Role of the calibration process in reducing model predictive error

Catherine Moore and John Doherty

School of Engineering, University of Queensland, Brisbane, Queensland, Australia

Received 16 July 2004; revised 13 February 2005; accepted 28 February 2005; published 24 May 2005.

[1] An equation is derived through which the variance of predictive error of a calibrated model can be calculated. This equation has two terms. The first term represents the contribution to predictive error variance that results from an inability of the calibration process to capture all of the parameterization detail necessary for the making of an accurate prediction. If a model is ''uncalibrated,'' with parameter values being supplied solely through ''outside information,'' this is the only term required. The second term represents the contribution to predictive error variance arising from measurement noise. In an overdetermined system, such as that which may be obtained through ''parameter lumping'' (e.g., through the introduction of a spatial zonation scheme), this is the only term required. It is shown, however, that parameter lumping is a form of ''implicit regularization'' and that ignoring the implied first term of the predictive error variance equation can potentially lead to underestimation of predictive error variance. A model's role as a predictor of environmental behavior can be enhanced if it is calibrated in such a way as to reduce the variance of those predictions which it is required to make. It is shown that in some circumstances this can be accomplished through ''overfitting'' against historical field data. It can also be accomplished by giving greater weight to those measurements which carry the greatest information content with respect to a required prediction. This suggests that a departure may be necessary from the custom of using a single ''calibrated model'' for the making of many different predictions. Instead, model calibration may need to be repeated many times so that in each case the calibration process is optimized for the making of a specific model prediction.

Citation: Moore, C., and J. Doherty (2005), Role of the calibration process in reducing model predictive error, *Water Resour. Res.*, *41*, W05020, doi:10.1029/2004WR003501.

1. Introduction

[2] Models are often used to make predictions of environmental behavior. Used in that role, they support environmental management. A vital aspect of the model construction process is the calibration phase. During this phase, model parameters are adjusted until the model's replication of historical field measurements is judged to be ''reasonably good.'' It is then assumed that this constitutes sufficient justification to use the model to make predictions and that those predictions will also be reasonably good.

[3] Unfortunately, even predictions made by a model that matches historical data perfectly may be considerably in error. In fact, a model's predictive uncertainty will only be reduced by calibration if the information content of the calibration data set is able to constrain those parameters that have a significant bearing on that prediction. Thus, for example, if a model is built to make predictions of contaminant transport, then its performance in this regard will be better served if it is calibrated against historical contaminant concentration data than if it is calibrated solely against groundwater heads [see, e.g., *Harvey and Gorelick*, 1995; *Poeter and Belcher*, 1991; *Frederick and Doherty*, 2003; *Franssen et al.*, 2003; *Feyen et al.*, 2003]. For the same reason, *Tiedeman et al.* [2003] assert that model predictive uncertainties may not necessarily decrease with calibration to a larger data set.

[4] Model predictive uncertainty arises from a number of sources. In the present paper inadequacies in model equations, or in the numerical implementation of those equations, will be neglected, and only those pertaining to its parameterization will be considered.

[5] Information on model parameters comes from two sources. The first of these consists of inferences of system properties based on knowledge of the materials of which the system under study is composed, often supplemented by direct measurements of system properties at a limited number of locations. The latter are often accompanied by noise so that parameter inferences drawn from them are subject to a possibly large margin of uncertainty. Because such ''prior information'' on parameter values is often vague, point-based, and uncertain, it is often best expressed in stochastic terms; for models such as groundwater models where parameterization is spatially based, geostatistical characterizations of hydraulic properties are thus often employed. The second source of knowledge on model parameters arises from historical measurements of system state. This is a more indirect form of knowledge of system properties, the information content of which is ''tapped'' during the model calibration process. Unfortunately, this information is often contaminated by noise; thus parameter values inferred from such data have a stochastic component

0043-1397/05/2004WR003501

that originates in the uncertainty associated with field measurements as well as in so-called "structural noise," a term used to describe various unavoidable forms of model inadequacy. Another problem with inferring parameter values from field measurements is that such measurements, like direct measurements of system properties, are often sporadic in both space and time. Hence there is an upper limit to the level of parameterization detail that can be inferred from them; see *Backus and Gilbert* [1969], *Menke* [1984], *Kitanidis* [1997], *Guadagnini and Neuman* [1999], and *Gorokhovski* [1996] to mention just a few discussions of this important subject.

[6] Bayes theorem provides the means to assimilate these two sources of information on model parameters into a "posterior parameter distribution" that reflects both the constraining effects of the calibration process and knowledge of parameter values that originates from outside of this process. Model predictive probabilities can then be evaluated using the relationships between model parameters and model predictive outputs encapsulated in the model. If necessary, such an analysis can include geostatistical characterizations of prior parameter uncertainty [see, e.g., *Woodbury and Ulrich*, 2000; *Woodbury and Rubin*, 2000, and references therein]. A non-Bayesian, but nevertheless effective, means of constraining geostatistical characterizations of parameter spatial variability such that historical measurements of system state are respected by the model involves the "bending" or "warping" of stochastic seed fields, thus forcing these fields to satisfy calibration constraints. For examples of this methodology, see *RamaRao et al.* [1995], *LaVenue et al.* [1995], *Gómez-Hernández et al.* [2003], and *Doherty* [2003].

[7] An alternative methodology for analyzing model predictive uncertainty is presented in this paper. Use of this methodology is based on the premise that a model has been "calibrated" against a set of field measurements as a precursor to its deployment for making predictions of future system behavior, this being the most common strategy for using models in environmental management. It is further assumed that calibration takes place as an underdetermined inverse problem. This strategy allows a model to employ a level of complexity that is sufficient to represent all processes on which a prediction of interest depends. While parameters pertaining to that complexity may not be uniquely estimated, it is demonstrated herein that full characterization of predictive error variance requires that this complexity be represented. The theory is then extended to overdetermined parameter estimation (applicable, for example, where a model domain is subdivided into a small number of zones of piecewise parameter constancy in accordance with the principle of parsimony). It is demonstrated that estimates of model predictive error variance made as an adjunct to model calibration based on this principle can be seriously flawed unless the effects of such system simplification are taken into account in making these estimates.

[8] Solution of an underdetermined inverse problem is possible only if some regularization strategy is employed. The use of regularized inversion in the context of groundwater model calibration has been discussed by a number of authors, including *Vasco et al.* [1997], *Clemo et al.* [2003], and *Doherty* [2003]. Software such as PEST [*Doherty*, 2004] is freely available for its implementation. Through the use of regularized inversion, simplifications in parameterization necessary for the achievement of numerical stability of the inverse problem are undertaken by the parameter estimation process itself rather than through manual simplification as a precursor to that process. This allows maximum information content to be extracted from a given calibration data set.

[9] The discussion begins in section 2.1 by exploring model predictive uncertainty in contexts where parameterization is unassisted by calibration. After a brief discussion of regularized inversion, exploration of predictive uncertainty is extended to accommodate the imposition of calibration constraints on parameter values. Through an analysis of the equations so derived, some important points regarding the role of model calibration in reducing (or not reducing) model predictive uncertainty are discussed. With these points in mind the discussion then turns to how the model calibration process can be made to better serve the model predictive process, particularly with regard to the assignment of "measurement weights" to elements of the calibration data set. Finally, some of the concepts developed in the analysis are applied to a synthetic case to demonstrate their use.

[10] It must be pointed out that the equations derived in section 2.3 are based on an assumption of model linearity. Most models, of course, are nonlinear; hence these equations will be only approximations in many cases. Nevertheless, they are useful for the contribution that they make to our understanding of the calibration process. Furthermore, it is hoped that their use can extend farther than this to a semiquantitative analysis of calibration outcomes. Where such an analysis seeks ordering relationships rather than absolutes (e.g., in determining the relative contribution to uncertainty made by different parameter groups or the relative reduction in uncertainty that can be accrued through acquisition of different types of supplementary data), conclusions drawn through application of these equations are likely to be quite robust, notwithstanding the nonlinear nature of a model to which they may be applied.

2. Theory

2.1. Linear Predictive Uncertainty Analysis for an Uncalibrated Model

[11] Suppose that the $(m \times 1)$ vector $\mathbf{p}$ contains the values of parameters used by a model. Unless these values are accurately known at all places within a model domain, they must be described in probabilistic terms. Let the covariance matrix of $\mathbf{p}$ be denoted as $C(\mathbf{p})$. Let s (a scalar) designate a prediction made by the model; let the sensitivities of this prediction to model parameters be represented by the vector $\mathbf{y}$. Then s is calculable using the relationship

$$s = \mathbf{y}^t\mathbf{p}, \tag{1}$$

where the "t" superscript denotes the transpose operation. (Note that in this, and subsequent equations, parameter and prediction offsets are ignored for the sake of simplicity so that a $\mathbf{p}$ of $\mathbf{0}$ results in a zero-valued prediction. Thus $\mathbf{p}$ can be considered to represent parameter perturbations from some known average value, while s can be considered to

represent the resulting perturbation of the model prediction; this makes no difference to the equations and concepts derived in this section and in sections 2.2–2.6.) Through basic matrix manipulation it is easily shown that the variance of s (i.e., σ_s^2) is given by

$$\sigma_s^2 = \mathbf{y}^t C(\mathbf{p})\mathbf{y}. \quad (2)$$

Equation (2) was used by *El Harrouni et al.* [1997] in calculating output uncertainties for their dual reciprocity boundary element method groundwater model on the basis of spatially correlated model parameters.

