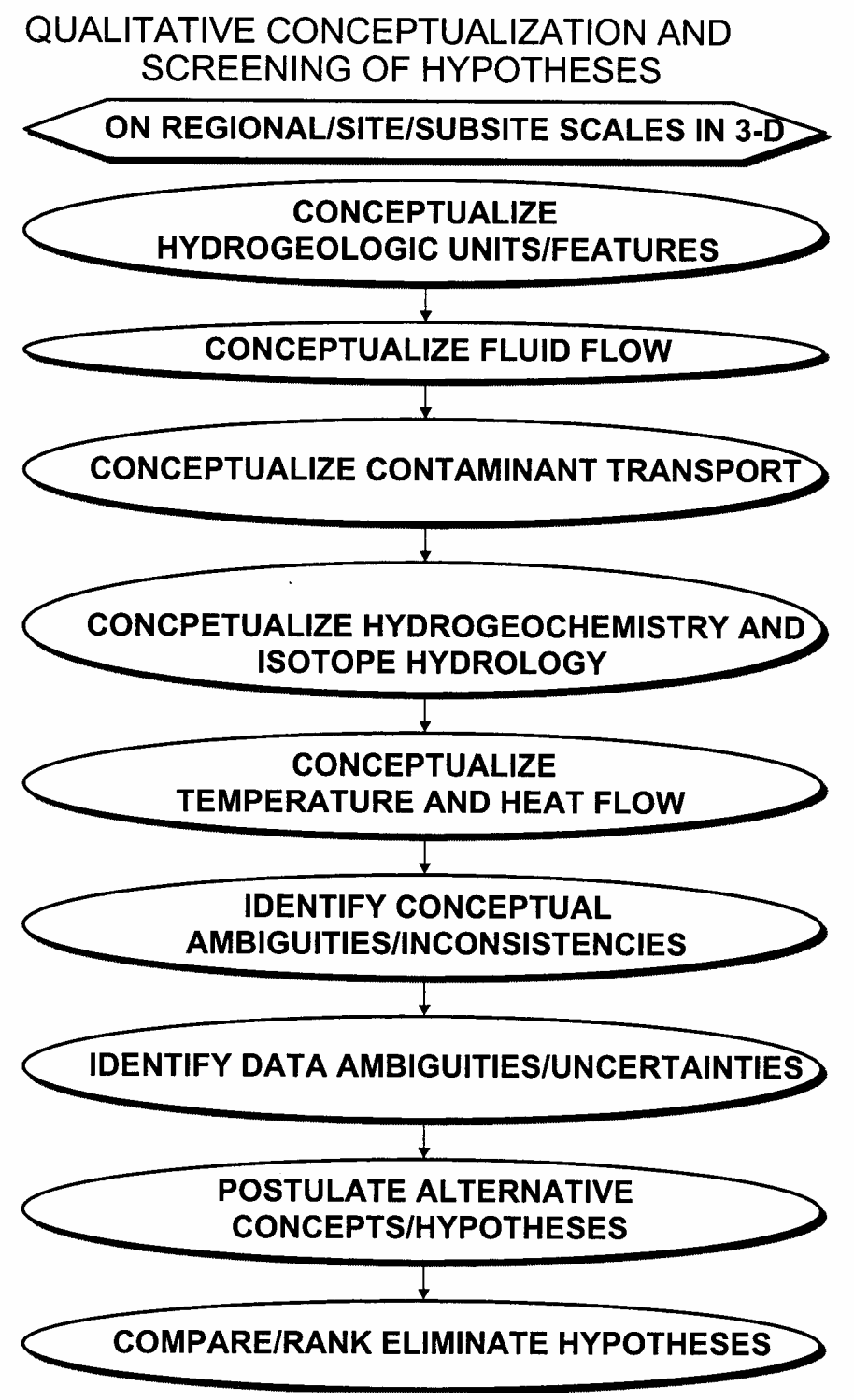


Figure 5-2. Qualitative conceptualization and screening of hypotheses.

To conceptualize contaminant transport on regional, site and subsite scales, one must describe the space-time distribution of major contaminants in the soil, vadose zone and groundwater on these scales. The description must include the space-time distributions of contaminant sources; mechanisms and rates of source contaminant mobilization and leaching; active transport phenomena such as average and fast advection, diffusion, dispersion, radioactive or biochemical decay, sorption, colloid transport; space-time distribution of migration, spreading, and dilution patterns; and overall mass and ionic balance for key contaminants.

Among tools that may be suitable for this task we list records, maps, cross-sections and panel diagrams of existing contamination, sources and groundwater flow patterns; published discussions and case studies of mechanisms that cause contaminant release, migration, spreading, and dilution along average and fast flow paths; simple formulae to assess corresponding rates and parameters; hand-drawn horizontal and vertical average and fast migration pathlines, velocity arrows, and isochrones of travel time; and simple formulae of mass and ionic balance.

A hydrogeologic conceptualization is not complete without a description of hydrogeochemistry and isotope hydrology on regional, site and subsite scales in three-dimensions. This includes the space-time distributions of major hydrochemical constituents, environmental isotopes and their sources above and below the water table; space-time distribution of groundwater ages; implications concerning flow between and within hydrogeologic, structural and textural units and features (including infiltration, evapotranspiration, recharge, discharge, directions and rates); and implications concerning transport (directions and velocities, possible compartmentalization, isolation and mixing of groundwater bodies, chemical reactions, and water-rock interactions).

Possible tools include maps, cross-sections, panel and compositional diagrams of constituents, isotopes and groundwater flow patterns; published discussions and case studies of mechanisms causing spatial-temporal variations in water chemistry, isotopes and ages; simple formulae to assess corresponding rates and parameters; hand-drawn horizontal and vertical, average and fast migration pathlines, velocity arrows and isochrones; and simple mixing and chemical balance formulae.

It often helps to conceptualize temperature and heat flow on regional, site and subsite scales in three-dimensions. This may include the space-time distributions of temperature and heat flow above and below the water table, and their implications concerning flow between and within hydrogeologic, structural and textural units and features (including infiltration, evapotranspiration, recharge, discharge, directions and rates).

Among tools that may facilitate such conceptualization are maps and cross-sections that include measured values and contours of temperature and heat flow; published heat conductivity and capacity values for soils, rocks and moisture conditions similar to those found at the site; published discussions and case studies of mechanisms causing spatial and temporal variations in water temperature and heat flow; and simple formulae that associate heat flow with fluid convection.

5.4 Why Formulate Multiple Conceptual Site Models?

Hydrogeologic systems are open and complex and the corresponding knowledge base is invariably incomplete and imprecise. Therefore, such systems almost always lend themselves to multiple conceptualizations and the postulation of several alternative hypotheses. It is therefore important to explore varied conceptual frameworks and assumptions through a comprehensive evaluation of a broad range of regional and site data, their translation into coherent and internally consistent conceptual models or hypotheses, and an in-depth examination of these hypotheses in light of the available knowledge base. The more experts with a wider range of earth and environmental specialties are given access to the knowledge base, the larger and more varied are the alternative site descriptions they may identify.

The conceptualization is not complete without a clear articulation of ambiguities and uncertainties associated with each alternative description and interpretation (conceptual model or hypothesis) of site hydrogeology.

5.5 How to Formulate Alternative Conceptual Models?

To develop alternative conceptual models for a site, one should consider (among others) alternative representations of space-time scales; number and type of hydrogeologic units such as layers and structures such as faults; flow and transport properties (their values and statistics, spatial distribution and geostatistics, internal heterogeneity, anisotropy); location and type of system boundaries; space-time distribution of fluids and their states (pressure, density, saturation, temperature); space-time distribution of saturated, vadose and perched zones; space-time distribution of driving forces (infiltration, recharge, discharge, initial system states, boundary conditions); space-time distribution of flow patterns; existence and nature of fast flow paths; overall water balance; space-time distribution of contaminants; space-time distribution of contaminant sources; mechanisms and rates of source contaminant mobilization and leaching; controlling transport phenomena; migration, spreading and dilution patterns of contaminants; contaminant mass balance; space-time distribution of groundwater ages; space-time relationships between major chemical constituents, isotopes, temperatures and heat flows; and their implications regarding flow and transport on regional, site and subsite scales in three dimensions and time.

The alternative conceptualizations should be firmly grounded in the available knowledge base. Each alternative conceptualization should be supported by key data. Conceptualizations that are contradicted by key data should be avoided. A major attempt should be made to articulate conceptual models that contain a minimum number of inconsistencies, anomalies and ambiguities with the lowest possible amount of remaining uncertainty about the site and the corresponding flow and transport regimes.

5.6 How to Assess and Rank Alternative Conceptual Models?

Once a number of alternative hydrogeologic conceptualizations have been articulated, they must be systematically examined, compared, screened and ranked according to acceptance criteria that include logical consistency and coherence, and the extent to which they are supported or contradicted by available observations and data. Among otherwise equal conceptual models, we

favor the least complex based on the principle of parsimony. Models that do not meet reasonable acceptance criteria of internal consistency, coherence and correspondence with the available data should be eliminated from further consideration at this stage of the analysis.

5.6.1 Example: Proposed High-Level Nuclear Waste Site

As an example of how one may qualitatively formulate alternative conceptual hydrogeologic models, assess them, select the best among them, and confirm or refute some of them through additional site exploration we consider the case of a large apparent hydraulic gradient (LAHG) identified on the basis of water level measurements in boreholes on the north side of Yucca Mountain, Nevada. The site is being considered as a potential repository of high-level nuclear waste (Figure 5-3). The material that follows is taken from Neuman (1997b) and represents his personal view.

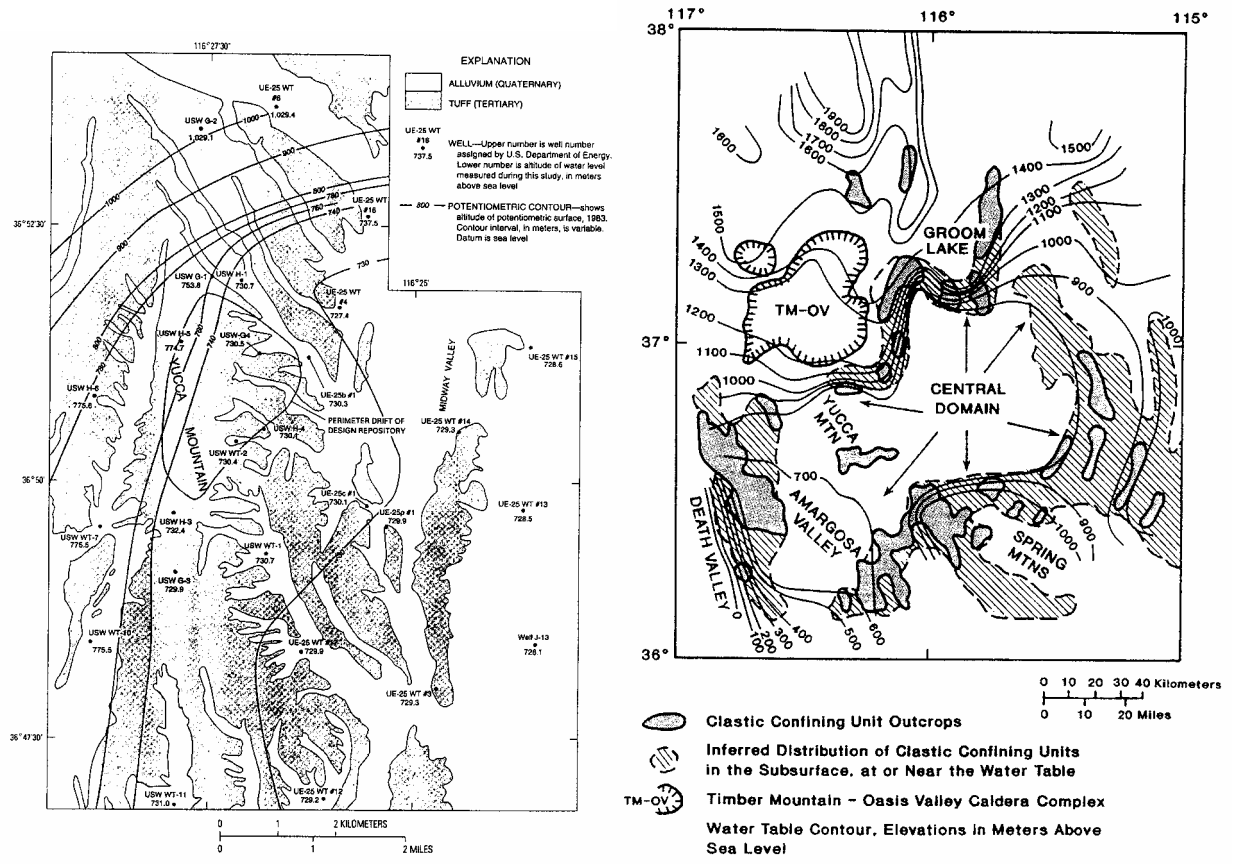


Figure 5-3. Local and regional water levels in the area (after Fridrich *et al.*, 1994).

The LAHG is defined by only two boreholes on its north (G-2, WT-6) and three boreholes on its south (G-1, H-1, WT-16). There are no data to confirm that water level highs recorded in G-2 and WT-6 persist further to the north, northwest or northeast. The LAHG is on the order of (200 m - 280 m)/2,000 m = 0.01 - 0.14 in contrast to a moderate gradient of 0.02 - 0.04 across the Solitario Canyon Fault, and small gradient

of 0.0001 - 0.0004 south and east of the former. Regionally, steep hydraulic gradients tend to be associated with known geologic or topographic features such as edges of thick confining units, faults with major offsets, caldera boundaries and mountain range fronts. A large gradient on the west side of Yucca Flat coincides with the edge of the Eleana Formation (Figure 5-3). The LAHG under consideration is unique in that it does not correspond to any obvious geologic or topographic feature.

Several conceptual models have been proposed for the LAHG. They attribute the LAHG to:

1. A perched system such that water levels in boreholes G-2 and WT-6 on the north side of the LAHG reflect the upper volcanic aquifer, those on the south reflect the lower volcanic aquifer, and downward vertical leakage takes place from the former to the latter through the intervening Calico Hills aquitard;
2. A semiperched system which is similar but considers the Calico Hills aquitard to be saturated;
3. Topographic control on the regional and local water table;
4. A drain model according to which the LAHG coincides with the effective northern limit of the deep carbonate aquifer and a fault buried under the Calico Hills Formation, marking the northern boundary of a buried graben; water drains through the fault downward from the volcanic system into the underlying carbonate aquifer, then returns in part by upwelling along the Solitario Canyon, Bow Ridge and Paintbrush Faults;
5. A spillway model in which a buried fault delineates the effective northern boundary of the lower volcanic aquifer, causing water to spill from the upper into the lower volcanic aquifer;
6. A model according to which a buried fault acts as a barrier to flow from north to south due to the juxtaposition of low- and high-permeability layers and/or the presence of low-permeability gouge material within the fault;
7. A reduction in permeability due to increased rock alteration and decreased fracturing to the north;
8. Presence of the Eleana formation which causes the thinning of overlying volcanics and acts to reduce flow rates from north to south;
9. Neotectonic phenomena such as rotational extension and/or increased stress to the north causing elevated water levels in this region.

For purposes of performance assessment, it is important to adopt an appropriate interpretive model for the LAHG so as to demonstrate an understanding of site hydrogeology and reduce uncertainty in both the qualitative (conceptual) and quantitative (computational) representations of flow and transport in the area. A reliable conceptual framework for the LAHG is especially important to correctly define inflow boundary conditions for the site groundwater flow model and to answer questions such as: Is inflow into the area from the north significant and, if so, where does it occur and at what rates? Is flow from the north diverted into Crater Flat or along Fortymile Wash? If so, how and to what extent?

In selecting among alternative conceptual models for the LAHG, Neuman favored those which are supported by the largest, most reliable and relevant set of observations and experimental data. Among all alternative models that conform to a given set of observations and experimental

data, he favors the least complex (this is known as the principle of parsimony).

Based on these principles, Neuman tentatively favored the perched system model because it is conceptually straight forward and, as shown below, is supported by numerous direct and indirect observations and/or data while contradicting none. The same is not true about any other model with the possible exception of the semiperched concept which is easy to accommodate jointly with the perched interpretation (more on this later). A single additional borehole, drilled strategically into the LAHG area and logged as well as sampled appropriately, should in his view suffice to confirm or deny that the LAHG is an artifact of perched conditions, as Neuman proposed. In the absence of new data from such a borehole (at the time of Neuman's writing), he tentatively associated a probability of 0.95 with the perched system model, 0.4 with the semiperched model, and 0.1 with all remaining conceptual models for the LAHG.

The following illustrates how observations and data (Figures 5-4 – 5-6, Table 5-1) support, more-or-less in the order listed, a perched interpretation of the LAHG: (1) Recorded water levels in G-2 and WT-6 are near the contact between the Topopah Spring basal vitrophyre and the underlying Calico Hills Formation; perched conditions are known to exist near this contact in other wells (UZ-14, NRG-7A, SD-7, SD9); (2) recorded water levels in G-2 and WT-6 (1,020 - 1,030 m) are not anomalous when compared to perched levels (from north to south) in UZ-14 (960 m), NRG-7A (860 m), SD-9 (890 m) and SD-7 (860 m); (3) the upper volcanic confining unit is much thicker in G-2 (326 m) and WT-6 than in G-1 (156 m), H-1 (135 m) and WT-16; these wells define the LAHG; (4) geophysical logs suggest that rock saturation along G-2 is at and/or slightly below unity at altitudes above 730 m which coincide with the Calico Hills aquitard, but at unity in the underlying lower volcanic aquifer; water levels in G-1 (750 m), H-1 (731 m) and G-4 (731 m) further south are just below the aquifer-aquitard contact (27 m, 5 m and 2 m, respectively); thus, the top of the saturated zone in G-2 is not anomalous when compared to G-1, H-1 and G-4; (5) water levels in G-2 declined by 12 m between 1981 and 1994, while those in WT-6 rose by 4 m; (6) thermal gradients in G-2 decreased gradually with time; (7) pumping in April 1996 resulted in an asymptotic residual drawdown of about 0.5 m in G-2 by December 17, 1996; (8) wet walls and dripping was observed in the air-filled part of this borehole above the Topopah Spring basal vitrophyre; and (9) pulsed-heat flow meter logs have indicated downward flow in the water-filled part of the borehole, suggesting leakage into the lower volcanic aquifer. It is important to point out that other observations employed to support alternative models neither support nor contradict the perched system concept. In particular, since the perched system is relatively shallow, it should not be expected to explain phenomena and/or geophysical anomalies which are associated with deeper parts of the Yucca Mountain geologic environment.

Since the Calico Hills aquitard appears to be either at or just below full saturation in the area of the LAHG, a minute addition of water to it would render this unit fully saturated. It is therefore very likely that conditions in the LAHG area vary temporally, and spatially, between perched and semiperched. If full saturation occurs only intermittently within this unit (as current data suggest), then one can expect flow within it to take place vertically

downward under a near-unit hydraulic gradient and the hydraulic conductivity of the aquitard to be near its maximum saturated value. If, on the other hand, semiperched conditions were allowed to persist, flow within the aquitard would develop horizontal components and the LAHG would dissipate to form a milder lateral hydraulic gradient. Since the observed lateral gradient is not mild but steep, the semiperched model does not offer a viable standalone interpretation for LAHG.

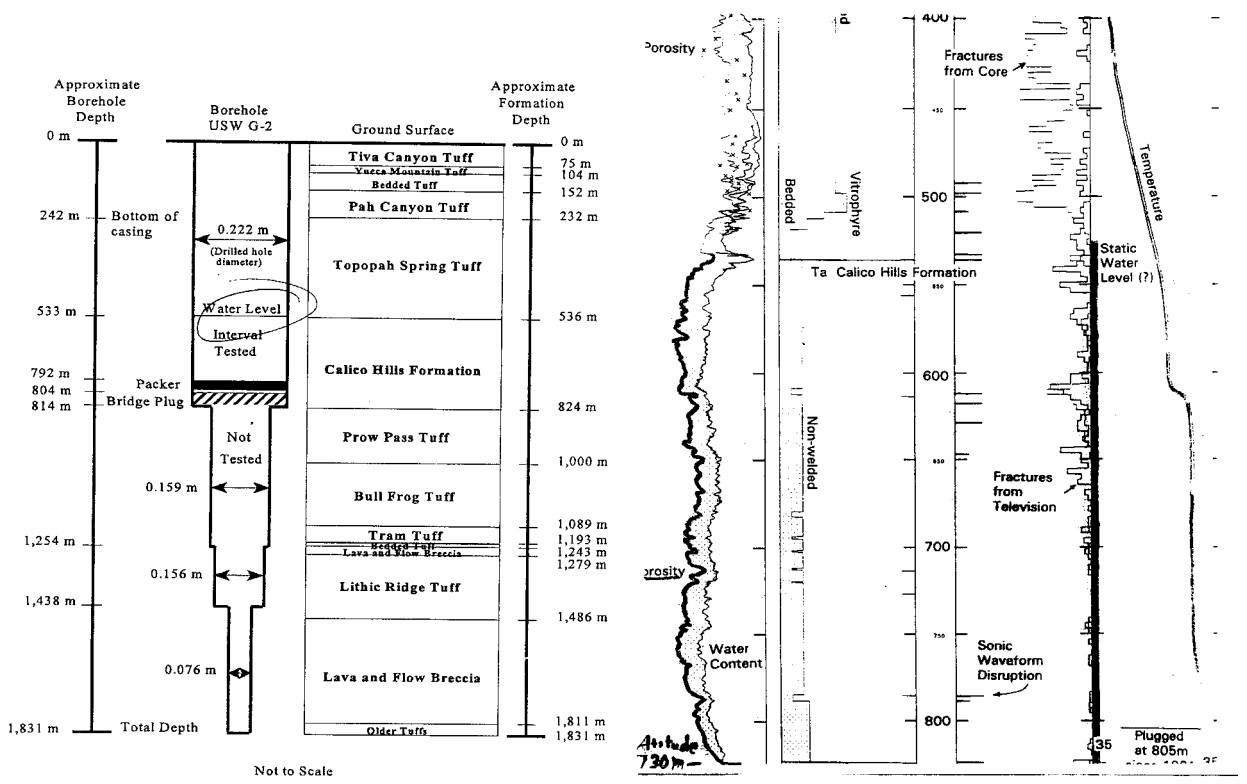


Figure 5-4. Details of conditions in borehole G-2 (presented to Expert Elicitation Panel by Czarnecki, 1997).

None of the remaining conceptual models for the LAHG are consistent with key observations and data used earlier to support the perched system concept. In particular, these models are inconsistent with the G-2 geophysical log which suggests partial saturation within the Calico Hills Formation; with the strong correspondence between recorded water levels in G-2, and confirmed perched conditions along the contact between the Calico Hills Formation and the Topopah Spring basal vitrophyre in other wells; and with the striking correspondence between the geophysically indicated top of the saturated zone in G-2 and water table elevations recorded in other wells. The idea that topography controls the LAHG is additionally contradicted by the lack of any correlation between recorded water levels and topographic elevations of wells across the LAHG (Figure 5-7); topographic control appears to be evident only on a much larger, regional scale and may help explain the general increase in water levels from their approximate values of 730 - 780 m in much of the Mountain area to about 1,200 m in the upper Fortymile Wash and 1,400 m at Pahute Mesa.

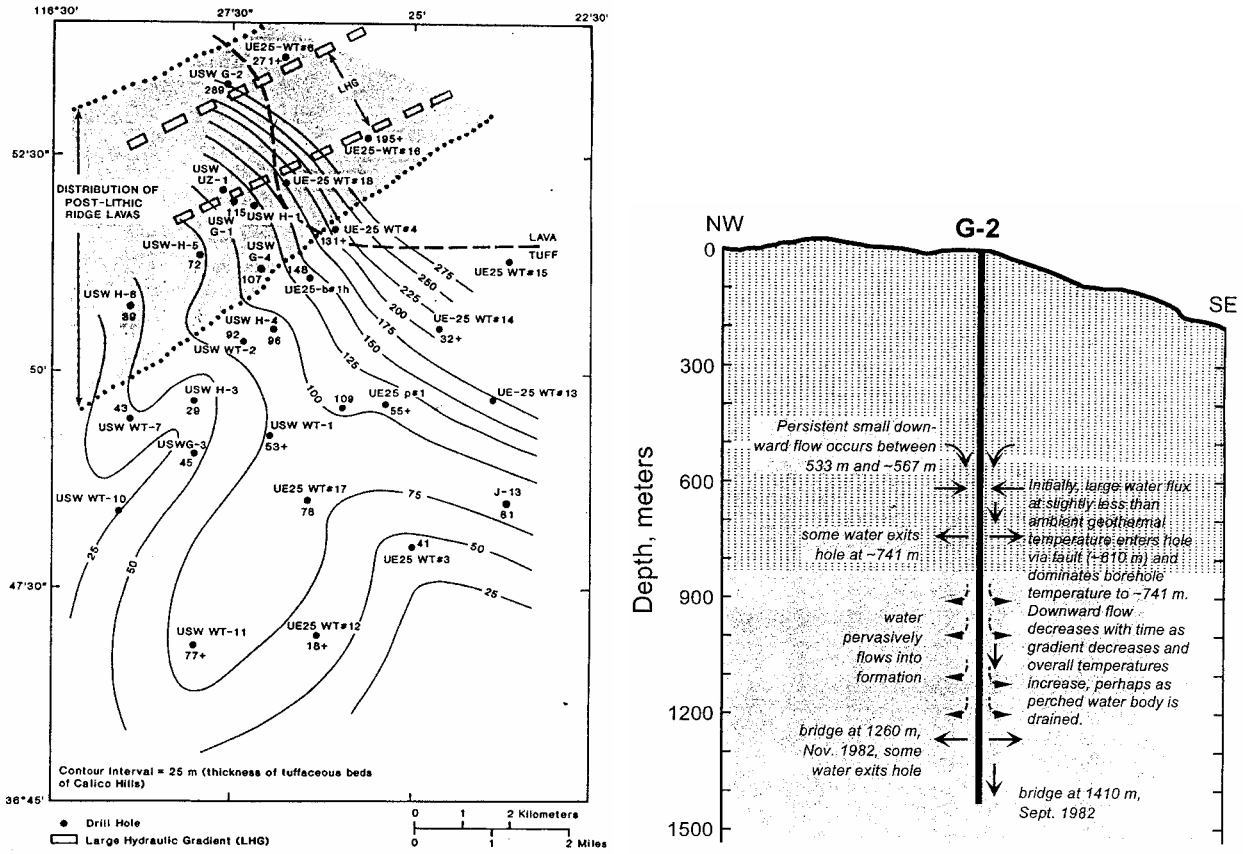


Figure 5-5. Isopach map of Calico Hills (after Fridrich *et al.*, 1994)
and flow conditions in borehole G-2 (presented to Expert Elicitation Panel by Czarnecki, 1997).

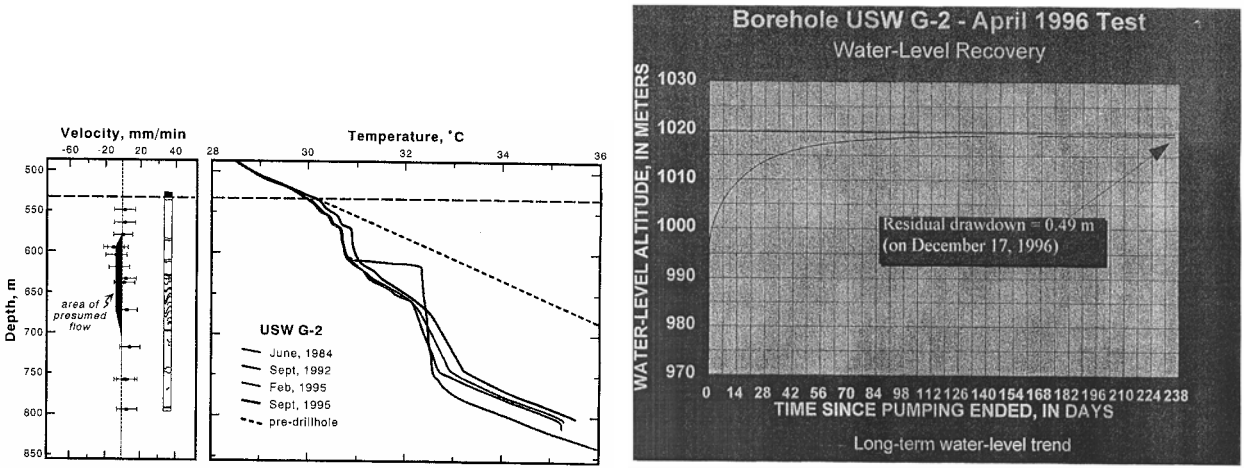


Figure 5-6. Temperature profiles and water level recovery in borehole G-2
(presented to Expert Elicitation Panel by Czarnecki, 1997).

Table 5-1. Altitudes (and thicknesses) of hydrogeologic units in deep boreholes at the mountain (after Luckey *et al.*, 1996; modified by Neuman, 1997b).

[All altitudes are in meters above sea level; water level is for uppermost interval monitored; base of hydrogeologic unit is shown even if it is above the saturated zone; >TD indicates that at the total depth borehole was still in or had not reached this hydrogeologic unit]

Borehole name	Altitude of water level	Altitude of bottom of borehole	Altitude of base of hydrogeologic unit					
			Upper volcanic aquifer	Upper volcanic confining unit	Thickness UVCU	Lower volcanic aquifer	Thickness LVA	Lower volcanic confining unit
USW G-1	750	-503	933	777	156	252	525	>TD
USW G-2	1,020	-277	1,056	730	326	361	369	>TD
USW G-3	730	-53	1,119	1,005	114	307	698	>TD
USW G-4	731	355	875	733	142	>TD	>378	>TD
USW H-1	731	-526	871	736	135	200	536	>TD
USW H-3	731	264	1,119	1,030	89	387	643	>TD
USW H-4	730	30	888	753	135	94	659	>TD
USW H-5	776	259	996	886	110	439	447	>TD
USW H-6	776	82	931	843	88	427	416	>TD
UE-25 b#1	731	-19	806	631	175	12	619	>TD
UE-25 c#1	730	216	758	615	143	>TD	>399	>TD
UE-25 p#1	752	-691	780	678	102	241	437	-130
J-13	728	-52	612	481	131	36	445	>TD

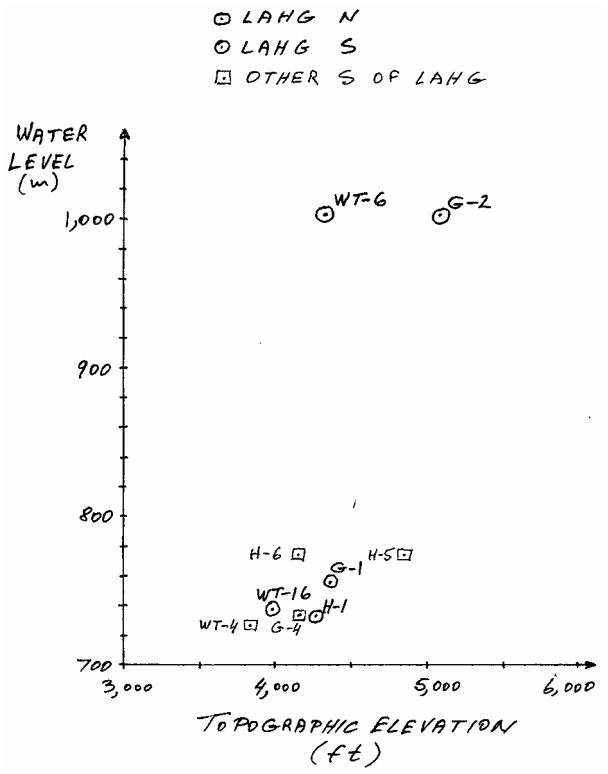


Figure 5-7. Water levels versus topographic elevations (Neuman, 1997b).

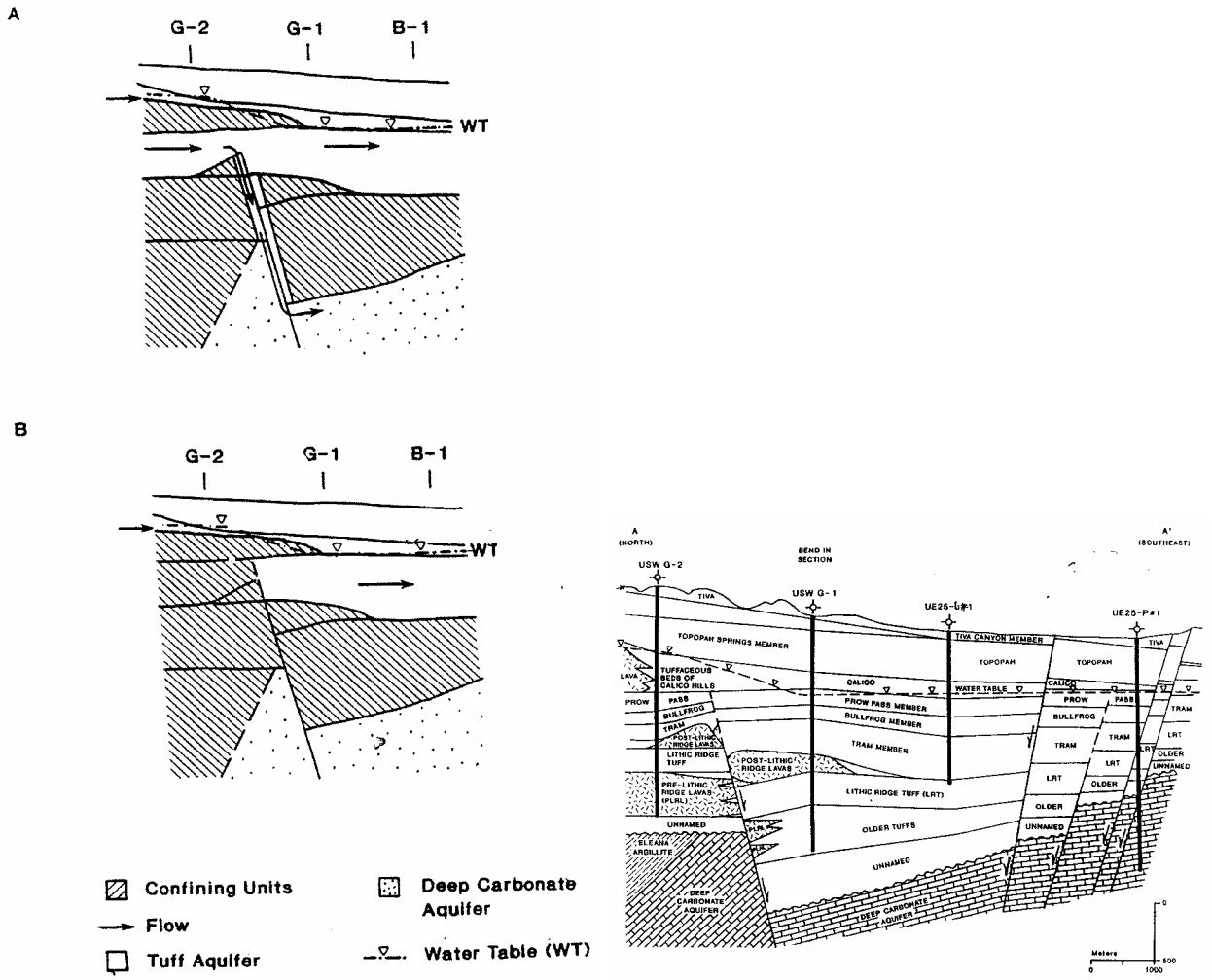


Figure 5-8. Drain (A) and spillway (B) models with section showing postulated buried graben (after Fridrich *et al.*, 1994).

The drain model (Figure 5-8A) relies on a conjectured fault and graben (Figure 5-8) for which there is no direct evidence, although it is consistent with a measured gravity anomaly (Figure 5-9) and the thickening of the Crater Flat Group between G-2 and G-1. It also postulates a complex and improbable flow system which requires that the drain (fault), and the entire volcanic system north of its inlet, be hydraulically isolated from the same system south of the inlet; what else would prevent water in the volcanic system from bypassing the inlet of the drain and thereby causing the LAHG to dissipate? The model further attributes the low heat flow "anomaly" in the unsaturated zone (Figure 5-9) to cooling of the deep carbonate aquifer by water draining into it from above, which does not explain the observed strong negative correlation between heat flow in the unsaturated zone and the thickness of this zone (Figure 5-10), suggesting that the proposed mechanism may not be the cause of the anomaly. Finally, the model attributes relatively high water table temperatures along the Solitario Canyon, Bow Ridge, and Paintbrush Faults (Figure 5-10) to upwelling of relatively warm water from the carbonate aquifer through these faults, which however does not require postulating a drain.

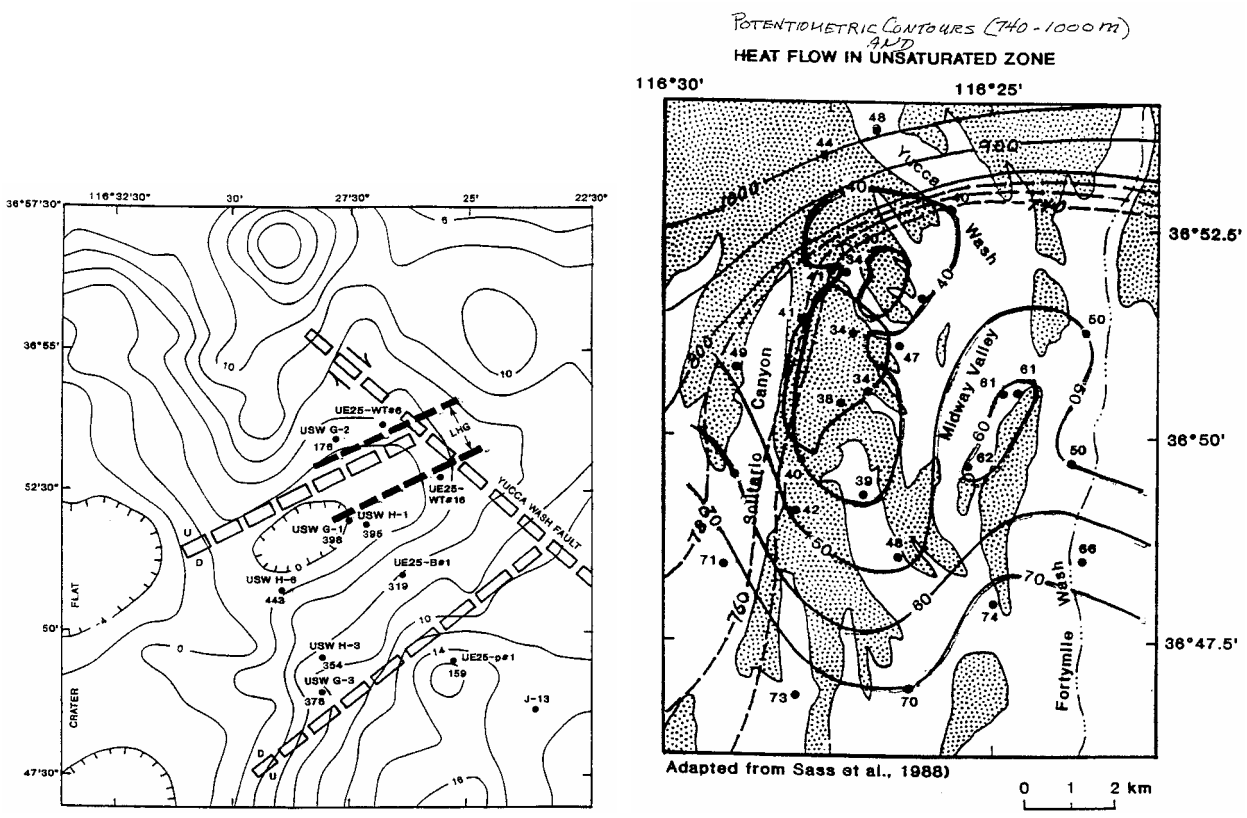


Figure 5-9. Residual gravity contours and heat flow distribution in unsaturated zone (after Fridrich *et al.*, 1994; modified by Neuman, 1997b).

The spillway model (Figure 5-8B) likewise relies on a conjectured fault and graben for which there is no direct evidence but which is consistent with a measured gravity anomaly and the thickening of the Crater Flat Group between G-2 and G-1. It further proposes an effective termination of the Crater Flat Group as an aquifer north of the LAHG without offering any direct evidence in support of such a termination. There also is no direct evidence for the conjectured reduction in permeability from south to north due to enhanced alteration and reduced fracturing, or for the proposed presence of the Eleana Formation north of the mountain; though such presence is consistent with an aeromagnetic high (Figure 5-11) attributed to magnetite-bearing argillites, it is consistent neither with the above mentioned gravity anomaly nor with the stratigraphy observed in G-2; the Eleana Formation is not present in UE-25p1. The proposed neotectonic models require postulating complex and unproven tectonophysical effects on hydrogeologic conditions and flow.

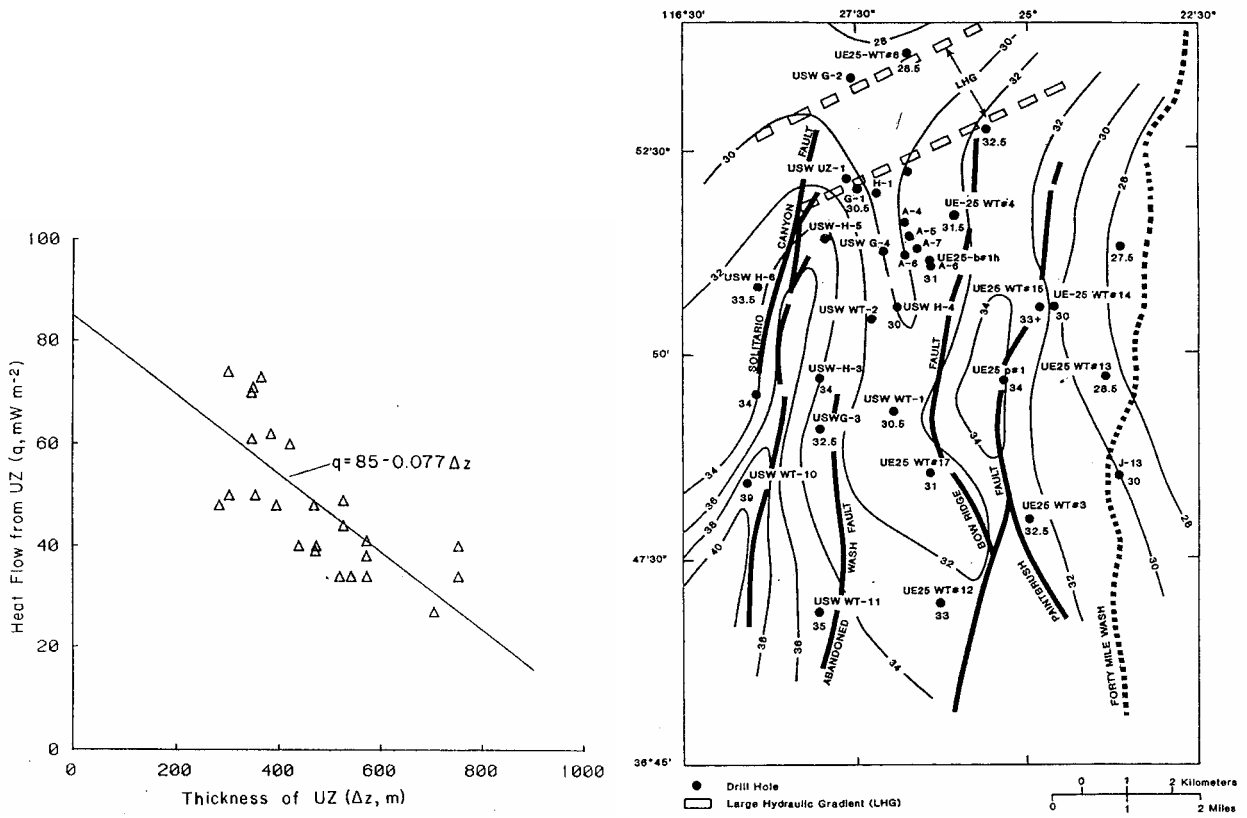


Figure 5-10. Heat flow variation with thickness of unsaturated zone (after Sass *et al.*, 1988) and distribution of isotherms at water table (after Fridrich *et al.*, 1994).

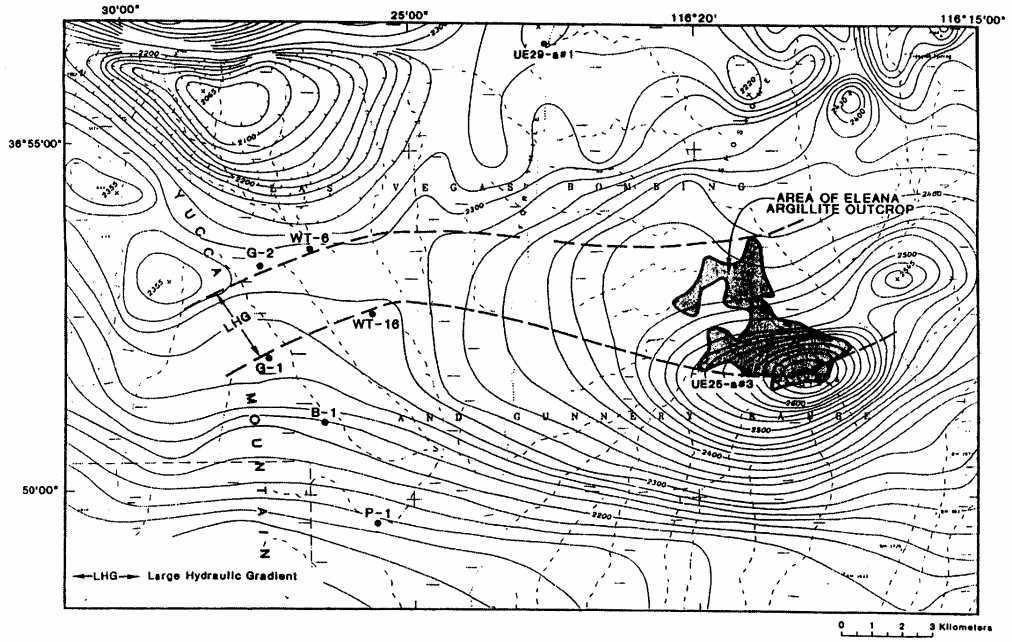


Figure 5-11. Aeromagnetic survey map (after Fridrich *et al.*, 1994).

Members of the Expert Elicitation Panel were not unanimous about whether the large hydraulic gradient is important to understanding the amount of inflow to Yucca Mountain from the north, and whether it is associated with a perched system or is connected to the regional saturated system. They, however, all agreed that if the hydraulic high is caused by a perched zone, then the underlying unsaturated zone is near saturation. They also agreed that a carefully drilled well should go a long way toward resolving the issue.

In 1998, a new well (USW WT-24) was completed just southeast of well G-2. During drilling, a water-bearing fracture was encountered at about 760 m below land surface in the lower part of the Calico Hills unit, and water rose in the borehole to an altitude of about 840 m above sea level (Tucci, 2001). According to Tucci, “drilling continued for another 104 m below the water-bearing fracture without any significant change in the water level ... Because the potentiometric level persisted as the borehole was considerably deepened, and because the water level remained relatively stable after completion of the well, the 840 m level is assumed to represent the regional potentiometric level and not a perched level."

We note that there appears to be nothing in the new information from well WT-24 to indicate that the Calico Hills unit is saturated between elevations 840 m (the water level recorded within the underlying upper volcanic aquifer in WT-24) and 1,020 - 1,030 m (the water levels recorded in the nearby wells G-2 and WT-6). It therefore remains very likely that conditions in the LAHG area vary temporally, and spatially, between perched and semi-perched as proposed by Neuman (1997b).

6 MATHEMATICAL CONCEPTUALIZATION AND QUANTITATIVE EXPLORATION OF HYPOTHESES

Once alternative qualitative descriptions of site hydrogeology have been articulated and examined, the next step is to formulate corresponding conceptual-mathematical models of flow and transport at the site under existing and anticipated conditions. These models can then be used to explore and screen quantitatively alternative hypotheses regarding the hydrogeologic makeup and behavior of the site.

6.1 What is a Conceptual-Mathematical Model of Site Hydrogeology?

As stated earlier, a conceptual-mathematical model is obtained upon formulating a qualitative conceptual model in mathematical language. Its purpose is to help define and describe the hydrogeologic system in terms of space-time dimensions; topology; geometry; interactions (called processes) between kinematic (mass, concentration, flux, velocity) and dynamic (energy, force, stress) quantities; parameters and forcing terms (sources, initial and boundary conditions). Whether such a model is analytical or numerical (written in the language of calculus or algebra) is merely a technical, not a fundamental, distinction. The model ultimately allows one to explain and interpret existing observations, and to predict new observations, quantitatively.

A conceptual-mathematical model embodies the descriptive component of a hydrogeologic model, cast in the form of mathematical equations suitable for system evaluation. The equations represent the symbolic framework, or structure, of the conceptual-mathematical model. Ideally, they provide a mathematical description of all physico-chemical (and other) processes that are considered relevant to flow and transport at the site on a given range of space-time scales subject to well-defined forcing. Defining the scales of the system includes specifying its spatial and temporal dimensions. Forcing includes sources, initial and boundary conditions. Suitability for system evaluation means that the symbolic conceptual-mathematical framework, or structure, is cast in a form that lends itself to computation and subsequent presentation (numerical, graphical) and analysis (deterministic, statistical) of the results.

One cannot legitimately disassociate hydrogeologic interpretations, or conceptual models, from corresponding mathematical and computational models for purposes of quantitative environmental impact and performance assessments of a given nuclear facility or site; these models are intimately linked and define a single conceptual-mathematical framework for quantitative site analyses.

Conceptual-mathematical modeling may, but need not, be accompanied by the selection of a corresponding computational code (or several codes). We mentioned earlier that a code is a tool rather than a model. However, selecting a particular code implies identifying specific processes that may govern flow and transport at a site, their symbolic mathematical representation, and their numerical approximation. Selecting the space-time dimensions and size of a computational grid determines the dimensions and scale of the system being modeled. Choosing the sizes of space-time discretization intervals defines the scales at which flow and transport processes are resolved. Specifying the location and type of sources, initial and boundary conditions identifies the forcings. Both material

properties and forcing terms are associated with parameters that must eventually be assigned numerical values. Choosing the modes of their representation (parameterization) defines the resolution scales of these parameters (and thus of material properties and forcing terms) in space-time. Once this has been accomplished in a way that is supported by all relevant regional and site data, the code has been transformed from a mere tool to a *bona fide* conceptual-mathematical model of site hydrogeology.

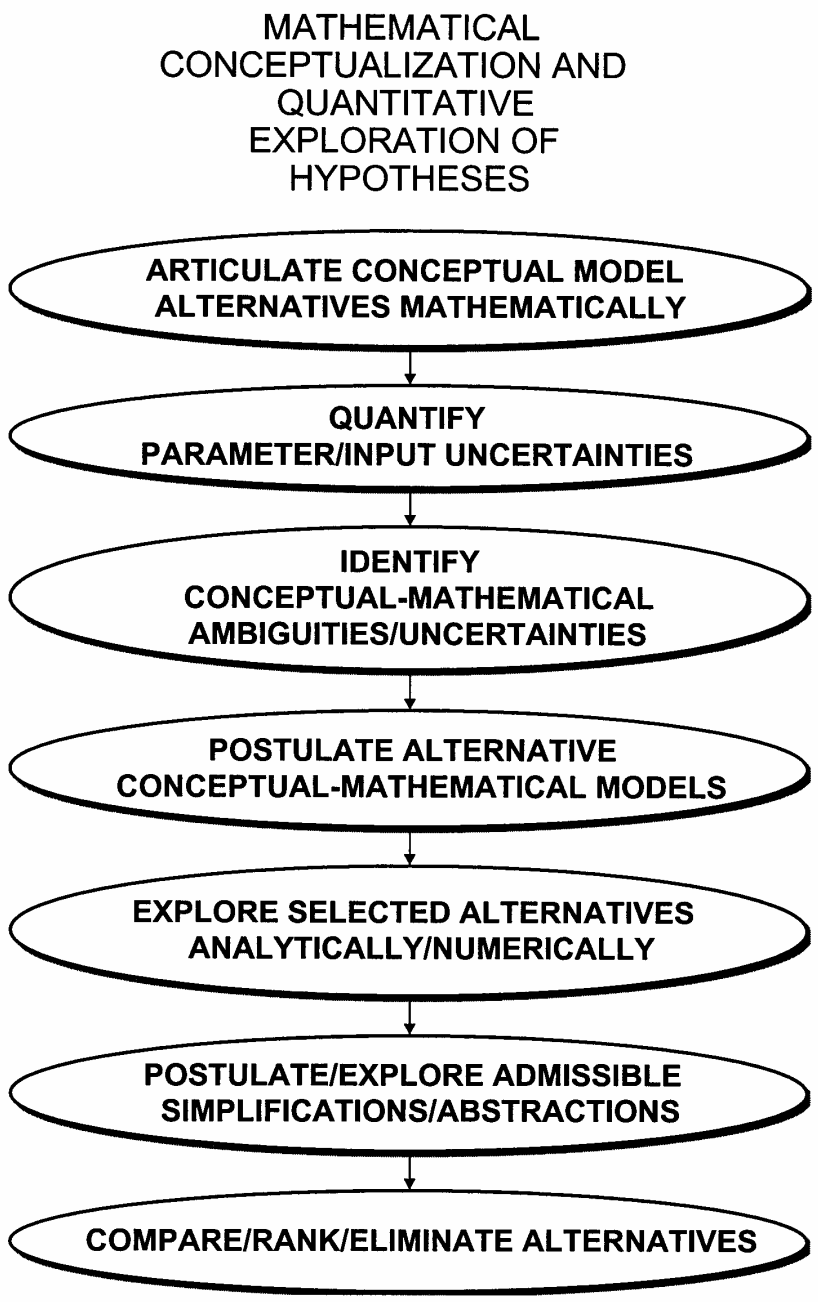


Figure 6-1. Mathematical conceptualization.

6.2 How to Formulate a Conceptual-Mathematical Model for a Site?

6.2.1 What to Specify?

Mathematical conceptualization and the quantitative exploration and screening of hypotheses are encapsulated in Figure 6-1. A prerequisite is the availability, or acquisition, of expertise in the theoretical and mathematical interpretation and analysis of hydrogeologic field data. For each alternative conceptual model identified during the qualitative stage of the process, one needs to define the hydrogeologic units and features (such as folds, faults, offsets and intrusions) that are to be explicitly modeled; their three-dimensional topology and geometry in terms of location, shape, size and relationship to other units and features (extent of external and internal boundaries and structural elements); equations that govern flow and transport phenomena included in the model; corresponding equations of state, in the form of functional relationships between parameters and state variables in the model; spatial variability of parameters that enter into the governing and state equations, within each hydrogeologic unit and feature; spatial variability of initial states in each unit and feature; boundary equations for flow and transport; and the space-time distribution of boundary and source parameters and values.

The quantification of system states includes functional and quantitative representations of fluid saturation, pressure, temperature, density, solute mass and concentration and regional as well as perched water tables on a range of scales (between and within hydrogeologic, structural and textural units and features). Governing and boundary equations include mathematical definitions and descriptions of active and anticipated flow and transport phenomena such as advection, convection, diffusion, dispersion and sorption for selected modes (discrete features; single, dual or multiple continua) and scales of manifestation, in the interior of units and features as well as on their boundaries and interfaces, including source terms, parameters, functional relationships between these quantities, and their space-time distributions. Examples of this process are given in Figures 6-2 – 6-3.

6.2.2 How to Determine a Suitable Level of Model Complexity?

The purpose of conceptual-mathematical models is to help quantify alternative hypotheses regarding the hydrogeologic makeup and behavior of a site. Since such hypotheses are always cast in three spatial dimensions, ideally so should be the corresponding conceptual-mathematical models. Both model types must allow for the hydrogeologic regime to evolve in time.

In reality, it is much more difficult and time consuming to set up and run mathematical models in three than in two spatial dimensions. Working in two spatial dimensions is often feasible because eliminating the third dimension usually entails a far lesser phenomenological change than reducing the dimensionality of the problem from two to one. Modeling and visualization of complex flow and transport phenomena is much easier in two than in three dimensions, but not much harder than in one dimension. Hence the strategy in this report supports two-dimensional flow and transport analyses (whether mathematical-analytical or computational-numerical) for exploratory purposes. Such simulations and analyses are illustrated in Figures 6-4 – 6-7. The strategy supports a similar approach for performance assessment purposes in cases where the effect of the third dimension is

demonstrably minor. However, the strategy discourages one-dimensional analyses unless a very strong and convincing hydrogeologic argument is made in their favor. This is so because flow and transport behaviors predicted by one- and multi-dimensional models often differ from each other in fundamental ways.

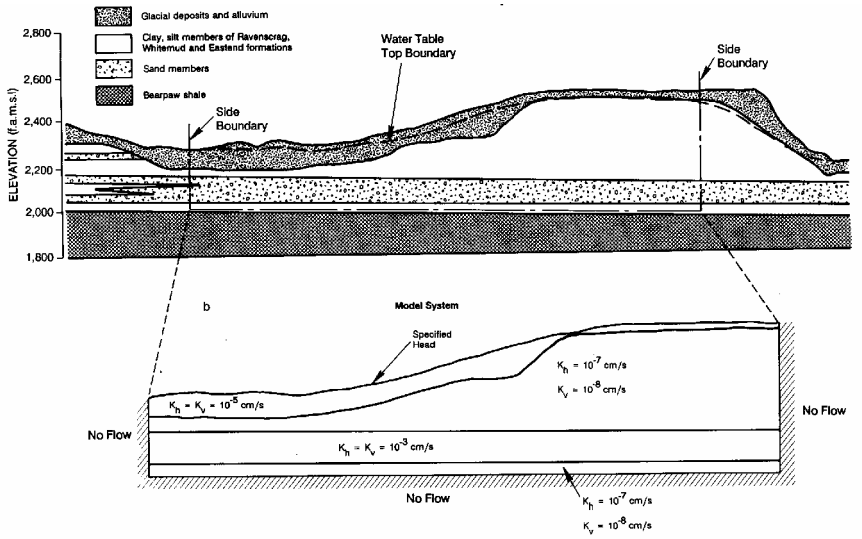


Figure 6-2. Example of hydrogeologic schematization for purposes of mathematical modeling (after NRC, 1990). With permission, National Academies Press.

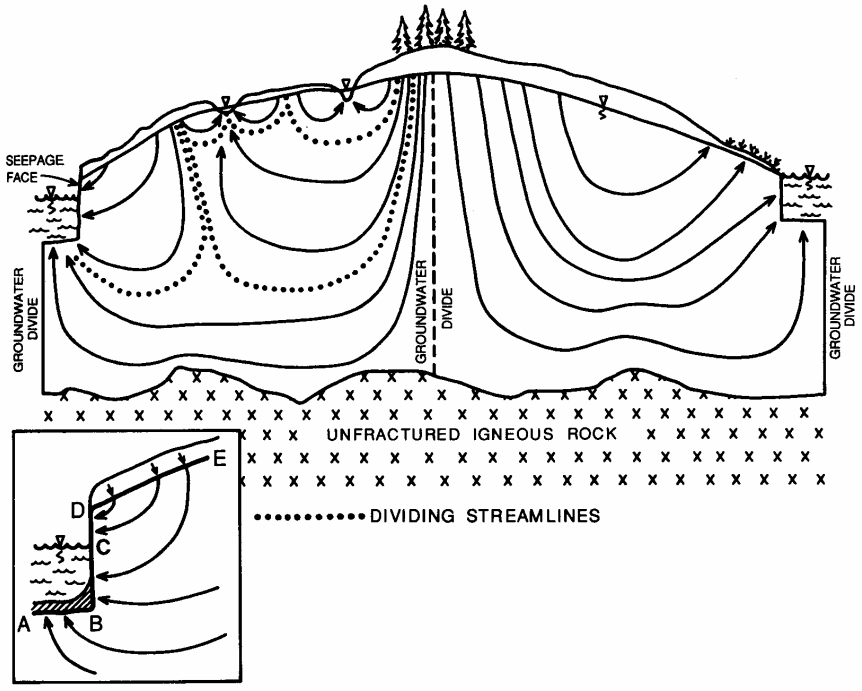


Figure 6-3. Example of hydrogeologic schematization for purposes of mathematical modeling (after Anderson and Woessner, 1992). With permission, Academic Press.

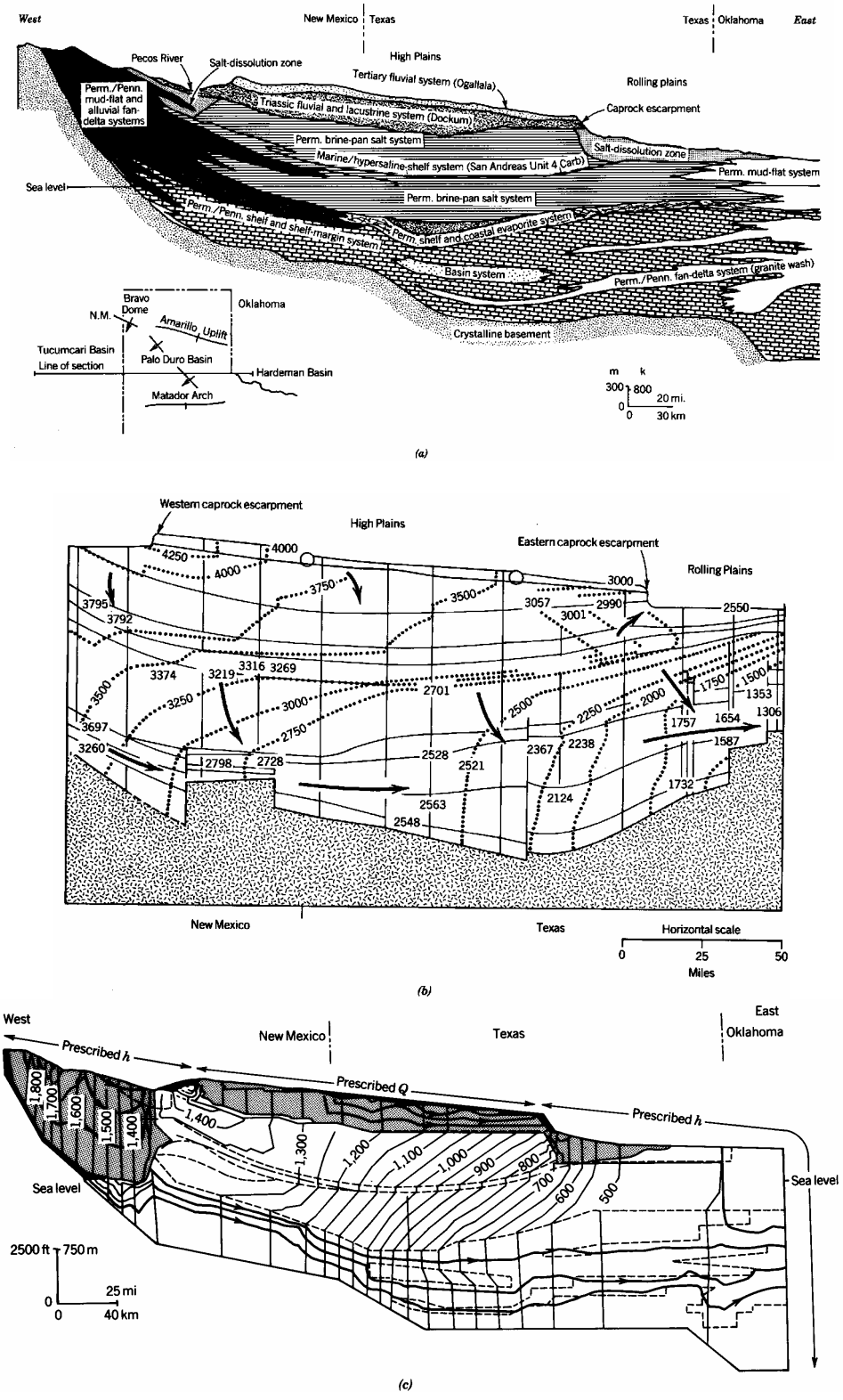


Figure 6-4. Geology (a), regional flow conceptualization (b) and simulation (c) (after Domenico and Schwartz, 1990).

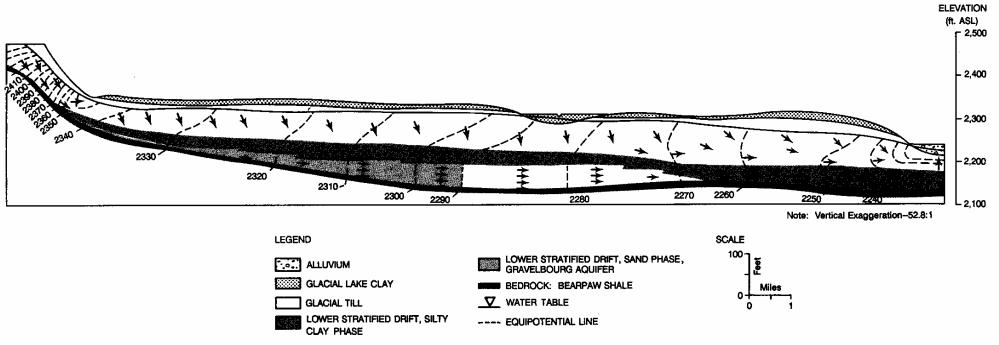


Figure 6-5. Regional groundwater flow conceptualization (after NRC, 1990). With permission, National Academies Press.

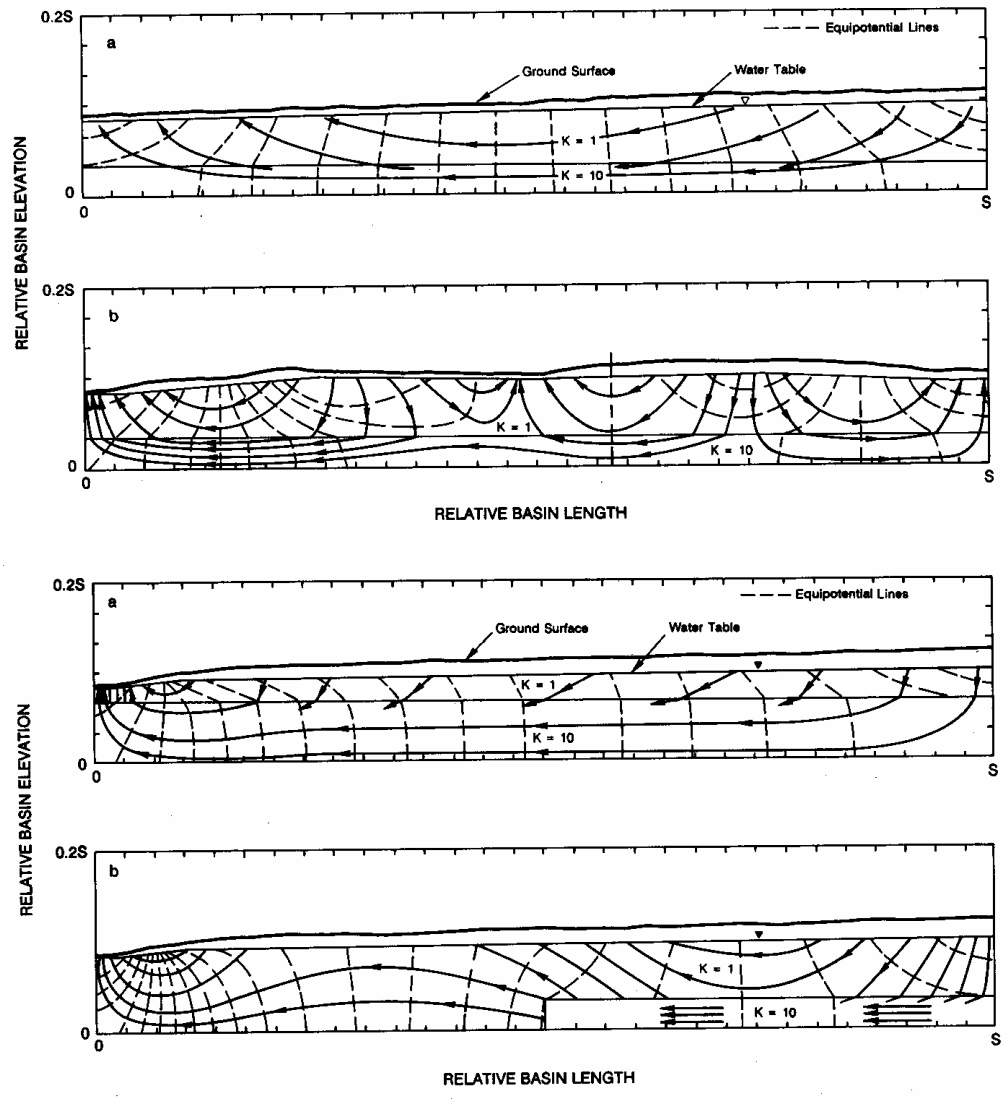


Figure 6-6. Vertical flow nets (after Freeze, 1969). With permission, AGU.

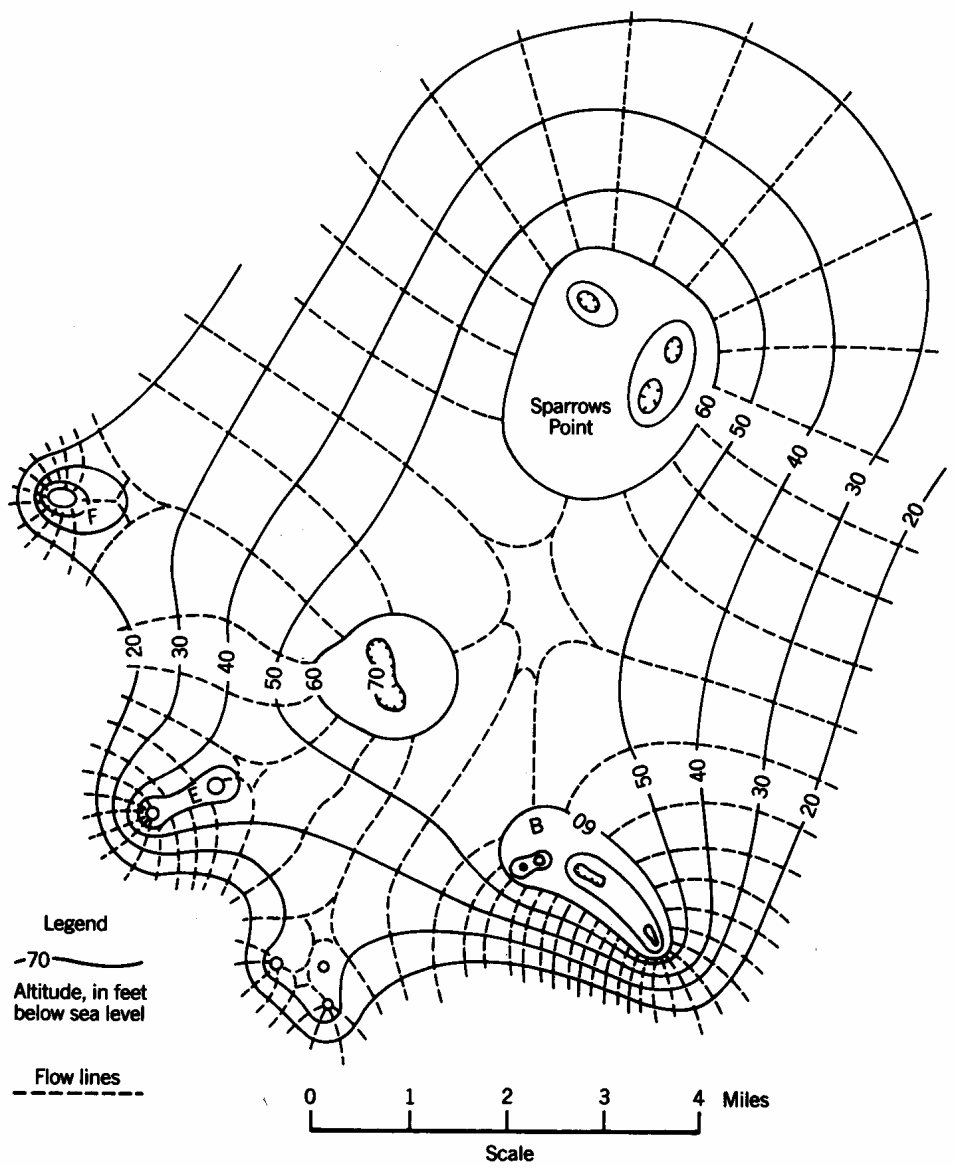


Figure 6-7. Horizontal flow net (after Bennet and Meyer, 1952).

The spatial and temporal scales at which flow and transport phenomena are modeled depend in part on the contextual framework (area, depth and time-frame of concern; space-time scales on which performance measures are defined). These phenomena are affected to a large extent, but not exclusively, by hydrogeologic complexities, heterogeneities and driving-mechanisms that manifest themselves on similar scales. It is therefore important that conceptual-mathematical modeling start by identifying these "site-scale" features, factors and/or phenomena and by incorporating them directly as explicit elements in the model. Models that do not incorporate such elements explicitly, but account for them implicitly, are considered here to be simplified or abstracted; methods to develop them are discussed in the next section.

Models that fail to account for site scales complexities, either explicitly or implicitly, are oversimplified or would generally be too crude to provide a reliable description of site hydrogeology for most purposes.

Site-scale hydrogeologic complexities, heterogeneities and driving-mechanisms may be influenced and/or controlled by larger- or regional-scale features and factors such as regional recharge, discharge, and flow mechanisms and patterns on various time scales. A model must account for these large-scale influences and controls, as well as for associated uncertainties, through the appropriate assignment of initial conditions, boundary conditions, and source terms such as those that describe infiltration, recharge, discharge and leakage across aquitards.

Site-scale hydrogeologic complexities, heterogeneities and driving mechanisms may be significantly influenced by smaller- or subsite-scale features and factors such as space-time irregularities and fluctuations in external and internal boundary and source shapes or conditions; internal heterogeneities within site-scale hydrogeologic units, faults, dikes or other features; as well as smaller-scale units, faults, dikes, fractures or preferential flow channels. Determining what is the nature and extent of these influences is an integral part of developing a conceptual-mathematical model for a site. If features and factors that manifest themselves on subsite scales are deemed important for the modeling of site-scale phenomena (as in the case of dispersion in Figure 6-8), one must account for them directly (explicitly, by embedding such features and factors discretely in the site model) or indirectly (implicitly, by formally integrating these features and factors into the site model equations and parameters, such as dispersivity) in the model.

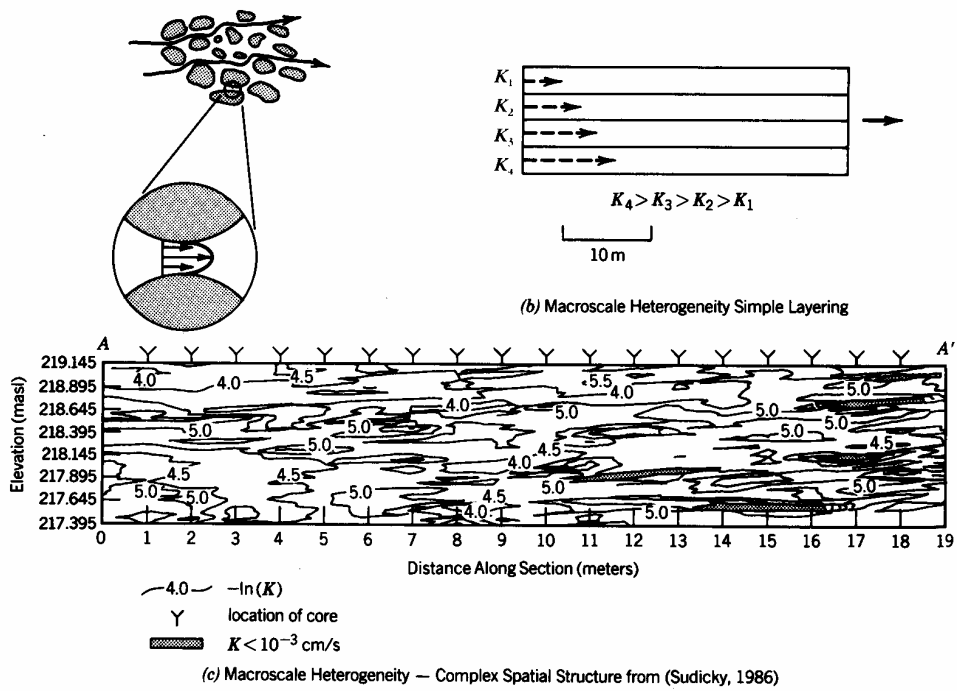


Figure 6-8. Dispersion affected by heterogeneities on various scales (after Domenico and Schwartz, 1990).

Embedding discrete small-scale features and factors in a model renders it relatively complex. Integrating small-scale features and factors into the model equations and parameters renders it less complex than embedding. However, to compensate for loss of information, the equations change form and phenomenology (from Stokes to Darcy, isotropic to anisotropic, juxtaposed to overlapping dual or multiple continua, differential local to integro-differential nonlocal) and acquire new phenomenological parameters (permeability, dispersivity, integral kernels) that differ in nature and magnitude (from scalar to tensor, local to nonlocal, well-defined to scale-dependent) from the original parameters. Only in special cases can the original form and phenomenology be recovered, and even then the parameters usually change (due to upscaling). It is not presently clear which of these two approaches is better suited for their intended task. The strategy in this report considers both options. Methods of embedding are discussed in the next section.

Ignoring the influence of regional- or subsite-scale features or factors on site-scale flow or transport without due justification, or failing to insure that the treatment of all scales is self-consistent without demonstrating that some of them are not relevant to the problem at hand, constitute oversimplifications which may cast doubt on the reliability of the model. A large number of both steady-state and transient two- and three-dimensional analytical solutions and computational as well as visualization codes are available and easily accessible to general users. Many of them are well suited for the development, exploration, screening and selection of alternative conceptual-mathematical models of site hydrogeology. Among these are the Groundwater Modeling System (GMS) mentioned in Chapter 4, which includes the forward and inverse groundwater flow and transport codes MODFLOW, MODPATH, MT3D, RT3D, FEMWATER, SEEP2D, SEAM3D, PEST, UCODE and UTCHEM; FEFLOW, an interactive graphics-based finite element simulation system for groundwater flow, contaminant and heat transport in two- and three-dimensions; *etc.*

6.2.3 How to Simplify Models?

Narrowly defined contextual or regulatory criteria, limited data or resources, and a quest for simplicity or transparency may motivate the adoption of hydrogeologic flow and transport models that are less than three-dimensional, ignore time, and include few details on limited scales. While such motivation for simplification and abstraction of hydrogeology and flow or transport dynamics may sometimes be justified on practical grounds, it does not in itself turn simplified and/or abstracted models into scientifically valid tools of performance assessment. Only a formal demonstration that such models capture the essential features and capabilities of their more complex and complete counterparts, and that they thereby provide comparable performance assessments or conservative bounds thereof, might justify their use for regulatory purposes. As already stated, to ensure that relevant aspects of hydrogeologic complexity are reflected in modeled system behavior and performance, and that the model can be rendered compatible with site data, it is important that the process of simplification be done systematically and objectively. This can be done by filtering out undesirable details through formal averaging of the governing equations in space-time or in probability space, in a way which retains and renders their influence on the model implicit. This was called embedding in the previous section.

