Efficient nonlinear predictive error variance for highly parameterized models

Matthew Tonkin,[1,2] John Doherty,[3] and Catherine Moore[4]

Received 16 July 2006; revised 9 April 2007; accepted 19 April 2007; published 18 July 2007.

[1] Predictive error variance analysis attempts to determine how wrong predictions made by a calibrated model may be. Predictive error variance analysis is usually undertaken following calibration using a small number of parameters defined through a priori parsimony. In contrast, we introduce a method for investigating the potential error in predictions made by highly parameterized models calibrated using regularized inversion. Vecchia and Cooley (1987) describe a method of predictive error variance analysis that is constrained by calibration data. We extend this approach to include constraints on parameters that lie within the calibration null space. These constraints are determined by dividing parameter space into combinations of parameters for which estimates can be obtained and those for which they cannot. This enables the contribution to predictive error variance from parameterization simplifications required to solve the inverse problem to be quantified, in addition to the contribution from measurement noise. We also describe a novel technique that restricts the analysis to a strategically defined predictive solution subspace, enabling an approximate predictive error variance analysis to be completed efficiently. The method is illustrated using a synthetic and a real-world groundwater flow and transport model.

Citation: Tonkin, M., J. Doherty, and C. Moore (2007), Efficient nonlinear predictive error variance for highly parameterized models, *Water Resour. Res.*, *43*, W07429, doi:10.1029/2006WR005348.

1. Introduction

[2] This paper presents an efficient method for investigating the potential error in predictions made by a calibrated model. On the assumption that accurate simulation of past system response is a necessary (but not sufficient) condition for making reliable predictions, model parameters are calibrated using measurements of system state (observations) and direct or indirect measurements of system properties [*Carrera and Neuman*, 1986a, 1986b; *Carrera et al.*, 2005]. Calibration is an underdetermined and nonunique inverse problem [*Carrera and Neuman*, 1986a, 1986b]. Nonetheless, the existence of a calibrated model is often inferred as meaning that the inverse problem has been solved. Although an infinite number of solutions may exist, a simple solution is usually adopted. Simplification may be undertaken using a priori parsimony, such as a small number of parameter zones, or using regularized inversion [*Tonkin and Doherty*, 2005]. In either case the calibrated parameters do not represent the true system properties; instead, they are a complex average of the true system properties [*Guadagnini and Neuman*, 1999a, 1999b; *Menke*, 1989].

[3] Put simply, calibration can only capture as much of reality as there is information in the available data. Where data are insufficient to fully characterize the system, predictions made with the calibrated model are liable to be wrong [*Moore and Doherty*, 2006]. If the model accurately represents the relevant processes, the error depends on parameter details that are not represented in the model. Under some circumstances, calibration can provide no improved capacity for making predictions: For example, *Moore and Doherty* [2005] demonstrate that transport predictions made by a model that calibrates perfectly to groundwater elevations can be 100% wrong due to parameter simplifications. In recognition of this, research has been undertaken to develop means for evaluating the error of model predictions. The two broad methodological approaches may be categorized as (1) predictive uncertainty analysis and (2) predictive error variance analysis.

[4] Predictive uncertainty analysis acknowledges that many parameter sets enable the model to reproduce the available observations. Methods of predictive uncertainty analysis include generalized likelihood uncertainty estimation (GLUE) [*Beven and Binley*, 1992]; techniques that condition stochastic realizations using property measurements alone [*Deutsch and Journel*, 1992; *Gutjahr et al.*, 1994]; calibration-constrained Monte Carlo, Markov-Chain Monte Carlo (MCMC), and other methods that propagate prior stochastic parameter descriptions through the model to develop posterior parameter and prediction probabilities [*Kitanidis*, 1996; *Yeh et al.*, 1996; *Oliver et al.*, 1997; *Kuczera and Parent*, 1998; *Zimmerman et al.*, 1998; *Woodbury and Ulrych*, 2000; *Carrera et al.*, 2005]; and methods based on stochastic equations that solve directly

[1] S. S. Papadopulos and Associates, Bethesda, Maryland, USA.
[2] Also at Department of Civil Engineering, University of Queensland, Brisbane, Queensland, Australia.
[3] Watermark Numerical Computing, Corinda, Queensland, Australia.
[4] Lincoln Ventures, Ltd., Lincoln University, Christchurch, New Zealand.

0043-1397/07/2006WR005348

for posterior parameter and prediction probabilities [*Rubin and Dagan*, 1987a, 1987b; *Guadagnini and Neuman*, 1999a, 1999b; *Hernandez et al.*, 2006]. Non-Bayesian methods, which deform stochastic parameter fields until a desired objective function is achieved, have been described by *Lavenue and de Marsily* [2001] and *Gomez-Hernandez et al.* [1997, 2003] among others.

[5] Though a wealth of research has taken place into various methods of uncertainty analysis, it is the authors' experience that common-practice groundwater modeling does not reflect this accumulated knowledge and expertise. That is, outside of research or well-funded applications, modeling is usually conducted within a fairly simple decision-based framework: Rigorous uncertainty analyses are rarely undertaken, and those who make decisions on the basis of models are often, for better or for worse, presented with a single model rather than an ensemble.

[6] Predictive error variance analysis attempts to answer the following question: Given that the inverse problem was solved to obtain a single (simplified) parameterization, how wrong could predictions made using the model be? Traditional linear [*Draper and Smith*, 1981] and nonlinear [*Vecchia and Cooley*, 1987; *Christensen and Cooley*, 1999] error variance analyses are used with a small number of parameters that are defined prior to calibration to guarantee a unique solution. Nonlinear predictive error analysis is undertaken by maximizing and minimizing a prediction while constraining parameter variability so that model-to-measurement misfit corresponds with measurement noise. Using a priori parsimony assumes that reality is only as complicated as the model, and that measurement noise has been characterized accurately. *Cooley* [2004] and *Cooley and Christensen* [2006] recognize that parameter simplifications contribute to predictive error, and quantify the contribution of parsimony to measurement noise for incorporation in traditional error analyses.

[7] This paper describes a new method of predictive error variance analysis. In contrast to the work of *Vecchia and Cooley* [1987], among others, we use regularized inversion for model calibration. Regularized inversion also achieves uniqueness through parameter simplification. However, simplification is undertaken formally and can be as little as necessary to solve the inverse problem uniquely [*Moore and Doherty*, 2006]. Regularized inversion is used widely for data interpretation [e.g., *Engl et al.*, 1996; *Haber et al.*, 2000; *Vogel*, 2002], and has seen increasing application in groundwater model calibration. Conceptually, there is no limit to the number of parameters that can be used in the parameter estimation process: hence the scale of simulated parameter variability could theoretically approach that of the real world. The ability to include greater parameter variability than traditional parsimonious analyses may lead to more accurate estimates of predictive error variance [*Moore and Doherty*, 2005]; help to explore the range of values that the prediction may take; and identify mechanisms leading to extreme values of the prediction.

[8] The computational burden of calibrating highly parameterized models may be reduced through the estimation of "superparameters" [*Tonkin and Doherty*, 2005]. The concept of superparameters rests on the premise that parameter space can be subdivided into parameter combinations that are estimable through calibration (the calibration solution space) and those that are not (the calibration null space). Using this approach, errors in solution space parameter combinations arise from noise in the measured data, whereas errors in null space parameter combinations are solely due to innate parameter variability, since calibration cannot constrain these parameter combinations. However, traditional predictive error variance analysis cannot be applied directly to a model calibrated using regularized inversion since it does not enforce constraints on these null space parameter combinations, despite the fact that the prediction may be sensitive to them. As a result, unlikely predictions can be produced by varying null space parameter combinations by unrealistic amounts. A second difficulty is that calculating derivatives for many parameters incurs a large computational burden, though in rare circumstances this burden can be reduced through the use of adjoint methods.

[9] We address these issues by reformulating the method of *Vecchia and Cooley* [1987] to analyze predictive error in highly parameterized models calibrated using regularized inversion. This is accomplished by (1) extending the theory presented by *Moore and Doherty* [2005] to define constraints on parameter combinations that lie within the calibration null space, penalizing variability in parameters that are not constrained by available data; and (2) extending the superparameter concept to define a low-dimensional predictive solution subspace comprising parameter combinations to which the prediction is (ideally, entirely) sensitive. This enables nonlinear predictive error variance analysis of highly parameterized models to be completed efficiently following a hybrid truncated singular value decomposition (TSVD)-Tikhonov calibration [*Tonkin and Doherty*, 2005]. To our knowledge, this is the first presentation of a method for undertaking nonlinear predictive error variance analysis in the regularized inversion context.

[10] This paper is presented as follows. First, the theory is presented for quantifying predictive error variance following regularized inversion. Equations are presented that describe (1) how the calibration is reformulated from a traditional overdetermined problem into a regularized inversion; (2) how predictive error variance is calculated following calibration using subspace-based techniques; (3) the method of *Vecchia and Cooley* [1987]; and (4) how constraints on null space parameter combinations are developed for enforcement during nonlinear predictive analysis. Following these equations, the novel method of surrogate predictions is described, which may be used to develop a predictive solution subspace. The possible impacts of assumptions on the performance of the method are then discussed. The new method is then demonstrated by calculating upper confidence limits for predictions simulated by a synthetic and a real-world model. The results of these analyses are discussed together with concluding remarks.

