

Case Study/

Practical Postcalibration Uncertainty Analysis: Yucca Mountain, Nevada

by Scott C. James¹, John E. Doherty², and Al-Aziz Eddebbarh³

Abstract

The values of parameters in a groundwater flow model govern the precision of predictions of future system behavior. Predictive precision, thus, typically depends on an ability to infer values of system properties from historical measurements through calibration. When such data are scarce, or when their information content with respect to parameters that are most relevant to predictions of interest is weak, predictive uncertainty may be high, even if the model is "calibrated." Recent advances help recognize this condition, quantitatively evaluate predictive uncertainty, and suggest a path toward improved predictive accuracy by identifying sources of predictive uncertainty and by determining what observations will most effectively reduce this uncertainty. We demonstrate linear and nonlinear predictive error/uncertainty analyses as applied to a groundwater flow model of Yucca Mountain, Nevada, the United States' proposed site for disposal of high-level radioactive waste. Linear and nonlinear uncertainty analyses are readily implemented as an adjunct to model calibration with medium to high parameterization density. Linear analysis yields contributions made by each parameter to a prediction's uncertainty and the worth of different observations, both existing and yet-to-be-gathered, toward reducing this uncertainty. Nonlinear analysis provides more accurate characterization of the uncertainty of model predictions while yielding their (approximate) probability distribution functions. This article applies the above methods to a prediction of specific discharge and confirms the uncertainty bounds on specific discharge supplied in the Yucca Mountain Project License Application.

¹Corresponding author: Sandia National Laboratories, Thermal/Fluid Science & Engineering, PO Box 969, Livermore, CA 94551-0969; (925) 294-2616; fax (925) 294-3410; scjames@sandia.gov

²Watermark Numerical Computing, 336 Clivenden Avenue, Corinda 4075 Queensland, Australia; +61(7)3379-1664; johndoherty@ozemail.com.au

³Los Alamos National Laboratory, Earth and Environmental Sciences, EES-6 Group Mail Stop T003, PO Box 1663, Los Alamos, NM 87545-0001 aeddebba@lanl.gov

Corrections added after online publication September 10, 2009: on page 2, "parameterization" should have been "parameterization"; on page 3, Equation 3 was incorrect and has now been fixed; on page 4, Equation 15 was incorrect and has now been fixed; on page 5, two instances of "Tarameter" should have been "Parameter"; on page 6, "the composite objective function described rises not higher" should have been "the composite objective function described rises no higher"; on page 7, a comma was missing after "Tiedeman et al. (2003, 2004)". We apologize for these errors.

Received January 2009, accepted August 2009.

Copyright © 2009 The Author(s)

Journal compilation © 2009 National Ground Water Association.

doi: 10.1111/j.1745-6584.2009.00626.x

Introduction

Modeling is routine in environmental management investigations, especially at sites where heavy investment may follow important decisions. Models provide the bridge between our understanding of environmental processes formulated through mathematical descriptors, such as Darcy's law, and predictions of system behavior, perhaps subject to remediation efforts. In doing this, they provide important decision-making support.

Models serve another important role in environmental management—a role that is often insufficiently recognized. Specifically, this role is to process *all* available site data, both hard and soft. Examples of such data include qualitative impressions of the disposition of geological layering supported by borehole intersections at a small number of locations, geophysical data, point measurements of hydraulic properties, historical measurements of system state at different locations, geochemical inference, and so on. A good model will extract maximum information content from site data during the calibration process

while also reproducing historical site measurements. This leads to:

1. Predictions of future system behavior with decreased uncertainty.
2. The ability to quantify uncertainty.
3. The ability to examine the contribution of dataset members in reducing uncertainty to its current level.
4. The ability to quantify contributions to uncertainty from model parameterization.
5. The ability to quantify how additional data collection activities could reduce model predictive uncertainty.

Where a modeling exercise addresses all of the issues listed above, it is indeed paying returns on the investment of model development. For this to be accomplished, however, the model must be used in conjunction with appropriate parameter-estimation and uncertainty analysis software that maximizes the potential of the model in relation to the goals listed. It is the authors' contention that the use of such software in conjunction with complex, site-specific models should be viewed as an indispensable component of model-based decision making and management.

This article illustrates state-of-the-art uncertainty analysis techniques used in conjunction with a saturated zone groundwater model of the Yucca Mountain site

in Nevada (Figure 1). All of the techniques discussed can be implemented with relatively little computational effort. Although they have all been outlined elsewhere, documentation of their combined application as an adjunct to calibration of a complex three-dimensional model is hereby presented for the first time. Herein, we specifically:

1. Record the outcomes of a suite of uncertainty analyses that collectively implement optimum data interpretation in conjunction with a complex, three-dimensional model.
2. Demonstrate the benefits gained by each of these analyses both individually and in concert.

Methods discussed herein are applied to the Yucca Mountain (Nye County, Nevada) site-scale saturated zone flow model (YMPSZFM or the model) documented by SNL (2007, search DN2002478808 at <http://www.lsnnet.gov>). This model was developed as part of the licensing requirements for the Yucca Mountain nuclear waste repository. The model is briefly described; a more detailed description of its construction and calibration is provided in a comprehensive Yucca Mountain Analysis and Model Report (SNL 2007). All software used for these analyses are part of the PEST suite (Doherty

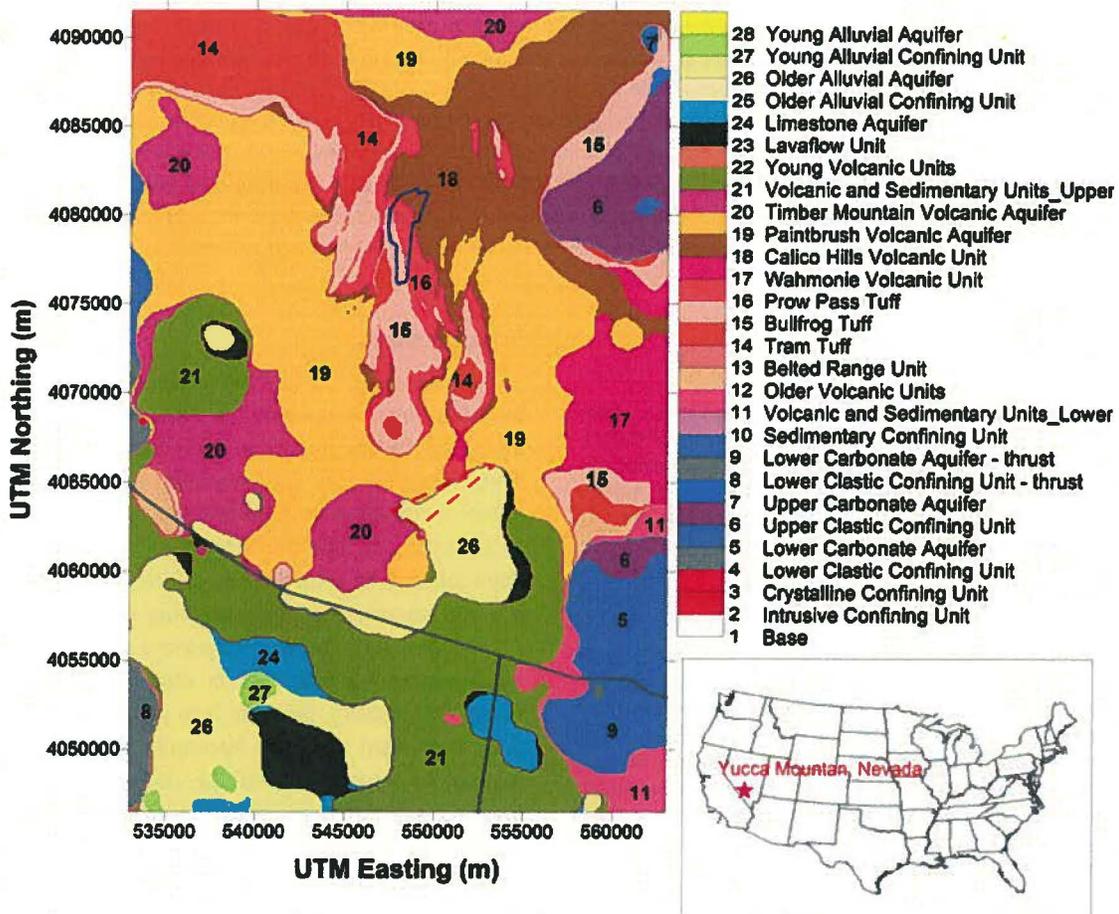


Figure 1. Geologic units (sliced at the water table) featured in the YMPSZFM (from SNL [2007]).

2009a, 2009b), which is in the public domain and is extensively documented.

This case study provides insights into the identifiability and sensitivity of permeability in major hydrogeologic units and structural features in the central part of the Death Valley regional groundwater flow system. In addition, uncertainty in groundwater-specific discharge, which is relevant to the rates of migration of radionuclides from the proposed repository at Yucca Mountain and from underground nuclear testing at the Nevada Test Site, is quantitatively evaluated.

This article is organized as follows. First, a brief review is provided on the theoretical underpinnings of the methodologies. A short description of the model and of data used in its calibration and construction follows. Then, a number of uncertainty-related analyses as applied to the YMPSZFM are described. Finally, conclusions and ramifications for general modeling practice are provided.

Theoretical Background

General

Hunt et al. (2007) and Moore and Doherty (2006) suggest that simulation of complex environmental processes by numerical models requires a level of parameterization complexity commensurate with the scale of process and hydrogeologic heterogeneity within a study area. They note that this does not ensure that the model will make correct predictions. What it does provide, however, is the ability to compute confidence intervals that embrace the full range of predictive possibilities. Meanwhile, these confidence intervals are no larger than they need to be because they are conditioned on the available calibration data set, and because the calibration process has extracted maximum information from that dataset.

Several methods have been developed for calibration and uncertainty analyses of highly parameterized models (Gómez-Hernández et al. 1997 and papers cited therein; Kitanidis 1996; Woodbury and Ulrych 2000). In this article, we use the term “highly parameterized” to refer to models where parameter numbers are such that their estimation requires solution of an ill-posed inverse problem. Although the YMPSZFM does not possess as many parameters as many other models, it does possess many more parameters than are capable of unique estimation on the basis of the available dataset. Methodologies used in the present study are based on concepts outlined for linear model behavior by Moore and Doherty (2005) and for nonlinear model behavior by Tonkin and Doherty (2005, 2009) and Tonkin et al. (2007). These comprise a family of techniques rooted in the theory of mathematical inversion as outlined in various texts (Aster et al. 2005; Menke 1989). In the quasi-linear context, these methodologies allow rapid statistical characterization of model parameter and predictive error, parameter contributions to predictive error, and the relative efficacy of different existing or hypothetical data in reducing that error. In the nonlinear context, these techniques facilitate a complete

characterization of model predictive error variance, taking into full account the complex nature of hydrologic property variability and of other processes affecting flow and transport. A brief overview of pertinent theory is presented next.