Averaging three- or two-dimensional equations across one spatial dimension renders them two- or one-dimensional, respectively. An example of this procedure is presented in Appendix B. Averaging transient equations over time may (but need not) render them representative of a steady state. Averaging can also be done over subdomains of the site being modeled, and over multiple time intervals. In each case, the averaging results in governing flow and transport equations that contain upscaled quantities (such as the relative permeability curves in Figure 6-9). If the space-time scales of these quantities differ from those of the available site data, then either the model or the data must be rescaled to render them compatible and comparable with each other.

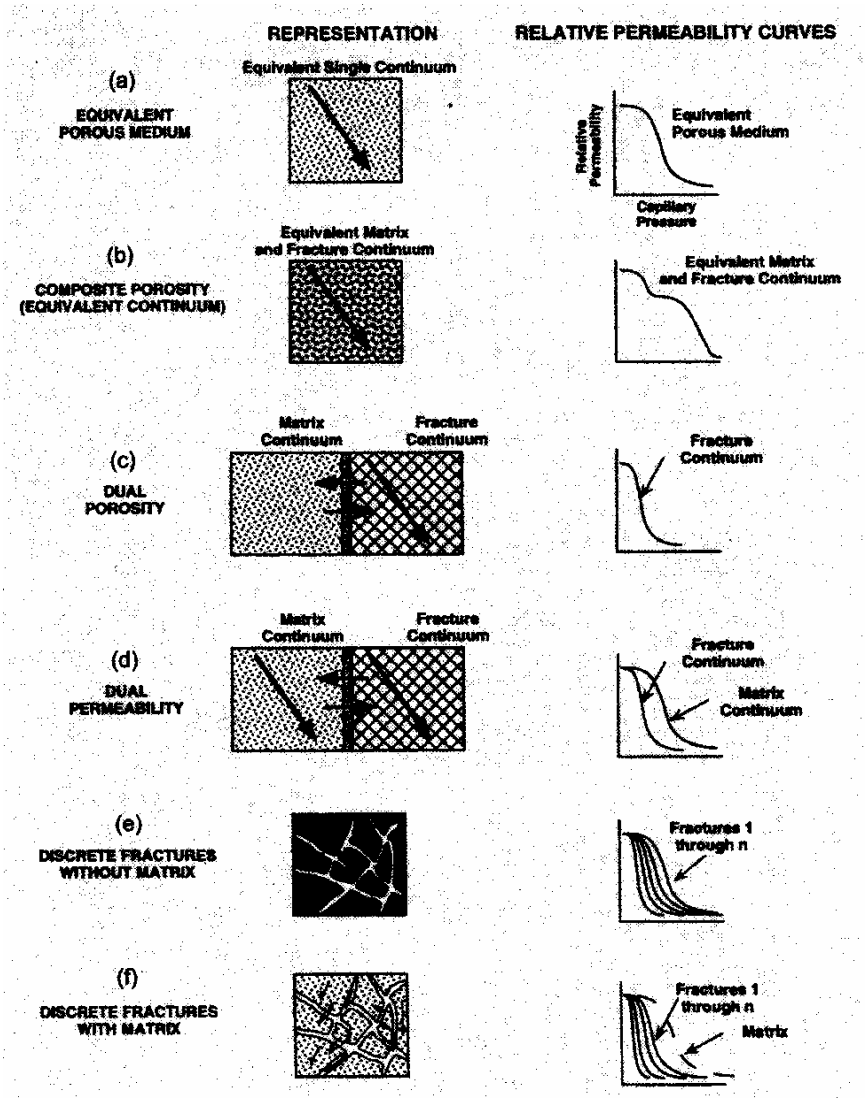


Figure 6-9. Alternative conceptual models and their relative permeability characteristics (after Altman *et al.*, 1996).

It has been traditional to assume that spatially and temporally averaged hydrologic quantities are less prone to predictive uncertainty than quantities that are localized in space-time. While this may be true for systems that are statistically homogeneous and stationary, such systems are the exception rather than the rule. Spatial and temporal averages are not necessarily more predictable as traditionally believed if the averaging covers a scale that contains a strong transition or change in behavior (Entekhabi, 2000).

This and other scale-related problems are generally avoided if the averaging is done in probability space. Mathematical illustrations of averaging in probability space are provided in Appendices C and D. Such "ensemble" averaging leads to stochastic equations that contain statistical moments of hydrogeologic variables (considered random), most commonly the mean and variance-covariance. The mean is a predictor of system behavior or performance, and the variance-covariance is a measure of predictive error. Both are smoother (vary more slowly in space-time) than their random counterparts and, in this sense, render the model relatively "simple." Despite their smoothness, both moments are defined on the same space-time scales as are the random hydrogeologic variables on which they are based. Stochastic models thus achieve smoothness and simplicity without any need to average or upscale in space-time. As they are typically conditioned on site measurements (*i.e.*, they honor the data), stochastic models are compatible with these measurements both in scale and magnitude. Yet another advantage of the stochastic method over space-time averaging is that it yields measures of predictive uncertainty. Stochastic approaches are increasingly recognized as offering a way to deal with complex, scale-dependent heterogeneous systems by means of relatively simple models (Hunt and Zheng, 1999).

Whenever possible, simplified and/or abstracted models should be deduced systematically and objectively from, be comparable with, and be judged and validated against their more comprehensive (complex and complete) conceptual-mathematical counterparts. The strategy strongly suggests postulating and evaluating the latter (at least on paper, not necessarily implementing them on the computer) before attempting to develop and validate the former.

6.2.4 How to Factor Uncertainty Into a Conceptual-Mathematical Model?

Information about site- and regional-scale hydrogeology is typically sparse and uncertain. To account for this uncertainty, it is necessary to treat hydrogeologic quantities probabilistically (as random variables) or stochastically (as correlated random fields or processes).

The traditional approach to hydrologic model uncertainty analysis has been to postulate a deterministic conceptual-mathematical model structure and treat its parameters as being imperfectly known. The present strategy considers this approach as one of two options.

The other option is to postulate a stochastic rather than a deterministic conceptual-mathematical model structure. It considers that hydrogeologic environments consist of natural soils and rocks, which exhibit both systematic and random spatial variations in hydraulic and transport properties on a multiplicity of scales. Groundwater flow and solute transport take place under the action of forces whose exact nature, magnitude and

space-time distribution is generally uncertain. Traditional deterministic methods of analyses capture at best the larger-scale, systematic components of these variations and distributions. They however fail to resolve smaller scale variations or account for their uncertain nature. Yet capturing hydrologic complexity on a wide range of scales, and quantifying the associated uncertainty, is proving to be of paramount importance for a wide array of applications such as the characterization and remediation of contaminated sites, risk analysis and monitoring design for sites subject to potential future contamination, and performance assessment of nuclear facilities and sites.

Under the stochastic option, one describes the spatial variability and scaling of hydrogeologic medium properties geostatistically, and analyzes subsurface fluid flow and solute transport stochastically. A key component of geostatistical characterization is the assessment of spatial covariance structure (in terms of a covariance or variogram function). In the absence of site characterization data that are amenable to geostatistical analysis, one can treat the structural parameters (of the covariance or variogram function) as free parameters to be estimated by model calibration.

The strategy supports two general approaches to predict stochastic groundwater flow and transport on the basis of geostatistical input: High-resolution numerical Monte Carlo simulation and direct (deterministic) prediction of mean behavior. In the more widely known Monte Carlo approach, one generates a large number of equally likely parameter input fields, feeds them into a standard deterministic flow and transport simulator that however uses a very fine computational grid, and produces a large number of model outputs representing equally likely system behaviors. These nonunique outputs are summarized in terms of statistically averaged quantities, their variance-covariance, and perhaps higher moments of the corresponding sample probability distributions. Results that honor measured values of medium properties are said to be conditioned on these data. Upon conditioning the simulations on measured values of parameters in space, one obtains (among others) conditional mean flow and transport variables that constitute optimum unbiased predictors of these unknown random quantities. One also obtains conditional second moments (variance-covariance) that provide a measure of the associated prediction errors. To condition the predictions not only on measured input variables (derived from site characterization) but also on monitored system behavior tends to improve their reliability. This is accomplished by either discarding random simulations that do not reproduce the observations, or employing a suitable inverse procedure.

The more novel and less widely known direct approach relies on stochastically derived deterministic ensemble mean flow and transport equations in which the dependent variables represent not actual system states (such as head, concentration or flux) but rather their (ensemble) mean values or statistical expectations. These mean or expected values represent optimum unbiased predictors of the unknown, actual system states (similar to those obtained upon averaging the results of numerous Monte Carlo simulations). To condition the predictions on system monitoring data, one must employ a special inverse procedure.

The variance-covariance of high-resolution (conditional) Monte Carlo simulations serves to quantify predictive uncertainty. Another option is to compute this variance-covariance directly (deterministically) by means of corresponding moment equations.

Various software packages are available to help one implement the geostatistical and Monte Carlo options in two and three spatial dimensions. For up-to-date information about software and literature concerning the statistics and geostatistics of spatial data the reader is referred to AI-GEOSTATS (www.ai-geostats.org), a central server for GIS and spatial statistics run by the University of Lausanne in Switzerland. A number of publicly or commercially available software packages are listed below.

GMS: This system was described in Chapter 4.

UNCERT: UNCERT is a public domain uncertainty analysis and geostatistical software package for groundwater flow and contaminant transport modeling, available from the Department of Geology and Geological Engineering at the Colorado School of Mines (uncert.mines.edu). It includes a number of software modules that allow one to 1) input raw field data or data from a pre-existing database, 2) analyze the data using classical statistics, 3) evaluate trends, 4) evaluate the data using geostatistical techniques such as semivariogram analysis, various kriging techniques (simple, ordinary, indicator, and Bayesian), and stochastic simulation. When the data are analyzed, or prepared from other sources, graphical tools are available to view the results in two-, two-and-a-half-, and three-dimensions. Once the spatial variation of materials has been determined, tools are available to automatically generate finite-difference grids for groundwater flow and contaminant transport codes such as MODFLOW andMT3D, 5) run these models, and 6) evaluate the results of individual runs, as well as the composite results of multiple model simulations. MODFLOW and MT3D are widely used by hydrogeologists to simulate groundwater flow and solute transport, respectively, in three-dimensional saturated porous media.

UNCERT is written primarily in C language and is compatible with Unix and Linux computational platforms. Code options include logarithmic transformation of data; generation of histograms, probability plots, variograms, cross-variograms, covariance functions, indicator variograms and covariance functions; jackknifing; manual or automatic modeling of standard spatial correlation functions; estimation via trend surface analysis as well as ordinary, simple and indicator kriging; mathematical manipulation of numerical series associated with two-dimensional and three-dimensional grids; and Monte Carlo simulation of random fields conditioned on both hard and soft data. In addition to a manual interface, UNCERT connects with GIS packages such as Arc/Info and Grass, and offers a graphical interface for MODFLOW and MT3D.

The philosophy behind UNCERT is that there are a multitude of possible interpretations of the subsurface which honor any given set of data. To evaluate the alternatives manually would take

considerable time and still only a small portion of the possibilities could be evaluated. This is true even when the subsurface configuration is relatively simple. In order to evaluate this inherent uncertainty, computers can be used to create multiple alternative realizations of the subsurface. The process can be forced to honor hard data (well logs, etc.) by using indicator kriging techniques, and incorporate more uncertain data (soft data - data with a range of uncertainty, *e.g.,* seismic information, geophysical well logs, expert opinion) through Bayesian kriging. By automating this process, much of the uncertainty can be characterized with comparatively little time invested by the hydrogeologist. Once multiple realizations are created, groundwater flow and contaminant transport models can be executed to compare modeled and field conditions. When a model response clearly doesn't match field observations, this possible subsurface configuration can be disregarded; of the remaining realizations (invalidating 90% of the realizations might not be unreasonable) that appear reasonable, the distribution of contaminants may be evaluated, for the time already modeled, or for future conditions. Based on the results of flow and transport modeling in these remaining configurations, the probable locations of contamination may be identified. Also, the probable effectiveness of remediation facilities designed to contain the contamination can be evaluated. A computer can evaluate only a limited number of realizations, but the number is so large, relative to that which can be accomplished manually, that a representative assessment of the reasonable alternatives will be realized.

The modeler can start with hard and/or soft field data and be guided through statistical analysis of the data, generation of multiple realizations of each data property, development of model grids, kriging of data properties into model grids, generation of input files for flow and contaminant transport models, execution of models, and visualization of model results.

Gstat: Gstat is a program for the modeling, prediction and simulation of geostatistical data in one, two or three dimensions. It is made available free of charge by the Department of Geography at Utrecht University in the Netherlands (www.gstat.org) and works on various platforms including Linus and win32.

Geostatistical modelling comprises calculation of sample variograms and cross variograms (or covariograms) and fitting models to them. Sample (co-) variograms are calculated from ordinary, weighted or generalized least squares residuals. Nested models are fitted to sample (co-)variograms using weighted least squares, and during a fit each single parameter can be fixed. Restricted maximum likelihood estimation of partial sills and interactive variogram modeling are possible.

Gstat provides estimation and prediction using a model that is the sum of a trend modeled as a linear function of polynomials of the coordinates or of user-defined base functions, and an independent or dependent, geostatistically modeled residual. This allows simple, ordinary and universal kriging; simple, ordinary and universal cokriging; standardized cokriging; kriging with external drift, block kriging and trend kriging; as well as

uncorrelated, ordinary or weighted least squares regression prediction. Gstat allows univariate or multivariate conditional or unconditional multi-Gaussian sequential simulation of point values or block averages, or (multi-) indicator sequential simulation.

Gslib: The Gslib code is included on CD with the book of Deutsch and Journel (1998). It is written in Fortran language for DOS and Unix operating systems. Code options include the generation of histograms, probability plots, two- and three-dimensional variograms and cross-variograms; point or block estimation by ordinary, simple and universal kriging, kriging with an external drift, cokriging, nonlinear kriging, indicator principal component kriging, and soft kriging; cross-validation and jackknifing; and conditional Monte Carlo simulation by indicator principal components, LU-decomposition, simulated annealing, sequential Gaussian simulation, sequential indicator simulation of a categorical or continuous variable, and the turning bands method. Output is in postscript format but can be displayed in a graphical raster format by means of the UPFILE and 3Plot codes that are available without charge.

FSS: This set of tools for DOS, Unix and Linux platforms is available from FSS International Europe, a consulting firm based in Valencia, Spain (www.fssintl.com). It uses the Gslib data format to perform univariate and bivariate statistical analyses of data; to generate variograms, covariance and correlation functions, H-scattergram and variogram plots; to estimate random field values by ordinary and simple kriging with global or local mean, cokriging, collocated cokriging, universal kriging, indicator ordinary and simple kriging, and multi-Gaussian kriging; and to conduct Monte Carlo simulations by sequential Gaussian and indicator methods.

The direct stochastic modeling option is supported by a menu of exact conditional moment equations developed for steady state (Neuman and Orr, 1993; Neuman *et al.*, 1996) and transient (Tartakovsky and Neuman, 1998, 1999) groundwater flow in saturated porous media, steady state flow in a certain class of unsaturated soils (Tartakovsky *et al.*, 1999; Lu *et al.*, 2002), and advective (Neuman, 1993) or advective-dispersive (Zhang and Neuman, 1996) transport of a nonreactive solute.

Even though the underlying stochastic flow and transport regimes are taken to obey Darcy's law and Fick's analogy, these relations are generally not obeyed by the conditional mean flow and transport regimes except in special cases or as localized approximations. Such localized approximations yield familiar-looking differential equations which, however, acquire a non-traditional meaning in that their parameters (hydraulic conductivity, seepage velocity, dispersivity) and state variables (hydraulic head, concentration, fluid and solute fluxes) are information-dependent and thus nonunique (Guadagnini and Neuman, 1999a-b, 2001; Ye *et al.*, 2002). Whereas nonlocal moment equations contain information about predictive uncertainty, localized (differential) equations generally do not.

There are presently no publicly available computer codes to implement the conditional moment approach, though some prototype codes exist and other are being developed. To implement the approach, one

must start by adopting a suitable closure approximation such as perturbation (Guadagnini and Neuman, 1999a-b, 2001; Ye *et al.*, 2002; Lu *et al.*, 2002) or an assumption of Gaussianity (Neuman *et al.*, 1999; Amir and Neuman, 2001; Wang *et al.*, 2002). Though perturbative closure approximations are not guaranteed to work for strongly heterogeneous media, they have proven to work remarkably well in two-dimensional finite element analyses of superimposed mean uniform and convergent steady state (Guadagnini and Neuman, 1999a-b) and transient (Ye *et al.*, 2002) flows in saturated porous media and for similar type steady state flow in unsaturated media (Lu *et al.*, 2002). A different perturbative approach to conditional moment analysis and its various applications can be found in the recent book of Zhang (2001).

6.2.5 How To Interpret Traditional Deterministic Models Stochastically?

A comparison of stochastic moment equations of flow and transport with traditional deterministic equation shows that the latter can be interpreted stochastically. This is seen upon "localizing" conditional mean equations that are otherwise nonlocal (integro-differential) in a way which renders them approximately differential (Appendices C and D). Localized conditional mean flow and transport equations can be solved with the aid of standard flow and transport codes or modifications thereof. While this is an important advantage, a major disadvantage of localized equations is that they do not yield information about predictive uncertainty.

The prevailing tradition has been to model flow and transport by means of standard (local) deterministic equations. In Appendices C and D we examine this tradition from a stochastic viewpoint and comment on its validity and meaning. More specifically, we ask what is the meaning of hydrogeologic variables that enter into traditional deterministic groundwater flow and transport models, when the latter are applied to randomly heterogeneous media? What is the significance of parameters obtained by calibrating a deterministic flow model against randomly varying data? What do corresponding parameter estimation errors imply about predictive uncertainty?

We show, in a rigorous manner, that applying traditional deterministic flow and transport models to randomly heterogeneous media is valid at best as an approximation, due to localization. The following are some important implications of interpreting traditional deterministic flow and transport models stochastically: (1) their parameters are inherently nonunique, regardless of how well or poorly posed an associated inverse problem may be; (2) these parameters differ from their optimum (geostatistical) estimates as obtained on the basis of field measurements; and (3) uncertainties in these parameters reflect uncertainties in predicted (conditional) mean system behavior, not uncertainties associated with actual system behavior.

It follows that analyzing the effect of parameter uncertainty on system behavior and performance by means of traditional deterministic models generally underestimates (often severely) behavior and performance uncertainty. Therefore, the strategy in this report supports the use of traditional deterministic models only to the extent that they are interpreted and implemented in their proper stochastic context, as illustrated in Appendices C and D.

6.3 How to Formulate and Explore Alternative Conceptual-Mathematical Models?

6.3.1 Where to Start?

The qualitative stage of conceptualization should lead to several alternative hypotheses concerning the hydrogeologic makeup of a site, the prevailing flow and transport regimes, and the way in which these regimes are expected to evolve under future scenarios of interest. With each of these, one must associate at least one mathematical structure to form a corresponding conceptual-mathematical model. This structure should initially include a direct representation of site-scale hydrogeologic complexities, heterogeneities and driving-mechanisms in the form of explicit model elements. It should account directly for larger- or regional-scale features and factors through the appropriate assignment of initial conditions, boundary conditions, and source terms such as those that describe infiltration, recharge, discharge and leakage across aquitards. If features and factors that manifest themselves on subsite scales are deemed important for the modeling of site-scale phenomena, one must account for them explicitly or implicitly at this initial stage of conceptual-mathematical model building. This may result either in a deterministic or a stochastic model structure.

Ignoring the influence of regional-, site- or subsite-scale features or factors at this initial stage, without having demonstrated that they are not relevant to the problem at hand, constitutes an oversimplification which renders the model *a priori* suspect of being invalid and unreliable. This is true even if the model can be made to reproduce observed system behavior, because there is reason to suspect that it would be unable to predict reliably behaviors under future scenarios, which differ in significant ways from that under which the observations have been made.

The initial conceptual-mathematical model should identify the sources and nature of uncertainties that may impact its predictive power. This includes uncertainties in the very structure of the model, as well as uncertainties in its input parameters (material properties and forcings). Both deterministic and stochastic models are expected to have uncertain structural elements and parameter values.

Having associated a deterministic and/or stochastic mathematical structure with each hypothesis that had been postulated for a site, the next step is to explore them quantitatively and graphically. The aim of this initial exploration is qualitative: To identify the potential of each conceptual-mathematical model to explain observed system behavior and predict it under future scenarios. For this, it is useful to run the model with a plausible range of input parameters and study its output in light of the available site information. This may reveal ambiguities and uncertainties that may prompt a reexamination of the data and/or a revision of the model (which may, but need not, render the model more complex). It may also reveal differences between the various conceptual-mathematical models that may help eliminate some of them and rank the rest in the order of their apparent ability to explain and predict system behavior.

The initial exploration of alternative conceptual-mathematical models can often be done in the two-dimensional horizontal and/or vertical planes. This is so because two-dimensional analyses differ from their three-dimensional counterparts to a much lesser extent than do one-dimensional

analyses, yet are much less demanding than three-dimensional analyses. Two-dimensional scoping analyses should yield contour and/or color representations of hydrogeologic parameter distributions; system states (hydraulic head, pressure, saturation, density, temperature, solute concentration); flowlines (streamlines, pathlines, streaklines); and vectors of fluid and solute fluxes and velocities (Figures 6-2 – 6-7). If the conceptual-mathematical model involves transients, the analyses should consider variations of hydrogeologic variables with time. Some three-dimensional exploratory analyses may also be possible at this stage, including the preliminary quantification of fluid, solute and energy balances. Three-dimensional analyses may also be required if the two-dimensional results are insufficient to provide insight into relevant processes or fail to resolve key ambiguities and uncertainties.

6.3.2 How to Proceed?

Once an initial set of conceptual-mathematical models have been formulated and explored in a preliminary fashion, once has the option of postulating and exploring less complex alternatives to each. There is no reason to contemplate more complex alternatives at this stage of the modeling process. Model simplification may entail a reduction in dimensionality (from three spatial dimensions to two or one; from transient to steady state), model size (smaller area and/or reduced depth); or details of various features, events and processes (fewer layers, faults or fracture zones; a less detailed representation of internal heterogeneity; single rather than dual continuum or discrete representation of a fractured rock; fewer discrete rainfall or infiltration events; a less detailed delineation of contaminant sources; constant rather than scale-dependent dispersion; equilibrium rather than kinetic sorption). Various types and levels of simplification can be entertained, leading to a number of simplified model structures for each initial (and more complex) conceptual-mathematical model.

As stated earlier, the strategy discourages the use of one-dimensional models unless a very strong and convincing hydrogeologic argument is made in their favor. This is so because system behavior and performance predicted by a one-dimensional model tend to be fundamentally different from those predicted by two- and three-dimensional models.

Some of the details lost in the simplification process may have an important impact on the predictive capabilities of the model and some may have a lesser impact or none at all. This should be established by comparing each simplified model with its more complex initial version *vis-a-vis* their potentials to explain observed system behavior and predict it under anticipated future scenarios. For this, it is useful to run both models with a comparable range of plausible input parameters and compare their outputs in light of the available site information. To do so requires that any differences in scale between the input and output variables of the two models be resolved so as to render the models mutually compatible and comparable. Inability to do so in a scientifically defensible way should disqualify the simplified model. Other factors that should disqualify a simplified model are a much reduced ability (in comparison to the initial model) to explain observed system behavior that is considered important for the problem at hand, or an inability to predict future system behavior and/or performance that are qualitatively and quantitatively similar to those predicted by the initial model.

To render a simplified model compatible and comparable with its more complex (initial) alternative, it helps to derive it formally and objectively from the latter. This can be done through averaging or filtering in space, time and/or probability space. Appendix B shows in detail how the three-dimensional equations that govern flow in a water-table (unconfined) aquifer are formally averaged over the vertical, based on the well-known Dupuit assumptions, so as to yield the corresponding Boussinesq equation of horizontal flow. This formal analysis reveals with clarity under what conditions is one justified simplifying unconfined flow by restricting it to the two-dimensional horizontal plane. Appendix C explains and illustrates how the steady state stochastic groundwater flow equation, and corresponding hydrogeologic variables, are formally averaged over an ensemble of random hydraulic conductivity, head and flux realizations. Appendix D does the same for transport.

Yet another important example is provided by Domenico and Robbins (1984) who demonstrate formally that characterizing the two- or three-dimensional spread of a solute with a one-dimensional transport model requires an apparent dispersivity, which increases as a function of distance. This is so even when the dimensionally correct (two- or three-dimensional) transport model is associated with a constant dispersivity. It follows that replacing the latter model by a one-dimensional version with constant dispersivity is an error, which does not necessarily lead to conservative results.

A fifth example is provided by the theoretical and field study of Neuman and Witherspoon (1972) on multiple (leaky) aquifer systems. A pumping test conducted by these authors near Oxnard, California, has shown that when drawdowns in the pumped Oxnard aquifer are interpreted by means of the well-known and widely-used Theis model, the aquifer transmissivities appear to increase systematically with distance from the pumping well. The authors demonstrate theoretically that this apparent aquifer heterogeneity is, at least in part, an artifact of having simplified a complex multiaquifer flow regime by means of a model, which disregards interactions between the Oxnard aquifer and surrounding hydrogeologic units. Other examples abound, which collectively demonstrate that simplifying hydrogeologic flow and transport models on the basis of intuition or convenience, rather than on the basis of a more comprehensive formal analysis, may lead to serious errors of omission and commission.

7 RENDERING A MODEL COMPATIBLE WITH DATA

Once an acceptable set of conceptual-mathematical models and corresponding simplifications have been identified, the next step is to render them compatible with available site characterization and monitoring data. Models that cannot be so rendered are unsuitable for quantitative representation of site hydrogeology and must be eliminated from further consideration.

7.1 Why Worry About Compatibility With Data?

A fundamental premise of this strategy is that a model must be operational, *i.e.*, that it must be based on quantities which are either measurable, or can be derived from measurements, on well-defined space-time scales. The quantities we speak of include all parameters, initial and forcing terms (including sources and boundary conditions), and state variables that enter into the model (such as pressure and concentration). For example, if the flow model is based on Darcy's law, and this law is deemed relevant on a field length scale of 1 m (equivalently, area scale of 1 m^2 or volume scale of 1 m^3), then the Darcy flux, permeability, and pressure or its gradient, must be amenable to direct or indirect measurement or evaluation on this same scale. The latter does not imply that it must be practical to perform such measurements within every cubic meter of the rock, only that it must be possible to do so in a sufficiently large sample of similar rock volumes to yield a statistically significant sample. Only then can a model be meaningfully tested against experiment and/or observations.

A model that includes nonmeasurable quantities, or quantities that cannot be evaluated on the basis of measurements, can neither be applied directly to real data nor confirmed experimentally. Such a model is therefore non-operational and can neither be tested nor confirmed.

7.2 How to Achieve Compatibility in Scale and Magnitude?

To render a model compatible with site data, one starts by asking whether the scales of parameterization and resolution in the model are consistent with the support (measurement) scales of relevant data (site characterization data that define model structure and parameters, and site monitoring data that record observed state variables) and their statistics. If the answer is negative, one needs to either rescale the data and their statistics to fit the model, or to rescale the model to fit the available (as well as anticipated) data and their statistics.

In many cases, the support scale of hydrogeologic data is smaller than the scale of parameterization and resolution of corresponding hydrogeologic models. In recent years, there has been a concerted effort on the part of many researchers to develop methods for the scaling up of hydrogeologic variables, most notably permeability. As explained by Neuman (1997a), the search has focused in large part on methods of upscaling that ascribe equivalent parameters to the grid blocks of numerical flow and transport models on the basis of smaller-scale random or nonrandom parameter values. One approach has been to postulate more-or-less *ad hoc* rules for upscaling based on numerically determined criteria of equivalence. Another approach has been to develop upscaling rules analytically through volume or stochastic averaging of smaller-scale values. Both sets

of rules are still evolving and it may be premature to recommend specific methods of upscaling for the purposes of this strategy. Instead, we mention two examples.

Neuman and Depner (1988) have shown how one can upscale hydraulic conductivity data from single-hole hydraulic injection tests in fractured crystalline rocks at the Oracle site near Tucson, Arizona, as well as their spatial statistics. Their approach is based on a stochastically derived steady state formula due to Gelhar and Axness (1983) that applies to statistically homogeneous and infinite continua (a similar formula for finite three-dimensional blocks, under either steady state or transient conditions, has been developed by Tartakovsky and Neuman, 1998). The authors compare their upscaled values successfully with larger-scale hydraulic conductivities from cross-hole injection tests at the site.

The above approach does not work at the Apache Leap Research Site (ALRS) in Arizona, where air permeability is not statistically homogeneous but exhibits the properties of a random fractal field (Chen *et al.*, 2000). A method to rescale the permeability of such a field has recently been proposed by Di Federico *et al.* (1999) and used by Hyun *et al*. (2002) to interpret a pronounced permeability scale effect at the ALRS.

A rule of thumb that often works reasonably well in practice is to upscale permeability by considering the geometric mean of smaller-scale values. This is known to be exact (Neuman and Orr, 1993) for a statistically homogeneous Gaussian log permeability field in an infinite planar domain, but is less accurate for three-dimensional or finite domains (and entirely unsuitable for one-dimensional domains, for which the harmonic mean is more appropriate). Despite this potential inaccuracy, the geometric mean is often used in conjunction with block kriging for the purpose of ascribing equivalent permeabilities to finite subdomains of a site on the basis of smaller scale measurements.

Rubin *et al.* (1999) used a stochastic approach to develop analytical expressions for a grid-block effective dispersivity tensor due to subgrid-scale random variations in permeability. Their expressions can be conditioned on measurements of permeability.

Efendiev *et al.* (2000) developed a strategy for incorporating subgrid effects in coarse-scale numerical models of flow in heterogeneous media by upscaling a deterministic fine-grid permeability description and then solving the flow equation over the coarse grid to obtain coarse-scale velocities. A coarse-grid saturation equation is formed through a volume average of the fine-scale equations and includes terms involving both the average and fluctuating components of the velocity field. The terms involving the fluctuating components are subgrid effects that appear as length- and time-dependent dispersivities. A simplified model for the coarse-scale dispersivity, in terms of these subgrid velocity fluctuations, was proposed and a numerical scheme based on it was implemented in two dimensions.

Some practitioners set the ratio between the longitudinal dispersivity assigned to a numerical grid block, and the length scale of this block, proportional to 10. This is based on a generalized scaling rule proposed by Neuman (1990) as a way of explaining an observed increase in apparent longitudinal dispersivity with the scale of observation in a large number of tracer studies, in diverse

hydrogeologic environments under varied circumstances, worldwide. It is not clear that the same scaling rule applies locally to specific sites under particular conditions of flow and transport.

Numerical methods to upscale unsaturated soil hydraulic properties have been discussed by Bagtzoglou *et al.* (1994) and Desbarats (1995, 1999). Rockhold *et al.* (1999) use geostatistical indicator simulation techniques for spatial interpolation of field-measured water contents and porosities, and a conditional simulation method based on similar media scaling for estimating hydraulic properties from a set of scale-mean parameters and the initial water content and porosity distributions. They then apply an upscaling algorithm to determine effective grid-scale parameters in three-dimensions.

As pointed out by Neuman (1997a), one major difficulty with upscaling is that it postulates local relationships between upscaled driving forces and fluxes (Darcy's and Fick's laws) when in fact these relationships are generally nonlocal. Another conceptual difficulty with upscaling is that it requires the *a priori* definition of a numerical grid and resolution scale. We add that improper upscaling may rendered the prior parameter estimates biased and their uncertainty mischaracterized.

Upscaling can often be avoided if one adopts a stochastic approach to groundwater modeling. One choice under the stochastic option is to use a high-resolution conditional Monte Carlo method. In this method, the grid is fine in comparison to the spatial correlation scale of the hydrogeologic input variables, most importantly the permeability. Ideally, one would want the length scale of each grid cell to be of the same order as that of the input variable support scale. In practice, it is usually enough to set the length scale of each grid cell equal to or less than one fifth the spatial correlation scale of the input variable. Experience has shown that with such a grid, one is often able to reproduce random spatial fluctuations in the input variable with high fidelity. As the correlation scales of output variables (such as head, concentration and flux) are generally larger, this also allows resolving the latter with equal or higher fidelity.

Another choice under the stochastic option is to employ conditional moment equations. Here scale-related problems are generally avoided because the input as well as the output variables represent ensemble averages in probability space rather than volume and/or time averages in real space. Though they are smoother (vary more slowly in space-time) than their random counterparts, these variables are nevertheless defined on the same space-time scales. Hence stochastic moment equations achieve smoothness and simplicity without any need to average or upscale hydrogeologic variables in space-time. By being smoother than their random counterparts, ensemble moments can in principle be resolved on grids that are coarser than those required for high-resolution Monte Carlo simulation.

Stochastic Monte Carlo and moment equation models honor site characterization data by being conditioned upon them. They are thus *a priori* compatible with site data both in scale and magnitude. Their output can be made similarly compatible with site monitoring data by conditioning the models *a posteriori* on observed behavior through appropriate inverse procedures. By doing so, the models may still honor the site characterization data and so remain conditional on both data types.

Site characterization data may be used to derive prior input parameters for a deterministic model through appropriate upscaling and estimation procedures. In this limited sense, the deterministic model can be made to honor the site characterization data and so being conditional upon them as well as compatible with them *a priori* in both scale and magnitude. To render the same model compatible with (conditional on) site monitoring data, it is almost always necessary to calibrate them *a posteriori* against such data. This invariably modifies the model input parameters to generate posterior estimates, which often differ significantly from their prior counterparts. It is therefore generally not possible to render a deterministic model compatible in magnitude with both site characterization and monitoring data at the same time and to the same degree. The best one can do is to seek an optimum tradeoff between model fit (a measure of compatibility with the monitoring data) and parameter plausibility (a measure of compatibility with the characterization data) either by viewing the problem in a multiobjective context as proposed by Neuman (1973), or by incorporating the priors into a statistical statement of the inverse problem as proposed by Carrera and Neuman (1986a-b). Another (we think less desirable) possibility is to calibrate the model without considering prior parameter estimates (and, by implication, site characterization data on which such estimates are based) at all, as recommended by Hill (1998). More on these and other inverse modeling options later.

7.3 How to Infer Prior Input Parameters and Statistics?

7.3.1 What to Infer from Site Characterization Data and How?

Since hydrogeologic medium properties and forcing functions are always uncertain, so are the model input variables. In other words, the input parameters of groundwater flow and transport models are inherently uncertain regardless of whether the model is deterministic or stochastic. The difference is that whereas in deterministic models the input parameters are viewed as imperfectly known deterministic quantities, in stochastic models they are viewed as correlated random fields or processes that may be perfectly or imperfectly known at discrete sampling locations in space-time. In both cases, imperfect knowledge is characterized by random errors that may, but often are not, considered to be mutually correlated. This renders the input parameters of deterministic models correlated or uncorrelated random variables, and those of stochastic models correlated random fields or processes conditional on either exact or random measurements, which may themselves be correlated or uncorrelated among themselves. Whereas the second moment of uncorrelated random variables is a diagonal matrix of their respective variances, that of correlated random variables is a full square symmetric and positive definite matrix of their respective covariances.

Estimating input parameters for a model (deterministic or stochastic) on the basis of incomplete and/or uncertain data is equivalent to inferring their mean values from these data. If the mean values are based on site characterization data, they represent prior parameter estimates. If they are based on site monitoring data, they represent posterior parameter estimates. In this section we consider only prior inference of model input parameters.

If the inferred parameter estimates vary in space and/or time in a way which reflects similar variations in the underlying site characterization data, they are said to be conditional on these data. If the inferred mean values do not reflect such spatial and/or temporal variability, they are said to be unconditional. Unconditional inference may take place when the data are too few or too clustered to allow defining their variability in space-time.

Assume that a set of site specific measurements are available, which represent the same hydrogeologic variable (say permeability or porosity) as a corresponding set of model input parameters. Such measurements are considered to be "hard." Ideally, the data would include probabilistic information about errors of measurement and test interpretation that suffer from a known amount (ideally zero) of statistical bias. If one has a statistically significant set of such data, one should be able to estimate a prior set of model input parameters on their basis. One should also be able to postulate a probabilistic model of prior parameter uncertainty based on statistics derived from these data. Such a probabilistic model is known to be of Type A.

A key measure of parameter uncertainty is the second statistical moment, or variance-covariance, of their estimation errors. Since input variables into both deterministic and stochastic computational models are specified in terms of a discrete set of input parameters, the corresponding variance-covariance forms a matrix. In the conditional case, the off-diagonal covariance terms of the matrix may reflect spatial and/or temporal correlation between parameters of a given type (say permeability or porosity) as well as cross-correlations between parameters of different types (say permeability and porosity). In the unconditional case, there is no inferred space-time variability and the covariance terms represent at most cross-correlations between different types of parameters.

Prior parameter estimates can be obtained from clustered data by means of common statistical methods, and from spatially distributed data by means of standard geostatistical techniques. This is true regardless of whether the parameters are intended for a deterministic or a stochastic model. The main difference is that in the deterministic case, there may be a need to upscale the parameters whereas in the stochastic case, this may not be necessary. The simplest and most practical form of geostatistical inference with upscaling is block kriging. While this may not always be the most accurate and sophisticated way to proceed, it is quite adequate for many purposes. This is especially true in situations where enough site monitoring data are available to later modify the parameter estimates through model calibration.

To characterize prior estimation uncertainty one should, as a minimum, infer from the data a variance-covariance matrix of prior estimation errors. Only in rare circumstances would there be enough data to permit inferring from them higher statistical moments of these errors.

Standard geostatistical analysis consists of identifying the spatial autocovariance structure of each variable being analyzed (in terms of an autocovariance or variogram function) and providing a smooth (kriged) conditional estimate of this variable on a two- or three-dimensional grid, as well as the associated estimation (kriging) variance at each grid point. A more advanced analysis may also yield an autocovariance matrix for the estimate across the grid. Another level of sophistication may be achieved by identifying cross-covariance functions or cross-variograms for two or more variables, estimating them simultaneously by cokriging, and computing their co- and cross-covariances across the grid. Both kriging and cokriging may be used to estimate average values of the variables over finite blocks or subdomains of a two- or three-dimensional grid.

7.3.2 How to Incorporate Soft Data?

If there are insufficient hard data of a given hydrogeologic variable to conduct a meaningful statistical or geostatistical analysis, then the use of soft (qualitative) data coupled with indicator geostatistical analysis are recommended. This yields an uncertainty model for the prior parameters that is intermediate between Type A and Type B (defined in the next section). Soft or indirect information about the parameters may include (a) off-site measurements of the parameters proper (quite often on scales other than those corresponding to the intended scale of model resolution) and/or (b) surrogate measurements on site that are known to correlate with the parameters of interest (for example, porosities or geophysical signatures that correlate in known ways with permeabilities, water contents or fracture densities). Statistics derived from off-site data must be considered potentially biased (due to a lack of site-specific information about mean parameter values and incompatibility of geology and scale). The associated variance may be too small or too large, depending on the quantity and quality of such data. Statistics derived from surrogate data may suffer from poorly defined correlations and incompatibility of scale.

Indicator geostatistics yields indicator variograms for various classes of the variable, a smooth (kriged) estimate of this variable on a two- or three-dimensional grid, and the probability that the variable is larger (or smaller) than specified at each grid point. Most geostatistical software packages include an indicator option and allow one to generate random realizations of the variable on a grid by means of indicator Monte Carlo simulation.

An example of soft data use is that of pedotransfer functions, which allow one to estimate soil hydraulic characteristics on the basis of soil textural data. One such software package is Rosetta (www.ussl.ars.usda.gov), developed by Schaap at the U.S. Salinity Laboratory in Riverside, California. The Rosetta code predicts van Genuchten (1980) water retention parameters and saturated hydraulic conductivity on the basis of soil textural class, textural distribution, bulk density and one or two water retention data points. It follows a hierarchical approach that is based on five models, depending on the available data. The simplest model consists of a lookup table of average hydraulic parameters for each textural class such as sand, silty loam, clay loam, etc. The other models rely on neural network analyses (Schaap *et al.*, 1999) of over 2,000 soil samples for water retention, over 1,000 samples for saturated hydraulic conductivity (Schaap and Leij, 1998), and over 200 samples for unsaturated hydraulic conductivity (Schaap and Leij, 2000).

The samples represent a large number of agricultural and non-agricultural soils in temperate climate zones of the northern hemisphere, most from the U.S. and some from Europe. The authors warn the user that the application of Rosetta to other climate zones, in which pedogenic processes are different, could lead to inaccurate predictions. Included in the output are uncertainty estimates of the predicted hydraulic parameters, which can be used to assess the reliability of the predictions.

We have used Rosetta in our analysis of data from the Maricopa Agricultural Center in Arizona as illustrated later in this report. Rossini *et al.* (1994) and Rovellini *et al.* (1998) utilize the concept of facies as the basis for petroleum reservoir characterization. They treat facies as the elementary units of the reservoir, determine them through a log-based cluster analysis, and associate them with petrophysical properties from cores. In the case study of Rossini *et al.*, the reservoir shows extreme variability, comprising all the transitional lithologies from sand to dolomite. Due to the limited continuity of its lithologic facies, their spatial distribution is difficult to identify. The authors started by identifying deterministically three hydraulically separated layers in a vertical cross-section, distinguished in part by their diverse original gas-oil and oil-water contacts. After developing a petrophysical criterion to distinguish between sandy and dolomitic facies, they established a frequency distribution of porosities, and a correlation between log permeabilities and porosities, for each facies on the basis of core data. Likewise, they developed horizontal and vertical indicator semivariograms of porosity for each facies. Next, the authors divided the three layers into 1,645,000 cells measuring 50 m horizontally and 0.5 m vertically. They then generated ten equally likely random images of the facies across this grid by conditional stochastic indicator simulation, and assigned random porosities and permeabilities to grid blocks within each facies. To simulate flow through each of the ten generated reservoirs, the authors superimposed a coarse grid over the original fine grid and assigned an upscaled porosity and permeability to each coarse grid block. Rovellini *et al.* (1998) also used a stochastic model to distribute facies and related petrophysical parameters (porosity, permeability and irreducible water saturation) within the volume of a reservoir. They then generated several equally probable realizations of the reservoir and ran it through a flow simulator. The realization providing the best history match was selected for production forecasts.

7.3.3 How to Use Generic and Subjective Probabilities?

If the available hard and soft data are not amenable to geostatistical analysis (due to insufficient information about their spatial location, an inappropriate spatial pattern, insufficient number and/or poor quality of data), an alternative is to rely on generic and/or subjective probabilities and statistics. Doing so is equivalent to postulating a Type B statistical model of prior parameter uncertainty. Such a model should always be suspected of suffering from an unknown amount of statistical and personal bias. Statistical bias is introduced due to lack of site-specific information about mean values of the parameters in question. A personal bias tends to manifest itself in the form of assigned uncertainty measures (most importantly bias and error variance) that are either too small or too large. The first is a manifestation of over-confidence in the model parameters, the second of unduly low confidence in their values.

Generic probabilities for unsaturated hydraulic properties associated with various soil textures are given by Carsel and Parish (1988) and Meyer *et al.* (1997). An example of generic probability distributions of soil hydraulic properties for sand by the latter authors is given in Figure 11-10. The same authors also describe a Bayesian procedure that makes it possible to update these probabilities, and reduce the corresponding uncertainty in hydraulic parameters, on the basis of site-specific data.

We have used both sets of generic probabilities, and the Bayesian updating procedure of Meyer *et al.* (1997), in our analysis of data from the Maricopa Agricultural Center in Arizona as illustrated later in this report.

7.4 How to Calibrate Traditional Deterministic Models?

7.4.1 Where to Start for what Purpose?

The traditional approach to hydrogeologic modeling has been to postulate a deterministic model structure and treat its parameters as being imperfectly known. One would then derive prior estimates of these parameters from site characterization data (if available) and modify them so as to achieve an acceptable fit between model outputs and available monitoring data (if such exist). The process is known as model calibration, parameter estimation, history matching or inverse modeling. The parameters that yield the best match between observed and recorded system behavior form posterior estimates. It is common to consider the latter to be more suitable for predictive purposes than prior parameter estimates.

Although this is not always done, it is generally advisable to start the model calibration process by postulating a prior parameter uncertainty model. We saw how to do this and mentioned that it may lead to a Type A probability model when prior parameter uncertainty is characterized on the basis of a statistically significant set of site specific measurements that represent the parameter; a Type B model when the uncertainty is characterized by generic and/or subjective probabilities; and an intermediate type model when parameter uncertainty is inferred from indirect information about the parameters. We noted that a Type B model should always be suspected of an unknown amount of statistical and personal bias. A model that is intermediate between Type A and Type B must also be considered potentially biased, with variances that may be too small or too large, poorly defined correlations, and incompatibility of scale.

We noted earlier that improper upscaling may render the prior parameter estimates biased and their uncertainty mischaracterized, regardless of what type probability model one adopts. It is generally believed that model calibration tends to reduce both the bias and the uncertainty in prior parameter estimates by insuring that it reproduces adequately observed system behavior. This should render a calibrated model more reliable as a predictor than an uncalibrated model. Appendix C points out that calibrating a traditional deterministic flow model against measured values of head and flux is tantamount to conditioning it on such measurements. As the model hydraulic parameters are by nature conditional, the very act of adding measured heads and fluxes to the database alter their values. This inherent nonuniqueness persists regardless of whether the inverse problem is well- or ill-posed. It explains why model parameter estimates tend to change every time one redefines the underlying database.

7.4.2 How to Calibrate?

The last thirty years have seen major advances in the development of theories and algorithms for the estimation of deterministic model parameters. Many (though not all) of these theories and algorithms are "statistical" in that they include analyses of parameter estimation uncertainty. Such analyses typically accept, but do not necessarily require, information about prior parameter statistics as input. The output includes *posterior* statistics of parameter estimation errors, which tend to be less biased and smaller than the prior estimation errors. A recent summary and comparison of various statistical inverse methods for groundwater flow models has been published by Zimmerman *et al.* (1998). A detailed set of guidelines for the effective calibration of deterministic groundwater flow models has been prepared by Hill (1998). Among inverse codes that can be used to estimate the parameters of standard deterministic flow and transport models we mention the following:

UCODE: This is a public domain universal inverse modeling tool developed by Poeter and Hill (1998) (www.mines.edu/igwmc/freeware/ucode/) at the Department of Geology and Geological Engineering of the Colorado School of Mines in Golden, and the U.S. Geological Survey in Lakewood, Colorado, respectively. The code is available for DOS, Windows and Unix operating systems. The code is universal in that it works with any model that has suitable numerical (ASCII or text only) input and output files. The code generates sensitivity coefficients and statistics for parameter estimates. It is included as an option in GMS (described earlier).

PEST: This is a universal parameter estimation software developed by Doherty *et al.* (1994). The code is marketed by the International Ground-Water Modeling Center (www.mines.edu/igwmc/) at the Colorado School of Mines in Golden, the Scientific Software Group (www.scisoftware.com) and other vendors. It uses text input-output files and provides estimation statistics including confidence intervals, covariance and correlation matrices between the parameters. A parallel version of the code can run simultaneously on several processors or on multiple machines across a PC network. PEST is an optional component of GMS.

PEST has been used successfully by Vesselinov and Neuman (2001) and Vesselinov *et al.* (2001a-b) to interpret multi-rate single-hole and cross-hole pneumatic injection tests at the Apache Leap Research Site near Superior, Arizona. Our inverse interpretation of the cross-hole data yields tomographic images of air permeability and porosity variations in three-dimensional space. The latter represent geostatistical (kriged) parameter estimates based on pressure monitoring data collected during each cross-hole test. We coupled PEST with a three-dimensional finite volume flow simulator, FEHM (Zyvoloski *et al.*, 1997), an automatic mesh generator, X3D (Trease *et al.*, 1996), and the geostatistical program GSTAT (Pebesma and Wesseling, 1998) mentioned earlier. Our computations were conducted in parallel on 32 processors of the SGI Origin 2000 supercomputer at The University of Arizona in Tucson.

ITOUGH2: ITOUGH2 (inverse TOUGH2) is an inverse code developed by Finsterle (1999a-c) specifically for the TOUGH2 model, which simulates multiphase, multicomponent, non-isothermal flows in multidimensional fractured-porous media. The inverse code is available for a fee from the Earth Sciences Division of Lawrence Berkeley National Laboratory at esd.lbl.gov/ITOUGH2/. Any TOUGH2 input parameter can be estimated based on any observation for which a corresponding TOUGH2 output can be calculated. ITOUGH2 conducts sensitivity analyses and provides statistical information about estimation residuals, uncertainties, and the ability to discriminate among model alternatives. It also allows one to propagate parameter uncertainties so as to quantify prediction errors. Our group has used ITOUGH2 to help interpret infiltration and tracer experimental data from the Maricopa Agricultural Center, as discussed elsewhere in this report.

HYDRUS-2D: This is a Windows-based software package developed by Simunek *et al.* (1999) at the U.S. Salinity Laboratory in Riverside, California (www.ussl.ars.usda.gov). The program simulates water, heat and solute movement in variably saturated media in a vertical plane, a horizontal plane, or a three-dimensional region with a vertical axis of symmetry. The package includes a mesh generator, a graphic interface, and an automatic parameter estimation module for the estimation of selected soil hydraulic and/or solute transport parameters. It is available for a fee from the International Ground-Water Modeling Center at the Colorado School of Mines in Golden (www.mines.edu/igwmc/).

7.4.3 How to Select Model Fit Criteria?

Differences between values of model simulated system states (most commonly hydraulic head, pressure or concentration) at discrete points in space-time, and observed values of these same state variables, are termed "residuals" in model calibration parlance. The residuals are considered to represent errors that are distributed randomly about the simulated model output. An underlying (often tacit) assumption is that the conceptual-mathematical model is exact and associated with an unknown set of "true" parameters. If these parameters were known, the model (structure plus parameters) would be exact. To the extent that it would produce nonzero residuals, this would be entirely due to errors in the monitoring record. These data errors (or their logarithmic transform, as is usually done with permeability and transmissivity) are typically taken to have zero mean and be normally distributed. Hence ideally (in the theoretical event that the true parameters were known), the residuals would exhibit a multivariate Gaussian distribution with zero mean (*i.e.*, they would be unbiased) and a variance-covariance identical to that of the monitoring data.

Model calibration is seen as the process of estimating the model true parameters without bias and as closely as possible. This is accomplished by defining an appropriate calibration criterion (or objective function) in terms of the residuals, and "optimizing" the parameters in a way which comes closest to satisfying this criterion (achieving the objective). The extent to which the criterion is satisfied (the objective achieved) becomes a measure of model fit. Theoretically, the optimized parameters are associated with estimation errors that have zero mean and minimum variance.

The most commonly used measure of model fit is the weighted sum of squared residuals. This is meaningful in the above statistical sense provided the monitoring data errors are mutually uncorrelated and each weight is inversely proportional to the corresponding error variance. Model calibration then reduces to a weighted least squares fit of the model to the data (or regression of the parameters on the data). Since groundwater model parameters are usually related to state variables in a nonlinear fashion, the weighted least squares fit is nonlinear. This is why model calibration must usually be done iteratively.

If all the weights (and variances) are equal, the model fit criterion can be normalized to form a simple sum of squared residuals. Model calibration then reduces to an ordinary nonlinear least squares fitting process (or nonlinear regression).

A less common but still widely used measure of model fit is the generalized sum of squared residuals. Whereas in the previous cases the weights formed a diagonal matrix proportional to the inverse of a diagonal matrix of observational error variances, here the weights form a full matrix proportional to the inverse of a (usually full) variance-covariance matrix. This is statistically meaningful in the previous sense when the errors are mutually correlated. Calibration now becomes a generalized nonlinear least squares or regression process.

If some of the statistical parameters that define the observational errors are unknown, they may sometimes be estimated jointly with the other model parameters by the maximum likelihood method. This entails a likelihood function, which is the likelihood of the parameters given a set of (conditional on) observational data. The latter is the probability density of the (error corrupted and so random) data given (conditional on) the parameters. The objective would be to maximize the likelihood function.

Since the data (or their log transform) are assumed to be normally distributed, the likelihood function is multivariate Gaussian. As such, it includes an exponential term whose negative logarithm is equal to half the weighted (by the inverse covariance matrix) sum of theoretical square residuals. Parameters are estimated by minimizing the negative logarithm of the likelihood function. In the special case where all statistical parameters are known, the negative log likelihood function reduces to the standard generalized least squares criterion.

A maximum likelihood approach to model calibration, which incorporates information about prior parameter statistics into the statement of the inverse problem, was proposed by Carrera and Neuman (1986a-b). It yields a negative log likelihood criterion that includes two weighted square residual terms instead of one. The first is the usual generalized sum of squared differences between simulated and observed state variables. The second is a generalized sum of squared differences between posterior and prior parameter estimates. The corresponding weight matrix is proportional to the inverse covariance matrix of prior parameter estimation errors. The constant of proportionality, λ, is treated as a free statistical parameter that may be estimated jointly with the remaining parameters by maximum likelihood. The authors have done so by calibrating the model for various values of λ, plotting the negative log likelihood (S) against λ, and finding graphically the value of λ that minimizes S (Figure 7-1). Allowing λ to be initially

unspecified means that neither the covariance matrix of the observational data nor that of the prior parameters need be fully specified; it is enough to specify each of them up to its own constant of multiplication. A method to estimate these constants of multiplication on the basis of λ has been described by the authors (see also Carrera *et al.*, 1997).

Including prior information in the calibration criterion allows Carrera and Neuman (1986a-b) 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).

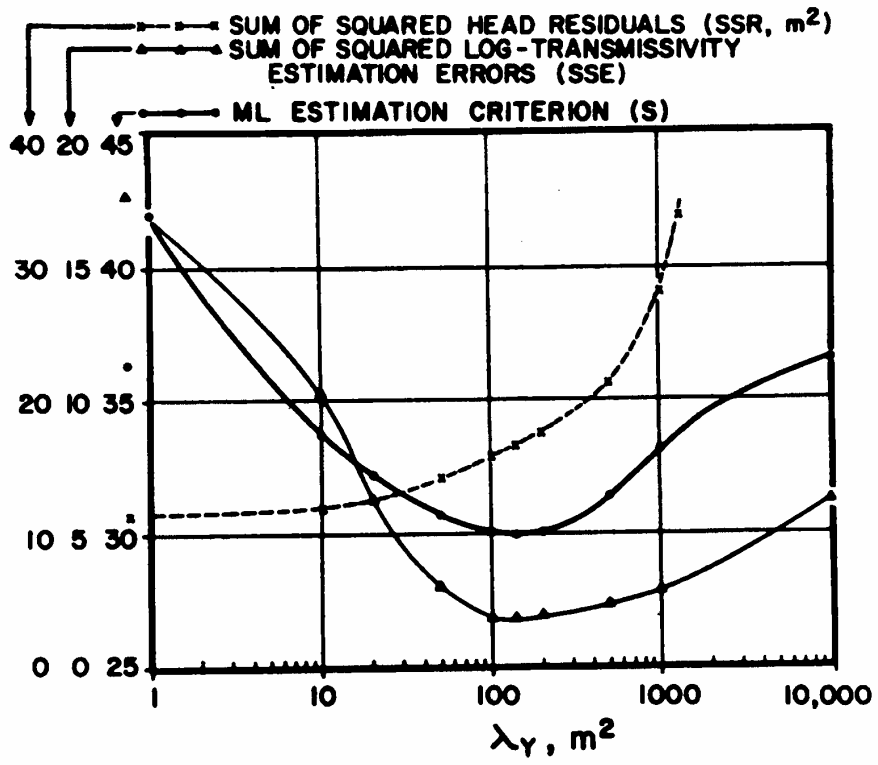


Figure 7-1. Variation of log-likelihood criterion, sum of squared head residuals, and sum of squared log-transmissivity estimation errors with λ (after Carrera and Neuman, 1986b). With permission, AGU.

7.4.4 What to Do with Insufficient Data?

When either set of data is too small, clustered or otherwise unsuitable for a meaningful assessment of prior statistics, it may still be possible to come up with prior parameter estimates and weigh them subjectively relative to each other, based on their perceived reliability. Similar weights may be assigned to site monitoring data. One can then adopt a calibration criterion $J + \lambda P$ equal to a weighted sum (J) of squared residuals (differences between simulated and observed state variables), plus

the product of λ with a weighted sum (P) of squared differences between posterior and prior parameter estimates. One can then calibrate the model for various choices of λ and plot the weighted sum of squared residuals versus $1/\lambda$ to see how the latter affects the former. As shown by Neuman (1973), one would typically find that the sum of squared residuals is largest when $1/\lambda = 0$ and decreases more-or-less monotonically to an asymptote as $1/\lambda$ increases (Figure 7-2). This is clear considering that $1/\lambda = 0$ corresponds to $\lambda = \infty$, which is equivalent to giving infinite weight to the prior parameter estimates and forcing the posterior estimates to coincide with them. As $1/\lambda$ is allowed to increase, the weight placed on the prior parameters decreases and their constraining effect on the posterior estimates gradually diminishes. Eventually $1/\lambda$ becomes large enough to virtually eliminate any effect of the prior on the posterior estimates. The calibration is now unconstrained by site characterization data and relies entirely on site monitoring data.

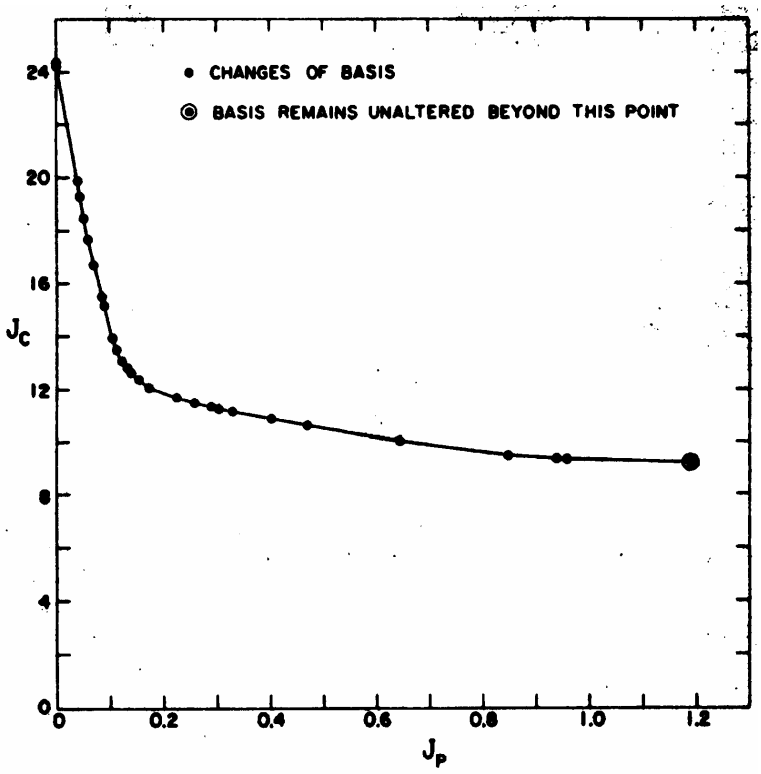


Figure 7-2. Residual versus parameter plausibility criteria (after Neuman, 1973). With permission, AGU.

Rather than one calibration criterion one now has two, the weighted sum of squared residuals (J) and the weighted sum of squared differences between posterior and prior parameter estimates (P). These two criteria are mutually incompatible in that to satisfy one, it is necessary to sacrifice the other (to achieve the objective of minimizing one, the other must be allowed to remain arbitrarily large, as seen in Figure 7-2). In most cases, the optimum solution to this multiobjective or Pareto problem is not

to minimize either objective but to find an acceptable tradeoff between them on the basis of subjective value judgment. The idea is discussed in detail by Neuman (1973).

In the absence of sufficient or reliable site monitoring data, the model must rely entirely on prior parameter estimates and is only as good as are the latter. The model remains uncalibrated, untested and unconfirmed.

7.5 How to Calibrate Stochastic Models?

7.5.1 How to Calibrate High-Resolution Monte Carlo Models?

High-resolution Monte Carlo modeling entails generating multiple random realizations of flow and transport parameters on a fine grid, solving standard deterministic flow and transport equations with these parameters on the same grid, averaging the results and analyzing them statistically. Conditional Monte Carlo simulations honor measured values of the parameters at discrete points in space. They however do not guarantee that the simulations produce results which correspond to actual records of system behavior. To render Monte Carlo simulations compatible with observational data, the approach has been to simply discard realizations that do not conform to such data. This is the approach that underlies the GLUE (Generalized Likelihood Uncertainty Estimation) strategy of Beven and Binley (1992; see also Beven, 2000) and the Bayesian approach of Gaganis and Smith (2001). The problem with it is that it usually requires a tremendously large number of simulations without ever guaranteeing full compatibility with the observations.

Sahuiquillo *et al.* (1992), Gómez-Hernández *et al.* (1997) and Capilla *et al.* (1997) have developed an inverse method that allows one to condition individual Monte Carlo realizations not only on measured hydrogeologic input parameters (notably transmissivity) but also on observed system states (notably hydraulic head). In the version described by Gómez-Hernández *et al.* (1997), first a random transmissivity field is generated conditional only on measurements of the same (via pumping tests). Next, the transmissivity field (and possibly the boundary conditions) is modified iteratively, without destroying the spatial pattern of its variability, until head simulations based upon it come close to honoring the observational data, by yielding a sufficiently small sum of weighted squared residuals. The transmissivity is parameterized geostatistically in terms of discrete values associated with actual and fictitious "master" (analogous to de Marsily's "pilot", 1978) points of measurement. Values associated with real measurements are taken to represent known transmissivities. Values associated with pilot points are treated as hydrogeologic parameters to be estimated by the inverse algorithm. Both sets of values are projected onto a computational grid via kriging to yield parameter estimates for the entire model. An option is provided whereby values associated with real measurement points can also be modified during the parameter estimation process within their respective uncertainty ranges.

To allow variations in boundary conditions, head and flux are also parameterized in terms of values at a few master locations on the boundary, which are estimated jointly with the transmissivities. In the first iteration, the flow equation is linearized to allow direct solution of the optimization problem by quadratic programming. In subsequent iterations, the equation is solved in its nonlinear form and optimization proceeds using a gradient method.

Capilla *et al.* (1997) have applied the method to two-dimensional steady state flow using a finite difference numerical scheme. In their synthetic examples, two or three iterations have sufficed to yield excellent results even for cases where natural log transmissivity variance is as high as 10.

Gómez-Hernández *et al.* (1997) stress that their aim is not to obtain "the optimal transmissivity field" but *any transmissivity field* that satisfies the conditions of honoring both transmissivity and head data while displaying a plausible pattern of spatial variability. It is clear that their optimal field is only one among many equally likely random fields that satisfy the same conditions.

In principle, one should be able to generate a large number of random fields that satisfy the above conditions by combining the proposed inverse method with the Monte Carlo approach. Though this might require considerable computer time and resources, it would most probably be much more efficient than merely discarding standard conditional Monte Carlo simulations that do not honor the observational data.

7.5.2 How to Calibrate Moment Equation Models?

Inverse algorithms based on the moment equation approach are presently under development by our group. The algorithms yield directly (without Monte Carlo simulation) optimized unbiased predictors of groundwater flow and transport variables for randomly heterogeneous hydrogeologic environments, under the action of uncertain source and boundary terms. They also yield the variance-covariance of associated estimation and prediction errors. The algorithms do so while accounting explicitly for the multiscale (*e.g.*, fractal) nature of hydrogeologic heterogeneity. They allow optimum use of field information through joint conditioning on measured values of hydraulic parameters, hydraulic heads and solute concentrations.

An inverse algorithm recently developed by us for steady state flow in saturated media (Hernandez *et al.*, 2002) is based on leading finite element approximations of exact first and second conditional moment equations. It parameterizes log permeability geostatistically in terms of discrete values associated with actual and fictitious "pilot" (de Marsily, 1978) points of measurement. Values associated with real measurements are taken to represent prior log permeability estimates. If and when enough such values are available, they are projected onto a computational grid via kriging to yield prior parameter estimates for the model. Values associated with pilot points are treated as unknown parameters to be estimated by the inverse algorithm. An option is provided whereby values associated with real measurement points can also be modified during the parameter estimation process. This is useful when they are uncertain and/or suspected of statistical bias.

Free statistical parameters may be introduced such as the variance or correlation scale of the underlying random field and/or its kriging statistics. These are then estimated jointly with the hydrogeologic parameters by using the nonlinear maximum likelihood approach of Carrera and Neuman (1986a-b). The negative log likelihood function is minimized by means of the Gauss-Newton nonlinear optimization method (Carrera *et al.*, 1997).

7.6 How to Gauge Quality of Model Fit?

To gauge the extent to which calibration has helped improve the quality of model fit, it is helpful to compare values of the calibration criterion prior to and after calibration. The first is computed with prior parameter estimates, and the second with posterior estimates. A significant improvement would manifest itself in a large relative move toward the objective (say of maximizing a likelihood function or minimizing a negative log likelihood criterion).

Other useful ways to gauge the quality of model fit, before and after calibration, include comparing contours, time records, peaks and trends exhibited by observed and simulated state variables. Systematic differences between observed and simulated behaviors are often an indication of systematic errors (biases) in the underlying model structure and/or parameters. Calibration should help eliminate systematic errors in the parameters. It follows that the persistence of such errors in a properly calibrated model may be taken as an indication of bias in model structure (provided the observational data are considered to be reliable). This may call for a reevaluation and revision or abandonment of the underlying conceptual-mathematical model.

It is also very important to perform an analysis of residuals with the aim of checking how closely they satisfy the assumptions on which the statistical inverse approach is predicated. We recall that these assumptions include multivariate Gaussian distribution, zero mean, and a variance-covariance similar to that of the observational data. Several statistical tests and graphical procedures to perform such checks are discussed by Hill (1998).

The following example illustrates how examining the quality of calibrated model fit may help one disqualify the central hydrogeologic hypothesis that underpins the model.

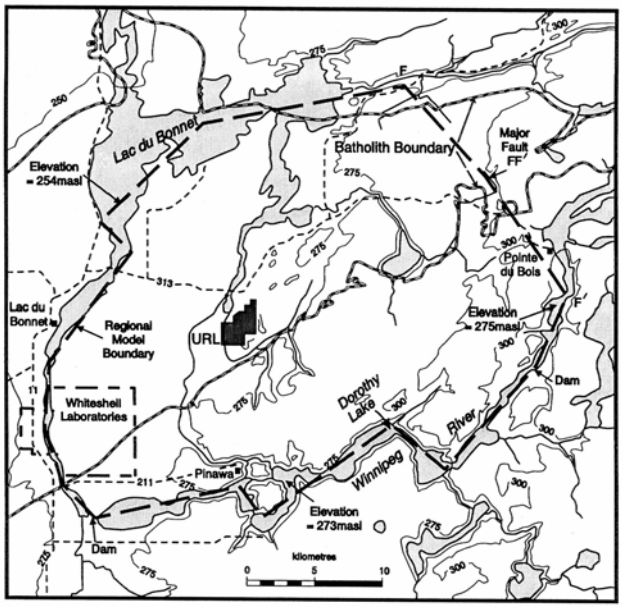


Figure 7-3. Whiteshell Research Area (after Ophori *et al.*, 1996).

7.6.1 Example: Whiteshell Research Area

The Whiteshell Research Area (WRA) near Pinawa, Manitoba (Figure 7-3), and the underground research laboratory (URL) contained within it, have been used as a generic field site to investigate the potential suitability of crystalline rocks in the Canadian shield to host a repository for spent nuclear fuel.

As part of his service on the Scientific Review Group (SRG) for the Canadian concept of geologic spent fuel disposal, Neuman (Scientific Review Group, 1996) has examined a recently revised conceptual and mathematical model that has been proposed for the area by Atomic Energy for Canada Limited (AECL). This model rests on a number of fundamental premises, which state that

1. Large intact bodies of sparsely fractured rock (SFR) are found at the site;
2. Uniformly low permeabilities and hydraulic gradients in the SFRs render transport through them predominantly diffusive;
3. There is a consistent decrease in permeability with depth in all rock units at the site, including fracture zones (FZs) of relatively high permeability;
4. Vertical and horizontal hydraulic communication between permeable rock units, including FZs, is limited;
5. Ambient flow in all rock units takes place at steady state.

The review revealed that none of these simplifying assumptions are supported, and most appear to be contradicted, by the available data. A reinterpretation of these data by Neuman has led him, and the SRG (1996), to a fundamentally different conceptual-mathematical model of site hydrogeology that is considerably more complex.

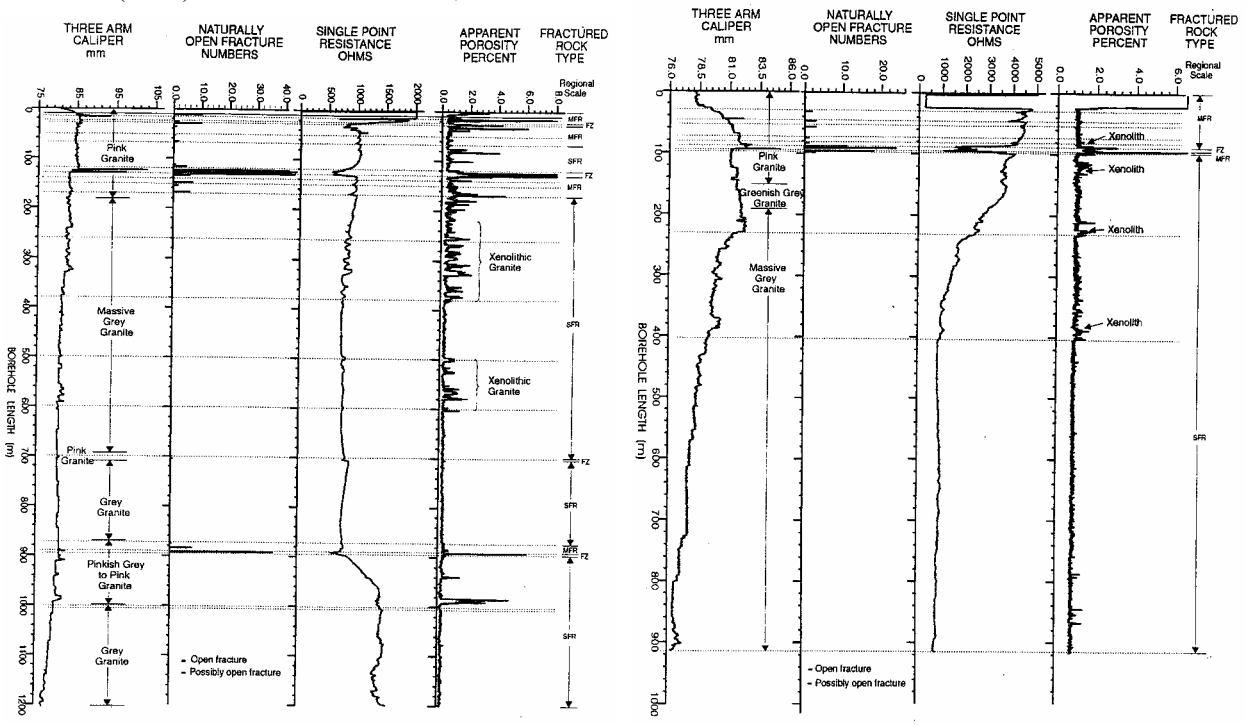


Figure 7-4. Selected logs of boreholes WD3 (left) and WG4 (right) (after Stevenson *et al.*, 1996).

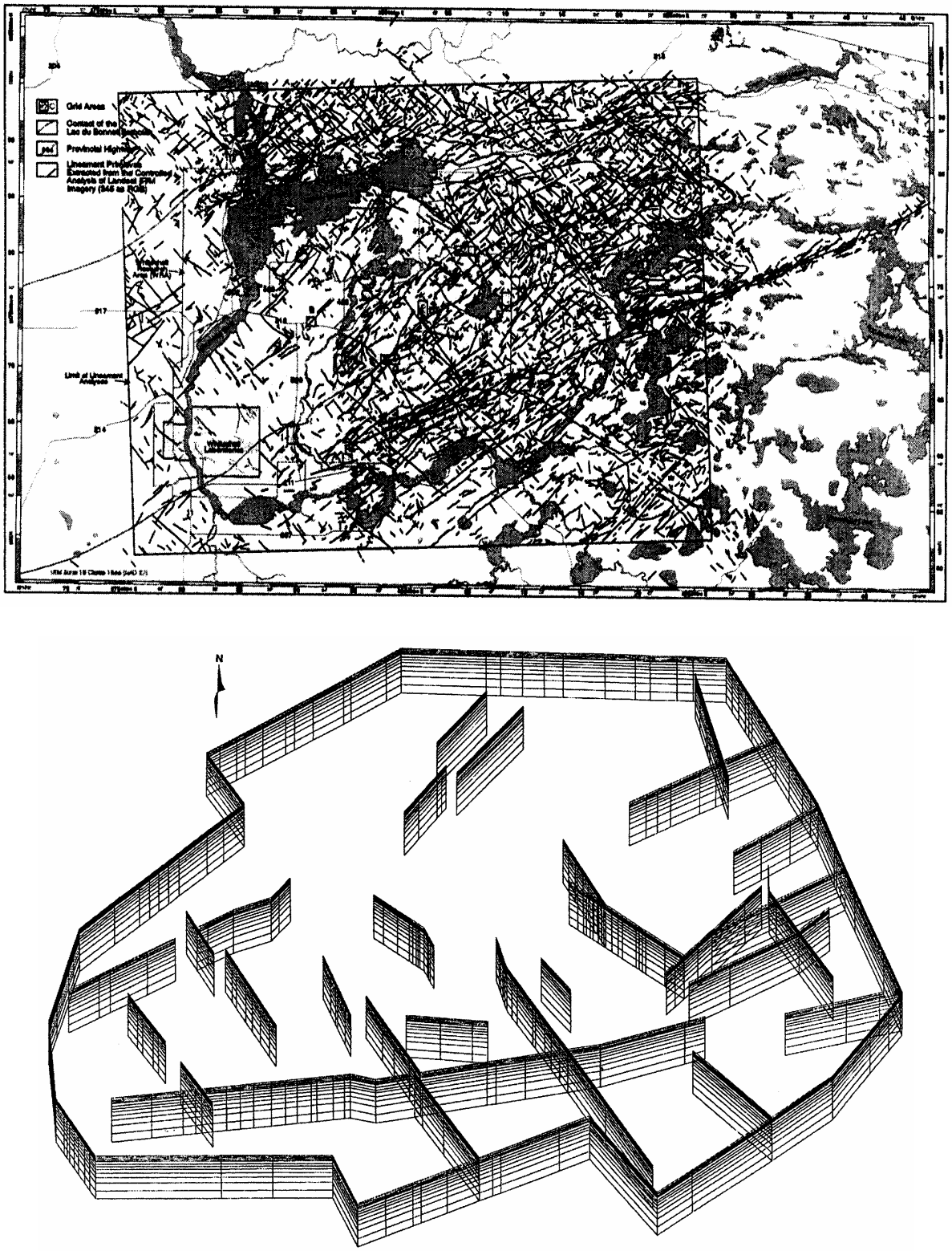


Figure 7-5. Landsat lineaments in WRA and outline of model finite element grid with embedded fracture zones (after Ophori *et al.*, 1996).

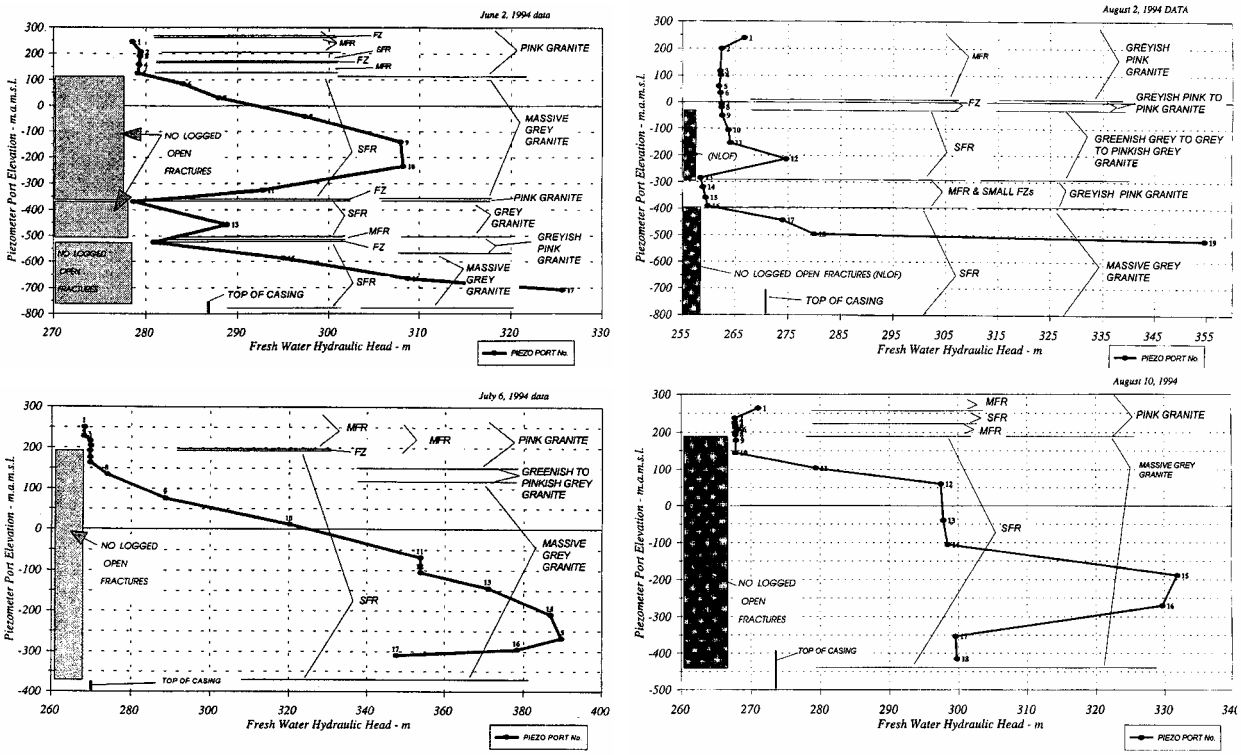


Figure 7-6. Equivalent fresh water head profiles in boreholes WD3 (upper left), WG4 (upper right), WA1 (lower left), and WG1 (lower right) (after Stevenson *et al.*, 1995, and personal communication).