2. Theory

2.1. Parameter Error

[11] Suppose that a model is parameterized with a spatial and/or temporal density that reflects true system heterogeneity. Let $\mathbf{p}$ represent the true parameters, and let $\overline{\mathbf{p}}$ represent

the estimated parameters. Typically $\bar{\mathbf{p}}$ contains more elements than can be estimated uniquely, and regularization is used in the inversion process. For a linear model, $\bar{\mathbf{p}}$ is calculated from measurements $\mathbf{h}$ using

$$\bar{\mathbf{p}} = \mathbf{G}\mathbf{h}, \tag{1}$$

where $\mathbf{G}$ depends on the regularization method employed. For nonlinear models, $\mathbf{G}$ changes as parameters are updated throughout the iterative solution process. If the system response is linear its action can be represented by a matrix $\mathbf{X}$ so that

$$\mathbf{h} = \mathbf{X}\mathbf{p} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{X}$ contains the sensitivity of each simulated equivalent of each observation with respect to each parameter and $\boldsymbol{\varepsilon}$ represents measurement noise. Regularization causes $\bar{\mathbf{p}}$ to be a smoothed representation of $\mathbf{p}$. That is, each element of $\bar{\mathbf{p}}$ is a linear combination of the true parameters [*Menke*, 1989]. Parameter error, $\mathbf{p} - \bar{\mathbf{p}}$, incurred through calibration is expressed by [*Moore and Doherty*, 2005]

$$\mathbf{p} - \bar{\mathbf{p}} = (\mathbf{I} - \mathbf{R})\mathbf{p} - \mathbf{G}\boldsymbol{\varepsilon}, \tag{3}$$

where $\mathbf{R}$ is the resolution matrix. Expressions for $\mathbf{R}$ and $\mathbf{G}$ for different regularization techniques that are relevant to this study are

Tikhonov regularization [*Tikhonov and Arsenin*, 1977]

$$\mathbf{R} = \left(\mathbf{X}^t\mathbf{Q}\mathbf{X} + \beta^2\mathbf{T}^t\mathbf{S}\mathbf{T}\right)^{-1}\mathbf{X}^t\mathbf{Q}\mathbf{X} \tag{4a}$$

$$\mathbf{G} = \left(\mathbf{X}^t\mathbf{Q}\mathbf{X} + \beta^2\mathbf{T}^t\mathbf{S}\mathbf{T}\right)^{-1}\mathbf{X}^t\mathbf{Q} \tag{4b}$$

Truncated singular value decomposition (TSVD) [*Aster et al.*, 2005]

$$\mathbf{R} = \mathbf{V}_1\mathbf{V}_1^t \tag{5a}$$

$$\mathbf{G} = \mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q} \tag{5b}$$

Hybrid TSVD-Tikhonov regularization

$$\mathbf{R} = \mathbf{V}_1\left(\mathbf{Z}^t\mathbf{Q}\mathbf{Z} + \beta^2\mathbf{T}^t\mathbf{S}\mathbf{T}\right)^{-1}\mathbf{Z}^t\mathbf{Q}\mathbf{X} \tag{6a}$$

$$\mathbf{G} = \mathbf{V}_1\left(\mathbf{Z}^t\mathbf{Q}\mathbf{Z} + \beta^2\mathbf{T}^t\mathbf{S}\mathbf{T}\right)^{-1}\mathbf{Z}^t\mathbf{Q}, \tag{6b}$$

where

- $\mathbf{Q}$ observation weight matrix;
- $\mathbf{I}$ identity matrix;
- $\mathbf{T}$ vector of Tikhonov regularization constraints on parameters;
- $\mathbf{S}$ regularization weight matrix;
- β^2 regularization weight factor;
- $\mathbf{Z}$ sensitivity matrix of model outputs with respect to superparameters used in the hybrid TSVD-Tikhonov inversion [*Tonkin and Doherty*, 2005];

and $\mathbf{V}_1$ and $\mathbf{E}_1$ are obtained through singular value decomposition (SVD) of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ so that

$$\mathbf{X}^t\mathbf{Q}\mathbf{X} = [\mathbf{V}_1\mathbf{V}_2]\begin{bmatrix}\mathbf{E}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{E}_2\end{bmatrix}[\mathbf{V}_1\mathbf{V}_2]^t, \tag{7}$$

where

- $\mathbf{V}_1$ matrix of orthogonal pretruncation unit eigenvectors;
- $\mathbf{V}_2$ matrix of orthogonal post-truncation unit eigenvectors;
- $\mathbf{E}_1$ diagonal matrix of eigenvalues corresponding to $\mathbf{V}_1$;
- $\mathbf{E}_2$ diagonal matrix of eigenvalues corresponding to $\mathbf{V}_2$.

[12] The regularization weight factor, β^2, is recalculated each iteration of the inversion, in an effort to achieve a user-specified level of model-to-measurement misfit that ideally reflects measurement noise. In the absence of Tikhonov regularization the terms on the right-hand side of (3) are orthogonal. The first term expresses the contribution to parameter error from the calibration null space, arising from combinations of parameters to which model outputs are insensitive. The second term accounts for the fact that estimates of parameter combinations occupying the calibration solution space are contaminated by measurement noise. If Tikhonov regularization is employed in the inversion, this partitioning is approximate.

2.2. Parameter and Predictive Error Variance

[13] Let $\mathrm{C}(\mathbf{p})$ be the covariance matrix of the true system properties $\mathbf{p}$, and let $\mathrm{C}(\boldsymbol{\varepsilon})$ represent the covariance matrix of measurement noise. $\mathrm{C}(\mathbf{p})$ encapsulates precalibration knowledge of the values that a property may assume, and the correlation that may exist between these values. $\mathrm{C}(\mathbf{p})$ may be estimated using geostatistics or related methods. For a linear model the covariance matrix of parameter error can be calculated from (3) as

$$\mathrm{C}(\mathbf{p} - \bar{\mathbf{p}}) = (\mathbf{I} - \mathbf{R})\mathrm{C}(\mathbf{p})(\mathbf{I} - \mathbf{R})^t + \mathbf{G}\mathbf{C}(\boldsymbol{\varepsilon})\mathbf{G}^t. \tag{8}$$

[14] The error variance of a model prediction $\bar{s}$ for which the true value is s, and whose sensitivity to the vector of parameters $\bar{\mathbf{p}}$ is expressed by the vector $\mathbf{y}$, can be calculated using [*Moore and Doherty*, 2005]

$$\sigma^2_{s-\bar{s}} = \mathbf{y}^t(\mathbf{I} - \mathbf{R})\mathrm{C}(\mathbf{p})(\mathbf{I} - \mathbf{R})^t\mathbf{y} + \mathbf{y}^t\mathbf{G}\mathrm{C}(\boldsymbol{\varepsilon})\mathbf{G}^t\mathbf{y}, \tag{9}$$

where $\sigma^2_{s-\bar{s}}$ is the prediction error variance. The first term on the right-hand of (9) represents the contribution to predictive error variance that arises from the inability of the calibration to capture detail on which s depends. The second term represents the contribution to predictive error variance arising from the contamination by measurement noise of parameter combinations that can be inferred through calibration. Unfortunately, if a model is even moderately nonlinear, (9) can lead to poor estimates of predictive error variance. Another drawback of (9) is that it does not identify the parameter values that give rise to a prediction at a certain confidence level. This is particularly unfortunate since modelers are often interested in not just the value of a prediction at a certain confidence level, but also in the mechanism leading to that value.

2.3. Predictive Error Variance Analysis Through Constrained Optimization

[15] Confidence intervals on model predictions are defined by two confidence limits, i.e., the minimum (lower limit) and maximum (upper limit). In linear models, intervals can be computed easily and the point estimates (i.e., means) lie at the center of the computed intervals. In nonlinear models the intervals are asymmetric, i.e., the point estimates do not lie at the center of the intervals, and calculating the lower and upper limits is more intensive. *Vecchia and Cooley* [1987], *Christensen and Cooley* [1999], and *Cooley and Christensen* [2006] formulate this search for a predictive extreme as a constrained optimization in which the prediction is maximized or minimized while ensuring that model-to-measurement misfit is consistent with measurement noise.

[16] The *Vecchia and Cooley* [1987] method is applicable in the overdetermined context where parsimony eliminates the calibration null space. Under these circumstances, $\mathbf{R} = \mathbf{I}$ and the first terms of equations (8) and (9) are zero. Hence parameter and predictive likelihood depend only on model output likelihood under calibration conditions: The closer that model outputs match field measurements, the greater the likelihood of the corresponding parameters. As a prediction is maximized or minimized, the objective function rises until a limiting value is reached that indicates parameter values are unlikely at that confidence level. The prediction is unlikely at the same confidence level. If predictive intervals are used instead of confidence intervals, predictive noise is included in the objective function.

[17] Let Φ_0 represent the objective function constraint on parameters and predictions as these are maximized or minimized. Its value depends on the confidence interval sought and on whether individual or simultaneous confidence intervals are calculated [*Cooley*, 2004]. *Vecchia and Cooley* [1987] show that parameters $\bar{\mathbf{p}}$ corresponding to the maximized or minimized prediction $\bar{s}$ constrained by this objective function can be calculated using

$$\bar{\mathbf{p}} = (\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\left\{\mathbf{X}^t\mathbf{Q}\mathbf{h} - \frac{\mathbf{y}}{2\lambda}\right\}, \tag{10}$$

where

$$\left(\frac{1}{2\lambda}\right)^2 = \pm\frac{\Phi_0 - \mathbf{h}^t\mathbf{Q}\mathbf{h} + \mathbf{h}^t\mathbf{Q}\mathbf{X}(\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Q}\mathbf{h}}{\mathbf{y}^t(\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\mathbf{y}}. \tag{11}$$

[18] The positive or negative solution to (11) is chosen depending on whether the maximum or minimum is sought. If Φ_0 includes predictive error, (10) is still employed but the term $(w_e e)^2$ is incorporated where w_e is the weight assigned to predictive noise and e is the predictive error. Equation (11) then becomes

$$\left(\frac{1}{2\lambda}\right)^2 = \pm\frac{\Phi_0 - \mathbf{h}^t\mathbf{Q}\mathbf{h} + \mathbf{h}^t\mathbf{Q}\mathbf{X}(\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Q}\mathbf{h}}{\mathbf{y}^t(\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\mathbf{y} + w_e^{-2}}, \tag{12a}$$

while e is calculated as

$$e = -w_e^{-2}/2\lambda. \tag{12b}$$

[19] For nonlinear models these equations are solved iteratively with $\bar{\mathbf{p}}-\bar{\mathbf{p}}_0$ replacing $\bar{\mathbf{p}}$, where $\bar{\mathbf{p}}_0$ represents parameters calculated during the previous iteration; $\bar{\mathbf{o}}_0$ listing model outputs corresponding to $\bar{\mathbf{p}}_0$; and $\mathbf{h}$ replaced by the residual vector $\mathbf{r}$ so that

$$\mathbf{r} = \mathbf{h} - \bar{\mathbf{o}}_0. \tag{13}$$

2.4. Regularized Inversion Constraints

[20] Computing predictive error variance in the regularized inversion context requires an objective function that includes constraints imposed on parameters by the calibration data, and constraints imposed on parameters in accordance with their (estimated) innate variability. TSVD divides parameter space into orthogonal null and solution subspaces, the former spanned by $(\mathbf{I} - \mathbf{R})$ and the latter spanned by $\mathbf{G}$. Since constraints on parameter combinations occupying these subspaces are orthogonal they can be applied independently. In the solution subspace, calibration constraints are based on the stochastic character of $C(\varepsilon)$ in a manner similar to those enforced in the overdetermined case. In the null subspace, innate variability constraints are based on the stochastic character of the parameters, described by $C(\mathbf{p})$. If calibration is implemented using TSVD, then according to (5) $\mathbf{R}$ is equal to $\mathbf{V}_1\mathbf{V}_1^t$, and since $\mathbf{V}_1\mathbf{V}_1^t + \mathbf{V}_2\mathbf{V}_2^t = \mathbf{I}$, the null space is described by $\mathbf{V}_2\mathbf{V}_2^t$. Therefore equation (3) becomes [*Moore and Doherty*, 2005]

$$\mathbf{p} - \bar{\mathbf{p}} = \mathbf{V}_2\mathbf{V}_2^t\mathbf{p} + \mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q}\mathbf{h}. \tag{14}$$

