A hybrid regularized inversion methodology for highly parameterized environmental models

Matthew James Tonkin

S.S. Papadopulos and Associates, Inc., Yarmouth Port, Massachusetts, USA

John Doherty

Watermark Numerical Computing, University of Queensland, Corinda, Queensland, Australia

Received 27 January 2005; revised 4 June 2005; accepted 15 June 2005; published 22 October 2005.

[1] A hybrid approach to the regularized inversion of highly parameterized environmental models is described. The method is based on constructing a highly parameterized base model, calculating base parameter sensitivities, and decomposing the base parameter normal matrix into eigenvectors representing principal orthogonal directions in parameter space. The decomposition is used to construct super parameters. Super parameters are factors by which principal eigenvectors of the base parameter normal matrix are multiplied in order to minimize a composite least squares objective function. These eigenvectors define orthogonal axes of a parameter subspace for which information is available from the calibration data. The coordinates of the solution are sought within this subspace. Super parameters are estimated using a regularized nonlinear Gauss-Marquardt-Levenberg scheme. Though super parameters are estimated, Tikhonov regularization constraints are imposed on base parameters. Tikhonov regularization mitigates over fitting and promotes the estimation of reasonable base parameters. Use of a large number of base parameters enables the inversion process to be receptive to the information content of the calibration data, including aspects pertaining to small-scale parameter variations. Because the number of super parameters sustainable by the calibration data may be far less than the number of base parameters used to define the original problem, the computational burden for solution of the inverse problem is reduced. The hybrid methodology is described and applied to a simple synthetic groundwater flow model. It is then applied to a real-world groundwater flow and contaminant transport model. The approach and programs described are applicable to a range of modeling disciplines.

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

1. Introduction

[2] The inverse problem in groundwater modeling is generally ill posed and nonunique [*Carrera and Neuman*, 1986]. Parsimony is often employed to achieve a unique solution. This typically involves parameterizing the model with zones of piecewise constancy such that the number of parameters (n) is small and the number of observations (m) is larger than n. This makes the inverse problem numerically tractable but often results in high levels of model-to-measurement misfit. Further, where parameter zones are based on arbitrary geometry that does not accurately reflect geology, levels of misfit reflect the parameterization scheme, not measurement noise. This occurs, at least in part, because this approach precludes the inverse process from inferring heterogeneity where calibration data supports its inclusion. This can result in high variance for predictions that are sensitive to that detail [*Moore and Doherty*, 2005]. Additional sources of error, such as structural inaccuracy, are outside the scope of this study.

[3] An alternative approach is to parameterize the model with more parameters than are uniquely determinable from the data [*Jackson*, 1972]. This can enable variations in spatially distributed parameters to be specified at a scale commensurate with field-scale heterogeneity. However, this may lead to an under-determined problem, the solution to which is nonunique. In these circumstances regularization can be employed to stabilize the inversion of the under-determined problem and guarantee convergence to a unique solution [*Tikhonov and Arsenin*, 1977; *Engl et al.*, 1996]. Though regularization can be implemented by a variety of means, a regularization parameter, μ, is employed that governs the strength with which regularization constraints are imposed. Weak imposition of regularization constraints can result in close agreement between model outputs and observations at the cost of parameter reasonableness and numerical stability. Strong imposition of regulariza-

0043-1397/05/2005WR003995

tion constraints can result in greater numerical stability but a decreased ability of the calibration process to identify small-scale parameter variations [*Engl et al.*, 1996; *Haber*, 1997; *Vogel*, 2002].

[4] Two forms of regularization are described in this study. These are as follows.

[5] 1. The first is Tikhonov regularization, where the normal matrix is amended by the addition of equations relating parameters to each other or to preferred values. Here μ is a multiplier for the regularization equation weights [*Tikhonov and Arsenin*, 1977; *Vogel*, 2002]. In this formulation a preferred condition is assumed to prevail from which deviations occur to the extent that they enable model-to-measurement misfit to be reduced to a level determined by the modeler.

[6] 2. The second is truncated *s*ingular value decomposition (TSVD) where the inverse problem is truncated to a subspace of the full problem [*Lawson and Hanson*, 1995; *Weiss and Smith*, 1998]. Here μ is a filter that excludes eigenvectors of the normal matrix that do not contribute to objective function minimization.

[7] Regularized inversion is common in geophysical data interpretation where efficient methods have been applied to obtaining parameter sensitivities, such as the adjoint state [*Tsourlos and Ogilvy*, 1999], and solving the large matrix equations that result from estimating a large number of parameters, such as conjugate gradients [*Mackie and Madden*, 1993; *Haber et al.*, 2000] and matrix factorizations [*Portniaguine*, 1999; *Portniaguine and Zhdanov*, 2002]. In groundwater flow-and-transport modeling difficulties are encountered in undertaking regularized inversion that do not exist to the same extent in geophysical data interpretation. These include the following.

[8] 1. Model complexity and nonlinearity lead to forward simulation times of minutes, hours or greater. This creates a high computational burden for obtaining the sensitivities of more than just a few parameters.

[9] 2. Sensitivity calculations are not typically embedded in the simulation code hence obtaining parameter sensitivities is an additional computational step.

[10] 3. Groundwater flow and contaminant transport models comprise a wide variety of parameter types. Each of these parameter types may be best represented using different regularization strategies.

[11] For these reasons a priori parsimony, such as is provided by a small number of zones, is often used to parameterize groundwater models. Yet, a priori parsimony can inhibit the calibration process from extracting information from observations that pertains to the parameters on which model predictions depend. For example, where a model is used to predict contaminant movement, predictions may depend upon aquifer heterogeneity [*Moore and Doherty*, 2005]. This is particularly important where concentration data are included in the calibration data set since groundwater elevation data may contain little information on this heterogeneity [*Franssen et al.*, 2003].

[12] Regularized inversion of a highly parameterized model has greater potential to infer detail from a calibration data set. When Tikhonov regularization is employed this can be accomplished while ensuring that a preferred condition prevails. However the advantages of using a large number of parameters come at a cost of high calibration times and a tendency toward numerical instability. An ideal calibration methodology would combine the rapid inversion times and numerical stability that result from employing a small number of parameters, with the potential to simulate and infer small-scale parameter variations that comes from using highly parameterized models.

[13] This paper introduces a hybrid approach to the regularized inversion of highly parameterized models and demonstrates the approach using synthetic and real-world groundwater models. Underpinning the approach is the combination of subspace and Tikhonov regularization methods. The paper is organized as follows: an overview of nonlinear least squares minimization is followed by a description of the regularization methods employed in the hybrid approach. The hybrid approach is described, and applied to a simple synthetic groundwater flow model. The approach is then applied to a real-world groundwater flow-and-transport model designed to support the operation of a pump-and-treat (PT) groundwater remedy. The bases of decisions intrinsic to the hybrid approach are discussed, including the selection of regularization parameters.

2. Nonlinear Least Squares

[14] The equations presented below assume that model outputs are a linear function of the model parameters. In fact, environmental models are rarely linear. However, these equations form the basis of the iterative gradient-based parameter estimation process that underpins traditional least squares minimization and the hybrid methodology.

[15] For a linear model, the relationship between the n model parameters, $\mathbf{p}$, and the m model outputs for which there are corresponding field observations, $\mathbf{y}$, can be represented by the matrix equation:

$$\mathbf{Xp} = \mathbf{y} \tag{1}$$

where $\mathbf{X}$ is an $m \times n$ matrix of independent variables that for linear problems is equivalent to the Jacobian matrix, describing the forward action of the model. $\mathbf{X}$ contains the sensitivity of each simulated equivalent of each observation with respect to each parameter. $\mathbf{X}$ can be constructed using perturbation sensitivities, requiring $n + 1$ forward model executions; sensitivity equations, requiring the solution of $n + 1$ sensitivity equations; or adjoint sensitivities, requiring the solution of $m + 1$ adjoint sensitivity equations [*Townley and Wilson*, 1985; *Carrera et al.*, 1990; *Sun*, 1994]. Where $n > m$, adjoint methods require the fewest model runs to form $\mathbf{X}$. Further, adjoint and sensitivity equation sensitivities may be more accurate than perturbation sensitivities [*Sun*, 1994; *Hill*, 1998]. However, perturbation sensitivities can be obtained for any model because no changes are required to the model code to implement their calculation. Perturbation sensitivities are used in the hybrid approach. Because the hybrid approach reduces the dimensionality of the inverse problem, i.e., the ratio of n to m, the computational advantages of adjoint methods may often be eliminated.

[16] The hybrid regularization approach is founded upon traditional weighted least squares (L_2 norm) minimization of a model-to-measurement misfit objective function defined as

$$\Phi_m = \left(\mathbf{d} - \mathbf{X}\left(\underline{\mathbf{p}}\right)\right)^t \mathbf{Q}_m \left(\mathbf{d} - \mathbf{X}\left(\underline{\mathbf{p}}\right)\right) \quad (2)$$

where $\underline{\mathbf{p}}$ is an n-row vector of current parameter values, $\mathbf{d}$ is an m-row vector containing measurements (observations), and $\mathbf{Q}_m$ is a square $m \times m$ weight matrix, ideally proportional to the inverse of the covariance matrix of observation errors [*Bard*, 1974; *Hill*, 1998]. Other objective functions exist, such as the L_1 norm [*Xiang et al.*, 1993] and L_p norm [*Sun*, 1994]. However, the L_2 norm has desirable qualities including the ability to assess parameter and predictive uncertainty [*Bard*, 1974; *Cooley*, 2004].

[17] For a linear model Φ_m of (2) is minimized when

$$\underline{\mathbf{p}} = (\mathbf{X}^t\mathbf{Q}_m\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Q}_m\mathbf{d} \quad (3a)$$

In linear models where Φ_m is quadratic minimization of (2) may be achieved in a single step. In nonlinear models $\underline{\mathbf{p}}$ cannot be calculated directly from $\mathbf{d}$ in this fashion. Rather a parameter upgrade vector, $\Delta\underline{\mathbf{p}}$, is calculated using the equation:

$$\Delta\underline{\mathbf{p}} = (\mathbf{X}^t\mathbf{Q}_m\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Q}_m\mathbf{r} \quad (3b)$$

where $\mathbf{r}$ lists the residuals for the current parameter set. Repeated linearizations are required, necessitating repeated construction of $\mathbf{X}$ in an iterative process that converges to an optimal parameter set $\underline{\mathbf{p}}$. In highly parameterized models construction of $\mathbf{X}$ is the most computationally costly aspect of the inverse process [*Carrera et al.*, 1990; *Sun*, 1994].