The SRG review has found the available data to show and suggest the following**:**

1. Boreholes intersect intervals with up to 100+ m of SFR (Figure 7-4), which have been interpreted in the original model as thicknesses of laterally extensive SFR bodies. There is however compelling evidence (below) that numerous permeable fracture zones (FZ) and moderately fractured rock (MFR) bodies form a hydraulically interconnected network which envelopes less permeable SFR bodies on many if not all sides.
2. Whereas MFR dominates at shallow depths, it is sometimes found down to depths of 1,000 m (Figure 7-4).
3. 500+ lineaments and FZs have been identified on the surface, but relatively few have been encountered in boreholes or included in the model (Figures 7-4–7-5).
4. Only small differences have been found between equivalent fresh water head values in high-permeability zones at all depths across the region (Figure 7-6). This suggests that such zones form an interconnected regional network which allows rapid hydraulic communication and equilibration in all directions, thereby allowing flow within the network to be at a near-steady state.
5. Heads in high-permeability zones vary more in a horizontal direction (between boreholes) than in a vertical direction (within boreholes; Figure 7-6). This suggests that flow in the regional high-permeability network is predominantly horizontal.
6. Steep local hydraulic gradients prevail between high-permeability zones and SFR interiors (Figure 7-6). This suggests

that embedded within the regional high-permeability network are low-permeability SFR blocks from which water leaks into the surrounding network under steep gradients. Indeed, field salinity experiments (below) confirm such advective leakage. A plausible interpretation is that past heads were higher than those of today, possibly due to glacial loading. As the corresponding load relaxed, heads in the high-permeability network dissipated rapidly toward a new dynamic equilibrium (steady state) while heads within the low-permeability SFR blocks are still dissipating under a transient regime. Salts have been flushed from shallower parts of the network, but not yet fully from the SFR blocks.

7. The latter explains why salinity increases regionally with depth, and is locally higher within SFR blocks than within more permeable adjacent rock units.
8. Gascoyne *et al.* (1996) drilled three boreholes, up to 100 m in length, into the SFR at a depth 420 m (Figure 7-7). They packed off six sections which contained no apparent fractures, filled them with deionized water without pressurizing, and observed a continuous increase in salinity throughout the period 1992-1996. Core leaching tests and calculations indicated that ionic diffusion alone would give a salinity increase orders of magnitude lower than those actually observed. Gascoyne *et al.* therefore concluded that solute transport within the SFR, under sufficiently large hydraulic gradients, is predominantly advective rather than diffusive. As very steep hydraulic gradients between SFR interiors and adjacent high-permeability rocks have been recorded throughout the WRA (Figure 7-7), there appears to be no justification for the simplifying assumption that radionuclide transport through SFR blocks would be predominantly diffusive. This assumption is evidently nonconservative.

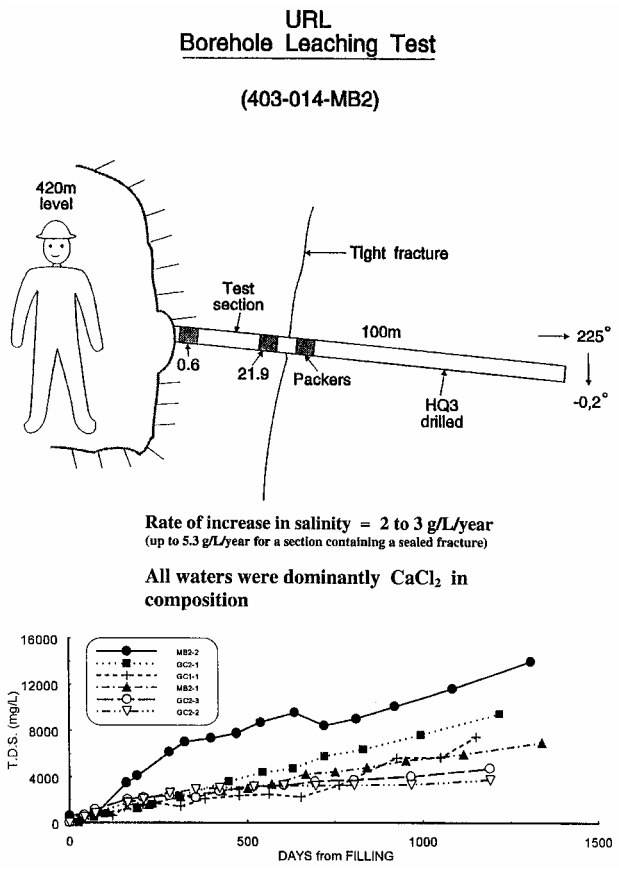


Figure 7-7. URL borehole leaching test setup and results (after Gascoyne *et al.*, 1996).

The numerical model used for performance assessment additionally assumed a downward trend in SFR permeability for which there is no evidence; disregarded the documented three-order variability in SFR permeabilities at depths below 200 m; underrated MFR permeabilities by many orders of magnitude at depths below 100 m; neglected spatial variability in FZ permeabilities; and set the latter 1-3 orders lower than is their mean at depths below 400 m. The groundwater flow and transport model was calibrated in the manner depicted in Figure 7-8. Though quantitatively the agreement between computed and observed equivalent fresh water heads was said by the modelers to be satisfactory, qualitatively the agreement is poor. This is clearly seen in Figure 7-9 where computed heads increase near-monotonically with depth while measures heads are much lower in high-permeability rock units than in low-permeability zones. The most probable reason for this mismatch is a conceptual-mathematical framework that fails to capture important complexities of hydrogeology at the URL site.

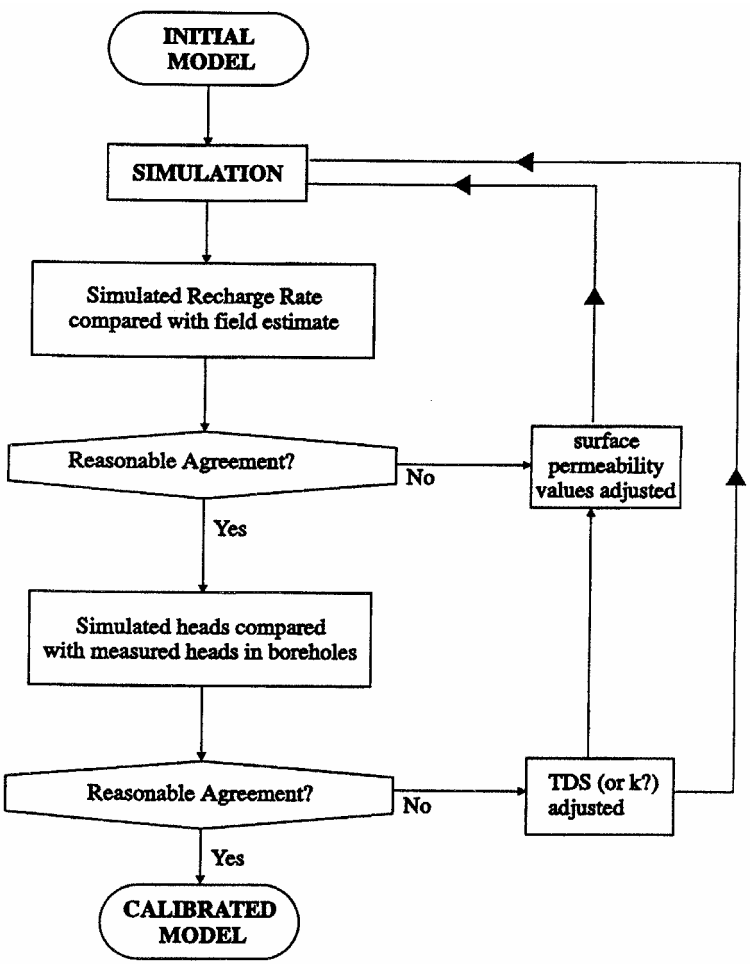


Figure 7-8. Flow chart of AECL model calibration procedure (after Ophori *et al.*, 1996).

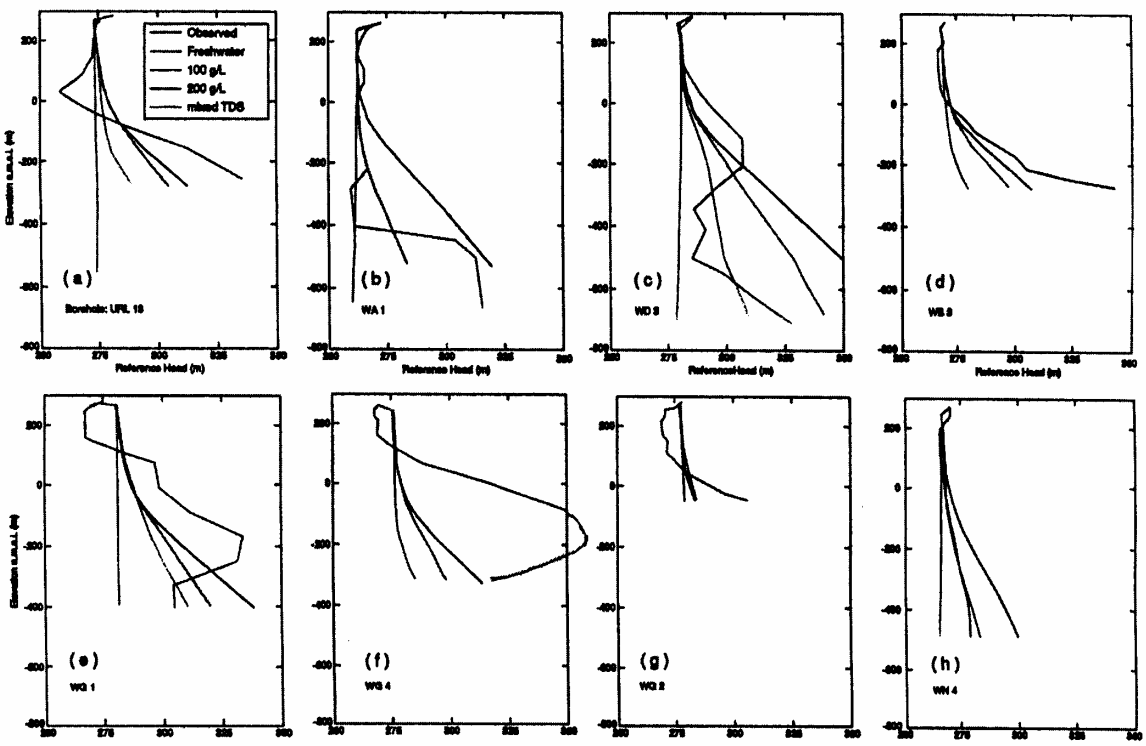


Figure 7-9. Comparison between computed and recorded equivalent fresh water head profiles (after Ophori *et al.*, 1996). Modified by Neuman (Scientific Review Group, 1996).

7.7 How to Quantify Posterior Parameter Uncertainty?

The statistical approach to deterministic model calibration allows one to quantify the uncertainty associated with posterior parameter estimates. One measure of posterior parameter estimation errors is the Cramer-Rao lower bound on their covariance matrix. This lower bound is given by the inverse Fisher information matrix (Carrera and Neuman, 1996a-b; Carrera *et al.*, 1997). Each term in the Fisher information matrix is one half the ensemble average of the second derivative of the negative log likelihood criterion with respect to a pair of parameters. It thus measures the average rate at which model sensitivity to one parameter is affected by changes in another parameter. From it, one can obtain a corresponding lower bound on the correlation matrix of the estimates. Ideally, the parameter estimates should be uncorrelated.

The Fisher information matrix is usually approximated by linearizing the relationship between system states and parameters when the latter are optimal. It is then expressed as a function of the calibration weights (inverse covariance matrices of observed and prior input data) and a Jacobian matrix, whose terms represent sensitivities of system states to the parameters. It is thus clear that an analysis of parameter uncertainty includes in it a sensitivity analysis of the model and its output.

A large variance associated with a given parameter indicates a high level of uncertainty in its estimate. A high degree of correlation between estimates implies that

they are linearly related in a statistical sense. This means that they cannot be estimated individually with the available data, and should either be lumped into one parameter or estimated anew when more data become available.

Linearization implies that if the prior (observation and parameter) errors are multivariate Gaussian (as one generally assumes), so are the posterior parameter (or log parameter) estimation errors. This allows one to associate them with linear confidence intervals, which can be viewed as corresponding error bounds.

More accurate measures of uncertainty, which compensate to some extent for nonlinearity, are discussed by Vecchia and Cooley (1987).

We illustrate these and related concepts later in the report in connection with our inverse analysis of experimental data from the Maricopa Agricultural Center. Another illustration related to inverse analysis of pneumatic pressure data from the Apache Leap Research Site is provided by Vesselinov *et al.* (2001a-b).

8 QUANTITATIVE COMPARISON AND RANKING OF ALTERNATIVE MODELS

Alternative conceptual-mathematical models that have been made compatible with the available database must now be compared among themselves to see how they rank on a relative scale of quality and complexity. Models deemed to rank low are iteratively reevaluated, revised or discarded. Those retained are ranked again on the basis of likelihood based model discrimination criteria and optionally eliminated from further consideration at this final stage of model testing and selection.

8.1 How to Establish Comparative Measures of Model Quality and Complexity?

Alternative conceptual-mathematical models that have been successfully calibrated can be compared on the basis of quality criteria such as model fit and posterior parameter uncertainty. Examples include:

- Likelihood or negative log likelihood of posterior parameter estimates, given the available data;
- Inverse variance-covariance weighted sums of squared residuals, which measure the overall lack of fit between model generated and observed system states;
- Various statistics of the residuals, which measure how closely they satisfy the assumptions of Gaussianity, unbiasedness, and similarity of their variance-covariance to that of the observational data;
- Consistent measures of systematic differences between contours, time records, peaks and trends exhibited by observed and simulated state variables;
- Invariant properties of the Cramer-Rao lower bound on the posterior covariance matrix of parameter estimation errors such as—
  - Trace, also known as A-optimality criterion (which measures the cumulative variance of all estimation errors);
  - Determinant, also known as D-optimality criterion (which provides an overall measure of the estimation error variance-covariance matrix);
  - Eigenvectors (the components of which represent relative contributions by the various parameter estimates, implying that parameters associated with a single eigenvector have uncorrelated estimation errors, and those associated with multiple eigenvectors have cross-correlated estimation errors, rendering them less amenable to discrimination); and
  - Eigenvalues (parameters associated with eigenvectors that have small eigenvalues being less uncertain than those associated with eigenvectors that have large eigenvalues); the maximum absolute eigenvalue is known as E-optimality criterion;
- Linear confidence intervals;
- Invariant properties of the Jacobian sensitivity matrix (of system states to the parameters); *etc*.

Models that have not been formally compared against monitoring data have indeterminate quality in terms of their ability to reproduce real system behavior. Calibrated models rank higher than uncalibrated models in terms of parameter quality criteria because (posterior) estimates tend to be less biased and uncertain than prior estimates, on which uncalibrated models are based.

If all models remain uncalibrated, their main comparative quality criteria are those that pertain to prior parameter uncertainty (if monitoring data are available against which model predictions can be compared, it is advisable to further calibrate the models against such data). These include

- The variance-covariance of prior parameter estimation errors;
- Various invariant properties thereof; and
- Corresponding linear confidence intervals.

All models can be compared on the basis of model complexity criteria such as

- Dimensionality;
- Number of simulated processes;
- Degree of nonlinearity;
- Number of parameters; and
- Number or size of grid cells.

8.2 How to Compare and Select Alternative Models Based on Quality and Complexity?

8.2.1 Iterative Model Testing and Selection

Alternative conceptual-mathematical models that have been made compatible with the available database must be compared among themselves to see how they rank on a relative scale of quality and complexity. Models deemed to rank low are iteratively reevaluated, revised or eliminated from further consideration.

In particular, models that do not appear to meet acceptable criteria of quality (by exhibiting excessive uncertainty or bias) or complexity (by being either too refined or too crude) are reevaluated, revised or discarded. This process of testing, comparison, revision and screening is repeated iteratively till no further improvements appear necessary or feasible. A set of models is identified which represent acceptable tradeoffs between quality and complexity. The retained models are gauged and ranked in terms of likelihood based model discrimination criteria, which account for both model fit and complexity. Models that rank low on the list can be eliminated from further consideration at this final stage of model testing and selection.

Examples of comparison and selection between models based on various criteria are included in our analysis of data from the Maricopa Agricultural Center discussed later in this report.

8.2.2 Multiobjective Approach

To help identify a set of models which represent acceptable tradeoffs between quality and complexity, it is helpful to employ a multiobjective approach.

Consider a single hydrogeologic conceptual model for a site, expressed in terms of several conceptual-mathematical models that differ in only one complexity measure, such as their number of input parameters (I). Suppose that the models have been calibrated without the benefit of prior parameter estimates and one decides to gauge their quality in terms of a single criterion, such as negative log likelihood (S). Upon plotting S versus I, one typically finds that S decreases as I increases, in a manner similar to Figure 7-2. In other words, model quality improves as the model becomes more complex, and deteriorates as the model becomes simpler. If the objective is to have a high-quality but simple model, then there is no obvious way to achieve this objective. Instead, one must decide subjectively how much quality one is willing to sacrifice for

the sake of simplicity (or vice versa), and based on this select from the graph a range of *S* and *I* values that represent an acceptable tradeoff between the two. This "feasible" range constitutes a "Pareto" solution to the dual-objective optimization (or Pareto) problem.

It is often advisable to gauge model quality in terms of more than a single criterion. For example, the calibration objectives for a transport model may include minimizing three criteria of model fit: negative log likelihood (*S*), a measure (*T*) of how well the model reproduces the observed peak arrival time of a plume at a monitoring point, and a measure (*P*) of how accurately it reproduces the magnitude of the peak. The objectives may also include minimizing two model complexity criteria: Its dimensionality (*D*) and number of parameters (*I*). The overall objective thus consists of minimizing 5 incommensurate criteria simultaneously. Not all of them, however, can be achieved without sacrificing some other objectives. The only solution is to find an acceptable tradeoff between them.

To do so, one must consider a range of models with different values of *D* (say 3, 2 and 1 corresponding to three-, two- and one-dimensional models, respectively) and *I* (say 100, 10 and 1 to represent highly heterogeneous, less heterogeneous and uniform media, respectively). One must then calibrate these models and determine their corresponding values of the model fit parameters *S*, *T* and *P*. Next, one must determine an optimum tradeoff between these various model choices.

Consider yet another case where each of several hydrogeologic conceptual models for a site is represented by a number of conceptual-mathematical models, the latter differing from each other in one or more measures of complexity (number of parameters, dimensionality, *etc.*). Upon considering one or more measures of quality for these models, one again faces a multiobjective problem that may entail many more potential choices than those obtained in the previous case.

Pareto problems with more than two objectives are difficult to solve by inspection, as we saw is possible in the dual-objective case. The alternative is to employ multiobjective programming techniques. Several methods have been developed for the articulation and solution of multiobjective decision problems which include the ELECTRE technique, compromise programming, multiattribute utility theory, surrogate worth trade-off, cooperative game theory, metagame analysis, Q-analysis, multiobjective simplex, and other (Duckstein *et al.*, 1991). The idea is to make it possible for the analyst to investigate a range of feasible solutions to the problem and solve it by determining (subjectively) an acceptable tradeoff between all of its conflicting and incommensurate objectives. A multi-objective programming method developed by Neuman and Krzystofowicz (1977) is especially well suited for this purpose. It provides a computationally efficient way to generate binary graphs of tradeoff between pairs of conflicting objectives, and to determine an optimum tradeoff between all objectives interactively, in an iterative manner that is guaranteed to converge under commonly encountered conditions.

8.2.3 Combined Measures of Model Fit and Complexity

The multiobjective approach helps one identify a set of models that represent acceptable tradeoffs between quality and complexity. To further discriminate between

these models, it is useful to gauge and rank them in terms of likelihood based model discrimination criteria, which account for both model fit and complexity. This applies only to calibrated models because uncalibrated models have unknown likelihood.

Consider several models that differ in their number of input parameters (*I*). The quality of their model fit is measured in terms of negative log likelihood (*S*). This common but special case has been studied intensively in connection with models of random time series. In particular, several criteria have been developed by Akaike (1974, 1977), Hannan (1980) and Kashyap (1982) to help one discriminate between such models. In recent years, the same criteria have been used with increasing frequency to discriminate between alternative geostatistical (Samper and Neuman, 1989a-b; Chen *et al.*, 2000) and dynamic (Carrera and Neuman, 1986a-b; Finsterle and Faybishenko, 1999) groundwater models. The model discrimination criteria are based on a combination of the model fit criterion (*S*) and the complexity criterion (*I*). By ranking the models in ascending order of their associated discrimination criteria, one favors the least complex among those that perform equally well in terms of model fit. Among models of equal complexity, those that fit the monitoring data better are favored.

The model discrimination criterion of Kashyap contains yet another term that gauges the information content of the available data. It thus allows one to consider models of growing complexity as the data base improves in quantity and quality. Stated otherwise, the criterion recognizes that when the data base is limited and/or of poor quality, one has little justification for selecting an elaborate model with numerous parameters. Instead, one should then prefer a simpler model with fewer parameters, which nevertheless reflects adequately the underlying hydrogeologic structure of the system, and the corresponding flow and transport regime. Kashyap's criterion favors that model which, among all alternatives considered, is least probable (or likely, in an average sense) to be incorrect. Stated otherwise, the criterion minimizes the average probability of selecting the wrong model among a set of alternatives.

Other model discrimination criteria can also be used, such as those employed by Hills and Wierenga (1994) in connection with INTRAVAL studies of experiments at the Las Cruces Trench Site in New Mexico.

The model testing and selection process ends with the final ranking of models on the basis of likelihood based discrimination criteria, and the optional elimination of low ranking models from further consideration. We do so for alternative models of the Maricopa experimental site, as illustrated later in the report.

8.2.4 Example: Chalk River Monzonitic Block

The above quantitative method of discriminating between conceptual-mathematical models has been applied successfully by Carrera *et al.* (1990) to the inverse modeling of transient flow during pumping tests conducted in fractured crystalline rock at Chalk River National Laboratories in Canada (Figure 8-1). They used formal model discrimination criteria to rank four different zonation patterns of hydraulic parameters, labeled Models 1 - 4 in Figure 8-2, to find that Model 4 is favored and indeed agrees extremely well with measurements.

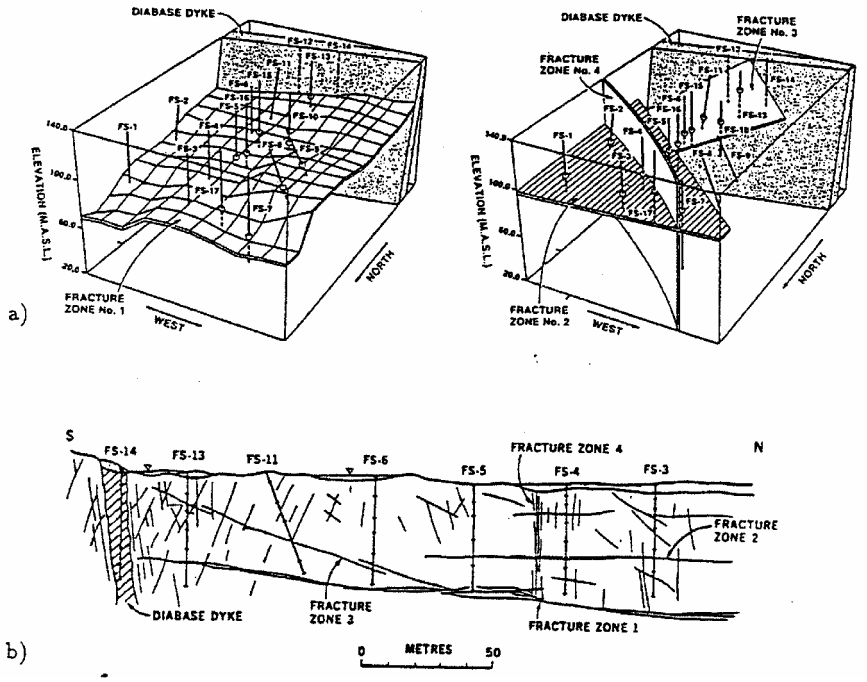


Figure 8-1. Block diagrams (a) and section (b) showing fracture zones and boreholes at Chalk River (after Carrera *et al.*, 1990). With permission, Swets & Zeitlinger Publishers.

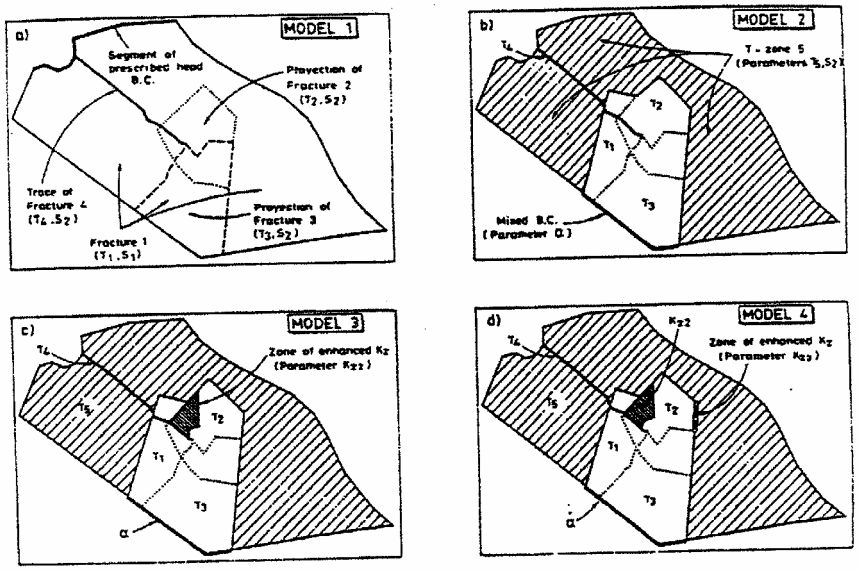


Figure 8-2. Four alternative parameter zonation patterns used in Chalk River model (after Carrera *et al.*, 1990). With permission, Swets & Zeitlinger Publishers.

8.2.5 *Example: Los Monegros Basin, Spain*

Samper-Calvete and Garcia-Vera (1999) applied similar criteria to the inverse modeling of groundwater flow in the semiarid basin of Los Monegros in Spain. They compared isotropic and anisotropic representations of two aquifers in the basin, and found that anisotropy is very important in one of them but not so much in the other.

8.2.6 *Example: Apache Leap Research Site*

Chen *et al.* (2000) compared three different conceptual-geostatistical models of spatial variability for single-hole natural log air permeability data collected on a nominal scale of 1 m at the Apache Leap Research Site in Arizona: (1) a fractal power-law variogram model, (2) an exponential variogram model with linear drift, and (3) an exponential variogram model with quadratic drift. Upon applying the maximum likelihood cross-validation method of Samper and Neuman (1989a-b) to these data, they found that whereas all three models yield comparable kriged estimates of log permeability in three dimensions, they yield very different estimation variances. This is illustrated along a vertical cross-section for models 1 and 3 in Figure 8-3. The negative log likelihood model fit criterion, NLL, in Table 8-1 is very similar for all three models, implying that all three fit the data equally well. Yet all four model discrimination criteria, AIC, MAIC, HIC and KIC (due to Kashyap), consistently rank model 1 first and model 3 last. The reason is that, even though all three models fit the data reasonably well, model 1 is the most parsimonious and model 3 the least parsimonious among the three, hence the former is the choice of preference.

Trend Model	Original Data	1st Order De-trend	2nd Order De-trend
NLL	665.801	665.080	655.849
Variogram Model of Residual	Power	Expotential	Expotential
Number of Parameter	2	6	12
Parameter c	1.4396	3.0788	2.6248
Parameter a	0.4475	1.6650	1.2602
AIC	669.801	677.08	679.849
Rank	1	2	3
MAIC	677.231	696.37	718.428
Rank	1	2	3
HIC	672.407	684.898	695.486
Rank	1	2	3
KIC	680.016	690.088	700.907
Rank	1	2	3

Table 8-1. Model fit and discrimination criteria for three geostatistical models applied to log air permeabilities from the ALRS.

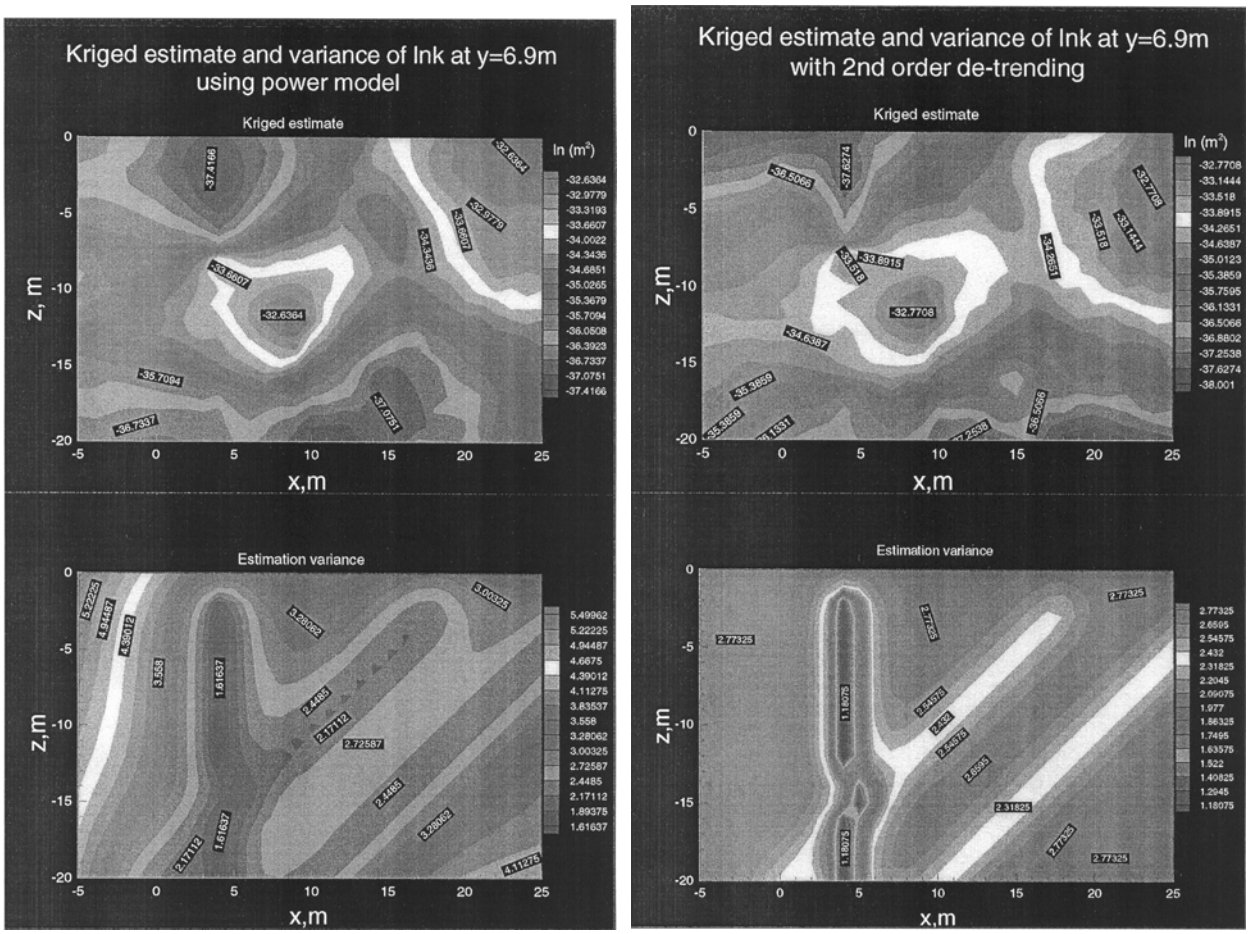


Figure 8-3. Gray scale images of kriged log air permeabilities and their variances along a vertical cross-section at the ALRS using two geostatistical models.

9 RENDERING PREDICTIONS AND ASSESSING PREDICTIVE UNCERTAINTY USING MULTIPLE MODELS

Alternative models that have been tested, compared, ranked and retained for further consideration are considered to constitute potentially valid simulators and predictors of site hydrogeology. There is no valid basis to prefer one of these models over another, and one must therefore use them in tandem. This means applying all of them to any given scenario and averaging their results to form a single "best" prediction. The prediction is best only in a relative and not in an absolute sense: It is conditional on the available models and data. This implies that better predictions could potentially be produced with other (as yet unidentified) models or data. The next task is to render a joint prediction by means of all the retained models and to identify their joint conditional predictive uncertainty.

9.1 How to Assess Predictive Uncertainty for a Single Deterministic Model?

9.1.1 Monte Carlo Method

Monte Carlo simulation is by far the most common method of assessing the predictive uncertainty of a model. The method is conceptually straight forward and has the advantage of applying to a very broad range of both linear and nonlinear flow and transport models. Given information about the statistical properties of the model input parameters (including those that represent forcing terms) or their log transformed values, one generates numerous equally likely realizations of the parameters (or log parameters, as in the case of permeability and transmissivity). If the model has been calibrated, the input statistics correspond to the posterior parameter estimates. If the model is uncalibrated, they correspond to the prior estimates. In both cases, the estimates are generally associated with a full variance-covariance matrix. Upon assuming that the estimation errors are multivariate Gaussian, one has all the information one needs to produce random realizations of the parameters.

One way to produce random realizations of a multivariate Gaussian vector (of parameters) is by Cholesky decomposition of its variance-covariance matrix. This LU-factorization method, introduced by Clifton and Neuman (1982), is well suited for full matrices and highly efficient when the latter are not too large. Its efficiency stems from the fact that only one decomposition is required for an unlimited number of realizations.

Once a parameter realization has been generated, one runs the model under a given scenario to generate a corresponding random prediction of system states. Upon averaging the predictions over all realizations, one obtains a sample mean prediction. As the number of such Monte Carlo runs increases, one hopes that the sample mean converges to a theoretical ensemble mean. While such convergence can neither be insured nor verified in most cases, it is important to at least plot a representative measure of the sample mean versus the number of runs to verify that it has reached a stable value. The stable sample mean is then taken to constitute the best prediction the model can produce.

In addition to computing the sample mean, one typically also computes the sample variance of the predictions and generates a corresponding frequency histogram. Both converge more slowly to their theoretical

ensemble equivalents than does the sample mean. It is therefore very important to verify that both have stabilized with the number of runs. Since most groundwater models are nonlinear, one need not expect the histogram of the predictions to resemble either a normal or a log normal distribution.

9.1.2 Linearization

Linearizing the flow and transport models allows one to establish approximate error bounds, or confidence limits, on model predictions on the basis of linear regression theory. Linearization expresses each observed value of a state variable as a linear combination of unknown "true" parameters plus a random error, ε. As the model is considered exact and its parameters true, the latter implies that ε represents an error of observation. For regression (calibration) purposes, ε is replaced by a corresponding residual and an optimum (posterior) estimate of the unknown parameters is obtained by minimizing a suitably weighted sum of squared residuals. This sum thus becomes a measure of the error (variance) with which observed state variables are represented by the calibrated model. The variance is generally larger than that of ε because the model is based on estimated rather than true parameters.

The weighted sum of squared residuals does not, however, measure the error with which state variables that have not yet been observed (or have not been considered for calibration) would be predicted by the model. This is so because future predictions would be done with parameter estimates that are themselves corrupted by an estimation error. Linearization expresses each predicted value of a state variable as a linear combination of unknown parameters plus a random error, ε. Each unknown parameter is the sum of a known posterior parameter estimate and an associated estimation error. Hence the error in predicted system states is the sum of two errors: one due to replacing the true parameters by their posterior (calibrated) estimates (proportional to the weighted sum of residuals), and the other due to rendering a prediction with the aid of uncertain posterior parameters (the uncertainty being characterized by the estimation variance-covariance of the parameters).

Hill (1998) explains how one can establish linearized confidence limits, or error bounds, for predictions rendered by models calibrated using nonlinear least squares with a diagonal weight matrix.

9.2 Why Consider Stochastic Models?

Appendix C explains that assessing the uncertainty of standard deterministic model parameter estimates may allow one to quantify the uncertainty in computed conditional mean system state predictors (mean head, concentration, flux *etc.*). This, however, says nothing about how actual (random and unknown) system states fluctuate about their predictors; such information is provided only by stochastic models that treat hydraulic parameters as random fields.

Appendix C concludes that traditional methods of assessing uncertainty in the output of a calibrated deterministic model, of the kind just described, fail to account for the stochastic component of prediction error. To do so, one must adopt a stochastic model.

9.3 How to Assess Predictive Uncertainty for a Single Stochastic Model?

9.3.1 High-Resolution Monte Carlo Method

High-resolution Monte Carlo simulation is by far the most common method to assess predictive uncertainty stochastically. It entails generating multiple random realizations of flow and transport parameters on a fine grid, solving standard deterministic flow and transport equations with these parameters on the same grid under scenarios of interest, averaging the results and analyzing them statistically. If based on prior parameter estimates, the Monte Carlo results honor measured values of the parameters at discrete points in space. If based additionally on the inverse method of Sahuiquillo *et al.* (1992) and Gómez-Hernández *et al.* (1997), they also honor observed values of state variables.

In all other respects, the analysis of predictive uncertainty by the stochastic Monte Carlo method is similar to that described previously in connection with deterministic models.

9.3.2 Moment Equation Method

We mentioned earlier that forward and inverse algorithms based on the moment equation approach are presently under development by our group. The algorithms yield directly (without Monte Carlo simulation) optimized unbiased predictors of groundwater flow and transport variables for randomly heterogeneous hydrogeologic environments, under the action of uncertain source and boundary terms. They also yield the variance-covariance of associated estimation and prediction errors. The algorithms do so while accounting explicitly for the multiscale (*e.g.*, fractal) nature of hydrogeologic heterogeneity. They allow optimum use of field information through joint conditioning on measured values of hydraulic parameters, hydraulic heads and solute concentrations. To see how this works for steady state flow in saturated media the reader is referred to Hernandez *et al.* (2002).

9.4 How to Render Joint Predictions with a Set of Alternative Deterministic Models?

Alternative models that have been tested, compared, ranked and retained for further consideration are considered to constitute potentially valid simulators and predictors of site hydrogeology. There is no valid basis to prefer one of these models over another, and one must therefore use them in tandem. This raises the question how to render an optimum joint prediction using multiple models, and how to assess their joint predictive uncertainty?

We propose to do so by adopting a novel Maximum Likelihood Bayesian Model Averaging (MLBMA) approach recently described by Neuman (2002). His paper is included in Appendix E. As mentioned earlier, the approach differs in fundamental ways from previous methods. It relies on probabilistic maximum likelihood (ML) concepts to (a) calibrate each model against observed space-time variations in system states (pressure, water content, concentration), considering prior information about relevant soil properties; (b) eliminate models that cannot be so calibrated with acceptable fidelity; (c) predict future system behavior or performance measures (travel times, concentrations, mass rates), and assess corresponding predictive uncertainty, using each model; and (d) average the results using posterior model probabilities as

weights. MLBMA supports the principle of parsimony in that among models that have similar predictive capabilities, it favors those having fewer parameters and being therefore simpler.

9.4.1 How to Render Joint Predictions?

Let $\mathscr{M} = (M_1,...,M_K)$ be the set of all retained models and let Δ be a quantity one wants to predict. The optimum joint prediction of Δ by means of all the models, given a discrete set of data $\boldsymbol{D}$, is given by its conditional ensemble mean (expectation) $E[\Delta|\boldsymbol{D}]$. According to MLBMA, the latter can be approximated by (Appendix E)

$$E[\Delta|\boldsymbol{D}] \approx \sum_{l=1}^{K} E\left[\Delta\middle|M_l,\hat{\boldsymbol{\theta}}_l,\boldsymbol{D}\right] p(M_l|\boldsymbol{D})$$

where $E\left[\Delta\middle|M_l,\hat{\boldsymbol{\theta}}_l,\boldsymbol{D}\right]$ is the conditional mean associated with model M_l and its maximum likelihood parameter estimates $\hat{\boldsymbol{\theta}}_l$, and $p(M_l|\boldsymbol{D})$ is the posterior (conditional) probability of this model. The term $E\left[\Delta\middle|M_l,\hat{\boldsymbol{\theta}}_l,\boldsymbol{D}\right]$ represents simply a prediction of Δ generated by model M_l after it has been calibrated against data $\boldsymbol{D}$ using the maximum likelihood method. The posterior model probability $p(M_l|\boldsymbol{D})$ can be calculated as a byproduct of this calibration using equation (6) in Appendix E. The joint prediction $E[\Delta|\boldsymbol{D}]$ is thus seen to be a weighted average of predictions obtained using individual calibrated models where the weights are the posterior (post calibration) probabilities of these models.

9.4.2 How to Assess Joint Predictive Uncertainty?

The joint posterior probability of predictions Δ is given approximately by

$$p(\Delta|\boldsymbol{D}) \approx \sum_{k=1}^{K} p\left(\Delta\middle|M_k,\hat{\theta}_k,\boldsymbol{D}\right) p(M_k|\boldsymbol{D})$$

where $p\left(\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right)$ is the posterior probability of Δ associated with model M_k after it has been calibrated against data $\boldsymbol{D}$ using the maximum likelihood method. The joint predictive probability is thus seen to be a weighted average of predictive probabilities associated with individual calibrated models where the weights are the posterior (post calibration) probabilities of these models.

The joint posterior variance of predictions is given approximately by

$$Var[\Delta|\boldsymbol{D}] \approx \sum_{l=1}^{K}\left\{Var\left[\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right] + E\left[\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right]^2\right\} \cdot p(M_l|\boldsymbol{D}) - E[\Delta|\boldsymbol{D}]^2$$

Methods to evaluate $\hat{\boldsymbol{\theta}}_k$ by calibrating a deterministic model M_k against hydrogeologic data $\boldsymbol{D}$, which may include prior information about the parameters, are described by Carrera and Neuman (1986a-b) and Carrera *et al.* (1997). The same can be done with a stochastic model based on moment equations in a manner patterned after Hernandez *et al.* (2002). These methods also yield an approximate covariance matrix for the estimation errors of $\hat{\boldsymbol{\theta}}_k$. Upon considering the parameter estimation errors of a calibrated deterministic model M_k to be Gaussian or log Gaussian, one easily determines

$p\left(\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right)$ by Monte Carlo simulation of Δ through random perturbation of the parameters. The simulation also yields corresponding approximations $E\left[\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right]$ of $E\left[\Delta\middle|M_k,\boldsymbol{D}\right]$, and $Var\left[\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right]$ of $Var\left[\Delta\middle|M_k,\boldsymbol{D}\right]$.

If M_k is a stochastic model based on moment equations, it can yield $E\left[\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right]$ and $Var\left[\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right]$ directly without Monte Carlo simulation (Hernandez *et al.* 2002). Such a model does not, however, yield $p\left(\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right)$.

9.5 How to Evaluate Hydrogeologic Performance?

Hydrogeologic performance of a site, under a given scenario, is evaluated in terms of corresponding performance measures. These in turn are computed on the basis of system states (such as travel time and concentration) that are predicted using groundwater flow and transport models. This implies that the same models and methods that one uses to predict system states, and to assess their predictive uncertainty, can also be used to predict performance measures and their uncertainty. No special methodology is needed for this purpose.

9.6 Conditional Nature of Hydrogeologic Predictions

We recall that predictions of hydrogeologic system behavior and performance are conditional on the available models and data. Many other predictions are potentially (and most probably) possible with other models and data that are equally valid and telling, perhaps more so. The uncertainty measures one associates with these predictions are likewise conditional. They therefore provide at best a lower bound on actual predictive uncertainty, which may (and most probably is) larger to an unknown degree.

10 MODEL CONFIRMATION AND ASSESSMENT OF DATA NEEDS

Once a set of predictive models have been developed and their predictive uncertainty assessed, one should try to confirm the models through comparison with observational data not previously used for calibration. This provides a suitable opportunity to evaluate the adequacy of the available database and make a decision whether or not to expand it, and how.

10.1 How to Confirm Calibrated Models?

In rare situations where enough observational data are available to allow using only some of them for model calibration, it may be possible to use the remaining data for purposes of model verification or confirmation. For this to be valid, the data used for confirmation must represent a different mode of system behavior that those used for calibration. Otherwise, the confirmation data are merely an extension of the calibration data and being able to reproduce them does not constitute a meaningful test of model predictive capability. As both the model and the data are generally uncertain, model confirmation must be cast in a suitable probabilistic and/or stochastic framework.

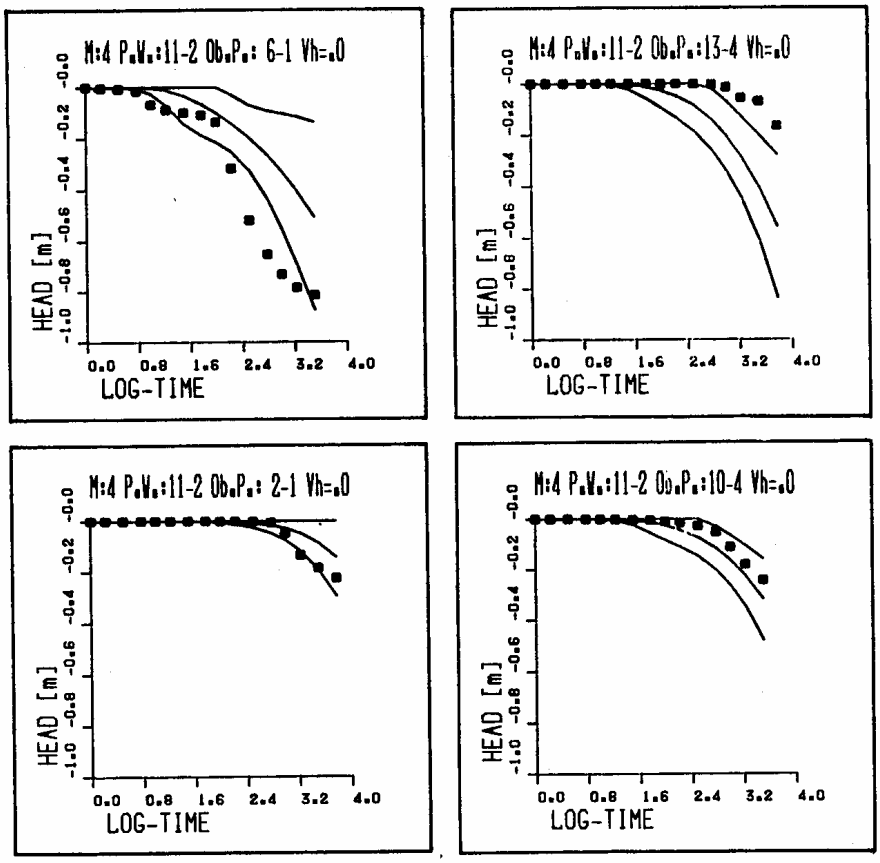


Figure 10-1. Drawdown versus log time in response to pumping from well FS-11-2. Continuous lines represent predictions and their 95% confidence intervals; dots represent measurements (after Carrera *et al.*, 1990). With permission, Swets & Zeitlinger Publishers.

10.1.1 Example: Chalk River Monzonitic Block

In section 8.2.4 we discussed briefly inverse modeling by Carrera *et al.* (1990) of transient flow during pumping tests conducted in fractured crystalline rock at Chalk River National Laboratories in Canada (Figure 8-1). The authors used formal model discrimination criteria to rank four different zonation patterns of hydraulic parameters, labeled Models 1 - 4 in Figure 8-2, to find that Model 4 is favored and indeed agrees extremely well with measurements taken while water was withdrawn from well FS-10. They then checked the extent to which the same model reproduces drawdowns during another pumping test when water is withdrawn from well FS-11-2. As shown in Figure 10-1, the predicted drawdown (dots) is within the 95% prediction confidence interval in some of the monitoring intervals but not in others.

10.2 How to Assess Additional Data Needs

10.2.1 Role of Sensitivity Analyses

To help evaluate what if any additional data might be worth collecting so as to materially reduce model uncertainty (by further constraining the range of alternative structures and parameters), one may conduct a sensitivity analysis to indicate what system behavior appears to be most sensitive to which parameters at what locations. The next step is to consider performing additional site characterization where existing parameter estimates are least certain and the model is relatively insensitive to their values, and monitoring system behavior where it is most sensitive to model parameters while prediction errors appear to be relatively large and consequential. This may indicate the type and quantity of additional site data that might materially enhance model reliability and credibility.

10.2.2 How to Decide When To Stop Collecting Data?

The question how much and what kind of data are enough for model development and evaluation is one of economics and policy, not of hydrogeologic analysis or modeling. It is therefore the responsibility of managers and decision-makers to answer, not hydrogeologists or modelers. The strategy in this report is designed to help one address the question by how much would additional site characterization and monitoring improve the prediction of performance measures derived from the hydrogeologic analysis. In other words, the strategy addresses the worth of data in terms of their contributions to the potential enhancement of model reliability and credibility, not in terms of their marginal cost-benefit. Its aim is to allow managers and decision-makers to make informed decisions about the time, manpower and budget that they deem worth allocating to these activities in light of such enhancement under existing administrative, budgetary and policy constraints.

11 CASE STUDIES

The case studies in this chapter illustrate selected components of the proposed strategy. These include the use of public, generic and site data in the conceptualization, characterization, and modeling of flow and transport, geostatistical analysis of hard and soft data, model calibration, assessment of prior and posterior parameter uncertainty, and model discrimination.

11.1 Maricopa Agricultural Center

This case study is based on the doctoral dissertation of Wang (2002).

11.1.1 Introduction

Field infiltration and tracer experiments were conducted at the Maricopa Agricultural Center (MAC) near Phoenix, Arizona (Young *et al.*, 1999; Yao *et al.*, 2003). We (Wang, 2002) have conducted a study to address the questions

- How accurately can site hydrogeology, with emphasis on the vadose zone, be conceptualized on the basis of public data and generic databases?
- How reliably can flow and transport within the vadose zone at a site be reproduced by means of simple models, and how simple can such models be?
- What methods can be used to quantify related uncertainties?
- What is the role of site data in bringing about a reduction in uncertainty?

To address these questions, we employ various methods to postulate, compare and rank alternative conceptual-mathematical models of unsaturated flow and transport during selected infiltration and tracer experiments at the MAC site. Our models include one- and two-dimensional flow and transport in a uniform soil, a soil consisting of uniform layers, and a stratified soil having laterally varying properties. Characterization of the soil as a uniform medium with a relatively deep regional water table is based on information obtained from public sources. Characterization of the soil as a layered medium with a relatively shallow water table is based on site data.

At the time of this study, only a handful of soil hydraulic properties were available for the MAC site. We use various ways to estimate these properties indirectly on the basis of soil pedologic data. We start by ascribing uniform soil hydraulic properties to each layer on the basis of soil type using mean values of three generic databases. We then ascribe variable soil hydraulic properties to individual soil samples based on soil type and bulk density using regression and neural network pedotransfer models. Upon treating these variable hydraulic property estimates as measurements, we incorporate them in the original databases to obtain Bayesian updates of their mean values and variances of the corresponding estimates. Next, we conduct a geostatistical analysis of soil pedologic and hydraulic properties which provides support for a layered conceptual model with relatively large-scale lateral variability in each layer.

In the next stage of our analysis, we use the above conceptual-mathematical models and hydraulic parameter estimates to compare simulated and observed water contents during one of the infiltration experiments at the MAC. We show that in order to reproduce observed behavior, it is necessary to further modify the hydraulic parameter estimates through inverse modeling. We

compare and rank the various conceptual-mathematical models and parameter estimates using likelihood-based model discrimination criteria and confirm our choice of best model by successfully simulating flow during an earlier infiltration experiment. Finally, we adopt one-dimensional inverse estimates of soil hydraulic parameters based on one infiltration experiment for one-dimensional solute transport modeling of bromide tracer data collected during an earlier experiment.

An advection-dispersion model with linear mass transfer to and from a zone of immobile water, or anion exclusion, is employed in forward and inverse modes.

Our study illustrates how a combination of methods and data sets can be used sequentially and in tandem to improve one's understanding of unsaturated flow and transport conditions at a site.

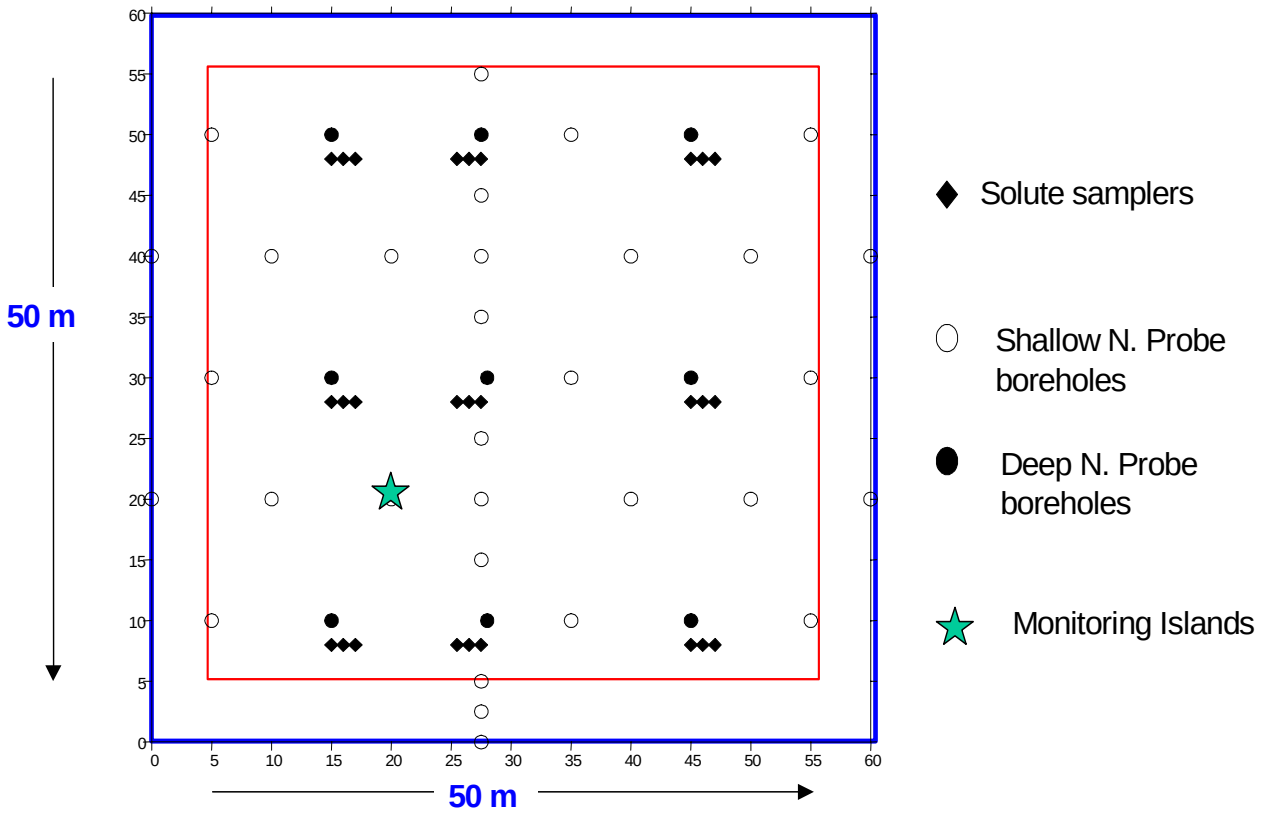


Figure 11-1. Location of monitoring equipment (Scale in meters).

11.1.2 Brief Description of Experiments 1–3

Three infiltration and tracer experiments were conducted at the MAC by Young *et al.* (1999). During these experiments, water was uniformly applied to a 50×50 m^2 area using a drip irrigation system at a controlled rate. The area was covered by a 60×60 m^2 thick Hypalon® pond liner to minimize evaporation. Monitoring took place along a trench, within two islands, and in boreholes (Figure 11-1). The backfilled trench was 65 m long, 1.5 m deep and contained various instrument clusters at 10 m intervals. The

islands consisted of highway culverts, 1.5 m in diameter, placed vertically in the soil down to 3 m depth. They were instrumented to measure soil water tension, soil water content and tracer concentration. Borehole monitoring equipment consisted of neutron probe access tubes, tensiometers and solution samplers. Nine neutron probe boreholes were drilled to 15 m depth and several others to 3 m depth. Solution samplers near the nine deep boreholes were installed at depths of 3, 5, and 10 m. Solution samplers in the monitoring islands were installed at depths from 0.5 to 3.0 m at 0.5 m intervals.

We focus on the first and last of the three experiments. Experiment 1 lasted 93 days starting April 28 and ending July 30, 1997. Water was applied at an average rate of 1.85 cm/day to the field for 24 days, with a bromide tracer added for the first 15 days at a mean concentration of 31.6 ppm. The water application period was followed by a redistribution period of 69 days. Experiment 3 lasted more than 200 days; in this study, we use data from the first 56 days starting April 24 and ending June 19, 2001. Water was applied at an average rate of 2.66 cm/day for 28 days and redistribution measured for the following 28 days. In experiment 3, monitoring was limited to the nine deep boreholes, yet data were collected more frequently, using fewer field monitoring devices (neutron probes and deep tensiometers).

11.1.3 Summary of Relevant Public Information

To develop a conceptual hydrogeologic model for the MAC, we started by collecting public data on geology, soils, aquifers, driller's logs, well logs, climate and weather, evaporation and evapotranspiration as well as related information from monographs and research papers. These data were obtained from various sources including United States Geological Survey (USGS), Arizona Department of Water Resources (ADWR), Arizona Department of Agriculture, Arizona Land Department, Arizona Land Resource Information System (ALRIS), and Arizona Meteorological Network (AZMET). They led to the following conceptual model of the experimental site.

The site lies within the upper fill of the Maricopa-Stanfield alluvial basin (Andersen *et al.*, 1992). The upper fill, which has the hydrostratigraphic designation of upper alluvial unit (UAU), consists mainly of gravel, sand and silt. It has a relatively uniform thickness of between 200 and 300 feet, becoming thinner as one nears mountain fronts. The unit was deposited by ancestral surface rivers during the final stages of basin development. It comprises alluvial channel, terrace, floodplain and alluvial fan deposits.

Historically, the unconfined UAU has been the most productive aquifer of the basin. However, groundwater pumping has greatly reduced its saturated thickness. Currently, groundwater levels are recovering throughout the basin. The rise in water levels results from an overall reduction in groundwater withdrawals, recharge following floods, and the introduction and widespread use of Central Arizona Project (CAP) water.

The State Soil Geographic Database (STATSGO), developed by the Natural Resource Conservation Service and accessible through the ALRIS website, includes a soil map for the State of Arizona. According to this map, soils in the MAC area consist of sandy loam.

Abundant well logs provide public information about the geology of the MAC area. Figure 11-2 shows a schematic log for a well about 200 meters southwest of the MAC. Figure 11-3 depicts our attempt to correlate several well logs along a west-east section near the site. Data about well locations, pumping rates and historical water levels are available for more than 1000 wells across the basin from the ADWR. Figure 11-4 shows water table elevations within the UAU aquifer across part of the basin in November 1988 (left) and 1993 (right), respectively. For reasons mentioned earlier, water levels in the MAC area have increased by about 50 feet between 1988 and 1993.

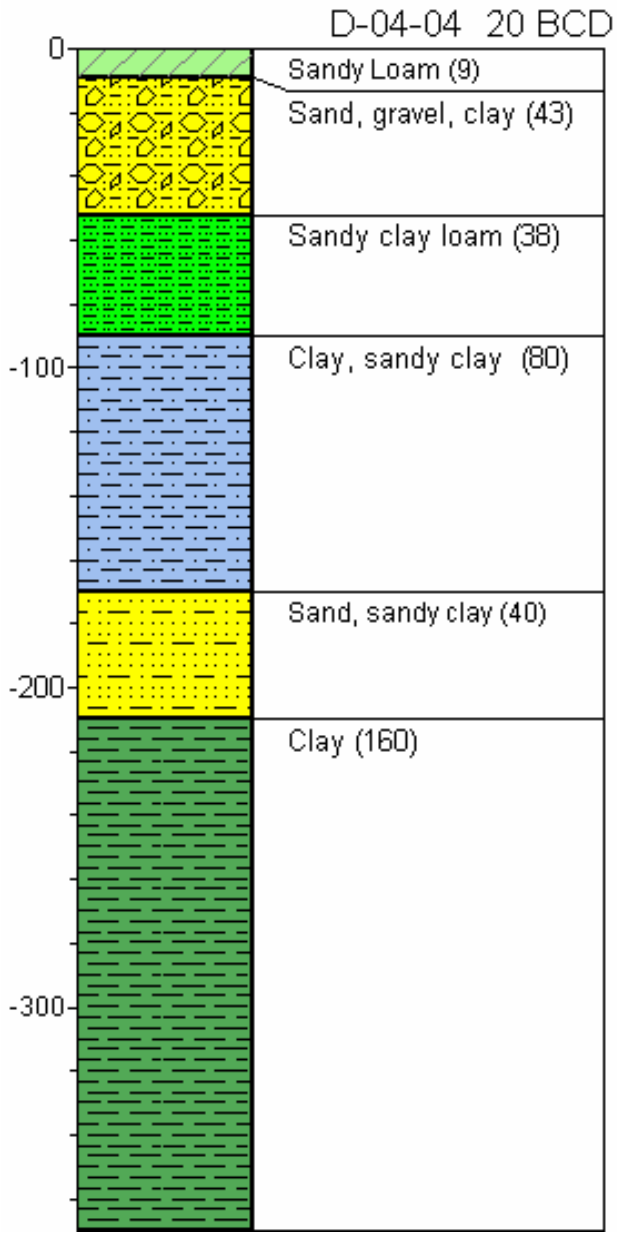


Figure 11-2. Well log of D-04-04 20 BCD near the MAC (depth and thickness in feet).

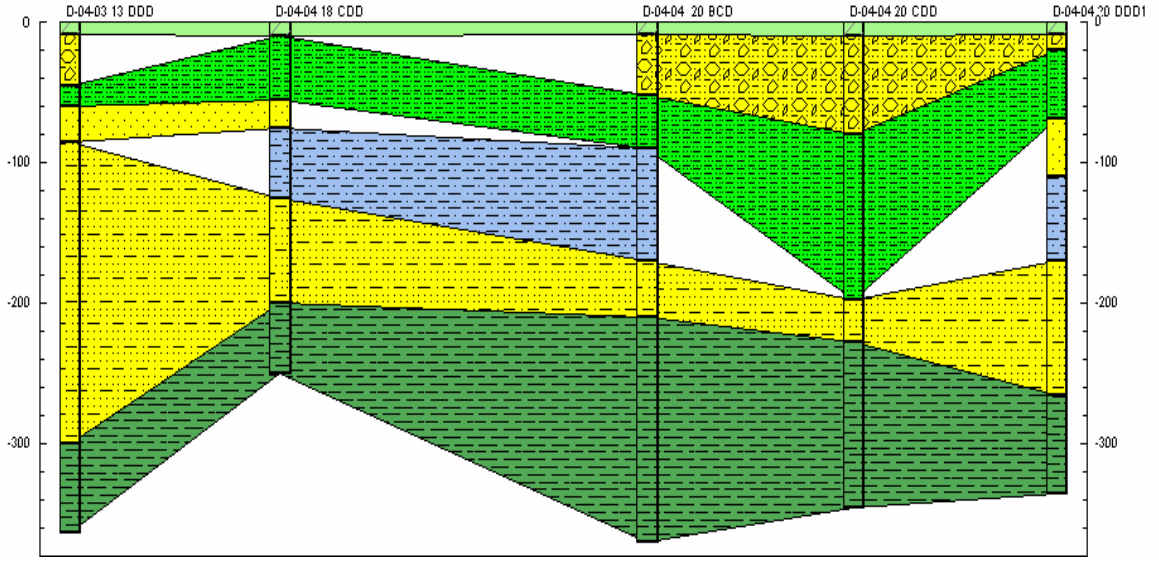


Figure 11-3. An East-West cross-section near the MAC
(depth and thickness in feet; see legend in Figure 2).

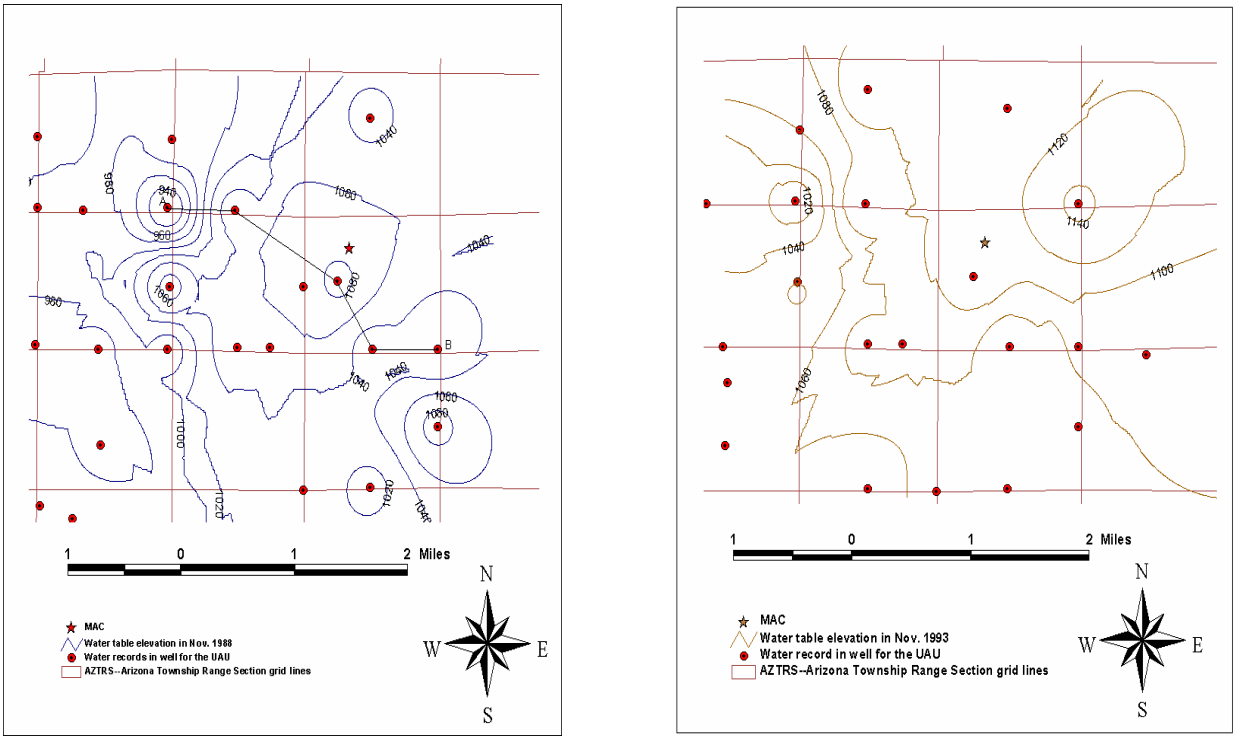


Figure 11-4. Elevation contours (in feet) of water level in the UAU aquifer
in Nov. 1988 (left) and Nov. 1993 (right).

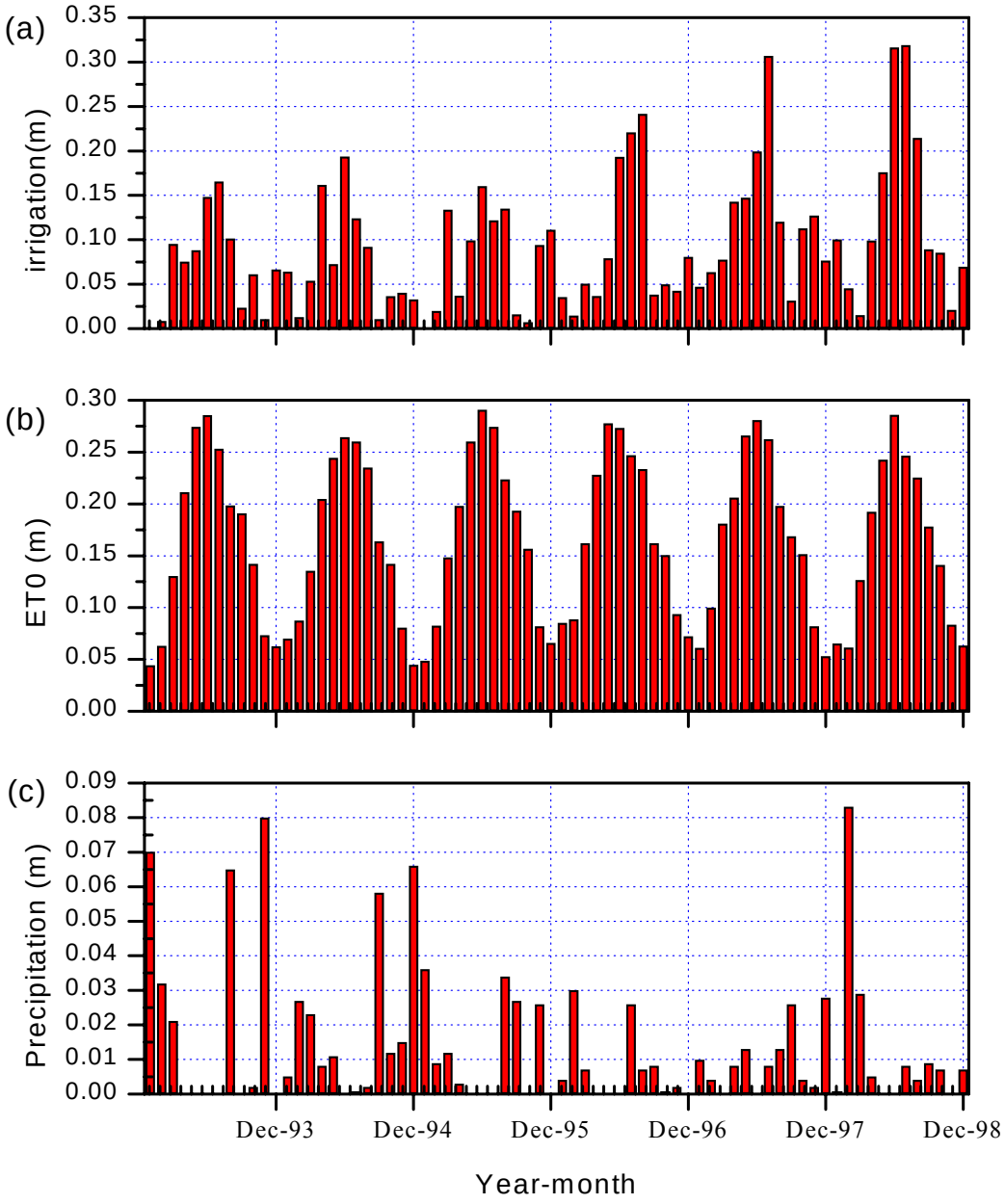


Figure 11-5. Monthly irrigation (a), reference evapotranspiration ET0 (b), and precipitation (c) in MAC area.

Meteorological data from the AZMET include air temperature, relative humidity, solar radiation, precipitation, soil temperatures and wind velocity. Information about agricultural irrigation is available from the MAC Administration and ADWR. Figure 11-5 shows how monthly irrigation, reference evapotranspiration (ET_0) and precipitation varied at the MAC during 1993-1998. ET_0 is evapotranspiration from

3–6 inches tall cool season grass that completely covers the ground and is supplied with adequate water. Average annual irrigation during 1989 - 1998 over an area of 430 acres was 1.09 m, average annual reference evapotranspiration during 1990 - 1998 was 1.98 m, and average annual precipitation during 1988 - 1998 was 0.18 m. Information about the distribution of irrigated land is available from the ALRIS website.

11.1.4 Conceptual-Mathematical Modeling Based on Public Data

The available public data suggest the following conceptual hydrogeologic model for the MAC experimental site:

- Soils at the experimental site consist of sandy loam down to a depth of about 16 m.
- The UAU aquifer is unconfined with a regional water table locally at a depth of 22 m.
- At the site, groundwater within the UAU flows from northeast to southwest under a gradient of about 0.3 %.
- Local recharge of the UAU is due to irrigation and precipitation; floods are a source of recharge elsewhere in the basin. At the covered experimental site, recharge is entirely due to applied irrigation.
- Local discharge from the UAU is due to pumping.
- Loss of water from the unsaturated zone is due to evapotranspiration, which is suppressed at the covered experimental site.

To simulate flow in the unsaturated zone, we adopt the following constitutive relationships between soil hydraulic variables (Van Genuchten, 1980)

$$\frac{\theta-\theta_r}{\theta_s-\theta_r}=\left[\frac{1}{1+\left(\alpha|h|\right)^n}\right]^m$$

$$K=K_s\frac{\left[1-\left(\alpha|h|\right)^{n-1}\left[1+\left(\alpha|h|\right)^n\right]^{1/n-1}\right]^2}{\left[1+\left(\alpha|h|\right)^n\right]^{0.5(1-1/n)}}$$

or equivalently

$$K=K_s\left(\frac{\theta-\theta_r}{\theta_s-\theta_r}\right)^{1/2}\left\{1-\left[1-\left(\frac{\theta-\theta_r}{\theta_s-\theta_r}\right)^{1/m}\right]^m\right\}^2$$

where θ is volumetric water content; θ_r and θ_s are residual and saturated water content, respectively; h is capillary pressure head; α and n are van Genuchten parameters with $m=1-1/n$; K is unsaturated hydraulic conductivity; and K_s is saturated hydraulic conductivity.

The available public information does not include soil hydraulic properties for the site. We start by estimating soil hydraulic parameters indirectly on the basis of soil class. Several generic databases have been developed for this purpose including RAWLS (Rawls et al., 1982), ROSETTA (Schaap and Leij, 1998) and CARSEL (Carsel and Parrish, 1988).

The RAWLS database includes 5401 soil samples from across the United States. Rawls et al. (1982) published a table of mean parameter estimates and their standard deviations for 11 USDA soil texture classes, based on the constitutive model of Brooks and Corey (1964). The parameters include a pore size distribution index, λ, and the air-entry pressure head, h_b. These are related to parameters of the van Genuchten (1980) constitutive model through $\lambda=n-1$ and $h_b=1/\alpha$.

The ROSETTA database is pooled from part of the AHUJA (Schaap and Leij, 1998), UNSODA (Leij *et al.*, 1996) and RAWLS databases. It contains water retention parameters for 2134 soil samples and K_s for 1306 samples. Schaap and Leij (1998) tabulate mean values of these parameters for 12 USDA soil classes.

The CARSEL database contains 15,737 soil textural samples collected by the Natural Resources Conservation Service (formerly Soil Conservation Service) from 42 of the United States. The database does not contain measured hydraulic parameters. Based on a regression model due to Rawls and Brakensiek (1985) coupled with Monte Carlo simulations, Carsel and Parrish (1988) derived from the same database probability distributions for saturated volumetric water content θ_s, residual volumetric water content θ_r, saturated hydraulic conductivity K_s, and van Genuchten's parameters α and n for twelve USDA soil textural classes. Meyer *et al.* (1997) extended their results to include probability distributions for the Brooks-Corey parameters λ and h_b, effective porosity, field capacity, wilting point, and available water content.

The distribution of samples among soil classes in RAWLS, ROSETTA and CARSEL is quite uneven. Whereas the sand portion of samples in RAWLS and ROSETTA is much larger than in CARSEL, the sandy loam portion in RAWLS is smaller and there are fewer fine-textured samples in RAWLS and ROSETTA than in CARSEL. We apply all three databases to the MAC and assess the extent to which they are supported by site data.

11.1.5 Variogram Analysis of Soil Composition at the MAC Site

Pedologic data at the MAC site include soil composition down to a depth of 15 m and bulk density down to 5 m (Young *et al.*, 1999). Soil samples were collected from the trench, monitoring islands and boreholes at depth intervals of about 0.3 m. Most samples are concentrated in the upper 1.8 m (Table 11-1) and we therefore focus our variogram analysis on this shallow depth interval. Samples below 1.8 m originate mainly from boreholes. Histograms and statistics of soil composition at the site can be found in Wang (2002).

Table 11-1. Distribution of Soil Samples with Depth.

Sampling Depth	Number of Samples
0-30 cm	70
31-60 cm	33
61-90 cm	46
91-120 cm	39
121-150 cm	88
151-180 cm	74
181-1545 cm	198

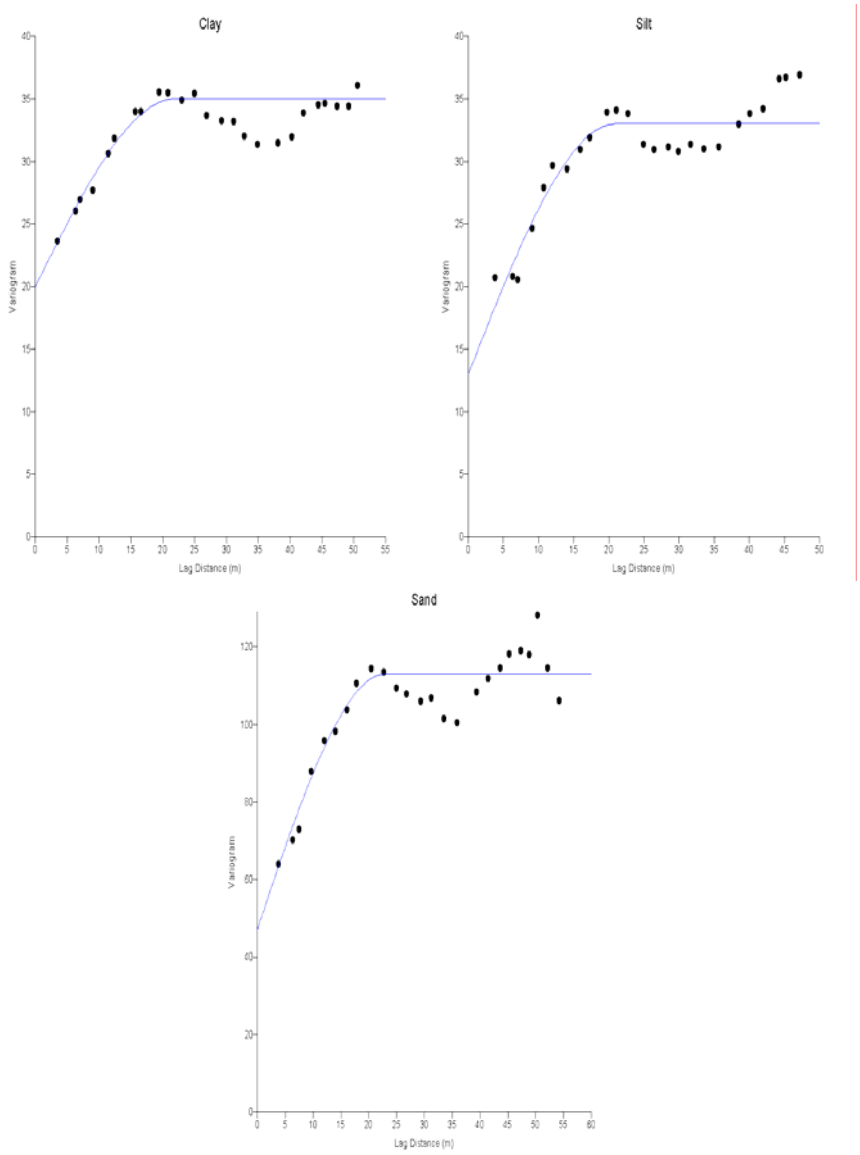


Figure 11-6. Omni-directional sample variograms and fitted spherical models for percent sand, silt, and clay at depths 0 to 30 cm.