[21] Let $\mathbf{m}$ be the projection of $\mathbf{p} - \bar{\mathbf{p}}$ onto the calibration null space. Thus

$$\mathbf{m} = \mathbf{V}_2\mathbf{V}_2^t(\mathbf{p} - \bar{\mathbf{p}}). \tag{15a}$$

[22] Because $\mathbf{V}_1$ and $\mathbf{V}_2$ are orthogonal, $\mathbf{V}_1\mathbf{V}_2 = \mathbf{0}$ and $\mathbf{V}_2\mathbf{V}_1 = \mathbf{0}$. Therefore premultiplication of (14) by respective matrices can be shown to give the following relationships:

$$\mathbf{m} = \mathbf{V}_2\mathbf{V}_2^t\mathbf{p}. \tag{15b}$$

[23] Thus

$$C(\mathbf{m}) = \mathbf{V}_2\mathbf{V}_2^tC(\mathbf{p})\mathbf{V}_2\mathbf{V}_2^t, \tag{16}$$

where $C(\mathbf{m})$, the covariance matrix of $\mathbf{m}$, expresses the stochastic character of parameter combinations occupying the calibration null space. Let $\mathbf{n}$ be the projection of $\mathbf{p} - \bar{\mathbf{p}}$ onto the calibration solution space so that

$$\mathbf{n} = \mathbf{V}_1\mathbf{V}_1^t(\mathbf{p} - \bar{\mathbf{p}}), \tag{17a}$$

which from (14) becomes

$$\mathbf{n} = \mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q}\mathbf{h}. \tag{17b}$$

[24] Thus

$$C(\mathbf{n}) = \mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t\mathbf{X}^t\mathbf{Q}\mathbf{C}(\varepsilon)\mathbf{Q}\mathbf{X}\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1^t. \tag{18}$$

[25] Use of (18) to constrain solution subspace parameter combinations in the nonlinear context is cumbersome and inaccurate, since $\mathbf{X}$ cannot be used in place of the true nonlinear model to evaluate changes in model outputs resulting from changes in parameters. However, if the true nonlinear model is employed, the likelihood of changed model outputs can be assessed using $C(\varepsilon)$, and constraints can be imposed on this basis.

[26] To accomplish this, let $\bar{\mathbf{o}}$ represent model outputs corresponding to calibrated parameters $\bar{\mathbf{p}}$. Parameters $\bar{\mathbf{p}}$ that give rise to model outputs $\bar{\mathbf{o}}$ are unlikely if $\bar{\mathbf{o}}$ differs significantly from $\bar{\mathbf{o}}$ at a specified confidence level that is assessed in terms of $C(\varepsilon)$. In other words, if $\bar{\mathbf{p}}$ differs from calibrated parameters $\bar{\mathbf{p}}$ in such a way that $\bar{\mathbf{o}}$ differs from $\bar{\mathbf{o}}$ by an amount that is greater than can be explained by measurement noise, then parameters $\bar{\mathbf{p}}$ are unlikely. This assessment is exact if $\mathbf{G}$ of (1) is invariant with $\mathbf{p}$; but for nonlinear models $\mathbf{G}$ does change as $\mathbf{p}$ changes. Nonetheless, *Christensen and Cooley* [1999] employ a similar approximation when computing confidence intervals in the overdetermined context and show that this approximation does not incur large errors.

[27] An objective function constraint or target objective function, Φ_0, must be formulated for use in equations (10)–(12) in the regularized inversion context. Define the vector $\mathbf{j}$ as

$$\mathbf{j} = \begin{bmatrix} \mathbf{V}_2\mathbf{V}_2^t(\bar{\mathbf{p}} - \bar{\bar{\mathbf{p}}}) \\ (\bar{\mathbf{o}} - \bar{\bar{\mathbf{o}}}) \end{bmatrix}. \tag{19a}$$

[28] The first part of $\mathbf{j}$ projects differences between the current and calibrated parameters onto the calibration null space. The second part of $\mathbf{j}$ comprises differences in current model outputs from those that correspond to the calibrated parameters. Assuming that $C(\varepsilon)$ is independent of $C(\mathbf{p})$, the covariance matrix of $\mathbf{j}$ is given by

$$C(\mathbf{j}) \begin{bmatrix} \mathbf{V}_2\mathbf{V}_2^t C(\mathbf{p}) \mathbf{V}_2\mathbf{V}_2^t & \mathbf{0} \\ \mathbf{0} & C(\varepsilon) \end{bmatrix}. \tag{19b}$$

[29] If $C(\mathbf{p})$ and $C(\varepsilon)$ have multi-Gaussian distributions, then an objective function, Φ, can be formulated for the current parameters $\bar{\mathbf{p}}$ in order to evaluate the likelihood of various realizations of $\mathbf{j}$:

$$\Phi = \mathbf{j}^t C^{-1}(\mathbf{j})\mathbf{j}. \tag{20}$$

[30] If $\mathbf{h}$ is replaced by $\mathbf{j}$ in equations (10)–(12), it is Φ as calculated using (20) that is constrained by Φ_0 in equations (10)–(12). In this reformulation, $\mathbf{X}$ contains the sensitivities of the elements of $\mathbf{j}$ to parameters $\bar{\mathbf{p}}$. Because the number of elements of $\mathbf{j}$ exceeds the number of elements of $\bar{\mathbf{p}}$, $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ is invertible. As $\bar{\mathbf{p}}$ varies through the prediction maximization or minimization, Φ of (20) can be compared with the square of the normal variate to assess confidence. For example, if Φ_0 of (11) and (12) is set to 9.0, this corresponds to a confidence limit equal to three standard deviations of the normal variate, or about 99.7% if a two-sided confidence interval is sought through successive maximization and minimization of the prediction of interest.

[31] If pure TSVD is used to calibrate nonlinear models, $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ is decomposed each iteration of the inverse process, and hence the null and solution subspaces alter each iteration. Tikhonov regularization does not subdivide parameter space into orthogonal subspaces, although if measurement noise is small and the target objective function is low, an approximate subdivision occurs. When the TSVD-Tikhonov hybrid [*Tonkin and Doherty*, 2005] is used, the solution and null subspaces are defined at the commencement of the parameter estimation and do not change as parameter values change. Therefore, at the completion of the TSVD-Tikhinov hybrid calibration, calculation of $\mathbf{R}$ and $(\mathbf{I} - \mathbf{R})$ produces an approximate subdivision into orthogonal solution and null spaces.

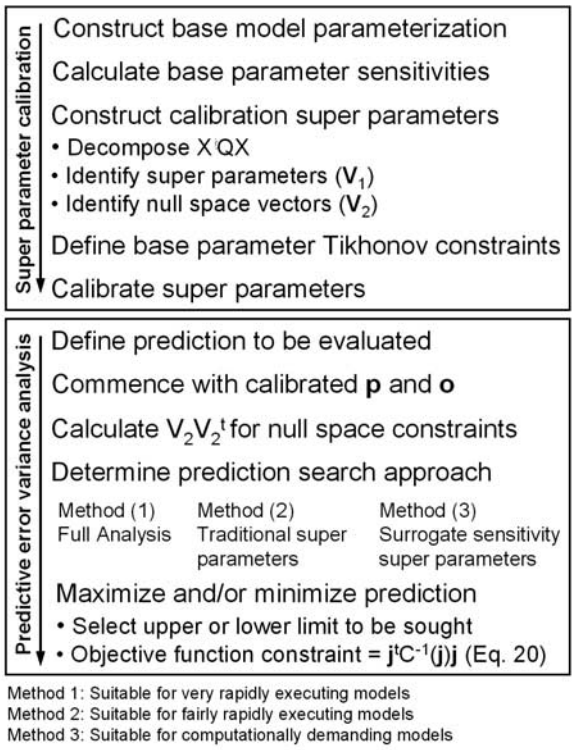

Figure 1. Schematic of the regularized inversion calibration and predictive analysis process.

2.5. Increasing Efficiency by Defining a Predictive Solution Subspace

[32] For a linear model the predictive solution space comprises one dimension, vector $\mathbf{y}$ of (9), and computing confidence intervals is efficient. If adjoint sensitivities are available, one model run suffices to compute $\mathbf{y}$. If adjoint methods are unavailable, computing $\mathbf{y}$ requires as many model runs as there are parameters. For a nonlinear model, $\mathbf{y}$ is computed each iteration of the predictive error analysis. For highly parameterized models, this burden usually prevents predictive analysis from being undertaken. Three parameterization methods can be employed for predictive error variance analysis with highly parameterized nonlinear models (Figure 1): method 1, using all available (base) parameters; method 2, using predictive superparameters constructed in a manner similar to that described by *Tonkin and Doherty* [2005]; or method 3, defining a predictive solution subspace of small dimensions by inspecting the problem at hand. To the authors' knowledge, none of these approaches has been demonstrated previously.

[33] Method 1 is the most computationally intensive, since sensitivities must be calculated for all base parameters each iteration. Method 1 is employed in the synthetic example described later, where it forms a "base case" for evaluation of the potential benefits of method 3. Method 2 is based directly upon the method of *Tonkin and Doherty* [2005], who show that estimating superparameters can increase model calibration efficiency. A similar strategy could be employed in the predictive error variance analysis, but with solution space and null space constraints applied using the methods described above. Solving (10) through (12) for these superparameters instead of for all base parameters could increase model run efficiency without compromising the inference possible from the analysis.