2.2. Regularized Inversion

[12] "Inversion" is the process by which model parameter values are inferred from measurements of system state by matching model outputs to these measurements. In the discussion which follows these inferred parameter values are designated as $\hat{\mathbf{p}}$ to distinguish them from "true" model parameters $\mathbf{p}$. Much of the purpose of this study is to quantify the difference between these two. (Note that the difference between "model parameters" and "system properties" that arises from the fact that even the most complex model is a simplification of reality is ignored in this study. Hence, for the present purpose the elements of $\mathbf{p}$ can also be thought of as representing true system properties, of which the elements of $\hat{\mathbf{p}}$ are their inferred equivalents.)

[13] Let the (assumed linear) relationship between the m model parameters $\hat{\mathbf{p}}$ and n model outputs $\mathbf{o}$ be represented by the matrix equation

$$\mathbf{X}\hat{\mathbf{p}} = \mathbf{o}, \quad (3)$$

where $\mathbf{X}$ is the model "sensitivity" or "Jacobian" matrix. Let $\mathbf{h}$ be a vector of field measurements corresponding to the model output vector $\mathbf{o}$; $\mathbf{h}$ is expressible as

$$\mathbf{h} = \mathbf{X}\mathbf{p} + \boldsymbol{\epsilon}, \quad (4)$$

where, as stated above, $\mathbf{p}$ represents the true parameters of the model (which we will never know) and ϵ represents measurement and structural noise associated with $\mathbf{h}$. Let the covariance of this noise be represented by the $n \times n$ matrix $C(\boldsymbol{\epsilon})$; for better or for worse this is normally assumed to be a diagonal matrix.

[14] Let the extent of model-to-measurement misfit be represented by an objective function Φ defined as

$$\Phi = (\mathbf{X}\hat{\mathbf{p}} - \mathbf{h})^t \mathbf{Q} (\mathbf{X}\hat{\mathbf{p}} - \mathbf{h}), \quad (5)$$

where $\mathbf{Q}$ is a positive definite "cofactor matrix." This is normally chosen to be proportional to the inverse of $C(\boldsymbol{\epsilon})$, that is,

$$C(\boldsymbol{\epsilon}) = \sigma_h^2 \mathbf{Q}^{-1}, \quad (6)$$

where the constant of proportionality σ_h^2 (the so-called "reference variance") can be estimated through the calibration process as

$$\sigma_h^2 = \Phi/(n-m), \quad (7)$$

where Φ is the objective function corresponding to an acceptable level of model-to-measurement fit.

[15] Φ of equation (5) is minimized when

$$\mathbf{X}^t\mathbf{Q}\mathbf{X}\hat{\mathbf{p}} = \mathbf{X}^t\mathbf{Q}\mathbf{h}. \quad (8)$$

If the matrix $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ is not of full rank and the inverse problem is thus underdetermined, there is no unique solution to equation (8).

[16] In highly parameterized contexts it is often possible to find parameter sets $\hat{\mathbf{p}}$ which reduce Φ to almost zero, resulting in exquisite fits between model outputs and field measurements. However, achieving such excellent fits just because they are possible is not necessarily good practice, for this does not take into account the fact that measurements are contaminated by noise. Unrealistic parameter values, and high levels of parameter spatial variability, are often the outcomes of such overfitting. Nevertheless, with due account taken of measurement noise the modeler is justified in seeking a suitably low level of model-to-measurement misfit and can thus define an appropriate objective function Φ_n that reflects the measurement and structural noise content of the observation data set on which calibration is based and seek a $\hat{\mathbf{p}}$ which satisfies

$$(\mathbf{X}\hat{\mathbf{p}} - \mathbf{h})^t \mathbf{Q} (\mathbf{X}\hat{\mathbf{p}} - \mathbf{h}) = \Phi_n. \quad (9)$$

Whether or not $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ is of full rank, the solution of equation (9) is nonunique if Φ_n is greater than the minimized objective function. Thus it must be solved through some kind of regularized inversion process. The present discussion focuses on "truncated singular value decomposition" as a regularization mechanism. However, the conclusions, and many of the equations, derived in this section and section 2.3 are just as applicable to other regularization methods such as "constrained minimization regularization," otherwise known as "Tikhonov regularization."

[17] Singular value decomposition (SVD) can be used to determine the eigenvalues and eigenvectors of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$, whether or not this matrix is of full rank. Thus

$$\mathbf{X}^t\mathbf{Q}\mathbf{X} = \mathbf{V}\mathbf{E}\mathbf{V}^t, \quad (10)$$

where $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ and $\mathbf{E}$ is a diagonal matrix of eigenvalues of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$. Where $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ has less than full rank, some of the eigenvalues in $\mathbf{E}$ are zero; in fact, $\mathbf{E}$ has as many zero-valued eigenvalues as the rank deficiency of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$. Because $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ is positive semidefinite, its eigenvalues are real, and its eigenvectors are orthogonal. Thus

$$\mathbf{V}^t = \mathbf{V}^{-1}. \quad (11)$$

Let $\mathbf{V}$ be characterized as

$$\mathbf{V} = [\mathbf{V}_1 \mathbf{V}_2], \quad (12)$$

where $\mathbf{V}_1$ contains eigenvectors corresponding to the k largest eigenvalues of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ and $\mathbf{V}_2$ contains the remaining eigenvectors, including those whose eigenvalues are zero. In order to obtain a unique solution for $\hat{\mathbf{p}}$ at an acceptable level of model-to-measurement misfit, let all eigenvectors after the kth be assigned a value of zero (hence the term

"truncated" in the description of this regularization methodology). Premultiplication of equation (8) by $\mathbf{V}_1^t\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t$ then results in

$$\mathbf{V}_1^t\hat{\mathbf{p}} = \mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q}\mathbf{h}. \tag{13}$$

In equation (13), $\mathbf{E}_1$ is the diagonal matrix of pretruncation eigenvalues of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ (all of which are nonzero). The elements of the vector $\mathbf{V}_1^t\hat{\mathbf{p}}$ are the inner product of a parameter solution vector with each of the eigenvectors contained in $\mathbf{V}_1$. Thus equation (13) solves for the projection of solutions of equation (8) onto the subspace of parameter space spanned by the eigenvectors contained in $\mathbf{V}_1$. Because $\mathbf{E}_1$ has no diagonal elements equal or close to zero (which is ensured if k is selected low enough), a stable solution to the regularized inversion problem has been obtained:

$$\hat{\mathbf{p}} = \mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q}\mathbf{h}. \tag{14}$$

In forming equation (14) we have slightly adjusted our characterization of $\hat{\mathbf{p}}$ to be $\mathbf{V}_1\mathbf{V}_1^t\hat{\mathbf{p}}$, this being the (unique) $\mathbf{V}_1$ subspace parameter vector corresponding to the vector components calculable through equation (13). Thus the m-dimensional inverse problem has been transformed into a k-dimensional inverse problem confined to the subspace of parameter space spanned by $\mathbf{V}_1$. Furthermore, by choosing k appropriately a good, but not excessively good, fit can be obtained between model outputs and field data. Normally, k is chosen such that (9) is approximately obeyed.

[18] In general, eigenvectors of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ (i.e., columns of $\mathbf{V}$) corresponding to high eigenvalues show low spatial variability within the model domain, whereas those corresponding to low eigenvalues tend to show high variability [see, e.g., *Wiggins et al.*, 1976]. Thus the truncated SVD solution process tends to select smooth solutions to the inverse problem, reflecting the inherent incapacity of a calibration data set to furnish fine system detail in most modeling contexts.

2.3. Linear Predictive Uncertainty Analysis for a Calibrated Model

[19] If equation (4) is substituted into equation (14), we obtain

$$\hat{\mathbf{p}} = \mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q}(\mathbf{X}\mathbf{p} + \boldsymbol{\varepsilon}). \tag{15}$$

Expanding terms in this equation and substituting (10), it becomes

$$\hat{\mathbf{p}} = \mathbf{V}_1\mathbf{V}_1^t\mathbf{p} + \mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q}\boldsymbol{\varepsilon}, \tag{16a}$$

that is,

$$\hat{\mathbf{p}} = \mathbf{R}\mathbf{p} + \mathbf{G}\boldsymbol{\varepsilon}, \tag{16b}$$

where

$$\mathbf{R} = \mathbf{V}_1\mathbf{V}_1^{\mathrm{t}} \tag{17}$$

and

$$\mathbf{G} = \mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q}. \tag{18}$$

$\mathbf{R}$, the so-called "resolution matrix," describes the relationship between estimated parameters and true parameters. The difference between true and estimated parameters is given by

$$\mathbf{p} - \hat{\mathbf{p}} = (\mathbf{I} - \mathbf{R})\mathbf{p} - \mathbf{G}\boldsymbol{\varepsilon}. \tag{19}$$

Equation (19) expresses the "parameterization wrongness" of a calibrated model; unfortunately, this cannot be calculated because $\mathbf{p}$ is unknown. However, its expected value (i.e., $E(\mathbf{p} - \hat{\mathbf{p}})$) is proportional to the expected value of $\mathbf{p}$ (i.e., $E(\mathbf{p})$), assuming that $E(\boldsymbol{\varepsilon})$ is zero. As is stated in section 2.1, with the elements of $\mathbf{p}$ defined as parameter perturbations from their (assumed known) average values, $E(\mathbf{p} - \hat{\mathbf{p}})$ is zero.