Parameter and Predictive Error

Let model parameters comprise the vector \mathbf{p} . Let the action of a (linear) model on these parameters be characterized by the matrix \mathbf{X} . Let the vector \mathbf{h} represent observations of system state comprising the calibration dataset; these are contaminated by measurement noise $\boldsymbol{\varepsilon}$. Then:

$$\mathbf{h} = \mathbf{X}\mathbf{p} + \boldsymbol{\varepsilon} \quad (1)$$

Let model-to-measurement misfit be characterized by an objective function Φ defined as:

$$\Phi = (\mathbf{X}\mathbf{p} - \mathbf{h})^t \mathbf{Q} (\mathbf{X}\mathbf{p} - \mathbf{h}) \quad (2)$$

where \mathbf{Q} is an appropriate weight matrix. The superscript t indicates matrix transpose. \mathbf{Q} can be (but does not have to be) chosen such that it is proportional to the inverse of the covariance matrix of measurement noise $C(\boldsymbol{\varepsilon})$ so that:

$$\mathbf{Q} = \sigma_r^2 C^{-1}(\boldsymbol{\varepsilon}) \quad (3)$$

where σ_r^2 is a reference variance. If the number of parameters comprising the vector \mathbf{p} is small enough to allow their unique estimation on the basis of the current calibration dataset, then minimization of Φ leads to an estimated set of parameters $\underline{\mathbf{p}}$:

$$\underline{\mathbf{p}} = (\mathbf{X}^t \mathbf{Q} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Q} \mathbf{h} \quad (4)$$

However, suppose that there are more elements in \mathbf{p} than can be uniquely inferred on the basis of the current calibration dataset. Suppose further that a suitably low objective function, one not so low that “overfitting” has occurred, can be obtained on the basis of a vector $\underline{\mathbf{p}}$ calculated from \mathbf{h} as:

$$\underline{\mathbf{p}} = \mathbf{G}\mathbf{h} \quad (5)$$

(An appropriate \mathbf{G} will be derived shortly.) If Equation 1 is substituted into Equation 5, we have:

$$\underline{\mathbf{p}} = \mathbf{G}\mathbf{X}\mathbf{p} + \mathbf{G}\boldsymbol{\varepsilon} = \mathbf{R}\mathbf{p} + \mathbf{G}\boldsymbol{\varepsilon} \quad (6)$$

where $\mathbf{R} = \mathbf{G}\mathbf{X}$ is the resolution matrix. If there is no measurement noise, the “ i ”th row of \mathbf{R} provides the averaging function through which an estimated parameter p_i (the “ i ”th element of $\underline{\mathbf{p}}$) is related to the unknown real-world hydraulic properties encapsulated in \mathbf{p} . From Equation 6, parameter error is given by:

$$\underline{\mathbf{p}} - \mathbf{p} = -(\mathbf{I} - \mathbf{R})\mathbf{p} + \mathbf{G}\boldsymbol{\varepsilon} \quad (7)$$

from which the covariance matrix of parameter error $C(\underline{\mathbf{p}} - \mathbf{p})$ is readily computed as:

$$C(\underline{\mathbf{p}} - \mathbf{p}) = (\mathbf{I} - \mathbf{R})C(\mathbf{p})(\mathbf{I} - \mathbf{R})^t + \mathbf{G}C(\boldsymbol{\varepsilon})\mathbf{G}^t \quad (8)$$

where $C(\mathbf{p})$ (which must be supplied by the modeling and site characterization teams) bounds the variability of subsurface hydraulic properties. Specifically, it characterizes the current state of geological knowledge and geological uncertainty associated with a study site. Geologic uncertainty is expressed through nonzero diagonal elements. Geologic knowledge is expressed through nonzero off-diagonal elements (indicating that something is known of the spatial correlation of hydraulic properties) and finite diagonal elements (indicating that there are bounds on geological uncertainty).

Let a prediction s depend on parameters \mathbf{p} through the sensitivities recorded in the vector \mathbf{y} . That is:

$$s = \mathbf{y}^t \mathbf{p} \quad (9)$$

A model prediction, \underline{s} , is calculated from calibrated parameters $\underline{\mathbf{p}}$ as:

$$\underline{s} = \mathbf{y}^t \underline{\mathbf{p}} \quad (10)$$

so that predictive error is:

$$\underline{s} - s = \mathbf{y}^t (\underline{\mathbf{p}} - \mathbf{p}) \quad (11)$$

Predictive error can never be known because the true parameters \mathbf{p} of the system are not known. However, using Equation 8, predictive error variance $\sigma_{\underline{s}-s}^2$ can be calculated as:

$$\begin{aligned} \sigma_{\underline{s}-s}^2 &= \mathbf{y}^t C(\underline{\mathbf{p}} - \mathbf{p}) \mathbf{y} \\ &= \mathbf{y}^t (\mathbf{I} - \mathbf{R}) C(\mathbf{p}) (\mathbf{I} - \mathbf{R})^t \mathbf{y} + \mathbf{y}^t \mathbf{G} C(\epsilon) \mathbf{G}^t \mathbf{y} \quad (12) \end{aligned}$$

Gallagher and Doherty (2006) demonstrate the use of the preceding equation in computing the error variance associated with a variety of predictions made by a single-layer, water-resource-management groundwater model.

Singular Value Decomposition

Equation 4 cannot be used to calculate $\underline{\mathbf{p}}$ when more parameters that can be estimated uniquely on the basis of the current calibration dataset are sought, because $\mathbf{X}^t \mathbf{Q} \mathbf{X}$ is not invertible under these conditions. Hence, computation of a suitable value for $\underline{\mathbf{p}}$ requires modification of this equation.

Singular value decomposition (SVD) of $\mathbf{X}^t \mathbf{Q} \mathbf{X}$ leads to two matrices \mathbf{V} and \mathbf{S} :

$$\mathbf{X}^t \mathbf{Q} \mathbf{X} = \mathbf{V} \mathbf{S} \mathbf{V}^t \quad (13)$$

which, after partitioning of the \mathbf{S} and \mathbf{V} matrices, is equivalent to

$$\begin{aligned} \mathbf{X}^t \mathbf{Q} \mathbf{X} &= [\mathbf{V}_1 \quad \mathbf{V}_2] \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^t \\ \mathbf{V}_2^t \end{bmatrix} \\ &= \mathbf{V}_1 \mathbf{S}_1 \mathbf{V}_1^t + \mathbf{V}_2 \mathbf{S}_2 \mathbf{V}_2^t \quad (14) \end{aligned}$$

where \mathbf{S} is a diagonal matrix of singular values (arranged in decreasing order) and \mathbf{V} is a matrix of orthogonal unit vectors (which are also eigenvectors of $\mathbf{X}^t \mathbf{Q} \mathbf{X}$) spanning parameter space. Note that the second version of the

preceding equation explicitly separates the solution space (subscript 1) from the null space (subscript 2). Orthogonal unit vectors spanning the former subspace comprise the columns of \mathbf{V}_1 , whereas orthogonal unit vectors comprising the latter subspace comprise the columns of \mathbf{V}_2 . These subspaces are orthogonal to each other; collectively they comprise the entirety of parameter space. The latter subspace is associated with singular values that are zero or very low, and these being contained within the \mathbf{S}_2 matrix. These are eliminated through truncation of parameter space in the truncated SVD procedure through which the ill-posed inverse problem of estimating $\underline{\mathbf{p}}$ is solved.

The interested reader may review the work of Tonkin and Doherty (2005) and Gallagher and Doherty (2006) for more detailed descriptions of the roles of the solution and null spaces in the model calibration process. Mathematically, any vector $\delta \mathbf{p}$ for which $\mathbf{X} \delta \mathbf{p}$ is zero lies within the null space of the matrix \mathbf{X} . From the existence of a null space follows nonuniqueness of the inverse problem of model calibration. This arises out of the fact that if any vector \mathbf{p} satisfies Equation 1, so too does the vector $\mathbf{p} + \delta \mathbf{p}$. In calibrating a model, we normally search for the simplest \mathbf{p} that satisfies Equation 1; thus, we represent in the calibrated model no parameterization detail other than that required to explain the data. Parameterization complexity beyond this, which does not detract from the model's ability to replicate past system behavior (and hence by definition lies within the null space of \mathbf{X}), can only be represented in probabilistic terms (as we do herein).

It can be shown that if truncated SVD is used for solution of the ill-posed inverse problem of estimating $\underline{\mathbf{p}}$, then \mathbf{G} and \mathbf{R} of Equations 5 and 6 become

$$\mathbf{G} = (\mathbf{V}_1 \mathbf{S}_1^{-1} \mathbf{V}_1^t) \mathbf{X}^t \mathbf{Q} \quad (15)$$

$$\mathbf{R} = \mathbf{V}_1 \mathbf{V}_1^t \quad (16)$$

It is easily shown that Φ in Equation 2 is minimized if truncation is such as to relegate only zero, or extremely low, singular values to \mathbf{S}_2 . However, as Moore and Doherty (2005) show, it is best to truncate at singular values that are higher than zero, with the truncation level depending on the amount of noise associated with the measurement dataset \mathbf{h} . This avoids overfitting to the calibration dataset and consequential amplification of the contribution of measurement noise to parameter and predictive error (see the minimum of the bold curve in Figure 5).

Equation 12 indicates that there are two contributors to parameter and predictive error variance. The first term of this equation, comprising the null space contribution to predictive error variance, arises from parameter simplifications necessary to obtain a unique solution to the inverse problem (i.e., model calibration). In other words, it represents errors arising from the inability of calibrated model parameters (as estimated on the basis of the calibration dataset \mathbf{h}) to represent the innate complexity of the real world. The second term of Equation 12, the solution space

contribution to parameter and predictive error variance, arises from the fact that even the simplified parameterization used in a calibrated model is potentially in error because its estimation is based on a dataset contaminated by noise.

If $C(\mathbf{p})$ is diagonal, parameters are normalized with respect to their innate variability [so that the diagonal elements of $C(\mathbf{p})$ are 1], and noise ϵ is zero, it can be shown that calibrating a model using truncated SVD, where truncation is such that \mathbf{S}_2 is $\mathbf{0}$, leads to a minimum norm (and hence maximum likelihood) solution for \mathbf{p} . The situation is slightly different when the measurement dataset is accompanied by noise. As more singular values are included in the parameter-estimation process, and as the truncation point thereby shifts toward singular values of lower value (to the right in Figure 5), the first term on the right-hand side of Equation 12 falls, whereas the second term rises. The total error variance first falls, and then rises, as the number of pretruncation singular values increases. Truncation should take place at the singular value where the error variance for a prediction of interest is minimized.

Note that truncation at zero singular values is equivalent to using no data in the calibration process. Parameters thus maintain their initial values, and postcalibration parameter and predictive error variance are identical to precalibration parameter and predictive uncertainty.

Parameter Identifiability

The diagonal elements of the resolution matrix \mathbf{R} defined by Equation 16 have special significance and can be used as a measure of parameter identifiability (Doherty and Hunt 2009). The i th diagonal element of \mathbf{R} is the cosine of the angle between a vector in the direction of the i th parameter and its projection onto the calibration solution space. Its value can vary between zero and one. If it is zero, the parameter is completely *unidentifiable*; if it is one, the parameter is completely *identifiable*. If the identifiability of parameter i is designated as f_i , it can be computed as:

$$f_i = (\mathbf{V}_1 \mathbf{V}_1^t)_{ii} = \mathbf{i}^t (\mathbf{V}_1 \mathbf{V}_1^t) \mathbf{i} \quad (17)$$

where \mathbf{i} is a unit vector in the direction of the parameter.