Figure 11-6 shows omni-directional sample variograms and fitted spherical models for percent sand, silt and clay at depths 0 to 30 cm. Similar variograms for underlying 30-cm depth intervals, down to 1.8 m, can be found in Wang (2002). Though some appear to fit a linear or Gaussian model, most fit spherical models with ranges of 20 - 25 meters (Table 11-2). This implies that soil composition should be expected to vary laterally on scales exceeding 20–25 m. Figure 11-7 shows a vertical sample variograms and fitted spherical models of percent sand, silt and clay for depths 0–15 m. They reveal that the vertical correlation scale is about 2 m. In other words, soil composition should be expected to vary vertically on scales exceeding 2 m.

Table 11-2. Omni-directional variogram models fitted to soil composition data at various depths.

Depth (cm)	Component	Model	Nugget	Sill	Range (m)	Slope
0-30	Sand+gravel	Spherical	47.0	113.0	24.0	
	Silt	Spherical	13.0	33.0	21.0	
	Clay	Spherical	20.0	35.0	22.0	
31-60	Sand+gravel	Spherical	2.0	27.0	25.0	
	Silt	Spherical	3.5	23.5	20.0	
	Clay	Spherical	6.0	35.0	23.0	
61-90	Sand+gravel	Spherical	13.0	163.0	20.0	
	Silt	Spherical	5.0	56.0	21.0	
	Clay	Spherical	9.5	39.5	21.0	
91-120	Sand+gravel	Linear	40.0			1.5
	Silt	Linear	15.0			0.9
	Clay	Spherical	7.0	20.0	24.0	
121-150	Sand+gravel	Linear	40.0			0.2
	Silt	Linear	25.5			0.3
	Clay	Spherical	6.5	13.0	22.0	
151-180	Sand+gravel	Spherical	24.0	44.0	35.0	
	Silt	Gaussian	15.8	29.8	26.0	
	Clay	Spherical	5.7	13.0	33.0	

11.1.6 Postulation of Layered Structure

Our variogram analysis of soil composition supports a conceptual model whereby the soil consists of horizontal layers that are about 2 m thick and laterally heterogeneous on scales exceeding 20 – 25 m. A layered structure of similar average thickness is supported by neutron count ratios which correlate with soil compositional data as illustrated for borehole 402 by Figure 11-8; see Wang (2002) for similar correlations in other boreholes. Neutron readings were taken in 9 boreholes down to a depth of 14 meters at 0.25 m intervals. The two sets of data suggest a subdivision of the soil profile down to a depth of 20 m into 10 layers consisting of four different soil types (sandy loam, gravelly loam sand, sand and sandy clay loam). Figure 11-9 is a panel diagram showing our interpretation of site stratigraphy at the MAC based on neutron probe and soil compositional data in the 9 boreholes. In preparing this diagram, we have given more weight to the neutron than to the soil sample data. The neutron data have revealed a perched water table at a depth of about 13 m across the site.

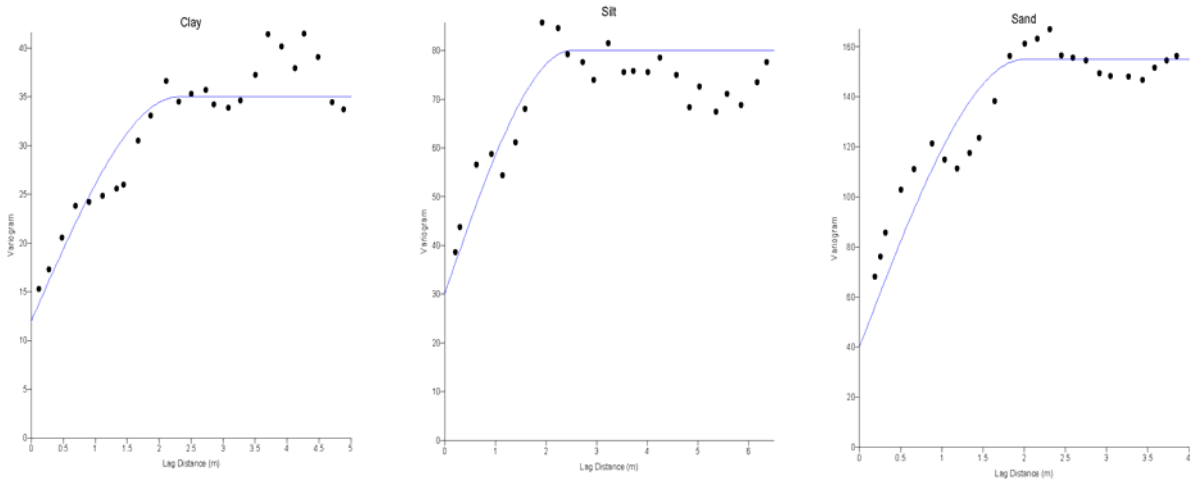


Figure 11-7. Vertical sample variograms and fitted models of percent sand, silt, and clay from the surface to 15 m depth.

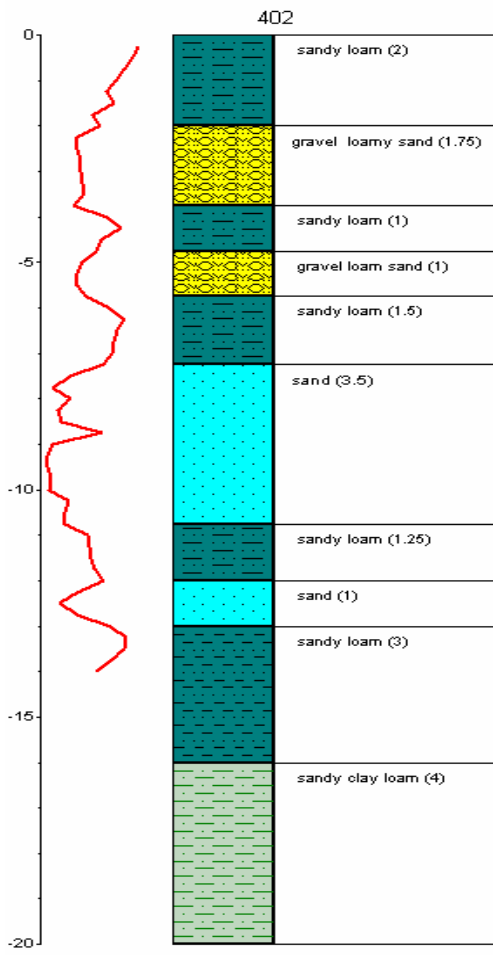


Figure 11-8. Neutron count ratio and soil class variations with depth at borehole 402.

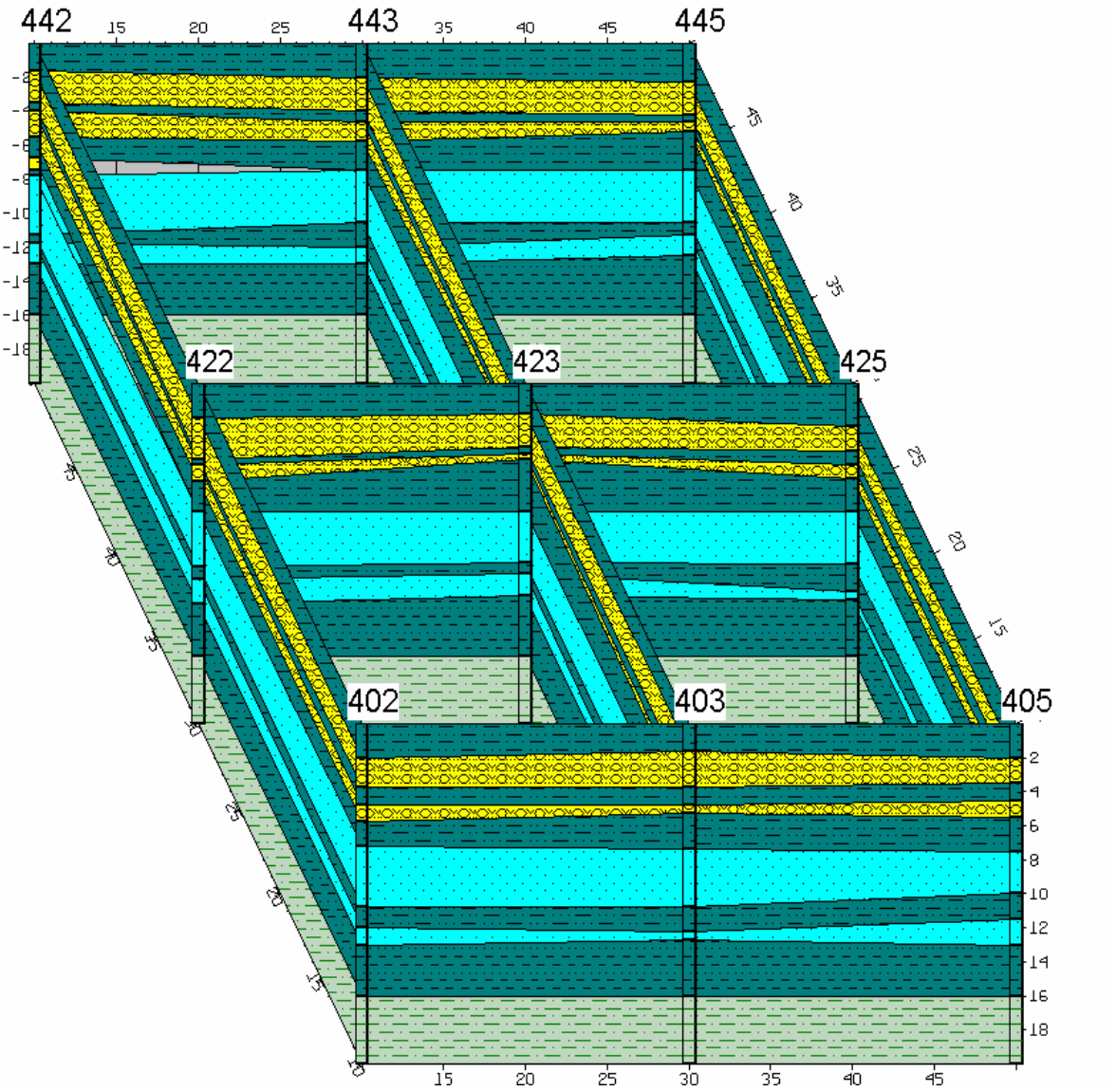


Figure 11-9. Local stratigraphy based on soil and neutron data (scale in meters; see legend in Figure 11-9).

11.1.7 Estimation of Hydraulic Parameters Using Pedotransfer Models

Various methods (called pedotransfer functions; Bouma and van Lanen, 1997) have been developed to translate soil pedologic data into hydraulic parameters. One approach is to translate soil class into mean hydraulic parameter values or corresponding probability distributions, based on generic databases, as we have done earlier. Another approach is to estimate soil hydraulic properties based on soil physical properties such as soil composition, bulk density, and organic matter content by regression or neural network models. This provides continuously varying soil hydraulic properties across the entire triangle of possible soil compositions. We use regression equations due to Rawls and

Brakensiek (1985) and a the Rosetta neural network software developed by Schaap *et al.* (1998) to estimate the hydraulic properties of individual samples across the MAC based on their particle size distribution and bulk density.

Meyer et al. (1997) provide probability distributions for the hydraulic parameters of various soil classes based on the CARSEL database. They use the Kolmogorov-Smirnov test of the null hypothesis H_o that a sample is drawn from a population having a particular probability distribution. Kenkel (1989) points out that this test is inappropriate when statistical parameters such as mean and variance are estimated from sample data. We therefore use the chi-square test to modify somewhat their probability distributions for hydraulic parameters of soil classes found at the MAC. Our approach consists of (a) postulating the null hypothesis H_o that the data are drawn from one of several theoretical probability distributions (we consider normal, lognormal and beta); (b) assigning the sample mean and variance to the selected theoretical distribution; (c) drawing random "observations" out of the sample distribution and grouping them into *K* bins; (d) calculating the Chi-square statistic

$$\chi^2 = \sum_{i=1}^{K} \frac{(o_i - e_i)^2}{e_i} \quad (4)$$

where $o_1, o_2, ..., o_K$ are the frequencies of observations in the K bins and $e_1, e_2, ..., e_K$ are corresponding theoretical frequencies; (e) minimizing χ^2 with respect to the mean and variance (we do so using the Levenberg-Marquardt algorithm); (f) repeating the process for all distributions; and (g) associating the data with that distribution which yields the smallest value of χ^2. We found in many cases that using sample mean and variance, the Kolmogorov-Smirnov test would not reject H_0 at a given level of significance α (0.05) for either the normal or lognormal distributions. However, using the above procedure we were always able to identify a single best-fit distribution among the three considered.

Figure 11-10 shows histograms and fitted distributions (solid curves) of hydraulic parameters for sand; histograms and fitted distributions for other soil components can be found in Wang (2002). The dashed curve of van Genuchten parameter *n* is adopted from the distribution of Meyer *et al.* (1997) based on a Kolmogorov-Smirnov test. Distributions based on the latter tests are associated with χ^2 values that tend to be an order of magnitude larger than those obtained by our procedure.

The regression model of Rawls-Brakensiek requires saturated water content as input. The latter is equal to porosity, which we infer from bulk density according to $\theta_s = 1 - \rho_b / \rho_s$ where ρ_b is bulk density and ρ_s is the density of the solids. Unfortunately, bulk density was determined only for shallow samples at depths of less than 5 m. Figure 11-11 shows corresponding estimates of saturated and residual water content, α, n and K_s at various depths. Saturated water content tends to decrease with depth due to a corresponding increase in bulk density.

To estimate hydraulic parameters at all depths we use the Rosetta neural network model. Results are shown in Figure 11-12. The statistics of these estimates differ from those obtained by regression (Table 11-3).

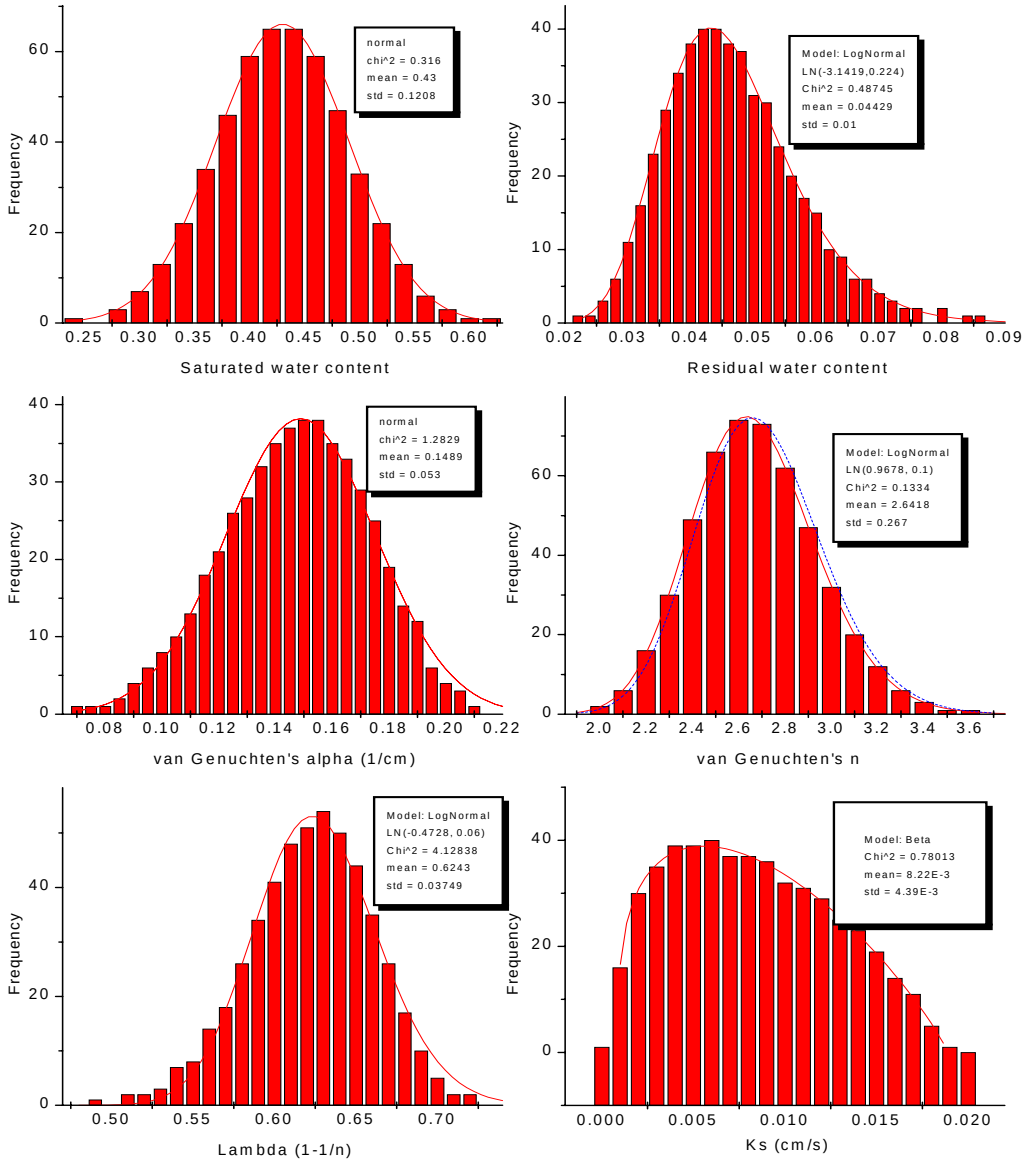


Figure 11-10. Distribution of hydraulic parameters for sand based on Meyer *et al.* (1997).

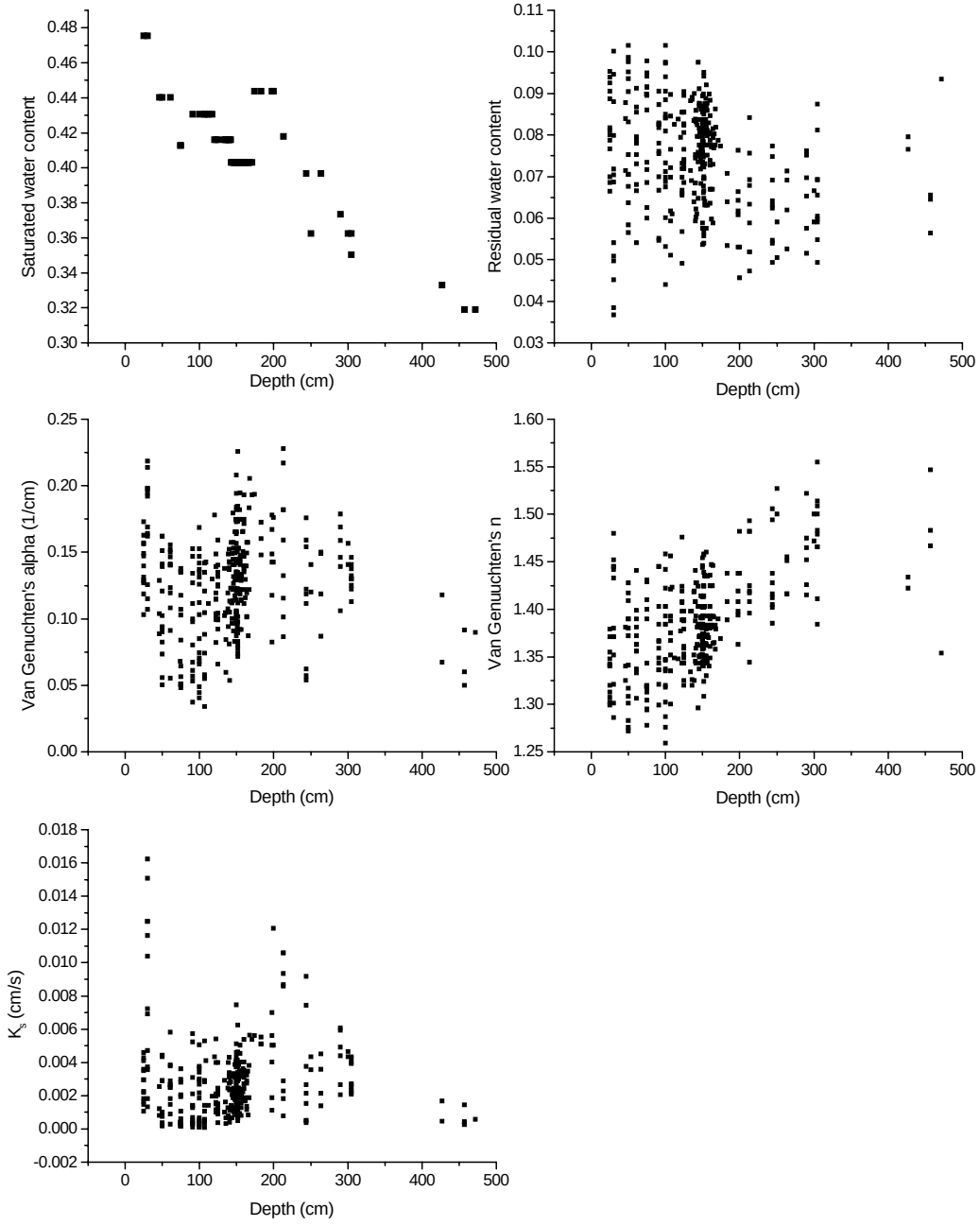


Figure 11-11. Hydraulic parameters at the MAC using Rawls-Brakensiek (1985) regressions with site-specific soil component and bulk density data at shallow depths of less than 5 m.

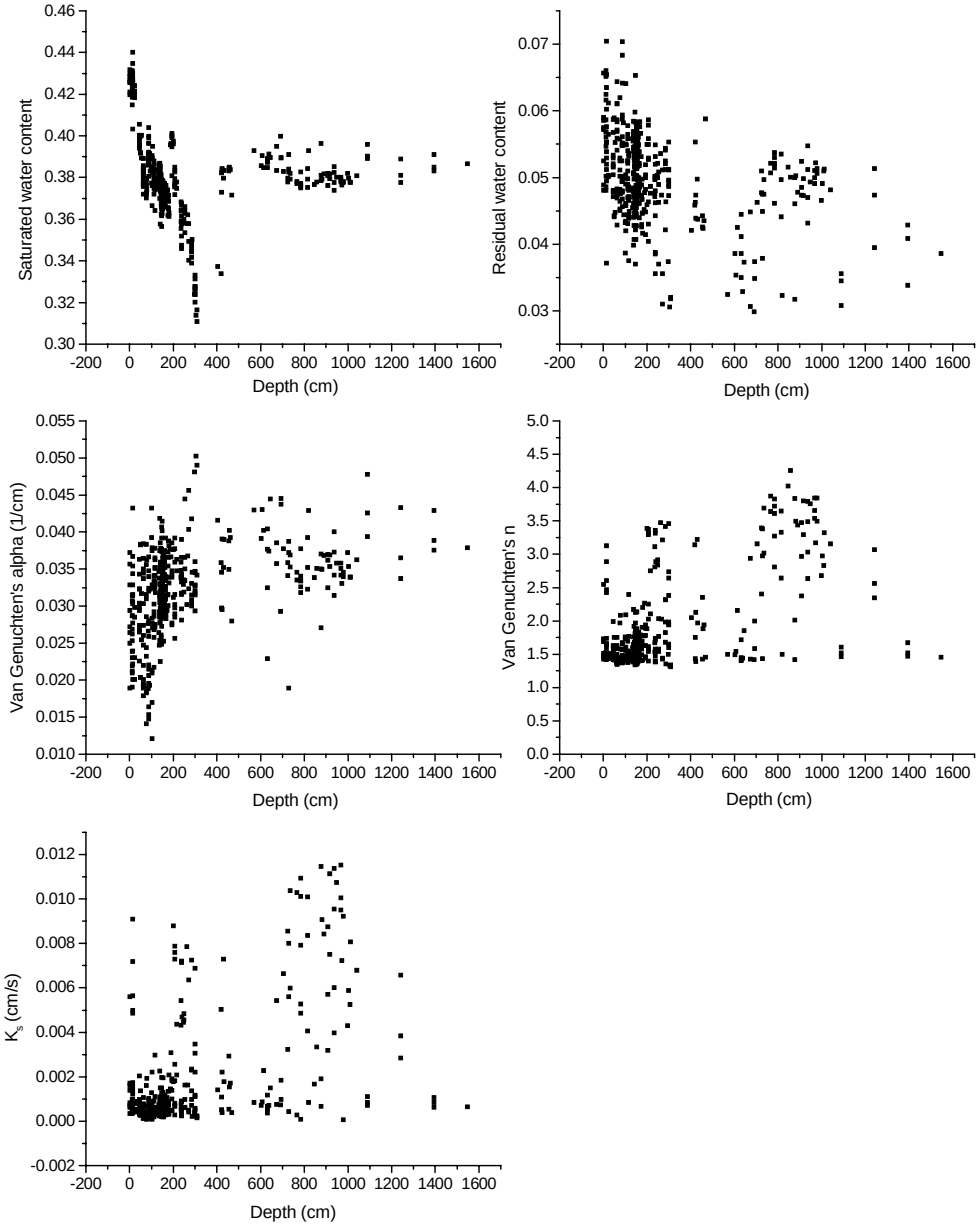


Figure 11-12. Hydraulic parameters at the MAC using Rosetta software with site-specific soil component (all the depths) and bulk density data (above 5 m).

Table 11-3. Comparison of hydraulic parameter estimates using two different pedotransfer models.

Estimates Using	Parameter	Mean	St.D.	Min	Max
Rosetta Neural Network	θ_r (cm^3 / cm^3)	4.98E-02	6.67E-03	2.99E-02	7.05E-02
	θ_r (cm^3 / cm^3)	3.80E-01	1.97E-02	3.11E-01	4.40E-01
	α (1/cm)	3.21E-02	5.81E-03	1.21E-02	5.03E-02
	n	1.84E+00	6.58E-01	1.31E+00	4.20E+00
	K_s (cm/s)	1.69E-03	2.49E-03	0.00E+00	1.15E-02
Rawls-Brakensiek Regression	θ_r (cm^3 / cm^3)	7.47E-02	1.25E-02	3.67E-02	1.02E-01
	θ_r (cm^3 / cm^3)	4.15E-01	2.78E-02	3.19E-01	4.75E-01
	α (1/cm)	1.26E-01	3.83E-02	3.42E-02	2.73E-01
	n	1.39E+00	5.07E-02	1.26E+00	1.56E+00
	K_s (cm/s)	2.79E-03	2.32E-03	9.68E-05	1.62E-02

We note in particular that whereas van Genuchten parameters *n* and α estimated by Rosetta differ from corresponding estimates using Rawls-Brakensiek regression, most other parameter estimates are comparable. The mean of Rosetta *n* estimates is much larger that of Rawls-Brakensiek estimates, the opposite being true for the mean of α estimates. The mean estimates of n from Rosetta and α from regression are comparable to mean values of corresponding CARSEL distributions for sandy loam, loamy sand and sand, the three dominant soil classes at the MAC.

11.1.8 Bayesian Updating of Hydraulic Parameter Estimates

Generic probability distributions of hydraulic parameters based solely on soil class can be updated by treating our pedotransfer estimates of these parameters as "measurements." In particular, we incorporate these measurements in the prior CARSEL database to obtain posterior estimates of the population mean values and the variance of these estimates. We do so by using a slightly modified version of a Bayesian updating code developed for this purpose by Meyer *et al.* (1997).

Updated mean hydraulic parameter distributions for sand obtained using "measurements" based on the Rosetta neural network model are depicted in Figure 11-13; corresponding distributions for this and other soil components, using either Rosetta or Rawls-Brakensiek regression, can be found in Wang (2002). The Bayesian approach of Meyer *et al.* (1997) equates the variance of prior mean values with that of soil hydraulic parameters in the CARSEL database. In this sense, Figures 11-10 can be viewed as representing the prior distribution of mean values. As expected, a comparison of Figures 11-10 and 11-13 reveals that Bayesian updating has brought about a change in the mean (reduction in bias) and a drastic reduction in its estimation variance. The same happens in the case of all soil classes at the MAC, regardless of what pedotransfer model one uses to obtain the "measurements" (Table 11-4).

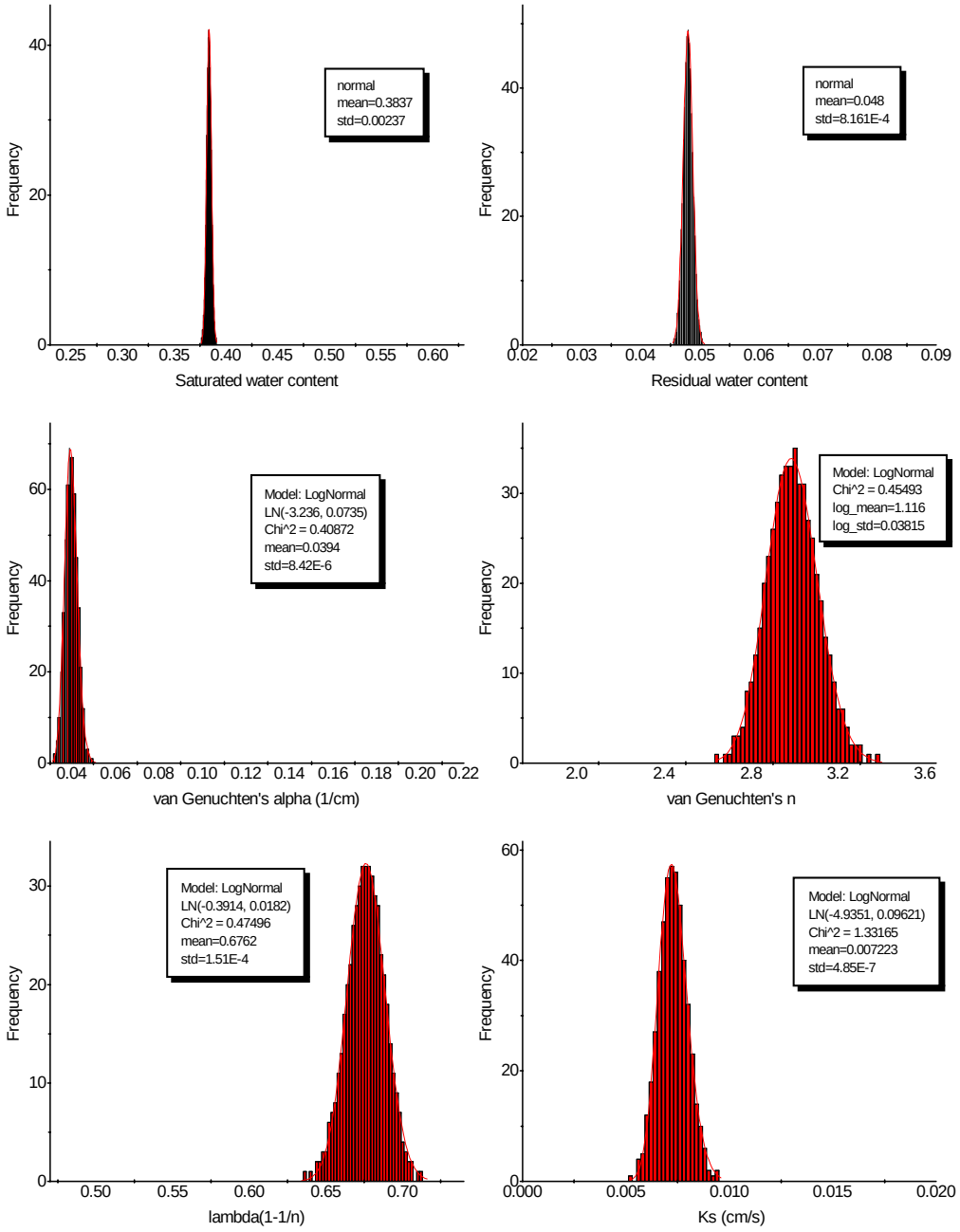


Figure 11-13. Bayesian updates of hydraulic parameters for sand using Rosetta estimates.

Table 11-4. Prior (based on CARSEL database) and posterior (following Bayesian updating) mean hydraulic parameter estimates and their standard deviations by soil class.

Soil Type	θ_s	St. D.	θ_r	St. D.	α	St. D.	n	St. D.	K_s	St. D.
unit					1/cm				cm/s	
Sandy loam										
Meyer et al	0.410	9.00E-02	0.065	1.70E-02	7.57E-02	3.68E-02	1.890	1.55E-01	1.17E-03	1.37E-03
Updated by Rawls-Brakensiek regression	0.422	1.30E-03	0.077	6.80E-04	1.24E-01	2.39E-03	1.380	1.13E-05	1.97E-03	9.30E-09
Updated by Rosetta	0.385	1.00E-03	0.052	3.51E-04	2.90E-02	1.47E-07	1.556	1.20E-04	5.61E-04	5.71E-10
Gravel Loamy Sand										
Meyer et al	0.410	9.00E-02	0.057	1.40E-02	1.26E-01	4.04E-02	2.270	2.70E-01	1.40E-04	3.17E-03
Updated by Rawls-Brakensiek regression	0.391	7.70E-03	0.062	3.00E-03	1.33E-01	1.00E-02	1.459	1.20E-04	3.45E-03	1.42E-10
Updated by Rosetta	0.365	3.24E-03	0.049	1.00E-03	4.00E-02	1.17E-05	1.882	1.15E-02	1.80E-03	9.98E-08
Sand										
Meyer et al	0.430	6.00E-02	0.047	1.00E-02	1.47E-01	2.55E-02	2.670	2.67E-01	8.22E-03	4.39E-03
Updated by Rosetta	0.384	2.37E-02	0.048	8.16E-04	3.94E-02	8.42E-06	3.055	1.35E-02	7.22E-03	4.85E-07

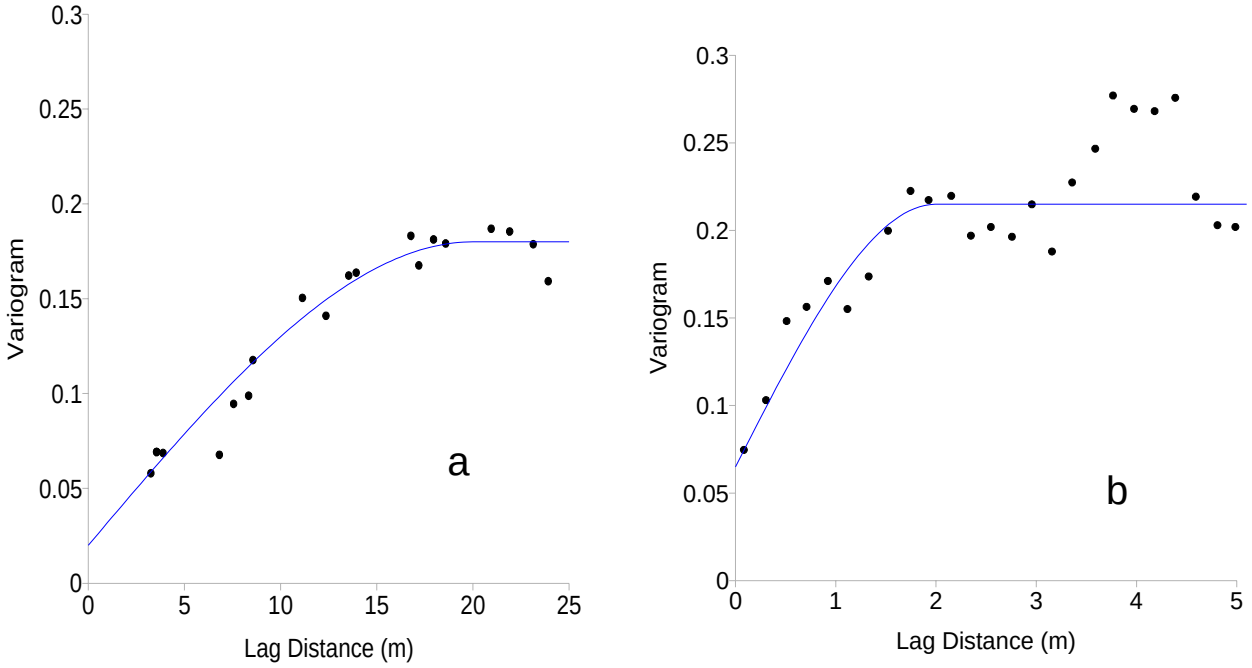


Figure 11-14. Omni-directional lateral (left, at depth 0–30 cm) and vertical (right) sample variograms and fitted models for log(K_s).

11.1.9 Variogram Analysis of Hydraulic Parameter Estimates

We conduct a variogram analysis of hydraulic parameter estimates obtained using the Rosetta neural network software. Figure 11-14a shows omni-directional sample variograms and fitted spherical models for log K_s at depths of 0–30 cm. Similar variograms are obtained down to a depth of 1.8 m for log K_s and saturated water content (θ_s) (Wang, 2002). Most of them fit spherical models and one fits a Gaussian model with horizontal range values between 20 and 36 m; some fit linear models that do not possess finite correlation scales (Table 11-5).

Vertical sample variogram and fitted spherical model of log K_s are depicted in Figure 11-14b. Similar results are obtained for saturated water content (θ_s). Both have a vertical range on the order of 1–2 m.

Our variograms of hydraulic parameter estimates closely corresponds to those of soil composition. This is not surprising considering that the former were estimated in part on the basis of the latter. Together with neutron probe data they support a conceptual model whereby the soil consists of horizontal layers that are about 2 m thick and laterally heterogeneous on scales of 20–30 m.

Table 11-5. Omni-directional variogram models fitted to hydraulic parameter estimates at various depths.

Parameter	Depth (cm)	Model	Nugget	Sill	Range (m)	Slope
θ_s	31-60	Spherical	1.70E-05	2.70E-05	25	
	61-90	Linear	3.60E-05			3.2E-07
	91-120	Spherical	0.00E+00	3.40E-05	30	
	121-150	Linear	3.84E-05			1.09E-07
	151-180	Gaussian	9.00E-06	2.30E-05	22	
K_s	0-30	Spherical	0.025	0.18	20	
	31-60	Spherical	0.009	0.119	20	
	61-90	Spherical	0.013	0.168	21	
	91-120	Linear	0.04			0.0015
	121-150	Spherical	0.029	0.055	24	
	151-180	Spherical	0.036	0.064	36	

11.1.10 Forward Flow Modeling Based on Public Data

As pointed out earlier, publicly available data suggest that soils at the experimental site consist of sandy loam down to a depth of about 16 m with a regional water table located at a depth of 22 m. In this paper we examine the extent to which such public information allows reproducing observed behavior during field infiltration experiment 3 at the MAC. We do so by postulating uniform vertical flow through the shallow subsurface, which we take to consist of uniform sandy loam. To characterize the sandy loam hydraulically, we ascribe to it mean parameter values taken from the three generic databases RAWLS (Rawls et al., 1982), ROSETTA (Schaap and Leij, 1998) and CARSEL (Carsel and Parrish, 1988; Meyer *et al.*, 1997).

Volumetric water contents are obtained from neutron probe readings using a calibration curve developed by Young (1999). The calibration is based on measured water contents from core samples and neutron probe readings at depths of 3.0 meters or less. We adopt water content measurements in borehole 422 prior to experiment 3 as initial water contents and set the water table as a constant head boundary. A constant flux of 2.66 cm/day is maintained at the soil surface during the first 28 days of the experiment. During the remainder of the experiment, the soil surface constitutes a no-flow boundary. We use the TOUGH2 code of Pruess *et al.* (1999) to simulate unsaturated flow.

Figures 11-15 – 11-18 compare measured and simulated water contents in borehole 422 using mean hydraulic parameters from ROSETTA, RAWLS and CARSEL, respectively. Only the latter yield acceptable matches with observations at depths of less than 4.0 m.

We conclude that

- Relying entirely on publicly available data leads to a poor reproduction of observed infiltration at the Maricopa site.
- Among the three generic databases examined, the best results are obtained with the CARSEL set.

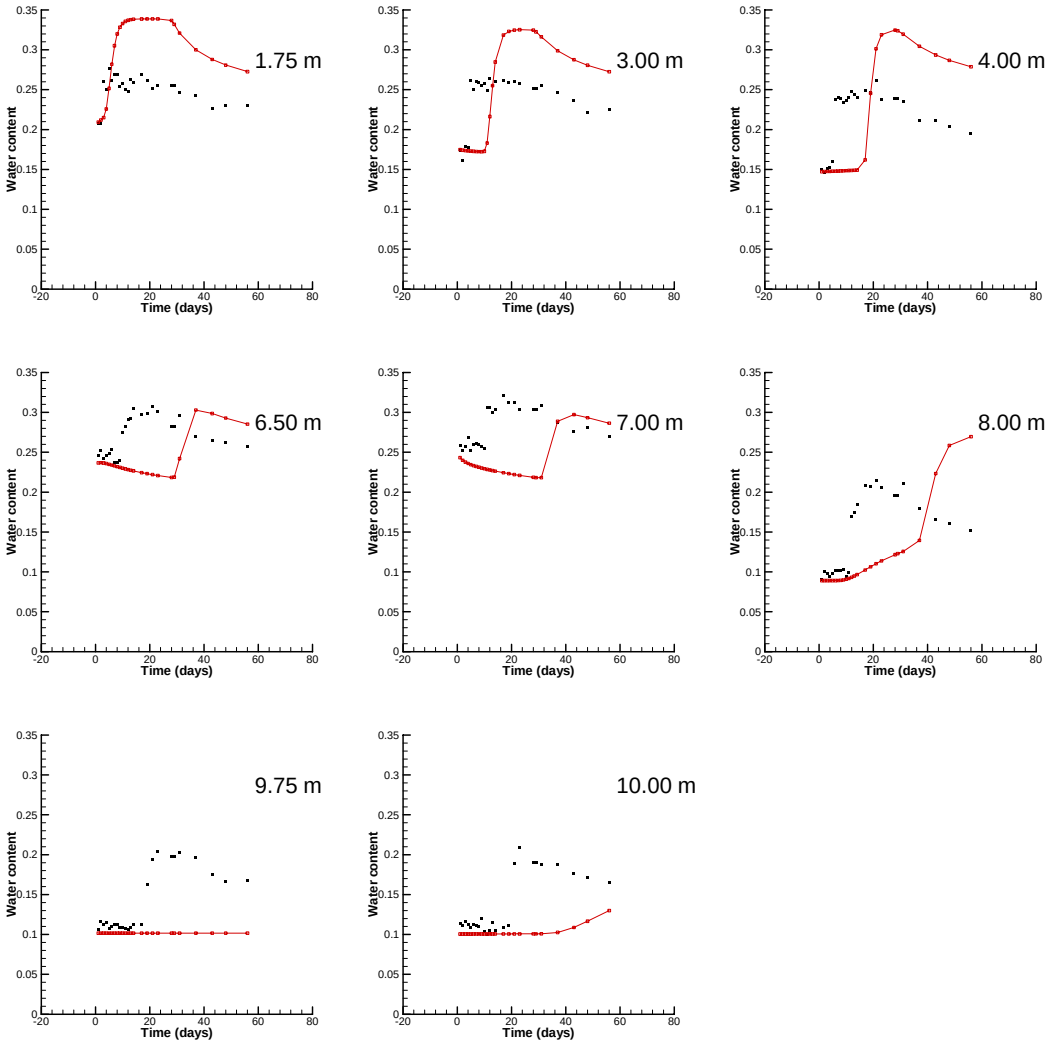


Figure 11-15. 1-D forward simulation of infiltration during experiment 3 at borehole 422 using mean hydraulic parameters from ROSETTA.

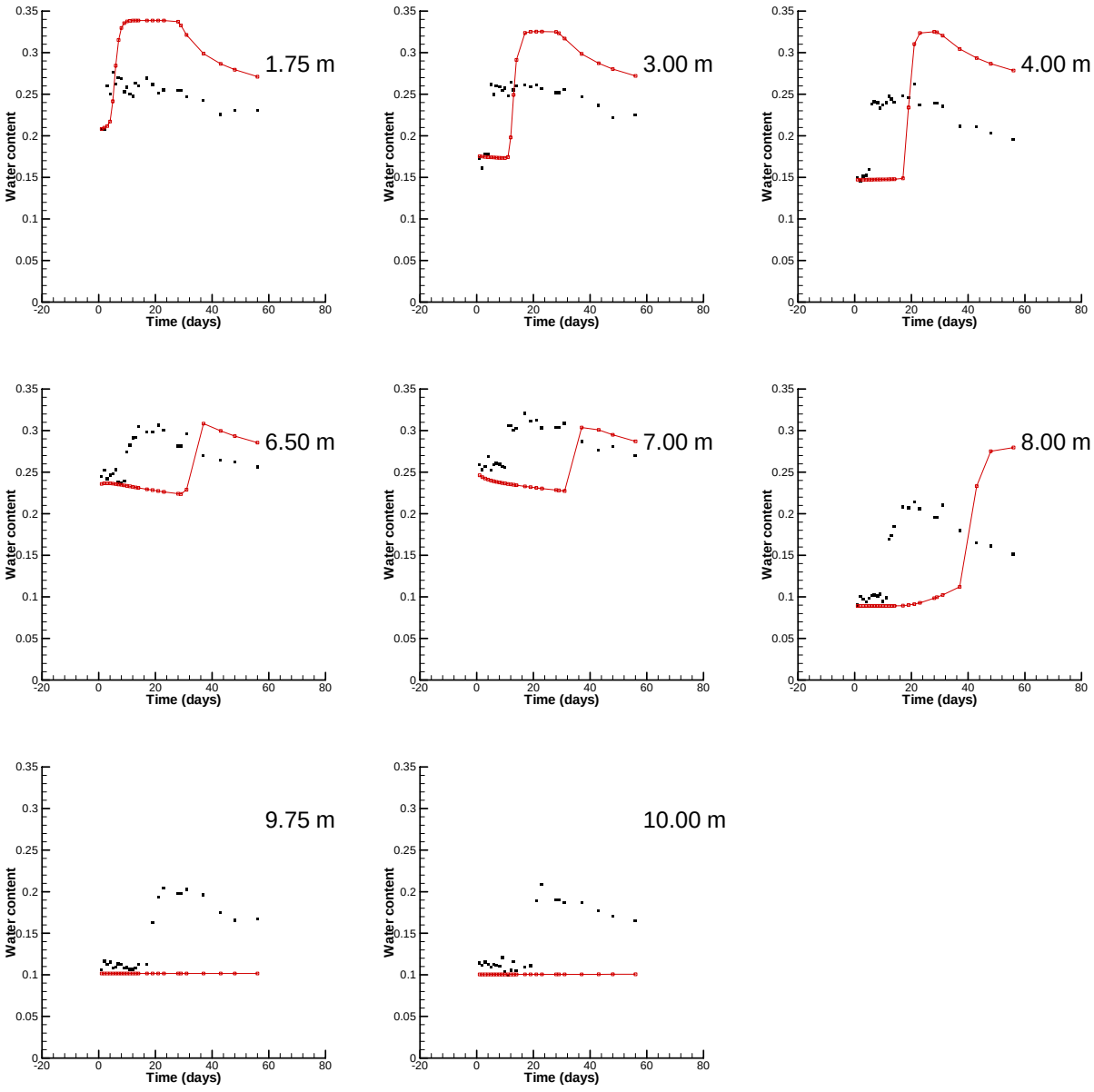


Figure 11-16. 1-D forward simulation of infiltration during experiment 3 at borehole 422 using mean hydraulic parameters from RAWLS.

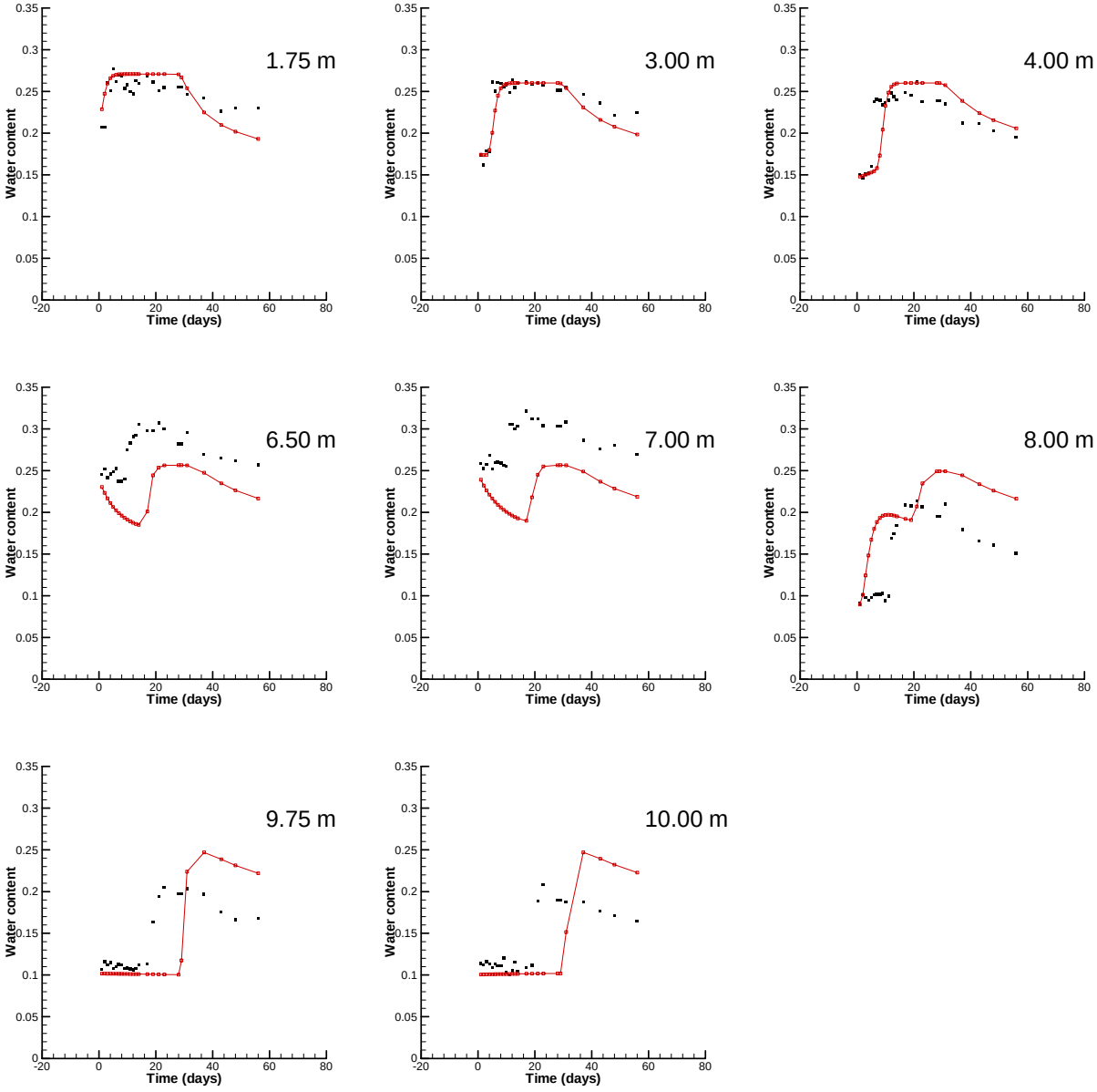


Figure 11-17. 1-D forward simulation of infiltration during experiment 3 at borehole 422 using mean hydraulic parameters from CARSEL.

11.1.11 Forward Flow Modeling Based on Site Data

Whereas public data imply that soils at the MAC consist of sandy loam down to a depth of about 16 m, site data indicate that the upper 20 m of soil consist of 10 layers composed of four different soil types (sandy loam, gravelly loam sand, sand and sandy clay loam), which are laterally heterogeneous on scales exceeding 20 – 30 m (see paper 1, this issue). Whereas publicly available data identify a regional water table that is locally about 22 m deep, site data indicate the additional presence of a perched water table at a depth of about 13 m across

the site. We conduct a series of one- and two-dimensional forward numerical simulations of flow during infiltration experiment 3 based on this site-specific conceptual model. Our computational domain consists of 9 - 11 layers in line with Figures 11-9 – 11-10. In one-dimensional simulations, the water table at depth 13 m constitutes a constant head boundary. In two-dimensional simulations, either a constant head boundary is prescribed at a depth of 13 m or a no-flow boundary at a depth of 20 m. The latter condition is motivated by the fact that soils below 16 m consist of low-permeability sandy clay loam or clay. Neutron-probe data suggest that the water table at the eastern and northern boundaries is higher by about 0.2 m than at the western and southern boundaries. This is reflected in the head boundaries prescribed on the sides of the two-dimensional domain. The latter extends 110 m horizontally. A constant flux of 2.66 cm/day is maintained at the soil surface during the first 28 days of the experiment within the irrigated plot, and a flux of zero outside this plot. During the remainder of the experiment, the soil surface constitutes a no-flow boundary.

Figure 11-18 shows the results of a one-dimensional forward simulation in borehole 402 using mean hydraulic parameters from CARSEL. The computed response (curves) captures in a very crude way the observed behavior (dots). However, there are large differences between measured and computed water contents and wetting-front arrival times. This suggests that the parameter estimates would have to be improved by model calibration against the observed water contents.

Figure 11-19 shows what happens when we adopt mean hydraulic parameters from ROSETTA. The results are much poorer than those in Figure 11-18. The same happens when we adopt mean hydraulic parameter values from RAWLS.

Figure 11-20 depicts results obtained using Bayesian updates of ROSETTA parameter estimates. Though there is some improvement in comparison to results obtained without updating, the results are still poorer than those obtained using mean hydraulic parameters from CARSEL. The same happens when we use updated parameter estimates from RAWLS.

These and similar simulations corresponding to borehole 422, summarized in Table 11-6, indicate that mean hydraulic parameters from CARSEL provide best fit simulations to measured water content data in all cases. Bayesian updates improve the accuracy of generic ROSETTA and RAWLS hydraulic parameters, but not significantly.

Figure 11-21 compares two-dimensional forward simulations and measurements along a N-S uniformly layered transect passing through boreholes 402, 422 and 442 (see Figure 11-10) using mean hydraulic parameter values from CARSEL. The quality of the results varies with depth, with simulated wetting front arrival times lagging by up to ten days behind those measured. Results obtained using mean parameter estimates from ROSETTA or RAWLS, as well as those obtained using Bayesian updates of these parameters, are much less satisfactory. The same is true for a central E-W transect through boreholes 422, 423 and 425 (Figure 11-10). A summary of all these runs in Table 11-7 indicates that hydraulic parameter estimates from CARSEL are generally superior to those from ROSETTA, RAWLS, or their Bayesian updates. However, none of the forward simulations are entirely satisfactory and there is an obvious need to calibrate the models against observed system behavior.

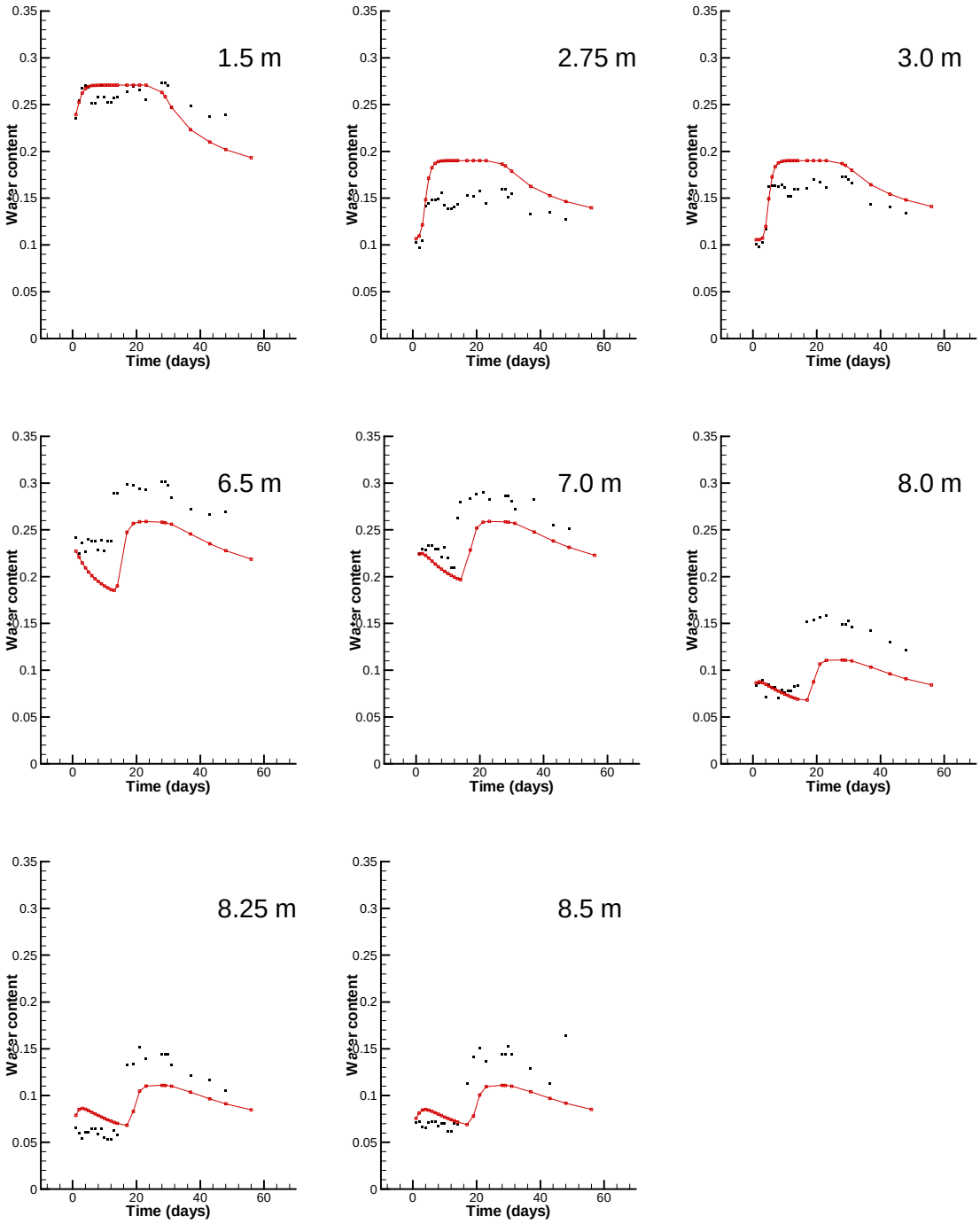


Figure 11-18. 1-D simulation of infiltration experiment 3 at borehole 402 using mean hydraulic parameter values from CARSEL at various depths.

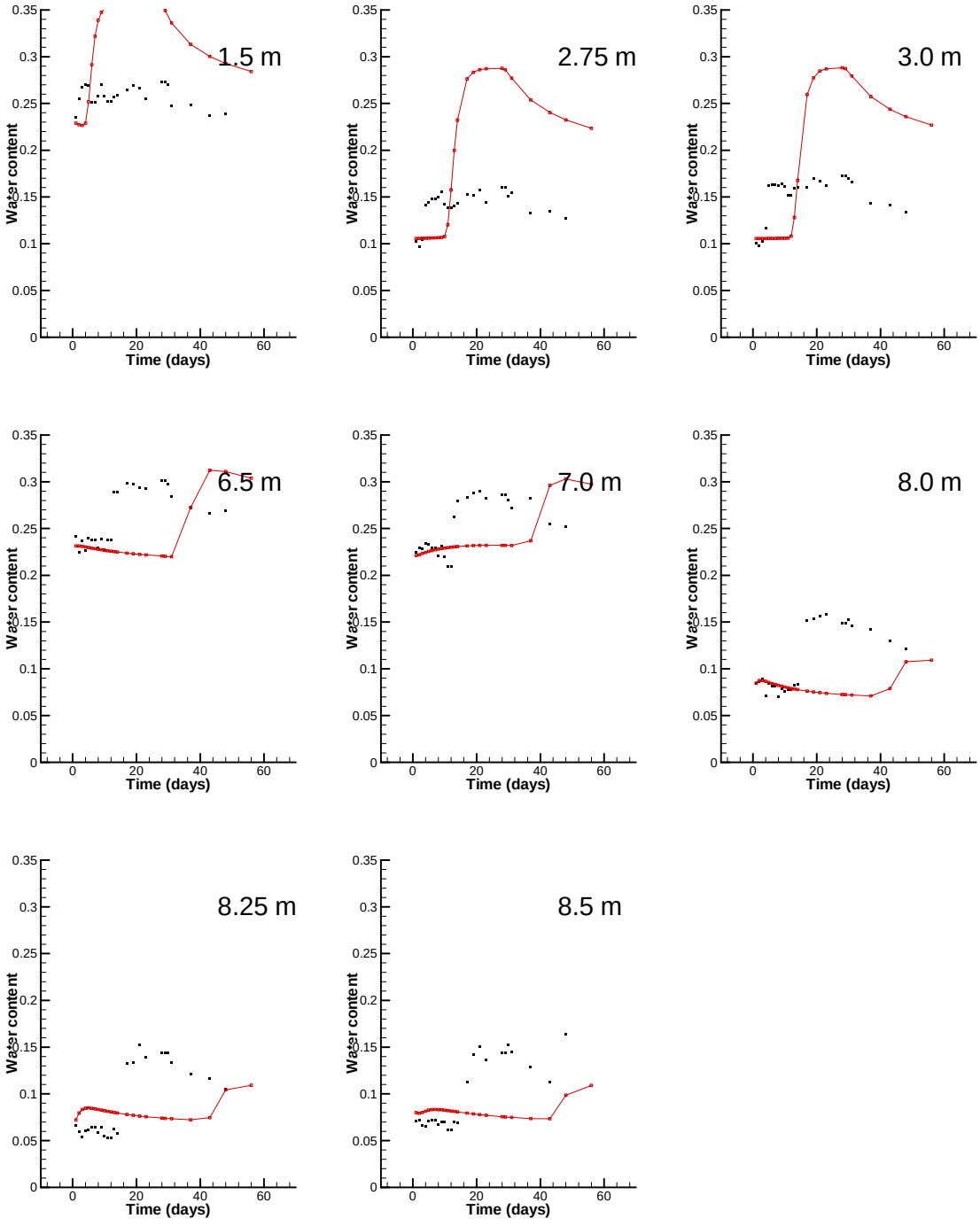


Figure 11-19. 1-D simulation of infiltration experiment 3 at borehole 402 using mean hydraulic parameter values from ROSETTA at various depths.

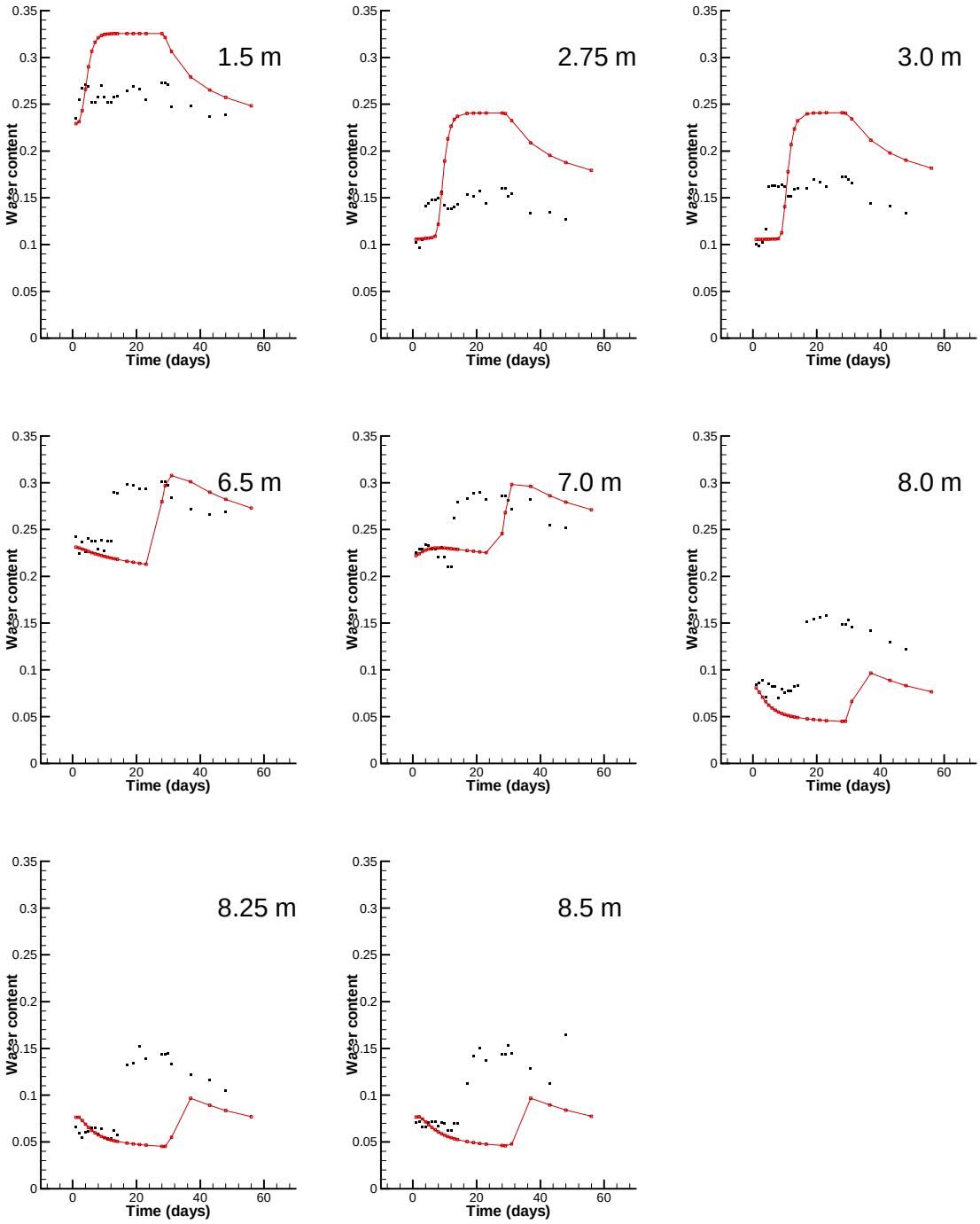


Figure 11-20. 1-D simulation of infiltration experiment 3 at borehole 402 using Bayesian updates of saturated hydraulic conductivity, van Genuchten's α and n based on Rosetta estimates.

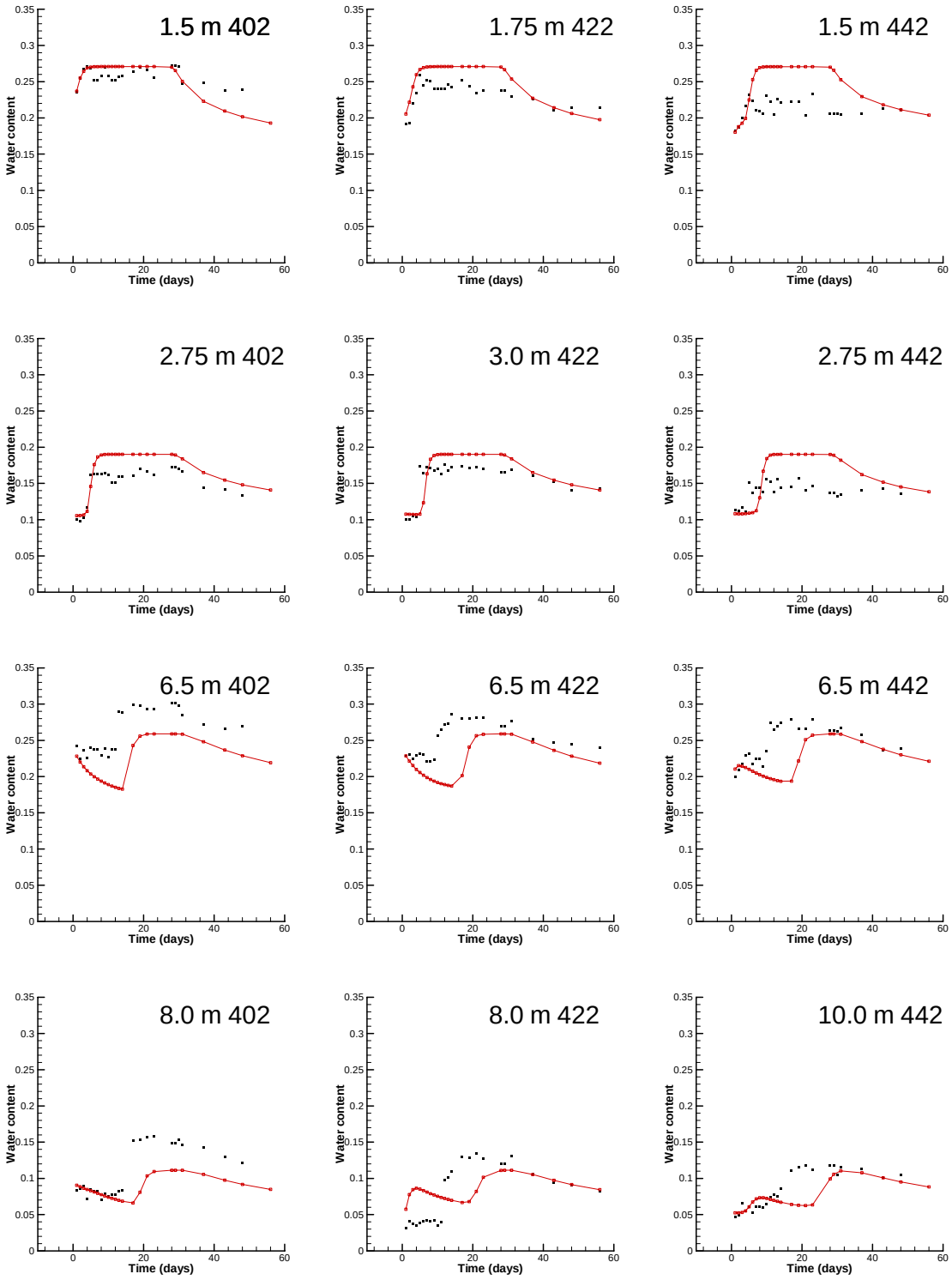


Figure 11-21. 2-D forward simulation along N-S uniformly layered transect using mean hydraulic parameter values from CARSEL at various depths.

Table 11-6. Results of forward simulations using 1-D layered models.

Borehole	Parameter source	Sum of weighted squared residuals
402	CARSEL	7.28E+02
402	ROSETTA	3.62E+03
402	RAWLS	3.35E+04
402	Bayesian updates based on Rosetta neural network	2.05E+03
402	Bayesian updates based on Rawls-Brakensiek regression	2.59E+04
422	CARSEL	1.53E+03
422	ROSETTA	3.74E+03
422	RAWLS	3.66E+03
422	Bayesian updates based on Rosetta neural network	2.30E+03
422	Bayesian updates based on Rawls-Brakensiek regression	3.21E+03

Table 11-7. Results of forward simulations using 2-D uniformly layered models.

Transect	Parameter source	Sum of weighted squared residuals
N-S	CARSEL	2.65E+03
N-S	ROSETTA	5.35E+03
N-S	RAWLS	5.53E+03
N-S	Bayesian updates from Rosetta	3.17E+03
N-S	Bayesian updates from Rawls-Brakensiek regression	4.06E+03
E-W	CARSEL	2.01E+03
E-W	ROSETTA	6.14E+03
E-W	RAWLS	6.73E+03
E-W	Bayesian updates from Rosetta	3.46E+03
E-W	Bayesian updates from Rawls-Brakensiek regression	4.72E+03

11.1.12 Inverse Flow Modeling Based on Site Data

We use the inverse code ITOUGH2 (Finsterle, 1999 a-b) to calibrate our one- and two-dimensional flow models against observed water contents at the MAC. The inverse code estimates hydraulic parameters by minimizing a negative log likelihood criterion equal to the weighted sum of squared water content and parameter residuals. Water content residuals are differences between simulated and observed water contents, the square of each being weighted by the inverse variance of the corresponding observation error. Parameter residuals are differences between posterior (inverse) and prior (input) parameter estimates, the square of each being weighted by the inverse variance of the corresponding prior estimation error.

Previously we described various ways to obtain prior hydraulic parameter estimates and error variances for the MAC site. Parameter estimates obtained on the basis of Bayesian updating have very small variances and therefore very large weights.

The latter allow only minimal departure of the posterior parameters from their initial (input) values, which is not enough to yield a significant improvement in model fit (reduction in the weighted sum of squared water content residuals). As unsaturated flow equations are highly nonlinear and soil hydraulic parameter estimates are correlated with each other, it is difficult to identify soil hydraulic properties uniquely using inversion. The more accurate are the input parameters, the higher is the prospect of obtaining meaningful inverse estimates. Since the CARSEL database has proven to yield best results in forward simulations, we adopt the mean and variance of hydraulic parameters from Meyer *et al.* (1997) as input into the inverse code.

Calibrating a one-dimensional model consisting of a single uniform layer against water content data in various boreholes during infiltration experiment 3 brings about only a minor improvement over the uncalibrated model. This model is clearly inferior to the multilayer models we consider below.

As the sandy loam at depths 0 - 2 m has a different bulk density than deeper sandy loam layers, we estimate its hydraulic properties separately. This yields a total of four materials for each sequence of layers: Sandy loam in the top layer, sandy loam in deeper layers, gravel loamy sand and sand. Sensitivity analysis about the prior parameter estimates suggests (see Wang, 2002, for details) that it should be possible to estimate independently the saturated hydraulic conductivity K_s and van Genuchten's n and α for each of these materials, and this is what we do.

Figure 11-22 shows matches between simulated (curves) and measured (dots) water contents in borehole 402 during infiltration experiment 3 following inversion. Inverse modeling is seen to have improved these matches significantly as compared to the forward modeling results in Figure 11-18. It has also brought about a significant change in the estimate (mean) and reduction in the estimation error (variance) of each parameter. The same happens when we calibrate our one-dimensional model against water content data from boreholes 422, 442, 423 and 425.

Figure 11-23 shows the results of two-dimensional inverse modeling along the western N-S transect by considering the soil to consist of horizontally uniform layers. The matches are seen to be much better than those obtained prior to inversion in Figure 11-21. While some simulation results fit the data well, other are systematically too low or too high. For example, water content in the top sandy loam layer in borehole 402 is systematically under-predicted, whereas in sandy loam and sand layers at depths 6 - 10 m in boreholes 422 and 442 it is systematically over-predicted. Figure 11-23 suggests indirectly that computed wetting-front arrival times in deeper sections of borehole 442 lag considerably behind the measured ones. Similar results are obtained for the central E-W transect. We conclude that it may be necessary to account for lateral variations in layer properties.

Variogram analysis has shown that the dominant horizontal correlation scale of soil hydraulic parameters at the MAC is 20 - 25 m. We therefore subdivide the transect into 3 horizontal segments, one per borehole. We ascribe to each segment initial parameter values equal to those previously obtained from corresponding one-dimensional inverse modeling results. This yields a total of 36 parameters per transect.

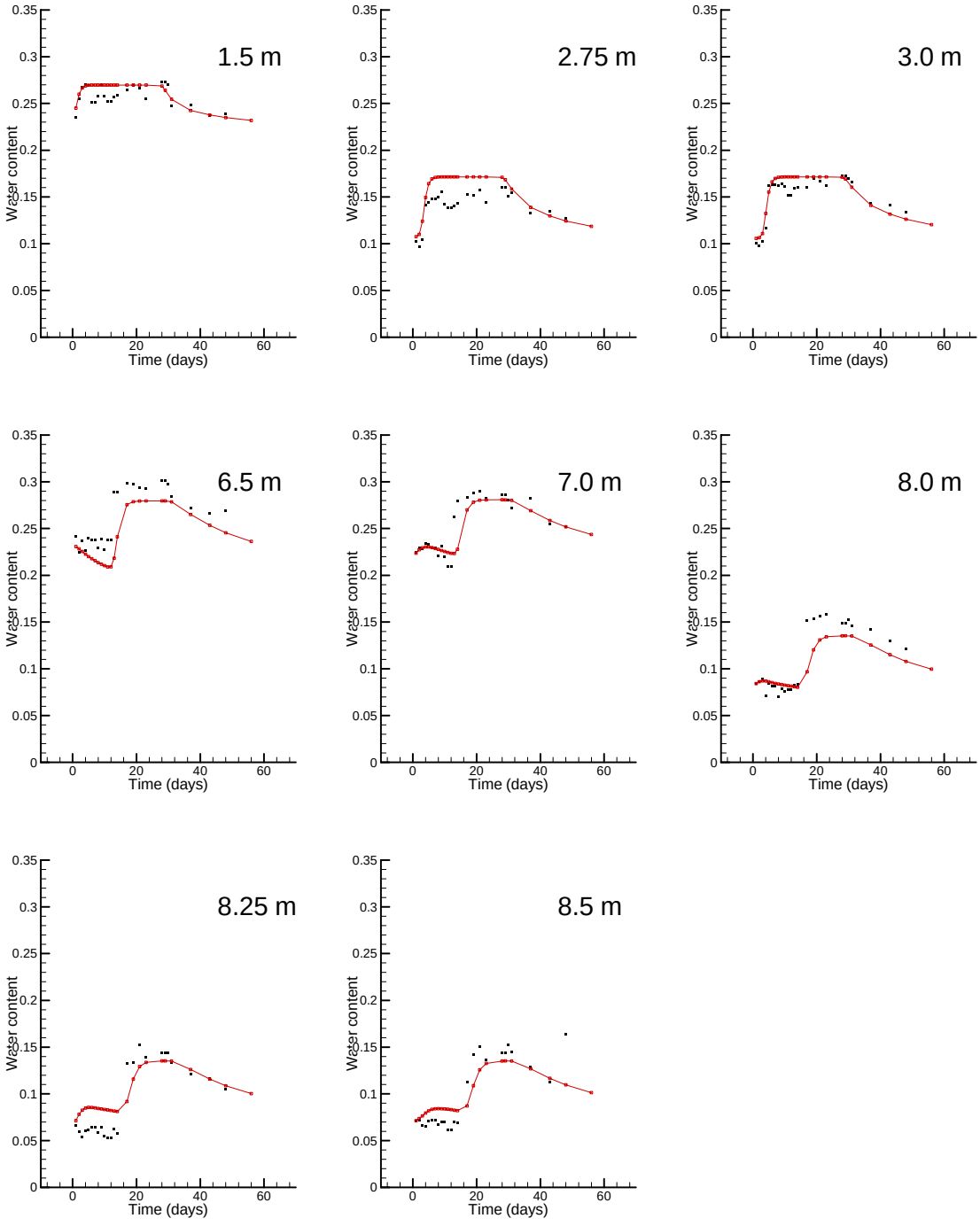


Figure 11-22. 1-D simulation of infiltration experiment 3 in borehole 402 using inverse estimates of saturated hydraulic conductivity and van Genuchten's α and n.