[34] Method 3 is described and evaluated in the examples below. Here the predictive solution space is defined using sensitivities computed for "surrogate predictions" that are computed more rapidly than the predictions themselves. A suitable source of surrogate sensitivities may be determined by inspection of the problem at hand; this is described further in the example applications. Once the surrogate sensitivities have been obtained, TSVD can be used to obtain orthogonal unit vectors to define a predictive solution subspace. If these vectors span enough of parameter space to define parameter combinations that are most influential to the surrogate predictions, they may be effective for investigating the actual prediction of interest. Scalar factors assigned to these vectors can be used as predictive superparameters in a nonlinear predictive error analysis, whose dimensionality is restricted to that which is most salient to the prediction.

[35] It must be clearly stated that when using method 3 the "true" model prediction is still used in the predictive error variance analysis. The analysis is simply restricted to a predictive subspace that is defined using linear combinations of the "true" model parameters. Surrogate predictions are used solely to define these parameter combinations and hence the predictive subspace. The limiting objective function of (20) is unchanged, so that constraints on model-to-measurement misfit and base parameter likelihood are the same as if method 1 or method 2 was used. Hopefully, the predictive subspace spans most (if not the entirety) of parameter space to which the prediction is sensitive. To the extent that the "true" prediction is sensitive to parameter combinations that are omitted from the analysis in order to make it numerically tractable, then the method may incur errors. Nonetheless, the approximation facilitates an otherwise intractable analysis, and whether the predictive subspace minimizes or maximizes the prediction to the extent necessary to support decision-making must be assessed following the analysis.

2.6. Approximations

[36] Some approximations involved in this nonlinear predictive error variance analysis procedure described above are now summarized.

[37] First, when regularized inversion is undertaken using subspace methods, the definition of the calibration null space is only approximate, since parameters $\bar{\mathbf{p}}$ deviate from calibrated values $\hat{\mathbf{p}}$ as the predictive variance analysis progresses. Hence the null space constraints that are based on $C(\mathbf{p})$ become inexact.

[38] Second, the application of constraints on $\hat{\mathbf{o}}-\bar{\mathbf{o}}$ at a particular confidence interval calculated on the basis of $C(\varepsilon)$ is only exact if the relationship between model parameters and model outputs is linear, i.e., if $\mathbf{G}$ is independent of $\mathbf{p}$. This is not a great disadvantage since the method results in a model that, following prediction maximization or minimization, is only "uncalibrated" to a level compatible with $C(\varepsilon)$ even if the relationship between the extent of "decalibration" and predictive confidence is compromised by nonlinearity.

[39] Third, the use of a limited dimensionality predictive search subspace may prevent the analysis from maximizing or minimizing a prediction to the full extent of its confidence range. This must be resolved in context by the modeler. Where run times are long and efficiency is vital, the cost of this approximation may be offset by the opportunity to investigate the approximate range of values that a prediction may take in a manner that is otherwise impracticable.

[40] Finally, Φ_0 is calculated using the square of a normal variate. However, since there is uncertainty associated with the magnitude of the multiplicative factor for $C(\varepsilon)$, i.e., the regression error variance, the target objective function might be calculated using the square of a Student's t distribution (assuming measurement noise is normally distributed). This would lead to wider confidence limits than would a normal variate, because uncertainty in $C(\varepsilon)$ is acknowledged at the same time that $C(\varepsilon)$ is employed as a parameter constraint. However, use of a Student's t variate may not be strictly correct, due to insights gained through regularized inversion [*Moore*, 2006]. Nevertheless, caution may dictate use of a Student's t to mitigate underestimation of predictive error variance.

[41] These approximations affect the accuracy of calculated confidence limits and intervals. It should be noted, however, that neither uncertainty analysis nor predictive error variance analysis provides completely rigorous confidence intervals since not all possibilities for predictive variability are explored. For example, the bulk of measurement noise may be structural noise induced by model inadequacy and parameter simplifications [*Cooley and Christensen*, 2006]. Even on occasions that rigorous geostatistical analyses have been completed, the structure of $C(\mathbf{p})$ is not known precisely. Nonetheless, these techniques often enable approximate analyses that can form a basis for decisions that take into account the possibility of error in model predictions. Furthermore, when the methods described in this study are applied, the possibility of predictions being in error, and the mechanism through which such errors may arise, can be investigated at relatively small numerical cost while respecting enforced constraints. This is now demonstrated.

3. Background to Applications

[42] The theory is first applied to a synthetic model that represents the major features of a real-world site. It is then applied to a real-world model, developed to evaluate the risk posed to a production well by a contaminant release. In each case, the error variance analysis is used only to calculate the upper confidence limit (i.e., the upper confidence interval) since the interest is in the maximum value of

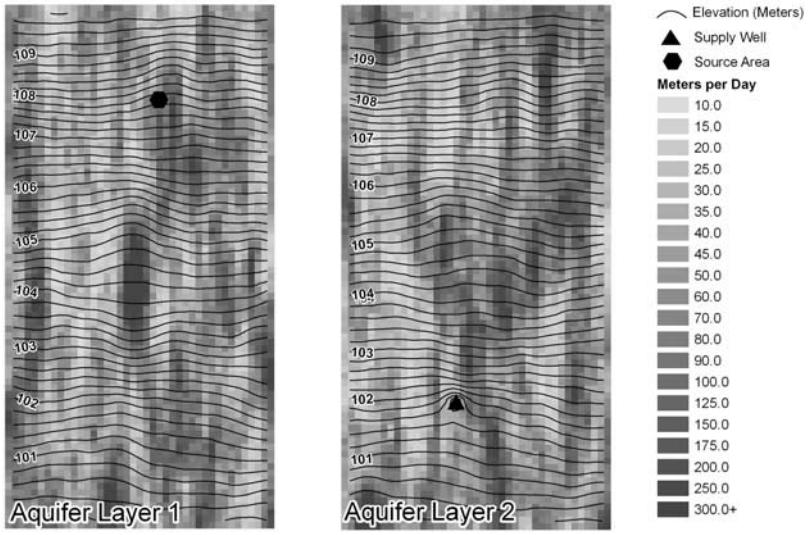

Figure 2. Synthetic model domain with true parameters and heads.

the prediction. This is consistent with applications that seek to determine "what is the worst that could happen, given current knowledge."

[43] Water quality in unconfined aquifers is threatened by contaminants released at the water table. The threat posed to deep production wells is uncertain because (1) it is cost-prohibitive to exhaustively characterize confining units between aquifers; (2) investigations focus on the perceived extent of the plume, rather than in the vicinity of production wells; and (3) it is difficult to characterize the effects of deep pumping on shallow groundwater. Additional uncertainties surround the source term. The modeler is often faced with a shallow plume migrating toward a production well that is separated vertically by tens or hundreds of meters of aquifer that may or may not include a contiguous confining layer. The potential value of coupled flow-and-transport model calibration in such circumstances is described by *Hendricks Franssen et al.* [2003] among others. The question posed here is, Could contaminants released to the shallow aquifer contaminate water recovered from the deep aquifer?

3.1. Synthetic Test Case

[44] The synthetic flow system is characterized by two aquifers separated by a semiconfining unit. The system is simulated using MODFLOW-2000 [*Harbaugh et al.*, 2000], using two layers separated by a confining bed. Hydraulic conductivity in the aquifers and aquitard are generated using the sequential field generator SGSIM [*Deutsch and Journel*, 1992] using exponential variograms with a range of 150 m and sill of 0.15 for the aquifers, and a range of 200 m and sill of 0.3 for the aquitard. Vertical anisotropy within the aquifers is set at 1:10. Steady state constant-transmissivity flow is simulated, using general head boundaries up and down gradient, and no-flow lateral boundaries. A production well is located in the deep aquifer (layer 2). A known time-varying contaminant source is simulated in the shallow aquifer upgradient and lateral to the production well. Contaminant transport is simulated for 4000 days using MT3DMS with the explicit total variation diminishing (TVD) solution scheme to preserve sharp fronts [*Zheng and Wang*, 1999].

3.1.1. Observation Data

[45] Figure 2 depicts the "true" multi-Gaussian hydraulic conductivities together with heads simulated using these conductivities. Of the many parameter fields generated, these were selected as reality because predictions based on the calibrated model are benign, though in the "true" case the well is contaminated. This presents a challenge to the predictive error variance analysis since, notwithstanding incorrect predictions made by the calibrated model, the right answer should lie within predictive error intervals. Figure 3 depicts the simulated plume after 800 days of transport. The model calculates a peak concentration at the production well of 27.5 μg/L, 1600 days following the contaminant release (t = 1600). Note that in both the synthetic application and the real-world application that follows, the timing of this peak is not included in the predictive analysis. Therefore the temporal coincidence of simulated peaks is not an assessment criterion. It is recorded because it is salient to the computational effort involved in the predictive error analyses.

[46] Heads and concentrations simulated by the model using the true parameters were used to generate data at 130 hypothetical monitoring wells, 65 wells in each model layer (Figure 4). In keeping with real-world practice, monitoring wells focus on the current extent of the plume and do not extend to the production well. Paired screens in the shallow and deep aquifers provide colocated water level and concentration data. Gaussian noise was added to the logs of simulated concentrations to generate concentration observation data; the standard deviation of this noise is equal to 15% of the mean of the natural logs of the concentrations. A hypothetical analytical method-reporting limit of 0.1 μg/L is mimicked in the regression. Gaussian noise with a standard deviation of 0.05 m was added to simulated heads. The calibration data was composed of one set (130) of head observations and three sets (390) of concentration observations, the latter mimicking sequential groundwater sampling events at t = 500, 650, and 800 days.

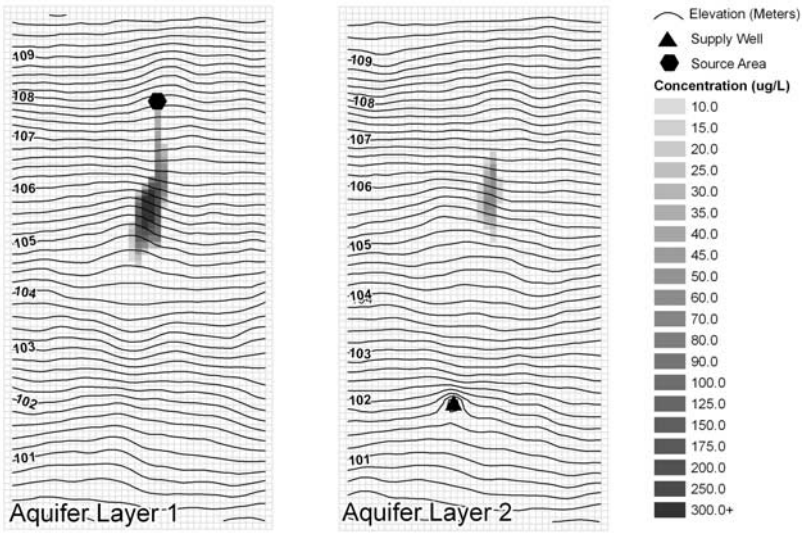

Figure 3. Contaminant plume in true model at time = 800 days.