If a parameter has an identifiability of one, it does not follow that it can be estimated without error because its estimation is based on a dataset contaminated by measurement noise. The error in an individual estimated parameter can be computed using Equation (12) with \mathbf{y} replaced by $p_i \mathbf{i}$, where p_i is the “ i ”th element of \mathbf{p} . Its “relative error reduction” (designated as e_i) is defined as:

$$e_i = 1 - \sqrt{\frac{\sigma_i^2}{[\sigma_i^2]_0}} \quad (18)$$

where $[\sigma_i^2]_0$ is the precalibration variance of the parameter [i.e., the diagonal element of $C(\mathbf{p})$ corresponding to this parameter]. Ideally, like identifiability, relative parameter error reduction should range between zero and one, with

a value of one corresponding to full solution space occupancy and estimation on the basis of no measurement noise, whereas a value of zero signifies the absence of identifiability. Unfortunately, e_i can sometimes become negative because it is possible that improper truncation (leading to overfitting) can actually increase the potential for error in some parameters (and also the predictions that depend on them) rather than decreasing it. It is noteworthy that amplification, rather than reduction, of parameter and predictive error can occur if regularization is undertaken recklessly, irrespective of whether regularization is achieved through mathematical means such as truncated SVD or manually through precalibration lumping, tying, or fixing of parameters.

As will be demonstrated below in the context of the YMPFSZFM, the linear theory discussed above provides useful insights into:

1. Parameter and predictive credibility.
2. Parameter contributions to predictive uncertainty.
3. Observation (both existing and hypothetical) contributions to predictive uncertainty reduction.

Next we demonstrate how to extend these linear methods to yield quantitative estimates of the error variance associated with predictions made by a complex nonlinear model.

Predictive Uncertainty and Predictive Error Variance

The methodologies discussed begin with the concept of a calibrated model and then explore the potential for error in both model parameter estimates and in predictions made by a calibrated model. An alternative approach is to compute parameter and predictive *uncertainty* rather than *error variance*, thus adopting a Bayesian rather than frequentist (Annis 2008) outlook. If the regularization methodology used in model calibration is carefully chosen and implemented, the two will yield similar estimates of postcalibration parameter and predictive variability. However, uncertainty is more of an intrinsic property of the available dataset than is error and is thus independent of the regularization device used to obtain a unique solution to the inverse problem of model calibration. The variance of parameter and predictive uncertainty will therefore be slightly lower than that of parameter and predictive error (much lower if regularization is carelessly applied).

For a linear model, if \mathbf{p} and ϵ are normally distributed, the variance of uncertainty of a prediction s conditioned by a calibration dataset \mathbf{h} is (Christensen and Doherty 2008):

$$\sigma_s^2 = \mathbf{y}^t C(\mathbf{p}) \mathbf{y} - \mathbf{y}^t C(\mathbf{p}) \mathbf{X}^t [\mathbf{X} C(\mathbf{p}) \mathbf{X}^t + C(\epsilon)]^{-1} \mathbf{X} C(\mathbf{p}) \mathbf{y} \quad (19)$$

The first term is the precalibration uncertainty of the prediction, this being solely a function of innate parameter variability expressed by $C(\mathbf{p})$ and the dependence of the prediction on model parameters expressed by the sensitivity vector \mathbf{y} . The second term embodies the reduction

in predictive uncertainty variance accrued by the conditioning effect of historical measurements of system state (i.e., calibration). Similar to Equation 12, Equation 19 can be used to explore the contribution of different parameter types to overall predictive uncertainty, as will be demonstrated shortly for the YMPSZFM. It can also be used to determine the relative efficacy of various existing or potential observations in reducing that uncertainty.

The relative uncertainty reduction of an individual parameter can be computed using Equation 18. However, in this case, σ_i^2 in that equation is computed using Equation 19 (with \mathbf{y} replaced by \underline{p}_i). Unlike relative error reduction, relative uncertainty reduction cannot fall below zero.

Before progressing to nonlinear analysis, it is relevant to point out a feature of linear analysis that can greatly contribute to its general use. Neither Equation 12 nor Equation 19 includes actual parameter or observation values; they only include parameter and observation sensitivities, or quantities derived from these. Hence, these equations can be used for exploration of model predictive error/uncertainty regardless of whether or not the model has actually been calibrated.

Nonlinear Error Variance Analysis

One means through which predictive error variance can be computed without the need for an assumption of model linearity is through modification of the constrained predictive maximization/minimization technique described for over determined parameter estimation by Vecchia and Cooley (1987). The following modifications allow this technique to be used in highly parameterized contexts when the inverse problem of model calibration is ill-posed. Implementation of the modified method requires the following steps to be taken:

1. Once calibrated, a model is reparameterized in terms of perturbations from the calibrated parameter set $\underline{\mathbf{p}}$.
2. A composite objective function is formed with two components. The first component measures deviations of an arbitrary parameter set \mathbf{p} from the calibrated parameter set $\underline{\mathbf{p}}$ after projection of these deviations onto the calibration null space. Weights are assigned on the basis of the $\mathbf{C}(\mathbf{p})$ matrix of innate parameter variability after projection of this matrix onto the calibration null space. The second objective function component measures deviations of model outputs from those calculated on the basis of $\underline{\mathbf{p}}$. Weights are assigned on the basis of the $\mathbf{C}(\epsilon)$ matrix of measurement noise.

A prediction of interest is then maximized/minimized subject to the constraint that the composite objective function described rises no higher than a user-specified limit. This limit determines the confidence level associated with the prediction.

Tonkin et al. (2007) detail a constrained maximization/minimization algorithm that implements this procedure and discuss the relationship between the magnitude

of the composite objective function limit and the concomitant confidence level of the maximized/minimized prediction.

Calibration-constrained Monte Carlo analysis constitutes a second means by which the predictive error variance of a nonlinear model can be explored. Like the constrained maximization/minimization methodology described, it is also based on Equation 8. Starting with a calibrated parameter set $\underline{\mathbf{p}}$, the following procedure is implemented:

1. Monte Carlo realizations of parameter fields are generated on the basis of the user-specified $\mathbf{C}(\mathbf{p})$ matrix.
2. Differences between these fields and the calibrated parameter field are calculated (thus generating realizations of $\mathbf{p}-\underline{\mathbf{p}}$).
3. These differences are projected onto the calibration null space using the $(\mathbf{I}-\mathbf{R})$ matrix achieved through the calibration process. This replicates parameter error variability described by the first term of Equation 8. These projected differences are then added to the calibrated parameter field $\underline{\mathbf{p}}$.
4. The model is then recalibrated, through adjustment of solution space parameter projections, to within a certain objective function threshold computed on the basis of $\mathbf{C}(\epsilon)$. This replicates parameter variability expressed by the second term of Equation 8.

The overall process is model-run-efficient because of the following:

1. Following null-space projection of parameter differences (step 3 above), the objective function is close to optimal, so that only minor recalibration is required.
2. Recalibration is undertaken by varying parameter combinations comprising \mathbf{V}_1 vectors spanning the calibration solution space, rather than individual parameters; only as many parameter combinations are varied as there are significantly nonzero singular values represented in the \mathbf{S}_1 matrix of Equation 14.
3. The first iteration of the recalibration process employs precalculated solution space sensitivities, these being reused for all Monte Carlo parameter realizations. For many realizations, only one iteration is required to achieve recalibration to the level of the specified objective function due to the close-to-optimal starting point of the recalibration process.

The outcome of this calibration-constrained Monte Carlo process is a suite of parameter realizations that are geologically realistic due to the fact that they were generated on the basis of a $\mathbf{C}(\mathbf{p})$ covariance matrix appropriate to the study site. Equally as important, these parameter realizations respect calibration constraints to within a threshold set by $\mathbf{C}(\epsilon)$. Any model prediction can then be made with all of these parameter realizations; the frequency histogram of that prediction characterizes the probability distribution of predictive error.

The Model

General

The YMPSZFM is outlined by SNL (2007); only those aspects of this model that are pertinent to the current analyses are summarized here.

Saturated zone water movement in the study area is simulated using the Finite Element Heat and Mass (FEHM) model (Bower and Zyvoloski 1997; Zyvoloski 1983). The YMPSZFM employs 69 layers and 956,345 nodes to simulate three-dimensional flow of groundwater within the study area. Steady-state flow conditions are assumed for calibration purposes.

The stratigraphy of the region (Figure 1) is based on a three-dimensional hydrogeologic framework model (HFM) that was developed from digital elevation models, geologic maps, borehole information, geologic and hydrogeologic cross sections, and other three-dimensional models to represent the geometry of the hydrogeologic units (HGUs). Structural features such as faults and fractures that affect groundwater flow also were added. The HFM represents Precambrian and Paleozoic crystalline and sedimentary rocks, Mesozoic sedimentary rocks, Mesozoic to Cenozoic intrusive rocks, Cenozoic volcanic tuffs and lavas, and late Cenozoic sedimentary deposits of the Death Valley regional groundwater flow system (DVRFS) region in 27 HGUs (Belcher 2004). In this case study, uncertainty in HGU permeability is reduced and quantified through application of tools available through the PEST (Doherty 2009a, 2009b) suite of codes by constraining permeability ranges assigned by expert opinion with site hydrologic data.

Groundwater flow in the Death Valley region is composed of several interconnected, complex groundwater flow systems. It occurs in three subregions in relatively shallow and localized flow paths that are superimposed on deeper, regional flow paths. Regional groundwater flow is predominantly through a thick Paleozoic carbonate rock sequence affected by complex geologic structures from regional faulting and fracturing that can enhance or impede flow (Belcher 2004).

Parameters

As discussed in detail by SNL (2007), the model grid was partitioned into 33 zones (including eight faults) of assumed hydraulic-property constancy (i.e., hydrogeologic units and features). Zonation was established through site characterization studies including drilling and geological mapping (Figure 1). Permeability multipliers for the 19 units in the altered northern region were deemed “adjustable” for the purpose of predictive uncertainty analyses. A single value of vertical anisotropy, assigned to all volcanic and alluvial units, was also deemed adjustable.