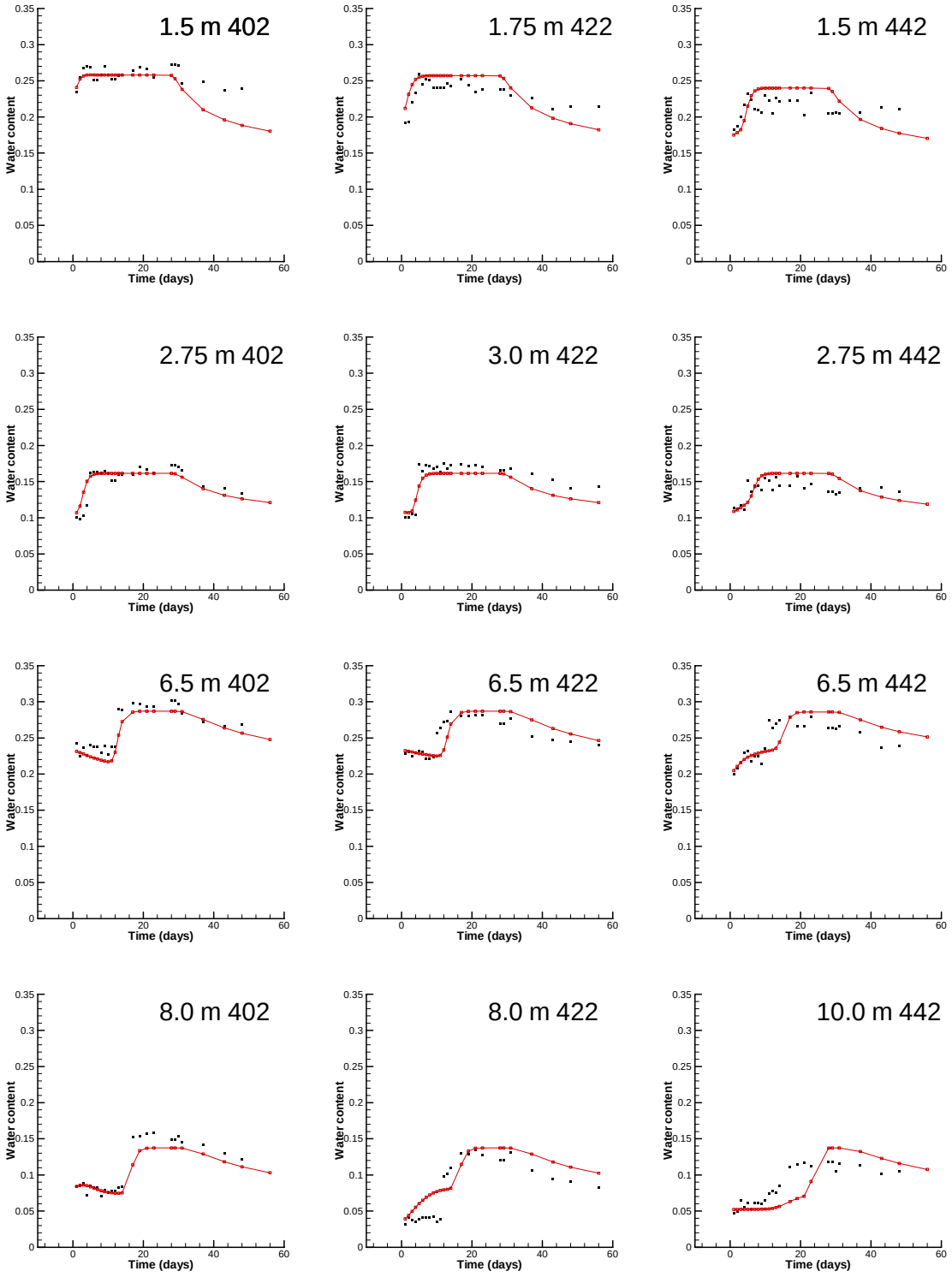


Figure 11-23. 2-D simulation of infiltration experiment 3 along western N-S transect (boreholes 402, 422, 442) using inverse estimates of saturated hydraulic conductivity and van Genuchten's α and n. Uniform soil layers.

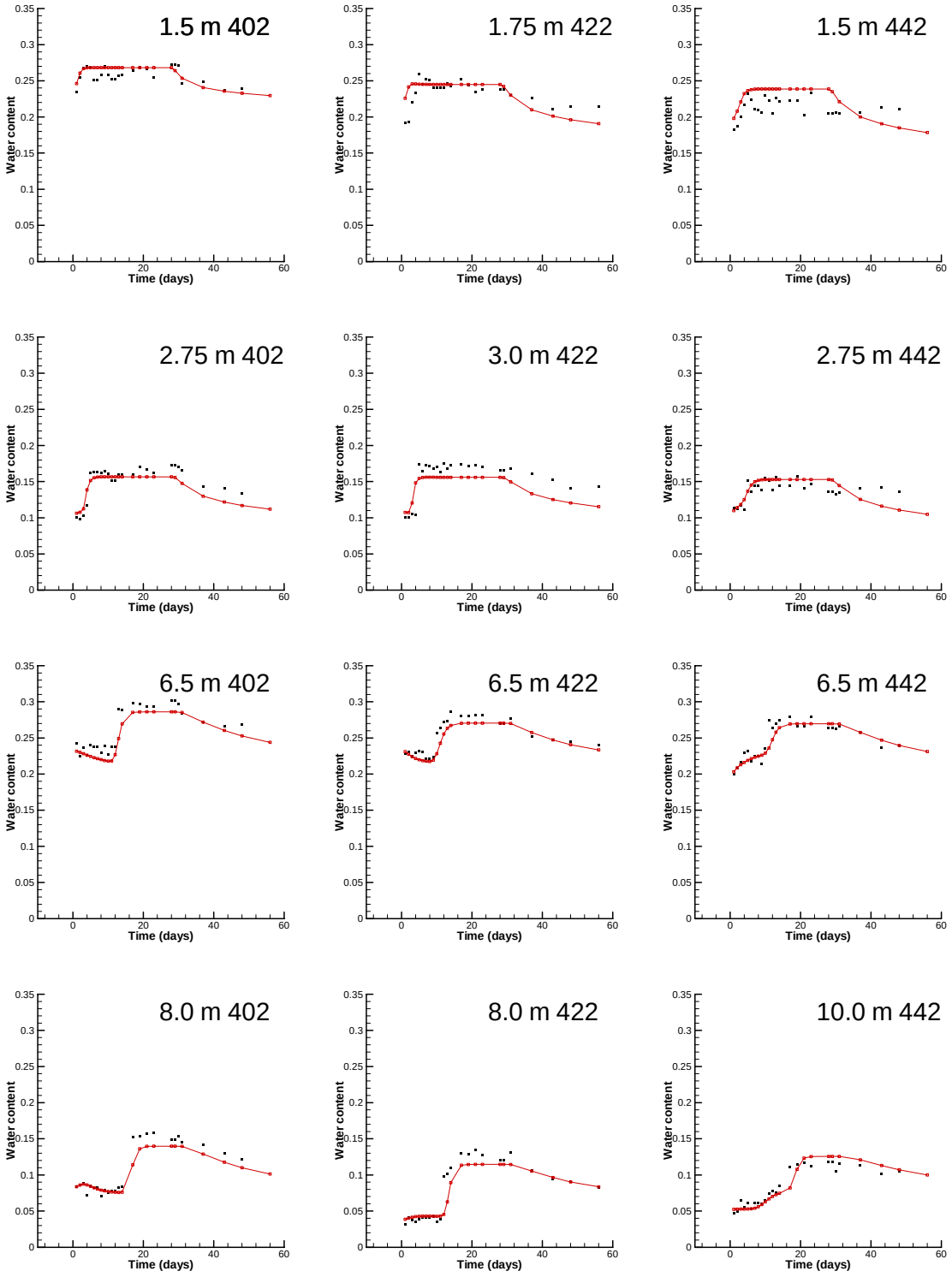


Figure 11-24. 2-D simulation of infiltration experiment 3 along western N-S transect (boreholes 402, 422, 442) using inverse estimates of saturated hydraulic conductivity and van Genuchten's α and n. Non-uniform layers.

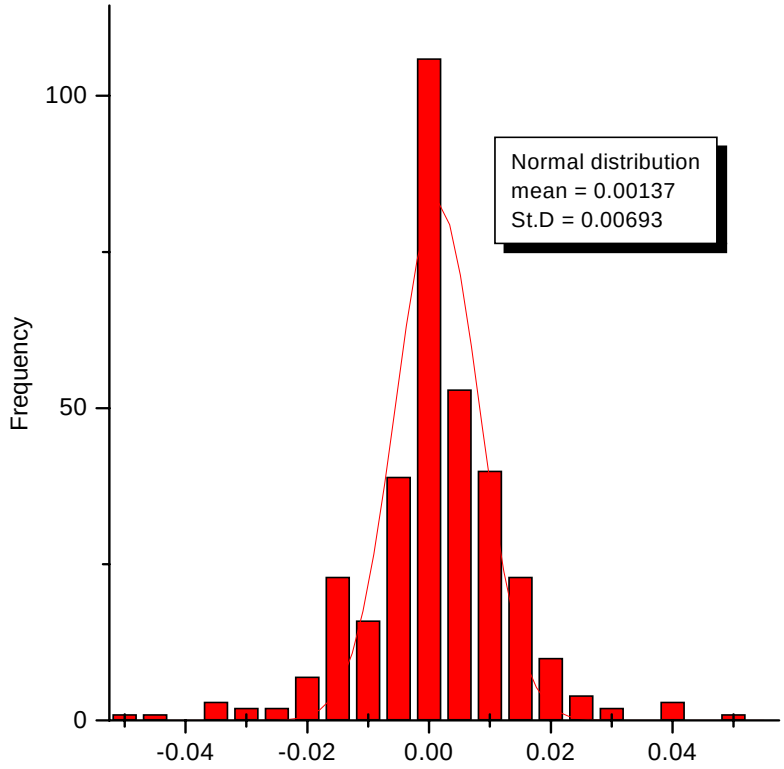


Figure 11-25. Histogram of differences between observed and simulated water contents along western N-S transect (boreholes 402, 422, 442) following inversion. Non-uniform layers.

Figure 11-24 compares simulated and observed water contents using inverse parameter estimates along the western N-S transect. The fit is seen to be good in all cases. A histogram of residuals (Figure 11-25) suggests that they are close to normal with a near-zero mean and small standard deviation. At a confidence level of 95%, only 14 out of the 300 residuals are identified as outliers. The parameter estimates differ only slightly from their one-dimensional counterparts. Similar results are obtained for the central E-W transect.

Tables 11-8 and 11-9 compare the various one- and two-dimensional models along the western N-S and central E-W transects, respectively, following inversion. The quality of model fit is compared on the basis of a D-optimality criterion equal to the determinant of the covariance matrix of parameter estimation errors, an A-optimality criterion equal to the trace of this matrix, an E-optimality criterion equal to the largest absolute eigenvalue of the same matrix (Steinberg and Hunter, 1984), the negative log likelihood criterion that is minimized during the inverse process, and the logarithm of the corresponding likelihood function (which is being maximized). These model fit criteria can be used to compare the quality of different models that have similar structure and number of parameters (such as our N-S and E-W two-dimensional models with uniform layers), but not models that have different structures or numbers of parameters (such as our two-dimensional models with uniform and non-uniform

layers). To validly compare the quality of all models in Tables 11-8 and 11-9, we employ likelihood-based model discrimination criteria due to Akaike (1974) and Kashyap (1982) as done previously by Carrera and Neuman (1986a-c). The smaller (or more negative) are these criteria, the better is the model. The model discrimination criteria consistently identify the uniform one-dimensional model as being the worst among those considered and the two-dimensional non-uniformly layered model as being the best.

Table 11-8. Model quality criteria for four 1-D and two 2-D models associated with N-S transect.

Model Quality Criteria	1-D models				2-D N-S models	
	422 uniform	422	402	442	Uniform layers	Non-uniform layers
D-optimality	4.79E-11	5.77E-68	5.65E-58	1.15E-58	5.83E-57	5.50E-179
A-optimality	9.21E-03	2.00E-04	4.80E-04	8.95E-04	1.96E-03	1.11E-03
E-optimality	9.00E-03	1.30E-04	2.21E-04	3.78E-04	8.49E-04	2.35E-04
Negative log likelihood	1.28E+03	2.34E+02	1.99E+02	2.29E+02	8.20E+02	2.70E+02
Log likelihood	-3.45E+01	6.28E+02	6.09E+02	6.96E+02	6.11E+02	1.00E+03
Akaike	7.51E+01	-1.23E+03	-1.19E+03	-1.37E+03	-1.20E+03	-1.93E+03
Kashyap	1.02E+02	-1.06E+03	-1.05E+03	-1.22E+03	-1.04E+03	-1.45E+03

Table 11-9. Model quality criteria for three 1-D and two 2-D models associated with E-W transect.

Model Quality Criteria	1-D Models			2-D E-W Models	
	422	423	425	Uniform layers	Non-uniform layers
D-optimality	5.77E-68	3.16E-74	4.06E-56	7.55E-62	1.44E-194
A-optimality	2.00E-04	9.79E-05	1.10E-03	1.04E-03	3.55E-04
E-optimality	1.30E-04	5.29E-05	6.19E-04	5.27E-04	8.80E-05
Negative log likelihood	2.34E+02	2.45E+02	2.61E+02	8.37E+02	4.54E+02
Log likelihood	6.28E+02	7.00E+02	6.99E+02	6.32E+02	9.55E+02
Akaike	-1.23E+03	-1.38E+03	-1.37E+03	-1.24E+03	-1.84E+03
Kashyap	-1.06E+03	-1.19E+03	-1.23E+03	-1.07E+03	-1.31E+03

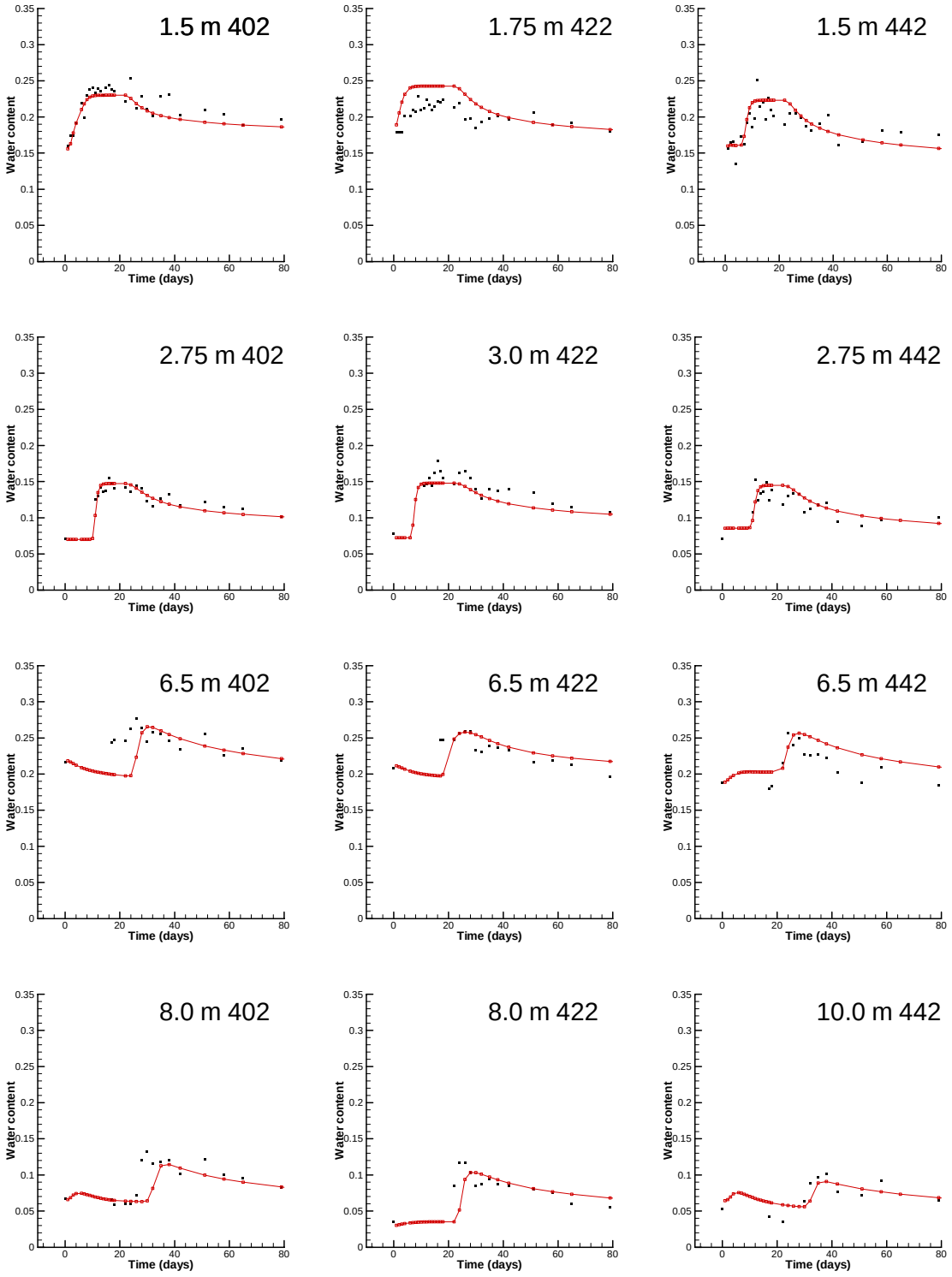


Figure 11-26. 2-D simulation of experiment 1 along the N-S transect (boreholes 402, 422, 442) using layered non-uniform conceptual model and hydraulic parameters obtained from experiment 3.

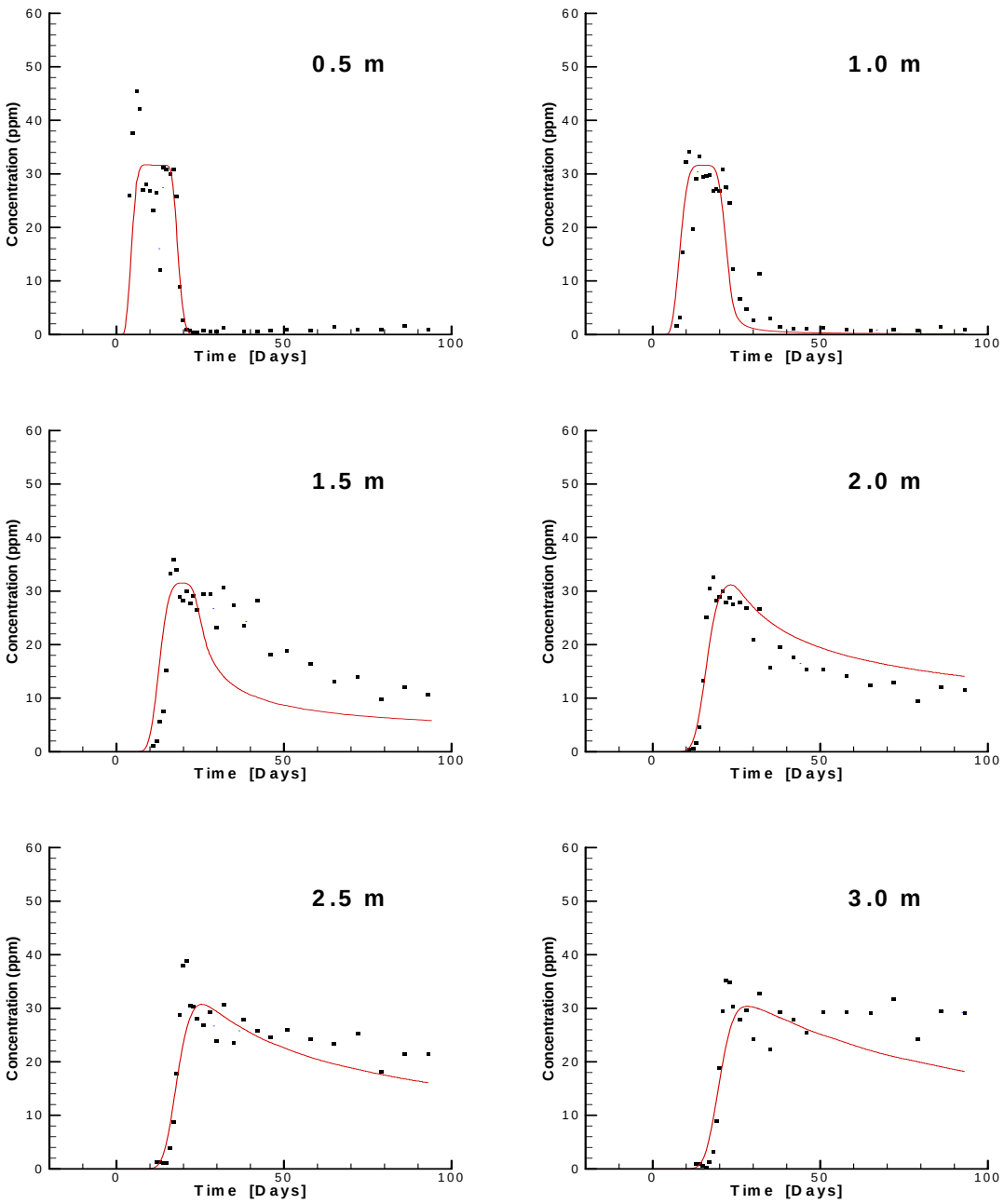


Figure 11-27. Simulated (curves) and measured (dots) bromide concentrations at the south monitoring island obtained using inverse estimates of dispersivity, immobile or excluded water content, and mass transfer coefficient.

11.1.13 Confirmation of Inverse Modeling Results

Our conceptual model and inverse hydraulic parameter estimates are based on data collected during infiltration experiment 3. We use them to simulate water contents at the MAC site during experiment 1.

Figure 11-26 shows a two-dimensional simulation of infiltration experiment 1 along the western N-S transect using the non-uniform layered model and inverse hydraulic parameters, which are obtained by calibrating against water contents observed during experiment 3. The good matches between simulated and observed water contents for experiment 1 constitute a confirmation of the calibrated model. Similar results were obtained for the central E-W transect.

11.1.14 Inverse Modeling of Solute Transport

One-dimensional inverse estimates of soil hydraulic parameters based on infiltration experiment 3 were adopted for one-dimensional solute transport modeling of bromide tracer data collected during experiment 1. We model transport as an advection-dispersion process with linear mass transfer to and from a zone of immobile water or, equivalently, anion exclusion.

Figure 11-27 shows simulated (solid curve) and measured (dots) bromide concentrations at the south monitoring island obtained using inverse estimates of dispersivity, immobile or excluded water content, and mass transfer coefficient. Varying the latter has no effect on our simulations.

11.1.15 Conclusions

Our work at the MAC site leads to the following conclusions:

- Publicly available data suggest that soils at the experimental MAC site consist of sandy loam down to a depth of about 16 m. Site data indicate instead that the soil profile down to a depth of 20 m consists of 10 layers composed of four different soil types (sandy loam, gravelly loam sand, sand and sandy clay loam), which are laterally heterogeneous on scales exceeding 20 – 30 m.
- Publicly available data identify a regional water table that is locally about 22 m deep. Site data indicate the additional presence of a perched water table at a depth of about 13 m across the site.
- Generic databases make it possible to associate each soil type at the MAC with prior probability distributions of relevant hydraulic parameters. The distributions vary with the database.
- Regression and neural network pedotransfer models allow translating information about the composition and density of discrete soil samples at the MAC into estimates of hydraulic parameters. The statistics of these estimates vary with the model and generic database on which it is based.
- Generic probability distributions of hydraulic parameters based solely on soil class can be updated by treating our pedotransfer estimates of these parameters as "measurements." Bayesian updating has brought about a change in the mean (reduction in bias) and a drastic reduction in its estimation variance.
- Our conclusion about the layered and laterally heterogeneous structure of soils at the MAC is supported in large part by

variograms of soil compositional data and hydraulic parameter estimates obtained using pedotransfer models.

- The same conclusion is supported by neutron probe readings, which correlate with soil compositional data and appear to be more consistent across the site than are these data.
- Relying entirely on publicly available data leads to a poor reproduction of observed infiltration at the MAC site.
- Mean hydraulic parameter estimates based on the generic CARSEL database, published by Meyer *et al.* (1997) on the basis of data assembled by Carsel and Parrish (1988), allow a much better reproduction of observed water contents at the MAC than do estimates based on the RAWLS (Rawls et al., 1982) or ROSETTA (Schaap and Leij, 1998) databases. Bayesian updating of mean hydraulic parameters based on RAWLS and ROSETTA leads only to a marginal improvement in their ability to reproduce observed water contents at the site.
- Regardless of our choice of database, pedotransfer function or flow model we were unable to reproduce observed water contents at the site without calibrating the flow model against such observations. We used the mean and variance of hydraulic parameters associated with the CARSEL database as inputs into our inverse flow models.
- Calibrating a one-dimensional model consisting of a single uniform layer against water content data in various boreholes during infiltration experiment 3 brings about only a minor improvement over the uncalibrated model. This model is clearly inferior to multilayer models we consider for the site.
- Inverse modeling brings about a significant improvement in our ability to reproduce observed water contents at the MAC using multilayer models regardless of whether the latter are one- or two-dimensional with uniform or laterally heterogeneous layers. Inverse modeling is accompanied by a significant change in the estimate (mean) and reduction in the estimation error (variance) of each hydraulic parameter.
- Likelihood-based model discrimination criteria consistently identify the uniform one-dimensional model as being the worst among those considered and the two-dimensional model with laterally heterogeneous layers as being the best.
- The latter model, calibrated against water content data observed during infiltration experiment 3, reproduces with fidelity water content data observed during infiltration experiment 1.
- One-dimensional inverse estimates of soil hydraulic parameters based on infiltration experiment 3 were adopted for one-dimensional solute transport modeling of bromide tracer data collected during experiment 1. An advection-dispersion model with linear mass transfer to and from a zone of immobile water, or anion exclusion, was employed in forward and inverse modes. Following calibration, the model reproduced reasonably well measured bromide breakthroughs at the monitoring island during experiment 1.

11.2 Apache Leap Research Site

11.2.1 Introduction

Issues associated with the site characterization of fractured rock terrains, the analysis of fluid flow and contaminant transport in such terrains and the efficient

handling of contaminated sites are typically very difficult to resolve. A major source of this difficulty is the complex nature of the subsurface "plumbing system" of pores and fractures through which flow and transport in rocks take place. This is especially true for fractured rocks that are only partially saturated.

One relatively convenient way to characterize such rocks is by means of single-hole and cross-hole pneumatic injection tests. Considerable experience with such testing of unsaturated fractured tuffs has been accumulated in recent years by the University of Arizona (U of A) at the Apache Leap Research Site (ALRS) near Superior, Arizona, under the auspices of the U.S. NRC (Chen *et al*., 2000; Guzman *et al*., 1996; Illman *et al*., 1998; Illman and Neuman, 2000, 2001; Neuman *et al*., 2001; Vesselinov *et al*., 2001a-b; Vesselinov and Neuman, 2001: Hyun *et al.*, 2002). This research has established the following conceptual framework for the ALRS:

- It is possible to model airflow through unsaturated fractured tuffs at the ALRS by treating water as if it was immobile, and the rock as if it was a porous continuum.
- During a pneumatic injection test, air moves primarily through fractures most of which contained relatively little water, and the test therefore yields permeabilities and porosities which reflect closely the intrinsic properties of the surrounding fractures. This is so because capillary forces tend to draw water from fractures into porous (matrix) blocks of rock, leaving the fractures saturated primarily with air, and making it difficult for air to flow through matrix blocks.
- Pneumatic permeabilities increase systematically with applied pressure, as air appears to displace water under two-phase flow. In a few single-hole tests, where the injection intervals were intersected by widely open fractures, air permeabilities decrease with applied pressure due to inertial effects. Two-phase flow and inertial phenomena decay rapidly with distance from the injection interval. Enhanced permeability due to slip flow (the Klinkenberg effect) appeares to be of little relevance to the interpretation of single-hole or cross-hole air injection tests at the ALRS.
- Flow in the vicinity of most relatively short pneumatic test intervals is three-dimensional regardless of the number or orientation of fractures in the surrounding rock. This implies that such flow is controlled by a single continuum, representative of a three-dimensional network of interconnected fractures, rather than by discrete planar features. Only in a small number of single-hole test intervals, known to be intersected by widely open fractures, do such features dominated flow. Some pressure records indicate radial flow during early and intermediate times, but none do so fully at late times.
- It is generally not possible to distinguish between the permeabilities of individual fractures, and the bulk permeability of the fractured rock in the immediate vicinity of a test interval, by means of pneumatic injection tests. Hence there is little justification for attempting to model flow through individual fractures at the site. The explicit modeling of discrete features appears to be justified only when one can distinguish clearly between layers, faults, fracture zones, or major individual fractures on scales not much smaller than the domain of interest.

- Air permeabilities obtained from single-hole tests are poorly correlated with fracture densities, as is known to be the case for hydraulic conductivities at many water-saturated fractured rock sites worldwide (Neuman, 1987). This provides further support for Neuman's conclusion that the permeability of fractured rocks cannot be reliably predicted from information about fracture geometry (density, trace lengths, orientations, apertures and their roughness) but must be determined directly by means of hydraulic and/or pneumatic tests.
- Core and single-hole measurements, conducted over short segments of a borehole, provide information only about a small volume of rock in the immediate vicinity of each measurement interval. They tend to vary erratically in space in a manner that renders the rock randomly heterogeneous and anisotropic.
- Local-scale air permeabilities from single-hole tests vary by orders of magnitude between test intervals across the site; their spatial variability is much more pronounced than their dependence on applied pressure.
- It is possible to interpolate some of the core and single-hole measurements at the ALRS between boreholes by means of geostatistical methods, which view the corresponding variables as correlated random fields defined over a continuum. This is especially true about air permeability, porosity, fracture density, water content, and the van Genuchten water retention parameter α, for each of which there are enough measurements to constitute a workable geostatistical sample. It supports the application of stochastic continuum flow and transport theories and models to unsaturated fractured porous tuffs at the ALRS on scales of one meter or more.
- Air-permeability is well characterized by a power variogram, which is representative of a random fractal field with multiple scales of spatial correlation.
- Cross-hole pneumatic injection test data from individual monitoring intervals are amenable to analysis by type-curve and numerical inverse models, which treat the rock as a uniform and isotropic fractured porous continuum. Analyses of pressure data from individual monitoring intervals by the two methods provide information about pneumatic connections between injection and monitoring intervals, corresponding directional air permeabilities, and air-filled porosities. All of these quantities vary considerably from one monitoring interval to another in a given cross-hole test on scales ranging from a few meters to several tens of meters. Thus, even though the analyses treat the rock as if it was pneumatically uniform and isotropic, they ultimately yield information about the spatial and directional dependence of pneumatic connectivity, permeability and porosity across the site.
- The pneumatic permeabilities and porosities of unsaturated fractured tuffs at the ALRS vary strongly with location and scale of measurement. The scale effect is most probably due to the presence in the rock of various size fractures that are interconnected on a variety of scales.
- As there is consistency between single-hole and cross-hole test results, the pronounced permeability scale effect at the ALRS is unrelated to the method of testing. As there is consistency between results obtained by means of diverse steady-state and transient, analytical and numerical methods of test interpretation,

the scale effect is unrelated to the method of interpretation. As neither the single-hole nor the cross-hole test results have been affected by any skin effect of consequence, the scale effect is unrelated to phenomena associated with borehole drilling and completion. The observed permeability scale effect at the ALRS appears to be real.

- The question was asked to what extent can one interpret the observed permeability scale effect at the ALRS by means of a stochastic scaling theory due to Di Federico and Neuman (1997) and Di Federico *et al.* (1999), which views log permeability as a truncated random fractal. There is considerable uncertainty about the magnitude of the observed scale effect at the ALRS, making it difficult to answer this question conclusively. However, there is sufficient correspondence between the site data and the theory to suggest that the latter may indeed provide a viable (if not complete) explanation of the observed scale effect.

These findings amount to a conceptual model of the pneumatic characteristics of, and airflow through, unsaturated fractured tuffs at the ALRS. It however leaves many fundamental questions unanswered, implying that the present conceptual model is incomplete and uncertain. Among conceptual issues that require resolution are the following:

- To what extent do conceptualizing the air permeability and air-filled porosity of unsaturated fractured tuffs at the ALRS as multiscale random fields, defined over a continuum, apply to saturated and unsaturated hydraulic conductivities and porosities?
- To what extent do bulk air permeabilities and air-filled porosities represent hydraulic conductivities and porosities of the fractured rock?
- How can unsaturated hydraulic rock properties be determined under field conditions?
- What roles do fractures play in controlling these properties?
- Would a (stochastic or deterministic) continuum representation of these properties be appropriate?
- How variable are these properties in space? Could their variability be described geostatistically, or would there be a need to consider discrete features such as fractures or channels in the conceptual model of medium heterogeneity?
- What role does preferential flow play under saturated and unsaturated conditions? Can preferential flow channels or unstable fingers be seen, identified and characterized in the field?
- Is there a need to identify fractures and preferential-flow-channels discretely? If so, is this feasible, and how?
- What role does dispersion play in the movement of tracers with water? Can it be predicted by means of continuum stochastic concepts?
- What role does matrix-fracture interaction play during saturated and unsaturated water flow and solute transport in fractured porous rocks? Are dual porosity, dual permeability and/or matrix diffusion active during such flow and transport?
- Can these interaction phenomena be detected in the field, and how?

These fundamental questions of conceptual understanding cannot be answered without conducting appropriate experiments with water and dissolved tracers at the ALRS.

To address them, a long-term ponding infiltration and tracer experiment was conducted at the site by Yao *et al.* (2002). Here we provide a brief synopsis of the experiment and our findings.

11.2.2 Experimental Setup

A 9 x 9 meter plot was divided into nine 3 x 3 meter subplots and instrumented as shown in Figure 11-28. An air compressor was used to expose the surface of the rock and major fractures. Ponding started on November 12, 1999 and bromide tracer was added on May 30, 2000. Ponding ceased on August 28, 2001. The advance of the water front was measured with tensiometers and a neutron probe within each subplot. Solute movement was determined based on solution samples collected from suction lysimeters at various locations and depths.

Figure 11-29 shows cumulative outflow from the pond over each subplot. Examples of neutron count and soil potential variations with depth are given for subplot 1 in Figures 11-30 and 11-31, respectively. Neutron counts were also available for vertical and slanted deep boreholes in and around the experimental plot which are not indicated in Figure 11-28. Figure 11-32 shows bromide concentrations versus time at 3 and 5m depth beneath pond 1.

11.2.3 Findings

The long-term ponding infiltration and tracer experiment at the ALRS has shown that arrivals of water fronts were clearly detectable from the tensiometer data but not as clearly from the neutron probe data. Higher infiltration rates were found in plots with apparent surface fractures than in other plots. After 650 days of flooding with bromide labeled water, 14 of the 45 suction lysimeters have indicated the arrival of surface applied bromide, but only 3 have indicated early arrival. Concentration spikes of bromide were detected in two lysimeters under plot 8 (at 2 and 3 m) and in one lysimeter under plot 9 (0.5 m). Though there was a clear increase in matric potential under all plots at depths of 0.5 and 1.0 m, there was no measurable bromide at 6 of the 9 lysimeters at these depths. High apparent dispersivity values were calculated from bromide breakthrough data at eight sampling points using a one-dimensional model with uniform velocity. The relatively high infiltration rates in some plots, early arrivals of bromide spikes at deep sampling points without showing up at shallow points, and high apparent dispersivity values at some points suggest the presence of a complex system of preferential flow paths at the site. Though fracture data beneath the plots are lacking, it stands to reason that these preferential flow paths are associated with fractures.

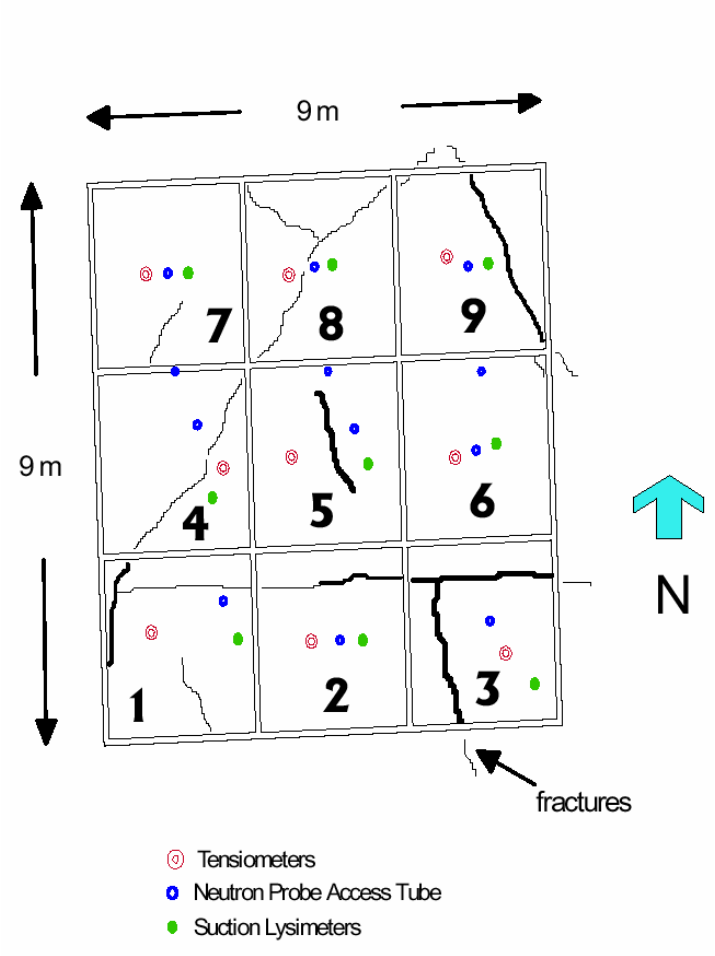


Figure 11-28. Experimental site with plots, instrumentation, and surface exposure of fractures.

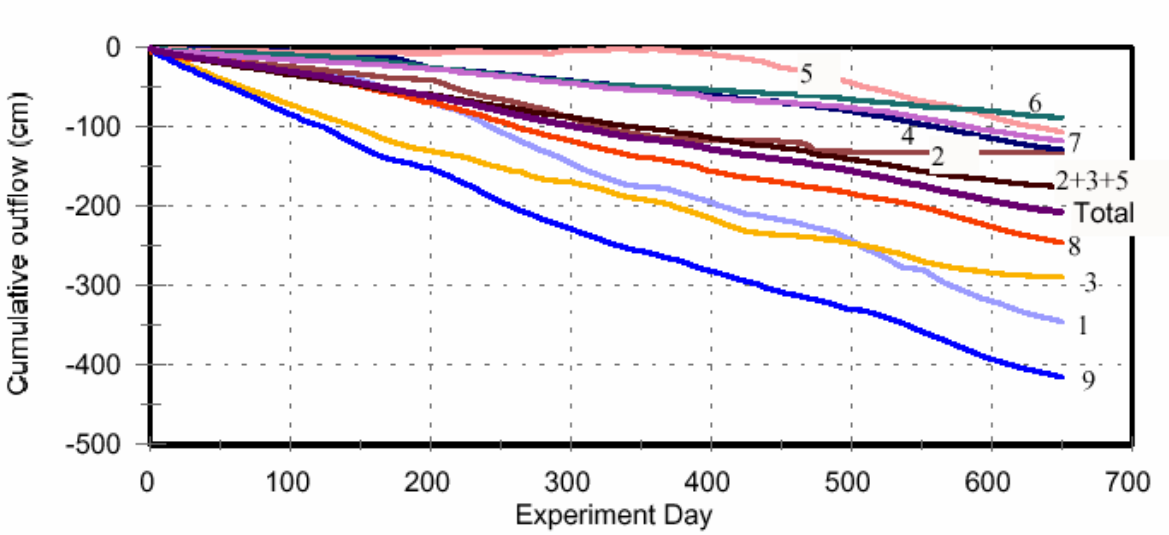


Figure 11-29. Cumulative outflow from each subplot pond.

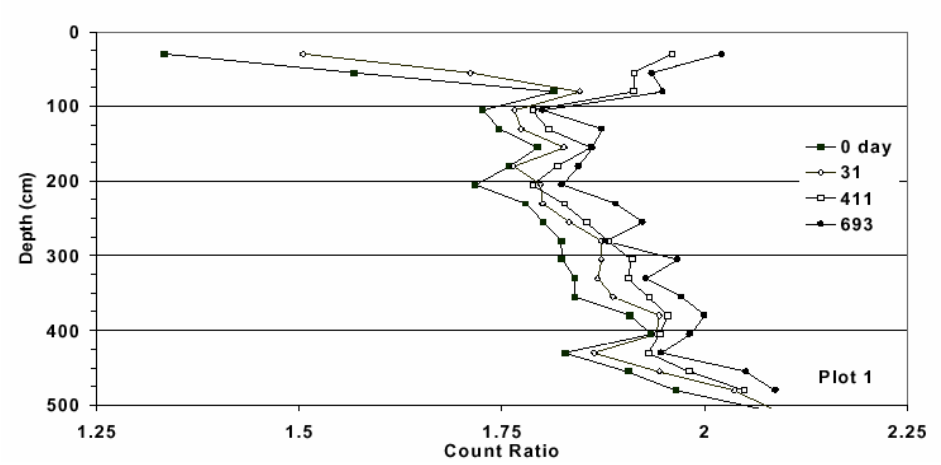


Figure 11-30. Plot 1 neutron probe count ratios versus depth.

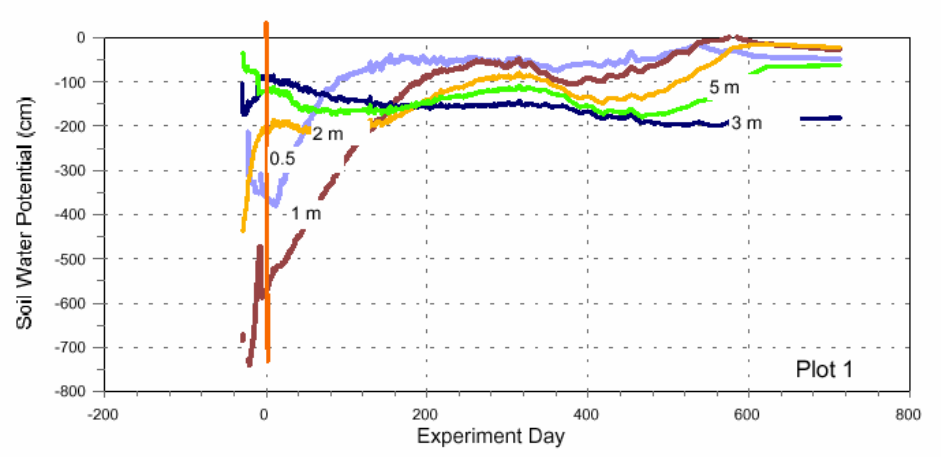


Figure 11-31. Plot 1 soil water potentials versus depth.

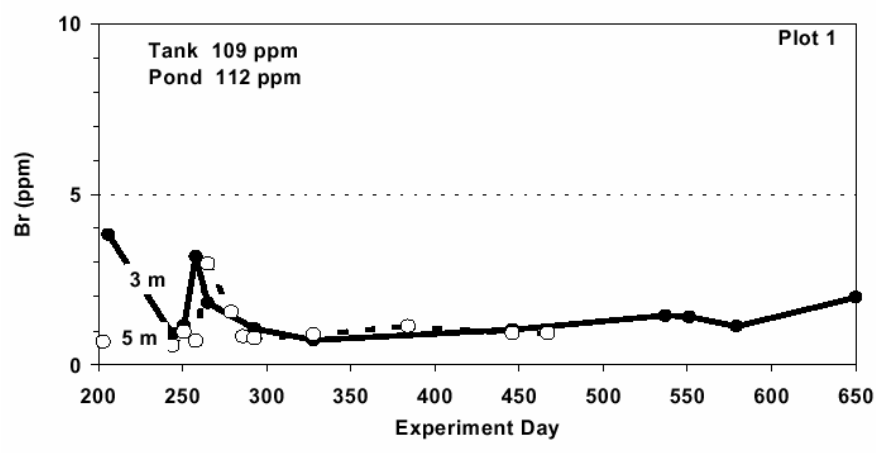


Figure 11-32. Plot 1 bromide concentrations versus time at 3 and 5 m depths.

11.3 Fanay-Augères

This case study is based on a paper by Ando *et al.* (2002).

11.3.1 Introduction

Cacas *et al.* (1990a-b) described flow experiments in a uranium mine at Fanay-Augères near Limoges, France. The experiments were conducted in a 100 m horizontal drift at a depth of 150 m within highly fractured granite. Ten 50 m long boreholes were drilled radially from the drift in three vertical planes at distances of 25.9 m and 23.7 m from each other (Figure 11-33). Planes 1 and 3 contain three boreholes each and plane 2 contains four boreholes. Fractures were mapped in the drift and observed in borehole cores. Water was injected into 180 packed-off chambers of length 2.5 m and diameter 7.6 cm at 10 MPa. Similar injection tests were conducted in 50 chambers of length 10 m and 10 chambers of length 50 m. Head was monitored in 68 chambers of length 5 m for over 400 days starting February 2, 1985 and ending March 3, 1986. They reveal the presence of a near-horizontal water table (zero-pressure isobar) a short distance above the drift. Below it, the rock is fully saturated with water.

Inflow into the drift was measured continually before and after partial desaturation due to borehole drilling. It exhibits two stages of near steady state with an intervening transient flow period (Figure 11-34). Two tracer tests were conducted by injecting different tracers into 5 m long chambers in boreholes F2 and F3 (Figure 11-35). Chambers are identified by symbols in which the first character refers to chamber number (increasing toward the drift, as illustrated in Figure 11-35) and the rest identify the borehole (for example, 4F3 is chamber 4 in borehole F3). During the first test that commenced on September 19, 1985, two tracers were injected into 7F2 and one into each of 6F2, 4F2, 1F2 and 4F3. During the second test that commenced on December 19, 1985, one tracer was injected into each of 3F2, 6F3, 3F3 and 1F3. Measurements included tracer concentrations at location "A" (a plastic sheet glued to the roof of the gallery near its intersection with borehole F2 at 50 m as measured from the north end of the drift); chambers 1 - 7 of borehole F4; rhodamine WT, injected into 4F3 at 50 m, along the drift intervals 43.5 m - 66.5 m on October 13 and 70.0 m - 99.0 m on October 30, 1985; location "B" in a 25m long drain along the gallery extending from the injection plane of boreholes F1 - F4 to the plane of boreholes F5 - F7; and location "C" along a 25 m - 50 m segment of the gallery downstream of the injection plane (Figure 11-35).

A stochastic discrete fracture network model was developed for the site by Cacas *et al.* (1990a-b). In their model, fractures were represented by circular disks with randomly prescribed radia and orientations based on statistics inferred from in situ geometric observations (Figurer 11-36). Flow in each disk was one-dimensional, purportedly to represent channels. The probability distribution of flow rates through fractures across the network was assumed to be log-normal, in analogy to the distribution inferred from steady state flow rates (and hydraulic conductivities) measured during 2.5 m scale constant-pressure packer tests. This inference ignored about 1/3 of the test values, which fell below the detection limit of $K = 6 \times 10^{-10}$ m/s (Figure 11-37). In postulating the analogy, no notice was taken of the fact that correlation between the intensity of fracturing (as represented by fracture density, measured by number of

fractures per unit length of borehole) and hydraulic conductivities from the 2.5 m scale tests is highly tenuous (Figure 11-38). A similar lack of clear correlation between fracturing and permeability has been noted by us at other sites (Neuman, 1997; Chen *et al.*, 2000), throwing into question the validity of models that relate the latter to the former.

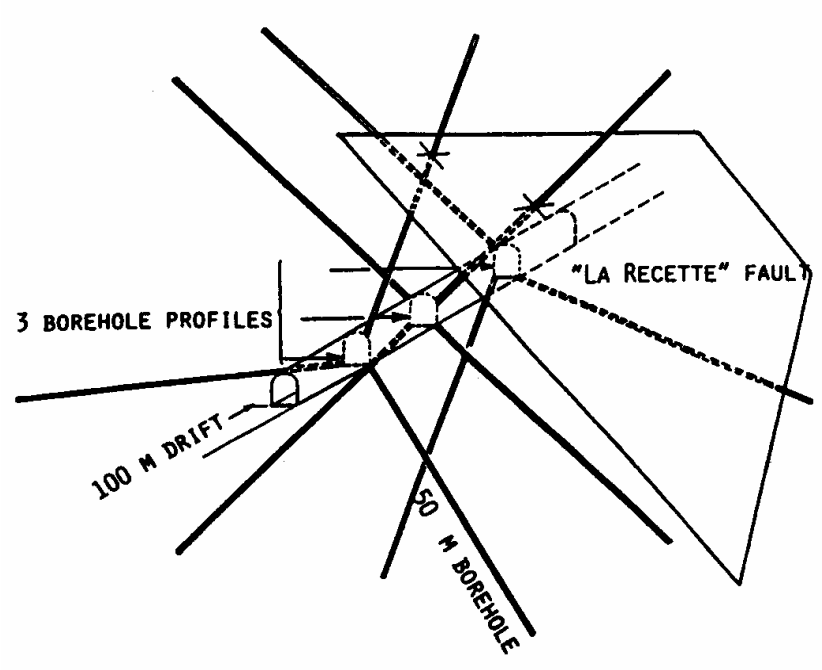


Figure 11-33. Layout of drift and boreholes (after Cacas *et al.*, 1990a). With permission, AGU.

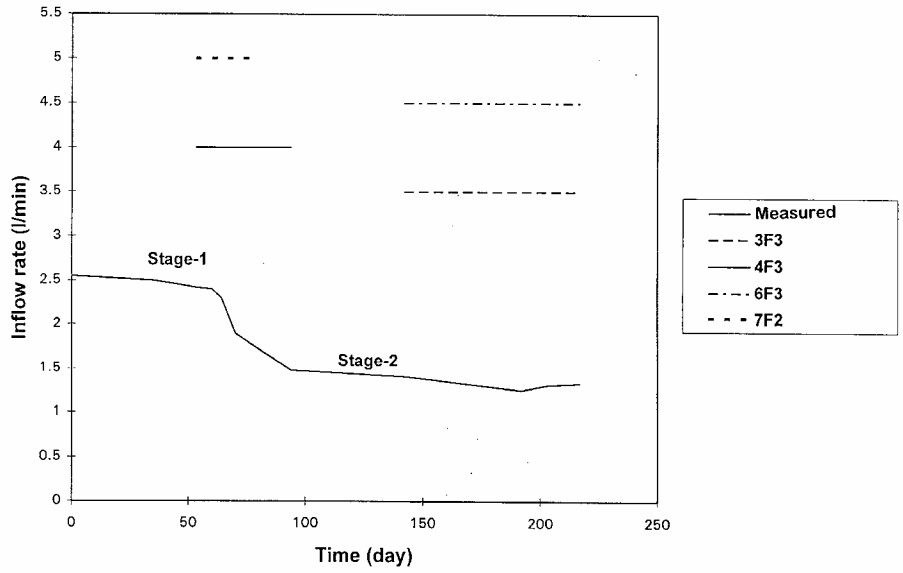


Figure 11-34. Measured inflow rates into drift and periods of tracer tests.

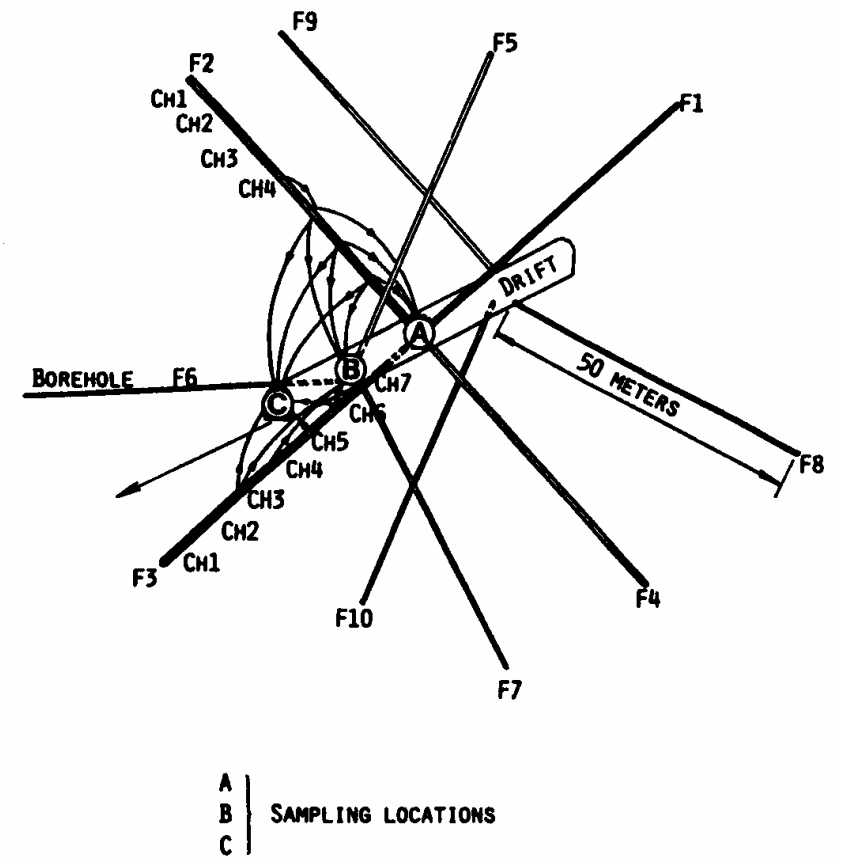


Figure 11-35. Tracer release intervals and monitoring locations in drift (after Cacas *et al.*, 1990b). With permission, AGU.

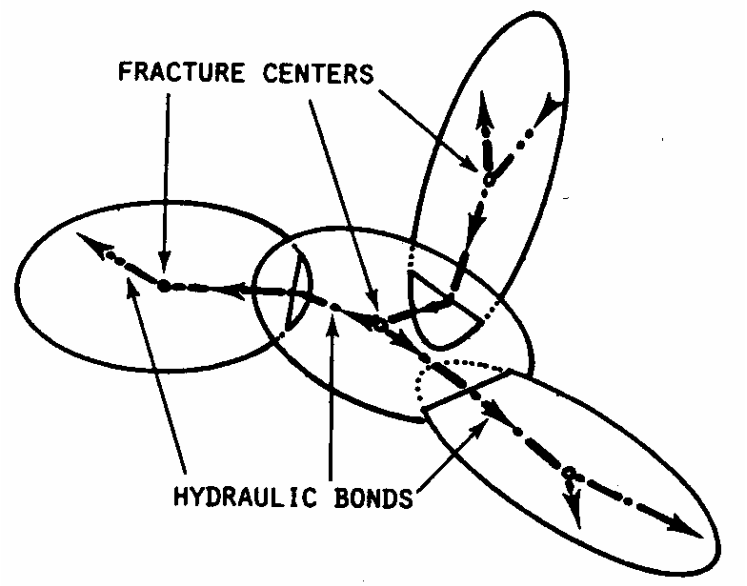


Figure 11-36. 3-D network of circular disks with 1-D flow and transport paths (after Cacas *et al.*, 1990a). With permission, AGU.

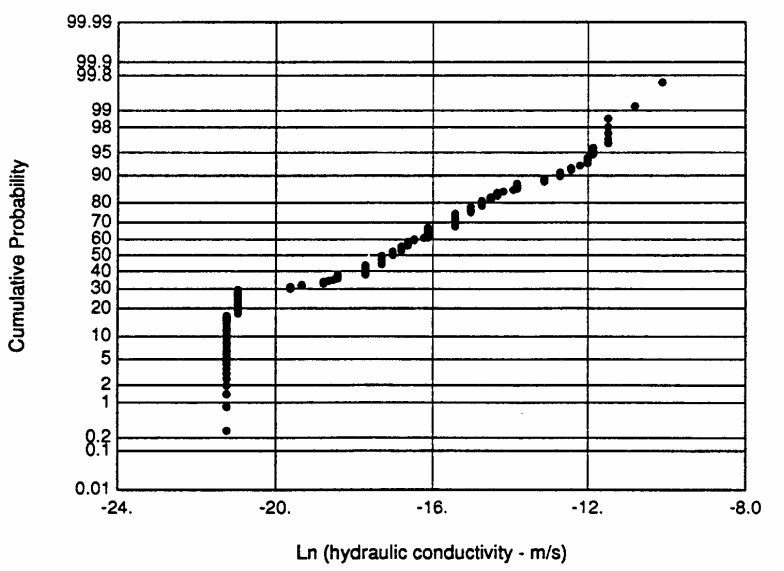


Figure 11-37. Normal probability plot of natural log-hydraulic conductivity from 2.5 m packer tests.

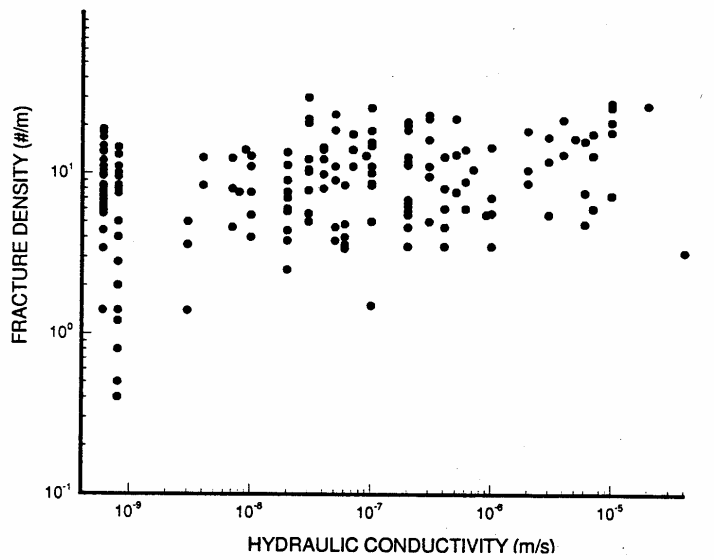


Figure 11-38. Scatter plot of fracture density (number per unit length of boreholes) versus hydraulic conductivity from 2.5 m scale packer tests.

Cacas *et al.* (1990a) generated 17 random networks of fractures within 10 m scale cubes. No spatial correlation between the fractures was considered even though the 2.5 m scale hydraulic conductivities (to which the fracture statistics were deemed related) are strongly auto-correlated, as we show later. By computing steady state flow across each cube under an externally imposed hydraulic gradient (and noting, as should have been anticipated by construction, that cube orientation has little effect on the results) the authors had obtained 17 scalar equivalent hydraulic conductivity values K_e that they considered representative of 10-m scale rock volumes.

The K_e values were treated as a sample from a statistically homogeneous and isotropic multivariate log-normal random field across the experimental site. An approximate three-dimensional stochastic continuum formula was then used to calculate an effective scalar hydraulic conductivity $K_{eff} \approx K_g\left(1+\sigma^2/6\right) \approx 1.6\times10^{-8}$ m/s for the rock mass on scales much larger than 10 m. Next, a circular finite element model was constructed on a plane transverse to the drift. Pressure head in the drift was set equal to zero and head was prescribed along 7 concentric circular boundaries with radii increasing from 5.5 m to 47.5 m, based on borehole measurements projected onto the model plane (Figure 11-39). For each of the 7 boundaries, the two-dimensional model was calibrated against measured flow into the drift (it is not clear how well did the calibrated models reproduce heads projected onto the interior of the model plane, a question we address later in the context of our own model). The calibrations yielded 7 equivalent uniform scalar hydraulic conductivity estimates that tended to decrease with domain size. Even though the estimates had not fully stabilized at the maximum radius of 47.5 m, their smallest value of 1.8×10^{-8} m/s was taken to represent the large-scale hydraulic conductivity of the rock. That this estimate is within the range $1.5-2.3\times10^{-8}$ m/s of 17 K_e values generated by means of the three-dimensional discrete network model, and is close to the stochastic average $K_{eff} \approx 1.6\times10^{-8}$ m/s of these values, was considered by Cacas *et al.* (1990a) to constitute a validation of their fracture network model with respect to flow.

To simulate tracer transport, Cacas *et al.* (1990b) generated 20 random three-dimensional networks of one-dimensional channels interconnected at 10,000 nodes. The network covered part of the domain, which was subjected to prescribed head boundary conditions (Figure 11-40) that do not appear to be entirely consistent with those observed (Figure 11-39). Tracer was advected through each network by particle tracking. Local-scale dispersion was simulated by routing particles randomly to one of several downstream channels. Large-scale dispersion was simulated by drawing random particle residence times in channels from a normal distribution with variance proportional to an (apparently arbitrary) dispersivity of 0.8 m. The mean residence time was made proportional to a retardation coefficient that was calibrated against peak arrival times at location A for tracers injected into 7F2 and 6F2, and location B for tracers injected into 6F3 and 4F3. As injection into 3F3 at a distance of 35 m from the drift would have required including 25,000 fractures in the model, it was not simulated. Measured and simulated mass recovery fractions did not match well and the paper includes no comparison between simulated and observed tracer breakthrough curves. Their finding that the model was able to reproduce observed breakthrough durations with some fidelity was taken by Cacas *et al.* to constitute a validation of their fracture network model with respect to solute transport.

Cacas *et al.* (1990b) stated that it would not be realistic to interpret the tracer experiments at Fanay-Augères without accounting for discrete channels in the model. In the view of Cacas *et al.* (1990a), continuum models are not capable of adequately interpreting small scale measurements; at best, they may be used at some distance where only average behavior is required. Judging by their papers, Cacas *et al.* considered the continuum approach to be invalid at Fanay-Augères on scales smaller than 10 m.

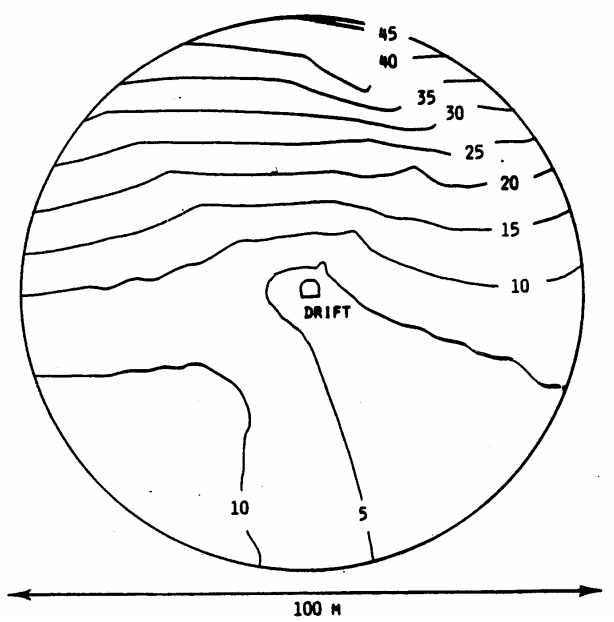


Figure 11-39. Projection of observed piezometric heads (in meters) onto transverse model plane (after Cacas *et al.*, 1990a). With permission, AGU.

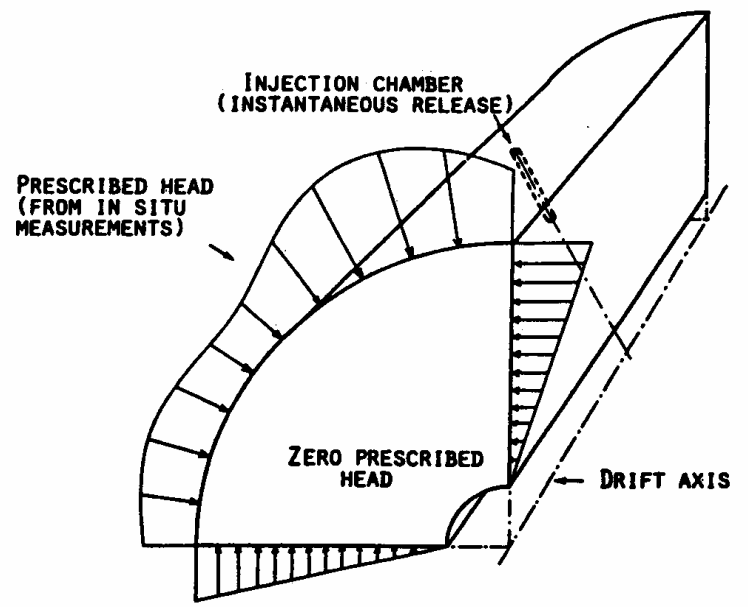


Figure 11-40. Boundary conditions used for simulating tracer injections (after Cacas *et al.*, 1990b). With permission, AGU.

We obtained data relating to the flow and transport experiments at Fanay-Augères courtesy of Professor G. de Marsily of the University of Paris in France. In the present paper, we utilize these data to model flow and transport at Fanay-Augères by viewing the fractured rock as a stochastic continuum on the nominal 2.5 m support (measurement) scale of available packer test data. This approach, originally proposed by

Neuman (1987, 1988), is based on the recognition that (a) it is expensive and difficult to characterize the geometry, hydraulic and transport properties of individual fractures on such small scales across a site, (b) fracture geometry tends to be a poor indicator of how such small scale hydraulic and transport properties vary across a site, and therefore (c) it is neither feasible nor necessary to collect detailed information about individual fractures on such small scales for the purpose of analyzing flow and transport on similar or larger scales. Instead, flow and transport in many fractured rock environments are amenable to analysis by continuum models that account adequately for medium heterogeneity. This includes the stochastic continuum approach, which treats in situ measurements of hydraulic conductivity (and possibly other rock properties) as a correlated random field and obviates the need for either detailed information about fracture geometry or assumptions about how individual fractures control flow and transport. The concept, and methods of analysis based upon it, have been applied successfully to fractured granites by Neuman and Depner (1988) and to unsaturated fractured tuffs by Chen *et al.* (2000), Vesselinov *et al.* (2001) and Hyun *et al.* (2002).

When flow or transport are affected by a few dominant features (fracture zones, faults or dykes) on relatively large scales, one has the option of embedding them as discrete slabs within the model and treating the internal properties of each feature as spatially auto-correlated random fields (Neuman, 1997). Though a fault zone has been identified at the Fanay-Augères experimental site (Figure 1), neither Cacas *et al.* (1990a-b) nor we have found it necessary to include this fault in our models. In fact, we show that it is possible to reproduce a larger selection of experimental results than those considered by Cacas *et al.* merely upon modeling the rock as a statistically homogeneous continuum in two dimensions. Our results demonstrate that a continuum approach may be well suited for the analysis of flow and transport in fractured rock. This does not constitute a validation of the continuum approach, just as the results of Cacas *et al.* fall short of validating the discrete fracture approach. Instead, the two sets of results illustrate jointly the well established principle that an open system, especially one as complex as fractured hydrogeologic environments tend to be, cannot be described uniquely on the basis of sparse data and need not be described in great detail to capture its salient behavior by a model.

The following is an abbreviation and synthesis of material contained in the M.S. theses of Kostner (1993) and Ando (1995). Readers interested in details beyond those presented here are encouraged to consult these theses.

11.3.2 Indicator Analysis and Simulation of Hydraulic Conductivities

We start by subjecting hydraulic conductivities (K) derived at Fanay-Augères from packer tests in 2.5 m borehole intervals to indicator geostatistical analysis. Such analysis treats the data as a sample from a spatially auto-correlated random field defined on a continuum. We use indicator rather than standard geostatistical analysis to account for the spatial distribution of hydraulic conductivities below the detection limit (Figure 11-37), which may have considerable impact on the movement of tracers by forcing it to migrate around areas of low permeability.

The first step in the analysis is to transform $K(\mathbf{x})$ into a binary indicator function $I(\mathbf{x};K_c)$ equal to 1 if $K(\mathbf{x}) \le K_c$ and 0 if $K(\mathbf{x}) > K_c$, where K_c is a specified indicator (cutoff) value and **x** is a vector of coordinates. We subdivide the available range of *K* data into 7 classes defined by 6 indicator cutoffs, listed in Table 11-10 together with the percent of *K* data that that lie at or below each cutoff. The first cutoff is the detection limit and the third is the median of the data. Indicator analysis requires no distributional assumptions about *K*. Instead the expectation of $I(\mathbf{x};K_c)$, conditional on a set of measurements $\boldsymbol{K}^T = (K_1, K_2, ...K_N)$ where $\boldsymbol{K}$ is a vector and *T* denotes transpose, is an estimate of the conditional cumulative probability distribution of $K(\mathbf{x})$,

$$E\left[I(\mathbf{x};K_c) \middle| \boldsymbol{K}^T \right] = P\left[K(\mathbf{x}) \le K_c \middle| \boldsymbol{K}^T \right].$$

Figure 11-37 shows that if one ignores values of $Y = \ln K$ at the detection limit, the remaining cumulative distribution is near-normal with slight positive skewness. The indicator approach makes it possible to extrapolate this distribution below the first and above the sixth cutoffs with the aid of suitable parabolic and hyperbolic models, respectively. The augmented distribution yields a variable probability of encountering a hydraulic conductivity below the detection limit and a finite probability of encountering one above the highest value measured. It increases slowly above the sixth cutoff and thus preserves the statistical properties of the data including their slight positive skewness.

The next step is to infer from the indicator-transformed data sample and theoretical (model) indicator variograms. Variogram models for all the cutoffs fit isotropic spherical models with ranges (measures of spatial auto-correlation) that vary from 18 m to 23 m. Variogram models for cutoffs 1 – 5 have sills (total variances) varying form 0.16 to 0.25 and nuggets (variances of white background noise) varying from 0.10 to 0.16. The variogram of cutoff 6 has much smaller sill (0.05) and nugget (0.02) values which, however, are based on a relatively small sample (Table 11-10) and are therefore less reliable. Normalizing all 6 variograms with respect to the sill of the median variogram shows that they are quite similar (Figure 11-41). As the sill has no effect on indicator kriging weights, we calculate them on the basis of a single omni-directional variogram corresponding to the median cutoff (Figure 11-42).

As noted by Cacas *et al.* (1990a), the spatial distribution of near steady state heads in the three parallel planes that contain the boreholes are sufficiently similar to justify modeling flow into the drift in two dimensions. We do so by projecting all the available *K* and head data onto a single plane transverse to the drift.

Indicator kriging yields a best linear unbiased estimate of the conditional probability $P\left[K(\mathbf{x}) \le K_c \middle| \boldsymbol{K}^T \right]$ for each cutoff (Gómez-Hernández and Srivastava, 1990). We estimate the latter on a 44 × 47 grid of 2 m × 2 m cells in a plane transverse to the drift using an algorithm found in Deutsch and Journel (1998). For example, Figure 11-43 shows how our estimate of the probability that hydraulic conductivity exceeds the fifth cutoff varies across the plane. One can use the results to map out estimates of the conditional mean of *K* or other statistical moments such as median or mode.

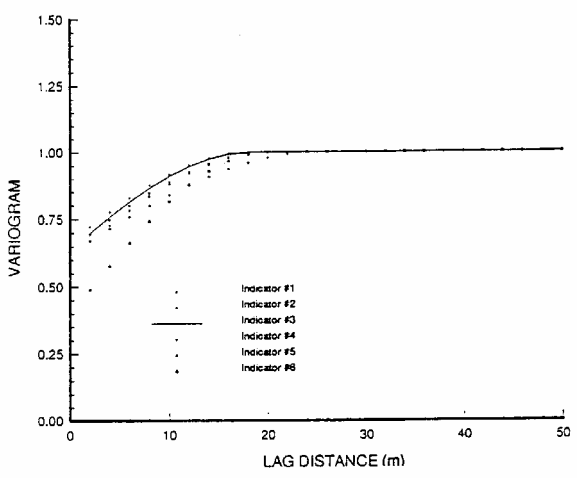


Figure 11-41. Omni-directional indicator variogram models of $Y = \ln K$ normalized with respect to sill of variogram model corresponding to indicator no. 3.

Table 11-10. Indicator cutoffs of hydraulic conductivity data and percent of data not exceeding each cutoff.

Indicator	Value of K (m/s)	Percent of data
1	8.0E-10	27
2	1.3E-08	40
3	4.0E-08	53
4	2.0E-07	68
5	4.0E-07	81
6	6.0E-06	95

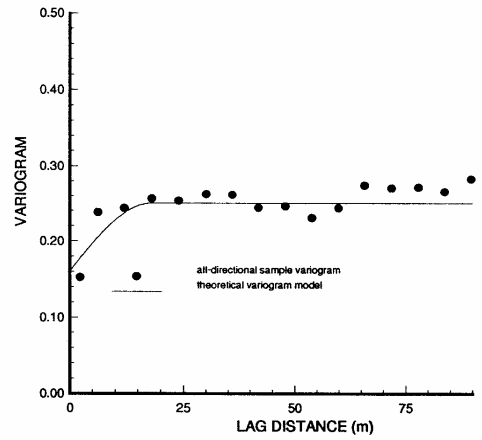


Figure 11-42. Omni-directional sample and spherical model semivariogram for indicator no. 3.

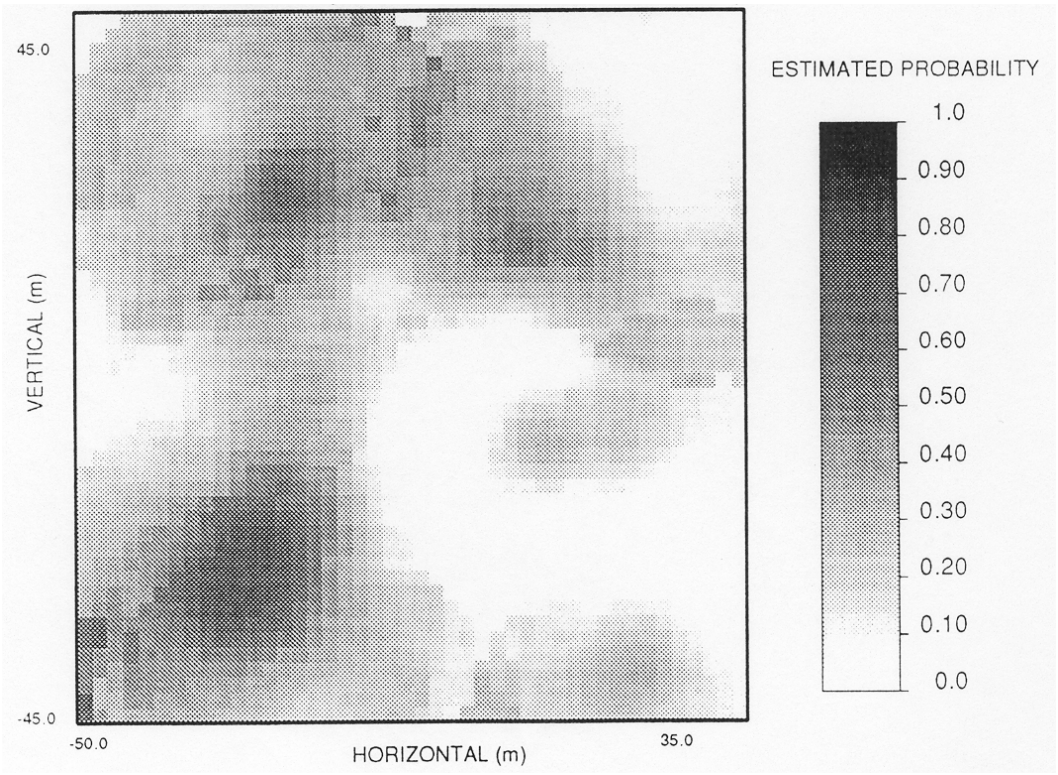


Figure 11-43. Kriged estimate of probability that hydraulic conductivity exceeds the fifth cutoff.

Finally, we use conditional Monte Carlo simulation to generate 100 stochastic realizations of the hydraulic conductivity field on the transverse plane. We do so by means of a sequential algorithm (Deutsch and Journel, 1998) that honors hard indicator data, hard indicator constraints, the augmented probability distribution of the data, and the indicator variogram of the median cutoff. The simulations are performed on a rectangular grid of 34×36 cells measuring 2.5 m $\times$ 2.5 m corresponding to the nominal support scale of the packer test data. Both the mean and the variance of the generated fields stabilize after only about eighty realizations. Two of the realizations are depicted in Figure 11-44. Whereas the sampled data do not fall below the detection limit, many of the simulated values do. Their histograms and probability plots (illustrated for the above two realizations in Figure 11-45) show bimodal distributions and a slight positive skewness as do the sampled data. The only notable difference between the original and simulated histograms is the wider spread (tail) of the latter below the detection limit. The generated hydraulic conductivity data under the detection limit are log-normally distributed.

11.3.3 Geostatistical Analysis and Simulation of Flow

To simulate flow and transport at Fanay-Augères we limit ourselves to the saturated zone. We work with pressure heads rather than hydraulic heads to allow easy identification of the water table as a zero-pressure isobar. Near steady state pressure head data collected during the first and second stages of the experiment, prior to and following a transitional nonsteady state period (Figure 11-34), are represented well

by omni-directional residual semivariograms with second-order polynomial drifts in the transverse (vertical) plane (depicted for stage 1 in Figure 11-46). We use them to estimate pressure head on the above grid of 34×36, 2.5 m $\times$ 2.5 m cells by kriging, and to identify the position of the water table at each stage (as shown for stage 1 in Figure 11-47). Kriging estimation errors (represented by their standard deviations) are seen to increase with distance from the borehole monitoring intervals.

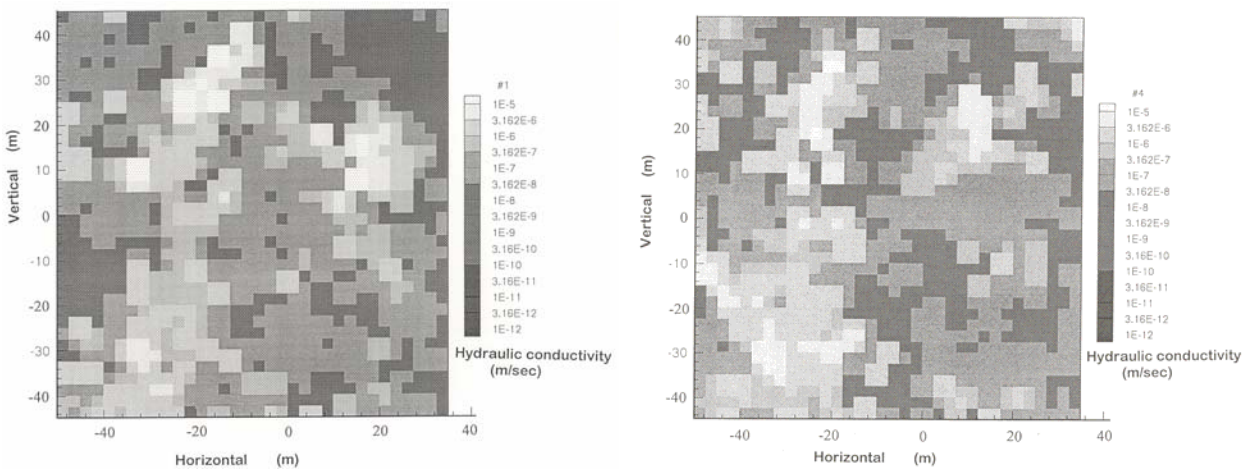


Figure 11-44. Two random realizations of hydraulic conductivity honoring the data.