[20] Let it be assumed that the covariance matrix of $\mathbf{p}$ (i.e., $C(\mathbf{p})$ as featured in equation (2)) and the covariance matrix of measurement and structural noise (i.e., $C(\boldsymbol{\varepsilon})$ as featured in equation (6)) are known; let it be further assumed that $\mathbf{p}$ and $\boldsymbol{\varepsilon}$ are independent. Then $C(\mathbf{p} - \hat{\mathbf{p}})$ is easily calculated from (19) as

$$C(\mathbf{p} - \hat{\mathbf{p}}) = (\mathbf{I} - \mathbf{R})C(\mathbf{p})(\mathbf{I} - \mathbf{R})^t + \mathbf{G}C(\boldsymbol{\varepsilon})\mathbf{G}^t. \tag{20}$$

If (17) and (18) plus (6) are now substituted into (20) and use is made of the relationship

$$\mathbf{V}\mathbf{V}^t = \left(\mathbf{V}_1\mathbf{V}_1^t + \mathbf{V}_2\mathbf{V}_2^t\right) = \mathbf{I}, \tag{21}$$

we obtain

$$C(\mathbf{p} - \hat{\mathbf{p}}) = \mathbf{V}_2\mathbf{V}_2^tC(\mathbf{p})\mathbf{V}_2\mathbf{V}_2^t + \sigma_h^2\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t. \tag{22}$$

This expression can be made even simpler if the precalibration probability distribution $C(\mathbf{p})$ of model parameters is such that they are all independently variable and have the same variance σ_p^2. In this case, equation (20) becomes

$$C(\mathbf{p} - \hat{\mathbf{p}}) = \sigma_p^2\mathbf{V}_2\mathbf{V}_2^t + \sigma_h^2\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t. \tag{23}$$

If a model prediction s is calculated from model parameters $\mathbf{p}$ using equation (1), we can compare the "model-calculated prediction" (i.e., the prediction made on the basis of calibrated parameters $\hat{\mathbf{p}}$, designated herein as $\hat{s}$) with the true prediction (i.e., the prediction made on the basis of true parameters $\mathbf{p}$, designated herein as s) using the formula

$$s - \hat{s} = \mathbf{y}^t(\mathbf{p} - \hat{\mathbf{p}}). \tag{24}$$

Equation (24) expresses the "wrongness" or "error" of a model prediction. Once again, this can never be known. For the same reasons as those already stated with respect to parameter wrongness its expected value is zero. Its variance can be calculated from (24) as

$$\sigma_{s-\hat{s}}^2 = \mathbf{y}^tC(\mathbf{p} - \hat{\mathbf{p}})\mathbf{y}. \tag{25}$$

Substitution of this relationship into (20), (22), and (23) yields

$$\sigma^2_{s-\hat{s}} = \mathbf{y}^t(\mathbf{I}-\mathbf{R})C(\mathbf{p})(\mathbf{I}-\mathbf{R})^t\mathbf{y} + \mathbf{y}^t\mathbf{G}C(\varepsilon)\mathbf{G}^t\mathbf{y}, \tag{26a}$$

$$\sigma^2_{s-\hat{s}} = \mathbf{y}^t\mathbf{V}_2\mathbf{V}_2^tC(\mathbf{p})\mathbf{V}_2\mathbf{V}_2^t\mathbf{y} + \sigma_h^2\mathbf{y}^t\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{y}, \tag{26b}$$

$$\sigma^2_{s-\hat{s}} = \sigma_p^2\mathbf{y}^t\mathbf{V}_2\mathbf{V}_2^t\mathbf{y} + \sigma_h^2\mathbf{y}^t\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{y}. \tag{26c}$$

From the above derivation it is apparent that equations (26b) and (26c) are special cases of (26a), derivable from it by substitution of appropriate expressions for **R** and **G**. In fact, equation (26a) is perfectly general and pertains to no regularization method in particular. Thus exactly the same formula can be used to calculate predictive error variance where regularization is achieved through a constrained minimization process. However, in that case, equations (17) and (18) are replaced by

$$\mathbf{R} = \left(\mathbf{X}^t\mathbf{Q}\mathbf{X} + \beta^2\mathbf{Z}^t\mathbf{Q}_r\mathbf{Z}\right)^{-1}\mathbf{X}^t\mathbf{Q}\mathbf{X} \tag{27a}$$

$$\mathbf{G} = \left(\mathbf{X}^t\mathbf{Q}\mathbf{X} + \beta^2\mathbf{Z}^t\mathbf{Q}_r\mathbf{Z}\right)^{-1}\mathbf{X}^t\mathbf{Q}, \tag{27b}$$

where **Z** is a matrix of regularization constraints on parameter values (assumed to be linear), $\mathbf{Q}_r$ is the "regularization weight matrix," and β^2 is the "squared regularization weight factor," which can also be considered to be a Lagrange multiplier in the constrained minimization process; see *Doherty* [2003] for a description of this type of regularization and for an example of its use in the groundwater modeling context.

2.4. Significance of Equations

[21] Equations (26) are of great importance. Formulation of predictive error variance using these equations has the benefit that the contributions made to this variance by two different aspects of the model parameterization process are made explicit. The second term of equations (26) is the component of model predictive uncertainty that arises from model-to-measurement misfit. In an overdetermined system (where parameters are outnumbered by observations) this is the only source of model predictive uncertainty considered to exist, for under these conditions the resolution matrix **R** is actually the identity matrix, and the first term of equations (26) vanishes. Model predictive error analysis based on this term has been undertaken by a number of authors in the groundwater modeling context [see, e.g., *Hill*, 1989; *Christensen and Cooley*, 1999; *Vecchia and Cooley*, 1987].

[22] The first term of equations (26) accommodates the fact that the calibration process cannot capture all of the hydraulic detail prevailing within a study area. The further removed the resolution matrix **R** is from the identity matrix **I**, the larger this term is. In general, both data scarcity and high data noise content promote "blurry" resolution matrices and hence loss of system detail in a calibrated model. This can lead to grossly inaccurate model predictions where these predictions depend on that detail (i.e., when parameters are sensitive to that detail and hence **y** is nonorthogonal to $(\mathbf{I}-\mathbf{R})\mathbf{p}$).

[23] For an uncalibrated model the second term of equations (26) is zero, and the resolution matrix **R** becomes the null **0** matrix (this can be demonstrated in equation (27) by setting **Q** to zero, effectively giving all observations a weight of zero in the inversion process). Equation (26a) then becomes equation (2), as it should.

[24] The second term of equations (26) also becomes zero when a perfect fit is obtained between model outputs and field measurements because of the absence of any measurement or structural noise (i.e., $C(\varepsilon)$ is **0**). The fact that predictions made by a "perfectly calibrated" model can be substantially in error is readily apparent from an inspection of the $\mathbf{y}^t(\mathbf{I}-\mathbf{R})$ portion of the remaining term. For an underdetermined system, **R** is rank deficient; therefore its columns span only a subspace of parameter space. It is thus possible for $\mathbf{y}^t\mathbf{R}$ to be zero; under these circumstances the calibration process does nothing whatsoever to reduce the uncertainty of that particular model prediction, for equation (26a) then yields the same result as equation (2).

[25] This analysis can be taken a step further. When model-to-measurement misfit is zero, **R** becomes a projection operator. This is easily demonstrated by observing that

$$\mathbf{R}\mathbf{R}\mathbf{p} = \mathbf{R}\hat{\mathbf{p}} \tag{28a}$$

(from $\hat{\mathbf{p}} = \mathbf{R}\mathbf{p}$) and

$$\mathbf{R}\hat{\mathbf{p}} = \hat{\mathbf{p}} \tag{28b}$$

and therefore

$$\mathbf{R}\mathbf{R}\mathbf{p} = \hat{\mathbf{p}} = \mathbf{R}\mathbf{p} \qquad \mathbf{R}\mathbf{R} = \mathbf{R}. \tag{29}$$

Equation (28b) follows from the fact that $\mathbf{X}\mathbf{p} = \mathbf{X}\hat{\mathbf{p}}$. Also following from this is the fact that

$$\mathbf{X}\mathbf{R} = \mathbf{X}. \tag{30}$$

From equation (17) it can be shown that if

$$\mathbf{X}\mathbf{y} = \mathbf{0}, \text{ then } \mathbf{R}\mathbf{y} = \mathbf{0}. \tag{31}$$

From (31) it follows that $\mathbf{y}^t\mathbf{R}$ will be zero if $\mathbf{X}\mathbf{y}$ is **0** and **R** is symmetrical, as it always is when regularized inversion is implemented using truncated SVD. From this it follows that if, for a particular model prediction, the vector of predictive sensitivities (i.e., **y**) is perpendicular to all observation sensitivity vectors (i.e., to all rows of the **X** matrix), then the calibration process does nothing to decrease the uncertainty of this prediction. Thus if a model is calibrated against data types which bear little relation to the types of predictions that a model will be required to make, then there can be no guarantee that the calibration process will reduce the uncertainties of these predictions at all.

[26] Examination of equation (26c) allows further insight to be gained into the role of the calibration process in reducing (or not) predictive uncertainty. Consider that a prediction sensitivity vector **y** is parallel to an eigenvector of **V**. That particular eigenvector must feature in either the first term or the second term of (26c), depending on whether it belongs to $\mathbf{V}_1$ or to $\mathbf{V}_2$. If it belongs to $\mathbf{V}_2$, then potential

wrongness in the model prediction arises from the fact that the calibration process provides no information that is relevant to that prediction. Potential predictive error is governed entirely by σ_p^2, the inherent (precalibration) uncertainty of system properties, for the second term in (26c) is zero because of orthogonality of $\mathbf{y}$ to all members of $\mathbf{V}_1$ (because the eigenvectors composing the columns of $\mathbf{V}$ are all orthogonal to each other). Thus the uncertainty of this prediction is undiminished from that which prevailed prior to model calibration.