Model parameters are listed and briefly described in Table A1. Precalibration parameter uncertainties in Table A1 represent the authors’ best estimate of the standard deviations (in log space) of all adjustable parameters based on all available data (SNL 2007, section 7). If

no data were available for a geologic unit’s permeability (e.g., for parameter *uccu*), then the standard deviation was set to 1.5. For units with multiple measurements of permeability (e.g., parameter *cfta*), the standard deviation of the log of the measurements was used (SNL 2007, section 7.2.2). Because the uncertainty of permeability multipliers is to some extent already accommodated in the uncertainties assigned to pertinent unit permeabilities, these were set to 0.5 (except for *icum* and *xcum* which were set to 0.17 to ensure low permeabilities for these units in the altered northern region). Fault uncertainties were set to 0.67 (except for 0.33 for *4wfz*, reflecting the fact that it is known to be a conductive feature). Anisotropy uncertainty was set to the reasonably large value of 0.5 to reflect the range of vertical anisotropies that are possible in this alluvial and fractured volcanic system. Vertical head gradient multiplier uncertainties were specified as 0.02 to avoid unrealistically large vertical gradients. The infiltration multiplier uncertainty was set to 0.5 to reflect uncertainty and complexity of the infiltration processes. It should be noted that in specifying the uncertainty of parameters in this manner, exact characterization is not necessary; however, it is important that uncertainty be geologically reasonable (e.g., a crystalline unit at depth would not have a permeability overlapping that of surface sandstone). An advantage of working in the log domain is that variability is implicitly expressed in terms of factors rather than absolutes, which prevents implicit parameter negativity where parameter values are small and uncertainties are large.

Additional parameters (not adjusted during the model calibration process) were introduced in the linear analysis described subsequently to assess their contributions to model predictive error/uncertainty. To the extent that any of the values that were assumed for these additional parameters are incorrect, any prediction that depends directly on them has the potential for error. Moreover, any parameter whose values were misassigned during the calibration process because of inadvertent misassignment of fixed values to these five parameters also has the potential for error. These five parameters are (1) four multipliers that yield vertical gradients on the fixed heads assigned to the northern, southern, eastern, and western boundaries of the model domain, and (2) a multiplier applied to global recharge through the uppermost active layer of the model.

Assignment of values to parameters and/or boundary conditions independently of the calibration process is not uncommon when constructing groundwater models. However, notwithstanding advice provided by authors such as Tiedeman et al. (2003, 2004), testing the repercussions of these assumptions by way of their impact on predictive error is rare. Gallagher and Doherty (2006) point out that there are many occasions when arguments over the accuracy of particular boundary conditions between modelers and reviewers can be quantitatively resolved through demonstrating whether the potential for predictive uncertainty incurred through misassignment of these conditions is large or small (i.e., the impacts of such assumptions

on model predictions are identified). It is precisely such a line of inquiry that motivated us to consider whether vertical gradients on the boundaries of the YMPSZFM or a recharge multiplier over the potentiometric surface of the YMPSZFM had significant impacts on specific discharge near Yucca Mountain. (Past reviews of the model have questioned whether vertical gradients on the boundaries or changes to infiltration could significantly impact estimates of specific discharge.) Those assumptions (such as that of no vertical gradient on the model boundaries) that have limited capacity to induce error in critical model predictions require little attention in this process. For those that do (like the recharge multiplier), the impact of different hydrogeological opinions can be quantified.

A total of 58 model parameters were deemed as adjustable for the purpose of linear analysis (i.e., all parameters listed in Table A1). This was decreased to 53 for nonlinear analyses.

Observations

Steady-state heads from wells in the model domain were used for model calibration (SNL 2007). Historical head data for these wells are of varying quality, with some suspected of being perched and thus not being truly representative of regional groundwater conditions. During calibration, relative weighting of head measurements was such that those believed to be of higher credibility received greater weights. Moreover, those wells indicating an upward gradient (flow from the deep carbonate aquifer to the overlying volcanic units) were more heavily weighted, as were those along the inferred flowpath from beneath the proposed repository. Figure 2 shows the calibrated water table and the corresponding residuals for each calibration well from SNL (2007).

Groundwater fluxes at YMPSZFM boundaries computed by the Death Valley Regional Flow System (DVRFS) model (Belcher 2004) also comprised part of the calibration dataset; see SNL (2007) for further details of the role that the DVRFS played in setting boundary conditions for the YMPSZFM. A total of seven such “measurements” were included in the calibration dataset, one pertaining to the northern model boundary, another pertaining to the western model boundary, and five pertaining to the eastern model boundary (indicated with differently styled lines in Figure 2). Flow through the southern boundary was not targeted because it is necessarily the sum of all other flows into (or out of) the system (SNL 2007). These flux observations were not weighted heavily because they provide only general guidelines for flow at this scale.

Model Calibration

Contributions that different groups of observations (Table 1) make to the overall reduction in predictive uncertainty accrued through the process of model calibration is discussed subsequently. As described by SNL (2007), parameter adjustment during calibration of the YMPSZFM was undertaken using the PEST (Doherty 2009b) package, supplemented by some manual adjustment where deemed necessary (SNL 2007).

The $C(p)$ and $C(\epsilon)$ Matrices

Equations 8, 12, and 19 cite a $C(p)$ matrix expressing the innate variability of parameters employed by the model. For parameters that represent permeability, the respective diagonal elements of this matrix express the range of permeabilities that, from hydrogeological considerations alone, these units are likely to possess before constraints on their variability are imposed by the necessity that model outputs fit the calibration dataset. Similar considerations apply to other parameter types. As stated previously, the $C(p)$ matrix is an expression of the state of precalibration hydrogeological knowledge and hydrogeological uncertainty.

To make use of Equations 12 and 19, all parameters were assumed to possess log-normal probability distributions (a typical assumption for geologic materials); the standard deviation of each distribution (in \log_{10} space) appears in the final column of Table A1. As no statistical correlation is assumed to exist between any pair of parameters, $C(p)$ is diagonal.

An implicit attempt was made to estimate relative values for the elements of the (assumed diagonal) $C(\epsilon)$ matrix prior to model calibration, by assigning weights to observations in inverse proportion to the suspected level of noise associated with each. On completion of the calibration process, $C(\epsilon)$ was computed from the (diagonal) weight matrix Q using Equation 3, with the reference variance σ_r^2 chosen such that the postcalibration objective function Φ was equal to 161, which is the number of observations comprising the calibration dataset.

The Prediction

The uncertainty and error variance of only one prediction is analyzed using particle tracking simulations; specifically the specific discharge (path length divided by travel time) is averaged over 100 particles released at random locations below the repository and tracked until they cross UTM northing 4,073,761 m (approximately 5 km south of the repository midpoint). For the calibrated model, this is 0.36 m/year (SNL 2007). End members of the 100-particle plume computed by the calibrated model have specific discharge values of 0.11 and 0.66 m/year. Although this analysis focuses on only one prediction of interest to provide a specific example for analysis, it could be applied equally well to any other model prediction.

Results

Calculation of Sensitivities

Implementation of both linear and nonlinear uncertainty and predictive error analysis requires that sensitivities of the model prediction, and of all model outputs for which there are corresponding field measurement, be computed with respect to all model parameters (i.e., a Jacobian matrix). This was achieved here using finite parameter (central) differences, with each parameter being varied by 1% of its current value when calculating these differences.

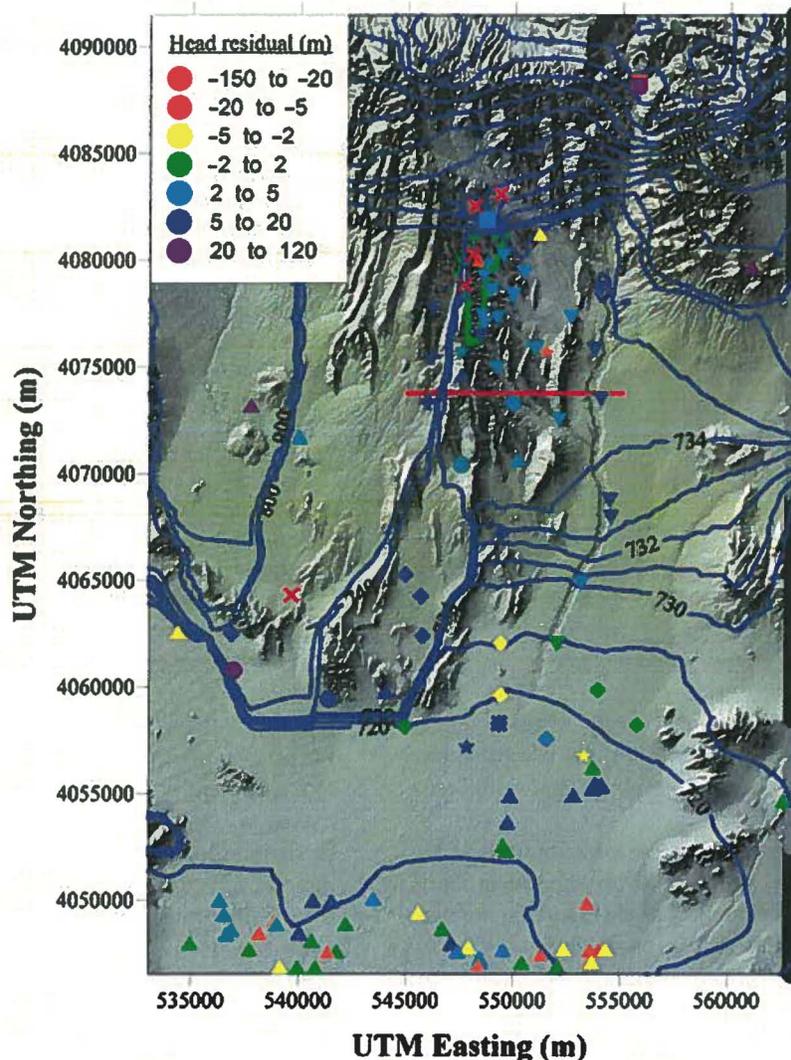


Figure 2. Head residuals and the calibrated potentiometric surface (SNL 2007). The *CraterFlat* group is indicated with circles, *HighHead* with squares, *ModGrad* with pluses, *NyeCounty* with diamonds, *Other* with upward triangles, *Path* with downward triangles, *Perched* with crosses (large negative residuals), and *UpGrad* with stars. The lateral extents of the five eastern boundary flux target segments are indicated with differently styled lines. A distance of 5 km south of the middle of the repository is illustrated with a red line.

Linear Analysis

Predictive Uncertainty

Use of Equation 19 to compute postcalibration uncertainty of specific discharge predictions reveals that:

- The precalibration standard deviation of predictive uncertainty is 1.46 m/year, equivalent to a variance of 2.12 (m/year)^2 .
- The postcalibration standard deviation of predictive uncertainty is 1.06 m/year, equivalent to a variance of 1.12 (m/year)^2 . These results are consistent with the uncertainty range of factors of 1/8.93 and 8.93 (selected from a log-transformed piecewise uniform distribution) used in other Yucca Mountain uncertainty analyses (SNL 2008, Section 6.5.2.1[a]).

It is apparent from the results that constraints on parameter values imposed by asking the model to replicate site data reduce the uncertainty of predicted specific

discharge (and, by inference, predictions of long-term contaminant movement that are potentially important to performance assessment). However, relatively speaking, the reduction in uncertainty is not great.