[18] Use of (3b) is effective for over-determined systems comprising a small number of parameters. However, if $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$ is rank deficient or (near) singular, i.e., at least one of its eigenvalues is (near) zero, it cannot be inverted. Eigenvectors corresponding to the (near) zero eigenvalues of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$ span parameter combinations that do not contribute to minimizing (2) and are not estimable using the available observations. In this paper the term calibration null space is used to describe the subspace that is spanned by eigenvectors that do not contribute to minimizing (2), in distinction from the strictly defined term model null space, which is the space for which $\mathbf{Xp} = 0$. The calibration null space includes the model null space in addition to the space spanned by eigenvectors corresponding to low eigenvalues, since inclusion of these eigenvectors in the solution may amplify noise [*Lines and Treitel*, 1984; *Lawson and Hanson*, 1995; *Aster et al.*, 2005]. The calibration solution space is the complement of the calibration null space.

[19] The likelihood of encountering a (near) singular matrix increases as n increases. $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$ may be rescued from singularity using the Levenberg-Marquardt parameter, λ [*Levenberg*, 1944; *Lines and Treitel*, 1984]. In fact λ is often employed irrespective of whether the inverse problem is ill conditioned since its use enables the parameter upgrade vector to be determined along a continuum between the Gauss-Newton and steepest descent methods [*Marquardt*, 1963]. Despite the ability of λ to contribute stability to an otherwise unstable inversion, robust parameter estimation requires regularization. Tikhonov regularization, which has seen increasing application in groundwater modeling in recent years, is now described.

3. Tikhonov Regularization

[20] Tikhonov regularization is founded upon supplementing the observation data set with information pertaining directly to parameters. This takes the form of a series of regularization equations, the weights for which are determined during the inverse process. Regularization equations are typically based upon differencing schemes, such as the difference of a parameter value from a preferred value, or the difference in value between parameters [*Ory and Pratt*, 1995; *Liu and Ball*, 1999]. Example regularization equations are described by *Doherty* [2003] and may be similar in form to prior information equations described by *Cooley* [1982]. The measure of misfit pertaining to the regularization equations can be described by a regularization objective function that is similar in form to the measurement misfit objective function defined by (2), i.e.,

$$\Phi_r = \left(\mathbf{e} - \mathbf{R}\left(\underline{\mathbf{p}}\right)\right)^t \mathbf{Q}_r \left(\mathbf{e} - \mathbf{R}\left(\underline{\mathbf{p}}\right)\right) \quad (4)$$

where $\mathbf{R}$ is a regularization operator that encapsulates the regularization equations; $\mathbf{e}$ expresses a preferred condition; and $\mathbf{Q_r}$ incorporates (relative) weights assigned to the regularization equations.

[21] Inclusion of Tikhonov regularization in the inverse problem results in an approach referred to as penalized least squares [*Engl et al.*, 1996] in which a composite objective function is defined as

$$\Phi_g = \Phi_m + \mu_T\Phi_r \quad (5)$$

where μ_T is the Tikhonov regularization parameter and Φ_g is the composite or global objective function, i.e., the summation of the measurement objective function described by (2) and the regularization objective function described by (4). μ_T is calculated during the inversion process such that a target level of model-to-measurement fit is achieved. The target measurement objective function, Φ_t, is supplied by the modeler in accordance with the anticipated magnitude of noise in the calibration data. When formulated with a target measurement objective function, the inversion can be viewed as a constrained minimization in which μ_T performs a similar role to that of a Lagrange multiplier [*Haber*, 1997].

[22] *Neuman and Yakowitz* [1979] formed a composite least squares objective function in terms of log transmissivity, using a parameter plausibility criterion that is of similar form to (5). They recognized that information on aquifer transmissivity is typically insufficient to specify an optimal value (or weight) for prior information, and solved a series of generalized least squares problems to estimate it. *Jacobson* [1985] adopted this approach for incorporating prior information determined from an analysis of aquifer test data, reducing the condition number of the inverse problem and enabling estimation of a fairly large number of parameters using a smaller number of observations. The approach of *Neuman and Yakowitz* [1979] and *Jacobson* [1985] was not developed as a constrained minimization,

and was demonstrated only for the estimation of aquifer transmissivity.

[23] Tikhonov regularization can be implemented for all parameter types encountered in groundwater models. Though commonly employed for spatially varying parameters Tikhonov regularization has been employed in the reconstruction of a time-varying contaminant source [*Liu and Ball*, 1999]. Tikhonov relationships can be linear or nonlinear, and are often designed to yield smooth parameter fields. Smoothness is not always desirable [*Vogel*, 1997; *Portniaguine*, 1999] but has been demonstrated to produce numerical stability and conceptually agreeable results in groundwater models [*Liu and Ball*, 1999; *Doherty*, 2003].

[24] Because Tikhonov regularization enables the specification of an acceptable misfit and a preferred condition in the inverse problem definition, its inclusion can lead to estimated parameter fields that are in accordance with a prior conceptualization. However, inversion based solely on Tikhonov regularization is not unconditionally stable. This occurs because it is difficult to specify regularization constraints that are enforced strongly where the information content of the calibration data is low, and less strongly where the information content of the calibration data is high. This is particularly true for nonlinear models that include a variety of parameter types. A regularization method that is unconditionally stable, Truncated Singular Value Decomposition, is now described.

4. Truncated Singular Value Decomposition

[25] Singular value decomposition (SVD) is an alternative approach to full matrix inversion for determining $\Delta\underline{\mathbf{p}}$ from $\mathbf{X^tQ_mX}$ [*Lawson and Hanson*, 1995; *Anderson et al.*, 1999]. SVD decomposes an arbitrary matrix $\mathbf{A}$ into:

$$\mathbf{A} = \mathbf{UBV^t} \tag{6}$$

where $\mathbf{U}$ and $\mathbf{V}$ contain the left and right eigenvectors of $\mathbf{A}$, respectively [*Lawson and Hanson*, 1995]. In the general case the matrix that undergoes decomposition need not be square. In the special case of the square symmetric matrix $\mathbf{X^tQ_mX}$, $\mathbf{B}$ is diagonal and lists the n singular values or eigenvalues of $\mathbf{X^tQ_mX}$. In addition $\mathbf{U} = \mathbf{V}$ and the n column vectors of matrix $\mathbf{V}$ are the eigenvectors of $\mathbf{X^tQ_mX}$. Thus

$$\mathbf{X^tQ_mX} = \mathbf{VBV^t} \tag{7}$$

Because $\mathbf{X^tQ_mX}$ is symmetric positive semidefinite its eigenvectors are orthogonal, hence [*Lawson and Hanson*, 1995; *Seber and Wild*, 1989]:

$$\mathbf{V^t} = \mathbf{V^{-1}} \tag{8}$$

Combining (3b) and (7), and noting that $\mathbf{B}$ is diagonal and that the inverse of a product of two matrices $(\mathbf{YZ})^{-1}$ can be expressed as $\mathbf{Z^{-1}Y^{-1}}$, it is evident that $(\mathbf{VBV^{-1}})^{-1} = \mathbf{VB^{-1}V^t}$ and that the parameter upgrade vector $\Delta\underline{\mathbf{p}}$ can be determined from SVD of $\mathbf{X^tQ_mX}$ using

$$\Delta\underline{\mathbf{p}} = \mathbf{VB^{-1}V^tX^tQ_mr} \tag{9}$$

One evident advantage of solving for $\Delta\underline{\mathbf{p}}$ using (9) rather than (3b) is that SVD obviates the task of inverting $\mathbf{X^tQ_{\underline{m}}X}$ which can be difficult where $\mathbf{X^tQ_{\underline{m}}X}$ is (near) singular.

[26] The full complement of eigenvectors of $\mathbf{X^tQ_mX}$ is a set of mutually orthogonal (linearly independent) directions in parameter space. When using SVD and (9) to solve for $\Delta\underline{\mathbf{p}}$ the parameter upgrade vector is proportional to the weighted sum of these eigenvectors. Typically, the eigenvectors $\mathbf{V}$ are normalized, i.e., $\mathbf{V^tV} = 1.0$, and the corresponding eigenvalues $\mathbf{B}$ are proportional to the steepness (L_2 norm) of the slope of the objective function Φ_m in the direction of the corresponding eigenvector [*Shewchuk*, 1994].

[27] For a square positive semidefinite $n \times n$ matrix such as $\mathbf{X^tQ_mX}$ there exist n real-valued eigenvalues. In general a small number of eigenvalues are dominant, corresponding to a small number of eigenvectors representing dominant directions in parameter space. The remaining eigenvalues are vanishingly small and correspond to eigenvectors that span the calibration null space. Projections of the solution vector onto these eigenvectors may be dominated by noise and lead to inversion instability because inversion of $\mathbf{B}$ in (9) amplifies any noise present [*Lawson and Hanson*, 1995, p. 198; *Aster et al.*, 2005, p65]. Stable inversion is achievable where parameter combinations belonging to the calibration solution space are estimated, and parameter combinations corresponding to the calibration null space are eliminated from the inverse process.

[28] Truncated singular value decomposition (TSVD) is a mechanism for determining $\Delta\underline{\mathbf{p}}$ from the k most dominant eigenvectors, where $k < n$ is a subset of the full complement of eigenvectors. Truncation is accomplished using a regularization parameter, μ_S. Diagonal entries of $\mathbf{B}$ beyond the truncation limit established by μ_S are assigned a value of zero [*Lawson and Hanson*, 1995; *Aster et al.*, 2005], i.e.,

$$\mathbf{B} = \begin{bmatrix} B_1 & \mathbf{0} \\ \mathbf{0} & B_2 \end{bmatrix} \tag{10}$$

where $\mathbf{B}_2$ contains eigenvalues set to zero. The dimensions of the calibration solution space is determined by specifying $\mu_S = k$, corresponding to the k largest eigenvalues; or by specifying μ_S as the ratio of the lowest to highest eigenvalue that is to be retained. In the latter instance the dimensions of the subspace are unknown a priori and may vary through the parameter estimation process. In either circumstance the n-dimensional problem is transformed to a k-dimensional approximation.