[27] On the other hand, if the prediction sensitivity vector $\mathbf{y}$ is parallel to a $\mathbf{V}_1$ eigenvector, the first term of equation (26c) is zero and the second term is nonzero. The magnitude of this second term depends on two factors. One is the goodness of model-to-measurement fit as encapsulated in the term σ_h^2; the other is the magnitude of the eigenvalue corresponding to the eigenvector to which the prediction sensitivity vector is parallel. If this eigenvalue is small, the contribution to uncertainty arising from the second term can be very large, possibly larger than if the model had not been calibrated at all. In this case the calibration data set says less about the parameter combinations that define prediction sensitivity than can be said on the basis of knowledge about system properties from outside of the calibration process altogether. Under these circumstances, once again, the calibration process provides no assistance in reducing predictive uncertainty below that which exists if the model has not been calibrated at all. On the other hand, if the pertinent eigenvalue is large, then the calibration process may reduce predictive error variance substantially, the extent to which it does this being dependent on the measurement error variance σ_h^2.

2.5. Tailoring the Calibration Process to Reduce Predictive Error

[28] The notion of a calibrated model conveys the idea that a model, once calibrated, can be used to make a variety of different predictions of system behavior. It also suggests that the calibration process is independent of the prediction process. The analysis presented in section 2.4, however, suggests that the usefulness of environmental models in making critical predictions of system behavior can be enhanced if the calibration process is undertaken with predictions required by the model kept specifically in mind.

[29] When undertaking regularized inversion based on truncated SVD, eigenvectors can be shifted from $\mathbf{V}_2$ to $\mathbf{V}_1$ of equations (26) to increase goodness of fit to a level considered acceptable. This is normally done in order of decreasing respective eigenvalue; that is, eigenvectors corresponding to high eigenvalues (and hence respecting broad-scale hydraulic property distributions) are normally shifted to $\mathbf{V}_1$ while those with low eigenvalues (reflecting system detail) are retained in $\mathbf{V}_2$. Where no eigenvectors are shifted to the second term, the predictive variance is the same as that of an uncalibrated model (see equation (2)). As eigenvectors are shifted from $\mathbf{V}_2$ to $\mathbf{V}_1$, the first term of equation (26c) falls monotonically, while the second term rises monotonically. (The same will be generally true of equations (26a) and (26b), though departures from this may occur if $C(\hat{\mathbf{p}})$ entails high correlation between individual parameters and/or regularization is not undertaken using truncated SVD.)

[30] In most underdetermined calibration contexts the sum of the two terms of equation (26c) will show a minimum as eigenvectors are transferred from the first to the second term. This occurs because of the fact that under these circumstances, $\mathbf{E}_1$ possesses zero or near-zero diagonal elements which greatly magnify the second term as corresponding eigenvectors are transferred to it. Here it is assumed that there is enough salient information within the observation data set (e.g., enough high-valued elements of $\mathbf{E}_1$) for the calibration process to reduce the variance of the prediction at least a little from that pertaining to an uncalibrated model before it commences to rise, thus causing the minimum of the curve. Where it exists, the location of this minimum will be dependent on the particular prediction being investigated. However, there is no certainty that the number of eigenvalues at which this predictive uncertainty minimum is achieved corresponds to the number of eigenvalues required to achieve (but not to undercut) a suitably defined Φ_n. In fact, as is demonstrated in section 3.2, minimization of error variance for a particular prediction may require use of many eigenvalues beyond that which is required to achieve a suitable value for Φ_n and can thus result in what classical analysis would perceive as overfitting. Furthermore, the truncation level for minimizing the uncertainty of one specific prediction may not be the same as that required to minimize the uncertainty of another prediction, lending weight to the assertion made in section 1 that model calibration may need to be prediction-specific.

[31] So how can model calibration be "tuned" to the prediction that it must make? A number of options exist. One is to actively seek the minimum in the predictive variance curve, even if this leads to overfitting according to the classical view of model calibration. Another option is to vary from the traditional practice of ranking eigenvalues in decreasing order of their magnitude when deciding on a level of truncation. This traditional strategy always leads to the loss of high eigenvalues from $\mathbf{V}_2$ before low eigenvalues, regardless of the disposition of corresponding eigenvectors with respect to a model prediction of particular interest. An alternative strategy is to take account of the orientation of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ eigenvectors with respect to predictive sensitivity when deciding on the order of eigenvector excision from $\mathbf{V}_2$. Thus, for example, eigenvector A may have a lower eigenvalue than eigenvector B, yet its inner product with $\mathbf{y}$ may be higher. If its eigenvalue is not so low that its presence in the second term of equation (26) results in higher predictive uncertainty than its presence in the first, it should be included in the second term (i.e., the $\mathbf{V}_1$ term), thus avoiding excision in the truncation procedure.

[32] An alternative, and simpler, strategy can be employed to increase the likelihood that eigenvectors of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ which are parallel to a key model prediction belong to $\mathbf{V}_1$ rather than to $\mathbf{V}_2$, thus reducing the probability of model error in making that prediction. This strategy is to increase the weights associated with observations whose sensitivities are more aligned with a particular prediction relative to those that are orthogonal to it. (Greater alignment of a prediction with an observation is identified through greater normalized inner product of the respective vectors, normalization being achieved by dividing by the product of the magnitude of the two vectors.) This procedure results in

a $\mathbf{Q}$ matrix for which equation (6) no longer applies. The second term in equations (26b) and (26c) becomes more complicated as a result; equations (32) repeat equations (26) where $C(\varepsilon)$ is no longer proportional to $\mathbf{Q}^{-1}$:

$$\sigma^2_{s-\hat{s}} = \mathbf{y}^t(\mathbf{I}-\mathbf{R})C(\mathbf{p})(\mathbf{I}-\mathbf{R})^t\mathbf{y} + \mathbf{y}^t\mathbf{G}C(\varepsilon)\mathbf{G}^t\mathbf{y}, \tag{32a}$$

$$\sigma^2_{s-\hat{s}} = \mathbf{y}^t\mathbf{V}_2\mathbf{V}_2^tC(\mathbf{p})\mathbf{V}_2\mathbf{V}_2^t\mathbf{y} + \sigma_h^2\mathbf{y}^t\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q}C(\varepsilon)\mathbf{Q}\mathbf{X}\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{y}, \tag{32b}$$

$$\sigma^2_{s-\hat{s}} = \sigma_p^2\mathbf{y}^t\mathbf{V}_2\mathbf{V}_2^t\mathbf{y} + \sigma_h^2\mathbf{y}^t\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q}C(\varepsilon)\mathbf{Q}\mathbf{X}\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{y}. \tag{32c}$$

With $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ thus reformulated its eigenvectors and eigenvalues also change. However, now the eigenvectors which are more parallel to $\mathbf{y}$ will tend to have higher eigenvalues and hence will be less likely to be truncated in a calibration procedure that orders eigenvalues by decreasing magnitude prior to truncation. Care must be taken in implementing this procedure, however, to ensure that the second term of equations (32) is not unduly amplified by providing high weights to observations that are inherently unreliable.

[33] This philosophy of weights assignment violates traditional least squares practice. This traditional practice is based on predictive error variance minimization [*Bard*, 1974] for overdetermined systems, in which all parameters to which a prediction is sensitive are assumed to be individually estimable through the calibration process. However, it is worth noting that the strategy of placing increased emphasis on observation types that most resemble the types of predictions that a model will be required to make is certainly in harmony with the philosophical underpinnings of manual calibration which is often based on the simple, but effective, premise that "if you can't fit everything, then at least fit the things that matter most." The effectiveness of this strategy is demonstrated in section 3.3 using a synthetic example.

[34] It must also be pointed out that adherence to traditional weights assignment practices that are recommended by authors such as *Hill* [1998] presupposes that $C(\varepsilon)$ is known or can be estimated. Where calibration data noise is dominated by model structural error (as is mostly the case), $C(\varepsilon)$ is not known nor can it be easily estimated. Furthermore, in many contexts such "noise" may show considerable spatial correlation of unknown magnitude. Thus the common practice of assuming independence of measurement errors (and thus a diagonal $\mathbf{Q}$ matrix) is, in fact, a violation of the precepts espoused in guidelines such as these. In view of this fact the assignment of weights in a manner that places greater emphasis on observations that are more closely related to key model predictions is probably no less in violation of these precepts than many other commonly implemented methods of weights assignment; however, in view of its probably beneficial outcome of reducing predictive error variance it can be far more effective.

[35] Where regularized inversion is undertaken using methods other than truncated SVD (e.g., constrained minimization), it is also possible to tailor the model calibration process such that the variance of model predictive error is minimized. In equations (27) the reciprocal of the squared regularization weight factor β^2 plays a similar role to that of k, the eigenvalue truncation number employed by the truncated SVD method; higher values of β^2 result in smoother calibrated fields and higher values of Φ. Normally, a value of β^2 is sought which results in an "adequate" level of model-to-measurement misfit, that is, an objective function equal to a suitably chosen Φ_n. However, as will be demonstrated in section 3.2, upward variation of β^2 results in a monotonic lowering of the second term of equation (26a) and a monotonic rise in the first term. Hence there is a value of β^2 for which the error variance of a particular prediction is minimized. This value will rarely coincide with that required to exactly achieve Φ_n. Hence one option for tuning the calibration process for prediction optimization is to seek that β^2 which minimizes predictive error variance. Alternatively (or as well), higher weights could be assigned to those members of a calibration data set that are most pertinent to a particular prediction, as was discussed above in relation to truncated SVD. A further strategy may be to tailor the regularization constraint matrix $\mathbf{Z}$ to best accommodate the prediction that the model is required to make.

2.6. Effect of Parameter Lumping

[36] Classical approaches to model calibration undertake "preemptive" or "implicit" regularization using some form of parameter lumping. For a groundwater model this often takes the form of spatial parameter definition using a limited number of zones of piecewise parameter uniformity. This approach to regularization has the advantage that if zones are few enough and are defined in accordance with spatial data density, numerical stability of the inversion process is guaranteed. However, it has the disadvantage that some of the information content of the calibration data set may be lost because the chosen parameterization methodology has limited means of expressing hydraulic property complexity that may become apparent through the calibration process itself. (It should be noted in passing that regularized inversion and geologically based parameter zonation are not mutually exclusive [see, e.g., *De Groot-Hedlin and Constable*, 1990]. The advantage of combining zones with regularized inversion based on a large number of parameters is that intrazonal heterogeneity can be accommodated in the model at the same time as interzonal hydraulic property contrasts if the calibration data set provides a strong enough indication that such intrazonal complexity exists.)