It is noteworthy that the computed uncertainties stated above (e.g., 1.06 m/year) are greater than the value of the prediction to which they pertain (i.e., 0.36 m/year). Greater linearity in the relationship between this prediction and model parameters might have been achieved if the log of specific discharge was designated as the prediction of interest rather than the specific discharge itself. That is, the predicted value of 0.36 m/year with a postcalibration standard deviation of 1.06 m/year might be better represented as $-0.44 \log(\text{m/year})$ with a standard deviation of $0.02 \log(\text{m/year})$, but the intuitive feel of the magnitude would be lost. Furthermore, because the focus of linear analysis of this type is less to quantify

Table 1
Names of Groups into Which Observations Were Subdivided

Observation Group Name	Number of Measurements	Description
<i>CraterFlat</i>	7 head data	Measurements taken in the Crater Flat geologic units
<i>HighHead</i>	4 head data	Heads in northern portion of model domain that are important in establishing an upward gradient in the carbonate aquifer
<i>ModGrad</i>	4 head data	Measurements taken in west of Yucca Mountain in a region of moderate hydraulic gradient
<i>NyeCounty</i>	25 head data	Wells in Nye County
<i>Other</i>	50 head data	Other head measurements
<i>Path</i>	35 head data	Wells along the inferred flow path from below the repository
<i>Perched</i>	15 head data	Water levels suspected of being perched (minimally weighted) often in the northern region
<i>UpGrad</i>	10 head data	Measurements showing an upward hydraulic gradient
<i>Flux_N</i>	1 flux datum	Flux through northern model boundary
<i>Flux_W</i>	1 flux datum	Flux through western model boundary
<i>Flux_E</i>	5 flux data	Flux through eastern model boundary (indicated in Figure 2)

predictive uncertainty/error than to compare these quantities (e.g., before and after calibration), the use of native rather than log values probably does not have a significant detrimental effect in this study.

Precalibration and postcalibration contributions of all parameters used in the YMPSZFM to the uncertainty variance of predicted specific discharge (Figure 3) are discussed next. The contribution to predictive uncertainty variance made by a particular parameter is calculated by repeating the predictive uncertainty computation while employing an assumption that the parameter whose contribution is sought is perfectly known [and hence that the corresponding element of the $C(\mathbf{p})$ matrix is zero]. For ease of description, the resulting decrease in predictive uncertainty variance is defined as that parameter's contribution to uncertainty for the prediction.

The following features of Figure 3 are of interest:

1. It is not impossible for the postcalibration contribution to predictive uncertainty by a particular parameter to be greater than its precalibration contribution. This is an outcome of the definition of "contribution" as the loss in predictive uncertainty variance accrued through acquiring perfect knowledge of a parameter's value through means outside of the calibration process. If a certain parameter is such that it can only be estimated in conjunction with another parameter (i.e., it is correlated), and if a prediction is sensitive to that other parameter, then acquisition of perfect knowledge of the first parameter can yield better estimates for the second parameter through the calibration process, and hence of any prediction that depends on that second parameter.
2. Uncertainties in vertical gradients attributed to fixed heads along model boundaries make only a small contribution to the uncertainty of predicted specific discharge. That is, the sensitivity of the prediction of

interest to these head gradients is not as great as it is for other model parameters.

3. Not surprisingly, uncertainty in the recharge multiplier makes a significant contribution to uncertainty in predicted specific discharge because increased recharge directly increases groundwater flux.
4. Postcalibration predictive uncertainty is dominated by contributions from permeability parameters, particularly *cfppa*, *ovu*, and *uccu*. The *cfppa* unit is important because it is present at the water table below the repository where most particles are introduced into the model. The *ovu* is regionally extensive in the vicinity of the repository. Although less clear as to its impact, the *uccu* serves as an aquitard between the lower carbonate aquifer and the overlying volcanic units and is salient to establishing the regional upward gradient. Fault permeabilities *scfz* and *stfz* are also large contributors to uncertainty because they are important in establishing the gradient in the vicinity of Yucca Mountain.

We assess in two ways the extent to which individual observations, or families of observations, contribute to the calibration process by reducing the uncertainties of key model predictions. The first method is to repeat the calculation of predictive uncertainty using Equation 19 with selected observations removed, thereby calculating the increase in predictive uncertainty variance incurred through their omission. The second method is to compute predictive uncertainty after removing all observations except those comprising the group of interest, and then computing the reduction from precalibration predictive uncertainty accrued through use of a calibration dataset comprising only these observations. Figure 4 depicts outcomes of these calculations for each of the observation groups listed in Table 1.

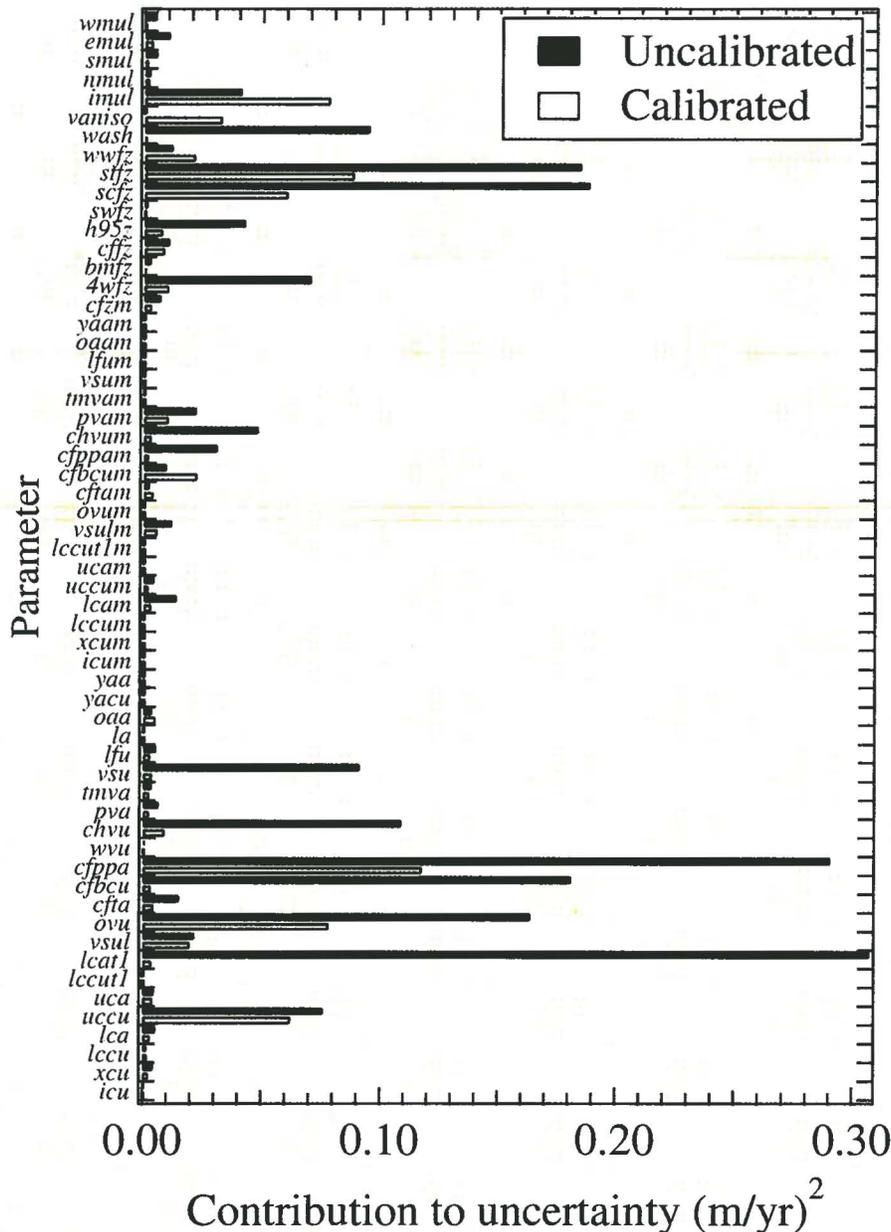


Figure 3. Precalibration (black) and postcalibration (white) contributions by model parameters to the uncertainty of predicted specific discharge. Total precalibration predictive uncertainty variance is 2.12 (m/year)^2 , whereas total postcalibration predictive uncertainty variance is 1.12 (m/year)^2 . Parameter abbreviations are defined in Table A1.

Figure 4a in particular indicates the importance of the *Flux_e* and *HighHead* observation types in enhancing the model's ability to predict specific discharge. This is not surprising, given that the high-head region in the north of the model domain drives flow through the system and establishes an upward gradient in the carbonate aquifer, and that the flux through the eastern boundary is relatively large (SNL 2007).

It is of interest to note that Equation 19 can be employed not only to compare the worth of existing observation types but also to evaluate the relative efficacy of future data-acquisition strategies. The actual values of measured data are not specified in Equation 19—only their noise covariance matrix must be estimated. This

means that there is no requirement for a measurement to have been made before its effectiveness in lowering the uncertainty of a key prediction is assessed. A similar approach was pursued by Hill et al. (2001) and Tiedeman et al. (2003, 2004) in assessing the relative adequacy of different data-acquisition strategies. However, Equation 19 provides a more rigorous basis for analysis of data worth in highly parameterized contexts.

Predictive Error Variance

Figure 5 shows the error variance of the specific discharge computed using Equation 12 plotted against the number of nontruncated singular values used in computation of the **R** matrix through Equation 13. Also

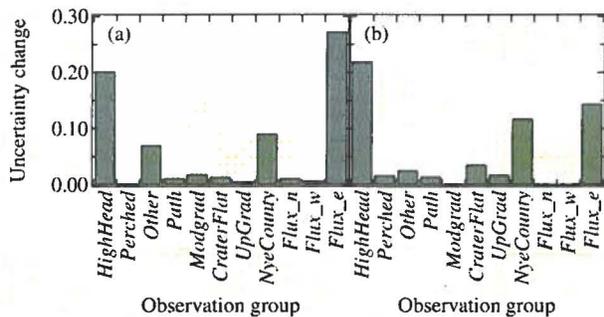


Figure 4. (a) Increase in predictive uncertainty variance accrued through omission of different observation types from the calibration process. (b) Reduction in predictive uncertainty variance from its precalibration level for different observation types as sole members of the calibration dataset.

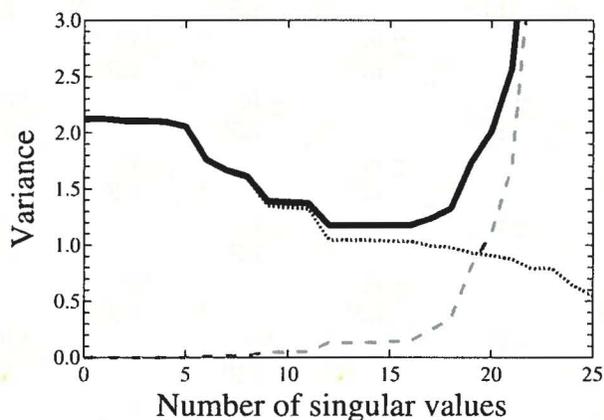


Figure 5. Error variance (solid) of predicted specific discharge for different numbers of pretruncation singular values. Also shown are the contributions to predictive error variance made by the null-space (short-dashed) and solution space (long-dashed) terms of Equation 12.

shown are the contributions to predictive error variance from the null-space term of this equation (the short-dashed curve falling with increasing number of singular values) and from the solution space term of this equation (the long-dashed curve rising with increasing number of singular values). A minimum error variance of 1.17 (m/year)^2 is achieved at 16 singular values; however, the error variance is nearly constant between 12 and 16 singular values. As expected, the value of 1.17 (m/year)^2 is slightly above the postcalibration predictive uncertainty variance of 1.12 (m/year)^2 computed using Equation 19. However, the proximity of these values illustrates that for the YMPSZFM, calibration using SVD with between 12 and 16 pretruncation singular values leads to predictions whose variances approach their theoretical lower limits based on the available calibration dataset.