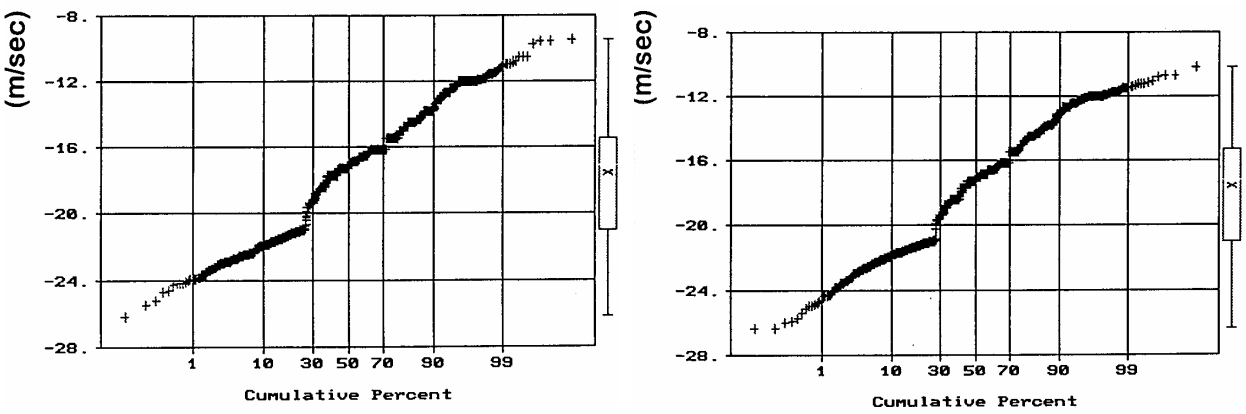


Figure 11-45. Cumulative distribution of hydraulic conductivity estimates in Figure 11-44.

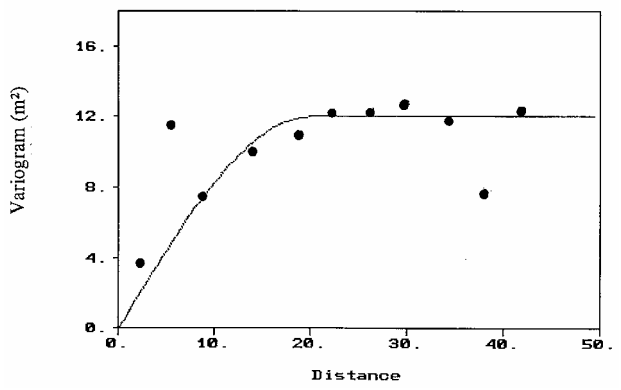


Figure 11-46. Sample (dots) and spherical model (solid curve) omni-directional variogram of pressure head at stage 1 with quadratic polynomial drift.

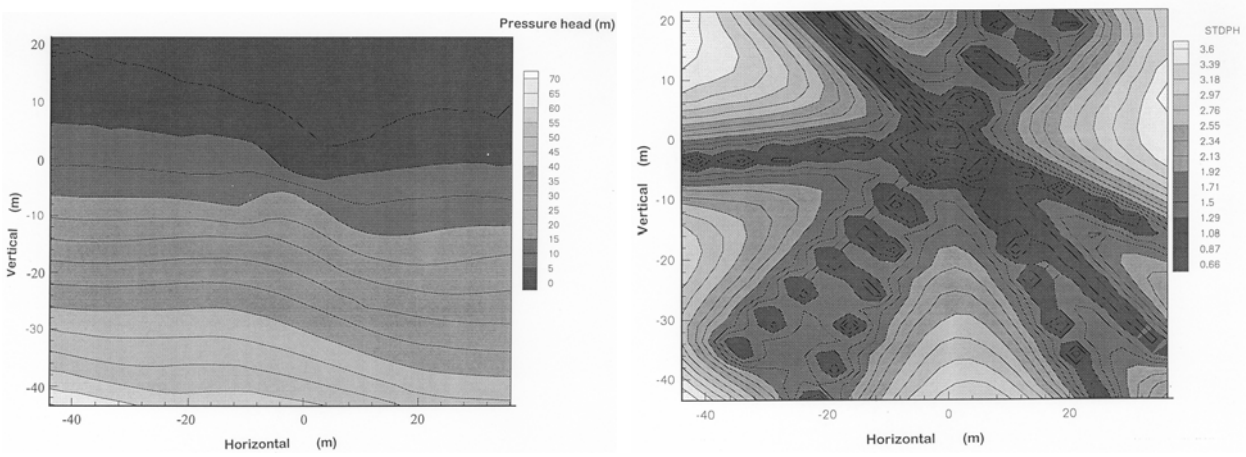


Figure 11-47. Kriged pressure head at stage 1 (left) and standard deviations of corresponding kriging errors (right).

Next we prescribe zero pressure head along the water table and the drift, and nonzero estimated pressure heads along the remaining boundaries of the saturated zone (Figure 11-48). We then simulate flow through this zone in each of the 100 generated permeability fields at each stage using the finite element code SUTRA (Voss 1984, 1990). Pressure and total heads corresponding to one of these simulations (no. 48 at stage 1) are depicted in Figure 11-49.

Figure 11-50 shows histograms of simulated inflow rates into the 100 meter long drift at stages 1 and 2. Corresponding statistics are given in Table 11-11. Cumulative frequency plots (not shown) suggest that the rates are close to being log-normally distributed with geometric mean values of 2.09 l/min and 1.56 l/min, respectively. The measured inflow rates of 2.42 l/min (September 1985) and 1.48 l/min (December 1985) during the two stages correspond closely to the most frequently simulated values. Simulations using geometric mean hydraulic conductivities over all 100 realizations at each stage (which vary from grid cell to grid cell) yield inflow rates of 2.14 l/min and 1.35 l/min, respectively. These are very close to the measured rates.

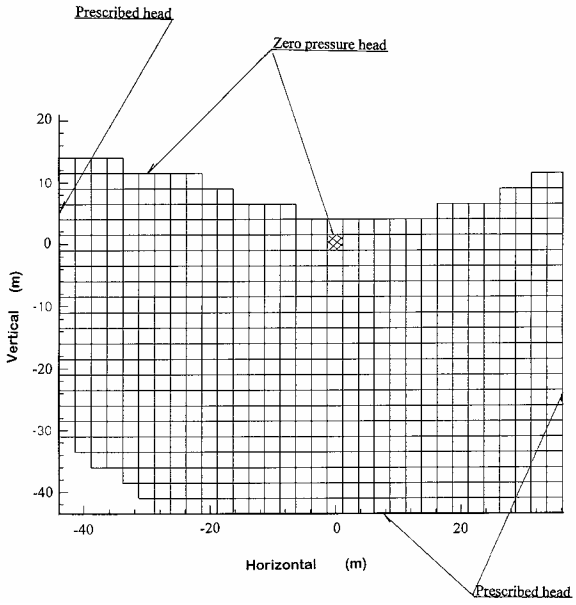


Figure 11-48. Flow model.

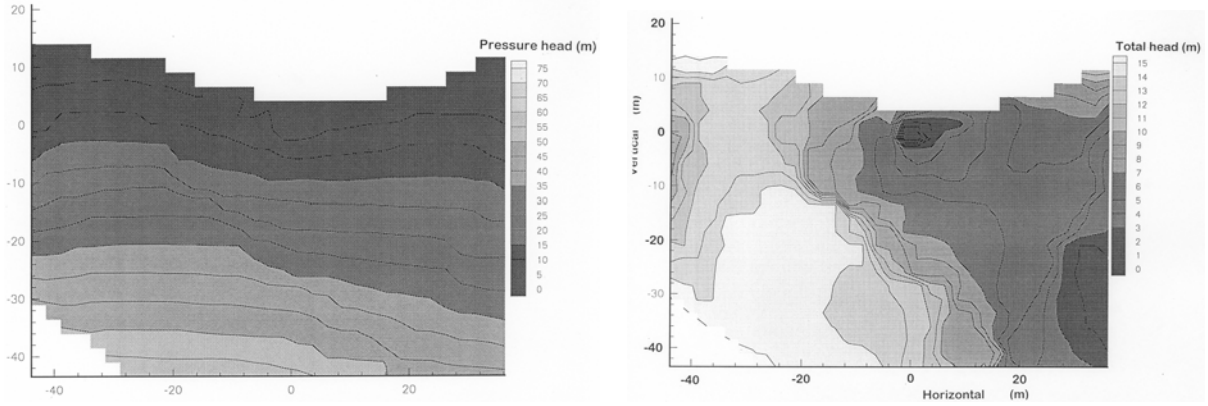


Figure 11-49. Conditional simulation no. 48 of pressure head (left) and total head (right) at stage 1.

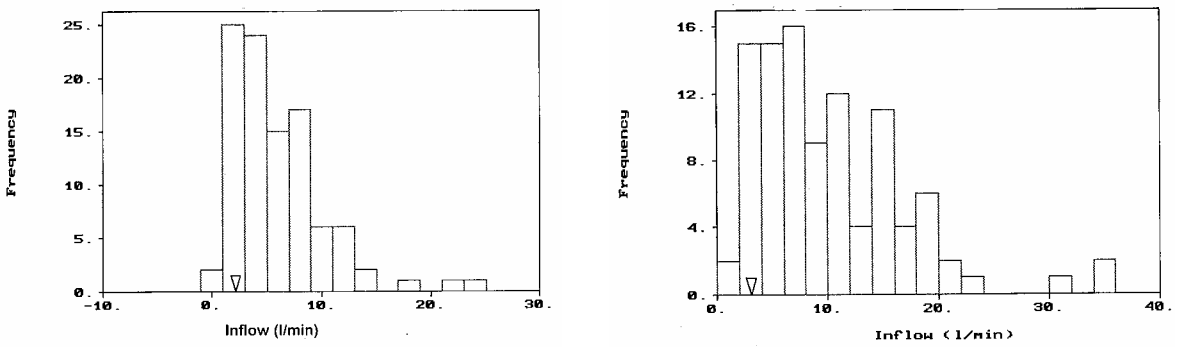


Figure 11-50. Histograms of simulated drift inflow rates at stages 1 (left) and 2 (right). ∇ indicates measured value.

Table 11-11. Statistics of simulated inflow rates into drift (in l/min).

	Stage-1	Stage-2
Mean of 100 realizations	10.09	5.97
Maximum of 100 realizations	34.49	23.18
Minimum of 100 realizations	1.57	0.73
Median of 100 realizations	8.25	4.90
Mode of 100 realizations	2 to 8	1 to 5
Using geometric mean hydraulic conductivities	2.14	1.35
Measured	2.42	1.48

Tables 11-12 lists statistics of differences between simulated and kriged pressure head values at stages 1 and 2 for five hydraulic conductivity realizations (nos. 15, 41, 48, 72, 95). Table 11-13 lists corresponding statistics of differences between simulated and measured pressure heads. The five simulations yield inflow rates into the drift, and pressure head residuals, that are closest to those actually measured. Among these five simulations, no. 48 yields the best fit to measured drift inflow rates at both stages (Figure 11-51). Corresponding scattergrams of simulated versus measured head residuals in Figure 11-52 show an acceptable fit (we do not know how well the model of Cacas *et al*. (1990a) reproduces observed heads in the interior of the flow domain). We therefore adopt conductivity realization no. 48 for the purpose of simulating transport.

Table 11-12. Statistics of pressure head residuals
(differences between simulated and kriged values in meters).

Realization	#15		#41		#48		#72		#95	
Stage	1	2	1	2	1	2	1	2	1	2
Mean	0.77	0.95	1.19	1.26	1.30	1.29	1.251.	1.28	1.19	1.19
Variance	2.39	1.78	3.09	2.42	2.72	2.13	1.89	1.58	3.17	2.25
Max	3.73	3.80	5.07	5.45	5.09	4.32	5.587	4.77	4.54	4.05
Min	-4.90	-3.34	-4.91	-3.81	-4.70	-3.49	-2.52	-2.23	-4.28	-2.64

Table 11-13. Statistics of pressure head residuals
(differences between simulated and measured values in meters).

Realization	#15		#18		#48		#72		#95	
Stage	1	2	1	2	1	2	1	2	1	2
Mean	1.00	1.30	1.00	1.27	1.00	1.38	1.55	1.71	1.60	1.68
Variance	4.70	2.95	6.38	2.09	6.83	1.96	1.55	1.72	5.13	3.44
Max	3.70	3.65	4.67	5.40	4.77	4.12	4.30	3.78	4.48	4.18
Min	-6.52	-4.62	-9.07	-6.82	-6.89	-4.82	-2.96	-1.72	-5.26	-3.52

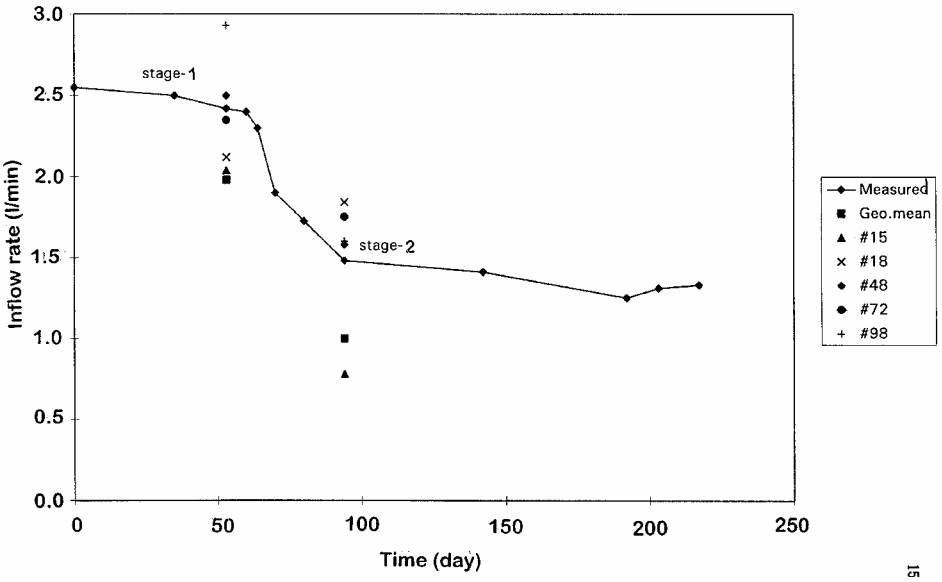


Figure 11-51. Measured and simulated flow rate into drift.

11.3.4 Simulation of Transport

For the simulation of transport we select a single realization (no. 48) that yields the best reproduction of inflow rates into the drift and a reasonable reproduction of measured heads at both stages of the analysis. We focus on four tracer tests during which injection took place into chambers 3F3, 4F3, 6F3 and 7F2. According to Cacas *et al.* (1990b), these four tests are the most reliable and offer the most complete database among all tracer tests at the site. Since their model was limited to a relatively small rock volume, Cacas *et al.* were unable to simulate test 3F3, analyzing instead test 6F2. Our model has no such limitation (mainly because we do not consider individual fractures, and to a lesser extent because we limit our analysis to two dimensions) and we therefore analyze test 3F3. We add that injection during test 6F2 took place above the water table, a fact not explicitly considered by Cacas *et al.*

The periods during which breakthrough of the four tracers injected into 3F3, 4F3, 6F3 and 7F2 were monitored are indicated by bars in Figure 2. Injection took place for 30 minutes at the start of each monitoring period. To mimic the transient stage during tracer tests 4F3 and 7F2, we interpolate linearly over time between the preceding steady state head configuration of stage 1 and the subsequent steady state head configuration of stage 2.

We use SUTRA and the two-dimensional grid in Figure11-48 to simulate each of the four tracer tests. Neglecting the third dimension in our model means that we do not allow the tracer to advect or disperse perpendicular to our planar flow domain. For example, the distance from 7F2 to observation site B is 5m in the two dimensional section, but in reality varies from 5m to 25.5m with distance along B, implying that transport at the site is three-dimensional. We however achieve sufficiently accurate results for our purposes with a two-dimensional model.

Maloszewski and Zuber (1993) estimated porosity indirectly to range between 1.25×10^{-4} and 1.10×10^{-3}. We estimate

effective porosity and longitudinal as well as transverse dispersivities in three steps:

1. Select dispersivities and modify effective porosity to fit peak arrival times (all tracers were conservative).
2. Modify longitudinal and transverse dispersivities to fit observed shapes of breakthrough curves.
3. Repeat steps 1 and 2 until a satisfactory fit to the data is obtained.

Figure 11-53 compares computed (curves) and measured (solid squares) breakthrough curves, normalized with respect to maximum concentration, for tracer tests 6F3, 4F3, 3F3 and 7F2, respectively. Concentration data for 4F3 and 3F3 were derived by conversion from fluorescence data. The corresponding best-fit effective porosities are 8.0×10^{-3}, 8.0×10^{-4}, 6.0×10^{-4} and 4.0×10^{-4}, all but the first of which are within the range proposed by Maloszewski and Zuber (1993). The large effective porosity in 6F3 suggests that transport may be taking place through a preferential path of high permeability, such as a channel or wide fracture. Longitudinal dispersivity is 17.5 m and transverse dispersivity is one tenth this value in all four cases. We attribute the high dispersivities in part to mixing within the drift and to neglect of the third dimension in our model.

Cacas *et al.* (1990b) did not provide simulated breakthrough curves. We therefore compare our simulated peak arrival times with those measured (Figure 11-54) and computed by Cacas *et al.* (Table 11-14). It appears that our two-dimensional stochastic continuum model reproduces observed behavior at least as well as does the three-dimensional stochastic discrete network model of these authors, if not better.

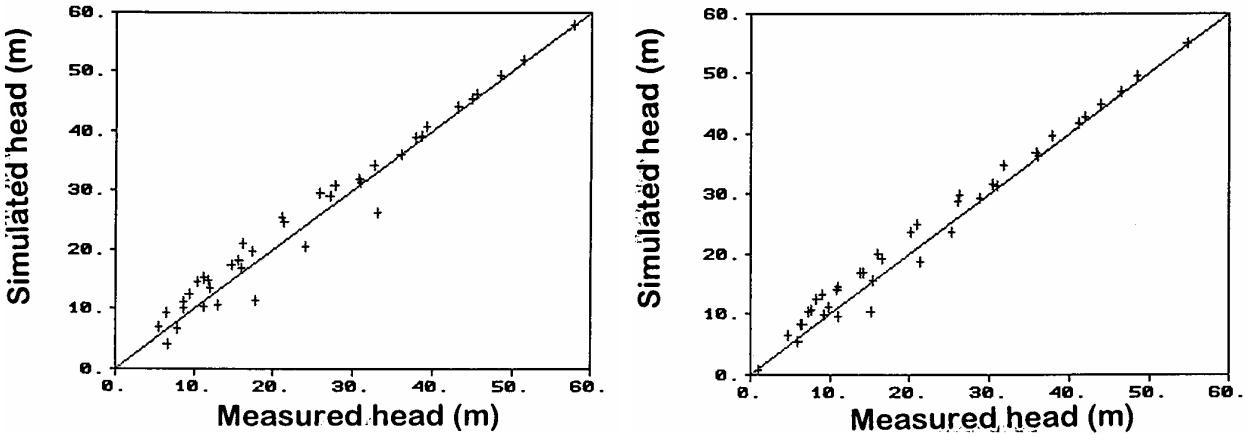


Figure 11-52. Scattergrams of simulated (using conductivity realization no. 48) versus measured heads at stages 1 (left) and 2 (right).

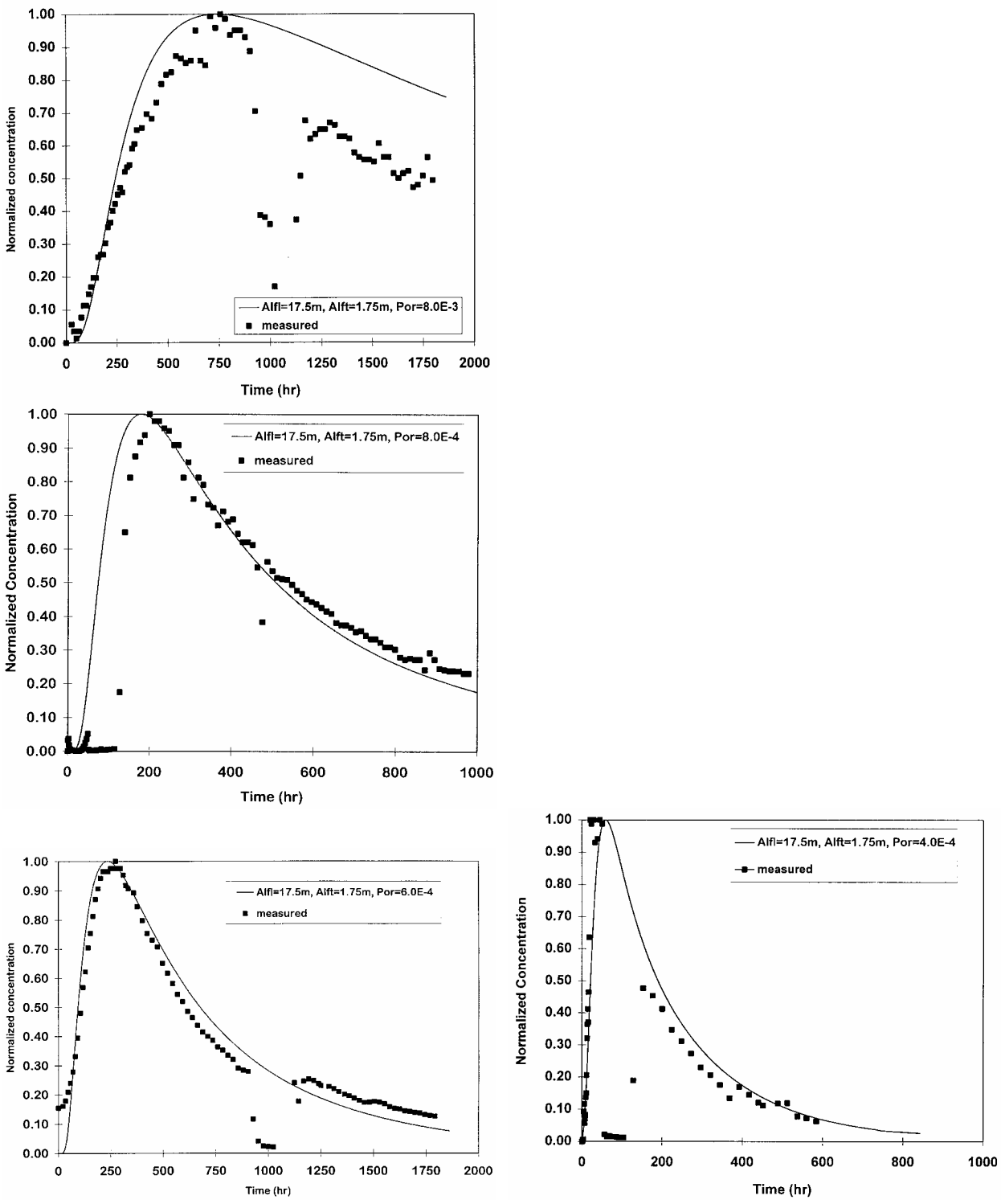


Figure 11-53. Computed versus measured concentrations normalized by maximum for tests 6F3 (upper left), 4F3 (upper right), 3F3 (lower left), and 7F2 (lower right). Por = effective porosity, Alfl = longitudinal dispersivity, Alft = transverse dispersivity.

Table 11-14. Simulated and measured tracer arrival times.

Tracer Number	Simulated peak arrival time (hr)	Measured peak arrival time (hr)	Peak arrival time from Cacas et al. (hr)
6F3	748	757	1 to 900
4F3	179	199	27 to 14900
3F3	236	274	out of computation

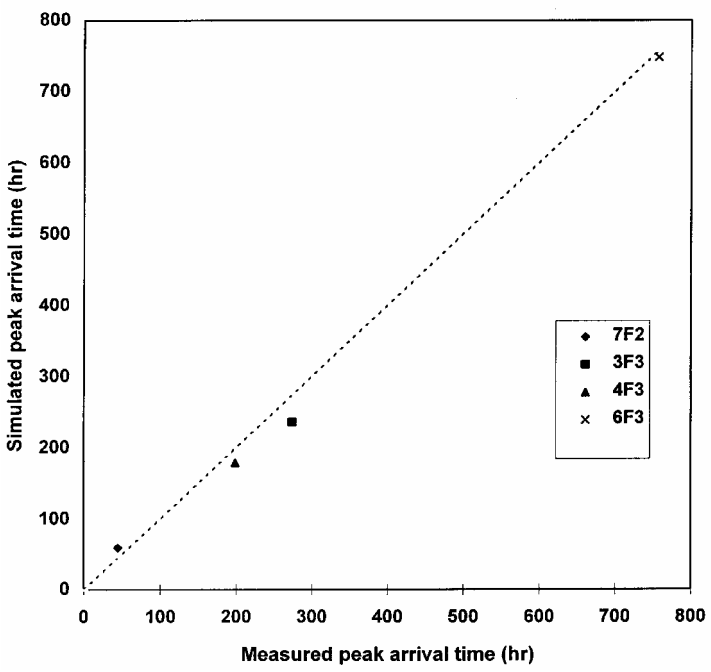


Figure 11-54. Simulated versus measured peak arrival times.

It might be possible to improve model fit further by conditioning a particular random conductivity field not only on measured conductivities as we have done but also on measured drift inflow rates and heads (Gómez-Hernández *et al.*, 1997, 2000) as well as tracer concentration data (Gómez-Hernández *et al.*, 2002) via an automated inverse procedure. We deem this unnecessary because the much simpler approach we use yields simulations of flow and transport that are sufficiently accurate for our purpose of juxtaposing the stochastic discrete network and continuum approaches.

11.3.5 Conclusions

1. Our work confirms the suitability of stochastic continuum modeling for the analysis of flow and transport in fractured crystalline rocks at Fanay-Augères in France. Claims in the literature that only discrete fracture network models are suitable for this purpose are unfounded.
2. Stochastic continuum conceptualization avoids the need for detailed information about fracture geometry and assumptions about the nature of flow within individual fractures. Embedding such information and assumptions in a discrete fracture network model has proven not to have any advantage over

the stochastic continuum approach in the case considered here.

Our results do not constitute a validation of the continuum approach just as those of Cacas *et al.* (1990a-b) fall short of validating the discrete fracture approach. Instead, the two sets of results illustrate jointly the well established principle that an open system, especially one as complex as fractured hydrogeologic environments tend to be, cannot be described uniquely on the basis of sparse data and need not be described in great detail to capture its salient behavior by a model.

12 GUIDELINES FOR REVIEW OF HYDROGEOLOGIC MODELING AND UNCERTAINTY ANALYSES

12.1 Introduction

The following guidelines for review of hydrogeologic modeling and uncertainty analyses are intended to help insure that documents reviewed conform in a broad way to the methodology described in this report. This means that such documents should reflect a systematic and comprehensive approach to hydrogeologic conceptualization, model development and predictive uncertainty analysis. The guidelines cover all stages of model building and uncertainty analysis. Model building stages include regional and site characterization, hydrogeologic conceptualization, development of conceptual-mathematical model structure, parameter estimation on the basis of monitored system behavior, and assessment of predictive uncertainty. Uncertainty analysis includes assessment of estimation and predictive errors arising from the conceptual framework that determines model structure, the definition of model parameters, spatial and temporal variations in hydrologic variables that are either not fully captured by the available data or not fully resolved by the model, and the scaling behavior of hydrogeologic variables.

The guidelines are generic but designed to be of practical use to NRC staff in their review of decommissioning plans and performance assessment of high-level and low-level radioactive waste disposal sites as well as uranium recovery facilities. For this purpose, the guidelines are cast in the context of a framework that is relevant to NRC staff review and performance evaluation needs. The context is defined in terms of corresponding performance measures, hydrogeologic analyses that are needed to assess them, the desired reliability of such assessments, and the expenditure (in time, effort and money) that is allowed to achieve it.

12.2 Summary of Guidelines

The guidelines are summarized below and explained further in subsequent sections of this chapter. The guidelines are arranged in an iterative sequence of steps that correspond to our proposed strategy of model building and uncertainty assessment.

- Define the **hydrogeologic context** of the problem.
- Collect a **broad range of regional and site data** from public and private sources that are relevant to the hydrogeologic context of the problem. Provide a detailed rationale for your choice of data.
- Interpret the data in a way that leads to **several plausible conceptual models** (hypotheses and assumptions) of regional and site hydrogeology in three spatial dimensions. Provide a detailed rationale for each interpretation and explain why no other hydrogeologic interpretations of the data would be plausible.
- Verify in detail that none of the selected conceptual models are in **conflict with any of the data**; discard those that are.

- **Rank the conceptual models** of regional and site hydrogeology in the order of their plausibility in light of the data, regardless of context. Provide a detailed rationale for your ranking and discard models that appear much less plausible than those retained.
- Starting with the most plausible conceptual model and continuing down the ranks:
  - Develop a complete, coherent and internally consistent **conceptual-mathematical model of three-dimensional transient flow and transport** that is as simple as warranted by the context of the problem and the data. Provide a detailed rationale for your schematization of regional and/or site hydrogeology; choice of governing, state and constitutive equations; type and space-time distribution of forcing terms (sources, initial and boundary conditions); parameterization (functional form and number of parameters); and all other conceptual-mathematical aspects of the model. **Discard models that do not lend themselves to such mathematical description**.
  - If mathematical modeling in three spatial dimensions and/or in time is considered unnecessary or computationally infeasible, provide a detailed **explanation of how the three-dimensional transient conceptual-mathematical model reduces to a two-dimensional and/or steady state model**. In particular, demonstrate (preferably through a mathematical formalism and/or computationally) how the lower-dimensional model captures all relevant aspects of the original three-dimensional transient model. **Avoid one-dimensional models at this stage**.
  - Select a suitable **computational code or algorithm** to represent the conceptual-mathematical model.
  - Associate objective or subjective **prior statistics** with all input parameters and forcing terms that are considered to be uncertain. Provide a detailed rationale for considering any of these inputs to be known with certainty.
  - Test the conceptual-mathematical model for **qualitative consistency with the available data** through computational exploration and visualization in at least two spatial dimensions, within a plausible range of input parameters and forcing terms. If inconsistent, repeat the previous steps in modified form (by altering and/or refining the hydrogeologic scheme and mathematical description of flow and/or transport) or discard the underlying conceptual model and select an alternative.
  - Use an inverse (preferably Maximum Likelihood) approach to **calibrate the computational model against observational (monitoring) data**. This should yield **posterior estimates of the model parameters**, **statistical measures of their estimation errors** and associated model quality criteria. If the model cannot be properly calibrated against reliable data, repeat the previous steps in modified form (by altering and/or refining the hydrogeologic scheme and mathematical description of flow and/or transport) or discard the underlying conceptual model and select an alternative.
  - If feasible, **confirm aspects of the model** by predicting system behavior under conditions other than those used for calibration and comparing with observed system behavior. **Quantify model predictive uncertainty** and check the extent to which the observations lie within corresponding uncertainty bounds. If not, repeat the previous steps in modified form (by altering and/or refining the hydrogeologic scheme and mathematical description of flow and/or transport) or discard the underlying conceptual model and select an alternative.

- Compare all retained calibrated models on the basis of suitable model quality criteria. **Discard models that are clearly inferior** to others.
- **Use each retained calibrated model to predict system behavior** as well as relevant hydrogeologic performance measures under various scenarios of relevance to the problem, and **quantify the corresponding predictive uncertainty**.
- **Use Maximum Likelihood Bayesian Model Averaging to render joint predictions by all the retained calibrated models, and to assess their joint predictive uncertainty**. Acknowledge the conditional nature of the predictions and their corresponding uncertainty measures.
- If none of the models are deemed acceptable, perform a **sensitivity analysis to help identify the type and quantity of additional site data** that might materially enhance their reliability and credibility. Decide based on the potential benefit (in terms of bias and uncertainty reduction) and cost (in terms of time, effort and money) of such data whether or not to collect them and how.
- If and when new data of significant weight are obtained, **repeat the iterative process till its cost-benefit ratio reaches a value that is not considered worth exceeding**.

The following provides additional detail concerning each guideline.

12.2.1 Define Hydrogeologic Context

The first step in model development is to define its context and purpose. Within the context of NRC staff criteria (see Appendix A), the purpose of hydrogeologic site models is to help one analyze, qualitatively and quantitatively, subsurface flow and transport at a site in a way that is useful for review of decommissioning plans and performance of high-level and low-level radioactive waste disposal sites and uranium recovery facilities. For this purpose, the strategy developed in this report is cast in the context of a framework that is useful to NRC staff review and performance evaluation needs. The framework is defined in terms of performance measures identified by the NRC staff (see Appendix A).

The contextual framework defines key questions to which groundwater flow and transport models are expected to provide answers; it helps narrow down the problem as well as the range and type of potential model applications. Each reviewed document should provide clear answers to the following questions:

- For what ***purposes*** will the model be used? The answer should include defining the nature and magnitude of an existing and/or potential problem, location of the problem area, its causes, potential short- and long-term solutions and/or remedies, their anticipated consequences and costs-benefits, issues needing resolution, and criteria by which a solution and/or remedy will be selected.
- To what ***hydrogeologic system*** will the models be applied? The answer should include defining the hydrogeologic environment that is or may potentially be affected by the problem and/or its solution or remedy, the corresponding site and surrounding environs of (potential) concern, and the regional hydrogeologic setting.
- Under what ***circumstances and scenarios*** will the modeled hydrogeologic system operate?

This requires defining predevelopment, current and potential undisturbed and disturbed site and regional conditions, and corresponding natural as well as anthropogenic influences, under which flow and transport would be modeled.

- What ***measures*** will be adopted to assess ***performance*** of the hydrogeologic system? This requires specifying measures and/or criteria related to site and regional hydrogeology that would be used to identify issues, potential solutions and/or remedies, their cost-efficiency and ability to meet regulatory requirements.
- What aspects of site and regional ***hydrogeology*** and ***flow/transport dynamics*** are expected to ***impact*** these ***performance measures***, and how? The answer involves defining key elements of the hydrogeologic system, key flow and transport mechanisms, key natural and anthropogenic influences, and space-time scales that may potentially impact these measures, as well as the manners in which such impacts might occur.
- With what ***reliability***, ***certainty*** and ***accuracy*** does one need to predict ***performance measures***? ***How important*** a role will such measures play in the decision process? What is the ***worth*** (in time/effort/resources) ***of assessing performance*** measures for the site ***at specified levels of reliability/certainty/accuracy***? These are **the most basic questions** that must be addressed before any site investigations and modeling efforts are initiated, as the answer would have a major impact on the time, effort and resources that could validly be expended in pursuit of performance assessment for the site.

Performance measures are typically articulated within the broader context of a system in which groundwater is only one among several components. However, our strategy focuses solely on groundwater aspects of these broader criteria. Groundwater-related performance measures that need to be assessed by means of hydrogeologic flow and transport models are listed in Chapter 3. All of them require developing a good understanding, and reliable models, of subsurface flow and transport. One cannot do so without recognizing and considering the full complexity of a site before attempting to represent it by means of a simplified conceptual-mathematical model. This is true regardless of the specific performance criteria one needs to address.

The process of model simplification or abstraction must consider the degree of reliability, certainty and accuracy with which given performance measures need to be predicted, their importance in the decision process, and the amount of time, effort and resources that assessing them at the specified levels of reliability, certainty and accuracy would justify in each specific case. These questions involve regulatory issues which might be best addressed by hydrologists through case-by-case dialogues with regulators. As such, they are not addressed by the strategy in this report. Instead, the strategy stipulates that these fundamental questions should be addressed and resolved through discussions with the NRC staff before any site investigation and modeling are initiated. This is so because the answers impact in a major way the degree to which one justifies the use of a simplified conceptual-mathematical model for what is ubiquitously a complex hydrogeologic system with three-dimensional, transient groundwater flow and contaminant transport.

Therefore, it is very important the hydrogeologists first articulate these processes in their naturally complex setting as best understood in light of available site data and the state of prevailing hydrogeologic knowledge prior to postulating a groundwater flow and transport for

performance at a given site. Even if performance assessment is ultimately conducted with the aid of highly simplified conceptual-mathematical models of groundwater flow and transport, the strategy deems it essential for the credibility of the assessment that these models derive objectively (and if possible formally) from a more complete description of site hydrogeology, based on clearly reasoned and properly defended arguments.

Rather than taking the attitude that a limited performance assessment goal justifies a limited view of hydrogeology, we propose that a comprehensive description of hydrogeology is required to properly adapt a hydrogeologic model to such a limited goal.

12.2.2 Collect a Broad Range of Regional and Site Data

Collect a broad range of regional and site data from public and private sources that are relevant to the hydrogeologic context of the problem. Provide a detailed rationale for your choice of data.

The guideline is based on the recognition that it is often possible to postulate hydrogeologic conceptual models or hypotheses for a site on the basis of a broad range of publicly available geologic and geographic information about its surroundings. Additional conceptualization can be done on the basis of generic data about similar regions and the properties of similar materials elsewhere. Several such regional and generic sources of information are identified and discussed in this report. Yet each site is unique and so virtually guaranteed to reveal additional features, properties and behaviors when characterized in some detail locally. Hence the strategy considers local characterization essential for the postulation of acceptably robust conceptual hydrogeologic models for a site. The broader is the available database, the more robust is the conceptualization.

Key data categories include site and regional physiography, topography, climate, meteorology, soils, vegetation, land use, geomorphology, geology, geophysics, surface and subsurface hydrology, inorganic and organic hydrochemistry, radiochemistry, natural and anthropogenic isotopes, remotely sensed data, *etc*. Of special relevance to hydrogeologic model development are regional and site data that allow one to define the distribution of hydrostratigraphic units on a variety of scales; their geologic structure; rock and soil types; their textural, physical, flow and transport properties; fluid types; their state of saturation, pressure, temperature and density; chemical constituents and isotopes; major contaminants in soil, rock and groundwater; and their sources.

Site characterization data form the foundation on which one postulates one or more conceptual-mathematical models for an area and assigns initial values to their input parameters (*i.e.*, material properties and forcing terms such as sources, initial and boundary conditions). To test and compare these models among themselves qualitatively and/or quantitatively, one also needs monitoring data that constitute observations of actual hydrologic behavior at and around the site. Only with such data can one evaluate the ability of models to mimic real system behavior (qualitatively at the conceptual level, quantitatively at the conceptual-mathematical level), improve their ability to do so through calibration against the monitoring data, determine their optimum degree of refinement or complexity, rank them and assess their cumulative impact on predictive uncertainty, and compare them with each other.

Site characterization and monitoring data are expensive and difficult to collect, leading to a ubiquitous scarcity of hard site information. It is therefore critically important to assess the role that such data play in rendering the hydrogeologic performance analysis credible. One cannot overemphasize the role of characterization and monitoring data in helping one identify and test alternative conceptual models, make rational choices among them, gauge and reduce model bias and uncertainty through proper model selection and calibration, assess the reliability of model predictions, and confirm the assessment through independent peer review as well as at least some degree of direct verification.

Regional and site characterization data tend to represent a wide range of measurement scales, not all of which are compatible with the intended scale of hydrogeologic model resolution. It is important to recognize this important issue explicitly and to make a conscious effort rendering the scale of measurement compatible with the scale of model resolution. This can be done by either rescaling the data to fit the scale of model resolution (which often entails averaging or upscaling over computational grid cells) or adapting model resolution to fit the scale of measurement (which often entails adapting the size of grid cells to the size of the data support).

12.2.3 Postulate Alternative Conceptual Models

Interpret available data in a way that leads to several plausible conceptual models (hypotheses and assumptions) of regional and site hydrogeology in three spatial dimensions. Provide a detailed rationale for each interpretation and explain why no other hydrogeologic interpretations of the data would be plausible.

A conceptual hydrogeologic model is a mental construct or hypothesis accompanied by verbal, pictorial, diagrammatic and/or tabular interpretations and representations of site hydrogeologic conditions as well as corresponding flow and transport dynamics. It should be presented in the form of written text accompanied by pictures, charts and diagrams (including maps, block-diagrams, cross-sections, panel diagrams, vertical/horizontal profiles) of system components, arrangements and relationships (flows) known collectively as structure. A conceptual model identifies relevant hydrogeologic units (soils, aquifers, aquitards, aquicludes, bedrock) and features (faults, intrusions, fractures), their makeup (mineralogy, petrography, texture, cementation, porosity, permeability, related petrophysical and geophysical properties), geometry (horizontal and vertical dimensions, boundary and internal configurations), system states under undisturbed and disturbed conditions (types of fluids that permeate the soils and rocks, their pressure, saturation, density and temperature, dissolved solutes and their concentrations), flow and transport dynamics and kinematics (active flow and transport processes; their driving mechanisms; fluid and solute fluxes and velocities; fluid, solute and energy balances), hydrochemistry and isotopes.

A hydrogeologic knowledge base typically includes information obtained from boreholes, soil and geologic outcrops, underground openings and geophysical surveys. This adds a depth dimension to geographic information about the site and its surroundings which is largely two-dimensional. A hydrogeologic knowledge base is thus inherently three-dimensional in space, and should allow one to describe the system in as many dimensions. This is important because hydrogeology is inherently three-dimensional (with many quantities depending additionally on

time), complex, and manifests itself on a multiplicity of scales. As such, it cannot be meaningfully captured in a single two-dimensional map or cross-section, nor can it be described adequately by a one-dimensional profile such as a stratigraphic column or a lithologic log, that show only selected details on a given scale. Instead, hydrogeologic data must be interpreted in three-dimensions and in some detail on a range of scales (from regional down to thin sections). There virtually always is at least some regional and site information that should allow one to do so; seldom can lack of data be validly quoted as a reason for doing less. This is true regardless of how complex or simple a set of models one ultimately employs for the assessment of site performance and the analysis of corresponding uncertainty.
A systematic approach to the qualitative conceptualization and screening of hydrogeologic hypotheses requires as a prerequisite the availability or acquisition of expertise in the qualitative and quantitative interpretation of hydrogeologic field data. With such expertise, and a healthy dose of hydrogeologic insight, one may proceed to conceptualize hydrogeologic units and features on the regional, site and subsite scales in terms of their hydrostratigraphy, lithology, vertical and horizontal boundaries, structural features such as folds, faults, offsets and intrusions, textural features such as grain size, cementation, microstructure and fracturing, flow and transport properties such as porosity, permeability, dispersivity and sorption coefficients, related pedological and petrophysical properties (spatial, directional and textural) as well as their distributions and variations (heterogeneity and anisotropy) on a range of scales, both between and within hydrogeologic units and relevant structural and textural features.

To help in this task, one must assemble or develop descriptions of geologic outcrops, well logs, air and satellite images, maps and cross-sections, panel diagrams, quantitative records of measured variables, and qualitative descriptions of observed phenomena. It is useful, but not necessary, to embed these in tools for the computer management of comprehensive hydrogeologic data of the kind discussed in Chapter 4.

The next step is to describe the space-time distributions of fluid types (water, air, nonaqueous phase liquids) and corresponding states such as saturation, pressure, temperature and density, followed by significant solutes and their concentrations, as well as the delineation of saturated, vadose and perched zones, all on a range of scales between and within hydrogeologic, structural and textural units and features. This in turn allows one to describe active and anticipated flow and transport phenomena such as advection (of solutes or particulates), convection (of heat), diffusion, dispersion and sorption, their modes (discrete features; single, dual or multiple continua) and scales of manifestation (regional, site, subsite, long-term, short-term) boundary and internal mechanisms, forces and sources that drive them (infiltration, evapotranspiration, recharge, discharge, pumping), as well as their relative intensities and significance.

The next stage is to delineate in a qualitative but internally consistent and coherent manner contours of equal hydraulic potential (head), pressure, saturation, density, temperature and solute concentration together with corresponding flowlines (streamlines, pathlines, streaklines), vectors of fluid and solute flux and velocity, and isochrones (of groundwater residence time and solute travel time). The potential for the development of fast flow paths, on scales smaller than are represented by these contours and vectors (focused and episodic infiltration, preferential wetting, high-permeability channels, instability of fluid fronts, fingering), must be articulated at this stage. One should also assess the overall balance of fluids, solutes and energy within the system.

To conceptualize contaminant transport on regional, site and subsite scales, one must describe the space-time distribution of major contaminants in the soil, vadose zone and groundwater on these scales. The description must include the space-time distributions of contaminant sources; mechanisms and rates of source contaminant mobilization and leaching; active transport phenomena such as average and fast advection, diffusion, dispersion, radioactive or biochemical decay, sorption, colloid transport; space-time distribution of migration, spreading, and dilution patterns; and overall mass and ionic balance for key contaminants.

A hydrogeologic conceptualization is not complete without a description of hydrogeochemistry and isotope hydrology on regional, site and subsite scales in three-dimensions. This includes the space-time distributions of major hydrochemical constituents, environmental isotopes and their sources above and below the water table; space-time distribution of groundwater ages; implications concerning flow between and within hydrogeologic, structural and textural units and features (including infiltration, evapotranspiration, recharge, discharge, directions and rates); and implications concerning transport (directions and velocities, possible compartmentalization, isolation and mixing of groundwater bodies, chemical reactions, and water-rock interactions).

It often helps to conceptualize temperature and heat flow on regional, site and subsite scales in three-dimensions. This may include the space-time distributions of temperature and heat flow above and below the water table, and their implications concerning flow between and within hydrogeologic, structural and textural units and features (including infiltration, evapotranspiration, recharge, discharge, directions and rates).

Hydrogeologic systems are open and complex and the corresponding knowledge base is invariably incomplete and imprecise. Therefore, such systems almost always lend themselves to multiple conceptualizations and the postulation of several alternative hypotheses. It is therefore important to explore varied conceptual frameworks and assumptions through a comprehensive evaluation of a broad range of regional and site data, their translation into coherent and internally consistent conceptual models or hypotheses, and an in-depth examination of these hypotheses in light of the available knowledge base. The more experts with a wider range of earth and environmental specialties are given access to the knowledge base, the larger and more varied are the alternative site descriptions they may identify.

To develop alternative conceptual models for a site, one should consider (among others) alternative representations of space-time scales; number and type of hydrogeologic units such as layers and structures such as faults; flow and transport properties (their values and statistics, spatial distribution and geostatistics, internal heterogeneity, anisotropy); location and type of system boundaries; space-time distribution of fluids and their states (pressure, density, saturation, temperature); space-time distribution of saturated, vadose and perched zones; space-time distribution of driving forces (infiltration, recharge, discharge, initial system states, boundary conditions); space-time distribution of flow patterns; existence and nature of fast flow paths; overall water balance; space-time distribution of contaminants; space-time distribution of contaminant sources; mechanisms and rates of source contaminant mobilization and leaching; controlling transport phenomena; migration, spreading and dilution patterns of contaminants; contaminant mass balance; space-time distribution of groundwater ages; space-time relationships

between major chemical constituents, isotopes, temperatures and heat flows; and their implications regarding flow and transport on regional, site and subsite scales in three dimensions and time.

The alternative conceptualizations should be firmly grounded in the available knowledge base. Each alternative conceptualization should be supported by key data.

The conceptualization is not complete without a clear articulation of ambiguities and uncertainties associated with each alternative description and interpretation (conceptual model or hypothesis) of site hydrogeology.

12.2.4 Avoid Conflict with Data

Verify in detail that none of the selected conceptual models are in conflict with any of the data; discard those that are.

Make sure that the selected conceptual models contain a minimum number of inconsistencies, anomalies and ambiguities with the lowest possible amount of remaining uncertainty about the site and the corresponding flow and transport regimes.

12.2.5 Rank Conceptual Models

Rank all alternative conceptual models in the order of their plausibility in light of the data, regardless of context. Provide a detailed rationale for the ranking and discard models that appear to be much less plausible than those retained.

Once a number of alternative hydrogeologic conceptualizations have been articulated, they must be systematically examined, compared, screened and ranked according to acceptance criteria that include logical consistency and coherence, and the extent to which they are supported or contradicted by available observations and data. Among otherwise equal conceptual models, we favor the least complex based on the principle of parsimony. Models that do not meet reasonable acceptance criteria of internal consistency, coherence and correspondence with the available data should be eliminated from further consideration at this stage of the analysis.

12.2.6 Develop Conceptual-Mathematical Models

For each alternative conceptual model, develop a complete, coherent and internally consistent conceptual-mathematical model of three-dimensional transient flow and transport within the schematized hydrogeologic system that is as simple as warranted by the context of the problem and the data. Provide a detailed rationale for your schematization of regional and/or site hydrogeology; choice of governing, state and constitutive equations; type and space-time distribution of forcing terms (sources, initial and boundary conditions); parameterization (functional form and number of parameters); and all other conceptual-mathematical aspects of the model. Discard models that do not lend themselves to such mathematical description.

A conceptual-mathematical model helps define and describe the hydrogeologic system in terms of space-time dimensions; topology; geometry; interactions (called processes) between kinematic (mass, concentration, flux, velocity) and dynamic (energy, force, stress) quantities; parameters and forcing terms (sources, initial and boundary conditions). Whether such a model is analytical or numerical (written in the language of calculus or algebra) is merely a technical, not a fundamental, distinction. The model ultimately allows one to explain and interpret existing observations, and to predict new observations, quantitatively.
For each alternative conceptual model identified during the qualitative stage of the process, one needs to define the hydrogeologic units and features (such as folds, faults, offsets and intrusions) that are to be explicitly modeled; their three-dimensional topology and geometry in terms of location, shape, size and relationship to other units and features (extent of external and internal boundaries and structural elements); equations that govern flow and transport phenomena included in the model; corresponding equations of state, in the form of functional relationships between parameters and state variables in the model; spatial variability of parameters that enter into the governing and state equations, within each hydrogeologic unit and feature; spatial variability of initial states in each unit and feature; boundary equations for flow and transport; and the space-time distribution of boundary and source parameters and values.

The quantification of system states includes functional and quantitative representations of fluid saturation, pressure, temperature, density, solute mass and concentration and regional as well as perched water tables on a range of scales (between and within hydrogeologic, structural and textural units and features). Governing and boundary equations include mathematical definitions and descriptions of active and anticipated flow and transport phenomena such as advection, convection, diffusion, dispersion and sorption for selected modes (discrete features; single, dual or multiple continua) and scales of manifestation, in the interior of units and features as well as on their boundaries and interfaces, including source terms, parameters, functional relationships between these quantities, and their space-time distributions.

12.2.7 Consider Simplifying the Model

If mathematical modeling in three spatial dimensions and/or in time is considered unnecessary or computationally infeasible, provide a detailed explanation of how the three-dimensional transient conceptual-mathematical model reduces to a two-dimensional and/or steady state model. In particular, demonstrate (preferably through a mathematical formalism and/or computationally) how the lower-dimensional model captures all relevant aspects of the original three-dimensional transient model. Avoid one-dimensional models at this stage.

The purpose of conceptual-mathematical models is to help quantify alternative hypotheses regarding the hydrogeologic makeup and behavior of a site. Since such hypotheses are always cast in three spatial dimensions, ideally so should the corresponding conceptual-mathematical models. Both model types must allow for the hydrogeologic regime to evolve in time.

Model simplification may entail a reduction in dimensionality (from three spatial dimensions to two or one; from transient to steady state), model size (smaller area and/or reduced depth); or details of various features, events and processes (fewer layers, faults or fracture zones; a less detailed representation of internal heterogeneity; single rather than dual continuum or discrete representation of a fractured rock; fewer discrete rainfall or infiltration events; a less detailed delineation of contaminant sources; constant rather than scale-dependent dispersion; equilibrium rather than kinetic sorption). Various types and levels of simplification can be entertained, leading to a number of simplified model structures for each initial (and more complex) conceptual-mathematical model.

In reality, it is much more difficult and time consuming to set up and run mathematical models in three than in two spatial dimensions. Working in two spatial dimensions is often feasible because eliminating the third dimension usually entails a far lesser phenomenological change than reducing the dimensionality of the problem from two to one. Modeling and visualization of complex flow and transport phenomena is much easier in two than in three dimensions, but not much harder than in one dimension. Hence the strategy in this report supports two-dimensional flow and transport analyses (whether mathematical-analytical or computational-numerical) in cases where the effect of the third dimension is demonstrably minor. However, the strategy discourages one-dimensional analyses unless a very strong and convincing hydrogeologic argument is made in their favor. This is so because flow and transport behaviors predicted by one- and multi-dimensional models often differ from each other in a fundamental way.

The spatial and temporal scales at which flow and transport phenomena are modeled depend in part on the contextual framework (area, depth and time-frame of concern; space-time scales on which performance measures are defined). These phenomena are affected to a large extent, but not exclusively, by hydrogeologic complexities, heterogeneities and driving-mechanisms that manifest themselves on similar scales. It is therefore important that conceptual-mathematical modeling start by identifying these "site-scale" features, factors and/or phenomena and by incorporating them directly as explicit elements in the model. Models that do not incorporate such elements explicitly, but account for them implicitly, are considered here to be simplified or abstracted. Models that fail to account for site scales complexities, either explicitly or implicitly, are oversimplified and would generally be too crude to provide a reliable description of site hydrogeology for most purposes.

Site-scale hydrogeologic complexities, heterogeneities and driving-mechanisms may be influenced and/or controlled by larger- or regional-scale features and factors such as regional recharge, discharge, and flow mechanisms and patterns on various time scales. A model must account for these large-scale influences and controls, as well as for associated uncertainties, through the appropriate assignment of initial conditions, boundary conditions, and source terms such as those that describe infiltration, recharge, discharge and leakage across aquitards.

Site-scale hydrogeologic complexities, heterogeneities and driving mechanisms may be significantly influenced by smaller- or subsite-scale features and factors such as space-time irregularities and fluctuations in external and internal boundary and source shapes or conditions; internal heterogeneities within site-scale hydrogeologic units, faults, dikes or other features; as well as smaller-scale units, faults, dikes, fractures or preferential flow channels. Determining

what is the nature and extent of these influences is an integral part of developing a conceptual-mathematical model for a site. If features and factors that manifest themselves on subsite scales are deemed important for the modeling of site-scale phenomena, one must account for them directly (explicitly, by embedding such features and factors discretely in the site model) or indirectly (implicitly, by formally integrating these features and factors into the site model equations and parameters) in the model.

Embedding discrete small-scale features and factors in a model renders it relatively complex. Integrating small-scale features and factors into the model equations and parameters renders it less complex than embedding. However, to compensate for loss of information, the equations change form and phenomenology (from Stokes to Darcy, isotropic to anisotropic, juxtaposed to overlapping dual or multiple continua, differential local to integro-differential nonlocal) and acquire new phenomenological parameters (permeability, dispersivity, integral kernels) that differ in nature and magnitude (from scalar to tensor, local to nonlocal, well-defined to scale-dependent) from the original parameters. Only in special cases can the original form and phenomenology be recovered, and even then the parameters usually change (due to upscaling). It is not presently clear which of these two approaches is better suited for their intended task. The strategy in this report considers both options.

Ignoring the influence of regional- or subsite-scale features or factors on site-scale flow or transport without due justification, or failing to insure that the treatment of all scales is self-consistent without demonstrating that some of them are not relevant to the problem at hand, constitute oversimplifications which may cast doubt on the reliability of the model.

Narrowly defined contextual or regulatory criteria, limited data or resources, and a quest for simplicity or transparency may motivate the adoption of hydrogeologic flow and transport models that are less than three-dimensional, ignore time, and include few details on limited scales. While such motivation for simplification and abstraction of hydrogeology and flow or transport dynamics may sometimes be justified on practical grounds, it does not in itself turn simplified and/or abstracted models into scientifically valid tools of performance assessment. Only a formal demonstration that such models capture the essential features and capabilities of their more complex and complete counterparts, and that they thereby provide comparable performance assessments or conservative bounds thereof, might justify their use.

To insure that relevant aspects of hydrogeologic complexity are reflected in modeled system behavior and performance, and that the model can be rendered compatible with site data, it is important that the process of simplification be done systematically and objectively. This can be done by filtering out undesirable details through formal averaging of the governing equations in space-time or in probability space, in a way which retains and renders their influence on the model implicit. This is equivalent to embedding.

Averaging three- or two-dimensional equations across one spatial dimension renders them two- or one-dimensional, respectively. Averaging transient equations over time may (but need not) render them representative of a steady state. Averaging can also be done over subdomains of the site being modeled, and over multiple time intervals. In each case, the averaging results in governing flow and transport equations that contain upscaled quantities. If the space-time scales

of these quantities differ from those of the available site data, then either the model or the data must be rescaled to render them compatible and comparable with each other.

Many scale-related problems are avoided if the averaging is done in probability space. Such "ensemble" averaging leads to stochastic equations that contain statistical moments of hydrogeologic variables (considered random), most commonly the mean and variance-covariance. The mean is a predictor of system behavior or performance, and the variance-covariance is a measure of predictive error. Both are smoother (vary more slowly in space-time) than their random counterparts and, in this sense, render the model relatively "simple." Despite their smoothness, both moments are defined on the same space-time scales as are the random hydrogeologic variables on which they are based. Stochastic models thus achieve smoothness and simplicity without any need to average or upscale in space-time. As they are typically conditioned on site measurements (*i.e.*, they honor the data), stochastic models are compatible with these measurements both in scale and magnitude. Yet another advantage of the stochastic method over space-time averaging is that it yields measures of predictive uncertainty. Stochastic approaches are increasingly recognized as offering a way to deal with complex, scale-dependent heterogeneous systems by means of relatively simple models.

12.2.8 Select Computational Code or Algorithm

Select a suitable computational code or algorithm to represent the conceptual-mathematical model.

A computational code is a tool rather than a model. However, selecting a particular code implies identifying specific processes that may govern flow and transport at a site, their symbolic mathematical representation, and their numerical approximation. Selecting the space-time dimensions and size of a computational grid determines the dimensions and scale of the system being modeled. Choosing the sizes of space-time discretization intervals defines the scales at which flow and transport processes are resolved. Specifying the location and type of sources, initial and boundary conditions identifies the forcings. Both material properties and forcing terms are associated with parameters that must eventually be assigned numerical values. Choosing the modes of their representation (parameterization) defines the resolution scales of these parameters (and thus of material properties and forcing terms) in space-time. Once this has been accomplished in a way that is supported by all relevant regional and site data, the code has been transformed from a mere tool to a *bona fide* conceptual-mathematical model of site hydrogeology.

12.2.9 Define Prior Statistics

Associate objective or subjective prior statistics with all input parameters and forcing terms that are considered to be uncertain. Provide a detailed rationale for considering any of these inputs to be known with certainty.

Since hydrogeologic medium properties and forcing functions are always uncertain, so are the model input variables. In other words, the input parameters of groundwater flow and transport models are inherently uncertain regardless of whether the model is deterministic or stochastic.

The difference is that whereas in deterministic models the input parameters are viewed as imperfectly known deterministic quantities, in stochastic models they are viewed as correlated random fields or processes that may be perfectly or imperfectly known at discrete sampling locations in space-time. In both cases, imperfect knowledge is characterized by random errors that may, but often are not, considered to be mutually correlated. This renders the input parameters of deterministic models correlated or uncorrelated random variables, and those of stochastic models correlated random fields or processes conditional on either exact or random measurements, which may themselves be correlated or uncorrelated among themselves. Whereas the second moment of uncorrelated random variables is a diagonal matrix of their respective variances, that of correlated random variables is a full square symmetric and positive definite matrix of their respective covariances.

Estimating input parameters for a model (deterministic or stochastic) on the basis of incomplete and/or uncertain data is equivalent to inferring their mean values from these data. If the mean values are based on site characterization data, they represent prior parameter estimates. If they are based on site monitoring data, they represent posterior parameter estimates. Here we consider only prior inference of model input parameters.

If the inferred parameter estimates vary in space and/or time in a way which reflects similar variations in the underlying site characterization data, they are said to be conditional on these data. If the inferred mean values do not reflect such spatial and/or temporal variability, they are said to be unconditional. Unconditional inference may take place when the data are too few or too clustered to allow defining their variability in space-time.

Assume that a set of site specific measurements are available, which represent the same hydrogeologic variable (say permeability or porosity) as a corresponding set of model input parameters. Such measurements are considered to be "hard." Ideally, the data would include probabilistic information about errors of measurement and test interpretation that suffer from a known amount (ideally zero) of statistical bias. If one has a statistically significant set of such data, one should be able to estimate a prior set of model input parameters on their basis. One should also be able to postulate a probabilistic model of prior parameter uncertainty based on statistics derived from these data. Such a probabilistic model is known to be of Type A.

A key measure of parameter uncertainty is the second statistical moment, or variance-covariance, of their estimation errors. Since input variables into both deterministic and stochastic computational models are specified in terms of a discrete set of input parameters, the corresponding variance-covariance forms a matrix. In the conditional case, the off-diagonal covariance terms of the matrix may reflect spatial and/or temporal correlation between parameters of a given type (say permeability or porosity) as well as cross-correlations between parameters of different types (say permeability and porosity). In the unconditional case, there is no inferred space-time variability and the covariance terms represent at most cross-correlations between different types of parameters.

Prior parameter estimates can be obtained from clustered data by means of common statistical methods, and from spatially distributed data by means of standard geostatistical techniques. This is true regardless of whether the parameters are intended for a deterministic or a stochastic

model. The main difference is that in the deterministic case, there may be a need to upscale the parameters whereas in the stochastic case, this may not be necessary. The simplest and most practical form of geostatistical inference with upscaling is block kriging. While this may not always be the most accurate and sophisticated way to proceed, it is quite adequate for many purposes. This is especially true in situations where enough site monitoring data are available to later modify the parameter estimates through model calibration.
To characterize prior estimation uncertainty one should, as a minimum, infer from the data a variance-covariance matrix of prior estimation errors. Only in rare circumstances would there be enough data to permit inferring from them higher statistical moments of these errors.

Standard geostatistical analysis consists of identifying the spatial autocovariance structure of each variable being analyzed (in terms of an autocovariance or variogram function) and providing a smooth (kriged) conditional estimate of this variable on a two- or three-dimensional grid, as well as the associated estimation (kriging) variance at each grid point. A more advanced analysis may also yield an autocovariance matrix for the estimate across the grid. Another level of sophistication may be achieved by identifying cross-covariance functions or cross-variograms for two or more variables, estimating them simultaneously by cokriging, and computing their co- and cross-covariances across the grid. Both kriging and cokriging may be used to estimate average values of the variables over finite blocks or subdomains of a two- or three-dimensional grid.

If there are insufficient hard data of a given hydrogeologic variable to conduct a meaningful statistical or geostatistical analysis, then the use of "soft" (qualitative) data coupled with indicator geostatistical analysis are recommended. This yields an uncertainty model for the prior parameters that is intermediate between Type A and Type B (defined below). Soft or indirect information about the parameters may include (a) off-site measurements of the parameters proper (quite often on scales other than those corresponding to the intended scale of model resolution) and/or (b) surrogate measurements on site that are known to correlate with the parameters of interest (for example, porosities or geophysical signatures that correlate in known ways with permeabilities, water contents or fracture densities). Statistics derived from off-site data must be considered potentially biased (due to a lack of site-specific information about mean parameter values and incompatibility of geology and scale). The associated variance may be too small or too large, depending on the quantity and quality of such data. Statistics derived from surrogate data may suffer from poorly defined correlations and incompatibility of scale.

An example of soft data use is that of pedotransfer functions, which allow one to estimate soil hydraulic characteristics on the basis of soil textural data.

Indicator geostatistics yields indicator variograms for various classes of the variable, a smooth (kriged) estimate of this variable on a two- or three-dimensional grid, and the probability that the variable is larger (or smaller) than specified at each grid point. Most geostatistical software packages include an indicator option and allow one to generate random realizations of the variable on a grid by means of indicator Monte Carlo simulation.

If the available hard and soft data are not amenable to geostatistical analysis (due to insufficient information about their spatial location, an inappropriate spatial pattern, insufficient number

and/or poor quality of data), an alternative is to rely on generic and/or subjective probabilities and statistics. Doing so is equivalent to postulating a Type B statistical model of prior parameter uncertainty. Such a model should always be suspected of suffering from an unknown amount of statistical and personal bias. Statistical bias is introduced due to lack of site-specific information about mean values of the parameters in question. A personal bias tends to manifest itself in the form of assigned uncertainty measures (most importantly bias and error variance) that are either too small or too large. The first is a manifestation of over-confidence in the model parameters, the second of unduly low confidence in their values.

Methods are available to update generic statistics on the basis of site-specific data using a Bayesian approach.

12.2.10 Perform Preliminary Test of Conceptual-Mathematical Model

Test the conceptual-mathematical model for qualitative consistency with the available data through computational exploration and visualization in at least two spatial dimensions, within a plausible range of input parameters and forcing terms. If inconsistent, repeat the previous steps in modified form (by altering and/or refining the hydrogeologic scheme and mathematical description of flow and/or transport) or discard the underlying conceptual model and select an alternative.

Having associated a deterministic and/or stochastic mathematical structure with each hypothesis that had been postulated for a site, the next step is to explore them quantitatively and graphically. The aim of this initial exploration is qualitative: To identify the potential ability of each conceptual-mathematical model to explain observed system behavior and predict it under future scenarios. For this, it is useful to run the model with a plausible range of input parameters and study its output in light of the available site information. This may reveal ambiguities and uncertainties that may prompt a reexamination of the data and/or a revision of the model (which may, but need not, render the model more complex). It may also reveal differences between the various conceptual-mathematical models that may help eliminate some of them and rank the rest in the order of their apparent ability to explain and predict system behavior.

The initial exploration of alternative conceptual-mathematical models can often be done in the two-dimensional horizontal and/or vertical planes. This is so because two-dimensional analyses differ from their three-dimensional counterparts to a much lesser extent than do one-dimensional analyses, yet are much less demanding than three-dimensional analyses. Two-dimensional scoping analyses should yield contour and/or color representations of hydrogeologic parameter distributions; system states (hydraulic head, pressure, saturation, density, temperature, solute concentration); flowlines (streamlines, pathlines, streaklines); and vectors of fluid and solute fluxes and velocities. If the conceptual-mathematical model involves transients, the analyses should consider variations of hydrogeologic variables with time. Some three-dimensional exploratory analyses may also be possible at this stage, including the preliminary quantification of fluid, solute and energy balances. Three-dimensional analyses may also be required if the two-dimensional results are insufficient to provide insight into relevant processes or fail to resolve key ambiguities and uncertainties.

Once an initial set of conceptual-mathematical models have been formulated and explored in a preliminary fashion, once has the option of postulating and exploring less complex alternatives to each. There is no reason to contemplate more complex alternatives at this stage of the modeling process.

12.2.11 Calibrate Model

Use an inverse (preferably Maximum Likelihood) approach to calibrate the computational model against observational (monitoring) data. This should yield posterior estimates of the model parameters, statistical measures of their estimation errors and associated model quality criteria. If the model cannot be properly calibrated against reliable data, repeat the previous steps in modified form (by altering and/or refining the hydrogeologic scheme and mathematical description of flow and/or transport) or discard the underlying conceptual model and select an alternative.

If monitoring data are available to allow comparing model predictions with observed system behavior, there usually is no reason to avoid calibrating the model against these data. The strategy strongly suggests that this be done.

The traditional approach to hydrogeologic modeling has been to postulate a deterministic model structure and treat its parameters as being imperfectly known. One would then derive prior estimates of these parameters from site characterization data (if available) and modify them so as to achieve an acceptable fit between model outputs and available monitoring data (if such exist). The process is known as model calibration, parameter estimation, history matching or inverse modeling. The parameters that yield the best match between observed and recorded system behavior form posterior estimates. It is common to consider the latter to be more suitable for predictive purposes than prior parameter estimates.

Though this is not always done, it is generally advisable to start the model calibration process by postulating a prior parameter uncertainty model. We explained how to do this and mentioned that it may lead to a Type A probability model when prior parameter uncertainty is characterized on the basis of a statistically significant set of site specific measurements that represent the parameter; a Type B model when the uncertainty is characterized by generic and/or subjective probabilities; and an intermediate type model when parameter uncertainty is inferred from indirect information about the parameters. We noted that a Type B model should always be suspected of an unknown amount of statistical and personal bias. A model that is intermediate between Type A and Type B must also be considered potentially biased, with variances that may be too small or too large, poorly defined correlations, and incompatibility of scale.

Improper upscaling may render the prior parameter estimates biased and their uncertainty mischaracterized, regardless of what type probability model one adopts. It is generally believed that model calibration tends to reduce both the bias and the uncertainty in prior parameter estimates by insuring that it reproduces adequately observed system behavior. This should render a calibrated model more reliable as a predictor than an uncalibrated model.

Calibrating a traditional deterministic flow model against measured values of head and flux is tantamount to conditioning it on such measurements. As the model hydraulic parameters are by nature conditional, the very act of adding measured heads and fluxes to the database alter their values. This inherent nonuniqueness persists regardless of whether the inverse problem is well-posed or ill-determined. It explains why model parameter estimates tend to change each time one redefines the underlying database.

Our guidelines stipulate that one use "statistical" model calibration algorithms which include analyses of parameter estimation uncertainty. Such analyses typically accept, but do not necessarily require, information about prior parameter statistics as input. The output includes posterior statistics of parameter estimation errors, which tend to be less biased and smaller than the prior estimation errors. Several public domain or commercial codes are available for this purpose (see section 7.4.2).

Differences between values of model simulated system states (most commonly hydraulic head, pressure or concentration) at discrete points in space-time, and observed values of these same state variables, are termed "residuals" in model calibration parlance. The residuals are considered to represent errors that are distributed randomly about the simulated model output. An underlying (often tacit) assumption is that the conceptual-mathematical model is exact and associated with an unknown set of "true" parameters. If these parameters were known, the model (structure plus parameters) would be exact. To the extent that it would produce nonzero residuals, this would be entirely due to errors in the monitoring record. These data errors (or their logarithmic transform, as is usually done with permeability and transmissivity) are typically taken to have zero mean and be normally distributed. Hence ideally (in the theoretical event that the true parameters were known), the residuals would exhibit a multivariate Gaussian distribution with zero mean (*i.e.*, they would be unbiased) and a variance-covariance identical to that of the monitoring data.

Model calibration is seen as the process of estimating the model true parameters without bias and as closely as possible. This is accomplished by defining an appropriate calibration criterion (or objective function) in terms of the residuals, and "optimizing" the parameters in a way which comes closest to satisfying this criterion (achieving the objective). The extent to which the criterion is satisfied (the objective achieved) becomes a measure of model fit. Theoretically, the optimized parameters are associated with estimation errors that have zero mean and minimum variance.

The most commonly used measure of model fit is the weighted sum of squared residuals. This is meaningful in the above statistical sense provided the monitoring data errors are mutually uncorrelated and each weight is inversely proportional to the corresponding error variance. Model calibration then reduces to a weighted least squares fit of the model to the data (or regression of the parameters on the data). Since groundwater model parameters are usually related to state variables in a nonlinear fashion, the weighted least squares fit is nonlinear. This is why model calibration must usually be done iteratively.

If all the weights (and variances) are equal, the model fit criterion can be normalized to form a simple sum of squared residuals. Model calibration then reduces to an ordinary nonlinear least

squares fitting process (or nonlinear regression). A detailed set of guidelines for the effective calibration of such nonlinear regression models has been prepared by Hill (1998).

A less common but still widely used measure of model fit is the generalized sum of squared residuals. Whereas in the previous cases the weights formed a diagonal matrix proportional to the inverse of a diagonal matrix of observational error variances, here the weights form a full matrix proportional to the inverse of a (usually full) variance-covariance matrix. This is statistically meaningful in the previous sense when the errors are mutually correlated. Calibration now becomes a generalized nonlinear least squares or regression process.

If some of the statistical parameters that define the observational errors are unknown, they may sometimes be estimated jointly with the other model parameters by the maximum likelihood method. This entails a likelihood function, which is the likelihood of the parameters given a set of (conditional on) observational data. The latter is the probability density of the (error corrupted and so random) data given (conditional on) the parameters. The objective would be to maximize the likelihood function.

Since the data (or their log transform) are assumed to be normally distributed, the likelihood function is multivariate Gaussian. As such, it includes an exponential term whose negative logarithm is equal to half the weighted (by the inverse covariance matrix) sum of theoretical square residuals. Parameters are estimated by minimizing the negative logarithm of the likelihood function. In the special case where all statistical parameters are known, the negative log likelihood function reduces to the standard generalized least squares criterion.

The strategy in this report favors a maximum likelihood approach to model calibration, which incorporates information about prior parameter statistics into the statement of the inverse problem (see section 7.4.3). This approach yields a negative log likelihood criterion that includes two weighted square residual terms instead of one. The first is the usual generalized sum of squared differences between simulated and observed state variables. The second is a generalized sum of squared differences between posterior and prior parameter estimates. The corresponding weight matrix is proportional to the inverse covariance matrix of prior parameter estimation errors. The constant of proportionality, λ, is treated as a free statistical parameter that may be estimated jointly with the remaining parameters by maximum likelihood. This can be done most easily by calibrating the model for various values of λ, plotting the negative log likelihood (S) against λ, and finding graphically the value of λ that minimizes S. Allowing λ to be initially unspecified means that neither the covariance matrix of the observational data nor that of the prior parameters need be fully specified; it is enough to specify each of them up to its own constant of multiplication. It is then possible to estimate these constants of multiplication on the basis of λ.

Including prior information in the calibration criterion 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).

When either set of data is too small, clustered or otherwise unsuitable for a meaningful assessment of prior statistics, it may still be possible to come up with prior parameter estimates and weigh them subjectively relative to each other, based on their perceived reliability. Similar weights may be assigned to site monitoring data. One can then adopt a calibration criterion $J + \lambda P$ equal to a weighted sum (J) of squared residuals (differences between simulated and observed state variables), plus the product of λ with a weighted sum (P) of squared differences between posterior and prior parameter estimates. One can then calibrate the model for various choices of λ and plot the weighted sum of squared residuals versus $1/\lambda$ to see how the latter affects the former. One would typically find that the sum of squared residuals is largest when $1/\lambda = 0$ and decreases more-or-less monotonically to an asymptote as $1/\lambda$ increases. This is clear considering that $1/\lambda = 0$ corresponds to $\lambda = \infty$, which is equivalent to giving infinite weight to the prior parameter estimates and forcing the posterior estimates to coincide with them. As $1/\lambda$ is allowed to increase, the weight placed on the prior parameters decreases and their constraining effect on the posterior estimates gradually diminishes. Eventually $1/\lambda$ becomes large enough to virtually eliminate any effect of the prior on the posterior estimates. The calibration is now unconstrained by site characterization data and relies entirely on site monitoring data.

Rather than a single calibration criterion one now has two, the weighted sum of squared residuals (J) and the weighted sum of squared differences between posterior and prior parameter estimates (P). These two criteria are mutually incompatible in that to satisfy one, it is necessary to sacrifice the other. In most cases, the optimum solution to this multiobjective or Pareto problem is not to minimize either objective but to find an acceptable tradeoff between them on the basis of subjective value judgment (see section 7.4.4).

Methods to calibrate stochastic models have also been developed and are discussed in the report.

In the absence of sufficient or reliable site monitoring data, the model must rely entirely on prior parameter estimates and is only as good as are the latter. The model remains uncalibrated, untested and unconfirmed.

The statistical approach to deterministic model calibration allows one to quantify the uncertainty associated with posterior parameter estimates. One measure of posterior parameter estimation errors is the Cramer-Rao lower bound on their covariance matrix. This lower bound is given by the inverse Fisher information matrix. Each term in the Fisher information matrix is one half the ensemble average of the second derivative of the negative log likelihood criterion with respect to a pair of parameters. It thus measures the average rate at which model sensitivity to one parameter is affected by changes in another parameter. From it, one can obtain a corresponding lower bound on the correlation matrix of the estimates. Ideally, the parameter estimates should be uncorrelated.

The Fisher information matrix is usually approximated by linearizing the relationship between system states and parameters when the latter are optimal. It is then expressed as a function of the

calibration weights (inverse covariance matrices of observed and prior input data) and a Jacobian matrix, whose terms represent sensitivities of system states to the parameters. It is thus clear that an analysis of parameter uncertainty includes in it a sensitivity analysis of the model and its output.

A large variance associated with a given parameter indicates a high level of uncertainty in its estimate. A high degree of correlation between estimates implies that they are linearly related in a statistical sense. This means that they cannot be estimated individually with the available data, and should either be lumped into one parameter or estimated anew when more data become available.

Linearization implies that if the prior (observation and parameter) errors are multivariate Gaussian (as one generally assumes), so are the posterior parameter (or log parameter) estimation errors. This allows one to associate them with linear confidence intervals, which can be viewed as corresponding error bounds.

12.2.12 Confirm Model

If feasible, confirm aspects of the model by predicting system behavior under conditions other than those used for calibration and comparing with observed system behavior. Quantify model predictive uncertainty and check the extent to which the observations lie within corresponding uncertainty bounds. If not, repeat the previous steps in modified form (by altering and/or refining the hydrogeologic scheme and mathematical description of flow and/or transport) or discard the underlying conceptual model and select an alternative.

In rare situations where enough observational data are available to allow using only some of them for model calibration, it may be possible to use the remaining data for purposes of model verification or confirmation. For this to be valid, the data used for confirmation must represent a different mode of system behavior that those used for calibration.
Otherwise, the confirmation data are merely an extension of the calibration data and being able to reproduce them does not constitute a meaningful test of model predictive capability. As both the model and the data are generally uncertain, model confirmation must be cast in a suitable probabilistic and/or stochastic framework.

12.2.13 Compare Calibrated Models

Compare all retained calibrated models on the basis of suitable model quality criteria. Discard models that are clearly inferior to others.

Alternative conceptual-mathematical models that have been successfully calibrated can be compared on the basis of quality criteria such as model fit and posterior parameter uncertainty. Examples include:

- Likelihood or negative log likelihood of posterior parameter estimates, given the available data;

- Inverse variance-covariance weighted sums of squared residuals, which measure the overall lack of fit between model generated and observed system states;
- Various statistics of the residuals, which measure how closely they satisfy the assumptions of Gaussianity, unbiasedness, and similarity of their variance-covariance to that of the observational data;
- Consistent measures of systematic differences between contours, time records, peaks and trends exhibited by observed and simulated state variables;
- Invariant properties of the Cramer-Rao lower bound on the posterior covariance matrix of parameter estimation errors, such as—
  - Trace, also known as A-optimality criterion (which measures the cumulative variance of all estimation errors);
  - Determinant, also known as D-optimality criterion (which provides an overall measure of the estimation error variance-covariance matrix);
  - Eigenvectors (the components of which represent relative contributions by the various parameter estimates, implying that parameters associated with a single eigenvector have uncorrelated estimation errors, and those associated with multiple eigenvectors have cross-correlated estimation errors, rendering them less amenable to discrimination); and
  - Eigenvalues (parameters associated with eigenvectors that have small eigenvalues being less uncertain than those associated with eigenvectors that have large eigenvalues); the maximum absolute eigenvalue is known as E-optimality criterion;
- Linear confidence intervals;
- Invariant properties of the Jacobian sensitivity matrix (of system states to the parameters); *etc*.

Models that have not been compared with monitoring data have indeterminate quality in that nothing is known about their ability to reproduce real system behavior. Calibrated models rank higher than uncalibrated models in terms of parameter quality criteria because (posterior) estimates tend to be less biased and uncertain than prior estimates, on which uncalibrated models are based.

If all models remain uncalibrated, their only comparative quality criteria are those that pertain to prior parameter uncertainty.

These include—
- The variance-covariance of prior parameter estimation errors;
- Various invariant properties thereof; and
- Corresponding linear confidence intervals.