[29] TSVD contrasts with a priori parsimony since the subspace determined from TSVD of $\mathbf{X^tQ_mX}$ is determined from the information content of the observations, whereas the subspace constructed from an a prior parsimony strategy is not. A variety of approaches exist for selecting μ_S. Popular among these is plotting the singular value spectrum, i.e., the entries in $\mathbf{B}$, from which the subset of eigenvectors that dominates the problem is typically evident [*Lawson and Hanson*, 1995; *Haber*, 1997; *Aster et al.*, 2005]. As described by *Moore and Doherty* [2005], it can also be selected in such a manner as to minimize the error variance of a particular model prediction. Selection of μ_S in the hybrid regularization methodology is discussed later.

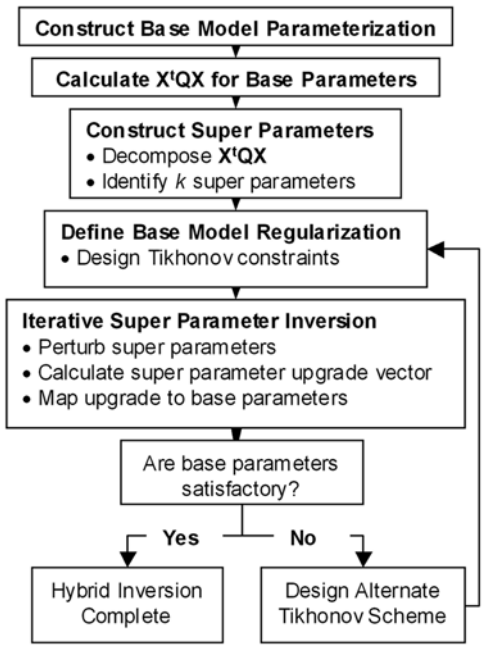

Figure 1. Schematic of the hybrid regularization process.

[30] The application of TSVD for model calibration is accompanied by three disadvantages: (1) Stability is not accompanied by accelerated solution to the inverse problem since it does nothing to reduce the burden of iteratively recalculating $\mathbf{X}$. (2) Using TSVD to improve the goodness of fit described by (2), the only mechanism for preventing over fitting is the selection of μ_S [*Aster et al.*, 2005], yet varying μ_S offers no guarantees for the reasonableness of estimated parameters. (3) Unlike Tikhonov methods there is no mechanism for setting a target measurement objective function or constraining the inversion in an effort to obtain reasonable estimated parameters and cease the inverse process. The method presented in this work addresses these issues.

5. Orthogonal Directions and Super Parameters

[31] The concept of super parameters is founded upon retaining the k eigenvectors obtained from TSVD of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$. As stated previously, solution of (9) using TSVD can be interpreted as a weighted sum of the k orthogonal eigenvectors retained from the truncated decomposition. This can be illustrated as

$$\Delta\underline{\mathbf{p}} = \mathbf{V}\underline{\mathbf{p}}_S \tag{11}$$

where $\underline{\mathbf{p}}_S$ provides the projection of the solution vector $\Delta\underline{\mathbf{p}}$ in the direction of the eigenvectors $\mathbf{V}$. Solution of the inverse problem in this form requires that $\underline{\mathbf{p}}_S$ is sought until $\underline{\mathbf{p}}_S$ is identified. It is possible to define and estimate $\underline{\mathbf{p}}_S$ as parameters, referred to here as super parameters. Together with the eigenvectors obtained from TSVD of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$, super parameters constitute the basis for reparameterizing the inverse problem.

[32] The use of super parameters is related to TSVD since the n-dimensional inverse problem is transformed into a k-dimensional problem. However, when using super parameters, the solution subspace is defined once through the initial TSVD of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$, and is constant throughout the inversion process. The reformulated problem is solved as a classical least squares problem but with parameters redefined as super parameters. Nonlinear inversion proceeds by perturbing super parameters to construct an $m \times k$ sensitivity matrix, $\mathbf{X}_S$, listing the sensitivity of each simulated equivalent of each observation with respect to each super parameter. The upgrade vector for the super parameters, $\Delta\underline{\mathbf{p}}_S$, can be determined using the same approach as (3b), i.e.,

$$\Delta\underline{\mathbf{p}}_S = \left(\mathbf{X}_S^t\mathbf{Q}_m\mathbf{X}_S\right)^{-1}\mathbf{X}_S^t\mathbf{Q}_m\mathbf{r} \tag{12}$$

When each super parameter is perturbed or upgraded, base parameters are perturbed or upgraded in the ratio in which they occur in the corresponding eigenvector $\mathbf{V}$ (equation (11)). Solving the reformulated problem provides the coordinates of the solution within a parameter subspace for which information is available from the calibration data. Since k executions of the model are required to construct $\mathbf{X}_S$, when $k \ll n$ the computational burden of each linearization in the inversion is reduced by the ratio k:n. Hence the number of base parameters can be large in an effort to capture the information content of the calibration data. Making n large gives the hybrid method flexibility in identifying informative directions in parameter space. However, the rank of the reformulated inverse problem is only as large as k and beyond a certain value of n this is unlikely to grow.

[33] *Jacobson* [1985] analyzed the singular value spectrum of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ and, where this spectrum suggested that use of prior information alone was insufficient to stabilize the problem, proposed reparameterizing aquifer transmissivity by (1) combining parameter zones or (2) constructing surrogate parameters. Surrogate parameters were constructed on the basis of the eigenvectors of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ in a manner similar to super parameters. However, *Jacobson* [1985] combined prior estimates on the log transmissivity of parameter zones to form prior estimates on the surrogate parameters. Further, *Jacobson* [1985] estimated a parameter plausibility scalar multiplier using stepwise residual and maximum likelihood analyses, rather than formulate the problem as a constrained minimization. Finally, the approach of *Jacobson* [1985] was demonstrated for estimation of aquifer transmissivity, whereas the approach described here is applicable to all parameter types.

[34] Reducing the dimensionality of the problem using super parameters stabilizes the inverse process and increases computational efficiency. This mitigates one drawback of traditional TSVD. However, two drawbacks remain, namely there is no guarantee of parameter reasonableness, and no mechanism for preventing over fitting. These are the principal strengths of the Tikhonov method.

6. Hybrid Regularization Methodology

[35] The hybrid methodology combines TSVD and Tikhonov regularization in an effort to rapidly invert under-determined systems in a stable manner and produce base parameter fields that are acceptable to the modeler. The following narrative describes the stepwise application of the hybrid methodology (Figure 1).

[36] 1. Construct and calibrate over-determined (parsimonious) model.

[37] 2. Construct highly parameterized model comprising n base parameters and m observations, where n may be greater than m. This is referred to as the base model parameterization. Initial values for base parameters are identified from the solution of the over-determined problem.

[38] 3. Calculate $\mathbf{X}$ for base model parameters and form $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$. $\mathbf{X}$ is only constructed once. Perturbations, sensitivity equations or adjoint sensitivities can be used.

[39] 4. Decompose $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$ into eigenvectors and their corresponding eigenvalues. Construct k super parameters, factors by which the dominant eigenvectors of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$ are multiplied to minimize the measurement objective function of equation (2).

[40] 5. Design Tikhonov regularization scheme(s) for base parameters, such as a scheme that preferentially minimizes the difference between closely spaced parameters of the same type. Identify a target objective function for the constrained minimization.

[41] 6. Estimate parameters by perturbing super parameters. The inversion is nonlinear since super parameter sensitivities are recalculated and super parameters are upgraded iteratively. Since Tikhonov regularization is assigned to base parameters and a target measurement objective function is specified, a regularization parameter is recalculated each iteration of the super parameter process.

[42] Methods for determining the truncation level for eigenvalues of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$, and hence identifying the number of super parameters to use in the reformulated problem, are not rigid. This is consistent with the findings of *Jacobson* [1985], who concluded from SVD analysis of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ that determining the truncation limit is not straightforward but may be accomplished from empirical analysis of the singular value spectrum.

[43] The maximum number of super parameters is equivalent to n, i.e., the number of base parameters. Defining n super parameters would be equivalent to retaining all eigenvectors of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$, and no computational gains would be made. The truncation upper limit could also be set as the number of eigenvalues that exhibit a ratio of greater than 1.0×10^{-6} to the largest eigenvalue, since our experience suggests this approximates the level of noise encountered in perturbation sensitivities. If the singular value spectrum is plotted it may appear that a small subset of the corresponding eigenvectors is sufficient to define the calibration solution space [*Aster et al.*, 2005; *Haber*, 1997]. However, there is no guarantee at the outset of the inverse process that that Φ_t can be achieved, though this does not typically deteriorate the Tikhonov scheme from ensuring reasonable parameters.

[44] In practice it is beneficial to select a number of super parameters that is greater than the minimum necessary to achieve Φ_t. Therefore we define a slightly ill-posed problem. Using more super parameters than are needed for unique parameter estimation at Φ_t allows the Tiknonov regularization flexibility in identifying reasonable parameters. If the resulting fit is unsatisfactory k may be increased and the super parameter inversion recommenced without recalculating and decomposing $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$. Alternatively, since super parameters are determined from TSVD of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$, i.e., before Tikhonov constraints are included, alternative Tikhonov schemes can be implemented without recalculating and decomposing $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$.

[45] In a nonlinear model $\mathbf{X}$ changes as the inversion progresses, however it is desirable that the super parameters, i.e., orthogonal directions determined through TSVD of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$, span the estimable parameter space throughout the inversion in spite of changing $\mathbf{X}$. In an effort to assure this, we define the base parameters used for calculating $\mathbf{X}$ on the basis of 'lumped' or averaged parameters obtained though solving a prior over-determined (parsimonious) inverse problem (step 1 above). This strategy is described real-world example below. If the inversion exhibits unsatisfactory convergence or fails to attain Φ_t, it may be indicative that the super parameters do not span the estimable subspace, or more seriously that the model suffers from pervasive structural or conceptual errors. The first of these possible causes can be evaluated by recalculating and decomposing $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$ using the current base parameters and recommencing the super parameter estimation process, whereas the second requires that fundamental model assumptions be revisited.