The dominant contribution from the null-space term to predictive error variance should be noted. The null-space term encapsulates contributions to potential predictive error arising from the substantial amount of parameter value detail that is simply beyond the reach of the calibration process. To the extent that a prediction (e.g., specific

discharge) is sensitive to this missing detail, its potential for error is not reduced during calibration.

Parameter Identifiability

Parameter identifiability, computed using Equation 17 on the basis of truncation at 13 singular values (Figure 6), shows that only five parameters possess an identifiability that approaches 1, these being the boundary head multiplier parameters (*wmul*, *emul*, *smul*, and *nmul*) and the *lcat1* parameter. About 15 other parameters have midrange identifiability, whereas the remaining parameters have small or zero identifiability. To the extent that a prediction is sensitive to any of this last group of parameters, its postcalibration potential for error is not reduced at all by the calibration process based on the available dataset.

Parameter Uncertainty and Error Variances Reductions

Figures 7a and 9b depict relative parameter uncertainty variance reduction (calculated using Equation 19) and relative parameter error variance reduction (based on 12), respectively, for all parameters used in this study of the YMPSZFM. These graphs are self-similar and some correlation to the identifiability graph of Figure 6 is obvious; however, some differences are also apparent:

1. As expected, parameter uncertainty variance reduction is greater than parameter error variance reduction.
2. As also expected, the relative uncertainty/error variance reduction of each parameter is less than its identifiability, with the difference between the two rising with the extent to which a parameter is associated with singular values of higher index.

Figure 7 encapsulates the insight into the geology gained through this case study. Specifically, the precalibration uncertainty in a HGU's permeability (Table A1) is reduced by the square root of the factor shown on the bar chart. For many of the HGUs, there is little to no reduction in uncertainty indicating that the current dataset does not provide information on that unit's permeability (based on the current conceptual model as implemented in the FEHM numerical model). For many of the faults and the Crater Flat units below Yucca Mountain (*cfta*, *cfbcu*, *cfppa* in Figure 1), there is some notable reduction in the precalibration uncertainty range listed in Table A1.

Nonlinear Analysis

Constrained Maximization/Minimization

One method through which the uncertainty associated with a prediction made by a nonlinear model can be examined is to maximize and then minimize that prediction (altering model hydraulic properties in the process), subject to the joint constraints that:

1. Model outputs corresponding to members of the calibration dataset do not differ from their calibration values by an amount that is greater than a plausibility threshold that is established by the covariance matrix of measurement/structural noise, $C(\epsilon)$.

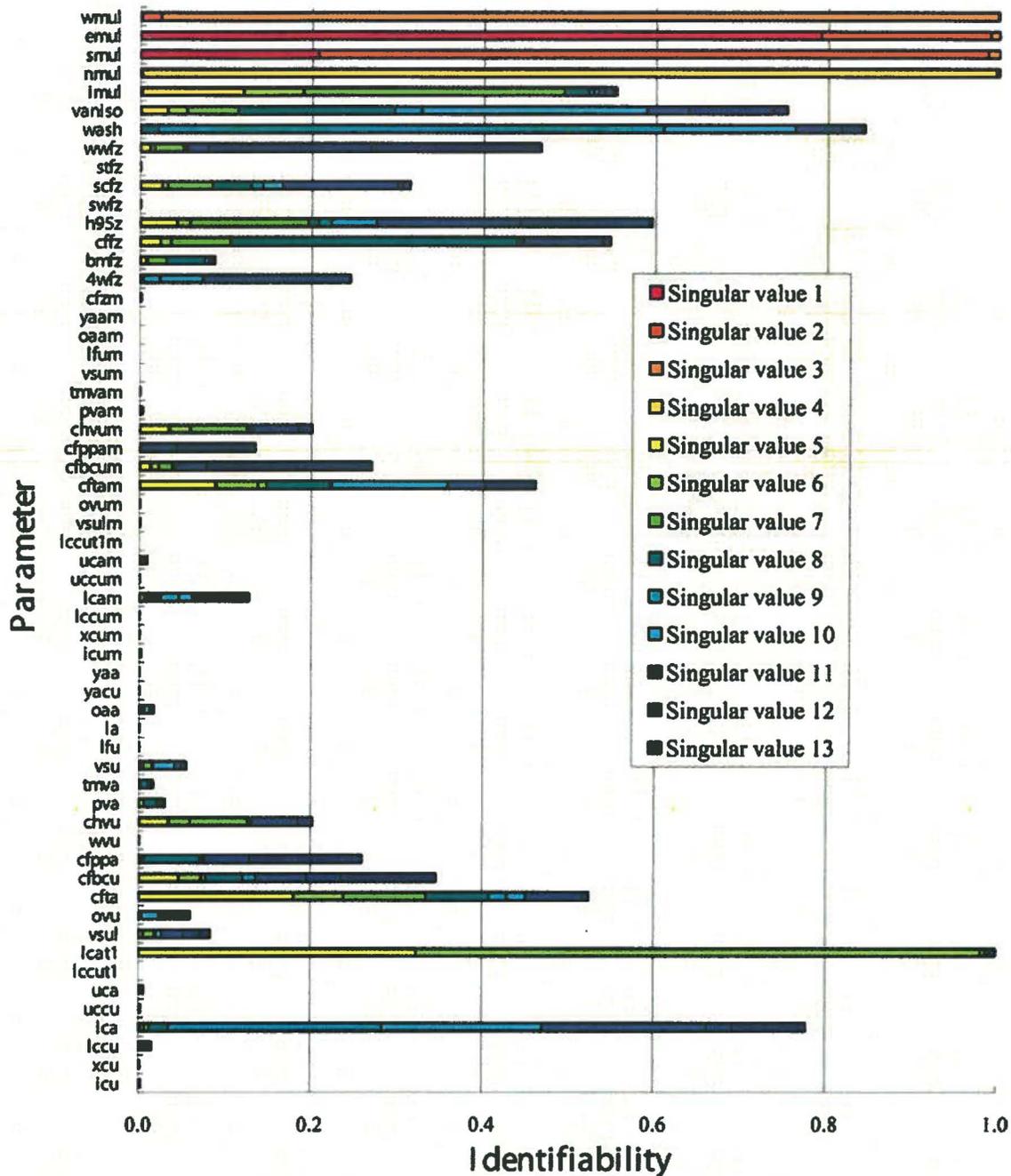


Figure 6. Parameter identifiability, calculated using Equation 17, for the YMPSZFM. Identifiability bars are stacked in this plot. The entire length of a particular bar represents that parameter's identifiability; segments within each bar represent contributions to identifiability by singular values up to 13 (where truncation occurs).

2. Null-space-projected differences between parameters corresponding to the maximized/minimized model prediction and those comprising the calibrated parameter set are compatible with the covariance matrix $C(p)$ of innate parameter variability.

These conditions must be met at a user-specified level of confidence, which for these analyses was set to 99.7%. This corresponds to 3 standard deviations on either side of the mean for a normal distribution, which both measurement/structural noise and model parameters are

assumed to follow. Thus, the prediction of interest (specific discharge) is maximized and minimized subject to the following constraint; the multicomponent objective function comprising squared differences between model outputs and their calibrated counterparts and the null-space-projected differences between parameter values computed as an outcome of prediction maximization/minimization and their calibrated counterparts (each of these standardized in terms of their respective covariance matrices) must rise not higher than 9 (the square of 3 standard deviations). See the work of Tonkin et al. (2007) for further details.

finite-differenced parameter derivatives during each maximization/minimization iteration requires a significant effort. The use of predictive super-parameters can ease this burden (Tonkin et al. 2007). However, using parameter subspaces for predictive maximization/minimization is less effective than when using them during calibration as they can compromise the ability of the maximization/minimization process to find predictive extremes.

Calibration-Constrained Monte Carlo

Calibration of the YMPSZFM was accomplished using PEST in conjunction with some hand-adjustment of parameter values (SNL 2007). The final value of the objective function (weighted sum-of-squares residual between modeled and measure heads) obtained through this process was 633.

When implementing calibration-constrained Monte Carlo analysis for the YMPSZFM, which is referred to herein as null space Monte Carlo (NSMC) to conform with terminology used in PEST documentation, a total of 200 parameter sets were first generated on the basis of the $C(p)$ matrix already discussed. Each set comprised 58 parameters, 52 of which were assigned different values in each realization. As for the constrained nonlinear maximization/minimization process discussed previously, *vaniso*, *imul*, *nmul*, *smul*, *emul*, and *wmul* were fixed at their nominal values during NSMC analysis; these variables were not considered important for this analysis because they were not part of the License Application.

For each randomly generated parameter set, a modified parameter set was then computed by (1) subtracting the calibration parameter set from the random parameter set; (2) projecting the resulting parameter difference set onto the calibration null space; and (3) adding the projected difference to the calibrated parameter set.

The calibration null space was defined through SVD of the $X'QX$ matrix of Equation 13 in the same manner as used in the linear analysis (with X calculated on the basis of the calibrated parameter set). Again, a solution space dimensionality of Equation 13 was used because it corresponds to the minimum of the predictive error variance curve depicted in Figure 5.

For most parameter sets computed in this fashion, the objective function was less than 1000, thus indicating a “nearly calibrated” model. To judge the efficacy of null-space projection in achieving near-calibration, a number of model runs were undertaken using prenull-space-projected random parameter sets; objective functions between 10^6 and 10^7 were computed for these parameter sets. As expected, null-space projection substantially (although not completely because of model nonlinearity) removes parameter combinations that decalibrate the model.

Next, each of the modified random parameter sets was used to parameterize the model and was then subjected to further modification as required to reduce the objective function to a level deemed to recalibrate the model (less

than 650). This process was made as efficient as possible in the following ways:

1. Only solution space parameter projections were adjusted to lower the objective function; thus, each calibration exercise required adjustment of only 13 super-parameters.
2. The first iteration of each calibration process was based on super-parameter sensitivities computed using the calibration parameter set; thus, the first iteration of each of the 200 calibration exercises required only a single model run—this being undertaken to compute the objective function for the modified parameter values.
3. The calibration process was immediately terminated if the objective function fell below 650, which is only slightly above the objective function of 633 achieved through calibration.

Of the 200 calibration exercises conducted in this manner (each based on a different null-space-projected random parameter set), all reached an objective function of less than 650 by the end of the second iteration. Nine results were discarded as they exhibited flow paths from the repository to the eastern model boundary; this same constraint was applied during the manual component calibration of the YMPSZFM model and is consistent with site hydrochemical data and analyses (SNL 2007). Thus, 191 calibration-constrained parameter sets were obtained. The average cost in model runs was a mere 4.3 runs per calibrated parameter set.