All models can be compared on the basis of model complexity criteria such as—

- Dimensionality;
- Number of simulated processes;
- Degree of nonlinearity;
- Number of parameters; and
- Number or size of grid cells.

To help identify a set of models which represent acceptable tradeoffs between quality and complexity, it is helpful to employ a multiobjective approach as described in section 8.2.2. To further discriminate between these models, it is useful to gauge and rank them in terms of likelihood based model discrimination criteria, which account for both model fit and complexity. This applies only to calibrated models because uncalibrated models are associated with zero likelihood. Several such criteria are discussed in section 8.2.3. The model testing and selection process ends with the final ranking of models on the basis of likelihood based discrimination criteria, and the optional elimination of low ranking models from further consideration.

12.2.14 Render Predictions Using Individual Models

Use each retained calibrated model to predict system behavior as well as relevant hydrogeologic performance measures under various scenarios of relevance to the problem, and quantify the corresponding predictive uncertainty.

Monte Carlo is by far the most common method of assessing the predictive uncertainty of a model. The method is conceptually straight forward and has the advantage of applying to a very broad range of both linear and nonlinear flow and transport models. Given information about the statistical properties of the model input parameters (including those that represent forcing terms) or their log transformed values, one generates numerous equally likely realizations of the parameters (or log parameters, as in the case of permeability and transmissivity). If the model has been calibrated, the input statistics correspond to the posterior parameter estimates. If the model is uncalibrated, they correspond to the prior estimates. In both cases, the estimates are generally associated with a full variance-covariance matrix. Upon assuming that the estimation errors are multivariate Gaussian, one has all the information one needs to produce random realizations of the parameters.

One way to produce random realizations of a multivariate Gaussian vector (of parameters) is by Cholesky decomposition of its variance-covariance matrix. This method is well suited for full matrices and highly efficient when the latter are not too large. Its efficiency stems from the fact that only one decomposition is required for an unlimited number of realizations.

Once a parameter realization has been generated, one runs the model under a given scenario to generate a corresponding random prediction of system states. Upon averaging the predictions over all realizations, one obtains a sample mean prediction. As the number of such Monte Carlo runs increases, one hopes that the sample mean converges to a theoretical ensemble mean. While such convergence can neither be insured nor verified in most cases, it is important to at least plot a representative measure of the sample mean versus the number of runs to verify that it has reached a stable value. The stable sample mean is then taken to constitute the best prediction the model can produce.

In addition to computing the sample mean, one typically also computes the sample variance of the predictions and generates a corresponding frequency histogram. Both converge more slowly to their theoretical ensemble equivalents than does the sample mean. It is therefore very important to verify that both have stabilized with the number of runs. Since most groundwater

models are nonlinear, one need not expect the histogram of the predictions to resemble either a normal or a log normal distribution.

Linearizing the flow and transport models allows one to establish approximate error bounds, or confidence limits, for model predictions based on linear regression theory, without Monte Carlo simulation. Linearization expresses each observed value of a state variable as a linear combination of unknown "true" parameters plus a random error (ε). As the model is considered exact and its parameters true, the latter implies that ε represents an error of observation. For regression (calibration) purposes, ε is replaced by a corresponding residual and an optimum (posterior) estimate of the unknown parameters is obtained by minimizing a suitably weighted sum of squared residuals. This sum thus becomes a measure of the error (variance) with which observed state variables are represented by the calibrated model. The corresponding variance is generally larger than that of ε because the model is based on estimated rather than true parameters.

The weighted sum of squared residuals does not, however, measure the error with which state variables that have not yet been observed (or have not been considered for calibration) would be predicted by the model. This is so because future predictions would be done with parameter estimates that are themselves corrupted by an estimation error. Linearization expresses each predicted value of a state variable as a linear combination of unknown parameters plus a random error (ε). Each unknown parameter is the sum of a known posterior parameter estimate and an associated estimation error. Hence the error in predicted system states is the sum of two errors: one due to replacing the true parameters by their posterior (calibrated) estimates (proportional to the weighted sum of residuals), and the other due to rendering a prediction with the aid of uncertain posterior parameters (the uncertainty being characterized by the estimation variance-covariance of the parameters). Hill (1998) explains how one can establish linearized confidence limits, or error bounds, for predictions rendered by models calibrated using nonlinear least squares with a diagonal weight matrix.

Assessing uncertainty in the output of a calibrated deterministic model, as just described, fails to account for the stochastic component of prediction error. Evaluating the uncertainty of standard deterministic model parameter estimates may allow one to quantify the uncertainty in computed conditional mean system state predictors (mean head, concentration, flux *etc.*). This, however, says nothing about how actual (random and unknown) system states fluctuate about their predictors; such information is provided only by stochastic models that treat hydraulic parameters as random fields.

To do so, one must adopt a stochastic model. High-resolution Monte Carlo simulation is by far the most common method to assess predictive uncertainty stochastically. It entails generating multiple random realizations of flow and transport parameters on a fine grid, solving standard deterministic flow and transport equations with these parameters on the same grid under scenarios of interest, averaging the results and analyzing them statistically. If based on prior parameter estimates, the Monte Carlo results honor measured values of the parameters at discrete points in space. If based additionally on a suitable inverse method, they also honor parameter observed values of state variables.

In all other respects, the analysis of predictive uncertainty by the stochastic Monte Carlo method is similar to that described previously in connection with deterministic models.

It is important to be aware that forward and inverse algorithms based on the moment equation approach are presently under development by our group. The algorithms yield directly (without Monte Carlo simulation) optimized unbiased predictors of groundwater flow and transport variables for randomly heterogeneous hydrogeologic environments, under the action of uncertain source and boundary terms. They also yield the variance-covariance of associated estimation and prediction errors. The algorithms do so while accounting explicitly for the multiscale (*e.g.*, fractal) nature of hydrogeologic heterogeneity. They allow optimum use of field information through joint conditioning on measured values of hydraulic parameters, hydraulic heads and solute concentrations. We expect these algorithms to be available for general use in the near future.

12.2.15 Render Joint Predictions by All Models

Use Maximum Likelihood Bayesian Model Averaging to render joint predictions by all the retained calibrated models, and to assess their joint predictive uncertainty. Acknowledge the conditional nature of the predictions and their corresponding uncertainty measures.

Alternative models that have been tested, compared, ranked and retained for further consideration are considered to constitute potentially valid simulators and predictors of site hydrogeology. There is no valid basis to prefer one of these models over another, and one must therefore use them in tandem. To render joint predictions using multiple models and assess their joint predictive uncertainty, we recommend adopting the Maximum Likelihood Bayesian Model Averaging (MLBMA) approach described in section 9.4. This novel approach relies on probabilistic maximum likelihood concepts to (a) calibrate each model against observed space-time variations in system states (pressure, water content, concentration), considering prior information about relevant soil properties; (b) eliminate models that cannot be so calibrated with acceptable fidelity; (c) predict future system behavior or performance measures (travel times, concentrations, mass rates), and assess corresponding predictive uncertainty, using each model; and (d) average the results using posterior model probabilities as weights. MLBMA supports the principle of parsimony in that among models that have similar predictive capabilities, it favors those having fewer parameters and being therefore simpler. Regulatory hydrogeologic performance of a site, under a given scenario, is evaluated in terms of corresponding performance measures. These in turn are computed on the basis of system states (such as travel time and concentration) that are predicted using groundwater flow and transport models. This implies that the same models and methods that one uses to predict system states, and to assess their predictive uncertainty, can also be used to predict performance measures and their uncertainty. No special methodology is needed for this purpose.

It is important to recognize implicitly that the set of predictions one produces with any given choice of alternative structural models and parameter sets, by whatever method, is conditional on the choice of models and the data used to support them. As such, these predictions do not represent all possibilities but only a limited range of such possibilities, associated with these

models and data. Any change in the latter would generally lead to a different assessment of predictive model uncertainty. There thus appears to be no way to assess the uncertainty of hydrologic predictions in an absolute sense, only in a conditional or relative sense.

12.2.16 Identify Additional Data Needs

If none of the models are deemed acceptable, perform a sensitivity analysis to help identify the type and quantity of additional site data that might materially enhance their reliability and credibility. Decide based on the potential benefit (in terms of bias and uncertainty reduction) and cost (in terms of time, effort and money) of such data whether or not to collect them and how.

To help evaluate what if any additional data might be worth collecting so as to materially reduce model uncertainty (by further constraining the range of alternative model structures and parameters), one may conduct a sensitivity analysis to indicate what system behavior appears to be most sensitive to which parameters at what locations. The next step is to consider performing additional site characterization where existing parameter estimates are least certain and the model is relatively insensitive to their values, and monitoring system behavior where it is most sensitive to model parameters while prediction errors appear to be relatively large and consequential. This may indicate the type and quantity of additional site data that might materially enhance model reliability and credibility.

The question how much and what kind of data are enough for model development and evaluation is one of economics and policy, not of hydrogeologic analysis or modeling. It is therefore the responsibility of managers and decision-makers to answer, not hydrogeologists or modelers. The strategy in this report is designed to help one address the question by how much would additional site characterization and monitoring improve the prediction of performance measures derived from the hydrogeologic analysis. In other words, the strategy addresses the worth of data in terms of their contributions to the potential enhancement of model reliability and credibility, not in terms of their marginal cost-benefit. Its aim is to allow managers and decision-makers to make informed decisions about the time, manpower and budget that they deem worth allocating to these activities in light of such enhancement under existing administrative, budgetary and policy constraints.

12.2.17 Repeat Iterative Process with New Data

If and when new data of significant weight are obtained, repeat the iterative process till its cost-benefit ratio reaches a value that is not considered worth exceeding.

This final guideline requires no further elaboration.

13 REFERENCES

Ahuja, L.R., J.W. Naney, and R.D. Williams, Estimating soil water characteristics from simpler properties or limited data, *Soil Sci. Soc. Am. Jour.*, 49, 1100–1105, 1985.

Akaike, H., A new look at statistical model identification, *IEEE Trans. Autom. Control*, AC19, 716–723, 1974.

Akaike, H., On entropy maximization principle, in *Application of Statistic*s, edited by P. R. Krishnaiah, 27–41, North-Holland, New York, 1977.

Altman, S.J., B.W. Arnold, R.W. Barnard, G.E. Barr, C.K. Ho, S.A. McKenna, and R.R. Eaton, Flow Calculations for Yucca Mountain Groundwater Travel Time (GWTT-95), Rep. SAND96-0189, Sandia National Laboratories, Albuquerque, New Mexico, 1996

Anderson, T.W. and G.W. Freethey, Simulation of Groundwater Flow in Alluvial Basins in South-central Arizona and Parts of Adjacent States, *U.S. Geological Survey Professional Paper* 1406-D, 1995.

Anderson, T. W., G. W. Freethey, and P. Tucci, Geohydrology and water resources of alluvial basin in south-central Arizona and parts of adjacent states. U.S. Geological Survey Professional Paper 1406-B, 1992.

Anderson, M.P. and W.W. Woessner, *Applied Groundwater Modeling, Simulation of Flow and Advective Transport*, Academic Press, San Diego, California, 1992.

Ando, K., A. Kostner, and S.P. Neuman, Stochastic continuum modeling of flow and transport in a crystalline rock mass: Fanay-Augères, France, *Hydrogeology Journal*, DOI 10.1007/s10040-003-0286-0, 2003.

Amir, O. and S.P. Neuman, Gaussian closure of one-dimensional unsaturated flow in randomly heterogeneous soils, *Transport in Porous Media*, 44(2), 355-383, 2001.

Bagtzoglou, A. C., S. Mohanty, A. Nedungadi, T.-C. J. Yeh, and R. Ababou, Effective hydraulic property calculations for unsaturated, fractured rock with semi-analytical and direct numerical techniques: Review and applications, *Rep. CNWRA 94-007*, Cent. for Nucl. Waste Regul. Anal., San Antonio, Texas, 1994.

Bennet, R.R. and R.R. Meyer, Geology and groundwater resources of the Baltimore area: Maryland Board Nat. Resources, Dept. Geology, Mines, Water Resources Bull 4., 1952.

Beven, K.J., Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, 16, 41- 51, 1993.

Beven, K., Uniqueness of place and non-uniqueness of models in assessing predictive uncertainty, 1085 – 1091, 785-792, *Computational Methods in Water Resources XIII*, edited by Bentley *et al.*, Balkema, Rotterdam, The Netherlands, 2000.

Beven, K.J. and A.M. Binley, The future of distributed models: model calibration and uncertainty prediction, Hydrological Processes, 6, 279 – 298, 1992.

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.

Beven, K., and P. Germann, Macropores and water-flow in soils, *Water Resour. Re*s., 18(5), 1311-1325, 1982.

Birkholzer, J. And C-F Tsang, Solute channeling in unsaturated heterogeneous porous media, *Water Resour. Res.*, 33(10), 2221-2238, 1997.

Bouma, J. and J. A. van Lanen, Transfer functions and threshold values: From soil characteristics to land qualities, in *Quantified land evaluation*, edited by K.J. Beek *et al.*, International Institute Aerospace Surv. Earth Sci., ITC Publ. 6, Enschede, The Netherlands, 106-110, 1987.

Bresler, E., Anion exclusion and coupling effects in nonsteady transport through unsaturated soils, 1, Theory, *Soil Sci Soc Am J.*, 37(5), 663-669, 1973.

Brooks, R. H., and A. T. Corey, Hydraulic properties of porous media, *Hydrol. Pap*., Colorado State Univ., Fort Collins, 3, 1–27, 1964.

Cacas, M.C., E. Ledoux, G. de Marsily, A. Barbreau, P. Calmels, B. Gaillard, and R. Margritta, Modelling fracture flow with a stochastic discrete fracture network: Calibration and validation, 1, The flow model, *Water Resour. Res.*, 26(3), 479-489, 1990a.

Cacas, M.C., E. Ledoux, G. de Marsily, A. Barbreau, P. Calmels, B. Gaillard, and R. Margritta, Modelling fracture flow with a stochastic discrete fracture network: Calibration and validation, 2, The transport model, *Water Resour. Res.*, 26(3), 491-500, 1990b.

Capilla, J.E., J.J. Gómez-Hernández, and A. Sahuquillo, Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, 2. Demonstration on a synthetic aquifer, Jour. Hydrology, 203, 175-188, 1997.

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: 2. Uniqueness, Stability and Solution Algorithms, *Water Resour. Res.* 22(2), 211-227, 1986b.

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, 1986c.

Carrera, J., J. Heredia, S. Vomvoris, and P. Hufschmied, Modeling of flow in a small fractured monzonitic gneiss block, pp. 115-167 in Neuman, S.P. and I. Neretnieks, *Hydrogeology of Low Permeability Environments, Hydrogeology, Selected Papers*, 2, Verla Heinz Heise, Hannover, Germany, 1990.

Carrera, J., A. Medina, C. Axness, and T. Zimmerman, Formulations and computational issues of the inversion of random fields, pp. 62-79 in Dagan, G. and S.P. Neuman (editors), *Subsurface Flow and Transport: A Stochastic Approach*, Cambridge University Press, Cambridge, United Kingdom, 1997.

Carsel, R.F. and R.S. Parish, Developing joint probability distribution of soil water retention characteristics, *Water Resour. Res.*, 24(5), 755-769, 1988.

Chen, G. and S.P. Neuman, Wetting front instability in randomly stratified soils, *Physics of Fluids*, 8(2), 353-369, 1996.

Chen, G., M. Taniguchi, and S.P. Neuman, An Overview of Instability and Fingering During Immiscible Fluid Flow in Porous and Fractured Media, Rep. NUREG/CR-6308, prepared for U. S. Nuclear Regulatory Commission, Washington, D. C., May, 1995.

Chen, G., W.A. Illman, D.L. Thompson, V.V. Vesselinov, and S.P. Neuman, Geostatistical, type curve and inverse analyses of pneumatic injection tests in unsaturated fractured tuffs at the Apache Leap Research Site near Superior, Arizona, pp. 73-98 in Dynamics of Flow and Transport in Fractured Rocks, edited by B. Faybishenko et al., AGU monograph series, 2000.

Clifton, P.M., and S.P. Neuman, Effects of Kriging and Inverse Modeling on Conditional Simulation of the Avra Valley Aquifer in southern Arizona, *Water Resour. Res.*, 18(4), 1215-1234, 1982.

Dagan, G. and S.P. Neuman (editors), *Subsurface Flow and Transport: A Stochastic Approach*, Cambridge University Press, Cambridge, United Kingdom, 1997.

D'Agnese, F.A., C.C. Faunt, A.K. Turner, and MC. Hill, Hydrogeologic Evaluation and Numerical Simulation of the Death Valley Regional Ground-Water Flow System, Nevada and California, U.S. Geological Survey Water-Resources Investigations Report 96-XXXX, Denver, Colorado, 1996.

de Marsily, G., De l'identification des systèmes en hydrogeologiques, *Doctor d'Etat dissertation*, L'Univ. Pierre et Marie Curie-Paris IV, Paris, 1978.

Desbarats, A. J., Upscaling capillary pressure-saturation curves in het-erogeneous porous media, *Water Resour. Re*s., *3*1(2), 281–288, 1995.

Desbarats, A.J., Upscaling constitutive relationships in unsaturated heterogeneous media: numerical experiments, 1381-1390, Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media, Proc. 1997 Intern. Workshop in Riverside, California, edited by M. Th. Van Genuchten, F.J. Leij, and L. Wu, University of California, Riverside, California, 1999.

Deutsch, C. V. and A. G. Journel, *GSLIB Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 1998.

Di Federico, V. and S.P. Neuman, Scaling of random fields by means of truncated power variograms and associated spectra, *Water Resour. Res.*, 33(5), 1075-1085, 1997.

Di Federico, V. and S.P. Neuman, Flow in multiscale log conductivity fields with truncated power variograms, *Water Resour. Res.*, 34(5), 975-987, 1998a.

Di Federico, V. and S.P. Neuman, Transport in multiscale log conductivity fields with truncated power variograms, *Water Resour. Res.*, 34(5), 963-973, 1998b.

Di Federico, V., S.P. Neuman, and D.M. Tartakovsky, Anisotropy, lacunarity, upscaled conductivity and its covariance in multiscale fields with truncated power variograms, *Water Resour. Res.*, 35(10), 2891-2908, 1999.

Distefano, N. and A. Rath, An identification approach to subsurface hydrological systems, Water Resour. Res., 11(6), 1005-1012, 1975.

Doherty, J., L. Brebber, and P. Whyte, *PEST: Model Independent Parameter Estimation*, Watermark Computing, 1994.

Domenico, P.A. and G.A. Robbins, A dispersion scale effect in model calibrations and field tracer experiments, *Jour. Hydrology*, 7, 121-132, 1984.

Duckstein, L., B. Bobée, and F. Ashkar, A multiple criteria decision modelling approach to selection of estimation techniques for fitting extreme floods, Stochastic Hydrology and Hydraulics 5, 227– 238, 1991.

Duckstein, L., M. Gershon, and R. McAniff, Model selection in multiobjective decision making for river basin planning, *Adv. Water Resources*, 5, 178–184, 1982.

Durand, E., Etude de l'efffet d'echelle en milieu fissure, Phase 2, Volume 1, Phase 2A: Etude des ecoulements sur le site de Fanay-Augères, Premier partie: Description des essais hydrauliques, Rapport final, Commission des Communautes Europeennes, Luxembourg, 1988.

Eakin, T.A. and others, Summary appraisals of the Nation's groundwater resources: Great Basin, U.S. Geological Survey Prof. Paper 813-G, 1976.

EES Division, Science and Technology Assessment and Division Review Committee Meeting, Los Alamos National Laboratory, Los Alamos, New Mexico, April 9–12, 2001.

Efendiev, Y., L.J. Durlofsky, and S.H. Lee, Modeling of subgrid effects in coarse-scale simulations of transport in heterogeneous porous media, *Water Resour. Res.*, 36(8), 2031-2041, 2000.

Emsellem, Y. and G. de Marsily, An automatic solution for the inverse problem, *Water Resour. Res.*, 7(5), 126-1283, 1971.

Entekhabi, D., Predictability and limits to prediction in hydrologic systems, *Eos Transactions*, AGU, 82(2), 14, January 9, 2000.

EPA, DOE, and NRC, A Technical Guide To Ground-Water Model Selection at Sites Contaminated With Radioactive Substances, EPA 402-R-94-012, Washington, D.C., 1994.

Finsterle, S., iTOUGH2 User's Guide, Report LBNL-40040, Lawrence Berkeley National Laboratory, Berkeley, Calif., 1999a.

Finsterle, S., iTOUGH2 Command Reference, Report LBNL-40041, (Update includes commands for latest iTOUGH2 version), Lawrence Berkeley National Laboratory, Berkeley, Calif., 1999b.

Finsterle, S., iTOUGH2 Sample Problems, Report LBNL-40042, (Updated reprint), Lawrence Berkeley National Laboratory, Berkeley, Calif., 1999c.

Finsterle, S. and B. Faybishenko, Inverse modeling of radial multistep outflow experiment for determining unsaturated hydraulic properties, Adv. Water Resour., 22(5), 431–444, 1999.

Fridrich, C.J., W.W. Dudley Jr., and J.S. Stuckless, Hydrogeologic analysis of the saturated-zone ground-water system, under Yucca Mountain, Nevada, *Jour. Hydrology*, 154, 133-168, 1994.

Freeze, A.R., Theoretical analysis of regional groundwater flow, Inland Water Branch, Department of Energy, Mines and Resources, Canada, Scientific Series, No. 3, 1969.

Gaganis P. and L. Smith, A Bayesian approach to the quantification of the effect of model error on the predictions of groundwater models, *Water Resour. Res.*, 37(9), 2309-2322, 2001.

Gascoyne, M., J.D. Ross, and R.L. Watson, Highly saline pore fluids in the rock matrix of a granitic batholith in the Canadian Shield, in Hydrogeology if Rocks of Low-Permeability, Proc. 30th Intern. Geol. Congress, Beijing, China, 1996.

Gautschi, A. The Kristallin programme: Investigations, results, future procedure, NAGRA Bull. 25, 10-21, Wettingen, Switzerland, 1995.

Gelhar, L.W. and C.L. Axness, Three-dimensional stochastic analysis of macrodispersion in aquifers, *Water Resor. Res.*, 19(1), 161-180, 1983.

Gómez-Hernández, J.J. and R.M. Srivastava, ISIM3D: An ANSI-C three-dimensional multiple indicator conditional simulation program, *Computers and Geosciences*, 16(4), 395-440, 1990.

Gómez-Hernández, J.J., A. Sahuquillo, and J.E. Capilla, Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, 1. Theory, *Jour. Hydrology*, 203, 162-174, 1997.

Gómez-Hernández, J.J., A. Sahuquillo, and J.E. Capilla, Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, 1. Theory, *Jour. Hydrology*, 203, 162-174, 1997.

Gómez-Hernández, J.J., H.J.W.M. Hendricks Franssen, and A. Sahuquillo, Calibration of 3-D transient groundwater flow models for fractured rock, pp. 185-194 in *Calibration and Reliability in Groundwater Modelling: Coping with Uncertainty, ModelCARE 99*, edited by F. Stauffer, W. Kinzelbach, K. Kovar and E. Hoehn, IAHS Publication no. 265, IAHS Press, Centre for Ecology and Hydrology, Wallingford, UK, 2000.

Gómez-Hernández, J.J., H.J.W.M. Hendricks Franssen, and A. Sahuquillo, Sequential self-calibrating multiple realizations to flow and transport data, *Proc. 4th Intern. Conf. on Calibration and Reliability in Groundwater Modelling (ModelCARE 2002)*, edited by K. Kovar and Z. Hrkal, Charles University, Prague, Czech Republic, 320-323, 2002.

Guadagnini, A. and S.P. Neuman, Nonlocal and localised analyses of conditional mean steady state flow in bounded, randomly nonuniform domains, 1, theory and computational approach. *Wat. Resour. Res*. 35(10), 2999-3018, 1999a.

Guadagnini, A. and S. P. Neuman, Nonlocal and localised analyses of conditional mean steady state flow in bounded, randomly nonuniform domains, 2, computational examples. *Wat. Resour. Res*. 35(10), 3019–3039, 1999b.

Guadagnini, A. and S.P. Neuman, Recursive conditional moment equations for advective transport in randomly heterogeneous velocity fields, *Transport in Porous Media*, 42(1/2), 37–67, 2001.

Gupta, H.V., S. Sorooshian, and P.O. Yapo, Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information, *Water Resour. Res.*, 334(4), 751–763, 1998.

Guzman, A.G. and S.P. Neuman, Field Air Injection Experiments, pp. 52-94 in Rasmussen, T.C., S.C. Rhodes, A. Guzman, and S.P. Neuman, Apache Leap Tuff INTRAVAL Experiments: Results and Lessons Learned, NUREG/CR-6096, U.S. Nuclear Regulatory Commission, Washington, D.C., 1996.

Hannan, E.S., The estimation of the order of an ARMA process, *Annals of Statistics*, 8, 1971–1981, 1980.

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. 4th Intern. Conf. on Calibration and Reliability in Groundwater Modelling (ModelCARE 2002)*, edited by K. Kovar and Z. Hrkal, Charles University, Prague, Czech Republic, 2002.

Hill, C.M., Methods and Guidelines for Effective Model Calibration, U.S. Geological Survey Water-Resources Investigations Report 98-4005, Denver, Colorado, 1998.

Hillel, D., and R. S. Baker, A descriptive-theory of fingering during infiltration into layered soils, *Soil Science*, 146(1), 51–56, 1988.

Hills, R.G. and P.J. Wierenga, INTRAVAL Phase II Model Testing at the Las Cruces Trench Site, NUREG/CR-6063, U.S. Nuclear Regulatory Commission, Washington, D.C., January 1994.
Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky, Bayesian model averaging: A tutorial, *Statist. Sci.*, 14(4), 382–417, 1999.

Hornberger, G.M. and R.C. Speer, An approach to the preliminary analysis of environmental systems, *Jour. Environ. Mgmt.*, 12, 7–18, 1981.

Hunt and Zheng, *Eos Transactions*, AGU, p. 29, January 19, 1999.

Hyun, Y., S.P. Neuman, V.V. Vesselinov, W.A. Illman, D.M. Tartakovsky, and V. Di Federico, Theoretical interpretation of a pronounced permeability scale effect in unsaturated fractured tuff, *Water Resour. Res.*, 38(6), 10.1029/2001WR000658, 2002.

Illman, W.A. and S.P. Neuman, Type-curve interpretation of multi-rate single-hole pneumatic injection tests in unsaturated fractured rock, *Ground Water*, 38(6), 899–911, 2000.

Illman, W.A. and S.P. Neuman, Type-curve interpretation of a cross-hole pneumatic injection test in unsaturated fractured tuff, *Water Resour. Res.*, 37(3), 583-603, 2001.

Illman, W.A., D.L.Thompson, V.V. Vesselinov, G. Chen, and S.P. Neuman, Single- and Cross-Hole Pneumatic Tests in Unsaturated Fractured Tuffs at the Apache Leap Research Site: Phenomenology, Spatial Variability, Connectivity and Scale, NUREG/CR-5559, 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.

James, R.V., J. Rubin, Transport of chloride ion in a water-unsaturated soil exhibiting anion exclusion, *Soil Sci Soc Am J*., 50(5), 1142–1149, 1986.

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.

Kenkel, J. L., *Introductory Statistics for Management and Economics*, PWS-KENT Publishing Company, 1989.

Kolm, K.E. and P.K.M. van der Heijde, Conceptualization and characterization of envirochemical systems, 267-275 in *Calibration and Reliability in Groundwater Modelling - ModelCARE 96*, edited by K. Kovar and P. Van der Heijde, IAHS Publication No. 237, Wallingford, Oxfordshire, UK, 1996.

Kostner, A., Geostatistical and Numerical Analysis of Flow In a Crystalline Rock Mass, M.S. thesis, Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, 1993.

Leij, F.J., W.J. Alves, and M.T. van Genuchten, The UNSODA unsaturated soil hydraulic database, EPA, Ada, Oklahoma, 1996.

Lein, K.J., Environmental Decision Making, Blackwell Science, Malden, Massachusetts, 1997.

Lu, Z., S.P. Neuman, A. Guadagnini, and D.M. Tartakovsky, Conditional moment analysis of steady state unsaturated flow in bounded, randomly heterogeneous soils, *Water Resour. Res.*, 38(4), 10.1029/2001WR000278, 2002.

Maloszewski, P. and A. Zuber, Tracer experiments in fractured rock: Matrix diffusion and validity of model, *Water Resour. Res.*, 29(8), 2723–2735, 1993.

Meyer, P.D. and G.W. Gee, Groundwater conceptual models of dose assessment codes, presented at U.S. NRC Workshop on Ground-Water Modeling Related to Dose Assessment, Rockville, Maryland, June 23–24, 1999.

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. 20555-0001, 1997.

National Research Council, *Ground Water Models: Scientific and Regulatory Applications*, National Academy Press, Washington, D.C., 1990.

National Research Council, *Groundwater & Soil Cleanup*, National Academy Press, Washington, D.C., 1999.

National Research Council, *Research Needs in Subsurface Science*, National Academy Press, Washington, D.C., 2000.

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

Neuman, S.P., Calibration of distributed parameter ground water flow models viewed as a multiple-objective decision process under uncertainty, *Water Res. Res.*, 9(4), 1006–1021, 1973.

Neuman, S.P., Stochastic Continuum Representation of Fractured Rock Permeability as an Alternative to the REV and Fracture Network Concepts, in *Rock Mechanics: Proceedings of the 28th U.S. Symposium*, Tucson, Arizona, edited by I.W. Farmer, J.J.K. Daemen, C.S. Desai, C.E. Glass, and S.P. Neuman, A.A. Balkema, Rotterdam, 533–561, 1987.

Neuman, S.P., A Proposed Conceptual Framework and Methodology for Investigating Flow and Transport in Swedish Crystalline Rocks, Arbetsrapport 88-37, 39 p., SKB Swedish Nuclear Fuel and Waste Management Co., Stockholm, September, 1988.

Neuman, S.P., Universal scaling of hydraulic conductivities and dispersivities in geologic media, *Water Resour. Res.*, 26(8), 1749–1758, 1990.

Neuman, S.P., Eulerian-Lagrangian theory of transport in space-time nonstationary velocity fields: Exact nonlocal formalism by conditional moments and weak approximation, *Water Resour. Res.*, 29(3), 633–645, 1993.

Neuman, S. P., Generalized Scaling of Permeabilities: Validation and Effect of Support Scale, *Geophys. Res. Let.*, 21 (5), 349–352, 1994.

Neuman, S.P., On advective dispersion in fractal velocity and permeability fields, *Water Resour. Res.*, 31(6), 1455–1460, 1995.

Neuman, S.P., Stochastic approach to subsurface flow and transport: A view to the future, pp. 231-241 in Dagan, G. and S.P. Neuman (editors), *Subsurface Flow and Transport: A Stochastic Approach*, Cambridge University Press, Cambridge, United Kingdom, 1997a.

Neuman, S.P., Elicitation Summary, in *Saturated Zone Flow and Transport Expert Elicitation Project*, Rep. WBW 1.2.5.7/SCPB: NA/QA: L, Prepared for U.S. Department of Energy, Yucca Mountain Site Characterization Office, by Geomatrix Consultants and TRW, San Francisco, CA, 1997b.

Neuman, S.P., Accounting for conceptual model uncertainty via maximum likelihood model averaging, 529-534, *Proc. 4th Intern. Conf. on Calibration and Reliability in Groundwater Modelling (ModelCARE 2002)*, edited by K. Kovar and Z. Hrkal, Charles University, Prague, Czech Republic, 2002.

Neuman, S.P. and G. Chen, On instability during immiscible displacement in fractures and porous media, *Water Resour. Res.*, 32(6), 1891–1894, 1996.

Neuman, S.P., and G. de Marsily, Identification of Linear Systems Response by Parametric Programming, *Water Resour. Res.*, 12(2), 253–262, 1976.

Neuman, S.P. and J.S. Depner, Use of Variable-Scale Pressure Test Data to Estimate the Log Hydraulic Conductivity Covariance and Dispersivity of Fractured Granites Near Oracle, Arizona, *J. Hydrology*, 102, 475–501, 1988.

Neuman, S.P. and V. Di Federico, Correlation, flow, and transport in truncated multiscale permeability fields, Chapter 13 in *Scale Dependance and Scale Invariance in Hydrology*, edited by G. Sposito, p. 354–397, Cambridge University Press, 1998.

Neuman, S.P. and A. Guadagnini, A new look at traditional deterministic flow models and their calibration in the context of randomly heterogeneous media, 189-197, Proc. ModelCARE 99, Intern. Conf. On Calibration and Reliability in Groundwater Modelling, ETH, Zürich, Switzeerland, September 1999.

Neuman, S.P. and R. Krzystofowicz, An iterative algorithm for interactive multi-objective programming, *Advances in Water Resources*, 1(1), 1–14, 1977.

Neuman, S.P. and S. Orr, Prediction of Steady State Flow in Nonuniform Geologic Media by Conditional Moments: Exact Nonlocal Formalism, Effective Conductivities and Weak Approximation, *Water Resour. Res.*, 29(2), 341–364, 1993.

Neuman, S.P., and P.A. Witherspoon, Theory of Flow in a Confined Two Aquifer System, *Water Resour. Res.*, 5(4), 803–816, 1969.

Neuman, S.P., and P.A. Witherspoon, Field Determination of the Hydraulic Properties of Leaky Multiple Aquifer Systems, *Water Resour. Res.*, 8(5), 1284–1298, 1972.

Neuman, S.P., D. Tartakovsky, T.C. Wallstrom, and C.L. Winter, Correction to "Prediction of Steady State Flow in Nonuniform Geologic Media by Conditional Moments: Exact Nonlocal Formalism, Effective Conductivities, and Weak Approximation" by S.P. Neuman and S. Orr, *Water Resour. Res.*, 32(5), 1479–1480, 1996.

Neuman, S.P., D.M. Tartakovsky, C. Filippone, O. Amir, and Z. Lu, Deterministic prediction of unsaturated flow in randomly heterogeneous soils under uncertainty without upscaling, in Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media, *Proc. 1997 Intern. Workshop in Riverside, California*, edited by M. Th. Van Genuchten, F.J. Leij, and L. Wu, University of California, Riverside, California, 1351–1365, 1999.

Neuman, S.P., W.A. Illman, V.V. Vesselinov, D.L. Thompson, G. Chen, and A. Guzman, Lessons learned from field studies at the Apache Leap Research Site in Arizona, in *Conceptual Models of Flow and Transport in the Fractured Vadose Zone*, National Academy Press, Washington, D.C., 295–334, 2001.

Ophori, D.U., A. Brown, T.Chan, C.C.Davison, M. Gascoyne, N.W. Scheier, F.W. Stanchell, and D.R.Stevenson, Revised model of regional groundwater flow of the Whiteshell Research Area, Atomic Energy of Canada Limited Report AECL-11435, COG-95-443, 1996.

Pebesma, E. J., and C. G.Wesseling, GSTAT: A program for geostatistical modelling, prediction and simulation. *Computers & Geosciences*, 24(1), 17–31, 1998.

Physics Today, The Constructs of Physics and the Role of Math — Revisited, 91–94, January 1998.

Poeter, E.P. and M.C. Hill, Documentation Of UCODE: A Computer Code For Universal Inverse Modeling, Water Resources Investigations Report 98-4080, U.S. Geological Survey, Lakewood, Colorado, 1998.

Pruess, K., C. Oldendurg, and G. Moridis, TOUGH2 user's guide: Version 2.0, Report LBNL-43134, Lawrence Berkeley National Laboratory, Berkeley, California, 1999.

Rawls, W.J. and D.L. Brakensiek, Prediction of soil water properties for hydrologic modeling, in *Proc. Watershed Management in the Eighties*, Am. Soc. Civil Engin., New York, 293–299, 1985.

Rawls, W.J., D.L. Brakensiek, and K.E. Saxton, Estimation of soil water properties, *Trans. ASAE,* 25, 1316–1320, 1982.

Rockhold, M.L., C.J. Murray, and M.J. Fayer, Conditional simulation and upscaling of soil hydraulic properties, 1391-1402, *Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media, Proc. 1997 Intern. Workshop in Riverside, California*, edited by M. Th. Van Genuchten, F.J. Leij, and L. Wu, University of California, Riverside, California, 1999.

Rossini, C., F. Brega, L. Piro, M. Rovellini, and G. Spotti, Combined geostatistical and dynamic simulations for developing a reservoir management strategy: A case history, *Jour. Petrol. Technology*, 46(11), 979–985, 1994.

Rovellini, M., F. Brega, and C. Monico, Facies related geological model: a reliable method to describe complex reservoirs, *Petroleum Geoscience*, 4 (3), 227–235, 1998.

Rubin, Y., A. Sun, R. Maxwell, and A. Bellin, The concept of block-effective macrodispresivity and a unified approach for grod-scal- and plume-scale-dependent transport, Jour. Fluid Mech., 395, 161–180, 1999.

Sahuiquillo, A., J.E. Capilla, J.J. Gómez-Hernández, and J. Andreu, Conditional simulation of transmissivity fields honouring piezometric head data, pp. 201-214 in *Hydraulic Engineering Software IV*, *Fluid Flow Modeling*, *Vol. II*, edited by W.R. Blair and E. Cabrera, Elsevier, Oxford, 1992.

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: 3. Application to hydrochemical and isotopic data, *Water Resour. Res.*, 25(3), 373–384, 1989b.

Samper-Calvete, F.J. and M.A. Garcia-Vera, Inverse modeling of groundwater flow in the semiarid evporitic closed basin of Los Monegros, Spain, *Hydrogeology Journal*, 6(1), 33–49, 1998.

Samper, J. and J. Molinero, Predictive uncertainty of numerical models of groundwater flow and solute transport, Abstract, AGU 2000 Fall Meeting Program, Washington, D.C., F500, 2000.

Schaap, M.G. and F.J. Leij, Database Related Accuracy and Uncertainty of Pedotransfer Functions, *Soil Science* 163:765–779,1998.

Schaap, M.G. and F.J. Leij, Improved prediction of unsaturated hydraulic conductivity with the Mualem-van Genuchten model, *Soil Sci Soc Am J.*, 64 (3), 843–851, 2000.

Schaap, M.G., F.J. Leij and M. Th. van Genuchten, A bootstrap-neural network approach to predict soil hydraulic parameters. *In*: van Genuchten, M.Th., F.J. Leij, and L. Wu (eds), Proc. Int. Workshop, *Characterization and Measurements of the Hydraulic Properties of Unsaturated Porous Media*, pp.1237–1250, University of California, Riverside, California, 1999.

Scientific Review Group (S.P. Neuman a member; Advisory to the Panel on Atomic Energy of Canada Limited's Concept for Disposal of Canada's Nuclear Fuel Waste), *Addendum to An Evaluation of Environmental Impact Statement on the Concept for Disposal of Canada's Nuclear Fuel Waste*, Ottawa, Canada, September 1996.

Simunek, J., M. Sejna, and M.Th. van Genuchten, The HYDRUS-2D Software Package for Simulating Two-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media, Version 2.0, U.S. Salinity Laboratory, ARS, Riverside, California, 1999.

Steinberg, D. M., and W. G. Hunter, Experimental design: Review and comment, *Technometrics*, 28, 71–97, 1984.

Stevenson, D.R., E.T. Kozak, C.C. Davison, M. Gascoyne, and R.A. Broadfoot, Hydrogeologic characteristics of domains of sparsely fractured rock in the granitic Lac Du Bonnet Batholith, southeastern Manitoba, Canada, Atomic Energy of Canada Limited Report AECL-11558, COG-96-117, 1996.

Sun, N.-Z., *Inverse Problems in Groundwater Modeling*, Kluwer, Dordrecht, The Netherlands, 1994.

Sun, N.-Z. and W. W-G Yeh, Identification of parameter structure in groundwater inverse problem, *Water Resour. Res.*, 21(6), 869–883, 1985.

Tartakovsky, D.M. and S.P. Neuman, Transient effective hydraulic conductivities under slowly and rapidly varying mean gradients in bounded three-dimensional random media, *Water Resour. Res.*, 34(1), 21–32, 1998.

Tartakovsky, D.M. and S.P. Neuman, Extension of Transient flow in bounded randomly heterogeneous domains: 1. Exact conditional moment equations and recursive approximations, *Water Resour. Res.*, 35(6), 1921–1925, 1999.

Tartakovsky, D.M., S.P. Neuman and Z. Lu, Conditional stochastic averaging of steady state unsaturated flow by means of Kirchhoff transformation, *Water Resour. Res.*, 35(3), 731–745, 1999.

Tartakovsky, A.M., S.P. Neuman, and D.M. Tartakovsky, Wetting front propagation in randomly heterogeneous soils, *Heat and Mass Transfer in the Natural Environment: A Tribute to J.R. Philip*, edited by P. Raats, D. Smiles, and A. Warrick, Geophysical Monograph 129, AGU, Washington, D.C., 87-94, 2002a.

Tartakovsky, A.M., S.P. Neuman, and D.M. Tartakovsky, Mean and variance of DNAPL finger development in a saturated, randomly heterogeneous porous medium, 1307-1314, *Computational Methods in Water Resources XIV*, edited by S.M. Hassanizadeh, R.J. Schotting, W.G. Gray, and G.F. Pinder, Elsevier, Amsterdam, The Netherlands, 2002b.

Tellman, B., Arizona Water Information Directory, Water Resources Research Center, College of Agriculture and Life Sciences, University of Arizona, Tucson, 2001.

Thorbjarnarson, K.W., D. Huntley, and J.J. McCarty, Absolute hydraulic conductivity estimates from aquifer pumping and tracer tests in a stratified aquifer, *Ground Water*, 36(1), 87–97, 1998.

Toth, J., A theoretical analysis of groundwater flow in small drainage basins, Jour. Geophys, Res., 68, 4795–4812, 1963.

Trease, H.E., D. George, C.W. Gable, J. Fowler, A. Kuprat, and A. Khamyaseh, The X3D grid generation system, in *Numerical Grid Generation in Computational Fluid Dynamics and Related Fields*, edited by Soni, B. K., J. F. Thompson, H. Hausser, and P. R. Eiseman, Engineering Research Center, Mississippi State Univ. Press, 1996.

Tucci, P., *Water-Level Data Analysis for the Saturated Zone Site-Scale Flow and Transport Model*, *ANL-NBS-HS-000034*, *Rev. 01*, prepared for US DOE by US Geological Survey, December 2001.

van Genuchten, M.Th., A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Am. J.*, 44, 892–898, 1980.

van Genuchten, M. T., and P. J. Wierenga, Mass transfer studies in sorbing porous media, I, Analytical solutions, *Soil Sci Soc Am J.*, 40(4), 473–480, 1976.

Vecchia, A.V. and R.L. Cooley, Simultaneous confidence and predicting intervals for nonlinear regression models, with application to a groundwater flow model, *Water Resour. Res.*, 23(7), 1237–1250, 1987.

Vesselinov, V.V. and S.P. Neuman, Numerical inverse interpretation of single-hole pneumatic tests in unsaturated fractured tuff, *Ground Water*, 39(5), 685–695, 2001.

Vesselinov, V.V., S.P. Neuman, and W.A. Illman, Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff: 1. Methodology and borehole effects, *Water Resour. Res.*, 37(12), 3001–3018, 2001a.

Vesselinov, V.V., S.P. Neuman, and W.A. Illman, Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff: 2. Equivalent parameters, high-resolution stochastic imaging and scale effects, *Water Resour. Res.*, 37(12), 3019–3042, 2001b.

Voss, C. I., SUTRA user's mannual, U.S. Geological Survey, 1984.

Voss C. I., SUTRA additional mannual for version 0690-2D, U.S.Geological Survey, 1990.

Wang, D., O. Amir, and S.P. Neuman, Gaussian finite element closure of steady state unsaturated flow in randomly heterogeneous soils, 1163–1170, *Computational Methods in Water Resources XIV*, edited by S.M. Hassanizadeh, R.J. Schotting, W.G. Gray, and G.F. Pinder, Elsevier, Amsterdam, The Netherlands, 2002.

Wang, W., Uncertainty Analysis of Groundwater Flow and Solute Transport in Unsaturated-Saturated Soil: Maricopa Case, Ph.D. Dissertation, University of Arizona, Tucson, 2002.

Wood, W.W., It's the Heterogeneity!, Editorial, *Ground Water*, 38(1), 1, 2000.

Yao, T., C.J. Mai, P.J. Wierenga, S.P. Neuman, and A.R. Graham, Results of Ponding Infiltration and Tracer Experiments at the Apache Leap Research Site, Arizona, NUREG/CR-XXXX, prepared for the U.S. Nuclear Regulatory Commission, Washington, D.C., 2002.

Yao, T., A. R. Graham, P. J. Wierenga, and S. P. Neuman, Unsaturated Infiltration and Tracer Experiments at the Maricopa Environmental Research Site, Arizona, NUREG/CR-XXXX, 2003.

Ye, M., S.P. Neuman, A. Guadagnini, and D.M. Tartakovsky, Nonlocal and localized analyses of conditional mean transient flow in bounded, randomly nonuniform domains, 1155-1162, *Computational Methods in Water Resources XIV*, edited by S.M. Hassanizadeh, R.J. Schotting, W.G. Gray, and G.F. Pinder, Elsevier, Amsterdam, The Netherlands, 2002.

Young, M.H., P. J. Wierenga, A. W. Warrick, L. L. Hofmann, S. A. Musil, M. Yao, C. J. Mai, Z. Zou, and B. R. Scanlon, Results of field studies at the Maricopa Environmental Monitoring Site, Arizona, pp. 260, U.S. Nuclear Regulatory Commission, Washington, D.C., 1999.

Yu, C., A.J. Zielen, J-J. Cheng, Y.C. Yuan, L.G. Jones, D.J. LePoire, Y.Y. Wang, C.O. Loureiro, E.K. Gnanapragasam, E. Faillace, A. Wallo III, W.A. Williams, and H. Peterson, Manual for implementing residual radioactive material guidelines using RESRAD version 5.0, ANL/EAD/LD-2, Argonne National Laboratory, Argonne, IL, 1993.

Zaslavsky, D. and G. Sinai, Surface hydrology: I –V, *Jour. Hydr. Div.*, ASCE, 107 (HYI), 1–93, 1981.

Zhang, D., *Stochastic Methods for Flow in Porous Media*, Academic Press, 2001.

Zhang, D. and S.P. Neuman, Effect of local dispersion on solute transport in randomly heterogeneous media, *Water Resour. Res.*, 32(9), 2715–2723, 1996.

Zimmerman, D.A., G. De Marsily, C.A. Gotway, M.G. Marietta, C.L. Axness, R.L. Beauheim, R. Bras, J. Carrera, G. Dagan, P.B. Davies, D.P. Gallegos, A. Galli, J. Gómez-Hernández, S.M. Gorelick, P. Grindrod, A.L. Gutjahr, P.K. Kitanidis, A.M. Lavenue, D. McLaughlin, S.P. Neuman, B.S. Rama Rao, 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.

Zyvoloski, G. A., B. A. Robinson, Z. V. Dash, and L. L. Trease, *Summary of the Models and Methods for the FEHM Application — A Finite-Element Heat- and Mass-Transfer Code*, *Tech. Rep. LA-13307-MS*, Los Alamos National Laboratory, New Mexico, 1997.

APPENDIX A

PERFORMANCE MEASURES AND HYDROLOGY ISSUES PERTAINING TO GROUNDWATER FLOW AND TRANSPORT MODELS OF LICENSED NUCLEAR SITES

APPENDIX A: PERFORMANCE MEASURES AND HYDROLOGY ISSUES PERTAINING TO GROUNDWATER FLOW AND TRANSPORT MODELS OF LICENSED NUCLEAR SITES

This note was prepared by the licensing staff in the Office of Nuclear Material Safety and Safeguards (NMSS) for the benefit of the research staff in connection with an ongoing generic research that is being conducted for NRC's Office of Nuclear Regulatory Research (RES) by the University of Arizona. The main purpose of the research is to develop strategy for formulation, development, and application of representative groundwater flow and transport models for licensed sites. More specifically, this information is provided to ensure that the ongoing research efforts identify methods and techniques that can be used by the licensing staff to better analyze uncertainties arid improve the decision-making procedures with regard to the selection and application of site-specific groundwater flow and transport models for licensed sites.

The note provides the following information concerning licensed nuclear facilities, including processing and waste disposal sites: (1) performance measures based on regulatory standards; (2) pertinent hydrology/groundwater issues that are related to performance; and (3) uncertainties commonly encountered in groundwater flow and transport modeling.

The performance measures and pertinent hydrology issues and uncertainties are provided for licensed sites in the following NMSS programs: low-level radioactive waste, decommissioning, high-level radioactive waste, and uranium recovery.

Although a variety of derivative hydrologic measures and hydrologic issues are stated, RES staff and their contractors should keep in mind that the end point of the analysis is the performance measure for the licensed facility obtained through a performance assessment. The hydrologic measures and issues need only be explored and dealt with to the extent required to have confidence in the technical basis supporting licensing decisions, taking into account the significant uncertainties and the hazard involved (note: some uncertainties may be satisfactorily addressed through bounding analyses).

Low-Level Waste Disposal Sites

The performance measure in the current regulations for low-level radioactive waste (LLW) disposal sites is provided by a standard for protection of the population and the environment in 10 CFR 61.41. According to this standard, concentrations of radioactive material which may be released to the general environment in groundwater, surface water, air, soil, plants, or animals must not result in an annual exceeding an equivalent of 25 millirems (mrem) to the whole body, 75 mrem to the thyroid, and 25 mrem to any other organ of any member of the public. In addition, the standard indicates that reasonable effort should be made to maintain releases of radioactivity in effluents to the general environment as low as reasonably is reasonably achievable (ALARA).

Staff views on performance assessment approaches for LLW sites is provided in NUREG-1573. As stated in NUREG-1573 (p. 3-57), the objective of the groundwater flow and transport analyses (including groundwater models) is to assess

concentrations of radionuclides released in the ground water at receptor locations so as to assess the potential annual dose. Staff recommended in NUREG-1573 that the groundwater transport analysis provide concentrations in well water at the site boundary representative of composite concentration of radionuclides for the average member of the critical group (critical group is defined as the group of individuals **reasonably** expected to receive the greatest dose from radioactive releases from the disposal facility over time, given the circumstances under which the analysis would be carried out - NUREG-1537, page 3-76).

Analysis of radionuclide concentrations in the groundwater at specific sites is carried out usually involves addressing such site-specific hydrologic issues as infiltration through the disposal site cover, release of radionuclides from the waste, and flow and radionuclide transport in the unsaturated zone and in the saturated zone to the receptor points.

Uncertainties commonly encountered in the hydrologic analysis of LLW disposal sites include both data as well as conceptual uncertainties. Data uncertainties include the hydraulic properties of the cover (mainly permeability/hydraulic conductivity, service life); the hydraulic properties of the formations beneath the site (mainly permeability/hydraulic conductivity, anisotropy and inhomogeneity, effective porosity); retardation properties and coefficients; and pH values. Conceptual uncertainties may include lateral flow versus vertical flow and possible development of perched water conditions in the unsaturated zone below the disposal facility; matrix versus fracture flow; structural controls on flow and radionuclide transport; inter-aquifer flow; and uncertainty related to changes in time-dependent variables (such as the water levels, concentrations, and pH). Data-related uncertainties are sometimes addressed by bounding analyses.

Decommissioning of Licensed Sites

The performance measure in the current regulations for decommissioning of licensed sites is provided by the standards in 10 CFR 20.1402 (for license termination including unrestricted use of the decommissioned site) and 10 CFR 20.1403 (license termination under restricted conditions for use of the decommissioned site). According to the standards in 10 CFR 20.1402, a site will be considered acceptable for unrestricted use if the residual radioactivity that is distinguishable from background radiation results in a Total Effective Dose Equivalent (TEDE) to an average member of the critical group at receptor locations or human access points that does not exceed 25 mrem per year, including that from groundwater sources of drinking water, and the residual radioactivity has been reduced to levels that are ALARA.

The standards in 10 CFR 20.1403 provide that a site will be considered acceptable for license termination under restricted conditions by satisfying certain provisions specified in the regulations. These include provisions pertaining to meeting the ALARA provision; legally enforceable institutional controls that provide reasonable assurance that the TEDE from the residual radioactivity will not exceed 25 mrem per year; financial assurance to assume and carry out any necessary control and maintenance of the site; submittal of a decommissioning or a license termination plan indicating intent to decommission in accordance with the regulations in IO CFR

Subparts 30.36(d), 40.42(d), 50.82(a) and (b), 70.38(d), or 72.54; and that if the institutional controls were no longer in effect, there is reasonable assurance that the residual activity to the average member of the critical group is ALARA, and would not exceed either 100 mrem per year, or 500 mrem per year provided that the licensee: (1) demonstrates that further reductions in residual radioactivity necessary to comply with the 100 mrem per year are not technically achievable, would be prohibitively expensive, or would result in net public or environmental harm; (2) makes provisions for durable institutional controls; and (3) provides sufficient financial assurance to enable a responsible government entity or independent third party, both to carry out periodic rechecks of the site to assure that the institutional controls remain in place as necessary to provide reasonable assurance that the TEDE from the residual radioactivity will not exceed 25 mrem per year.

Guidance for demonstrating compliance with the current regulations for decommissioning sites is provided in NUREG-1549. This guidance does not explicitly address how the groundwater analysis should be performed. However, guidance is provided for assessment of the dose for an individual located on site (e.g., using water extracted from a well located on the site with an intake point directly beneath the waste area), an individual located off site (e.g., using water extracted from a well located at the site boundary), or both. It should be pointed out, as reflected in NUREG-1549, the NRC recommended dose modeling approach is an iterative approach that involves a screening analysis initially, but eventually includes more site-specific analyses as warranted by the site conditions. The screening approach recommended in another NRC staff document, NUREG-5512 (vol. 1), has a predefined ground-water conceptual model.

Operational environmental monitoring of groundwater, although adequate for its intended purpose, may not be adequate for site characterization and to support dose assessments for decommissioning. As noted in NUREG-1727, "NMSS Decommissioning Standard Review Plan," Section 4.6, "Groundwater," "[T]he information supplied by the licensee should be sufficient to allow the staff to fully understand the types and movement of radioactive material contamination in groundwater at the facility, as well as the extent of this contamination." The actual number, location, and design of monitoring wells depend on the size of the contaminated area, the type and extent of contamination, the background quality, hydrogeologic system, and the objectives of the monitoring program.

As part of the guidance consolidation efforts, NUREG-1556, Vol. 1-20, staff is currently working on the consolidation of Decommissioning License Termination Guidance. As part of this effort, the guidance is being updated to risk inform. This guidance consolidation for decommissioning is planned to be completed within approximately 2 years and will be contained in NUREG-1757.

On October 9, 2002, NRC and EPA signed an MOU which outlined consultation and finality on decommissioning and decontamination of contaminated sites. The MOU requires NRC to seek EPA's expertise when NRC determines that there is radioactive groundwater contamination in excess of EPA's MCLs or for which restricted release or the use of alternate criteria for license termination is contemplated.

Uncertainties commonly encountered in the hydrologic analysis for decommissioning of licensed sites include both data as well as conceptual uncertainties that are similar to those encountered at LLW sites (see LLW Disposal Sites above).

High-Level Waste Disposal Sites

The performance measures for the potential high-level waste (HLW) repository at Yucca Mountain, Nevada in NRC's regulations at 10 CFR Part 63 are based on the EPA standards (40 CFR 197) which provide: (1) an overall performance measure (annual, individual dose limit of 15 mrem TEDE to a reasonably, maximally exposed individual); (2) a performance measure for a specified human intrusion scenario (annual, individual dose limit of 15 mrem TEDE to a reasonably, maximally exposed individual); and (3) separate limits for the protection of ground water (5 pCi/l for radium-226 and radium-228, 15 pCi/liter for gross alpha activity, and 4 mrem/year to the whole body or any organ from combined beta and photon emitting radionulcides). Compliance with these performance measures is demonstrated with a performance assessment using a compliance location of approximately 18 kilometers downgradient from the repository footprint and over a compliance period of 10,000 years. Additionally, the performance assessment is limited to certain conditions (e.g., features, events, and processes estimated to have less than one chance in 10,000 of occurring over 10,000 years are not to be considered in the performance assessment).

Hydrologic processes and flow and transport issues that may need to be considered for the Yucca Mountain site include: infiltration from the ground surface under present and future climates; deep percolation from the root zone into the waste emplacement drift (an unsaturated, fractured and anisotropic zone above the repository); thermal effects on the flow regime, hydraulic and transport properties of formations in the unsaturated zone; flow and transport of radionuclides in the unsaturated zone below the repository; and flow and transport of radionuclides in the saturated zone between the repository and the receptor group location.

Consideration of these complex these issues over the 10,000 year regulatory period is expected to require the evaluation of hydrologic uncertainties through the use of groundwater and transport models of Yucca Mountain.

Uranium Recovery/Tailings Disposal Sites

The performance measure for reviewing uranium mill and tailings sites can be divided into three areas: **Title I** dealing with DOE-remedial action programs of former mill tailings sites; **Title II** dealing with non-DOE mill tailings sites; and ***in situ* leach** (ISL) uranium solution mining sites. In all three areas, concentration limits of specified chemical and radionuclide constituents in groundwater are determined.

Title I

For Title I sites, the performance measures are covered in 40 CFR 192. Specifically Subparts A, B and C of Part 192 provide the regulatory requirements for water resources protection. Guidance for implementing performance of groundwater standards is covered in NUREG-1724 entitled "Draft Standard Review Plan for the Review of DOE Plans for Achieving Regulatory Compliance at Sites With Contaminated Ground Water Under Title I of the Uranium Mill Tailings Radiation Control Act."

Protection of water resources at Title I sites is a process that encompasses two strategies. The first strategy is to prevent or contain the spread of contaminants in the groundwater, surface water, and surrounding lands. The second strategy is to mitigate the threat to public health an the environment from contaminants that have already been mobilized. Performance of these strategies is maintained through groundwater monitoring programs and corrective action. Groundwater models are utilized to predict the fate and transport of the constituents and to design remedial systems. NRC staff utilize NUREG-1724 for specific guidance criteria in site characterization, groundwater monitoring, corrective action of groundwater contamination at Title I uranium mill tailings sites.

The hydrogeologic model plays a role in groundwater monitoring to demonstrate compliance with standards and in detection of contamination, site characterization of contamination, and corrective action. For example, in estimating the risk of a potential receptor (i.e. water supply well, stream) from groundwater contamination as a result of mill activities, models must be sufficiently detailed to provide a technical basis for regulatory decisions and assure the protection of human health and the environment. Factors such as contaminant velocity in groundwater through the geologic media, geochemical conditions, and potential preferential pathways must be factored into the transport of radionuclides and other constituents.

Title II

Regulations for the performance measures and standards for groundwater at Title II uranium mill and tailings sites is covered in 10 CFR Part 40, Appendix A Criterion 5, 7, and 13. Guidance for demonstrating performance for Title II uranium mill and tailings sites can be found in NUREG-1620 entitled "Draft, Rev. 1, Standard Review Plan for the Review of a Reclamation Plan for Uranium Mill Tailings Sites Under Title II of the Uranium Mill Tailings Radiation Control Act." The strategies with Title II sites are similar to Title I. The first strategy is to prevent or contain the spread of contaminants in the groundwater, surface water, and surrounding lands. The second strategy is to mitigate the threat to public health an the environment from contaminants that have already been mobilized. Groundwater protection standards are either: (1) Commission approved background of a constituent in the groundwater, (2) respective values given in 10 CFR Part 40, Appendix A, paragraph 5C [maximum concentration limits] if the constituent is listed in the table and if the background level of the constituent is below the value listed, or (3) an alternate concentration limit established by the Commission. Alternate concentration limits may be approved by the Commission if they do not present a significant hazard to human health or the environment, are as low as reasonably achievable, after considering practicable corrective actions. Groundwater detection programs are used to demonstrate compliance. If groundwater protection standards are exceeded, a corrective action program must be put into operation pursuant to 10 CFR Part 40, Appendix A, Criterion 5D. The objective of the program is to return the hazardous constituent concentration levels in groundwater to the standards.

In proposals for alternate concentration limits, licensee's often use groundwater fate and transport models to demonstrate that a concentration at the point of compliance will

not present a significant hazard at the point of exposure. The licensee can account for degradation of that constituent through the geologic media over time and distance. Staff criteria for reviewing these models can be found in NUREG-1620. Sufficient site characterization is paramount in building a technically adequate model. Uncertainties are encountered in the site characterization process and in the assumption that conditions, as they were characterized, will remain constant for 1000 years (performance period). Extended groundwater monitoring programs are implemented to lower uncertainty and to assure continued protection of human health and the environment.

Once a license for a Title II site is terminated, pursuant to 10 CFR Part 40.28, the site is transferred under general license to the Department of Energy, another Federal agency designated by the President, or a State where the disposal site is located. Groundwater monitoring is detailed in the Long-Term Surveillance Plan.

In Situ Leach (ISL) Uranium Extraction

Guidance for demonstrating compliance for *in situ* leach (ISL) uranium extraction license applications is provided in NUREG-1569. This guidance explicitly addresses the groundwater information and analysis that is specified in Regulatory Guide 3.46 "Standard Format and Content of License Applications, Including Environmental Report, for In Situ Uranium Solution Mining. NUREG-1569 identifies the NRC reviewer's proposed activities in reviewing a licensee submittal, specifically the areas of review, review procedures, acceptance criteria, evaluation findings and references. The ground-water issues in NUREG-1569 relate to ground- water quality restoration. The monitoring programs needed to assure ground-water quality restoration are discussed. The acceptance criteria for the ground-water quality are established based upon the background water quality prior to ISL mining.

NUREG-1569 states that restoration goals are established in the application for each of the monitored constituents. The applicant has the option of determining restoration goals for each constituent on a well-by-well basis, or on a well field average basis. Restoration goals should be established for the ore zone and for any overlying or underlying aquifer that remains affected by ISL solutions. Performance measures for ISL sites can be classified into two groups; primary restoration goals, and secondary restoration goals. For primary restoration standards, the primary goal for a restoration program is to return the water quality of the ore zone and affected aquifers to pre-operational (baseline) water quality. It is unlikely that after restoration activities the ground-water quality will be returned to the exact water quality that existed at every location in the aquifer before ISL operations. Therefore, it is acceptable to use standard statistical methods to set the primary restoration goal and to determine compliance with it. It is also acceptable for the applicant to propose that the baseline conditions for each chemical species be represented by a range of concentrations. The reviewer shall ensure that statistical methods used to determine such confidence intervals are properly applied.
For secondary restoration standards, since the ISL process requires changing the chemistry of the ore zone, it is reasonable to expect that ISL may cause permanent changes in water quality. For this reason, it is acceptable for the applicant to propose, as a secondary restoration standard, returning

the water quality to its pre-ISL class of use (e.g., drinking water, livestock, agricultural, or limited use). Applications should state that secondary standards will not be applied so long as restoration continues to result in significant improvement in ground-water quality. It is acceptable to the staff if, on a constituent-by-constituent basis, secondary goals are determined by applying the lower of the State or EPA secondary and primary drinking water standards.

If a ground-water parameter could not be restored to its secondary goal, an applicant could demonstrate to NRC that leaving the parameter at the higher concentration would not be a threat to public health and safety nor the environment, and that, on a parameter-by-parameter basis, water use would not be significantly degraded. Such proposed alternatives must be evaluated as a license amendment request only after restoration to the primary or secondary standard is shown not to be practical. This approach is consistent with the ALARA philosophy that is used broadly within NRC.

Uncertainties commonly encountered in the hydrologic analysis for Title I, Title II, and ISL sites include both data, as well as, conceptual model uncertainties. The data uncertainties are similar to those encountered in ground-water quality monitoring programs (e.g., sampling methods, well screen location, and laboratory analysis). The uncertainties in the conceptual models are those encountered in the site characterization process, as well as, the process for determining model assumptions used in estimating input parameters, and data analysis of field tests (e.g., pump and pilot study tests) and compliance monitoring.

APPENDIX B

VERTICAL AVERAGING OF WATER-TABLE EQUATION

APPENDIX B: VERTICAL AVERAGING OF WATER-TABLE EQUATION

Transient groundwater flow in a saturated porous medium is governed by the three-dimensional continuity equation

$$-\nabla_3 \cdot \boldsymbol{q} = S_s \frac{\partial h}{\partial t}$$

subject to Darcy's law $\boldsymbol{q} = -\boldsymbol{K}\nabla_3 h$ where $\nabla_3 = \left(\partial/\partial x, \partial/\partial y, \partial/\partial z\right)^T$ is the three-dimensional gradient operator (the superscript *T* indicating transpose), $\boldsymbol{q}$ is flux vector, S_s is specific storage, h is hydraulic head, t is time, and $\boldsymbol{K}$ is a 3 x 3 hydraulic conductivity tensor. The same equation applies at any point within a water-table aquifer. Let the bottom of this aquifer be at elevation $z_o(x,y)$, and the water table at elevation $\xi(x,y,t)$, above some horizontal datum. The same datum serves to define the three-dimensional head function $h(x,y,z,t)$.

Flow in unconfined aquifers is commonly described by a two-dimensional (rather than three-dimensional) flow equation in the horizontal plane. Here we derive this two-dimensional equation from its more comprehensive three-dimensional counterpart by formally averaging the above continuity expression over the vertical,

$$\int_{z_o}^{\xi}\left(\nabla_3 \cdot \boldsymbol{q} + S_s \frac{\partial h}{\partial t}\right) dz = 0$$

From Leibnitz's rule for the differentiation of integrals it follows that

$$\int_{z_o}^{\xi} \frac{\partial q_x}{\partial x} dz = \frac{\partial}{\partial x}\int_{z_0}^{\xi} q_x\, dz - q_x(\xi)\frac{\partial \xi}{\partial x} + q_x(z_o)\frac{\partial z_o}{\partial x}$$

$$\int_{z_o}^{\xi} \frac{\partial h}{\partial t} dz = \frac{\partial}{\partial t}\int_{z_0}^{\xi} h\, dz - h(\xi)\frac{\partial \xi}{\partial t}$$

This allows us to rewrite the vertically-integrated water balance equation as

$$\nabla \cdot \int_{z_o}^{\xi}\left(q_x, q_y\right) dz - \left(q_x, q_y\right)|_{z=\xi} \cdot \nabla \xi$$

$$+\left(q_x, q_y\right)|_{z=z_o} \cdot \nabla z_o + \int_{z_o}^{\xi} \frac{\partial q_z}{\partial z} dz$$

$$+S_s\left[\frac{\partial}{\partial t}\int_{z_o}^{\xi} h\, dz - h(\xi)\frac{\partial \xi}{\partial t}\right] = 0$$

where now $\nabla = \left(\partial/\partial x, \partial/\partial y\right)^T$ is a two-dimensional gradient operator in the horizontal plane, and S_s is considered to be independent of time. We define the vertically-averaged head $\bar{h}$ and horizontal flux $\bar{\boldsymbol{q}} = \left(\bar{q}_x, \bar{q}_y\right)^T$ as

$$\bar{h} = \frac{1}{\xi - z_o}\int_{z_o}^{\xi} h\, dz$$

$$\bar{\boldsymbol{q}} = \frac{1}{\xi - z_o}\int_{z_o}^{\xi}\left(q_x, q_y\right)^T dz$$

Next, we adopt the well-known Dupuit assumption that *flow in the aquifer is virtually horizontal*. This assumption implies that head does not vary much in the vertical direction and therefore $h(\xi) \approx h(z_o) \approx \bar{h}$. We can therefore write the vertically-integrated continuity equation as

$$\nabla \cdot \left[\left(\xi - z_o\right)\bar{\boldsymbol{q}}\right] - \left(q_x, q_y\right)|_{z=\xi} \cdot \nabla \xi$$

$$+\left(q_x, q_y\right)|_{z=z_o} \cdot \nabla z_o + q_z|_{z=\xi} - q_z|_{z=z_o}$$

$$S_s\left(\xi - z_o\right)\frac{\partial \bar{h}}{\partial t} \approx 0$$

where the last term follows from

$$\frac{\partial}{\partial t}\int_{z_o}^{\xi} h\, dz - h(\xi)\frac{\partial \xi}{\partial t} \approx \frac{\partial}{\partial t}\int_{z_o}^{\xi} h\, dz - \bar{h}\frac{\partial\left(\xi - z_o\right)}{\partial t}$$

$$= \frac{\partial}{\partial t}\int_{z_o}^{\xi} h\,dz - \frac{1}{\xi - z_o}\frac{\partial(\xi - z_o)}{\partial t}\int_{z_o}^{\xi} h\,dz$$

$$= (\xi - z_o)\frac{\partial}{\partial t}\left[\frac{1}{\xi - z_o}\int_{z_o}^{\xi} h\,dz\right] = (\xi - z_o)\frac{\partial \bar{h}}{\partial t}$$

We treat the water table as a free moving upper boundary of the saturated flow region, at which gauge pressure is zero (atmospheric) and therefore head *h* at the water table is equal to the elevation *z*. From the Lagrangian viewpoint of a fluid particle that moves with the water table, the quantity *h* - *z* remains fixed at zero. In other words, the Lagrangian derivative of *h* - *z* is zero,

$$\frac{D}{Dt}(h - z)|_{z=\xi}$$

$$\equiv \left[\frac{\partial h}{\partial t} + q_{sx}\frac{\partial h}{\partial x} + q_{sy}\frac{\partial h}{\partial y} + q_{sz}\left(\frac{\partial h}{\partial z} - 1\right)\right]_{z=\xi}$$

$$= 0$$

where $\boldsymbol{q}_s$ is the (seepage) velocity of the water table. By virtue of the Dupuit assumption, we can disregard $\partial h/\partial z$ in this expression and replace *h* in all the remaining terms by ξ to yield

$$\left(\frac{\partial \xi}{\partial t} + q_{sx}\frac{\partial \xi}{\partial x} + q_{sy}\frac{\partial \xi}{\partial y} - q_{sz}\right)_{z=\xi} = 0$$

Water balance at the water table requires that $(\boldsymbol{q} - \boldsymbol{r})\cdot\boldsymbol{n} = S_y\boldsymbol{q}_s\cdot\boldsymbol{n}$ where $\boldsymbol{r}$ is a vector representing recharge from above, and S_y is specific yield or drainable porosity. It is common to assume that <u>recharge is strictly vertical</u> so that $\boldsymbol{r} = (0, 0, -I)$ where *I* the rate of net vertical infiltration. Water balance at the free boundary then implies that $q_x = S_y q_{sx}$, $q_y = S_y q_{sy}$ and $q_z + I = S_y q_{sz}$. Multiplying the above equation throughout by S_y therefore yields

$$\left[S_y\frac{\partial \xi}{\partial t} + q_x\frac{\partial \xi}{\partial x} + q_y\frac{\partial \xi}{\partial y} - (q_z + I)\right]_{z=\xi} = 0$$

which can be rewritten as

$$(q_x, q_y)|_{z=\xi}\cdot\nabla\xi - q_z|_{z=\xi} = I - S_y\frac{\partial \xi}{\partial t}$$

Substituting this into the vertically-averaged continuity equation and rearranging yields

$$-\nabla\cdot\left[(\xi - z_o)\bar{\boldsymbol{q}}\right] + (q_x, q_y)|_{z=z_o}\cdot\nabla z_o + I$$

$$+ q_z|_{z=z_o} \approx \left[S_s(\xi - z_o) + S_y\right]\frac{\partial \bar{h}}{\partial t}$$

So far the vertical averaging has not affected the three-dimensional Darcy equation $\boldsymbol{q} = -\boldsymbol{K}\nabla_3 h$. It is important to note that setting $h \approx \bar{h}$ in accord with the Dupuit assumption eliminates the vertical derivative of head, but does not eliminate the vertical component of flux. For the vertical flux to vanish we must assume that **K** has a principal direction parallel to z. Then the vertical averaging of $\boldsymbol{q} = -\boldsymbol{K}\nabla_3 h$ yields a two-dimensional form of Darcy's law,

$$\bar{\boldsymbol{q}} = -\boldsymbol{K}_H\nabla\bar{h}$$

where $\boldsymbol{K}_H$ is an equivalent hydraulic conductivity tensor given by

$$\boldsymbol{K}_H = \frac{1}{\xi - z_o}\int_{z_o}^{\xi}\boldsymbol{K}\,dz$$

Note that $\boldsymbol{K}_H$ is a 2 x 2 tensor defined in the horizontal plane, which does not depend on *z* but may vary with the horizontal coordinates *x* and *y*. Note further that $\boldsymbol{K}_H$ is the arithmetic average of $\boldsymbol{K}$ over the saturated thickness $\xi - z_o$ of the aquifer. As the saturated thickness may vary with time, so can $\boldsymbol{K}_H$, even if $\boldsymbol{K}$ is a constant. The *variability of $\boldsymbol{K}_H$ with saturated thickness is commonly ignored*, even though it may be significant in heterogeneous (and especially stratified) aquifers.

If the *principal values of* $\boldsymbol{K}$ *in the horizontal plane do not vary strongly with elevation*, then the Dupuit assumption implies that $\left(q_x, q_y\right)^T |_{z=z_o} \approx \overline{\boldsymbol{q}}$. The vertically-averaged continuity equation can therefore be written as

$$-\nabla \cdot \left[\left(\xi - z_o \right) \overline{\boldsymbol{q}} \right] + \overline{\boldsymbol{q}} \cdot \nabla z_o + I + q_z |_{z=z_o}$$

$$\approx \left[S_s \left(\xi - z_o \right) + S_y \right] \frac{\partial \overline{h}}{\partial t}$$

Here $\left(\xi - z_o \right) \overline{\boldsymbol{q}}$ is horizontal flow rate across the entire saturated thickness of the aquifer per unit transverse horizontal distance, $\overline{\boldsymbol{q}} \cdot \nabla z_o$ is a term that accounts for variations in the elevation of the aquifer bottom, I is net vertical recharge at the water table, $q_z |_{z=z_o}$ is rate of leakage across the aquifer bottom, $S_s(\xi - z_o)$ is an internal (artesian) storage coefficient or storativity that depends on saturated thickness, and the specific yield S_y is storativity due to imbibition or drainage at the water as the latter rises or falls.

Introducing the vertically-averaged form of Darcy's law into the above expression, and replacing ξ by $\overline{h}$, yields the vertically-averaged flow equation

$$\nabla \cdot \left[\left(\overline{h} - z_o \right) \boldsymbol{K}_H \nabla \overline{h} \right] - \boldsymbol{K}_H \nabla \overline{h} \cdot \nabla z_o + I + q_z |_{z=z_o}$$

$$\approx \left[S_s \left(\overline{h} - z_o \right) + S_y \right] \frac{\partial \overline{h}}{\partial t}$$

which is seen to be nonlinear in the vertically-averaged head $\overline{h}$. Alternatively, one can replace $\overline{h}$ by ξ to obtain a similar equation in terms of water-table elevation,

$$\nabla \cdot \left[\left(\xi - z_o \right) \boldsymbol{K}_H \nabla \xi \right] - \boldsymbol{K}_H \nabla \xi \cdot \nabla z_o + I + q_z |_{z=z_o}$$

$$\approx \left[S_s \left(\xi - z_o \right) + S_y \right] \frac{\partial \xi}{\partial t}$$

The quantity $\left(\overline{h} - z_o \right) \boldsymbol{K}_H$ or $\left(\xi - z_o \right) \boldsymbol{K}_H$ is commonly referred to as transmissivity.

It is common to *set* $z_o \equiv 0$ and *disregard artesian storage*. Then one obtains the well-known Boussinesq equation either in terms of average head,

$$\nabla \cdot \left(\overline{h} \boldsymbol{K}_H \nabla \overline{h} \right) + I + q_z |_{z=z_o} \approx S_y \frac{\partial \overline{h}}{\partial t}$$

or in terms of water-table elevation,

$$\nabla \cdot \left(\xi \boldsymbol{K}_H \nabla \xi \right) + I + q_z |_{z=z_o} \approx S_y \frac{\partial \xi}{\partial t}$$

Our formal averaging of flow in an unconfined aquifer was accompanied by a number of simplifying assumptions. These state that

1. Flow in the aquifer is essentially horizontal.
2. Recharge above the water table is essentially vertical.
3. One principal direction of the three-dimensional hydraulic conductivity tensor $\boldsymbol{K}$ is vertical.
4. The horizontal components of $\boldsymbol{K}$ vary weakly with elevation (this is needed only if the aquifer bottom is uneven and/or one treats $\boldsymbol{K}_H$ as being independent of saturated thickness).

The standard form of the Boussinesq equation assumes additionally that

5. The aquifer bottom is horizontal and flat.
6. Artesian storage is negligible.

When these assumptions are violated, a two-dimensional description of flow in an unconfined aquifer in the horizontal plane may lead to a systematic bias in predicted heads and fluxes.

APPENDIX C

AVERAGING OF GROUNDWATER FLOW IN PROBABILITY SPACE

APPENDIX C: AVERAGING OF GROUNDWATER FLOW IN PROBABILITY SPACE

A new look at traditional deterministic flow models and their calibration
in the context of randomly heterogeneous media

SHLOMO P. NEUMAN
Department of Hydrology and Water Resources, The University of Arizona, Tucson, AZ, USA

ALBERTO GUADAGNINI
D.I.I.A.R., Politecnico di Milano, Piazza L. da Vinci, 32, 20133 Milano, Italy

Abstract We examine the tradition of modelling subsurface flow deterministically from a stochastic viewpoint. In particular, we show that traditional deterministic flow equations do apply to randomly heterogeneous media, albeit in an approximate manner, provided they are interpreted in a non-traditional manner. Our paper explains why parameter estimates obtained by traditional inverse methods tend to vary as one modifies the database. It also makes clear that the traditional Monte Carlo method of assessing uncertainty in the output of a calibrated deterministic model generally overestimates the predictive capabilities of the model. The only valid way to assess predictive uncertainty is by means of a stochastic model.

INTRODUCTION

Parameters such as hydraulic conductivity have been traditionally viewed as well-defined local quantities that can be assigned unique values at each point in space. Yet in practice they are deduced from measurements at selected well locations and depth intervals where their values depend on the scale and mode of measurement. Quite often, the support of the measurements is uncertain and the data are corrupted by experimental and interpretative errors. Estimating the parameters at points where measurements are not available entails an additional random error. It is therefore appropriate to think of the subsurface as being randomly heterogeneous and to describe fluid flow in geologic media by means of stochastic equations.

Yet, the tradition has been to model subsurface flow deterministically. In this paper we examine this tradition from a stochastic viewpoint and comment on its validity and meaning. More specifically, we ask what is the meaning of hydrogeologic variables that enter into traditional deterministic groundwater flow models when the latter are applied to randomly heterogeneous media? What is the significance of parameters obtained by calibrating a deterministic flow model against randomly varying data? What do corresponding parameter estimation errors imply about predictive uncertainty?