[37] Suppose that instead of estimating m parameters whose true values are encapsulated in the vector $\mathbf{p}$, j lumped parameters comprising the elements of a smaller vector $\hat{\mathbf{r}}$ are estimated in their stead. Suppose further that model outputs corresponding to observations are calculated using the relationship

$$\mathbf{o} = \mathbf{W}\hat{\mathbf{r}}; \tag{33}$$

$\hat{\mathbf{r}}$ can be determined through objective function minimization using the formula (same as equation (8))

$$\hat{\mathbf{r}} = (\mathbf{W}^t\mathbf{Q}\mathbf{W})^{-1}\mathbf{W}^t\mathbf{Q}\mathbf{h}. \tag{34}$$

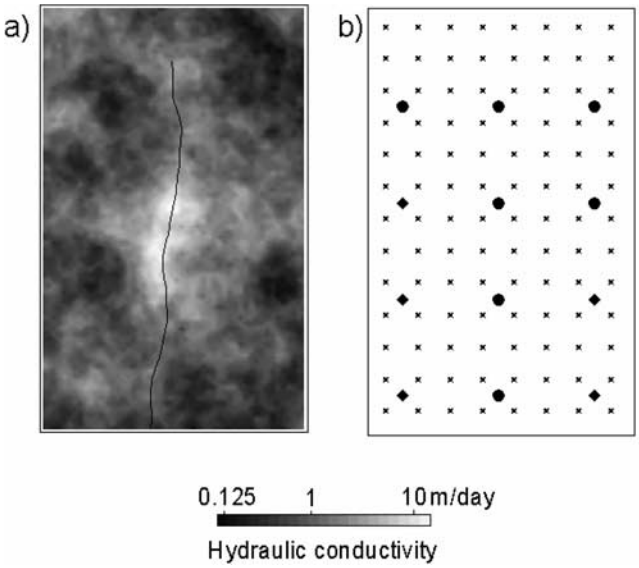

Figure 1. (a) Hydraulic conductivity distribution within rectangular model domain and trace of released particle. (b) Locations of observation wells (circles and diamonds) and pilot points (crosses). Model domain is 500 m × 800 m.

Substitution of (4) then yields

$$\hat{\mathbf{r}} = (\mathbf{W}^t\mathbf{Q}\mathbf{W})^{-1}\mathbf{W}^t\mathbf{Q}(\mathbf{X}\mathbf{p} + \boldsymbol{\varepsilon}) \tag{35a}$$

$$= \mathbf{R}'\mathbf{p} + \mathbf{G}'\boldsymbol{\varepsilon}, \tag{35b}$$

where the "modified resolution matrix" $\mathbf{R}'$ and the matrix $\mathbf{G}'$ of equations (35) are given by

$$\mathbf{R}' = (\mathbf{W}^t\mathbf{Q}\mathbf{W})^{-1}\mathbf{W}^t\mathbf{Q}\mathbf{X} \tag{36a}$$

and

$$\mathbf{G}' = (\mathbf{W}^t\mathbf{Q}\mathbf{W})^{-1}\mathbf{W}^t\mathbf{Q}. \tag{36b}$$

Let the relationship between a parameterization based on many parameters $\hat{\mathbf{p}}$ and that based on lumping of these parameters (i.e., $\hat{\mathbf{r}}$) be described by

$$\hat{\mathbf{p}} = \mathbf{L}\hat{\mathbf{r}}, \tag{37}$$

where $\mathbf{L}$ is a $m \times j$ "lumping matrix." In many cases each row of $\mathbf{L}$ will be composed of zero elements except for a single element of 1. For example, if zone-specific pilot points [see *Doherty*, 2003] are employed as the spatial parameterization basis for $\hat{\mathbf{p}}$, while the elements of $\hat{\mathbf{r}}$ are zonal parameter values, each element of $\hat{\mathbf{p}}$ will have the same value as the element of $\hat{\mathbf{r}}$ that pertains to the zone in which the corresponding pilot point lies.

[38] With $\mathbf{L}$, $\mathbf{R}'$, and $\mathbf{G}'$ defined as above it is easy to show that $\mathbf{R}$ and $\mathbf{G}$ matrices for use in equation (26a) can be calculated as

$$\mathbf{R} = \mathbf{L}\mathbf{R}' \tag{38a}$$

and

$$\mathbf{G} = \mathbf{L}\mathbf{G}'. \tag{38b}$$

With these definitions of $\mathbf{R}$ and $\mathbf{G}$, equation (26a) provides the full expression for predictive error variance of a lumped parameter model (and is equivalent to the linear part of *Cooley* [2004, equation 3.32]). For reasons already discussed in relation to underdetermined systems, omission of the first term of this equation (as is usually done in practice) can lead to significant underestimation of predictive error variance, especially where the underlying system is complex. Unfortunately, however, equation (26a) may be difficult to apply in practice. This is because its evaluation assumes that even though only lumped parameters are estimated, sensitivities of a much larger number of distributed parameters have also been calculated (for use in the $\mathbf{X}$ matrix). Nevertheless, this analysis demonstrates that estimation of predictive error variance based only on the statistics of measurement noise neglects an extremely important contributor to potential model error. Where lumping is significant and $\mathbf{R}$ is thus significantly different from $\mathbf{I}$ and/or where predictive sensitivities have a large component in the subspace of m-dimensional parameter space spanned by $(\mathbf{I} - \mathbf{R})$, neglect of this contribution may make such estimates almost meaningless.

3. Synthetic Example

3.1. Model Description

[39] The principles discussed in section 2 are illustrated with reference to a synthetic model. Figure 1a shows the 500 m × 800 m rectangular domain of a single-layer groundwater model of flow in a confined aquifer of 10 m thickness. A fixed inflow of 0.1 $m^3\ d^{-1}\ m^{-1}$ occurs through the upper boundary of the model; heads are fixed at 0 m along the lower boundary. A hydraulic conductivity field with a log average value of zero was generated using a log exponential variogram with a range (3 times the coefficient in the exponent in the variogram equation) of 600 m and a sill of 0.2. Diffuse recharge is zero. Flow within the domain was simulated using MODFLOW-2000 [*Harbaugh et al.*, 2000] using a finite difference grid consisting of 50 rows and 80 columns of 10 m square cells. The travel time and track of a particle released near the top boundary was simulated using the ADV package of MODFLOW-2000 [*Anderman and Hill*, 2001]; the path of the particle is depicted in Figure 1a.

[40] On the basis of the hydraulic conductivity field shown in Figure 1a, heads were generated at 12 wells, the locations of which are shown in Figure 1b; heads in these wells vary between 5.7 m in the upper part of the model domain and 1.1 m in the lower part of the domain. These heads were used for model calibration after the addition of Gaussian noise with a standard deviation of 0.3 m.

[41] Spatial parameterization was implemented using pilot points. As described by *Doherty* [2003], the calibration process assigns hydraulic conductivity values to these points; these values are then spatially interpolated to all cells of the model domain (using kriging in the present case).

[42] Regularized inversion of the head data was undertaken using PEST [*Doherty*, 2004], with assistance from the

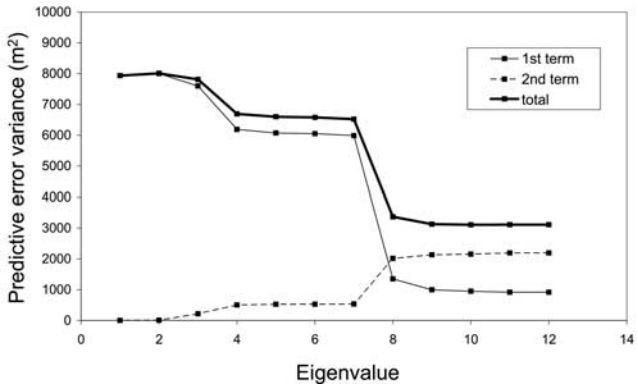

Figure 2. Terms of equation (26) and total model predictive error variance versus number of eigenvalues. Sensitivities were calculated on the basis of parameters estimated using four eigenvalues.

PEST groundwater utilities [*Doherty*, 2003]; both truncated SVD and constrained minimization regularization were employed. In the latter case, regularization constraints were of the "preferred value" type, with the log of each pilot point hydraulic conductivity being assigned a preferred value of zero; thus the **Z** matrix of equation (27) was an $m \times m$ identity matrix. $\mathbf{Q}_r$ was calculated as the inverse of an inter–pilot point covariance matrix; pilot point covariances were calculated using the same variogram as that employed for generation of the hydraulic conductivity field. The squared regularization weight factor β^2 was calculated by PEST as that required to achieve a user-supplied value for Φ_n, the objective function at which "adequate calibration" is deemed to occur in accordance with the level of measurement noise.

[43] Analyses in this section focus on a prediction of the particle exit location, the true value of this prediction being 206.78 m from the left side of the model. This prediction, rather than travel time, was chosen for the analyses described herein because of the fact that the latter prediction is relatively unconstrained by a calibration process that is based on heads alone.