3.1.2. Parameterization and Calibration

[47] The hydraulic conductivity of the system is parameterized using 792 pilot points [*Certes and de Marsily*, 1991; *Doherty*, 2003], 264 for each aquifer and for the aquitard, distributed evenly throughout the model domain (Figure 4). These pilot points constituted the base parameters of the superparameter scheme used to calibrate the model.

[48] Initial values for the pilot points were provided as the (known) mean of the corresponding aquifer hydraulic conductivity. When the model is run using these parameter values, a peak concentration of about 3 μg/L occurs at the well about 4000 days following the release. During calibration the model was executed for sufficient time to obtain simulated equivalents to the observations (i.e., at t = 800 days) requiring about one fifth the execution time required to simulate the peak concentration at the well. Base parameter sensitivities were calculated using these uniform parameter values using perturbations across a 20-node CPU network, providing matrix **X** with dimensions of 792 columns by 520 rows. **Q** was specified as a diagonal matrix with elements equal to the reciprocal of respective measurement variances. Thirty superparameters were constructed through TSVD of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$. Calibration of the superparameters was completed on the same CPU network.

[49] Calibrated aquifer hydraulic conductivities are illustrated in Figure 5. Figure 6 compares the natural logarithm of the simulated and observed concentrations. Concentrations measured and simulated below the hypothetical method reporting limit of 0.1 μg/L cluster in the vicinity of -11.5, the log of 0.1 times the observation weight. Executing the model with the best fit parameters results in a peak contaminant concentration at the production well of 11.1 μg/L that occurs about 3000 days following the release. Hence the predicted concentration in the well is about 40% of the true value. Predictive error variance analysis is now undertaken to determine if the predictive error extends to include elevated contaminant concentrations occurring at the well.

3.1.3. Definition of the Predictive Solution Subspace

[50] The initial and calibrated models predict that peak concentrations occur at the production well after 3000–4000 days. This requires a longer simulation than does the calibration (800 days). This presents problems for traditional predictive error variance analysis since the model must be executed to obtain the sensitivity of the prediction to each parameter each iteration of the predictive analysis. If all base parameters were included (Figure 1, method 1), this would require 793 model runs each iteration. If simulating 3000 days of transport, this translates to 15,000 CPU min each time **y** is formed. This might render predictive error variance analysis impracticable. Time could be saved by defining superparameters in a manner similar to that used

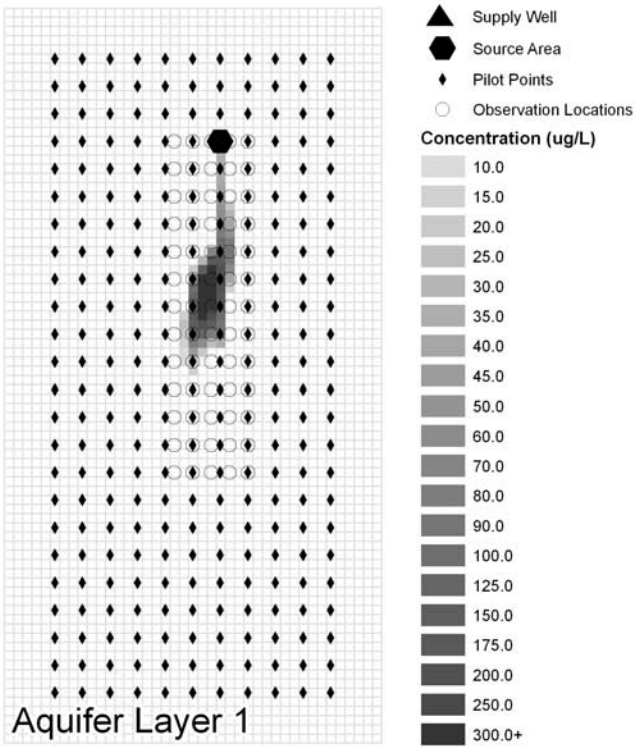

Figure 4. Hypothetical monitoring wells, and pilot points used to parameterize hydraulic conductivity.

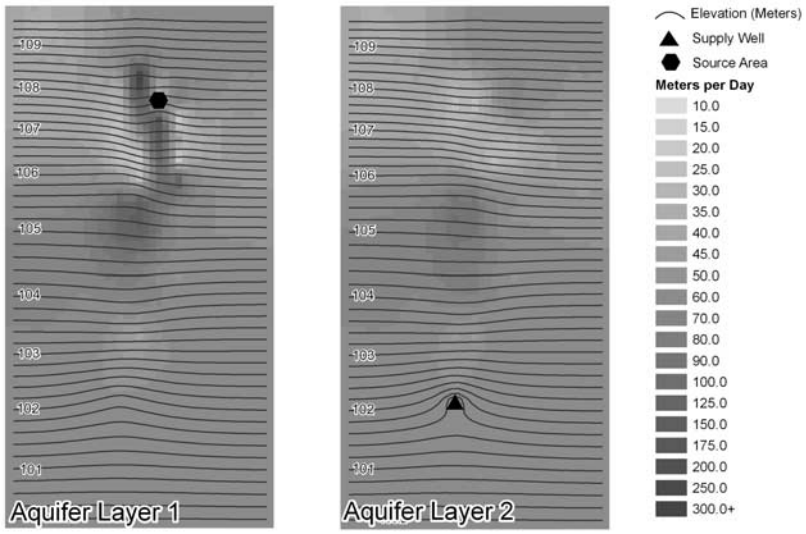

Figure 5. Calibrated aquifer hydraulic conductivities and simulated heads.

for calibration but including the prediction sensitivity as a row of a modified **X** matrix (Figure 1, method 2). This would reduce the computational burden; however, (1) 793 model runs are required to add this row to **X**; and (2) for a nonlinear model, the addition of a single prediction-pertinent direction to existing superparameters may be insufficient to explore predictive error variance.

[51] Instead, we define a solution subspace using sensitivity calculations that can be rapidly computed, which we term surrogate predictions (Figure 1, method 3). Predictive error variance analysis using (10) through (12) is then undertaken using this predictive solution subspace, which comprises linear combinations of the base parameters. Since the prediction is the peak concentration at a well, and since the source term is known, two approaches could be considered for rapidly obtaining sensitivities to define the predictive solution subspace using method 3: (1) advective transport as embodied in particle tracking; or (2) using an implicit finite difference (FD) advective-dispersive solution scheme. In each case the explicit TVD solution scheme would be used throughout the actual predictive error analysis.

[52] Here we use the first approach. This was accomplished by introducing 53 particles: 41 originating on the row containing the contaminant source, and 12 originating in a circle centered on the source. MODPATH [*Pollock*, 1994] was used to calculate the movement of these particles until they reached the model boundary or were captured by the well (Figure 7). Base parameter sensitivities were calculated on the basis of three criteria: (1) the time of arrival for particles captured by the well (19 particles); (2) for particles not captured by the well, the closest separation distance between the particle path and the well; and (3) the time at which this closest separation distance occurs.

[53] Since the flow and tracking models execute in seconds, the surrogate sensitivity matrix $\mathbf{X}_{sur}$ with dimensions of 792 columns by 159 rows (53 particles, each with three associated surrogate predictions), comprising sensitivities of all surrogate predictions with respect to all base parameters, was constructed on a single PC in several minutes. TSVD of $\mathbf{X}^t_{sur}\mathbf{Q}_{sur}\mathbf{X}_{sur}$ was completed using two diagonal weight matrices: the first, with weights of 1.0 for all captured particles and weights of 0.0 for all other particles (surrogate set A) and the second, with weights of 1.0 for all captured particles plus 22 particles with high sensitivities to the other two criteria, and weights of 0.0 for all other particles (surrogate set B).

3.1.4. Results

[54] Table 1 and Figure 8 summarize the outcomes of the predictive error analyses undertaken using the surrogate approaches, together with breakthrough curves simulated by the true, initial, calibrated, and full analysis (Figure 1, method 1) models. The initial model under-predicts the true

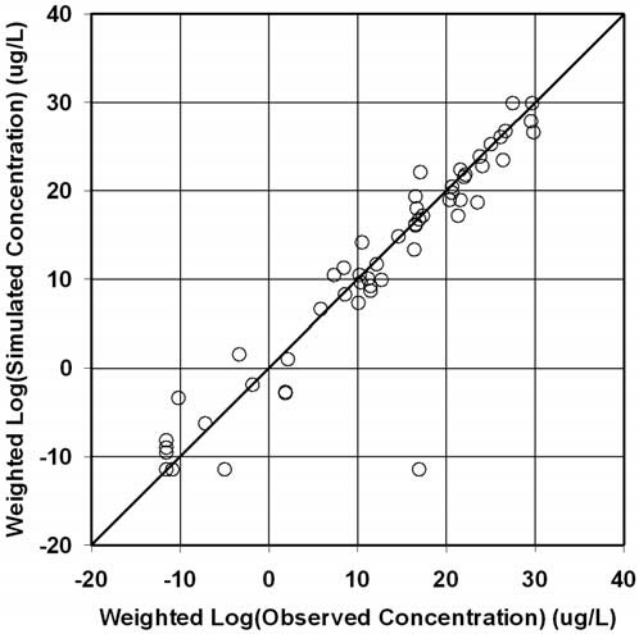

Figure 6. Simulated versus observed quantities in the synthetic model.