To address these questions, we consider steady state flow of groundwater in a randomly nonuniform domain, Ω. The flux $\boldsymbol{q}(\boldsymbol{x})$ and the hydraulic head $h(\boldsymbol{x})$ obey the continuity equation and Darcy's law, subject to appropriate boundary conditions. All parameters and state variables are defined on a consistent nonzero support volume, ω, which is small in comparison to Ω, but sufficiently large for Darcy's law to be locally valid. It has been shown (Neuman & Orr, 1993; Neuman *et al*., 1996; Guadagnini £ Neuman, 1999a, b; Tartakovsky & Neuman, 1998, 1999) that it is theoretically possible, and computationally feasible, to render optimum unbiased predictions of $h(\boldsymbol{x})$ and $\boldsymbol{q}(\boldsymbol{x})$ under ubiquitously nonuniform and uncertain field conditions by means of their first ensemble (statistical) moments (expected or mean values), $\langle h(\boldsymbol{x})\rangle_c$ and $\langle \boldsymbol{q}(\boldsymbol{x})\rangle_c$, conditioned on measurements of $K(\boldsymbol{x})$. The predictors $\langle h(\boldsymbol{x})\rangle_c$ and $\langle \boldsymbol{q}(\boldsymbol{x})\rangle_c$ satisfy the equations

$$-\langle \boldsymbol{q}(\boldsymbol{x})\rangle_c + \langle f(\boldsymbol{x})\rangle = 0 \qquad \boldsymbol{x} \in \Omega \tag{1}$$

$$\langle \boldsymbol{q}(\boldsymbol{x})\rangle_c = -\langle K(\boldsymbol{x})\rangle_c \nabla \langle h(\boldsymbol{x})\rangle_c + \boldsymbol{r}_c(\boldsymbol{x}) \qquad \boldsymbol{r}_c(\boldsymbol{x}) = -\langle K'(\boldsymbol{x})\, \nabla h'(\boldsymbol{x})\rangle_c \tag{2}$$

subject to the boundary conditions

$$\langle h(\boldsymbol{x})\rangle_c = \langle H(\boldsymbol{x})\rangle;\ \boldsymbol{x} \in \Gamma_D \qquad -\langle \boldsymbol{q}(\boldsymbol{x})\rangle_c \cdot \mathbf{n}(\boldsymbol{x}) = \langle Q(\boldsymbol{x})\rangle;\ \boldsymbol{x} \in \Gamma_N \tag{3}$$

where the subscript c implies "conditional"; primed quantities represent random fluctuations about (conditional) mean values; $K(\boldsymbol{x})$ is a random field of scalar hydraulic conductivities; $\boldsymbol{r}_c(\boldsymbol{x})$ is a residual flux; $\langle f(\boldsymbol{x})\rangle$, $\langle H(\boldsymbol{x})\rangle$, $\langle Q(\boldsymbol{x})\rangle$ are prescribed unconditional first moments of the statistically independent random source and boundary forcing terms $f(\boldsymbol{x})$, $H(\boldsymbol{x})$, $Q(\boldsymbol{x})$; and $\mathbf{n}(\boldsymbol{x})$ is a unit outward normal to $\Gamma = \Gamma_D \cup \Gamma_N$ where Γ_D and Γ_N are Dirichlet and Neumann boundaries, respectively. The residual flux $\boldsymbol{r}_c(\boldsymbol{x})$ is given implicitly by (Neuman *et al*., 1996)

$$\boldsymbol{r}_c(\boldsymbol{x}) = \int_\Omega \boldsymbol{a}_c(\boldsymbol{y}, \boldsymbol{x})\, \nabla_y \left\langle h(\boldsymbol{y})\right\rangle_c d\boldsymbol{y} + \int_\Omega \boldsymbol{d}_c(\boldsymbol{y}, \boldsymbol{x})\, \boldsymbol{r}_c(\boldsymbol{y})\, d\boldsymbol{y} \tag{4}$$

where the kernels

$$\boldsymbol{a}_c(\boldsymbol{y}, \boldsymbol{x}) = < K'(\boldsymbol{x})\, K'(\boldsymbol{y}) \nabla_x \nabla_y^{\mathrm{T}} G(\boldsymbol{y}, \boldsymbol{x}) >_c \tag{5}$$

$$\boldsymbol{d}_c(\boldsymbol{y}, \boldsymbol{x}) = < K'(\boldsymbol{x})\, \nabla_x \nabla_y^{\mathrm{T}} G(\boldsymbol{y}, \boldsymbol{x}) >_c \tag{6}$$

form a symmetric and a non-symmetric tensor, respectively. Here G is a random Green's function, or the solution of the random flow equations for the case where $f(\boldsymbol{x})$ is a point source of unit strength at point $\boldsymbol{y}$, subject to homogeneous boundary conditions $H(\boldsymbol{x}) \equiv Q(\boldsymbol{x}) \equiv 0$.

Due to the integro-differential nature of $\boldsymbol{r}_c(\boldsymbol{x})$, our conditional moment equations include nonlocal parameters that depend on more than one point in space (hence the equations are referred to as nonlocal). The traditional concept of an REV (representative elementary volume) is neither necessary nor relevant for their validity or application. The corresponding parameters are nonunique in that they depend not only on medium properties but also on the information one has about these properties (scale, location, quantity, and quality of data). It is clear that the flux predictor is generally nonlocal and non-Darcian, in that it depends on the residual flux $\boldsymbol{r}_c(\boldsymbol{x})$.

Guadagnini & Neuman (1999a, b) have developed corresponding integro-differential equations for the conditional variance-covariance of associated prediction errors in head and flux, and have shown how to solve both sets of equations by finite elements. Their solution entails expansion of the otherwise exact nonlocal moment equations in terms of a small parameter, σ_Y, which represents a measure of the standard deviation of natural log hydraulic conductivity, $Y(\boldsymbol{x}) = \ln K(\boldsymbol{x})$.

Since the flux predictors $\boldsymbol{r}_c(\boldsymbol{x})$ and $\langle \boldsymbol{q}(\boldsymbol{x})\rangle_c$ are generally nonlocal and non-Darcian, the traditional notion of effective conductivity looses meaning in the context of flow prediction by means of conditional ensemble mean quantities. In the following section we show how these quantities can be localised in an approximate manner, so as to render them Darcian.

LOCALIZATION OF CONDITIONAL MEAN FLUXES

If one treats the conditional mean hydraulic gradient $\nabla\langle h(\boldsymbol{y})\rangle_c$, and residual flux $\boldsymbol{r}_c(\boldsymbol{y})$, in (4) as if they were locally uniform, one can take them outside the corresponding integrals and express them as functions of $\boldsymbol{x}$. One can then express the residual flux in localised, Darcian form as

$$\boldsymbol{r}_c(\boldsymbol{x}) \approx -\boldsymbol{\kappa}_c(\boldsymbol{x})\,\nabla\langle h(\boldsymbol{x})\rangle_c \tag{7}$$

$$-\boldsymbol{\kappa}_c(\boldsymbol{x}) = \left[1 - \int_\Omega \boldsymbol{d}_c(\boldsymbol{y},\boldsymbol{x})\,d\boldsymbol{y}\right]^{-1} \int_\Omega \boldsymbol{a}_c(\boldsymbol{y},\boldsymbol{x})\,d\boldsymbol{y} \tag{8}$$

This leads to a familiar looking Darcian form for the mean flow predictor

$$\langle \boldsymbol{q}(\boldsymbol{x})\rangle_c \approx -\boldsymbol{K}_c(\boldsymbol{x})\,\nabla\langle h(\boldsymbol{x})\rangle_c \qquad \boldsymbol{K}_c(\boldsymbol{x}) = \langle K(\boldsymbol{x})\rangle_c\,\mathbf{I} + \boldsymbol{\kappa}_c(\boldsymbol{x}) \tag{9}$$

where $\mathbf{I}$ is the identity tensor and $\boldsymbol{K}_c(\boldsymbol{x})$ is a spatially-varying conditional effective hydraulic conductivity tensor. The latter is either symmetric positive definite or nonsymmetric, depending on whether $\boldsymbol{\kappa}_c(\boldsymbol{x})$ is symmetric or nonsymmetric (as $\boldsymbol{a}_c$ in (5) is symmetric while $\boldsymbol{d}_c$ in (6) is nonsymmetric, $\boldsymbol{\kappa}_c$ is symmetric only when $\boldsymbol{d}_c \equiv \mathbf{0}$).

Since $\boldsymbol{K}_c(\boldsymbol{x})$ is defined on the support scale ω, it is a local rather than an upscaled effective parameter and does not depend in any way on a grid of any kind. It constitutes a biased estimate of the actual ω-scale hydraulic conductivity, $K(\boldsymbol{x})$, which is represented more faithfully by its unbiased conditional mean value, $\langle K(\boldsymbol{x})\rangle_c$. Nevertheless, it is $\boldsymbol{K}_c(\boldsymbol{x})$ rather than $\langle K(\boldsymbol{x})\rangle_c$ that renders (more-or-less) unbiased predictions of head and flux, provided that the assumptions of local uniformity of $\nabla\langle h\rangle_c$ and $\boldsymbol{r}_c$ are approximately satisfied.

We see that the localisation of flow in a randomly nonuniform medium, under the influence of uncertain source and boundary terms, yields a deterministic system of differential equations, which is identical in form to traditional flow equations commonly written for fully deterministic systems. We can therefore comment about the widespread practice of applying such traditional equations to groundwater flow under ubiquitously nonuniform and uncertain field conditions. First, such applications are valid at best as an approximation, since it is clear that localisation of the mean flux introduces a modelling (localisation) error. Second, they are valid only provided one interprets the traditional equations (and associated analytical or numerical models of groundwater flow) in a non-traditional way. Whereas traditionally the hydraulic conductivity in deterministic flow equations has been considered to be a material property, we see that it is instead a nonunique effective parameter which depends on how closely the underlying property K (which enters into the stochastic, but not the deterministic, flow equations) has been defined on the basis of available measurements (*i.e.*, on their quantity, quality and scale ω). Whereas traditionally the hydraulic head and flux in deterministic flow equations have been considered to represent real system states, we see that they are in fact conditional predictors of these states. As such, they are data-dependent and therefore nonunique, relatively smooth, and differ generally from their true random counterparts. Calibrating a deterministic groundwater flow model against measured values of system states is tantamount to conditioning the model on such measurements. It is therefore easy to understand why the resulting parameters tend to vary continuously as more and more data are incorporated into the (calibration, inverse or parameter estimation) process: the very act of expanding the database alters the (data-dependent, and thus inherently nonunique effective) parameters which the process is designed to estimate. While nonlocal moment equations provide information about predictive uncertainty, localisation does not.

CONDITIONING ON HYDRAULIC CONDUCTIVITY DATA

Our nonlocal theory assumes that one has at his/her disposal a conditional unbiased estimate, $\langle K(\boldsymbol{x})\rangle_c$, of the randomly varying hydraulic conductivity function, $K(\boldsymbol{x})$, together with the second conditional moment of associated estimation errors of log hydraulic conductivities, $C_{Yc}(\boldsymbol{x}, \boldsymbol{y})$. When conditioning is performed on the basis of existing ω-scale measurements of K at a set of discrete points, $\langle K(\boldsymbol{x})\rangle_c$ and $C_{Yc}(\boldsymbol{x}, \boldsymbol{y})$ can be obtained (in principle) by means of geostatistical methods (*e.g.*, Deutsch & Journel, 1998). Geostatistical techniques can also be used to generate multiple conditional realisations of $K(\boldsymbol{x})$, which honour the measured data. Solving the stochastic flow equations for each of these realisations by conditional Monte Carlo simulation (CMS) allows one to calculate $\nabla\langle h(\boldsymbol{x})\rangle_c$ and $\langle \boldsymbol{q}(\boldsymbol{x})\rangle_c$. Based on these, one can use the localised equation (9) to evaluate $\boldsymbol{K}_c(\boldsymbol{x})$. The latter can also be evaluated directly by considering $K(\boldsymbol{x})$ to be multivariate log-normal and using stochastically derived formulae to approximate $\boldsymbol{K}_c(\boldsymbol{x})$. For example, if one treats $K(\boldsymbol{x})$ in a two-dimensional domain as if it was log-normal as well as statistically homogeneous and isotropic in the neighbourhood of $\boldsymbol{x}$, and the conditional mean flow as if it was locally uniform, one can use existing stochastic theory for flow in an infinite domain to deduce that $\boldsymbol{K}_c(\boldsymbol{x}) \approx K_G(\boldsymbol{x})\mathbf{I}$ where $K_G(\boldsymbol{x})$ is the (conditional) geometric mean of K at $\boldsymbol{x}$. In addition to localisation error, this introduces another modelling error due to the analytical approximation of $\boldsymbol{K}_c$.

We illustrate below the effect that these two modelling errors have on the flow predictors $\langle h(\boldsymbol{x})\rangle_c$ and $\langle \boldsymbol{q}(\boldsymbol{x})\rangle_c$ in a two-dimensional square domain, under superimposed mean uniform and convergent flows, subject to deterministic boundary conditions. For comparison purpose, we compute these conditional first moments by three methods: (a) CMS, which we consider to be the most reliable and accurate among the three methods; (b) recursive nonlocal finite elements, as proposed by Guadagnini & Neuman (1999a, b), which provide a direct solution to the stochastic flow problem to second order of approximation in σ_Y; and (c) localisation, by setting $\boldsymbol{K}_c(\boldsymbol{x}) \approx K_G(\boldsymbol{x})\mathbf{I}$. The flow domain measures $L_1 = L_2 = 10$ in arbitrary consistent length units. It is discretized into 50 x 50 rows and columns of $M = 2500$ square elements with sides $\Delta x_1 = \Delta x_2 = 0.2$. Deterministic no-flow boundary conditions are imposed at the bottom along $x_2 = 0$, and at the top along $x_2 = 10$. A uniform deterministic head $H_L = 10$ is prescribed on the left boundary along $x_1 = 0$, and $H_R = 0$ on the right boundary along $x_1 = 10$. We consider three test problems: TP1 without conditioning on hydraulic conductivity data, and TP2 and TP3 that involve conditioning on two different sets of K measurements. In all three test problems, the underlying (unconditional) log hydraulic conductivity field $Y(\boldsymbol{x})$ is multivariate Gaussian, statistically homogeneous and isotropic with an exponential spatial autocovariance.

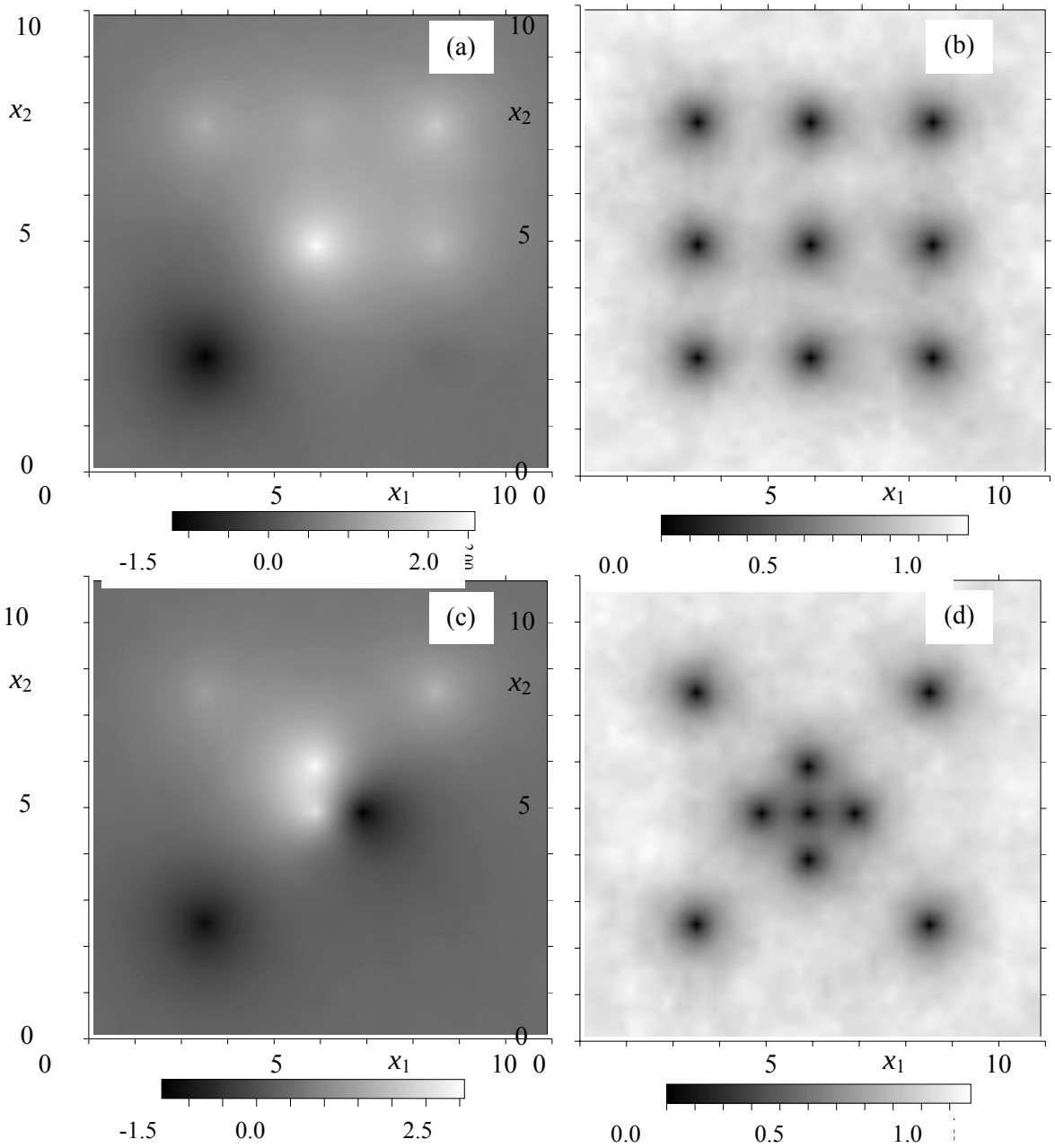


Fig. 1 Images (a) and (c) of $\langle Y(\boldsymbol{x})\rangle_c$, (b) and (d) of $\sigma_Y^2(\boldsymbol{x})$, obtained with NMC=2,000 for TP2 in (a) and (b), TP3 in (c) and (d).

Overall, the flow is uniform in the mean from left to right. We superimpose convergent flow on this mean uniform background by placing a well (point sink) at the centre ($x_1 = 5$, $x_2 = 5$) of the grid, which pumps at a volumetric flow rate of $Q = 5$ (measured in arbitrary consistent time units).

Unconditional random $Y(\boldsymbol{x})$ fields are generated by means of the simulator SGSIM (Deutsch & Journel, 1998), with unit log conductivity variance σ_Y^2 and autocorrelation scale λ. Figure 1 shows images of the conditional sample mean, $\langle Y(\boldsymbol{x})\rangle_c$, and variance, $\sigma_Y^2(\boldsymbol{x})$, obtained with NMC = 2000 Monte Carlo simulations for TP2 and TP3. The "measured" hydraulic conductivities for these test problems are extracted from a single unconditional random realisation of $Y(\boldsymbol{x})$. For purposes of flow analysis by CMS, we assign to each element in the grid a constant Y value, corresponding to the point value generated at the element centre. This is justified considering that our grid includes a minimum of five such cells per autocorrelation scale. The conditional $K_G(\boldsymbol{x})$ field, which plays the role of an effective parameter field for the localised mean flow equations, is obtained by averaging all conditionally generated Y realisation for each test problem.

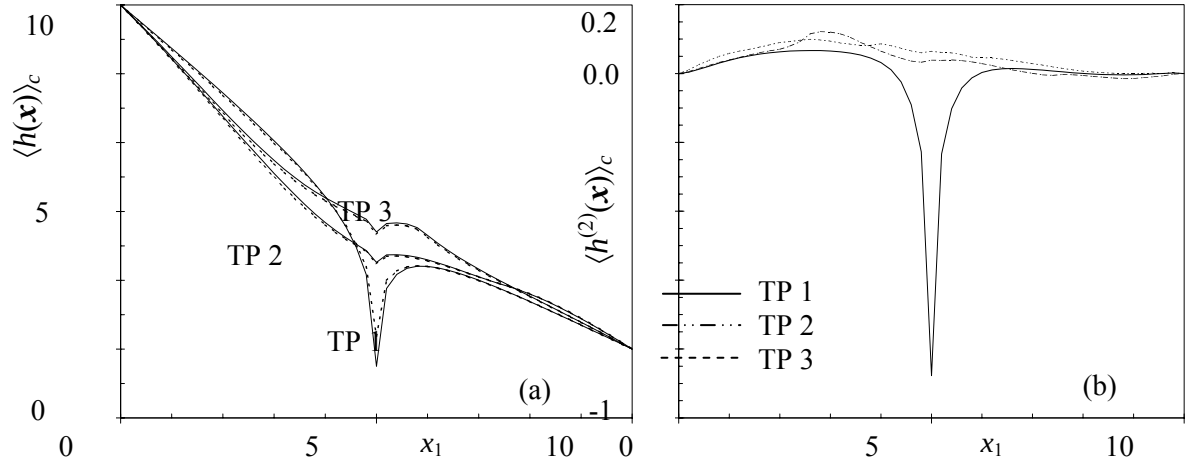


Fig. 2 (a) Nonlocal (solid, same as CMS) and localised (dashed) solutions for $\langle h(\boldsymbol{x})\rangle_c$ along longitudinal section (x_2 = 5) through well; (b) second-order component, $\langle h^2(\boldsymbol{x})\rangle_c$, of nonlocal $\langle h(\boldsymbol{x})\rangle_c$ solution along same section.

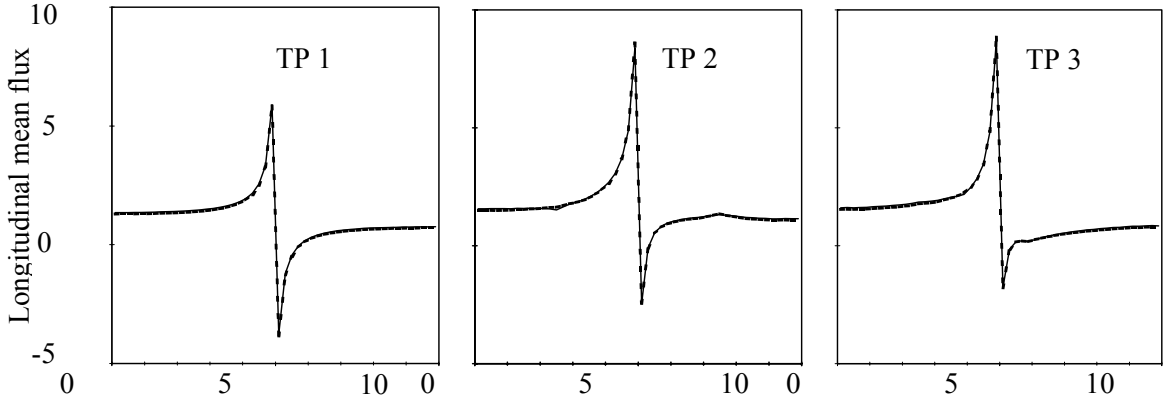


Fig. 3 Nonlocal (solid, same as CMS) and localised (dashed) solutions for longitudinal mean flux along longitudinal section (x_2 = 4.9).

Figure 2(a) depicts the conditional mean head, $\langle h(\boldsymbol{x})\rangle_c$, along a longitudinal section through the well for the three test problems. Values obtained by CMS and the nonlocal approach are virtually indistinguishable from each other, hence only nonlocal results are shown. Figure 2(b) shows the second-order (in σ_Y) component, $\langle h^{(2)}(\boldsymbol{x})\rangle_c$, of the nonlocal $\langle h(\boldsymbol{x})\rangle_c$ solution along the same section for the three test problems.

Figure 3 compares values of longitudinal mean flux, as obtained by the nonlocal and localised methods, along a longitudinal section passing through elements sharing the well node. Values obtained by the nonlocal and CMS approaches are virtually indistinguishable from each other, hence only nonlocal results are shown.

Both the nonlocal and localised conditional mean head solutions agree very well with those obtained by CMS. The localised results are slightly less accurate, especially at the well in the unconditional case, due to errors introduced through localisation and the approximation of $\boldsymbol{K}_c(\boldsymbol{x})$ by $K_G(\boldsymbol{x})\mathbf{I}$.

The incorporation of second-order terms in our nonlocal solution adds to its accuracy in all test problems. Figure 2(a) shows that the second-order head correction is negative at the well in the unconditional case (TP1), but remains positive in TP2 and TP3, in both of which the well acts as a conditioning point. It is worth noting that the "measured" K at the well is somehow larger than the unconditional mean conductivity $\langle K(\boldsymbol{x})\rangle$. Though both the nonlocal and localised solutions compare well with CMS results in all test problems, the nonlocal results do so uniformly better than the localised ones.

CONDITIONING ON HEAD DATA THROUGH MODEL CALIBRATION

$\boldsymbol{K}_c(\boldsymbol{x})$ in our localised equations can, in principle, be estimated by means of standard inverse methods. This is analogous to the common groundwater modelling practice of taking the standard deterministic flow equations for granted, and estimating the associated hydraulic conductivities by model calibration against measured head and (possibly) fluxes data. Such calibration is tantamount to conditioning the localised equations not only on measured values of hydraulic conductivity, as we did previously, but also on those of head and (possibly) flux. A typical goal of inverse modelling in this context would be to obtain a suitably parameterised estimate, $\hat{\boldsymbol{K}}_c(\boldsymbol{x})$, of $\boldsymbol{K}_c(\boldsymbol{x})$ so as to minimise, in some sense, the difference between computed and actual head values, $\langle \hat{h}(\boldsymbol{x}_i)\rangle_c$ and $h(\boldsymbol{x}_i)$, at a discrete set of measurement points, $\boldsymbol{x}_i$. If measurement of flux, $\boldsymbol{q}(\boldsymbol{y}_j)$, are available at discrete points $\boldsymbol{y}_j$, an equally important goal should be to minimise the difference between $\boldsymbol{q}(\boldsymbol{y}_j)$ and their computed counterparts, $\langle \hat{\boldsymbol{q}}(\boldsymbol{x}_j)\rangle_c = -\hat{\boldsymbol{K}}_c(\boldsymbol{x}_j)\nabla\langle \hat{h}(\boldsymbol{x}_j)\rangle_c$. Yet another goal could be to keep the parameter estimate, $\hat{\boldsymbol{K}}_c(\boldsymbol{x})$, close in some sense to a prior estimate, $\boldsymbol{K}_c^*(\boldsymbol{x})$, of $\boldsymbol{K}_c(\boldsymbol{x})$. In the examples considered earlier, a suitable prior might be $K_G(\boldsymbol{x})\mathbf{I}$. One way to accomplish this is to use the maximum likelihood method of Carrera & Neuman (1986a,b). This, and other related methods of parameter estimation, have been recently reviewed and compared by Zimmerman *et al*. (1998). By satisfying all three goals, one would end up with a localised model that is conditioned on a combination of hydraulic conductivity, head and flux data. Consider now the error,

$$\varepsilon_h(\boldsymbol{x}) \equiv h(\boldsymbol{x}) - \langle \hat{h}(\boldsymbol{x})\rangle_c \tag{10}$$

introduced upon using such a calibrated model to predict the actual head, $h(\boldsymbol{x})$, at any point $\boldsymbol{x}$ in the domain by means of $\langle \hat{h}(\boldsymbol{x})\rangle_c$. This error can be expressed as

$$\varepsilon_h(\boldsymbol{x}) = \varepsilon_{\langle h\rangle}(\boldsymbol{x}) + \varepsilon_{\langle \tilde{h}\rangle}(\boldsymbol{x}) + \varepsilon_{\langle \hat{h}\rangle}(\boldsymbol{x}) \tag{11}$$

where

$$\varepsilon_{\langle h\rangle}(\boldsymbol{x}) \equiv h(\boldsymbol{x}) - \langle h(\boldsymbol{x})\rangle_c \tag{12}$$

$$\varepsilon_{\langle \tilde{h}\rangle}(\boldsymbol{x}) \equiv \langle h(\boldsymbol{x})\rangle_c - \langle \tilde{h}(\boldsymbol{x})\rangle_c \tag{13}$$

$$\varepsilon_{\langle \hat{h}\rangle}(\boldsymbol{x}) \equiv \langle \tilde{h}(\boldsymbol{x})\rangle_c - \langle \hat{h}(\boldsymbol{x})\rangle_c \tag{14}$$

Here $\langle h(\boldsymbol{x})\rangle_c$ is the (generally) unknown true conditional mean head, $\langle \tilde{h}(\boldsymbol{x})\rangle_c$ is conditional mean head corrupted by localisation and other modelling errors, $\langle \hat{h}(\boldsymbol{x})\rangle_c$ is mean head corrupted by both localisation and parameter estimation errors, $\varepsilon_{\langle h\rangle}(\boldsymbol{x})$ is the error introduced by using the conditional mean $\langle h(\boldsymbol{x})\rangle_c$ to predict the unknown random variable $h(\boldsymbol{x})$, $\varepsilon_{\langle \tilde{h}\rangle}(\boldsymbol{x})$ is a modelling error due primarily to localisation, and $\varepsilon_{\langle \hat{h}\rangle}(\boldsymbol{x})$ is an additional error due to parameter estimation. Note that in our previous examples, the latter error was zero, and the modelling error was relatively small (except at the well).

Statistical analyses of parameter estimation errors, such as $\varepsilon_{\langle \hat{K}\rangle} \equiv \left\| \boldsymbol{K}_c(\boldsymbol{x}) - \hat{\boldsymbol{K}}_c(\boldsymbol{x}) \right\|$ where $\| \ \|$ is some consistent norm, typically yield a lower bound for the error (*c.f.*, Carrera & Neuman, 1996a, b). The corresponding error estimate is then commonly used to assess uncertainties in model output, by varying its parameters (in our case $\boldsymbol{K}_c$) randomly about their calibrated values (in our case $\hat{\boldsymbol{K}}_c$), and conducting Monte Carlo simulations of corresponding heads and fluxes. The approach yields at best information about errors such as $\varepsilon_{\langle \hat{h}\rangle}(\boldsymbol{x})$, which in our case stems from the parameter estimation error $\varepsilon_{\langle \hat{K}\rangle}(\boldsymbol{x})$. It provides no information about modelling errors such as $\varepsilon_{\langle \tilde{h}\rangle}(\boldsymbol{x})$, or prediction errors such as either $\varepsilon_{\langle h\rangle}(\boldsymbol{x})$ or $\varepsilon_h(\boldsymbol{x})$. As such, it underestimates the prediction error $\varepsilon_h(\boldsymbol{x})$, the only type of error that ultimately matters, and overestimates the predictive capabilities of the model. The reason is that $\boldsymbol{K}_c(\boldsymbol{x})$ relates conditional mean hydraulic gradient to conditional mean flux, and so $\varepsilon_{\langle \hat{K}\rangle}(\boldsymbol{x})$ affects only uncertainties associated with the evaluation of these conditional mean quantities. It contains no information about how actual heads and fluxes fluctuate about their conditional mean predictors; for that, one needs to evaluate higher conditional moments of head and flux, with the aid of stochastic methods that treat hydraulic parameters (in our case K or Y) as random fields. For a novel way of doing so, the reader is referred to Guadagnini & Neuman (1999a, b). Their work clearly demonstrates that the stochastic component, $\varepsilon_{\langle h\rangle}(\boldsymbol{x})$, of the total prediction error, $\varepsilon_h(\boldsymbol{x})$, can be very large in strongly heterogeneous media.

CONCLUSIONS

1. Applying traditional deterministic flow models to nonuniform media is valid at best as an approximation. Whereas traditionally the hydraulic conductivity in such models was considered to be a unique material property, in reality it is a nonunique, data-dependent, spatially-varying effective hydraulic conductivity tensor.

2. Whereas traditionally the hydraulic head and flux in deterministic flow equations have been considered to represent real system states, they must in fact be viewed at best as conditional predictors of these states, which differ from their unknown counterparts by random errors of prediction. Since standard deterministic flow models provide no information about such prediction errors, they yield solutions of indeterminate quality.

3. Calibrating a traditional deterministic flow model against measured values of head and flux is tantamount to conditioning it on such measurements. As the model hydraulic parameters are

by nature conditional, the very act of adding measured heads and fluxes to the database alters their values. This inherent nonuniqueness persists regardless of whether the inverse problem is well-posed or ill-posed.

4. Assessing the uncertainty of deterministic model parameters may allow one to quantify the uncertainty in computed head and flux predictors. This, however, says nothing about how actual heads and fluxes fluctuate about their predictors; such information is provided only by stochastic models that treat hydraulic parameters as random fields. The traditional Monte Carlo method of assessing uncertainty in the output of a calibrated deterministic model accounts neither for modelling errors nor for the stochastic component of the prediction error. As the latter error can be very large in strongly heterogeneous media, the traditional method may seriously overestimate the predictive capabilities of the calibrated groundwater flow model. The same holds for transport.

REFERENCES

Carrera, J. & Neuman, S. P. (1986a) Estimation of aquifer parameters under transient and steady state conditions, 1, Maximum likelihood method incorporating prior information. *Wat. Resour. Res.* **22** (2), 199-210.

Carrera, J. & Neuman, S. P. (1986b) Estimation of aquifer parameters under transient and steady state conditions, 2, Uniqueness, stability, and solution algorithms. *Wat. Resour. Res.* **22** (2), 211-227.

Deutsch, C. V. & Journel, A. G. (1998) *GSLIB Geostatistical Software Library and User's Guide*. Oxford University Press, New York.

Guadagnini, A. & Neuman, S. P. (1999a) Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains, 1, theory and computational approach. *Wat. Resour. Res*. **35** (10), 2999-3018.

Guadagnini, A. & Neuman, S. P. (1999b) Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains, 2, computational examples. *Wat. Resour. Res*. **35** (10), 3019-3039.

Neuman, S. P., & Orr S. (1993), Prediction of steady state flow in non uniform geologic media by conditional moments: Exact non local formalism, effective conductivities, and weak approximations. *Wat. Resour. Res.* **29** (2), 341-364.

Neuman, S. P., Tartakovsky, D. M., Wallstrom, T. C. & Winter, C. L. (1996) Correction to "Prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation". *Wat. Resour. Res.* **32**(5), 1479-1480.

Tartakovsky D. M. & Neuman, S. P. (1998) Transient flow in bounded randomly heterogeneous domains 1. Exact conditional moment equations and recursive approximations. *Wat. Resour. Res.* **34** (1), 1-12.

Tartakovsky, D. M. & Neuman, S. P. (1999) Extension of "Transient flow in bounded randomly heterogeneous domains: 1. Exact conditional moment equations and recursive approximations,". *Wat. Resour. Res.* **35**(6), 1921-1925.

Zimmerman, D. M., de Marsily, G., Gotway, C. A., Marietta, M. G., Axness, C. L., Beauheim, R. L., Bras, R. L., Carrera, J., Dagan, G., Davies, P. B., Gallegos, D. P., Galli, A., Gómez-Hernández, J., Grindrod, P., Gutjahr, A. L., Kitanidis, P. K., Lavenue, A. M., McLaughlin, D., Neuman, S. P., RamaRao, B. S., Ravenne, C. & Rubin, Y. (1998) A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Wat. Resour. Res*. **34**(6), 1373-1413.

APPENDIX D

MIXING, DILUTION, AND DISPERSION IN HYDROGEOLOGIC ENVIRONMENTS

APPENDIX D: MIXING, DILUTION, AND DISPERSION IN HYDROGEOLOGIC ENVIRONMENTS

The concepts of mixing, dilution and dispersion of dissolved contaminants are central to the assessment of groundwater pollution hazards. Yet there appears to be considerable confusion in the literature about the meaning of these concepts and their proper representation in contaminant transport models. Most notable is a lack of appreciation for the intimate link between dispersion and hydrogeologic uncertainty. The purpose of this section is to help shed light on these concepts and issues. A mathematical treatment of this topic can be found in Neuman (1993), Zhang and Neuman (1996) and Guadagnini and Neuman (2001).

It is common to assume for purposes of environmental impact and performance assessments of nuclear facilities and sites that when radionuclides exiting the base of the unsaturated zone and enter into the saturated, they rapidly (in fact instantaneously) mix with a body of groundwater that is many meters, even tens of meters, deep. Additional dilution is sometimes assumed to occur by mixing of different groundwater sources, either naturally along the flow path between the source of radionuclides and a user, or by the user tapping alternative sources of water as in the slotting of a well over different units. Some believe dilution to occur by the mixing of groundwaters from adjacent sub-basins due to water withdrawal in the vicinity of their mutual boundary and lateral as well as vertical dispersion. The issue of mixing, dilution and dispersion in hydrogeologic environments is clearly of major significance to performance assessment.

Yet, in reality, the only mechanisms by which dilution can actually take place in the absence of groundwater withdrawal for sampling or use are molecular diffusion, advective dispersion caused by space-time meandering of pathlines and variations in velocity along as well as among streamtubes, and turbulent eddies which are not likely to occur at most nuclear facilities and sites. Otherwise, dilution can occur when water is drawn into samplers or wells from multiple horizons and/or directions, and when waters in samplers or wells mix by diffusion and turbulence caused by shaking, stirring or rapid flow. Density effects due to space-time variations in solution chemistry, temperature and pressure may either enhance or prevent mixing (as in buoyant or gravity segregation of fluids having different densities). Another factor which may contribute to mixing is instability of fluid interfaces and resultant fingering.

There appears to be no scientific basis for the "stirred tank" model according to which contaminated waters from the unsaturated zone mix rapidly with pristine waters in the saturated zone down to some specified "mixing depth." Quite the contrary, small vertical dispersivity commonly observed in stratified materials will usually keep the incoming plume of radionuclides at shallow depth except where it is intercepted by major vertical fractures or faults within which there is significant downward flow. One must likewise question the scientific basis for the notion that waters from neighboring sub-basins mix naturally along a flow path; only where flow paths from the two sub-basins converge will their waters mix by dispersion under natural conditions.

Mixing by dispersion may occur, but it is important to recognize that dispersion cannot be interpreted as dilution except in special cases to be described later. On the laboratory scale, dispersion coefficients compensate for our inability to resolve the intricacies of advective transport and diffusion in the pore space, about which we thus have no direct information. On the field scale, dispersion coefficients compensate additionally for our inability to resolve the intricacies of advective transport and local-scale dispersion in a heterogeneous porous and/or fractured rock, about which we have only partial information. As space-time fluctuations in advective velocity are generally greater in the field than in the laboratory, the amount of information we generally lose through inadequate resolution of flow in the field tends to be greater than that which we lose through inadequate resolution of flow in a laboratory sample of rock. It follows that the coefficient of dispersion, which compensates for this loss of information, is generally larger on the field scale than on the laboratory scale. In fact, longitudinal dispersivities appear to increase consistently with the scale of field observations, a phenomenon we attribute today to a corresponding increases in the scale of medium heterogeneity, which our groundwater flow and transport models fail to resolve (Neuman, 1990, 1995; Di Federico and Neuman, 1997, 1998a,b; Neuman and Di Federico, 1998). Clearly, lack of resolution does not necessarily imply mixing and dilution; only if we sample on a scale comparable to that of unresolved heterogeneities should we expect dispersion to imply mixing and dilution. Hence confusing dispersion with mixing and dilution is generally inappropriate.

Given that field-scale dispersion is associated with incomplete resolution of medium heterogeneities, one can reduce it by including in the groundwater flow and transport models a greater amount of detail about the spatial variability of medium properties. Since one never has exhaustive information about all small-scale details of medium heterogeneity, the only hope one has to resolve such details is by means of statistical and geostatistical methods. A geostatistical description of medium heterogeneity introduces uncertainty into groundwater flow and transport models which renders these models stochastic. Hence stochastic approaches provide an appropriate framework for the analysis of flow and transport in heterogeneous media (Dagan and Neuman, 1997).

Briefly, the stochastic view of field-scale dispersion proposed here is as follows. Given a sufficiently large sample of local-scale medium properties at a site, it is often possible to generate by conditional Monte Carlo simulation random but potentially realistic spatial distributions or images (called realizations) of these properties which are equally likely and honor the data. Here local scale means any scale that (a) is much smaller than the rock volume under investigation and (b) allows measuring many if not all requisite flow and transport parameters (permeability, porosity, dispersivity) and state variables (heads, concentrations) by means of standard field techniques; the corresponding length scale is typically on the order of meters to tens of meters.

Given adequate computer resources, it is then also possible to perform random but equally likely, and potentially realistic, high-resolution flow and transport simulations on a computational grid with local-scale cells, using local-scale medium properties such as permeability, porosity and dispersivity. Since local-scale dispersivities are much smaller than their larger-scale counterparts, these simulations indicate much lesser (but more realistic) degrees of mixing and dilution than would simulations conducted on coarser grids with lesser spatial resolution.

Hence for purposes of investigating mixing and dilution, we propose that analysts follow the above approach of high-resolution, statistically-based conditional flow and transport simulations. Even a small number of corresponding realizations may be more telling with regard to dilution and mixing than would deterministic or stochastic simulations which do not achieve a comparable degree of resolution.

In the same context, it is important to recognize that local-scale dispersivities increase with their scale of definition; a crude rule of thumb (which does not consider numerical dispersion) is to set the longitudinal dispersivity equal to one tenth the length of a local-scale grid cell, and the transverse dispersivity a fraction thereof (one tenth to one third).

If a statistically significant number of conditional Monte Carlo flow and transport simulations are performed, their results can be averaged to yield conditional mean values of head, groundwater flux, groundwater velocity, solute concentration and solute mass flux at each grid point at many discrete time steps. These conditional mean values are deterministic (sure, certain) and vary much more smoothly in space-time than do their random (uncertain) counterparts (as represented by individual realizations). Most importantly, they constitute optimum unbiased predictors of the actual but unknown local-scale values of head, flux, velocity, concentration and mass flux at each point in space-time. The variance of the conditional Monte Carlo realizations provides a measure of the uncertainty associated with these predictions, which can in turn be translated into optimum predictions and uncertainties of performance measures. Even a small sample of realizations may yield meaningful, though probably not accurate, insight into the uncertainty associated with simulating flow and transport at a site, and assessing the corresponding system performance.

There is only one proper way to avoid running many high-resolution conditional Monte Carlo simulations of flow and transport: solve numerically, on a coarser computational grid than required for such simulations, a single set of deterministic flow and transport equations which control the space-time evolution of smooth conditional mean heads, fluxes, velocities, concentrations and mass fluxes (it is the relative smoothness of these deterministic functions which allows, in principle, computing them deterministically on a relatively coarse grid). Unfortunately, as was stated earlier, the corresponding conditional mean flow and transport equations are not differential but integro-differential. Fortunately, they can sometimes be approximated by differential equations which look like the familiar flow and transport equations (the same equations one would use, among others, for high-resolution conditional Monte Carlo simulations) but their coefficients and

variables now have different meanings and values. In particular, the conditional mean advection-dispersion equation now contains a conditional mean velocity vector which is much smoother, and a dispersion tensor which grows with time to become rapidly much larger, than their local counterparts. The large dispersion compensates for resolution lost in smoothing the velocity and concentration fields (recall that solving the conditional mean equation is analogous to averaging the results of many conditional Monte Carlo simulations, each one of which is associated with nonsmooth velocity and concentration fields, which however become smooth upon averaging). The more information about local-scale medium properties one builds into the model (i.e., the more strongly one conditions the model on local-scale data), the less smooth is the corresponding conditional mean velocity (as well as concentration) field and the smaller is the corresponding dispersion tensor. Clearly, building information into a model does not affect mixing and dilution, hence it is evident that the dispersion tensor does not generally reflect these phenomena.

Since there is always uncertainty about local-scale medium heterogeneities, a deterministic analysis of transport is never warranted unless the advection-dispersion model is viewed and interpreted in the manner just described. This means that the computed concentration and mass flux are recognized to represent not actual but smooth predicted values with which there is associated a quantifiable error of prediction and smoothing, and that the computed concentrations are recognized to potentially spread, or disperse, to a much greater extent than do their real but unknown counterparts (one can view this enhanced spread, or dispersion, not as that of real solute mass but that of information about the space-time distribution of this mass).

Only in special, so-called quasi-ergodic situations can such enhanced dispersion be interpreted to imply mixing and dilution. These situations arise as the mean travel distance becomes large enough for a plume to encounter (sample) heterogeneities of all relevant scales in the longitudinal direction of flow, and when the source of contamination is wide enough so that the plume can sample all such heterogeneities in the transverse direction. An extreme example of nonergodic transport is that of an imaginary solute "particle" of infinitesimal volume which never diffuses, disperses or dilutes. If the particle has unit mass normalized by porosity then its concentration, predicted deterministically with a relatively large "field-scale" dispersion coefficient, merely represents the probability of finding this particle in the immediate vicinity of any given point in space-time.

Despite significant advances in stochastic flow and transport theories over the last two decades as reflected in the recent book of Dagan and Neuman (1997), stochastically derived field-scale dispersivities are known to overpredict the vertical spread of solutes in stratified media under seemingly quasi-ergodic conditions.

Model abstraction through reduction of dimensionality (as when three-dimensional transport is represented by a network of one-dimensional flow tubes) is associated with a loss of spatial resolution which should be compensated for by a corresponding increase in dispersivities (Domenico and Robbins, 1984). Once again, such an increase must not be interpreted to imply enhanced dilution.

APPENDIX E

MAXIMUM LIKELIHOOD BAYESIAN MODEL AVERAGING

APPENDIX E: MAXIMUM LIKELIHOOD BAYESIAN MODEL AVERAGING

Accounting for conceptual model uncertainty
Via maximum likelihood Bayesian model averaging

S. P. NEUMAN
Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona 85721, USA
e-mail: neuman@hwr.arizona.edu

Abstract Analyses of groundwater flow and transport typically rely on a single conceptual model of site hydrogeology. Yet hydrogeologic environments are open and complex, rendering them prone to multiple interpretations. Adopting only one of these may lead to statistical bias and underestimation of uncertainty. A comprehensive strategy for constructing alternative conceptual-mathematical models, selecting the best among them, and using them jointly to render optimum predictions under uncertainty is being developed by the author. This paper proposes a Maximum Likelihood Bayesian Model Averaging approach, MLBMA, to rendering optimum predictions by means of several competing models and assessing their joint predictive uncertainty.
Key words conceptual models; model uncertainty; predictive uncertainty; Bayesian; maximum likelihood

INTRODUCTION

Analyses of groundwater flow and transport typically rely on a single conceptual model of site hydrogeology. Yet hydrogeologic environments are open and complex, rendering them prone to multiple interpretations. This is true regardless of the quantity and quality of available site data. Focusing on only one site interpretation may lead to Type I model errors, which arise 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 framework. Indeed, critiques of hydrogeologic analyses, and legal challenges to such analyses, typically focus on the validity of the underlying conceptual model. If severe, these may damage one's professional credibility; result in the loss of a legal contest; and lead to adverse environmental, economic and political impacts.

Analyses of model uncertainty based on a single hydrogeologic concept are prone to statistical bias (by committing a Type II error through reliance on an invalid model) and underestimation of uncertainty (by committing a Type I error through under sampling of the relevant model space). The bias and uncertainty that result from reliance on an inadequate conceptual model are typically much larger than those introduced through an inadequate choice of model parameter values. Yet most uncertainty analyses of flow and transport ignore the former and focus exclusively on the latter. This often leads to overconfidence in the predictive capabilities of the model, which the available site data do not justify.

It is argued by Beven and Freer (2001) "that, given current levels of understanding and measurement technologies, it may be endemic to mechanistic modeling of complex environmental systems that there are many different model structures and many different parameter sets within a chosen model structure that may be behavioural or acceptable in reproducing the observed behaviour of that system." They attribute to Hornberger and Speer (1981) the notion that this is not simply a problem of identifying a correct or optimal model given limited data. Instead, this is a generic problem which Beven (1993) calls *equifinality* and attributes to (Beven, 2000) limitations of current model structures in representing heterogeneous surface and subsurface flow systems, limitations of measurement techniques and scales in defining system characteristics including initial and boundary conditions for a model, and the uniqueness of individual sites. He points out that to do detailed measurements throughout a site is both impractical and unfeasibly expensive. The unique characteristics of a site are therefore inherently unknowable. All that can be done is to constrain the model representations of the site to those that are acceptably realistic, usually in the sense of being consistent with the data.

There is no established literature on ways to construct alternative conceptual models of site hydrogeology, select the best among them, use them jointly to render optimum predictions of groundwater flow and transport, and assess the uncertainty of such predictions. A comprehensive strategy for doing so is being developed by the author. This paper focuses on one aspect of the strategy, which addresses the question how to render optimum predictions by means of several competing deterministic or stochastic models and how to assess their joint predictive uncertainty.

PREDICTION AND UNCERTAINTY ANALYSIS BASED ON SINGLE MODEL

Deterministic model

The traditional approach to hydrologic prediction and uncertainty analysis has been to postulate a deterministic model structure and treat its parameters as being imperfectly known. To quantify this imperfect knowledge, one must postulate a *prior* parameter uncertainty model. When sufficient site parameter values are available, one could postulate a Type A probabilistic model of prior parameter uncertainty based on statistics derived from these data. When no such data are available in statistically significant quantities, one has the option of postulating a Type B model of prior parameter uncertainty on the basis of subjective probabilities. Such a model should always be suspected of suffering from an unknown amount of statistical and personal bias. Intermediate between Type A and Type B parameter uncertainty models is the case where indirect information about the parameters is available, from which relevant prior statistics can be derived formally. Such information may include off-site parameter measurements and/or surrogate data. Statistics derived from off-site data are potentially biased due to a lack of site-specific information about mean parameter values and incompatibility of geology and scale. The associated variance may be too small or too large, depending on the quantity and quality of such data. Statistics derived from surrogate data may suffer from poorly defined correlations and incompatibility of scale.

The traditional approach to reduce parameter bias and uncertainty has been to calibrate the model against observed system behavior by means of a suitable inverse method. The last thirty years have seen major advances in the development of theories and algorithms for the estimation of deterministic model parameters. Many (though not all) of these theories and algorithms are "statistical" in that they include analyses of parameter estimation uncertainty. Such analyses typically accept, but do not necessarily require, information about prior parameter statistics as input. The output includes *posterior* statistics of parameter estimation errors, which are generally less biased and smaller than the prior estimation errors. A recent summary and comparison of various statistical inverse methods for groundwater flow models has been published by Zimmerman *et al.* (1998). A detailed set of guidelines for the effective calibration of deterministic groundwater flow models has been prepared by Hill (1998).

The most common way to propagate input errors through an otherwise deterministic model is by means of Monte Carlo simulations. This is done by generating multiple, equally likely sets of randomized inputs; computing deterministically a set of corresponding model outputs for each; and analyzing the resultant multiple, equally likely random output sets statistically. Another approach is to associate the predictions with approximate error bounds, or confidence limits, computed on the basis of linear regression theory applied to the (typically nonlinear) groundwater inverse model (Hill, 1998).

Stochastic model

Hydrogeologic medium properties exhibit both systematic and random spatial variations on a multiplicity of scales. Traditional deterministic models capture at best the larger-scale, systematic components of these variations. They however fail to resolve smaller scale variations or account for their uncertain nature. The emphasis is therefore shifting from deterministic to probabilistic methods that are better suited for these needs. The trend has become to describe the spatial variability and scaling of hydrogeologic medium properties geostatistically, and to analyze subsurface fluid flow and solute transport stochastically. This trend has been documented in a number of recent books including those by Dagan and Neuman (1997) and Zhang (2001).

The most common method of stochastic analysis is high-resolution computational Monte Carlo simulation that produces a large number of equally likely results. These nonunique results are summarized in terms of statistically averaged quantities, their variance-covariance, and perhaps higher moments of the corresponding sample probability distributions. Results that honor measured values of medium properties are said to be conditioned on these data. Upon conditioning the simulations on measured values of parameters in space, one obtains (among others) conditional mean flow and transport variables that constitute optimum unbiased predictors of these unknown random quantities. One also obtains conditional second moments (variance-covariance) that provide a measure of the associated prediction errors. To condition the predictions on system monitoring data, one must either discard random simulations that do not reproduce the observations, or employ an inverse procedure of the kind developed for this purpose by Gómez-Hernández *et al.* (1997).

Monte Carlo analysis requires knowing the multivariate probability distribution of relevant hydrogeologic properties, which is difficult to infer from commonly available data. To achieve a high space-time resolution of relevant stochastic phenomena, it requires the use of large space-time grids with very small discretization intervals. To yield sample statistics that converge to their theoretical (ensemble) counterparts requires numerous repetitions (realizations). The net result is a large amount of computational time and storage, which are considered uneconomical for many practical applications.

This has given impetus to the development of alternative stochastic methods that allow one to compute the conditional mean, variance and covariance of groundwater flow and transport variables directly, without Monte Carlo simulation. This is done on the basis of moment equations conditional on hydraulic parameter measurements as illustrated for example by Guadagnini and Neuman (1999) and Ye *et al.* (2002). Conditioning additionally on observed hydraulic head values requires an inverse procedure (Hernandez *et al.*, 2002).

PREDICTION AND UNCERTAINTY ANALYSIS BASED ON MULTIPLE MODELS

Previous hydrologic approaches

Carrera and Neuman (1986b) have noted that an inadequate model structure (conceptualization) is far more detrimental to its predictive ability than is a suboptimal set of model parameters. This helps explain why the National Research Council (1999) has listed as second among four recommended research directions in subsurface science the development of tools and methodologies for conceptual modeling with emphasis on heterogeneity, scale and uncertainty bounds on the basis of field experimental data.

Recently, a panel was convened by the National Research Council (2001) to describe the process through which conceptual models of flow and transport in the fractured vadose zone are developed, tested, refined and reviewed. The panel concluded that development of the conceptual model is the most important part of the modeling process. The conceptual model is the foundation of the quantitative mathematical representation of the field site (i.e., the mathematical model), which in turn is the basis for the computer code used for simulation. Reasonable alternative conceptualizations and hypotheses should be developed and evaluated. In some cases, the early part of a study might involve multiple conceptual models until alternatives are eliminated by field results.

According to the panel, it is important to recognize that model predictions require assumptions about future events or scenarios, and are subject to uncertainty. Meaningful quantification of uncertainty should be considered an integral part of any modeling endeavor, as it establishes confidence bands on predictions given the current state of knowledge about the system. A suite of predictions for a range of different assumptions and future scenarios is more useful than a single prediction.

We have noted earlier that there is uncertainty not only about the parameter values that should enter into a given model (as characterized by its structure), but also about the very structure (conceptual and mathematical) of the model that should represent the hydrologic system of interest. The traditional approach to model uncertainty analysis, which considers only a single deterministic model structure, fails to adequately sample the complete space of plausible hydrologic models. As such, it is prone to modeling bias and underestimation of model uncertainty.

An example of how one could account quantitatively for structural model uncertainties was given by James and Oldenburg (1997). They investigated the uncertainty of simulated TCE concentrations, at the point of potential human exposure, due to uncertainty in the parameters (permeability, porosity, diffusivity, solubility, adsorption) and variations in the conceptual-mathematical model (injection rate of TCE source; initial TCE source saturation; regional groundwater flow; heterogeneity of permeability). The authors used the three-dimensional code T2VOC to simulate three-phase (gas, aqueous, NAPL), three-component (air, water, VOC) nonisothermal flow based on an actual site with a 25 m thick vadose zone and a saturated zone. To assess parameter uncertainty associated with a given model, they used the inverse code ITOUGH2. Their final step was to assess the range of outcomes that one obtains with the entire set of alternative conceptual-mathematical models. James and Oldenburg found that uncertainties in their model outcomes span orders of magnitude, and that both parameter and model uncertainty contribute significantly to this wide range of outcomes. They concluded that "risk assessment and remediation selection ... is meaningful only if analysis includes quantitative estimates of ... uncertainty" in both the parameters and the conceptual-mathematical models.

A similar approach has been advocated more recently by Samper and Molinero (2000). The authors consider the main uncertainties in predicting groundwater flow and transport to be those associated with the selection of future scenarios, choice of model structure and assignment of model parameters. The authors consider parameter uncertainty to be minor in comparison to structural (i.e. conceptual) model errors. They suggest to evaluate model predictive uncertainty by calibrating a number of conceptual-mathematical models against available monitoring data, to retain those calibrated models that can adequately reproduce past observations, to assess the predictive uncertainty of each model due to the uncertainty of its parameters, to treat the predictive uncertainty of each model as being equally likely, and to produce a single combined range of predictive uncertainties.

Rather than relying on model calibration and treating the outcomes of different structural models as being equally likely, Beven and Binley (1992) have proposed a strategy to which they refer as GLUE (Generalized Likelihood Uncertainty Estimation). The strategy calls for the identification of several alternative structural models and the postulation of a prior probabilistic model of parameter uncertainty for each. Each structural model, coupled with its corresponding parameter uncertainty model, is used to generate Monte Carlo realizations of past hydrologic behaviors and to compare the results with monitored system behavior during the same period. Likelihood measures are defined to gauge the degree of correspondence between each simulated and observed record of system behavior. If a likelihood measure falls below a subjectively defined "rejection criterion," the corresponding combination of model structure and parameter

set are discarded. Those combinations which pass this test are retained to provide predictions of system behavior under selected future scenarios. Each prediction is weighted by a corresponding normalized likelihood measure (so as to render the sum of all likelihood measures equal to one), to produce a likelihood-weighted cumulative distribution of all available predictions. For recent discussions of GLUE and its applications the reader is referred to Beven (2000) and Beven and Freer (2001).

A Bayesian approach to the quantification of errors in a single groundwater model was recently proposed by Gaganis and Smith (2001). Like GLUE, it relies on Monte Carlo simulations without model calibration and on subjective criteria of "model correctness."

It must be understood that the set of predictions one produces with any given choice of alternative structural models and parameter sets, by whatever method, is conditional on the choice of models and the data used to support them. As such, these predictions do not represent all possibilities but only a limited range of such possibilities, associated with these models and data. Any change in the latter would generally lead to a different assessment of predictive model uncertainty. There thus appears to be no way to assess the uncertainty of hydrologic predictions in an absolute sense, only in a conditional or relative sense.

Proposed approach

Bayesian model averaging

At the heart of the approach we propose is the concept of Bayesian Model Averaging (BMA), described with clarity in a recent tutorial by Hoeting *et al.* (1999). According to these authors, "standard statistical practice ignores model uncertainty … leading to over-confident inferences and decisions that are more risky than one thinks they are. … (BMA) provides a coherent mechanism for accounting for this model uncertainty." They introduce BMA by noting that if Δ is a quantity one wants to predict, then its posterior distribution given a discrete set of data $\boldsymbol{D}$ is

$$p\left(\Delta\middle|\boldsymbol{D}\right)=\sum_{k=1}^{K}p\left(\Delta\middle|M_k,\boldsymbol{D}\right)p\left(M_k\middle|\boldsymbol{D}\right) \tag{1}$$

where $\mathscr{M}=\left(M_1,...,M_K\right)$ is the set of all models (or hypotheses) considered, at least one of which must be valid. In other words, $p\left(\Delta\middle|\boldsymbol{D}\right)$ is the average of the posterior distributions $p\left(\Delta\middle|M_k,\boldsymbol{D}\right)$ under each model, weighted by their posterior model probabilities $p\left(M_k\middle|\boldsymbol{D}\right)$. The posterior probability for model M_k is given by Bayes' rule,

$$p\left(M_k\middle|\boldsymbol{D}\right)=\frac{p\left(\boldsymbol{D}\middle|M_k\right)p\left(M_k\right)}{\sum_{l=1}^{K}p\left(\boldsymbol{D}\middle|M_l\right)p\left(M_l\right)} \tag{2}$$

where

$$p\left(\boldsymbol{D}\middle|M_k\right)=\int p\left(\boldsymbol{D}\middle|\boldsymbol{\theta}_k,M_k\right)p\left(\boldsymbol{\theta}_k\middle|M_k\right)d\boldsymbol{\theta}_k \tag{3}$$

is the integrated likelihood of model M_k, $\boldsymbol{\theta}_k$ is the vector of parameters associated with model M_k, $p\left(\boldsymbol{\theta}_k\middle|M_k\right)$ is the prior density of $\boldsymbol{\theta}_k$ under model M_k, $p\left(\boldsymbol{D}\middle|\boldsymbol{\theta}_k,M_k\right)$ is the joint likelihood of model M_k and its parameters $\boldsymbol{\theta}_k$, and $p\left(M_k\right)$ is the prior probability that M_k is the correct model. All probabilities are implicitly conditional on $\mathscr{M}$.

The posterior mean and variance of Δ are (Draper, 1995)

$$E\left[\Delta\middle|\boldsymbol{D}\right]=\sum_{l=1}^{K}E\left[\Delta\middle|\boldsymbol{D},M_l\right]p\left(M_l\middle|\boldsymbol{D}\right) \tag{4}$$

$$Var\left[\Delta\middle|\boldsymbol{D}\right]=\sum_{l=1}^{K}\left\{Var\left[\Delta\middle|\boldsymbol{D},M_l\right]+E\left[\Delta\middle|\boldsymbol{D},M_l\right]^2\right\}p\left(M_l\middle|\boldsymbol{D}\right)-E\left[\Delta\middle|\boldsymbol{D}\right]^2. \tag{5}$$

According to Hoeting *et al.* (1999), there is considerable empirical evidence that averaging over *all* models in this fashion provides better average predictive ability than relying on a single model, M_k, conditional on $\mathscr{M}$. However, they list a number of factors that render the application of BMA to complex systems (such as those encountered in hydrology) difficult: (a) The number of potentially feasible models may be exceedingly large, rendering their exhaustive inclusion in (1) infeasible; (b) integrals of form (3) may be hard to compute; and (c) the specification of prior model probabilities $p\left(M_k\right)$ is challenging, having received little attention.

A practical way to eliminate the first difficulty is to average over a manageable subset of models that are supported by the data. The strategy being developed by the author promotes the idea of Occam's window (Madigan and Raftery, 1994) according to which averaging is limited to a relatively small set of the most parsimonious models that are most strongly supported by the data while remaining hydrologically plausible.

Maximum likelihood Bayesian model averaging

To render BMA computationally feasible, we adopt a suggestion by Taplin (1993) that $p\left(\Delta\middle|M_k,\boldsymbol{D}\right)$ in (1) be approximated by $p\left(\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right)$ where $\hat{\boldsymbol{\theta}}_k$ is the maximum likelihood estimate of $\boldsymbol{\theta}_k$ based on the likelihood $p\left(\boldsymbol{D}\middle|\boldsymbol{\theta}_k,M_k\right)$. Hoeting *et al.* (1999) note that Draper (1995), Raftery *et al.* (1996) and Volinsky *et al.* (1997) have shown this to be useful in the BMA context.

Methods to evaluate $\hat{\boldsymbol{\theta}}_k$ by calibrating a deterministic model M_k against hydrogeologic data $\boldsymbol{D}$, which may include prior information about the parameters, are described by Carrera and Neuman (1986a,b) and Carrera *et al.* (1997). The same can be done with a stochastic model based on moment equations in a manner similar to that of Hernandez *et al.* (2002). These methods also yield an approximate covariance matrix for the estimation errors of $\hat{\boldsymbol{\theta}}_k$. Upon considering the parameter estimation errors of a calibrated deterministic model M_k to be Gaussian or log Gaussian, one easily determines $p\left(\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right)$ by Monte Carlo simulation of Δ through random perturbation of the parameters. The simulation also yields corresponding approximations $E\left[\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right]$ of $E\left[\Delta\middle|M_k,\boldsymbol{D}\right]$, and $Var\left[\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right]$ of $Var\left[\Delta\middle|M_k,\boldsymbol{D}\right]$, in (4) and (5). If M_k is a stochastic model based on moment equations, it can yield $E\left[\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right]$ and $Var\left[\Delta\middle|M_k,\hat{\boldsymbol{\theta}}_k,\boldsymbol{D}\right]$ directly without Monte Carlo simulation (Hernandez *et al.*, 2002).

To eliminate the need for computer intensive integration according to (3), I propose to evaluate the weights $p\left(M_l\middle|\boldsymbol{D}\right)$ in (1) and (4) – (5) based on a result due to Kashyap (1982). The author considers a set $M_1,...,M_K$ of mutually exclusive models so that any set of observational data could have originated from only one of them. I interpret this to mean that only one of the models is correct even in the event that some yield similar predictions for a given set of data (in which event the degeneracy could be resolved by prior information about which model "makes most sense," parsimony, and/or additional data for which the predictions would differ). The models may be linear or nonlinear, Gaussian or non-Gaussian. Kashyap proves that, under some fairly standard conditions,

$$\ln p\left(M_k\middle|\boldsymbol{D}\right)=\ln C_k+\ln p\left(\boldsymbol{D}\middle|\hat{\boldsymbol{\theta}}_k,M_k\right)+\ln p\left(\hat{\boldsymbol{\theta}}_k\middle|M_k\right)+\frac{N_k}{2}\ln\left(\frac{2\pi}{N}\right)$$
$$-\frac{1}{2}\ln\left|\boldsymbol{F}_k\left(\boldsymbol{D}\middle|\hat{\boldsymbol{\theta}}_k,M_k\right)\right|+R(N) \qquad (6)$$

where $C_k=cp\left(M_k\right)$, $p\left(M_k\right)$ being the prior probability of model M_k and c a constant that can be evaluated from

$$\sum_{l=1}^{K}p\left(M_l\middle|\boldsymbol{D}\right)=1, \qquad (7)$$

N_k is the dimension of $\boldsymbol{\theta}_k$, N is the dimension of $\boldsymbol{D}$, $\boldsymbol{F}_k$ is the normalized (by N) observed (as opposed to ensemble mean) Fisher information matrix having components

$$F_{k,ij} = -\frac{1}{N}\frac{\partial^2 \ln p\left(\boldsymbol{D}\middle|\boldsymbol{\theta}_k, M_k\right)}{\partial\theta_i \partial\theta_j} \tag{8}$$

and $NR(N)$ tends to a constant almost surely as $N \to \infty$.

Kashayp (1982) suggests that, in the absence of any contrary information, the models be assigned equal prior probabilities, yielding $C_k = C = constant$ for all k. The assumption that all models are *a priori* equally likely is considered by Hoeting *et al.* (1999) to be a "reasonable 'neutral' choice" when there is insufficient prior reason to prefer one model over another. Draper (1999) and George (1999) express concern that if two models are near equivalent as concerns predictions, treating them as separate equally likely models amounts to giving double weight to a single model of which there are two slightly different versions, thereby "diluting" the predictive power of BMA. One way to minimize this effect is to eliminate at the outset models that are deemed potentially inferior. Another is to retain only models that are structurally distinct and non-collinear. Otherwise, one should consider reducing (diluting) the prior probabilities assigned to models that are deemed closely related.

Kashyap's (1982) purpose in developing (6) was to derive an optimum decision rule for selecting one among several competing models, unrelated to BMA. Since the first term on the right hand side is constant and the last is asymptotically zero, Kashyap proposed to select that model which minimizes the criterion

$$\hat{d}_k = -\ln p\left(\boldsymbol{D}\middle|\hat{\boldsymbol{\theta}}_k, M_k\right) - \ln p\left(\hat{\boldsymbol{\theta}}_k \middle| M_k\right) - \frac{N_k}{2}\ln\left(\frac{2\pi}{N}\right) + \frac{1}{2}\ln\left|\boldsymbol{F}_k\left(\boldsymbol{D}\middle|\hat{\boldsymbol{\theta}}_k, M_k\right)\right| \tag{9}$$

Increasing the number of parameters N_k allows $-\ln p\left(\boldsymbol{D}\middle|\hat{\boldsymbol{\theta}}_k, M_k\right)$ 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 $\hat{d}_k$ grows while $p\left(M_k \middle| \boldsymbol{D}\right)$ in (6) diminishes. This means that a more parsimonious model with fewer parameters is preferred by (9) and assigned a higher probability by (6). On the other hand $-\ln p\left(\boldsymbol{D}\middle|\hat{\boldsymbol{\theta}}_k, M_k\right)$ diminishes with N at a rate higher than linear so that as the latter grows, there may be an advantage to a more complex model with larger N_k.

The last term in (9) reduces the relative emphasis on model fit as the information content of the data diminishes. As illustrated by Carrera and Neuman (1986b), it may cause one to prefer a simpler model that leads to a poorer fit with the data over a more complex model that fits the data better. The term tends to a constant as N becomes large, so that $\hat{d}_k$ becomes asymptotically equivalent to the Bayes information criterion BIC derived by Akaike (1977), Rissanen (1978) and Schwarz (1978) on the basis of different considerations. Raftery (1993) proposed adopting the asymptotic BIC approximation

$$\ln p\left(\boldsymbol{D}\middle|M_k\right) \approx \ln p\left(\boldsymbol{D}\middle|\hat{\boldsymbol{\theta}}_k, M_k\right) - \frac{N_k}{2}\ln N \tag{10}$$

for BMA (see also Raftery *et al.* 1996; Volinsky *et al.* 1997; Hoeting *et al.* 1999).

To my knowledge, the nonasymptotic expression (6) has not been previously incorporated into BMA. I propose to do so because environmental models seldom satisfy the assumption that N is large. To render the use of (6) in BMA computationally feasible, I propose to follow the approach of Carrera and Neuman (1986a) who incorporate in $\boldsymbol{D}$ both observational data such as head and prior estimates of the parameters, and treat the two sets as being mutually uncorrelated while allowing internal correlations between members of each set. This allows them to incorporate $\ln p\left(\hat{\boldsymbol{\theta}}_k\middle|M_k\right)$ into the log likelihood function $\ln p\left(\boldsymbol{D}\middle|\hat{\boldsymbol{\theta}}_k, M_k\right)$ and to compute $\boldsymbol{F}_k\left(\boldsymbol{D}\middle|\hat{\boldsymbol{\theta}}_k, M_k\right)$ in a straightforward manner (also Carrera *et al.* 1997).

CONCLUSION

Bayesian model averaging provides an optimal way to combine the predictions of several competing conceptual-mathematical models and to assess their joint predictive uncertainty. It can be made computationally feasible by basing it on a maximum likelihood approximation due to Kashyap (1982) and the parameter estimation method of Carrera and Neuman (1986a).

Acknowledgment This work was supported by the U.S. Nuclear Regulatory Commission under contract NRC-04-97-056. Special credit is due to Thomas J. Nicholson and Ralph Cady of the USNRC for having conceived and championed this research. I am most grateful to Weihsueh Chiu of the U.S. Environmental Protection Agency for having brought to my attention the extensive work on BMA conducted by the "Seattle school" of statistics at the University of Washington, and for his helpful comments on this paper.

REFERENCES

Akaike, H. (1977) On entropy maximization principle. In: *Applications of Statistics* (ed. by P. R. Krishnaiah), 27-41. North Holland, Amsterdam, The Netherlands.

Beven, K. J. (1993) Prophecy, reality and uncertainty in distributed hydrological modelling. *Adv. Water Resour.* **16**, 41- 51.

Beven, K. J. (2000) Uniqueness of place and non-uniqueness of models in assessing predictive uncertainty. In: *Computational Methods in Water Resources XIII* (ed. by L. R. Bentley, J.F. Sykes, W.G. Gray, C.A. Brebbia & G.F. Pinder), 1085-1091. Balkema, Rotterdam, The Netherlands.

Beven, K. J. & Binley, A. M. (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**, 279 – 298.

Beven, K. J. & Freer, J. (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrology* **249**, 11-29.

Carrera, J. & Neuman S. P. (1986a) Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Water Resour. Res.* **22**(2), 199-210.

Carrera, J. & Neuman, S. P. (1986b) Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data. *Water Resour. Res.* **22**(2), 228-242.

Carrera, J., Medina, A., Axness, C. & Zimmerman, T. (1997) Formulations and computational issues of the inversion of random fields. In: *Subsurface Flow and Transport: A Stochastic Approach* (ed. by G. Dagan & S. P. Neuman), 62-79. Cambridge University Press, Cambridge, United Kingdom.

Dagan, G. & Neuman, S. P. (1997) *Subsurface Flow and Transport: A Stochastic Approach*. Cambridge University Press, Cambridge, United Kingdom.

Draper, D. (1995) Assessment and propagation of model uncertainty. *J. Roy. Statist. Soc. Ser. B* **57**, 45-97.

Draper, D. (1999) Comment. *Statist. Sci.* **14**(4), 405-409.

Gaganis, P. & Smith, L. (2001) A Bayesian approach to the quantification of the effect of model error on the predictions of groundwater models. *Water Resour. Res.* **37**(9), 2309-2322.

George, E.I. (1999) Comment. *Statist. Sci.* **14**(4), 409-412.

Gómez-Hernández J. J., Sahuquillo, A. & Capilla, J.E. (1997) Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, 1. Theory. *J. Hydrology* **203**, 162-174.

Guadagnini, A. & Neuman, S. P. (1999) Nonlocal and localised analyses of conditional mean steady state flow in bounded, randomly nonuniform domains, 2, Computational examples. *Wat. Resour. Res*. **35**(10), 3019-3039.

Hernandez, A. F., Neuman, S. P. Guadagnini, A. & Carrera, J. (2003) Conditioning steady state mean stochastic flow equations on head and hydraulic conductivity measurements. In: *Calibration and Reliability in Groundwater Modelling: Coping with Uncertainty, ModelCARE'2002* (ed. by Z. Hrkal, J. Krasny & K. Kovar), 158-162. IAHS Publication, IAHS Press, Wallingford, Oxfordshire, United Kingdom.

Hill, C. M. (1998) Methods and Guidelines for Effective Model Calibration. *U.S. Geological Survey Water-Resources Investigations Report 98-4005*, Denver, Colorado.

Hoeting, J. A., Madigan, D., Raftery, A.E. & Volinsky, C.T. (1999) Bayesian model averaging: A tutorial. *Statist. Sci.* **14**(4), 382-417.

Hornberger, G. M. & Speer, R. C. (1981) An approach to the preliminary analysis of environmental systems. *J. Environ. Mgmt.* **12**, 7-18.

James, A. L. & Oldenburg. C.M. (1997) Linear and Monte Carlo uncertainty analysis for subsurface contaminant transport simulation. *Water Resour. Res.* **33**(11), 2495-2508.

Kashyap, R. L. (1982) Optimal choice of AR and MA parts in autoregressive moving average models. *IEEE Trans. Pattern Anal. Mach. Intel. PAMI* **4**(2), 99-104.

Madigan, D. & Raftery, A. E. (1994) Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Amer. Statist. Assoc.* **89**, 1535-1546.

National Research Council (1999) *Groundwater & Soil Cleanup*. Nat. Acad. Press, Washington, D.C.

National Research Council (2001) *Conceptual Models of Flow and Transport in the Vadose Zone*. Nat. Acad. Press, Washington, D.C.

Raftery, A. E. (1993) Bayesian model selection in structural equation models. In: *Testing Structural Equation Models* (ed. by K. Bollen & J. Long), 163-180. Sage, Newbury Park, California.

Raftery, A.E., Madigan, D. & Volinsky, C. T. (1996) Accounting for model uncertainty in survival analysis improves predictive performance. In: *Bayesian Statistics* (ed. by J. Bernardo, J. Beerger, A. Dawid & A. Smith), 323-349. Oxford Univ. Press, United Kingdom.

Rissanen, J. (1978) Modeling by shortest data description. *Automatica* **14**, 465-471.

Samper, J. & Molinero, J. (2000) Predictive uncertainty of numerical models of groundwater flow and solute transport. In: AGU 2000 Fall Meeting Program, Washington, D.C., F500.

Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Statist.* **6**, 461-464.

Taplin, R. H. (1993) Robust likelihood calculation for time series. *J. Roy. Statist. Soc. Ser. B* **55**, 829-836.

Volinsky, C. T., Madigan, D., Raftery, A. E. & Kronmal, R. A. (1997) Bayesian model averaging in proportional hazard models: assessing the risk of a stroke. *J. Roy. Statist. Soc. Ser. C* **46**, 433-448.

Ye M., Neuman, S. P., Guadagnini, A. & Tartakovsky, D.M. (2002) Nonlocal and localized analyses of conditional mean transient flow in bounded, randomly nonuniform domains. In: *Computational Methods in Water Resources XIV* (ed. by M. Hassanizadeh, R.J. Schotting, W.G. Gray, G.F. Pinder & C.A. Brebbia), 1155-1162. Elsevier, Amsterdam, The Netherlands.

Zhang, D. (2001) *Stochastic Methods for Flow in Porous Media*. Academic Press.

Zimmerman, D. A., De Marsily, G., Gotway, C. A., Marietta, M. G., Axness, C. L., Beauheim, R. L., Bras, R., Carrera, J., Dagan, G., Davies, P. B., Gallegos, D. P., Galli, A., Gómez-Hernández, J., Gorelick, S. M., Grindrod, P., Gutjahr, A. L., Kitanidis, P. K., Lavenue, A. M., McLaughlin, D., Neuman, S. P., Rama Rao, B. S., Ravenne, C. & Rubin, Y. (1998) 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.

NRC FORM 335 (2-89) NRCM 1102, 3201, 3202

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

1. REPORT NUMBER
(Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)

NUREG/CR-6805

2. TITLE AND SUBTITLE

A Comprehensive Strategy of Hydrogeologic Modeling and Uncertainty Analysis for Nuclear Facilities and Sites

3. DATE REPORT PUBLISHED

MONTH	YEAR
July	2003

4. FIN OR GRANT NUMBER

N/A

5. AUTHOR(S)

S.P. Neuman and P.J. Wierenga

6. TYPE OF REPORT

Technical

7. PERIOD COVERED *(Inclusive Dates)*

September 1997 - December 2002

8. PERFORMING ORGANIZATION - NAME AND ADDRESS *(If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)*

University of Arizona
Tucson, AZ 85721

9. SPONSORING ORGANIZATION - NAME AND ADDRESS *(If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)*

Division of Systems Analysis and Regulatory Effectiveness
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

10. SUPPLEMENTARY NOTES

T.J. Nicholson, NRC Project Manager

11. ABSTRACT *(200 words or less)*

This report describes a strategy that embodies a systematic and comprehensive a pproach to hydrogeologic conceptualization, model development and predictive uncertainty analysis. The strategy is comprehe nsive in that it considers all stages of model building and accounts jointly for uncertainties that arise at each of them. The stages include regional and site characterization, hydrogeologic conceptualization, development of conceptual-mathematical model structure, parameter estimation on the basis of monitored system behavior, and assessment of predictive uncertainty. In addi tion to parameter uncertainty, the strategy concerns itself with uncertainties arising from incomplete definitions of (a) t he conceptual framework that determines model structure, (b) spatial and temporal variations in hydrologic variables that are either not fully captured by the available data or not fully resolved by the model, and (c) the scaling behavior of hydrogeologic variables. The strategy is generic but designed to be of practical use to NRC licensing staff in their review of decommissionin g plans, and performance assessment of high-level and low-level radioactive waste disposal sites as well as uranium re covery facilities. An important component of the strategy is a systematic sequence of logical questions, guidelines and criteria with analytical methods appropriate for NRC review and performance evaluation.

12. KEY WORDS/DESCRIPTORS *(List words or phrases that will assist researchers in locating the report.)*

conceptual model uncertainty
ground-water analysis
hydrogeologic modeling
model development
performance assessment
predictive uncertainty analysis
sensitivity
uncertainty analysis

13. AVAILABILITY STATEMENT

unlimited

14. SECURITY CLASSIFICATION

(This Page)

unclassified

(This Report)

unclassified

15. NUMBER OF PAGES

16. PRICE