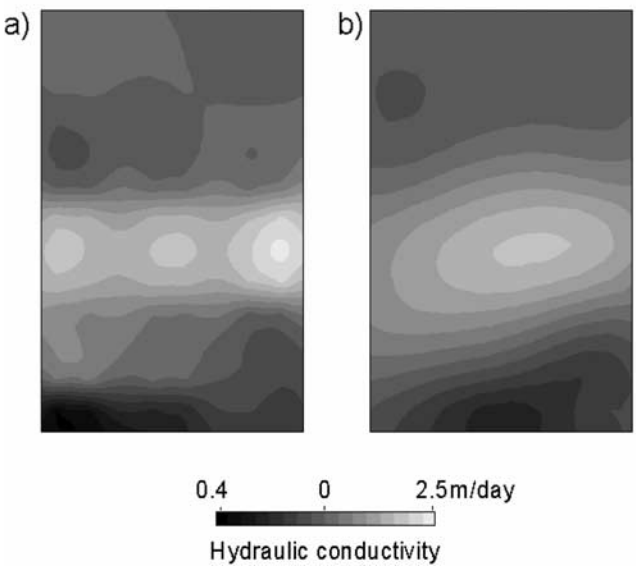

Figure 3. Calibrated hydraulic conductivity fields calculated using (a) truncated singular value decomposition with four eigenvalues and (b) constrained minimization regularization with Φ_n set to 12.0.

Table 1. Results of Truncated SVD Inversion With Varying Number of Eigenvalues[a]

Eigenvalues Before Truncation	Objective Function, m^2	Exit Point Prediction, m
1	34.71	245.3
2	21.09	244.7
3	13.08	257.9
4	11.18	251.9
5	11.19	247.3
6	5.5	264.2
7	3.15	187.0
8	3.087	172.3
9	3.35	183.5
10	1.64	123.4
11	3.155×10^{-3}	159.5
12	1.8×10^{-11}	155.7

[a]True exit point location is 206.8 m. SVD is singular value decomposition.

3.2. Predictive Variance Minimization

[44] With Gaussian noise of standard deviation 0.3 m added to the heads an objective function value (i.e., Φ_n of equation (9)) of 12.0 should be sought in a regularized inversion process in which measurement weights are all set to 3.33, this being the inverse of the noise standard deviation. (Note that weights are squared to form the diagonal elements of **Q**.)

[45] Table 1 summarizes the outcomes of undertaking regularized inversion using truncated SVD with a varying number of pretruncation eigenvalues. It is apparent from Table 1 that an objective function of 12 can be achieved with as few as four eigenvalues. If the terms of equations (26) are computed using sensitivities calculated on the basis of calibrated parameters (i.e., parameters calculated using four eigenvalues), the graphs of Figure 2 result; note that $C(\mathbf{p})$ in equations (26) was calculated using the same variogram as that which was used to generate the hydraulic conductivity field and thus properly represents the spatial characteristics of the true hydraulic property field. Figure 3a shows the calibrated parameter field.

[46] The monotonic rise of the second term of equations (26) and the monotonic fall of the first term are apparent from Figure 2. As the number of eigenvalues increases, the second term does not rise fast enough relative to the first term for the sum of the two terms to incur a minimum; however, if the graph were to be extended to 13 eigenvalues (at which stage the second term of equation (26) would be extremely high because of the fact that with 12 observations the rank of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ is only 12), a pronounced minimum at 12 eigenvalues would be apparent.

[47] Figure 2 demonstrates that even though the model can be considered to be calibrated with truncation occurring at 4 eigenvalues, minimum predictive error variance is achieved at 12 eigenvalues. It is interesting to note from Table 1 that the most accurate prediction of particle exit point location was made by the model calibrated using seven eigenvalues.

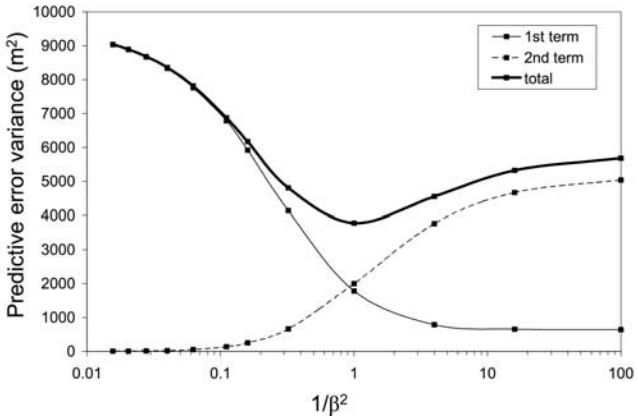

Figure 4. Terms of equation (26) and total model predictive error variance versus $1/\beta^2$. Sensitivities were calculated on the basis of parameters estimated using a β^2 value of 3.10 (i.e., a $1/\beta^2$ value of 0.32).

[48] The calibration process was repeated using constrained minimization regularization. For Φ_n equal to 12 the corresponding β^2 value is 3.10. The resulting calibrated parameter field is shown in Figure 3b.

[49] Using sensitivities calculated on the basis of the calibrated parameter field, the terms of equation (26) were calculated for varying values of β^2 (see Figure 4). Monotonicity of the two terms of equation (26) is clearly demonstrated in Figure 4. It is also apparent that the minimum predictive error variance is achieved at a β^2 value of $\sim$1.0. This is significantly less than the value of 3.10 required to achieve a Φ_n of 12, thereby allowing a closer fit between model outputs and field measurements to be attained if this value were used in a regularized inversion process. This demonstrates once again that model performance with respect to the prediction of particle exit location is optimized when the calibration process involves some degree of overfitting. (The fact that the contributions to total predictive error variance made by the terms of equation (26) are about equal at the point of minimum predictive variance is a coincidence.)

[50] Table 2 shows model-predicted exit points calculated on the basis of parameters achieved through regularized inversion using different values of β^2. The closest prediction to the actual exit point occurs at a β^2 value of between 0.74 and 1.25.

3.3. Observation Weights Adjustment for Prediction Optimization

[51] In section 3.2 it is demonstrated that a model can be considered calibrated yet may not be optimally parameterized for the making of a specific prediction. It is suggested in section 2 of this paper that if data are weighted in accordance with their relevance to a specific prediction, the calibrated model may be capable of making that prediction with a lower probability of error.

[52] For the analyses documented in section 3.2, observation weights were uniformly set at 3.33, this being the inverse of measurement uncertainty. These analyses were repeated with weights for 5 of the 12 observations being doubled. With this weighting strategy the model can be considered calibrated with an objective function of 27.0. The observations chosen for enhanced weighting were head values at the wells depicted as diamonds in Figure 1b. These observations were chosen for special treatment because the normalized inner products of the prediction sensitivity with the sensitivities of these observations were all greater than 0.1; for all other observations, normalized inner products were less than 0.1. Inner products were calculated using sensitivities pertaining to the parameters depicted in Figure 3b. (The fact that none of these observations lie along the particle path is of interest. These head observations are sensitive to the hydraulic conductivity of material occupying the boundary areas of the model domain; hydraulic conductivities are lower here than in the central part of the model domain and heads here are thus more sensitive to changes in their values. The direction in which the particle moves is also sensitive to changes in hydraulic conductivity in these boundary areas, for these low conductivities cause the fixed amount of water inflow through the top model boundary to be concentrated to some extent within the central part of the model domain. Any local reduction of these boundary area conductivities would thus result in the flow of more water into these areas and would thus result in movement of the particle toward them.)

Table 2. Results of Constrained Minimization Inversion Using Different Squared Regularization Weight Factors[a]

β^2	Objective Function, m^2	Exit Point Prediction, m
140.19	30.0	220.5
32.04	24.0	242.1
5.63	15.0	233.1
3.10	12.0	226.9
1.99	10.0	244.3
1.25	8.0	211.1
0.74	6.0	197.0
0.28	3.0	163.9
0.093	1.0	140.6
0.017	0.1	154.1

[a]True exit point location is 206.8 m.

[53] Table 3 shows the results of truncated SVD regularized inversion with a varying truncation limit. Once again, the model can be said to be calibrated if only four eigenvalues are employed. Figure 5 shows the terms of equation (32) (which must now be used instead of (26) because

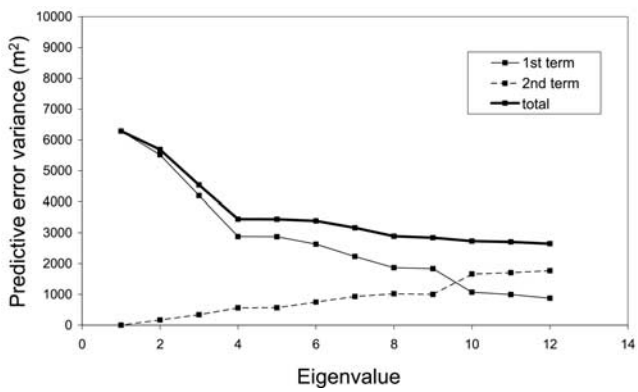

Figure 5. Terms of equation (32) and total model predictive error variance versus number of eigenvalues. Sensitivities were calculated on the basis of parameters estimated using four eigenvalues. Weights were doubled at wells shown as diamonds in Figure 1b.

Table 3. Results of Truncated SVD Inversion With Varying Number of Eigenvalues With Measurement Weights Doubled for Wells Shown as Diamonds in Figure 1b[a]

Eigenvalues Before Truncation	Objective Function, m^2	Exit Point Prediction, m
1	73.74	244.9
2	37.35	240.2
3	29.69	245.2
4	20.12	220.3
5	13.95	205.4
6	13.57	204.9
7	6.847	203.9
8	3.933	146.6
9	3.919	158.2
10	1.363	138.9
11	0.399	143.4
12	2.33×10^{-11}	155.7

[a]True exit point location is 206.8 m.

equation (6) no longer applies) calculated using sensitivities pertaining to the calibrated model. The minimum predictive error variance (once again obtained at 12 eigenvalues) is slightly lower than that obtained with uniform weights. What is more important, however, is that the predictive error variance at four eigenvalues is much lower than for uniform weights. Thus the calibrated model is a much better predictor of the exit point location. Calibration results listed in Table 3 support this conclusion.