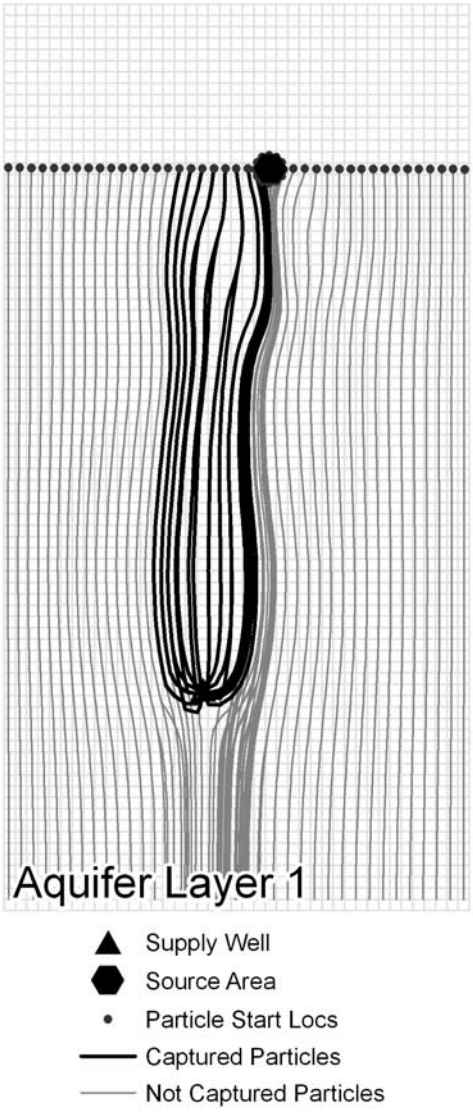

Figure 7. Particle starting locations and paths in the synthetic model.

peak concentration. The calibrated model predicts a peak concentration equal to about 40% of the true value. The full analysis completed using 792 base parameters predicts a peak concentration over twice the true value, but required over 2400 CPU hours to execute. In contrast, surrogate set B, using 41 superparameters defined from particle surrogates, predicts a peak concentration over 80% of that achieved using the full analysis, but required 125 CPU hours or about 5% of the time required for the full analysis. Surrogate set A, the most rapidly executing analysis, approaches the true prediction value but falls short of that computed using the full analysis and that computed using surrogate set B.

[55] The full analysis and the surrogate set B analysis include sufficient parameters, and are sufficiently stable that they identify a prediction that exceeds the true prediction, while maintaining calibration and null space parameter constraints. The surrogate set B analysis does not identify as high a concentration as does the full analysis. This is to be expected, since surrogate set B comprises only 41 prediction superparameters derived from the surrogate sensitivities, and illustrates the trade-off between computational burden and the ability to seek the full confidence range of a prediction. The fact that the initial and calibrated models simulate lower concentrations than the true model is unsurprising and was part of the motivation for this work.

3.2. Real-World Application

[56] We now investigate the possible fate of contaminants encountered in the sole-source glacial aquifers of Long Island, New York (the site). At the site, a plume of dissolved methyl-tertiary-butyl-ether (MTBE) is migrating downgradient of a refueling station, toward a public supply well (Figure 9). Investigations suggest that the top of the well screen is about 50 m below the plume. Simple mass flux calculations suggest that if the supply well should intercept the plume, relevant standards for MTBE could be exceeded. The pattern of MTBE concentrations is consistent with a finite-duration release. Further investigations revealed that an accident at the refueling station released an unknown volume of fuel 3 years prior to the discovery of groundwater contamination.

3.2.1. Model Development and Calibration

[57] Steady state groundwater flow and transient contaminant transport are simulated using an 11-layer model. General-head boundaries for the flow model are defined using a simple three-point regression scheme. The data available for calibration comprise 156 groundwater elevations and 278 measurements of MTBE concentration, ranging from nondetect to over 100,000 μg/L, collected in multilevel monitoring wells (Figure 9). Initial model

Table 1. Predictive Analysis Summary for Synthetic Model

Run	Description	Prediction		Approximate CPU Hours
		Peak Concentration[a]	Peak Time[b,c]	
True model	multi-Gaussian fields of known mean	27.52	1630	0.40
Initial model	uniform parameters based on known average	3.01	4050	n/a
Calibrated model	model calibrated with 30 superparameters	11.14	3080	120
Surrogate set A	19 particle-derived surrogates	25.55	2980	60[d]
Surrogate set B	41 particle-derived surrogates	54.50	2780	125[d]
Full analysis	792 base parameters	68.70	2750	2350[d]

[a]Concentration units, micrograms per liter (μg/L).
[b]Time units, days (d).
[c]Not explicitly incorporated in the prediction.
[d]In addition to the calibration.

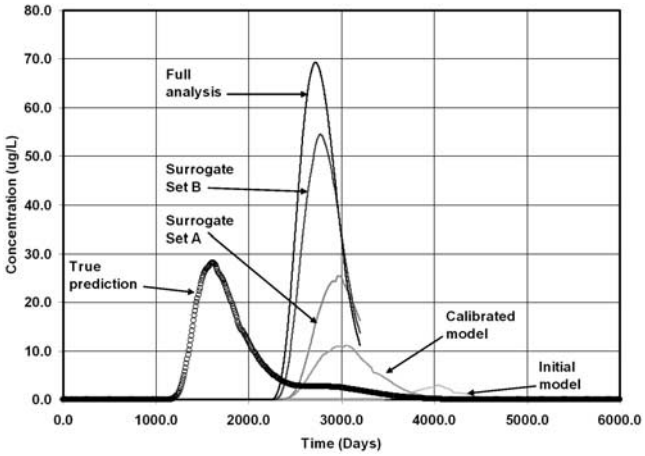

Figure 8. Peak concentrations and times in the synthetic model.

parameter values were based on published properties of the shallow (Upper Glacial) and deeper (Magothy) aquifers [*Soren and Simmons*, 1972; *Franke and McClymonds*, 1972]. The contaminant source term is unknown, though loading is assumed to commence at the date of the documented release. Its character is specified using lognormal (in time) mass loading defined by two parameters: the peak loading rate and the time at which the peak loading occurs.

[58] Updated estimates of the aquifer parameters, peak mass loading rate and timing for the contaminant mass loading function, and the general-head boundary parameters were estimated through manual and automated lumped-parameter calibration. This provided initial base parameter values for defining superparameters in the hybrid TSVD-Tikhonov calibration. This lumped-parameter model calculates a peak MTBE concentration at the potential receptor of about 36 μg/L, occurring about 3800 days following the release.

[59] For the calibration and predictive error variance analysis, horizontal and vertical hydraulic conductivities of model layers 1 through 9 (those layers in which contaminant transport is simulated) were parameterized using 1568 pilot points, 98 per layer for each parameter type. Pilot points were focused in the area encompassing the source and the well. A small number of pilot points outside this area extend the parameterization smoothly to model boundaries (Figure 10): Hydraulic conductivities assigned to these pilot points were fixed at their initial values. The remaining 1296 pilot points, together with nine parameters that describe recharge, porosity, and contaminant mass loading, constituted the base parameters.

[60] Initial values for all base parameters were defined from the lumped-parameter calibration. The model was executed for sufficient time during calibration to obtain simulated equivalents to measured data (t = 1200 days), this requiring about a third of the total execution time required to simulate contaminant migration to the production well. Base parameter sensitivities were calculated using perturbations across a 20-node CPU network, providing matrix **X** with dimensions of 1305 columns by 434 rows. **Q** was specified as a diagonal matrix with elements equal to the inverse of respective estimated measurement variances. Fifty superparameters were defined through TSVD of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$. Estimation of the superparameters was completed on the same CPU network.

[61] Calibrated hydraulic conductivities are illustrated in Figure 11 for model layer 3, corresponding with the middle elevations of the plume. Simulated versus observed concentrations are depicted in Figure 12. Simulated and observed concentrations below the method reporting limit of 0.1 μg/L cluster in the vicinity of −2.5, the log of 0.1 times the observation weight. (The highlighted data are from a source area well that shows variations over 4 orders of magnitude between sample events, and may represent the presence of free-phase contaminants not simulated by the model.) The model was executed with best fit parameters to obtain a peak concentration at the well of about 42 μg/L, this occurring about 3300 days following the release.

3.2.2. Definition of the Predictive Solution Subspace

[62] The initial and calibrated models suggest that peak concentrations occur at the well after nearly 4000 days of travel. As for the synthetic model, obtaining vector **y** listing the sensitivity of the prediction to all base parameters would be computationally costly. A full predictive analysis using base parameters commencing at their calibrated values would require 1306 executions of the model simulating nearly 4000 days of transport, equating to about 80,000 CPU min each time **y** is formed. Therefore we

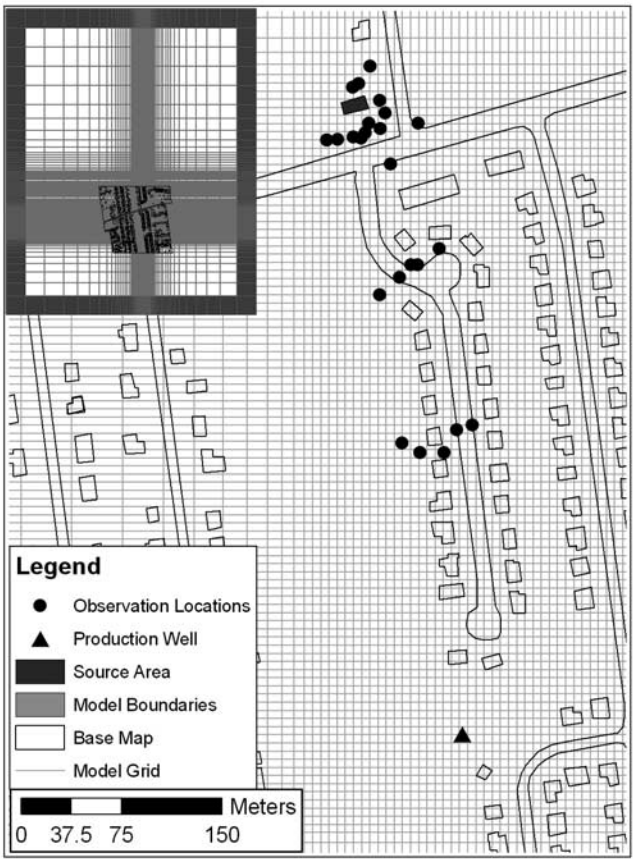

Figure 9. Real-world site showing the full model domain and boundaries (inset, top left), and local details from the finely discretized area of the model including the source area, production well, and observation locations.