7. Synthetic Case Study

[46] The hybrid methodology is first illustrated using a synthetic groundwater model based on that described by C. Moore and J. E. Doherty (The cost of uniqueness in groundwater model calibration, submitted to *Advances in Water Resources*, 2005, hereinafter referred to as Moore and Doherty, submitted manuscript, 2005). Figure 2a shows a regular finite difference domain with dimensions 500 m by 800 m, discretized using 50 columns and 80 rows. Boundaries are described by an upgradient Neumann (specified flux) condition of 0.1 m^3/d/m and downgradient Dirichlet (specified head) condition equivalent to 0.0 m. The aquifer is assumed confined. The only parameter considered for estimation is hydraulic conductivity. For purposes of parameter estimation the spatial distribution of hydraulic conductivity is represented using 104 pilot points placed throughout the model domain (Figure 2a) (see *Certes and de Marsily* [1991] and *RamaRao et al.* [1995, and references therein] for a discussion of pilot points as a parameterization device). The use of pilot points as a parameterization scheme in Tikhonov regularized inversion is described by *Doherty* [2003].

[47] The "true" hydraulic conductivity was generated by a three-step process.

[48] 1. A multi-Gaussian hydraulic conductivity field with an average log conductivity of 0.0 m/d (corresponding to a hydraulic conductivity of 1.0 m/d) was generated based on a log exponential variogram described by:

$$\gamma(h) = b\left(1 - e^{-ah}\right) \tag{13}$$

where the sill (b) equals 0.2 m^2 and the range coefficient (a) equals 200 m; and h is the separation (lag) distance (Figure 2a) [*Deutsch and Journel*, 1992].

[49] 2. The hydraulic conductivity at each pilot point was identified using bilinear interpolation from this multi-Gaussian field.

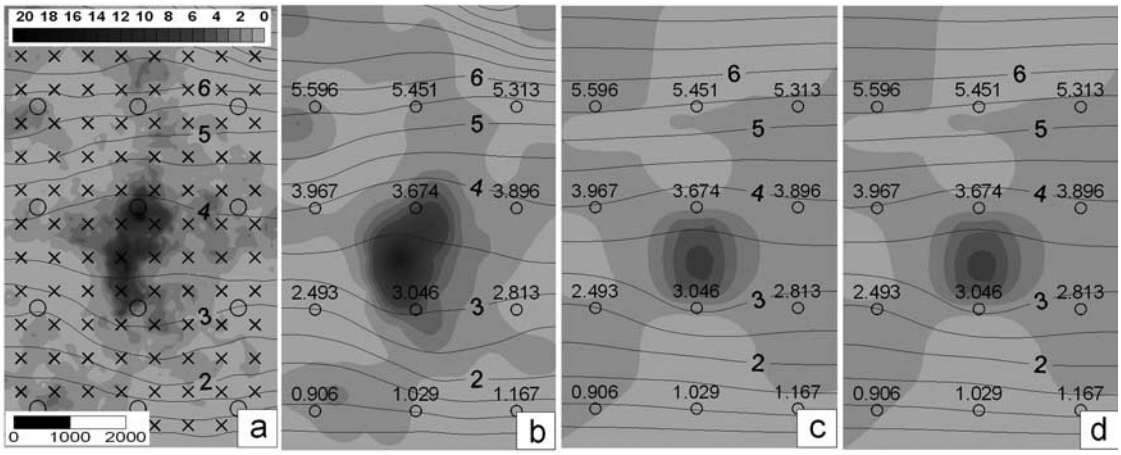

Figure 2. Parameters and observations in synthetic study. (a) Multi-Gaussian field, pilot points (crosses) and observations (circles), (b) true field and simulated heads, (c) Tikhonov field and simulated heads, and (d) hybrid field and simulated heads.

[50] 3. Interpolation of these values to the model nodes from the pilot points was undertaken using kriging with the variogram described by (13) (Figure 2b).

[51] Constructing the true hydraulic conductivity field in this manner ensures that the parameter field respects the defining variogram and that there are ''true values'' for the pilot point parameters. Steady state groundwater flow is simulated using MODFLOW-2000 [*Harbaugh et al.*, 2000], and the solution is used to generate 12 true groundwater elevation observations (Figure 2b). No synthetic noise is added to these observations, i.e., they are assumed to be known exactly. Using these observations hydraulic conductivity was calibrated using two methods: (1) Tikhonov regularized inversion as described by *Doherty* [2003] with each of the 104 pilot points identified as a parameter in the inversion and (2) the hybrid methodology. Following calculation of base parameter sensitivities, 12 super parameters were identified for estimation.

[52] In each case, interpolation of hydraulic conductivity from pilot points to the model nodes was undertaken using kriging based on the variogram described by (13). In each case, smoothing regularization was specified between pilot point pairs, whereby the difference between the logs of hydraulic conductivity values at pairs of pilot points is assigned a preferred value of zero (i.e., similarity) with a relative weight equivalent to the inverse of the square of the variogram value at that separation distance. Using this scheme Φ_g is penalized more severely for differences in values between closely spaced pilot points than for differences in values between more widely spaced pilot points. The initial value for each pilot point was equated to the mean of the true field, i.e., 1.0 m/d, and the target objective function, Φ_t, was set to 1.0×10^{-8} m^2.

[53] In both cases model-to-measurement misfit was reduced to Φ_t. The true hydraulic conductivity field (Figure 2b) is shown together with the fields estimated from the Tikhonov regularized inversion (Figure 2c) and the hybrid regularized inversion (Figure 2d). As would be expected, in neither case is the true hydraulic conductivity field recovered (Moore and Doherty, submitted manuscript, 2005). Given the paucity of observations both regularized inversions recover a muted or smoothed representation of

Table 1. Model Temporal Discretization and Calibration Data

Stress Period Type[a]	Time, days	Description	Calibration Targets
		Flow Model	
SS	0–2920	plume development (no PT remedy in place)	151 time-averaged steady state water levels
SS	2920–3060	simulation of the PT remedy; PT remedy start-up at t = 2920 days	50 time-averaged steady state water levels
		Transport Model	
TR	0–2920	plume development (no PT remedy in place)	351 time-varying MTBE concentrations throughout the plume
TR	2920–3060	simulation of the PT remedy; PT remedy start-up at t = 2920 days	128 time-varying MTBE concentrations at the recovery well

[a]SS, steady state; TR, transient.

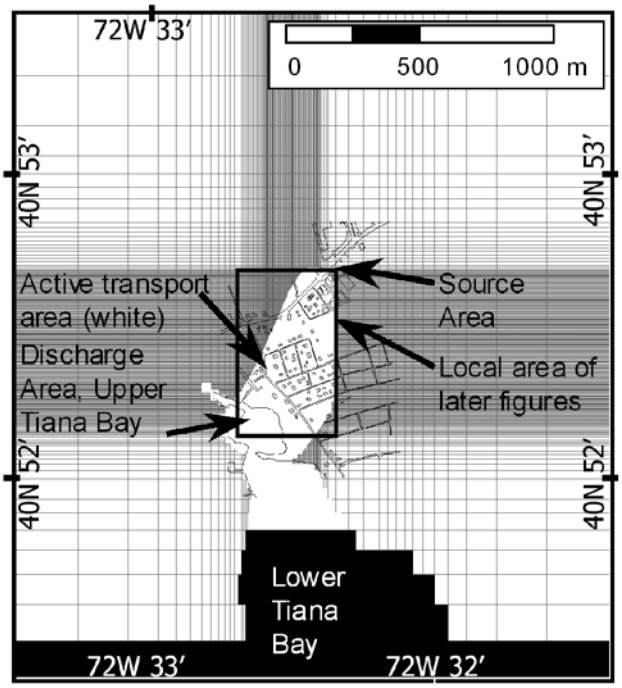

Figure 3. Area of interest and model domain.

the true hydraulic conductivity field. Comparison of Figures 2c and 2d reveals that the two estimated fields are very similar, and in fact the pilot point values estimated through both regularized inversions are essentially identical. However, the Tikhonov inversion required 526 model executions to achieve Φ_t whereas the hybrid inversion required only 209 model executions. Noting that the hybrid method required 105 model executions to calculate the base parameter sensitivity matrix **X**, only 104 model executions were required by the hybrid inversion in order to achieve the same target objective function as the full Tikhonov inversion. This represents a considerable time saving; however, in instances where the ratio of the number of base parameters to super parameters is greater, such as in the real-world problem described next, these savings may be considerably greater.

8. Real-World Application

[54] The hybrid methodology is applied to the inversion of a flow-and-transport model using MODFLOW-2000 and MT3DMS [*Zheng and Wang*, 1999]. The model was selected due to current requirements for predictions and the availability of a detailed calibration data set. The model is under development and no conclusions about its fitness for use should be drawn from the results of this study. Its use as the example model in this paper serves only to demonstrate that the hybrid approach is applicable to large, highly nonlinear models that exhibit lengthy execution times.

9. Site Description

[55] At the Hampton Bays Site, New York, releases from underground storage tanks created a plume of contaminants including the fuel oxygenate methyl-tertiary-butyl-ether (MTBE). The site investigation is described fully online at http://www.fueloxrem.com/ and only details relevant to this study are given here. The geology of the site is characterized by a sequence of highly transmissive glacial outwash sands referred to as the Upper Glacial Aquifer overlying a sequence of fine to medium sands comprising the Magothy Aquifer [*Franke and McClymonds*, 1972; *Soren and Simmons*, 1972]. Extensive characterization by New York State Department of Environmental Conservation (NYSDEC) contractor Environmental Assessment and Remediations (EAR) using 40 multilevel monitoring locations delineated a fuel oxygenate plume migrating through the Upper Glacial toward Tiana Bay, a sensitive saline water body.