The frequency distribution for the *wwfz* parameter (the permeability of the Windy Wash Fault Zone) arising from the 191 accepted model calibrations (Figure 8) shows that this parameter is moderately identifiable (as in Figure 6); hence, it is partly constrained by the calibration process, and partly “unconstrained.” It can vary as long as other parameters vary in harmony with it and thereby maintain model calibration. The precalibration standard deviation of the log of this parameter was assumed to be 0.67 (Table A1); its calibrated value is $4.8 \times 10^{-16} \text{ m}^2$, the log of which is -15.32 . Although the precalibration

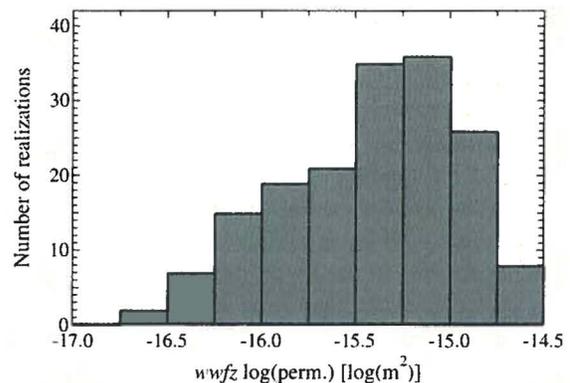


Figure 8. Frequency distribution of calibration-constrained values of the permeability of the Windy Wash Fault Zone, parameter *wwfz*.

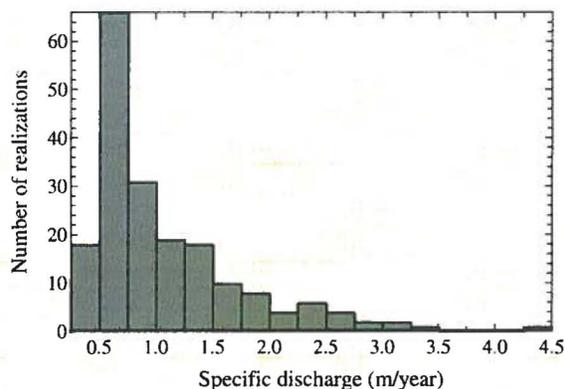


Figure 9. Frequency distribution for specific discharge.

standard deviation allows a broad permeability distribution, respect for calibration constraints forces a skewed distribution that does not allow large permeabilities. This is consistent with current understanding of the site where the Windy Wash Fault Zone acts as a barrier to flow.

Predictive model runs were made using the 191 calibration-constrained parameter sets. The frequency distribution of the specific discharge (Figure 9) shows that the long tail allows some large values of specific discharge. A pleasing result is that the length of this tail is of the same order as 3 standard deviations of predictive uncertainty as calculated through linear analysis.

Exact computation of predictive probabilities at the extremes of a probability distribution based on Monte Carlo analysis requires a large number of model runs. Nevertheless, enough calibration-constrained parameter fields have been generated to provide a strong indication of the range of predictive possibilities that are compatible with both hydrogeological knowledge and with observations of system state. Because of the model run efficiency achieved when generating calibration-constrained parameter sets, many more such sets could be produced relatively easily to further refine this analysis if required.

Discussion

This case study chiefly serves to demonstrate and illustrate the spectrum of powerful analyses that can be readily undertaken as an adjunct to highly parameterized model calibration. Several findings of potential relevance to the YMPSZFM are also noteworthy.

Uncertainty in predictions of specific discharge, which has historically been an important variable in total system performance assessment (TSPA), has minimal dependence on uncertainties in assignment of vertical gradients applied to the model boundaries. This is to be expected because the boundaries are far enough from the flow paths not to influence the specific discharge predictions near Yucca Mountain. Specific discharge shows moderate sensitivity to a model-wide infiltration multiplier and inherits uncertainty as a result of potential error assigned to recharge across the YMPSZFM. Although not unexpected, it is worth noting that sensitivity of model

predictions to recharge has in fact been taken into account in the TSPA through incorporation of the effects of an uncertain parameter, namely the groundwater-specific discharge multiplier.

Constrained maximization/minimization of specific discharge demonstrates that this can range from 0.08 to 1.34 m/year, which is well within the factors of 1/8.93 and 8.93 established by SNL (2008). However, NSMC analysis yields a range of 0.10 to 4.12 m/year. Although the highest of these values is in excess of the factor of 8.93 (nominal value of 0.36 m/year \times 8.93 = 3.21 m/year) established by SNL (2008), only the single 4.12-m/year value exceeds this.

Finally, although only the *wwfz* (Windy Wash Fault Zone) parameter was discussed in this article, other model parameters also demonstrated NSMC parameter ranges consistent with current understanding of the site. Through the calibration process, parameter variance ascribed to our lack of knowledge about a particular parameter is tempered with information gained through constraining parameters in accordance with historical observations of system state. Through this process, it was found that units or features thought to be barriers to flow show distributions skewed toward lower permeabilities. These sorts of results provide insight into the broader hydrogeology of the region by reinforcing hypotheses that some faults such as the Solitario Canyon Fault (*scfz*) and the Highway 95 Fault (*h95z*) act as barriers to flow.

Conclusions

Using the calibrated YMPSZFM, this article has demonstrated that the use of a model in performance assessment or environmental management can extend well beyond that of simply calibrating the model and then making a prediction. In particular, relatively simple analyses undertaken as adjuncts to model construction and calibration can provide the following insights:

1. Identification of observations that are most effective in achieving this reduction in uncertainty (e.g., *High-Head*, *Flux_e*, and *NyeCounty* are critical to reducing uncertainty in specific discharge, see Figure 4).
2. Assessment of the reduction in predictive uncertainty that is accrued through model calibration (see Figure 5).
3. Assessment of the identifiability of each model parameter (see Figure 6).
4. Identification of the parameters (including features such as poorly known boundary conditions) that are most instrumental in maintaining predictive uncertainty at its current level (e.g., *lcat1* and several of the fault parameters are important to reducing uncertainty in specific discharge, see Figure 7).
5. Objective assessment of the uncertainty associated with predictions of interest (see Figure 9).
6. Identification and ranking of observations that, if gathered, would afford the greatest reductions in the uncertainties of key model predictions.

A key feature of the analyses described herein is that they should be undertaken as adjuncts to parameterization of a model that reflects all aspects of system complexity that are relevant to predictions of interest; that is, the model must include all features and processes to which requested predictions are sensitive. This will normally result in the use of many more parameters than are uniquely estimable on the basis of most calibration datasets, resulting in formulation of an ill-posed inverse problem that must be solved during model calibration. The analyses documented herein do not require that parameter lumping and fixing strategies be implemented to ensure well-posedness of the inverse problem before undertaking calibration. Rather, these analyses readily accommodate a high-dimensional calibration null space and are able to quantify the limitations imposed on the model's ability to make accurate predictions by the existence of this null space. Calibration and uncertainty analyses that do not accommodate or acknowledge the null space, and which require preanalysis parameter lumping and fixing to artificially formulate a well-posed inverse problem, provide far fewer insights.

The methodologies demonstrated in this article are just as applicable to models that include hundreds or even thousands of parameters as they are to the YMP-SZFM that employs only tens of parameters. Heavily parameterized models may be crucial to risk assessment when significant geological heterogeneity exists, or where either continuous or discrete geological entities could rapidly convey contaminants to locations that threaten environmental health. However, with the exception of just one of the methodologies discussed herein, the large numbers of parameters involved in such analyses would require an increase in the numerical burden required to conduct them. That exception is the null space Monte Carlo methodology. Its relative immunity to the increased numerical burden of accommodating a far greater level of complexity in model parameterization is an outcome of its following features:

1. As Tonkin and Doherty (2009) demonstrate, the methodology can be used in conjunction with complex stochastic parameter fields (which may involve heterogeneity at the cell or element level), despite the fact that only a limited number of adjustable parameters are employed for recalibration of each field.
2. Because the dimensionality of the calibration solution space is a function of available data rather than the level of heterogeneity represented in a model, the numerical burden of recalibration based on solution space parameter eigencomponents does not increase as parameterization complexity increases.
3. Computation of the sensitivity of a prediction to all model parameters is not required.
4. Once a set of calibration-constrained stochastic fields has been obtained, these can be used to explore the uncertainty of any number of predictions made by the model.

Although useful quantities such as parameter contributions to current predictive uncertainty are not as easily obtained through nonlinear analysis as they are through linear analysis, Monte-Carlo-based methods have the advantage that parameter fields giving rise to predictions of rapid groundwater flow and contaminant transport become readily apparent, so that the link between low probability predictions and the features that may give rise to them is obvious.

Regardless of which of the methods documented herein augments understanding of predictions made by an environmental model, this article demonstrates that by (1) including all parameter complexity in a model that is relevant to predictions required of that model, and (2) undertaking analyses that expose the degree of uncertainty or error associated with model parameters and predictions, much more can be gained from a modeling exercise than simply making a few predictions. In fact, a comprehensive assessment can be made of the strengths and weakness of a current dataset in assessing environmental risk.

Overall, the thorough and detailed predictive uncertainty analyses for the Yucca Mountain site are strongly supportive of the independent assessment of uncertainty used in the repository performance assessment analyses for the YMP License Application where a Bayesian analysis yielded a range for the specific discharge multiplier of 1/8.93 to 8.93. In our analysis, only one NSMC realization of the 191 parameter fields exceeded this span indicating that the range used in the License Application is appropriate more than 99% of the time. Moreover, $8.93 \times 0.36 \text{ m/year}$ is 3.21 m/year , which is $0.36 + 2.64 \times 1.08(\bar{x} + n\sigma_x)$ or $n = 2.64$ standard deviations. Hence, 99.6% of the linearized specific discharge distribution is within the maximum specific discharge multiplier of 8.93 used in the License Application. Moreover, many of the faults included in this model were at least somewhat identifiable, and those hypothesized to be barriers to flow wfz , $scfz$, and $h95z$ were confirmed to behave as such through calibration-constrained null-space Monte Carlo runs.

Acknowledgments

The authors would like to acknowledge contributions made by staff from the USGS, the Los Alamos National Laboratory, and Sandia National Laboratories in characterizing the saturated zone at Yucca Mountain. This work was supported by the U.S. Department of Energy. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. The content of this presentation reflects the views of the authors and does not necessarily reflect the views or policies of the U.S. Department of Energy, Los Alamos National Laboratory, or Sandia National Laboratories. The authors also thank the Editor-in-Chief, Mary Anderson, and acknowledge three reviewers, Claire Tiedeman, Chris Muffels, and an anonymous reviewer.

Supporting Information

Supporting Information may be found in the online version of this article:

Parameter and observation contributions to predictive error variance are described in the supporting information and are shown in Figures S1 and S2. These are the frequentist counterparts to the Bayesian predictive uncertainty variance in Figures 3 and 4.

Figure S1. Precalibration (black) and postcalibration (white) parameter contributions to specific discharge error variance where calibration is notionally implemented using truncated SVD. Total precalibration predictive error variance is 2.12 (m/year)^2 , whereas total postcalibration predictive error variance is 1.17 (m/year)^2 .