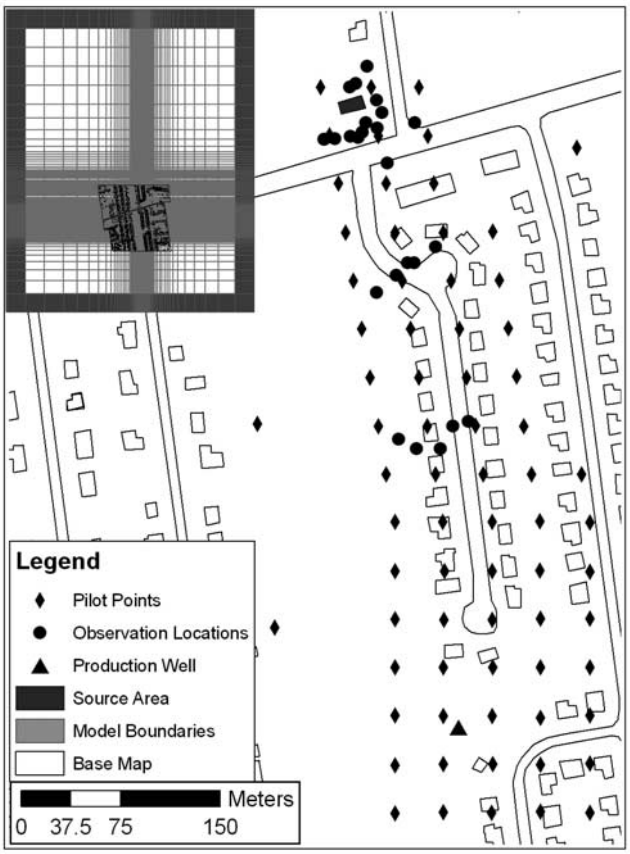

Figure 10. Real-world site showing the full model domain and boundaries (inset, top left), and local details from the finely discretized area of the model including the locations of pilot points used to parameterize hydraulic conductivity.

adopt the approach employed in the synthetic study and use surrogate sensitivities based on particle tracking to define a predictive solution subspace (Figure 1, method 3). This was accomplished using 5487 particles encircling the source and aligned along the row containing the source. MODPATH was used to simulate the movement of these particles until they reached the model boundary or were captured at the well. Base parameter sensitivities were calculated for three criteria analogous to those used in the synthetic model analysis.

[63] Because the flow and tracking models execute in seconds, matrix $\mathbf{X}_{sur}$ was constructed on a single PC in several minutes. Since particle tracking was used to obtain the surrogate prediction sensitivities, the sensitivity of parameters that define the contaminant source is zero. To ensure that base parameters defining source concentrations were represented in the predictive superparameters, $\mathbf{X}_{sur}$ was augmented with two row vectors containing zero entries except in the columns corresponding to the two source term parameters. Before undertaking TSVD of $\mathbf{X}^t_{sur}\mathbf{Q}_{sur}\mathbf{X}_{sur}$ to construct prediction superparameters, two alternative diagonal weight matrices were constructed. In each, weights of 1.0 were assigned for rows pertaining to captured particles and rows pertaining to particles with high sensitivities to the other two criteria, and weights of 0.0 were assigned for rows pertaining to the remaining particles. In the first weight matrix, weights assigned to rows pertaining to the source term were equal to 1.0 (surrogate set A); in the second weight matrix, weights assigned to rows pertaining to the source term were equal to 50,000 (surrogate set B). In each case, TSVD of $\mathbf{X}^t_{sur}\mathbf{Q}_{sur}\mathbf{X}_{sur}$ was used to define 54 prediction superparameters.

3.2.3. Results

[64] Table 2 and Figure 13 summarize the results of the predictive error analyses undertaken using the surrogate approaches, and depict breakthrough curves computed by the initial and calibrated models. The pattern is similar to that seen in the synthetic analysis. The calibrated lumped-parameter model predicts the lowest peak concentration. The calibrated superparameter model predicts a peak concentration about 19% higher than that of the initial model. Surrogate set A predicts a peak concentration about 50% higher than that of the initial model, and executes in a time comparable to that required for the calibration. Surrogate set B, constructed with higher weights assigned to surrogate sensitivities pertaining to source term parameters, predicts a peak concentration over 70% higher than that of the initial model, and executes in a time comparable to that required for the calibration. The full analysis (Figure 1, method 1) using 1305 base parameters was not completed since it is estimated that this would require over 10,000 CPU hours.

[65] The real-world analysis leads to smaller relative increases in the prediction than those computed in the synthetic analysis. This is unsurprising since (1) a confining unit is present in the synthetic model, about which the

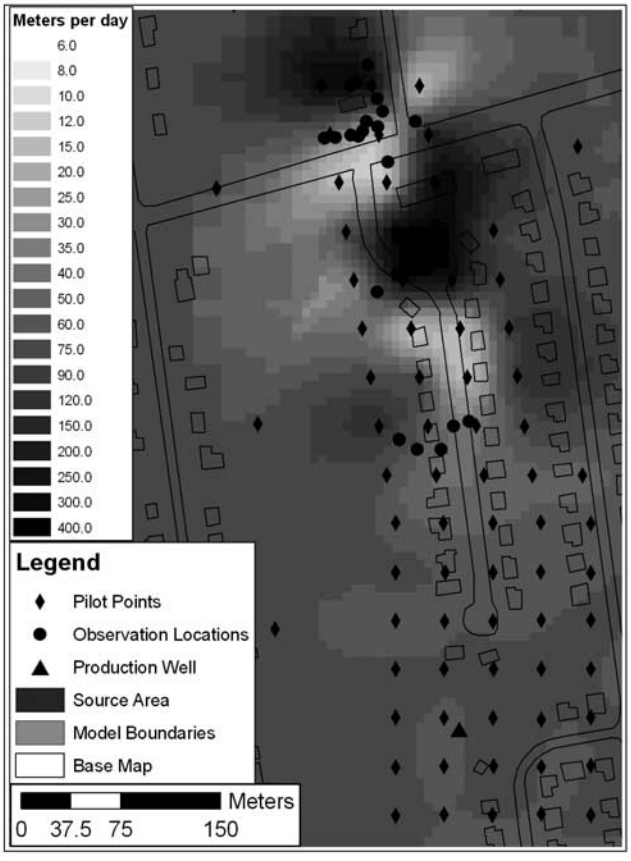

Figure 11. Calibrated hydraulic conductivity in real-world model layer 3.

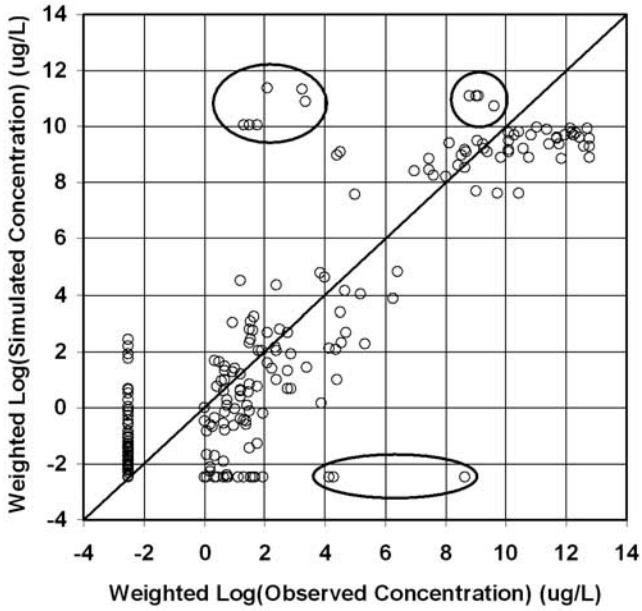

Figure 12. Simulated versus observed quantities in the real-world model. The highlighted data are from a source area well that may represent the presence of free-phase contaminants not simulated by the model.

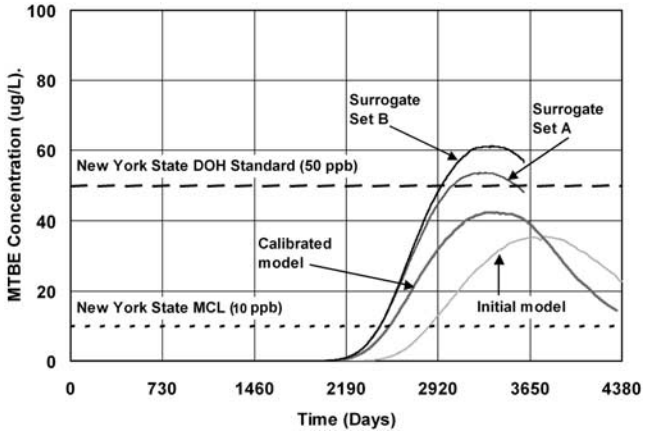

Figure 13. Peak concentrations and times in the real-world model.

measurement data provide little information, while no confining unit is simulated in the real-world model; (2) in the synthetic model the plume centerline is lateral to the well, whereas in the real-world model the plume centerline is aligned with the well; and (3) pumping in the real-world model leads to a broader capture zone than that in the synthetic model, so that more of the plume is captured following the lumped parameter calibration. Hence, in the synthetic model, alterations to parameter values during the predictive analysis aligned the simulated plume with the well in both vertical and horizontal directions, whereas this was already largely accomplished at the commencement of the real-world predictive analysis.

4. Discussion and Concluding Remarks

[66] This paper presents a method for investigating the potential error in predictions made with models calibrated using regularized inversion. Benefits of the approach include (1) the assurance of acceptable parameter values and fits to observed data, and (2) computational efficiency. Acceptable model-to-measurement fits and parameter reasonableness are ensured by enforcing calibration constraints on base parameter combinations that lie within the calibration solution space, together with constraints on base parameter combinations that lie within the calibration null space and are not constrained by available data. Efficiency is achieved using a problem-specific predictive solution subspace constructed using rapidly executed surrogate calculations.

[67] The example applications suggest that the method can help to assess the range of potential fates that are compatible with current knowledge of a contaminant's distribution. In particular, the method is suited to testing whether, given current knowledge, regulatory criteria may be exceeded at potential points of compliance. However, the method is general and may be employed with many models, for which outputs are differentiable with respect to the inputs, to explore the possible error associated with predictions.