[56] A flow-and-transport model was constructed to support the design of a PT interim remedial measure (IRM) comprising a single recovery well. The model simulates steady state flow and transient transport, with first-order decay of MTBE (Table 1). The model is composed of 14 layers, 203 rows, and 100 columns, with refinement of the grid to 5 m by 5 m in the area of interest (Figure 3). Multiple layers are necessary to simulate vertical transport of the plume that migrates 25 m below the water table before discharging at an elevation close to mean sea level. Model boundaries are described by a Cauchy (general head) condition representing upper Tiana Bay and a Dirichlet (specified head) condition representing lower Tiana Bay. Seventeen transport stress periods are simulated, representing 16 semiannual periods during plume development followed by a single stress period during mass recovery at the IRM. Selection of semiannual stress periods is based upon anecdotal observations of time-varying mass releases. The time-variable source concentration is estimated as part of the inverse process.

[57] Initial predictions of interest from the model were the extraction rate necessary to maintain hydraulic containment of contaminants and the expected timing of reductions in the

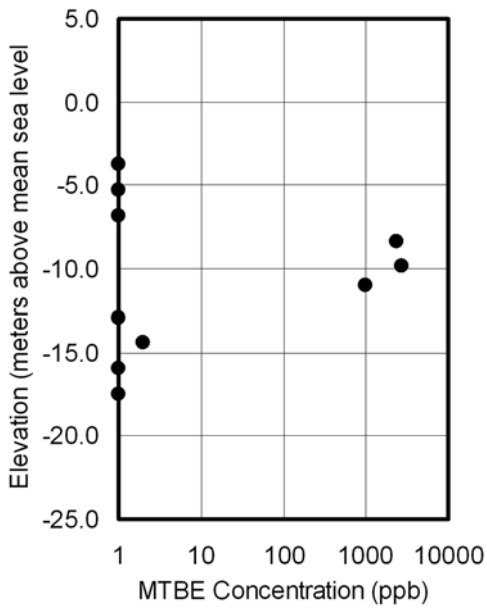

Figure 4. Typical concentration profile: Multilevel well ML-22.

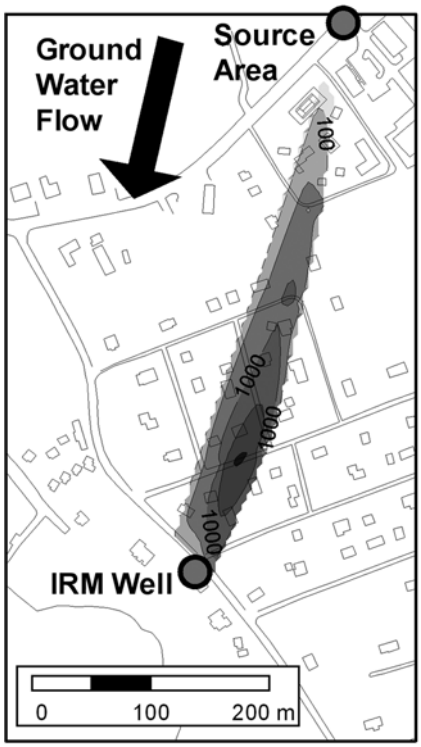

Figure 5. Contoured MTBE concentrations for May 2003 for an elevation of −10 m to −13 m mean sea level (msl).

flux of contaminants to Tiana Bay. Following the installation of a sensitive ex situ biologically activated Granular Activated Carbon (Bio-GAC) treatment system the principal prediction is the mass flux at the recovery well.

[58] Sharp fronts in the MTBE data suggest the problem is advection-dominated. Reported nondetects often lie adjacent to high concentrations at a separation approaching the model node spacing (Figures 4 and 5). Sharp fronted plumes within highly transmissive aquifers create two particular problems that render them poor candidates for highly parameterized calibration. First, the simulated plume can be close to the real plume yet sensitivities with respect to some observations may be zero, since the simulated plume does not encompass these locations. Second, the advection terms of the partitioned equation must be solved using a scheme that can resolve sharp fronts [*Zheng and Bennett*, 2002; *Barth and Hill*, 2005]. Solvers that mitigate numerical dispersion while achieving satisfactory mass balance, such as the Total Variation Diminishing (TVD) scheme used here, have stability constraints that lead to long simulation times.

[59] Some of these difficulties can be mitigated by first developing a model that closely approximates advective flow in the system. This was achieved here by calibrating a parsimonious (over-determined) flow-and-path line model to the estimated center of mass to obtain a starting point for flow-and-transport model calibration [*Pollock*, 1994]. Center of mass observations were determined using moments, by gridding contaminant data at several transects approximately perpendicular to the groundwater flow direction (Figure 6). This centers the simulated plume and provides parameter sensitivities that span the propensity of observations (Figure 7). This is important prior to calculating base parameter sensitivities in the hybrid methodology for reasons already outlined.

10. Observation Data

[60] Table 1 summarizes the calibration data, including water levels, MTBE concentrations in monitoring wells, and MTBE concentrations at the IRM. The importance of including concentrations in the objective function for flow-and-transport model calibration is discussed by *Strecker and Chu* [1986], *Wagner and Gorelick* [1987], *Gailey et al.* [1991], *Franssen et al.* [2003], and *Medina and Carrera* [1996, 2003]. These studies suggest that including observations in the calibration that are of the same type as the model prediction(s) may improve the reliability of the calibrated model. *Bard* [1974], *Sun* [1994], *Hill* [1998], and *Koch* [1999] among others detail the theoretical bases for assigning observation weights. Ideally, $\mathbf{Q}_m$ is proportional

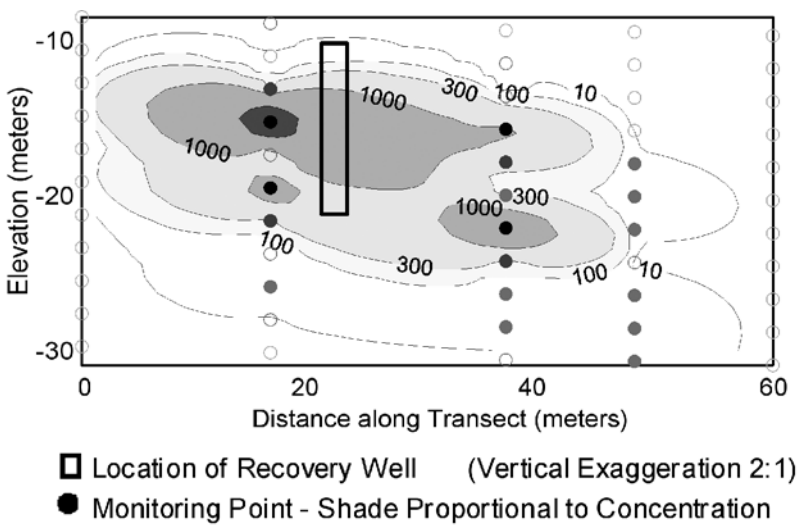

Figure 6. Example transect perpendicular to plume axis used in lumped parameter path line calibration.

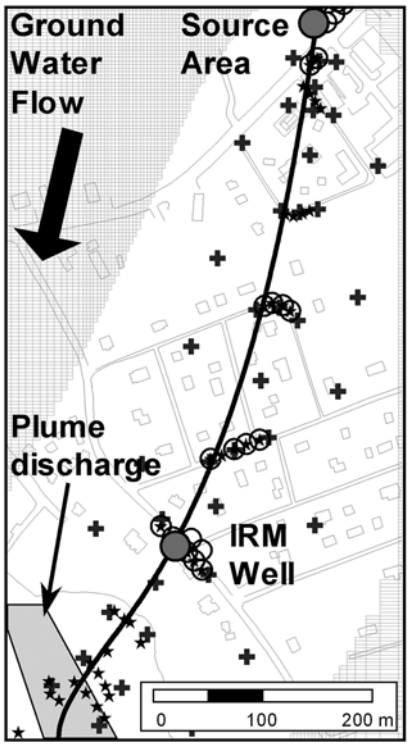

Figure 7. Active area of transport (white) showing observation locations, pilot points, source area, IRM well, and calibrated path line (circles, water level observations; stars, MTBE observations; crosses, pilot points).

to the inverse of the covariance matrix of measurement errors, yet information on the correlation structure of measurement errors is usually unavailable [*Sun*, 1994]. Specifying the weight and/or the observed value for concentration data can also be complicated since it is not evident if values below the reporting limit (RL) are slightly below the RL or represent a true zero outside the footprint of the plume [*Barth and Hill*, 2005].

[61] *Gailey et al.* [1991], *Sun* [1994], *Weiss and Smith* [1998], and *Barlebo et al.* [2004], among others, discuss alternative weighting strategies. *Moore and Doherty* [2005] discuss selective weight assignment for reducing the error variance of model predictions. In the present study weights are assigned so that the contribution to Φ_m by MTBE concentrations at the IRM well and MTBE concentrations in monitoring wells is similar at the commencement of the inverse process. Groundwater elevations are assigned low weights due to uncertainty in surveyed elevations. This is consistent with findings of *Barth and Hill* [2005].

11. Base Model Parameterization

[62] Table 2 summarizes the base parameterization. Parameters from the upper nine layers of the model are estimated, representing the Upper Glacial Aquifer. In total 1,195 base parameters are identified for estimation. This includes parameters assumed to be spatially variable yet constant in time (horizontal (HHK) and vertical (VHK) hydraulic conductivity, porosity (POR), and recharge (RCH)); parameters assumed constant within a model layer and constant in time (general head boundary conductance (GHB), dispersion (LDSP, TDSP), and decay rate (DEG)); and a time-varying source (SCON).

[63] The distribution of the base parameters HHK, VHK, POR and RCH is described using pilot points (Table 2 and Figure 7). Scalars describe the remaining base parameters, i.e., GHB, LDSP, TDSP, DEG and SCON. Interpolation is accomplished using kriging on the basis of an arbitrary exponential variogram. Pilot points can be used in conjunction with parameter zones if ancillary information supports the decision to do so. However, using pilot points ensures that estimated parameter fields are smoothly varying, which is generally consistent with the hydrogeology of the Upper Glacial Aquifer.

