A representer-based inverse method for groundwater flow and transport applications

Johan R. Valstar

Netherlands Institute of Applied Geoscience TNO–National Geological Survey, Utrecht, Netherlands

Dennis B. McLaughlin

Ralph M. Parsons Laboratory, Massachusetts Institute of Technology Cambridge, Massachussetts, USA

Chris B. M. te Stroet[1] and Frans C. van Geer

Netherlands Institute of Applied Geoscience TNO–National Geological Survey, Utrecht, Netherlands

Received 2 December 2003; revised 26 February 2004; accepted 30 March 2004; published 28 May 2004.

[1] Groundwater inverse problems are concerned with the estimation of uncertain model parameters, such as hydraulic conductivity, from field or laboratory measurements. In practice, model and measurement errors compromise the ability of inverse procedures to provide accurate results. It is important to account for such errors in order to determine the proper weight to give to each source of information. Probabilistic descriptions of model and measurement errors can be incorporated into classical variational inverse procedures, but the computational demands are excessive if the model errors vary over time. An alternative approach based on representer expansions is able to efficiently accommodate time-dependent errors for large problems. In the representer approach, unknown variables are expanded in finite series which depend on unknown functions called representers. Each representer quantifies the influence of a given measurement on the estimate of a particular variable. This procedure replaces the original inverse problem by an equivalent problem where the number of independent unknowns is proportional to the number of measurements. The representer approach is especially advantageous in groundwater problems, where the total number of measurements is often small. This approach is illustrated with a synthetic flow and transport example that includes time-dependent model errors. The representer algorithm is able to provide good estimates of a spatially variable hydraulic conductivity field and good predictions of solute concentration. Its computational demands are reasonable, and it is relatively easy to implement. The example reveals that it is beneficial to account for model errors even when they are difficult to estimate. *INDEX TERMS:* 1832 Hydrology: Groundwater transport; 1829 Hydrology: Groundwater hydrology; 1869 Hydrology: Stochastic processes; 1831 Hydrology: Groundwater quality; *KEYWORDS:* calibration, groundwater hydrology, groundwater transport, inverse modeling, representers, uncertainty

Citation: Valstar, J. R., D. B. McLaughlin, C. B. M. te Stroet, and F. C. van Geer (2004), A representer-based inverse method for groundwater flow and transport applications, *Water Resour. Res.*, *40*, W05116, doi:10.1029/2003WR002922.

1. Introduction

[2] Combined groundwater flow and transport models are frequently used to predict the movement and fate of subsurface contaminants. These models are prone to a number of errors that need to be considered in applications. First, some of the physical and biochemical processes included, such as biodegradation and adsorption, are not well understood and are frequently represented in a simplified way. Second, the models rely on poorly known inputs such as hydraulic conductivity, porosity, dispersivities, adsorption coefficients, and recharge values. Third, approximate upscaling procedures are generally used to represent the aggregate or large-scale effects of small-scale processes that cannot be explicitly resolved. Examples of spatial upscaling are macrodispersion theory [*Gelhar and Axness*, 1983] and the concept of dual porosity [*Harvey and Gorelick*, 2000], which account in somewhat different ways for enhanced spreading caused by geological heterogeneity. Although some model errors are time-invariant others, particularly those associated with groundwater recharge, solute mass fluxes, and upscaling, can vary over time.

[3] Time-dependent model errors can be accounted for probabilistically in various stages of the modeling process [*Schweppe*, 1973]. This helps to make model predictions, parameter estimates, and related uncertainty analyses more realistic since model limitations are acknowledged. Here we are particularly interested in using stochastic

[1] Also at Faculty of Civil Engineering, Delft University of Technology, Delft, Netherlands.