[68] This predictive error analysis approach is a logical sequel to the hybrid TSVD-Tikhonov calibration method [*Tonkin and Doherty*, 2005]. Both techniques obtain efficiency by defining a low-dimension solution subspace spanned by orthogonal unit vectors, and estimating the

Table 2. Predictive Analysis Summary for Real-World Model

Run	Description	Prediction		Approximate CPU Hours
		Peak Concentration[a]	Peak Time[b,c]	
Initial model	calibrated lumped base parameters	35.72	3770	n/a
Calibrated model	model calibrated with 50 superparameters	42.33	3320	450
Surrogate set A	54 prediction superparameters	53.52	3300	300[d]
Surrogate set B	54 prediction superparameters with high weight source term	61.22	3340	300[d,e]
Full analysis	1305 base parameters	n/a	n/a	>10,000[f]

[a]Concentration units, micrograms per liter (μg/L).
[b]Time units, days (d).
[c]Not explicitly incorporated in the prediction.
[d]In addition to the calibration.
[e]Converged in four optimization iterations.
[f]Estimated assuming six optimization iterations.

coefficients of these vectors in a reformulated calibration or prediction problem. Typically, different subspaces would be employed in each case, through TSVD of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ for calibration, and through TSVD of a matrix formed using rapidly computed surrogate sensitivities for the predictive error variance analysis. Under these circumstances, calibration and predictive error analysis can be accomplished efficiently despite the use of many parameters. Together, these approaches may provide some of the advantages of detailed parameterization that are outlined in the introduction.

[69] The surrogate approach for constructing the predictive solution subspace rests on the premise that the relevance of parameters to a prediction can be rapidly approximated. The surrogate approach may not always be applicable. In those contexts, Figure 1, method 1 or method 2, adjoint sensitivities, or other methods could be employed; this would still be consistent with the theory that is presented in the paper. In the instances that surrogate methods may apply, use of a prediction subspace may lead to underestimated predictive error limits. However, these limits may be more accurate than those computed using lumped-parameter models which cannot represent parameter detail that may lead to extreme predictions, with the result that such predictions can be erroneously interpreted as improbable. Whether the surrogates provide sufficient information to minimize/maximize the prediction to the extent necessary to support decision-making can be determined in the context of the question at hand.

[70] Predictive error variance analysis differs from predictive uncertainty analysis, since the former bases its analysis on a single model. Hence the method described is applicable where decisions rest on the use of a single calibrated model, and can be used to investigate the possible magnitude of errors in predictions made by that model. The example applications suggest that the computational effort required to complete a predictive error variance analysis can be commensurate with that required for calibration, and can be completed following calibration without undue burden. The examples also demonstrate that efficiency can be traded for predictive confidence width: Rapid computation of predictive confidence limits may lead to underestimation of those limits. Nevertheless, the approach offers a flexible and efficient means of investigating the feasibility of an unacceptable outcome, and if such is identified, the mechanism or mechanisms that may lead to that outcome.

[71] Although the approach is nonlinear, linearity is implied in the definition of predictive superparameters and in the definition of null space constraints. These assumptions may incur errors when computing the level of confidence associated with a certain prediction interval. Nonetheless, valuable outcomes of the analysis include (1) a parameter field that results in a calibrated model; (2) a parameter field that respects the stochastic character of the simulated area; (3) a parameter field devoid of unreasonable characteristics that lead to unduly pessimistic predictions; and (4) a parameter field corresponding to a prediction that is at, or approaching, its confidence limit. This parameter field can be used to make other predictions, such as the likely performance of a hypothetical groundwater remedy.

[72] Finally, caution is advised that when using any predictive error or uncertainty analysis method, the fact that the analysis did not identify an unwanted outcome does not mean that an unwanted outcome cannot or will not occur; that is, the absence of proof does not equate to proof of absence.

[73] **Acknowledgments.** The authors thank Joseph Haas (NYSDEC) for releasing data used in the real-world study, and Environmental Assessment and Remediations (EAR) for support compiling these data. The authors also thank H. J. Hendricks Franssen, two anonymous reviewers, and the editors for their comments which improved the final manuscript enormously. The approach described in this study is fully documented and is encapsulated in a set of programs supplied as part of the PEST [*Doherty*, 2005] suite, available free from www.sspa.com/pest.

References

Aster, R. C., B. Borchers, and C. H. Thurber (2005), *Parameter Estimation and Inverse Problems*, 301 pp., Elsevier, New York.

Beven, K. J., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*, 279–298.

Carrera, J., and S. P. Neuman (1986a), Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, *Water Resour. Res.*, *22*(2), 199–210.

Carrera, J., and S. P. Neuman (1986b), Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms, *Water Resour. Res.*, *22*(2), 211–227.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten (2005), Inverse problem in hydrogeology, *Hydrogeol. J.*, *13*, 206–222, doi:10.1007/s10040-004-0404-7.

Certes, C., and G. de Marsily (1991), Application of the pilot point method to the identification of aquifer transmissivities, *Adv. Water Resour.*, *14*(5), 284–300.

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

Cooley, R. L. (2004), A theory for modeling ground-water flow in heterogeneous media, *U.S. Geol. Surv. Prof. Pap., 1679*, 220 pp.

Cooley, R. L., and S. Christensen (2006), Bias and uncertainty in regression-calibrated models of groundwater flow in heterogeneous media, *Adv. Water Resour.*, *29*(5), 639–656.

Deutsch, C., and A. Journel (1992), *GSLIB: Geostatistical Software Library and User's Guide*, 340 pp., Oxford Univ. Press, New York.

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

Doherty, J. (2005), *User's Manual for PEST Version 10*, 339 pp., Watermark Numer. Comput., Brisbane, Australia.

Draper, N. R., and H. Smith (1981), *Applied Regression Analysis*, 2nd ed., 709 pp., John Wiley, Hoboken, N. J.

Engl, H. W., M. Hanke, and A. Neubauer (1996), *Regularization of Inverse Problems*, 321 pp., Springer, New York.

Franke, O. L., and N. E. McClymonds (1972), Summary of the hydrogeologic situation on Long Island, New York, as a guide to water-management alternatives, *U.S. Geol. Surv. Prof. Pap.*, *627-F*, 59 pp.

Gomez-Hernandez, J. J., A. Sahuquillo, and J. E. Capilla (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: 1. Theory, *J. Hydrol.*, *203*, 162–174.

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.

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

Guadagnini, A., and S. P. Neuman (1999b), Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains: 2. Computational examples, *Water Resour. Res.*, *35*(10), 3019–3039.

Gutjahr, A. L., B. Bullard, S. Hatch, and L. Hughson (1994), Joint conditional simulations and the spectral approach for flow modeling, *Stochast. Hydrol. Hydraul.*, *8*(1), 70–108.

Haber, E., U. M. Ascher, and D. Oldenberg (2000), On optimization techniques for solving non-linear inverse problems, *Inverse Probl.*, *16*(5), 1263–1280.

Harbaugh, A. W., E. R. Banta, M. C. Hill, and M. G. McDonald (2000), MODFLOW-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*, 121 pp.

Hendricks Franssen, H.-J., J. Gomez-Hernandez, and A. Sahuquillo (2003), Coupled inverse modeling of groundwater flow and mass transport and the worth of concentration data, *J. Hydrol.*, *281*(4), 281–295.

Hernandez, A. F., S. P. Neuman, A. Guadagnini, and J. Carrera (2006), Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media, *Water Resour. Res.*, *42*, W05425, doi:10.1029/2005WR004449.

Kitanidis, P. K. (1996), On the geostatistical approach to the inverse problem, *Adv. Water Resour.*, *19*(6), 333–342.

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*, *211*, 69–85.

Lavenue, M., and G. de Marsily (2001), Three-dimensional interference test interpretation in a fractured-unfractured aquifer using the pilot point inverse method, *Water Resour. Res.*, *37*(11), 2659–2676.

Menke, W. (1989), *Geophysical Data Analysis: Discrete Inverse Theory*, 2nd ed., 289 pp., Elsevier, New York.

Moore, C. (2006), The use of regularized inversion in groundwater model calibration and prediction uncertainty analysis, Ph.D. thesis, Univ. of Queensland, Brisbane, Australia.

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.

Moore, C., and J. Doherty (2006), The cost of uniqueness in groundwater model calibration, *Adv. Water Resour.*, *29*(4), 605–623.

Oliver, D. S., L. B. Cunha, and A. C. Reynolds (1997), Markov Chain Monte Carlo methods for conditioning a permeability field to pressure data, *Math. Geol.*, *29*, 61–91.

Pollock, D. W. (1994), User's guide for MODPATH/MODPATH-PLOT, version 3: A particle tracking post-processing package for MODFLOW, the U.S. Geological Survey finite-difference ground-water flow model, *U.S. Geol. Surv. Open File Rep*., *94-464*, 249 pp.

Rubin, Y., and G. Dagan (1987a), Stochastic identification of transmissivity and effective recharge in steady state groundwater flow: 1. Theory, *Water Resour. Res.*, *23*(7), 1185–1192.

Rubin, Y., and G. Dagan (1987b), Stochastic identification of transmissivity and effective recharge in steady state groundwater flow: 2. Case study, *Water Resour. Res.*, *23*(7), 1193–1200.

Soren, J., and D. L. Simmons (1972), Thickness and hydrogeology of aquifers and confining units below the Upper Glacial Aquifer on Long Island, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, *86-4175*, 3 pp.

Tikhonov, A., and V. Arsenin (1977), *Solutions of Ill-Posed Problems*, 258 pp., V. H. Winston, Washington, D. C.

Tonkin, M., and J. Doherty (2005), A hybrid regularized inversion methodology for highly parameterized models, *Water Resour. Res.*, *41*, W10412, doi:10.1029/2005WR003995.

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.

Vogel, C. R. (2002), *Computational Methods for Inverse Problems*, *Frontiers in Appl. Math. Ser.*, vol. 23, Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Woodbury, A. D., and T. J. Ulrych (2000), A full-Bayesian approach to the groundwater inverse problem for steady state flow, *Water Resour. Res.*, *36*(8), 2081–2094.

Yeh, T.-C. J., M. Jin, and S. Hanna (1996), An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields, *Water Resour. Res.*, *32*(1), 85–92.

Zheng, C., and P. Wang (1999), MT3DMS: A modular three-dimensional multi-species transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems: Documentation and user's guide, *Contract Rep. SERDP-99-1*, U.S. Army Corps of Eng. Eng. Res. and Dev. Cent., Vicksburg, Miss.

Zimmerman, D. A., et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resour. Res.*, *34*(6), 1373–1414.

J. Doherty, Watermark Numerical Computing, 336 Cliveden Avenue, Corinda 4075, Australia. (jdoherty@gil.com.au)

C. Moore, Lincoln Ventures Ltd., Lincoln University, P.O. Box 84, Christchurch, New Zealand. (moorec5@lvl.co.nz)

M. Tonkin, S. S. Papadopulos and Associates, Inc., 7944 Wisconsin Avenue, Bethesda, MD 20814, USA. (matt@sspa.com)