[64] Tikhonov regularization is employed for base parameters described by pilot points. In the present study pilot points representing parameters of the same type, within the same layer, are linked by regularization equations. Smoothness of the solution field is promoted using the approach described for the synthetic model above, i.e., by equating parameter differences to zero with weights that decay exponentially with increasing separation distance. Regular-

Table 2. Model Parameterization

Layer(s)	Parameter Type(s)[a]	Parameter Description	Number of Base Parameters
		Flow and Transport Parameters	
1–7	conductivities (HHK, VHK)	55 pilot points for each of HHK, VHK for each layer	770
1	recharge (RCH)	1 scalar for background recharge; 10 pilot points for irrigation/septic	11
1	general head boundary (GHB)	1 scalar for GHB conductance	1
		Transport Parameters	
1–7[b]	porosity (POR)	55 pilot points for each layer	385
1–9	decay (DEG)	1 scalar per active transport layer	9
1–9	dispersion (LDSP, TDSP)	2 scalars	2
1	source term (SCON)	1 scalar per stress period	17
Total			1,195

[a]HHK, horizontal hydraulic conductivity; VHK, vertical hydraulic conductivity.
[b]Note that parameters estimated for layer 7 are applied to layers 8 and 9 for the parameters HHK, VHK, and porosity.

Table 3. Summary of Hybrid Regularized Inversion Results[a]

Observation Group	Completion of Lumped Calibration	Completion of Hybrid Inversion	Percent Reduction
MTBE mass removal	5.27E+06	2.00E+06	62
MTBE concentrations	9.31E+06	1.68E+06	82
Water levels	6.19E+02	3.71E+02	40
Composite objective function	1.46E+07	3.68E+06	74

[a]Read 5.27E+06 as 5.27×10^6.

ization equations are not specified between parameters in different model layers due to the stratified nature of the aquifer materials. Pilot points are focused where observation data exist, with a small number of peripheral pilot points used for interpolation to the remainder of the model domain. This is consistent with the approach described by *RamaRao et al.* [1995] who located pilot points where their potential for reducing the objective function was greatest. In the case of recharge, pilot points are focused in an area of elevated return flows delineated by a consumptive use study, and preferred value regularization is added to stabilize the simultaneous estimation of HHK and RCH. In total, 18,000 regularization equations are specified.

[65] At the site, drilling indicated the presence of a contiguous clay extending beneath the upper Tiana Bay that is absent where incised by fresh water discharge during low tides. This stratigraphy causes the plume to discharge offshore, rather than at the shoreline as might otherwise be expected. However, no pilot points were fixed at preferred values, and no prior information was specified in response to this knowledge. Rather, it was hoped that the inverse process would identify the presence of this low permeability unit through the information content of MTBE concentrations sampled beneath Tiana Bay.

12. Model Inversion

[66] Solution of the transport equations was restricted to the area of interest to reduce the forward execution time below 60 minutes. Perturbation sensitivities were obtained across 24 PCs with typical processor speeds of 2.4 GHz using the parallelization capabilities of PEST [*Doherty*, 2005]. Calibration was completed in steps, in which the model progressed toward more computationally intensive methods to achieve a better fit to the data: (1) calibration of a flow-and-path line model using a subset of the parameters listed in Table 2, and lumping those parameters represented by pilot points (the parameter types and number of parameters used were HHK(1), VHK(1), POR(1), RCH(2), GHB(1), DEG(1) and SCON(17) for a total of 24 parameters), (2) calibration of the lumped parameters to the MTBE concentrations, IRM recovery data, and water levels, i.e., those observations that comprise the measurement objective function Φ_m, (3) completion of the base parameter

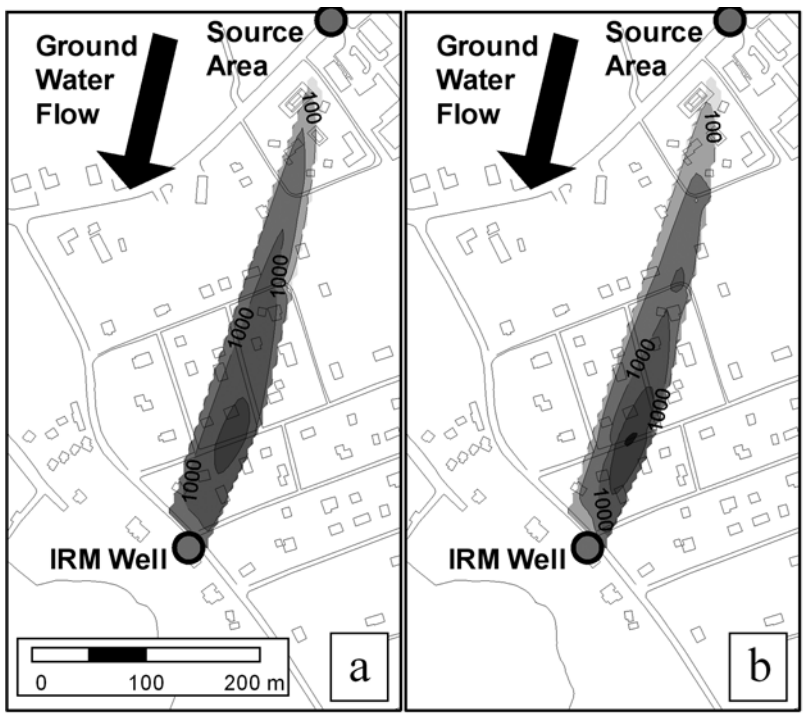

Figure 8. (a) Simulated and (b) observed MTBE concentrations for May 2003 for an elevation of −10 m to −13 m msl.

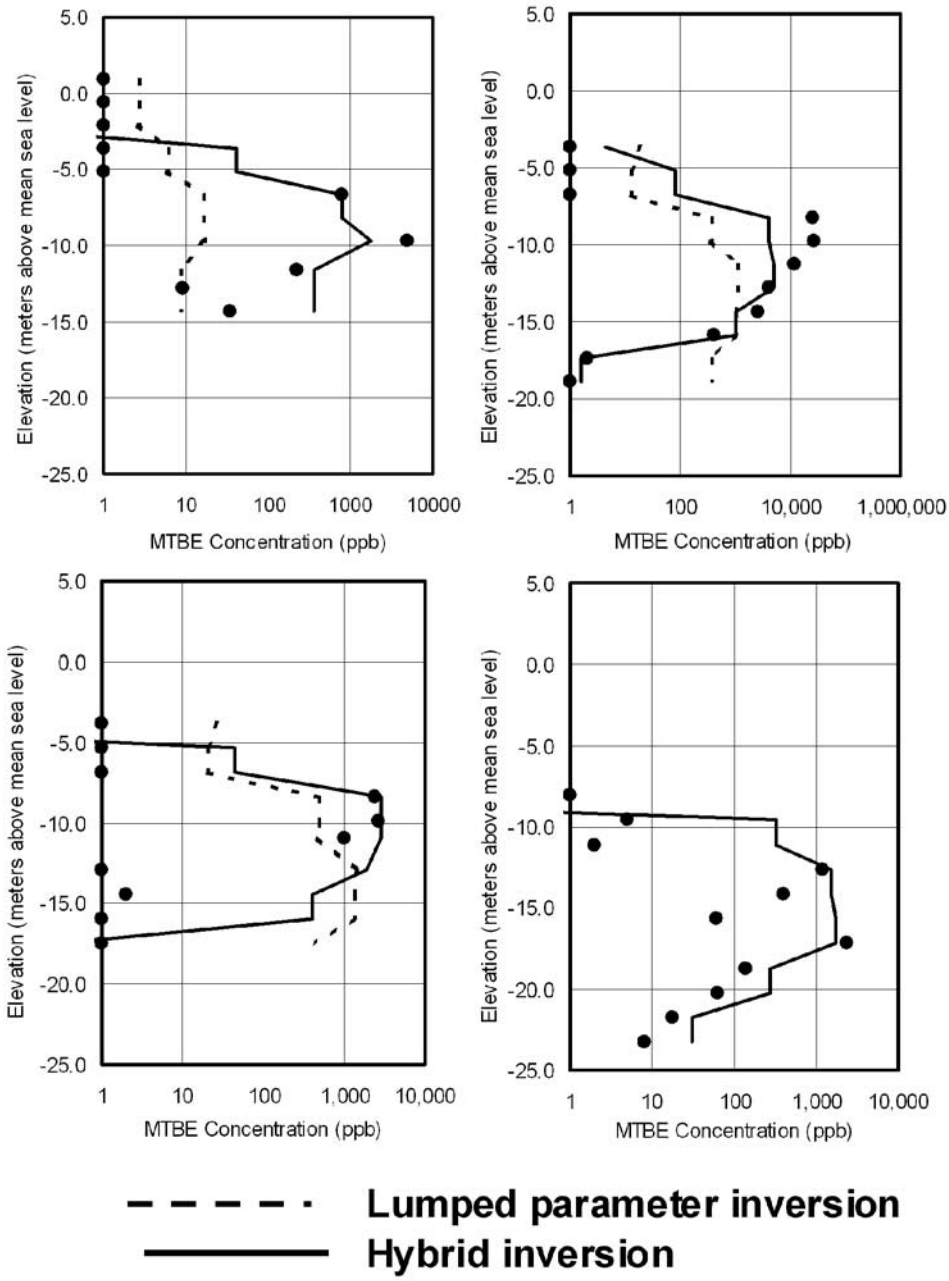

Figure 9. Simulated and observed MTBE concentration profiles: Multilevel wells ML-22, ML-9, ML-20, and ML-25.

sensitivity analysis using initial values for base parameters determined from step 2, and (4) decomposition of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$ into its component eigenvectors and eigenvalues, selection of super parameters for estimation, and commencement of the reformulated inverse problem.

[67] Calculation of $\mathbf{X}$ required 1195 executions of the model and completed in 44 hours. Decomposition of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$ indicated a fairly typical distribution of eigenvalues spanning many orders of magnitude and declining rapidly, suggesting the problem may be dominated by eigenvectors corresponding to about the 20 largest singular values. On the basis of the singular value spectrum and, pragmatically, on the number of PCs and relative processor speeds available for sensitivity calculations, 30 super parameters were defined for estimation. The hybrid inversion using super parameters proceeded for 6 iterations, requiring approximately 190 forward model executions and 9 hours to converge.