[54] The worth of selective observation weights enhancement was also tested using constrained minimization regularization. Figure 6 and Table 4 show the results. Calibration is achieved at a β^2 value of $\sim$10.9 (i.e., with $1/\beta^2$ equal to $\sim$0.092). Predictive error variance at the point of calibration is lower in Figure 6 than in Figure 4 where no weights enhancement was undertaken. Table 4 shows that the calibrated model is indeed a good predictor of particle exit point location.

[55] Before concluding this section it must be pointed out that the parameterization strategy employed in this example incorporates a form of lumping or "averaging," this being incurred through the kriging process by which pilot point values are transferred to the model grid. (The benefits of using a pilot points scheme for spatial parameter definition are explained by *Doherty* [2003].) Hence the "true parameter vector" **p** is, in fact, a slightly smoothed version of the real hydraulic conductivity distribution. In some circumstances this mechanism of spatial parameterization could place some constraints on the parameter field assigned to the model domain through the calibration process, potentially causing it to be smoother than it would otherwise be. In the present instance, however, parameter field smoothing incurred by kriging is small compared with that incurred by calibration, as an inspection of the calibrated fields in Figure 3 suggests. Thus the use of kriging in the parameterization process does not limit the ability of the calibration process to extract information from the calibration data set, this being readily verified by the fact that we can, if we wish, obtain a solution to the inverse problem with zero model-to-measurement discrepancy (through using 12 eigenvalues or a very low value of β^2).

[56] The presence of kriging-induced smoothing does, however, slightly compromise estimates of predictive error variance calculated in this example due to the sensitivity of travel path predictions to this lost detail. Underprediction of predictive error variance incurred through this mechanism has, however, been calculated to be slight; it could have been further reduced through the use of more pilot points than the 104 employed in this study. Alternatively, its contribution to predictive error variance could have been computed by other means, and its effect could have been included in the calculated total predictive error variance; this topic, however, is beyond the scope of the present analysis and will form the subject of a later paper.

Table 4. Results of Constrained Minimization Inversion Using Different Squared Regularization Weight Factors With Measurement Weights Doubled for Wells Shown as Diamonds in Figure 1b[a]

β^2	Objective Function, m^2	Exit Point Prediction, m
42.02	42	230.9
17.32	32	218.8
10.92	27	210.7
6.69	22	201.3
1.99	12	177.1
0.69	6	151.3
0.31	3	131.9
0.10	1	124.2

[a]True exit point location is 206.8 m.

3.4. Predictive Error Variance With Parameter Zonation

[57] As described in section 2.6, predictive error variance calculation based on lumped parameterization can be accommodated if equation (26a) incorporates **R** and **G** matrices described by equations (38). This is now demonstrated using parameters based on zones of piecewise uniformity.

[58] Figure 7 shows the zonation pattern chosen for this demonstration; boundaries for three (noncontiguous) zones were drawn against the background of the true hydraulic conductivity field depicted in Figure 1a (a luxury not available in normal modeling practice). Estimation of zonal hydraulic conductivities was undertaken using the same calibration data set as that used for previous analyses. However, weights enhancement was not employed, each head measurement thus being assigned a weight of 3.33.

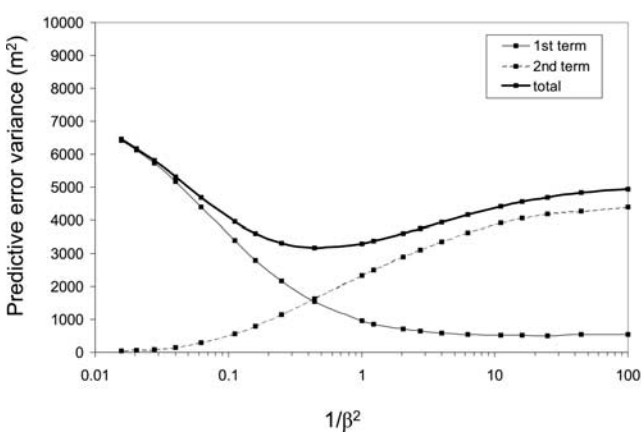

Figure 6. Terms of equation (26) and total model predictive error variance versus $1/\beta^2$. Sensitivities were calculated on the basis of parameters estimated using a β^2 value of 10.9 (i.e., a $1/\beta^2$ value of 0.092).

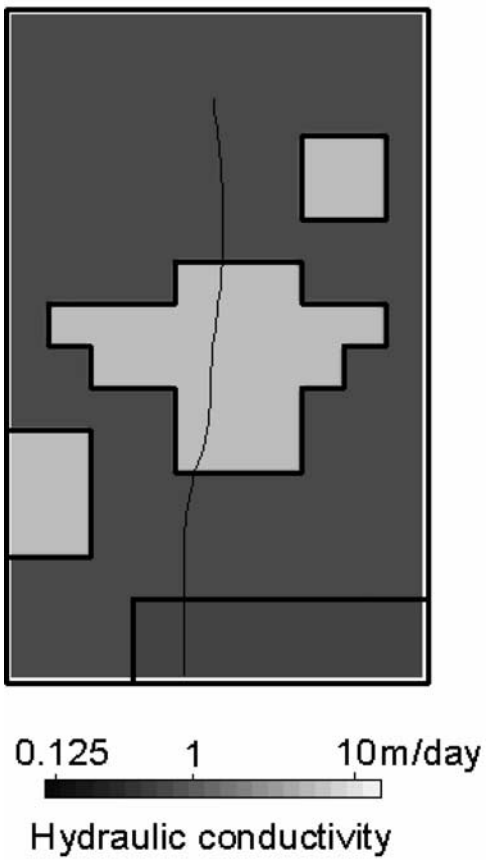

Figure 7. Model domain showing hydraulic conductivity field determined using zones. Predicted particle track line is also shown.

The **X** matrix of equation (36) was calculated on the basis of the pilot points depicted in Figure 1b, with calibrated values assigned to these points according to the zones in which they lie; however, in order to establish true correspondence between pilot point and zone parameters these sensitivities were calculated on the basis of a kriging procedure which prevented interpolation from pilot points in one zone to cells occupying another zone.

[59] Predictive error variance was calculated as 10,189 m^2. The contribution to this variance from the second term of equation (26) is 173 m^2, this being the predictive error variance that would have been calculated using traditional methods that ignore the implicit regularization of the lumping process. The actual model prediction was a surprisingly good 210 m. This level of accuracy is partly attributable to luck and partly attributable to the fact that zonation definition took place against the background of the true hydraulic conductivity field, constituting the addition of a significant amount of extra information to the calibration process. Nevertheless, the above calculation shows that the potential error associated with this model prediction is still quite high.

4. Discussion and Conclusions

[60] The concepts introduced in this paper have a number of important implications for the way in which models are calibrated and the way in which they are used to support the making of decisions in environmental management. Traditionally, model calibration is undertaken through solution of an overdetermined inverse problem based on a simplified representation of reality which involves some form of parameter lumping. Through this process a set of "average" parameter values is sought, with averaging normally taking place in a spatial sense. Weights ascribed to observations are calculated as the inverse of their "uncertainty." Though acknowledgment is often given to the fact that measurement noise is overshadowed by structural noise, the latter is mostly assumed to exhibit no spatial or temporal correlation or (in the case of hydrologic models) to possess a correlation structure that can be eliminated by fitting to an appropriate autoregressive moving average model [see, e.g., *Kuczera*, 1983; *Bates and Campbell*, 2001]. This allows the use of independent weights in place of a full weight matrix which is proportional to the inverse of the assumed measurement covariance matrix as required by traditional parameter estimation theory. Once the calibration process is complete, this single calibrated model is then used to make a variety of different predictions.

[61] This approach to calibration has some serious shortcomings. *Cooley* [2004] demonstrates that the values estimated for lumped parameters can only be interpreted as the outcomes of a user-specified averaging process of pertinent system properties when the measurement weight matrix is proportional to the inverse of the covariance matrix that properly characterizes the correlation structure of the "structural errors" induced by parameter lumping (e.g., by the replacement of a continuous hydraulic property field by a small number of zones of piecewise constancy). However, this can only be known if the true covariance structure of the hydraulic property whose spatial variability is lumped is accurately known. Where an incorrect hydraulic property statistical structure is assumed or where use of a full weight matrix is discarded in favor of independent observation weights, then an altogether different hydraulic property averaging process is implicitly undertaken through lumped parameter model calibration. The relationships between lumped parameters $\hat{\mathbf{r}}$ and true system parameters **p** implied by this averaging process are available through the $\mathbf{R}'$ matrix of equation (36). An examination of this matrix will often reveal that averaging of true parameter values to form lumped parameter values can take place in unintended ways and that where different parameter types are being simultaneously estimated, a certain degree of "parameter contamination" may be present where the averaging process extends across parameter types.

[62] Parameter estimation through regularized inversion also results in parameter value averaging (and with it, parameter contamination), the nature of which is available through postcalibration analysis of the resolution matrix **R**. Thus use of regularized inversion in model calibration has the disadvantage that the modeler is not able to specify in advance of the parameter estimation process the nature of the desired relationship between estimated and true parameters and, through use of an appropriate measurement weight matrix, to ensure that this relationship is respected by that process. However, in our opinion this disadvantage is more than offset by the fact that the solution of the inverse problem through regularized inversion based on a large number of parameters allows the calibration process to

probe as many different dimensions in parameter space as the calibration data set allows. In contrast, the solution of a lumped parameter inverse problem can occupy only as many dimensions in parameter space as there are lumped parameters. As a result the likelihood of a predictive sensitivity vector being orthogonal, or nearly orthogonal, to the subspace of parameter space occupied by the solution to the inverse problem is less when parameter estimation is achieved through regularized inversion than through estimation of lumped parameters. Thus the regularized calibration process is able to reduce the variance of a greater range of predictive types below their precalibration levels by greater amounts because of the greater likelihood of occupancy of at least part of the subspace of parameter space on which these predictions depend by the solution to the inverse problem. In addition to this, better fits can be obtained between model outputs and field data; hence lumping-induced structural noise is greatly reduced. Both of these contribute to increased predictive reliability of the calibrated model.