Figure S2. (a) Increase in predictive error variance accrued through omission of different observation types from the calibration process. (b) Reduction in predictive error variance from its precalibration level for different observation types as sole members of the calibration dataset.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

- Annis, C. 2008. Frequentists and Bayesians. http://mh1823.com/frequentists_and_bayesians.htm (accessed July 15, 2009).
- Aster, R. C., B. Borchers, and C. H. Thurber. 2005. *Parameter Estimation and Inverse Problems*. Amsterdam: Elsevier Academic Press.
- Belcher, W. R. 2004. Death Valley Regional Ground-Water Flow System, Nevada and California—Hydrogeologic Framework and Transient Ground-Water Flow Model. Reston, Virginia: USGS.
- Bower, K. M., and G. A. Zyvoloski. 1997. A numerical model for thermo hydro mechanical coupling in fractured rock. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts* 34, no. 8: 1201–1211.
- Christensen, S., and J. D. Doherty. 2008. Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration. *Advances in Water Resources* 31, no. 4: 674–700.
- Doherty, J. E. 2009a. *Addendum to the PEST Manual*. Brisbane, Australia: Watermark Numerical Computing.
- Doherty, J. E. 2009b. *Manual for PEST: Model Independent Parameter Estimation*. Brisbane, Australia: Watermark Numerical Computing.
- Doherty, J. E., and R. J. Hunt. 2009. Two easily-calculated statistics for evaluating parameter identifiability and error reduction. *Journal of Hydrology* 366, 119–127.
- Gallagher, M., and J. E. Doherty. 2006. Predictive error analysis for a water resource management model. *Journal of Hydrology* 34, no. 3–4: 513–533.
- Gómez-Hernández, J. J., A. Sahuquillo, and J. E. Capilla. 1997. Stochastic simulation of hydraulic conductivity fields conditional to both hydraulic conductivity and piezometric data. 1. Theory. *Journal of Hydrology* 20, 162–174.
- Hill, M. C., D. M. Ely, C. R. Tiedeman, F. A. D’Agnese, C. C. Faunt, and G. M. O’Brien. 2001. Preliminary evaluation of the importance of existing hydraulic-head observation locations to advective-transport predictions. Death Valley regional flow system, California and Nevada. Reston, Virginia: USGS.
- Hunt, R. J., J. E. Doherty, and M. J. Tonkin. 2007. Are models too simple? Arguments for increased parameterization. *Ground Water* 45, no. 3: 254–262.
- Kitanidis, P. K. 1996. The geostatistical approach to the inverse problem. *Advances in Water Resources* 19, no. 6: 333–342.
- Menke, W. 1989. *Geophysical Data Analysis: Discrete Inverse Theory*. San Diego, California: Academic Press.
- Moore, C., and J. E. Doherty. 2006. The cost of uniqueness in groundwater model calibration. *Advances in Water Resources* 29, no. 4: 605–623.
- Moore, C., and J. E. Doherty. 2005. Role of the calibration process in reducing model predictive error. *Water Resources Research* 41, no. 5: W05050.
- SNL. 2008. *Saturated Zone Flow and Transport Model Abstraction*. Las Vegas, Nevada: Sandia National Laboratories.
- SNL. 2007. *Saturated Zone Site-Scale Flow Model*. Las Vegas, Nevada: Sandia National Laboratories.
- Tiedeman, C. R., D. M. Ely, M. C. Hill, and G. M. O’Brien. 2004. A method for evaluating the importance of system state observations to model predictions, with application to the Death Valley regional groundwater flow system. *Water Resources Research* 40: W12411.
- Tiedeman, C. R., M. C. Hill, F. A. D’Agnese, and C. C. Faunt. 2003. Methods for using groundwater model predictions to guide hydrogeologic data collection, with application to the Death Valley regional groundwater flow system. *Water Resources Research* 39, no. 1: 5-1–5-17.
- Tonkin, M. J., and J. E. Doherty. 2009. Calibration-constrained Monte Carlo analysis of highly-parameterized models using subspace techniques. *Water Resources Research* 45: W00B10.
- Tonkin, M. J., Doherty, J. E., and Moore, C. 2007. Efficient nonlinear predictive error variance for highly parameterized models. *Water Resources Research* 43: W07429.
- Tonkin, M. J., and J. E. Doherty. 2005. A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resources Research* 41: W10412.
- Vecchia, A. V., and R. L. Cooley. 1987. Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model. *Water Resources Research* 23, no. 7: 1237–1250.
- Woodbury, A. D., and T. J. Ulrich. 2000. A full-Bayesian approach to the groundwater inverse problem for steady state flow. *Water Resources Research* 36, no. 8: 2081–2094.
- Zyvoloski, G. A. 1983. Finite element methods for geothermal reservoir simulation. *International Journal for Numerical and Analytical Methods in Geomechanics* 7, no. 1: 75–86.

Appendix

Model parameters are listed and briefly described in Table A1.

Table A1

Calibrated Permeabilities (third column) are in m^2 While Multipliers are Unitless. The Final Column Shows Precalibration Uncertainty for the Log of Each Parameter Representing the Authors' Best Estimate of a Parameter's Standard Deviation (in Log Space; These are the Diagonal Elements of $C(p)$).

Parameter	Description	Calibrated Value	σ_{\log}
<i>icu</i>	Intrusive Confining Unit (granite)	9.9×10^{-17}	0.78
<i>xcu</i>	Crystalline Confining Unit (granite)	1.0×10^{-16}	0.79
<i>lccu</i>	Lower Clastic Confining Unit	9.7×10^{-17}	0.78
<i>lca</i>	Lower Carbonate Aquifer	9.7×10^{-15}	0.3
<i>uccu</i>	Upper Clastic Confining Unit	9.8×10^{-16}	1.5
<i>uca</i>	Upper Carbonate Aquifer	1.1×10^{-12}	1.5
<i>lccutl</i>	Lower Clastic Confining Unit Thrust	9.8×10^{-19}	1.5
<i>lcatl</i>	Lower Carbonate Aquifer Thrust	5.6×10^{-12}	1.5
<i>vsul</i>	Volcanic and Sedimentary Units (lower)	1.1×10^{-14}	0.58
<i>ovu</i>	Older Volcanic Unit	9.8×10^{-16}	1.5
<i>cfta</i>	Crater Flat-Tram Aquifer	9.4×10^{-13}	0.62
<i>cfbcu</i>	Crater Flat-Bullfrog Confining Unit	5.2×10^{-14}	0.98
<i>cfppa</i>	Crater Flat-Prow Pass Aquifer	3.1×10^{-12}	0.55
<i>wvu</i>	Wahmonie Volcanic Unit	9.8×10^{-14}	1.5
<i>chvu</i>	Calico Hills Volcanic Unit	2.4×10^{-13}	0.8
<i>pva</i>	Paintbrush Volcanic Aquifer	6.5×10^{-14}	0.33
<i>tmva</i>	Timber Mountain Volcanic Aquifer	9.5×10^{-14}	0.33
<i>vsu</i>	Volcanic and Sedimentary Units (upper)	8.7×10^{-13}	1.5
<i>lfu</i>	Lava Flow Unit	8.9×10^{-14}	1.5
<i>la</i>	Limestone Aquifer	9.8×10^{-14}	1.5
<i>oaa</i>	Older Alluvial Aquifer	1.5×10^{-13}	0.58
<i>yacu</i>	Young Alluvial Confining Unit	9.9×10^{-15}	0.58
<i>yoa</i>	Young Alluvial Aquifer	9.8×10^{-13}	0.58
<i>icum</i>	<i>icu</i> Northern zone permeability multiplier	0.3	0.17
<i>xcum</i>	<i>xcu</i> Northern zone permeability multiplier	0.2	0.17
<i>lccum</i>	<i>lccu</i> Northern zone permeability multiplier	0.2	0.5
<i>lcam</i>	<i>lca</i> Northern zone permeability multiplier	0.2	0.5
<i>uccum</i>	<i>uccu</i> Northern zone permeability multiplier	9.7×10^{-3}	0.5
<i>ucam</i>	<i>uca</i> Northern zone permeability multiplier	2.0×10^{-2}	0.5
<i>lccutlm</i>	<i>lccutlm</i> Northern zone permeability multiplier	9.8×10^{-3}	0.5
<i>vsulm</i>	<i>vsul</i> Northern zone permeability multiplier	1.0×10^{-2}	0.5
<i>ovum</i>	<i>ovu</i> Northern zone permeability multiplier	9.9×10^{-3}	0.5
<i>cftam</i>	<i>cfta</i> Northern zone permeability multiplier	1.0×10^{-2}	0.5
<i>cfbcum</i>	<i>cfbcu</i> Northern zone permeability multiplier	9.1×10^{-3}	0.5
<i>cfppam</i>	<i>cfppa</i> Northern zone permeability multiplier	1.4×10^{-3}	0.5
<i>chvum</i>	<i>chvu</i> Northern zone permeability multiplier	2.3×10^{-3}	0.5
<i>pvam</i>	<i>pva</i> Northern zone permeability multiplier	9.6×10^{-3}	0.5
<i>tmvam</i>	<i>tmva</i> Northern zone permeability multiplier	9.8×10^{-3}	0.5
<i>vsuam</i>	<i>vsu</i> Northern zone permeability multiplier	1.0×10^{-2}	0.5
<i>lfum</i>	<i>lfu</i> Northern zone permeability multiplier	1.0×10^{-2}	0.5
<i>oaam</i>	<i>oaa</i> Northern zone permeability multiplier	1.0×10^{-2}	0.5
<i>yaam</i>	<i>yaa</i> Northern zone permeability multiplier	1.0×10^{-2}	0.5
<i>ymzm</i>	Yucca Mountain zone permeability multiplier	8.9	0.5
<i>4wfz</i>	Fortymile Wash Fault Zone	1.4×10^{-10}	0.33
<i>bmfz</i>	Bare Mountain Zone	9.9×10^{-16}	0.67
<i>cffz</i>	Crater Flat Fault Zone	9.7×10^{-17}	0.67
<i>h95z</i>	Highway 95 Fault Zone	1.0×10^{-14}	0.67
<i>swfz</i>	Sever Wash Fault Zone	9.8×10^{-18}	0.67
<i>scfz</i>	Solitario Canyon Fault Zone	5.0×10^{-16}	0.67
<i>stfz</i>	Stage Coach Fault Zone	4.7×10^{-16}	0.67
<i>wwfz</i>	Windy Wash Fault Zone	4.8×10^{-16}	0.67
<i>wash</i>	Lower Fortymile Wash	2.0×10^{-11}	0.67
<i>vaniso</i>	Vertical anisotropy (unitless)	0.1	0.5
<i>nmul</i>	Vertical head gradient multiplier on the northern boundary	1	0.02
<i>smul</i>	Vertical head gradient multiplier on the southern boundary	1	0.02
<i>wmul</i>	Vertical head gradient multiplier on the western boundary	1	0.02
<i>emul</i>	Vertical head gradient multiplier on the eastern boundary	1	0.02
<i>imul</i>	Infiltration multiplier	1	0.5