13. Results

[68] Comparison of the composite objective function and the contribution to the objective function from the three observation groups at the completion of the parsimonious calibration and hybrid inversion indicates decreases in all observation groups (Table 3). The most significant improve-

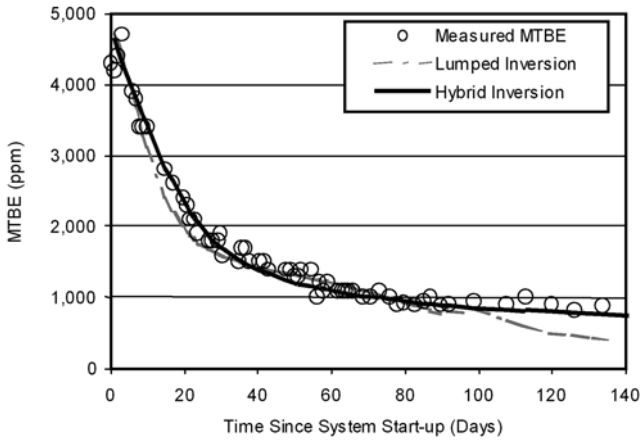

Figure 10. Simulated and observed MTBE concentrations at the IRM well.

ment is in the fit to MTBE concentrations at monitoring wells. This might be expected since the use of pilot points enables the hybrid methodology to introduce heterogeneity in the spatially variable parameters such as hydraulic conductivity and porosity.

[69] The fit is illustrated by visual comparison of simulated and observed concentrations at wells contoured using identical interpolation schemes, representing an elevation of approximately −10 to −13 m, i.e., close to the middle elevation of the plume and the midscreen elevation of the recovery well (Figure 8). Similarly, vertical plots of the data demonstrate reasonable fits, reproducing sharp vertical profiles and showing noticeable improvement with respect to the fit obtained from the lumped calibration (Figure 9) (note that the lumped calibration predicted concentrations below 1 ppb at ML-25). The improvement in fit to MTBE concentrations at the IRM is about 60% (Table 3 and Figure 10). Despite assigning low weights to water level observations, the fit to observed water levels was reasonable.

[70] Perhaps the most illuminating illustration of the improved fit to point data is provided by two-dimensional ''bubble plots'' of simulated and observed MTBE concentrations in monitoring wells (Figure 11). In Figure 11 the source area is located at right, and the discharge area, i.e., Tiana Bay, is located at left. The abscissa is the distance from the source area and the ordinate is the elevation above mean sea level. The area of each circle is proportional to concentration. Figure 11b illustrates the match of simulated and measured MTBE concentrations using the calibrated lumped parameter model. The general path of the plume is matched, i.e., the plume appears centered in the XZ plane. However, the simulated plume gives the appearance of simple spreading about a centerline, and areas of low or high measured concentrations are not reflected in the simulated equivalents. Figure 11c illustrates the match of simulated and measured MTBE concentrations using the optimized base parameters calculated through the hybrid inverse process. Two particular features of this plot are (1) the location of discharge (the hybrid inversion results shows the plume discharging further offshore than in the lumped parameter model) and (2) increased detail (the hybrid inversion identified some areas of higher (and lower) concentrations throughout the plume and is closer to matching these).

[71] Review of the base parameters estimated in the hybrid inversion suggests the improved fit may have been principally obtained by the introduction of spatial variability into hydraulic conductivity and porosity parameters (Table 4). This is expected since using a large number of pilot points together with Tikhonov regularization enables

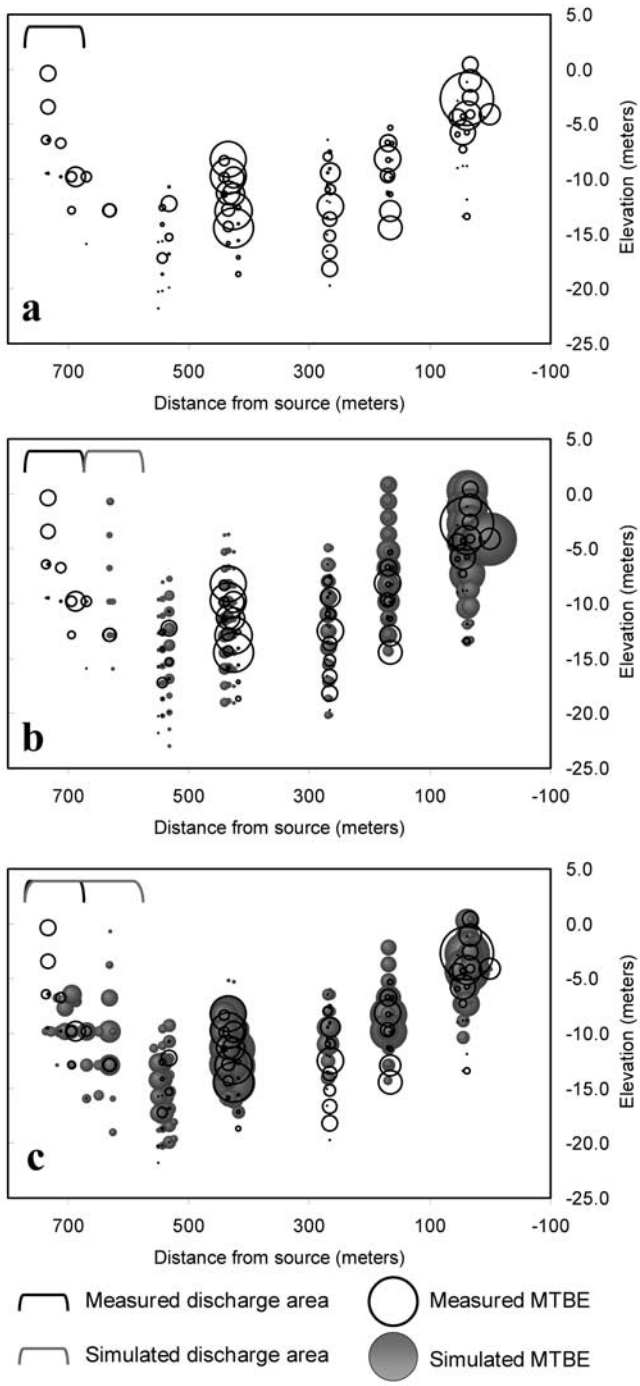

Figure 11. Profile of MTBE concentrations in wells from source area (right) to discharge location in Tiana Bay (left), showing (a) measured MTBE, (b) measured MTBE and simulated MTBE from lumped calibration, and (c) measured MTBE and simulated MTBE from hybrid calibration.

Table 4. Summary of Estimated Parameters Represented by Pilot Points

Parameter	Minimum	Mean	Geomean	Maximum	Comments
HHK pilot points, m/d	16.5	144.5	140.2	1670.[a]	range from slug/aquifer testsa: 33.5–120
VHK pilot points, m/d	1.07	1.43	1.40	2.84	n/a
Recharge pilot points, cm/yr	75.6	101.8	n/a	125.2	average recharge estimated from water use study[b]: 112
Porosity pilot points	0.12	0.21	0.20	0.40	n/a

[a]Note that this high value occurs adjacent to the incised channel area shown on Figure 13 corresponding with free-standing surface water in Tiana Bay. This value approximates the upper bound for clean sands and gravels [*Freeze and Cherry*, 1979].
[b]Environmental Assessment and Remediations (personal communication, 2003).

the hybrid scheme to introduce such variability. Parameter values estimated in the inversion are generally within the range of values estimated from field testing and are not unreasonable (Table 4). In particular the hybrid inversion appears to identify the area of low vertical hydraulic conductivity beneath the Tiana Bay, causing the simulated plume to discharge further offshore than in the lumped calibration (Figure 12).

14. Discussion and Concluding Remarks

[72] The hybrid method is based on the fact that though there may be many base parameters, a limited number of linear combinations of these parameters are estimable, these being determined from the information content of the observations. These combinations are identified and then estimated. Other combinations of parameters are assigned values that satisfy relationships supplied through Tikhonov regularization constraints specified in accordance with the modeler's notion(s) of parameter reasonableness.

[73] Super parameters accelerate model calibration because throughout the inversion process sensitivities are calculated not for individual base parameters but for combinations of these parameters identified as important to obtaining a good fit. This enables more rapid investigation of the parameter space defined by the available observations and current model design. In the real-world application the number of super parameters (30) is not much greater that the number of parameters in the prior parsimonious calibration (24) yet the improvement in fit is noteworthy. Assuming that the hybrid inversion, comprising 30 super parameters, converged to a similar composite objective function as might an inversion of the 1,195 base parameters, the hybrid inversion converged 35–40 times more rapidly than an inverse problem formulated in terms of the base parameters would.

[74] The most notable feature of the hybrid method is that it offers the potential for the regularized inversion of large models, in terms of execution time and parameterization, for which inversion is otherwise computationally prohibitive.