0043-1397/04/2003WR002922$09.00

descriptions of model errors to improve the solutions to parameter estimation (inverse) problems. Conventional inverse methods impose excessive computational demands for problems with time-dependent model errors since the problem dimensionality increases without limit as time increases. In this paper we describe a computationally efficient variational inverse procedure that accounts for time-dependent model errors. Our discussion begins with a general probabilistic problem formulation. We then introduce an inverse solution procedure based on representer expansions. This procedure is illustrated with a synthetic example of coupled groundwater flow and transport. We conclude with an assessment of the method's potential usefulness for field applications.

2. Formulation of the Inverse Problem

2.1. Forward Model

[4] Here we consider the problem of time-dependent solute transport in a time-invariant (steady state) velocity field. This problem is typically solved with numerical models based on finite element or finite difference discretizations of the governing flow and transport equations. The steady state discretized groundwater model may be written in a general matrix form as:

$$Ah - q = 0 \tag{1}$$

where A is a coefficient matrix that depends on the hydraulic conductivity at the nodes of the spatial grid, h is a vector of nodal heads, and q is a vector of corresponding forcing terms that depend on the flow boundary conditions. The time-dependent discretized groundwater transport model may be written as:

$$\begin{aligned} Bc^t - Dc^{t-1} - u^t = 0;\ t = 1, 2, \ldots, N_t \\ c^0 = C_{init} \end{aligned} \tag{2}$$

where B and D are coefficient matrices that depend on velocities and transport properties such as dispersivities and sorption coefficients, c^t is a vector of nodal solute concentrations at time t, u^t is a vector of corresponding forcing terms that depend on the transport boundary conditions, C_{init} is a vector of initial concentrations, and N_t is the number of time steps [*Pinder and Gray*, 1977]. The velocities included in B and D are derived from the heads and conductivities by means of Darcy's law. We suppose that all of the unknown parameters (hydraulic conductivities, dispersivities, etc.) used to derive A, B, D, and C_{init} are assembled in a time-invariant vector α. This is the parameter vector that we wish to estimate with the inverse procedure.

2.2. Accounting for Model Errors

[5] In order to account for model errors we must distinguish the model from reality. One way to do this is to suppose that the true heads and concentrations obey equations similar to equations (1) and (2), but with random variables added to account for imperfectly modeled inputs. Here we assume that the model errors act as additive forcing terms, similar in effect to unmodeled recharge or solute mass fluxes. The flow model errors are assumed to be time-invariant, as our flow model is time-invariant, while the transport model errors are time-dependent. This formulation of model errors can be represented mathematically if we suppose that the true heads and concentrations at the grid nodes obey the following stochastic state equations:

$$Ah - q - w_h = 0 \tag{3}$$

$$Bc^t - Dc^{t-1} - u^t - w_c^t = 0; \quad t = 1, 2, \ldots, N_t \tag{4}$$

$$c^0 = C_{init}$$

where w_h are the random flow equation model errors and w_c^t are the random transport equation model errors over time step t [*Schweppe*, 1973]. The head model errors are assumed to have a zero mean and a known covariance matrix P_h.

[6] When the true concentration approaches zero the transport model errors are likely to be smaller than when the concentration is large. To account for such state-dependent model errors we suppose that w_c^t has the form:

$$w_c^t = Y\left(c^{t-1}\right)\epsilon^t \tag{5}$$

where $Y(c^{t-1})$ is a diagonal matrix with diagonal element i proportional to the *ith* element of the c^{t-1} vector and ϵ^t are random errors with a zero mean and a known time-invariant covariance matrix P_ϵ. The values of ϵ^t at different times are assumed to be uncorrelated.

2.3. Measurements

[7] Inverse procedures derive estimates of uncertain model parameters from measurements of observable variables such as head and solute concentration. In the problem considered here the measurements are assumed to be interpolated values of the true nodal head and concentration values with random errors added. This is expressed mathematically by the following stochastic measurement equation:

$$z = M(h, c) + v \tag{6}$$

where z is the vector of all measurements taken in the region of interest over a time period spanning N_t time steps and c (with no superscript) is an extended vector that contains the true concentrations at all times. The interpolation function (or measurement operator) $M(h, c)$ describes how each measurement is related to the head and extended concentration vectors. In the groundwater problem considered it is reasonable to assume that $M(h, c)$ is linear, as the model prediction of either a head or a concentration measurement is linearly related to the nodal heads or concentrations respectively. The random measurement errors v are assumed to have a zero mean and known covariance matrix P_v. The vector v and the matrix P_v contain the measurement errors and their covariances for the measurements at all time steps.