[63] Where a model is calibrated through regularized inversion, the role of observation weights in the inversion process can differ from their traditional role in the calibration of lumped parameter models. In the regularized inversion context a judicious assignment of observation weights can allow the user to guide this process in ensuring that those dimensions of parameter space that are salient to a particular prediction are indeed represented in the solution of the inverse problem. Furthermore, the analysis presented herein can be used to ensure that if a particular parameter subspace can only be populated with parameter estimates that have more uncertainty (as a result of observation noise) than those which were available through prior knowledge of hydraulic property values, then such calibration-derived estimates can be excluded from the calibrated model. Judicious weights assignment can thus constitute a strategy for ensuring that the model contains enough parameterization detail for the making of a particular prediction without compromising the integrity of that prediction by contaminating it with any more "predictive noise" than is necessary.

[64] The theory presented in this paper can also provide a powerful basis for the making of important decisions regarding appropriate model complexity and the worth of data acquisition in supporting the parameterization of that complexity. ("Complexity" here is loosely equated to the number of parameters.) Proper assessment of these issues can lead to optimal deployment of financial and other resources in data acquisition and processing in support of environmental management. Use of equations (26) and (32) does not require that a model actually be calibrated; it only requires that all parameters that would be estimated through the calibration process be represented in the model. Once sensitivities of pertinent model outputs with respect to these parameters have been calculated, singular value decomposition of $(\mathbf{X}^t\mathbf{Q}\mathbf{X})$ allows estimation of the error variance associated with a particular model prediction. By comparing this variance with that achievable without calibration (calculable through equation (2)) the worth of the calibration process in reducing this variance can be judged. This, in turn, allows an assessment to be made of the utility of the calibration process in increasing the certainty of predictions of future environmental behavior before the calibration exercise is actually undertaken. Furthermore, through graphs such as those appearing in Figures 2, 4, 5, and 6 the level of model complexity that is optimal for the making of a particular prediction, given the nature of that prediction and the level of noise associated with field measurements, can be judged.

[65] What is even more useful, however, is that this analysis can then be repeated with the inclusion of notional additions to the calibration data set or with alterations made to $C(\mathbf{p})$ in order to reflect notional direct measurements of system properties. The reduction in model predictive variance that would be accrued through inclusion of this extra information in the model parameterization process can then be used to assess its worth. Various data acquisition strategies can thereby be compared in terms of their ability to increase model predictive reliability; that which provides the greatest return for investment can then be selected for implementation. Because the characterization of model predictive variance encapsulated in equations (26) and (32) includes the contribution to this variance made by both measurement error and environmental heterogeneity that cannot be captured by the calibration process (i.e., the second and first terms, respectively, of these equations), such a methodology for assessment of data worth is superior to those presented by *Tiedeman et al.* [2003] and *Hill et al.* [2001], who ignore the potentially large contribution that can be made to predictive variance by this latter source. The use of the methodology described herein in optimization of data acquisition will be demonstrated in a future publication.

References

Anderman, E. R., and M. C. Hill (2001), MODFLOW-2000, the U.S. Geological Survey modular ground-water model—Documentation of the advective-transport observations (ADV2) package, *U.S. Geol. Surv. Open File Rep.*, *01-54*.

Backus, G., and F. Gilbert (1969), The resolving power of gross Earth data, *Geophys. J. R. Astron. Soc.*, *16*, 169–205.

Bard, Y. (1974), *Nonlinear Parameter Estimation*, Elsevier, New York.

Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, *37*(4), 937–947.

Christensen, S., and R. L. Cooley (1999), Evaluation of prediction intervals for expressing uncertainties in groundwater flow predictions, *Water Resour. Res.*, *35*(9), 2627–2640.

Clemo, T., P. Michaels, and R. M. Lehman (2003), Transmissivity resolution obtained from the inversion of transient and pseudo-steady drawdown measurements, paper presented at MODFLOW and More 2003: Understanding Through Modeling, Int. Ground Water Model. Cent., Golden, Colo. (Available at http://www.mines.edu/igwmc/.)

Cooley, R. L. (2004), A theory for modeling ground-water flow in heterogeneous media, *U.S. Geol. Surv. Prof. Pap.*, *1679*.

De Groot-Hedlin, C., and S. Constable (1990), Occam inversion to generate smooth, 2-dimensional models from magnetotelluric data, *Geophysics*, *55*(12), 1613–1624.

Doherty, J. (2003), Groundwater model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170–177.

Doherty, J. (2004), PEST: Model-independent parameter estimation, user manual, version 5, Watermark Numer. Comput., Brisbane, Aust.

El Harrouni, K., D. Ouazar, L. C. Wrobel, and A. H.-D. Cheng (1997), Uncertainty analysis in groundwater flow with DRBEM, *Eng. Anal. Boundary Elem.*, *19*, 217–221.

Feyen, L., J. J. Gómez-Hernández, P. J. Ribeiro Jr., K. J. Beven, and F. De Smedt (2003), A Bayesian approach to stochastic capture zone delineation incorporating tracer arrival times, conductivity measurements, and hydraulic head observations, *Water Resour. Res.*, *39*(5), 1126, doi:10.1029/2002WR001544.

Franssen, H. H., J. J. Gómez-Hernández, and A. Sahuquillo (2003), Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data, *J. Hydrol.*, *281*, 281–295.

Frederick, D., and J. Doherty (2003), Regularized inversion using head and concentration measurements, paper presented at MODFLOW and More 2003: Understanding Through Modeling, Int. Ground Water Model. Cent., Golden, Colo. (Available at http://www.mines.edu/igwmc/.)

Gómez-Hernández, J. J., H. J. Hendricks Franssen, and A. Sahuquillo (2003), Stochastic conditional inverse modeling of subsurface mass transport: A brief review of the self-calibrating method, *Stochastic Environ. Res. Risk Assess.*, *17*, 319–328.

Gorokhovski, V. M. (1996), Problem-dependence of ground-water model identification: Significance, extent, and treatment, *Ground Water*, *34*(3), 461–469.

Guadagnini, A., and S. P. Neuman (1999), Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly uniform domains: 1. Theory and computational approach, *Water Resour. Res.*, *35*(10), 2999–3018.

Harbaugh, A. W., E. R. Banta, M. C. Hill, and M. G. McDonald (2000), The U.S. Geological Survey modular ground-water model: User guide to modularization concepts and the ground-water flow process, *U.S. Geol. Surv. Open File Rep.*, *00-92*.

Harvey, C. H., and S. M. Gorelick (1995), Mapping hydraulic conductivity: Sequential conditioning with measurements of solute arrival time, hydraulic head, and local hydraulic conductivity, *Water Resour. Res.*, *31*(7), 1615–1626.

Hill, M. C. (1989), Analysis of accuracy of approximate, simultaneous, nonlinear confidence intervals on hydraulic heads in analytical and numerical cases, *Water Resour. Res.*, *25*(2), 177–190.

Hill, M. C. (1998), Methods and guidelines for effective model calibration, *U.S. Geol. Surv. Water Resour. Invest. Rep*., *98-4005*.

Hill, M. C., M. Ely, C. R. Tiedeman, G. M. O'Brien, F. A. D'Agnese, C. C. and Faunt (2001), Preliminary evaluation of the importance of existing hydraulic-head observation locations to advective-transport predictions, Death Valley regional flow system, California and Nevada, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, *00-4282*.

Kitanidis, P. K. (1997), The minimum structure solution to the inverse problem, *Water Resour. Res.*, *33*(10), 2263–2272.

Kuczera, G. (1983), Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty, *Water Resour. Res.*, *19*(5), 1151–1172.

LaVenue, A. M., B. S. RamaRao, G. de Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated hydraulic conductivity fields: 2. Application, *Water Resour. Res.*, *31*(3), 495–516.

Menke, W. (1984), *Geophysical Data Analysis: Discrete Inverse Theory*, Elsevier, New York.

Poeter, E. P., and W. R. Belcher (1991), Assessment of porous media heterogeneity by inverse plume analysis, *Ground Water*, *29*(1), 56–62.

RamaRao, B. S., A. M. LaVenue, G. de Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated hydraulic conductivity fields: 1. Theory and computational experiments, *Water Resour. Res.*, *31*(3), 475–493.

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 Resour. Res.*, *39*(1), 1010, doi:10.1029/2001WR001255.

Vasco, D. W., A. Datta-Gupta, and J. C. S. Long (1997), Resolution and uncertainty in hydrologic characterization, *Water Resour. Res.*, *33*(3), 379–397.

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 Resour. Res.*, *23*(7), 1237–1250.

Wiggins, R. A., K. L. Larner, and R. Wisecup (1976), Residual statics analysis as a general linear inverse problem, *Geophysics*, *41*(5), 922–938.

Woodbury, A. D., and Y. Rubin (2000), A full-Bayesian approach to parameter inference from tracer travel time moments and investigation of scale effects at the Cape Cod experimental site, *Water Resour. Res.*, *36*(1), 159–171.

Woodbury, A. D., and T. J. Ulrich (2000), A full-Bayesian approach to the groundwater inverse problem for steady state flow, *Water Resour. Res.*, *36*(8), 2081–2093.

J. Doherty and C. Moore, School of Engineering, University of Queensland, Brisbane, Qld 4072, Australia. (jdoherty@gil.com.au; s4024009@student.uq.edu.au)