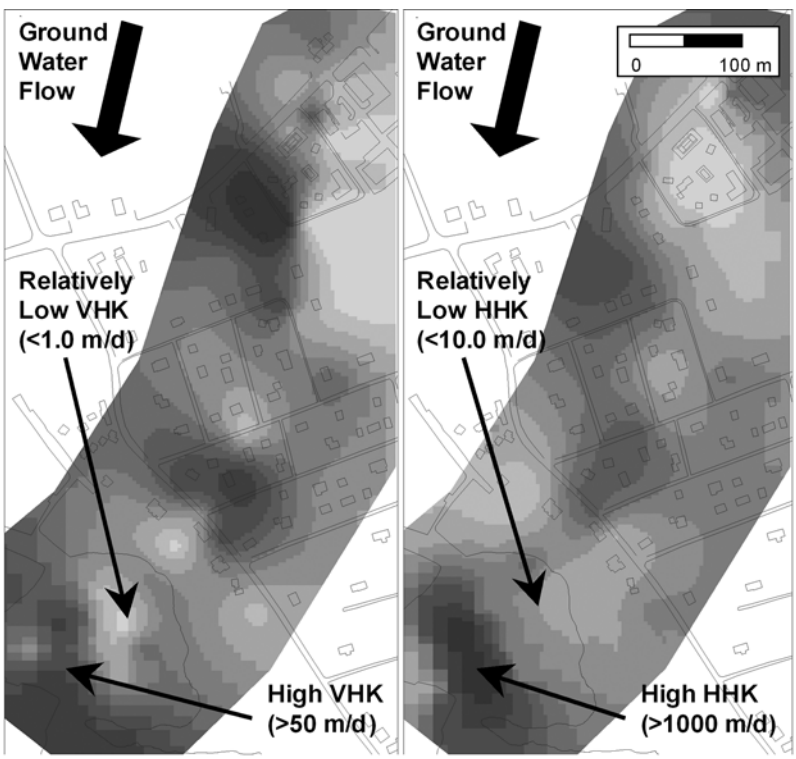

Figure 12. Calibrated vertical and horizontal hydraulic conductivity, model layer 1.

The hybrid method offers rapid inversion times and numerical stability for over-parameterized models and better fits than are achievable with a priori parameter parsimony. As such the capacity to extract information from observation data is greater than achieved through a priori parsimony. Model inversion with a large number of regularized base parameters, estimated using a smaller number of super parameters, may do justice to the increasingly detailed observation data sets available for model calibration.

[75] Since the hybrid method can use perturbation sensitivities it is applicable either where adjoint sensitivities are unavailable or where $m > n$ and adjoint sensitivities are not computationally beneficial. Where $m < n$ using perturbation sensitivities may still require fewer simulations than an adjoint approach, since the number of significant singular values of $\mathbf{X}^t\mathbf{Q}_m\mathbf{X}$, and hence the number of super parameters, may be less than the number of observations. Alternatively, where adjoint sensitivities are available and $m < n$, an adjoint approach could be used to form the base parameter Jacobian matrix when adjoint methods are the most rapid. In the reformulated problem where the subspace is defined by $k < m$, perturbation sensitivities could be employed. This would provide the most computationally efficient sensitivity calculation strategy for the hybrid scheme.

[76] *Carrera* [1993] among others points out that as the scale of simulated heterogeneity approaches the scale of true heterogeneity, mechanical dispersion approaches the scale of molecular dispersion. Noting that in the study model the longitudinal dispersivity estimated in the hybrid inversion (0.1 m) is smaller than that estimated in the lumped parameter calibration (0.14 m) it is interesting to consider if this occurs because one aspect of the field dispersion mechanism, i.e., heterogeneity, is better represented in the hybrid approach. Given the model discretization and the smooth parameter fields estimated using Tikhonov regularization, the model certainly does not capture small-scaled mechanical dispersion mechanisms; however, it may be less inhibited from doing so than parsimonious approaches.

[77] **Acknowledgments.** The authors thank Mary C. Hill, Jasper Urugt, one anonymous reviewer, and the Associate Editor Warren Bond for their contributions to the content and format of the manuscript. The authors also thank Mary C. Hill for bringing the unpublished works of E. Jacobson to their attention. 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 approach described in this study has been documented and is encapsulated in a set of programs referred to as "SVD Assist" supplied as part of the PEST suite, available for download from www.sspa.com/pest.

References

Anderson, E., et al. (1999), *LAPACK Users' Guide*, 3rd ed., 407 pp., Soc. for Ind. and Appl. Math., Philadelphia, Pa.

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

Bard, J. (1974), *Nonlinear Parameter Estimation*, 341 pp., Elsevier, New York.

Barlebo, H. C., M. C. Hill, and D. Rosbjerg (2004), Investigating the Macrodispersion Experiment (MADE) site in Columbus, Mississippi, using a three-dimensional inverse flow and transport model, *Water Resour. Res.*, *40*, W04211, doi:10.1029/2002WR001935.

Barth, G. R., and M. C. Hill (2005), Numerical methods for improving sensitivity analysis and parameter estimation of virus transport simulated using sorptive-reactive processes, *J. Contam. Hydrol.*, *76*, 251–277.

Carrera, J. (1993), An overview of uncertainties in modeling groundwater solute transport, *J. Contam. Hydrol.*, *13*, 23–48.

Carrera, J., and S. P. Neuman (1986), 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., F. Navarrina, L. Vives, J. Heredia, and A. Medina (1990), Computational aspects of the inverse problem, in *Computational Methods in Subsurface Hydrology: Proceedings of the Eighth International Conference on Computational Methods in Water Resources*, *Venice, Italy*, pp. 513–522, Springer, New York.

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.

Cooley, R. L. (1982), Incorporation of prior information on parameters into nonlinear regression groundwater flow models: 1. Theory, *Water Resour. Res.*, *18*(4), 965–976.

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

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), PEST version 9 users guide, Watermark Numer. Comput., Brisbane, Queensl., Australia.

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.

Franssen, H.-J. H., 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.

Freeze, R. A., and J. A. Cherry (1979), *Groundwater*, 604 pp., Prentice-Hall, Upper Saddle River, N. J.

Gailey, R. M., S. M. Gorelick, and A. S. Crowe (1991), Coupled process parameter estimation and prediction uncertainty using hydraulic head and concentration data, *Adv. Water Resour.*, *14*(5), 301–313.

Haber, E. (1997), Numerical strategies for the solution of inverse problems, Ph.D. thesis, Univ. of B. C., Vancouver, B. C., Canada.

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.

Hill, M. C. (1998), Methods and guidelines for effective model calibration, *U.S. Geol. Surv. Water Resour. Invest. Rep., 98-4005*, 90 pp.

Jackson, D. D. (1972), Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. R. Astron.*, 97–109.

Jacobson, E. A. (1985), Estimation method using singular value decomposition with application to Avra Valley Aquifer in southern Arizona, Ph.D. thesis, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson.

Koch, K.-R. (1999), *Parameter Estimation and Hypothesis Testing in Linear Models*, 333 pp., Springer, New York.

Lawson, C. L., and J. R. Hanson (1995), *Solving Least Squares Problems*, *Classics Appl. Math.*, vol. 15, 337 pp., Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Levenberg, K. (1944), A method for the solution of certain non-linear problems in least squares, *Q. Appl. Math*, *2*, 164–168.

Lines, L. R., and S. Treitel (1984), Tutorial: A review of least-squares inversion and its application to geophysical problems, *Geophys. Prospect.*, *32*, 159–186.

Liu, C., and W. P. Ball (1999), Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware, *Water Resour. Res.*, *35*(7), 1975–1985.

Mackie, R. L., and T. R. Madden (1993), Three-dimensional magnetotelluric inversion using conjugate gradients, *Geophys. J. Int.*, *115*, 215–229.

Marquardt, D. W. (1963), An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.*, *11*(2), 431–441.

Medina, A., and J. Carrera (1996), Coupled estimation of flow and solute transport parameters, *Water Resour. Res.*, *32*(10), 3063–3076.

Medina, A., and J. Carrera (2003), Geostatistical inversion of coupled problems: Dealing with computational burden and different types of data, *J. Hydrol.*, *281*(3), 251–264.

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.

Neuman, S. P., and S. Yakowitz (1979), A statistical approach to the inverse problem of aquifer hydrology: 1. Theory, *Water Resour. Res.*, *15*(4), 845–860.

Ory, J., and R. G. Pratt (1995), Are our parameter estimators biased? The significance of finite-difference regularization operators, *Inverse Probl.*, *11*, 397–424.

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.

Portniaguine, O. N. (1999), Image focusing and data compression in the solution of geophysical inverse problems, Ph.D. thesis, Dep. of Geol. and Geophys., Univ. of Utah, Logan.

Portniaguine, O. N., and M. S. Zhdanov (2002), 3-D magnetic inversion with data compression and image focusing, *Geophysics*, *67*(5), 1532–1541.

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

Seber, G. A. F., and C. J. Wild (1989), *Nonlinear Regression*, 792 pp., John Wiley, New York.

Shewchuk, J. R. (1994), An introduction to the conjugate gradient method without the agonizing pain, internal report, Sch. of Comput. Sci., Carnegie Mellon Univ., Pittsburgh, Pa.

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*.

Strecker, E. W., and W. Chu (1986), Parameter identification of a groundwater contaminant transport model, *Ground Water*, *24*(1), 56–62.

Sun, N.-Z. (1994), *Inverse Problems in Groundwater Modeling*, 352 pp., Springer, New York.

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

Townley, L. R., and J. L. Wilson (1985), Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models of groundwater flow, *Water. Resour. Res.*, *21*(12), 1851–1860.

Tsourlos, P. I., and R. D Ogilvy (1999), An algorithm for the 3-D inversion of tomographic and induced polarization data: Preliminary results, *J. Balkan Geophys. Soc.*, *2*(2), 30–45.

Vogel, C. R. (1997), Non-smooth regularization, in *Inverse Problems in Geophysical Applications*, pp. 1–11, edited by H. W. Engl, A. K. Louis, and W. Rundell, Soc. for Ind. and Appl. Math., Philadelphia, Pa.

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

Wagner, B. J., and S. M. Gorelick (1987), Optimal groundwater management under parameter uncertainty, *Water Resour. Res.*, *23*(7), 1162–1174.

Weiss, R., and L. Smith (1998), Parameter space methods in joint parameter estimation for groundwater flow models, *Water Resour. Res.*, *34*(4), 647–661.

Xiang, Y., J. F. Sykes, and N. R. Thomson (1993), A composite L1 parameter estimator for model fitting in groundwater flow and solute transport simulation, *Water Resour. Res.*, *29*(6), 1661–1674.

Zheng, C., and G. D. Bennett (2002), *Applied Contaminant Transport Modeling* (2nd ed.), 621 pp., John Wiley, Hoboken, N. J.

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 Eng. Res. and Dev. Cent., Vicksburg, Miss.

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

M. J. Tonkin, S.S. Papadopulos and Associates, Inc., 120 Main Street, Route 6A, Yarmouth Port, MA 02675, USA. (matt@sspa.com)