2.4. Objective Function

[8] With the state and measurement equations defined we can provide a statement of the inverse problem. We suppose that the parameter vector α is random with a known mean $\bar{\alpha}$ and covariance matrix P_α. These statistics summarize our prior knowledge about the parameter values (if the parameters are completely unknown the values in the prior parameter covariance matrix approach infinity). We seek an "optimal" estimate of the parameter vector α as well as

the model errors included in the vectors w_h and ϵ^t. If the errors and parameters included in the state and measurement equations are independent Gaussian random vectors the maximum a posteriori estimate is an attractive option. This estimate minimizes the following generalized least squares objective function [*Schweppe*, 1973]:

$$J = [z - M(h,c)]^T P_v^{-1}[z - M(h,c)] + (\alpha - \bar{\alpha})^T P_\alpha^{-1}(\alpha - \bar{\alpha}) + w_h^T P_{w_h}^{-1} w_h + \sum_{t=1}^{N_t} \epsilon^{tT} P_\epsilon^{-1} \epsilon^t \quad (7)$$

[9] The first term in this expression penalizes the misfit between the measurement and model prediction vectors. This term is a function of α, w_h and ϵ as the model predictions h and c depend on these parameters and model errors. The second term penalizes deviations of the model parameter vector from its prior mean. The third and fourth terms penalize the flow and transport model errors. The relative importance of each term is determined by the covariances, which act like weighting factors.

[10] In the synthetic example in this paper, we assume the covariance matrices P_v, P_α, P_{w_h} and P_ϵ and the prior mean of the parameters $\bar{\alpha}$ are known. For real world problems obtaining these statistical values is not a straightforward task. In practice, they are often based on the subjective choice of the modeler and are likely to differ from the actual statistical values.

[11] The generalized least squares objective function is very difficult to minimize when the number of parameters is large, especially when the minimization algorithm requires inversion of nondiagonal covariance matrices. Gradient-based search procedures such as conjugate gradient are the most commonly used minimization options. The gradients required by these procedures can be efficiently derived with a Lagrange multiplier approach (also commonly referred to as an adjoint or variational approach). Unfortunately, the computational effort can still be substantial for large problems since convergence can be slow. Moreover, for large size inverse problem the storage and inversion of the full matrices $P\alpha$, P_{w_h} and P_ϵ, needed for these gradient methods, is often too demanding. The number of time-invariant parameters can be reduced by using techniques such as zonation [*Cooley*, 1977; *Sun*, 1994; *Cooley*, 1982; *Carrera and Neuman*, 1986]. However, if time-dependent model errors are included in the problem formulation the total number of unknowns increases with time and the classical variational approach quickly becomes unmanageable. The representer technique described in the next section provides a computationally acceptable alternative for situations where time-dependent model errors are important.

3. Solution of the Inverse Problem

[12] The inverse problem outlined above can be solved with an extension of the representer algorithm, a least squares minimization procedure that has been extensively applied in oceanography [*Bennett*, 1992; *Eknes and Evensen*, 1995]. Hydrologic applications have been presented by *Reichle et al.* [2001], *Reid* [1996] and *Sun* [1998]. The representer method introduces an expansion that reduces the number of unknown parameters to the number of measurements. This expansion is exact when the state and measurement equations are linear and often provides a useful approximation when these equations are nonlinear. The method is most readily explained by first considering the necessary conditions for a local minimum.

3.1. Euler-Lagrange Equations

[13] The minimium of the generalized least squares objective function must be derived subject to the constraint that the state equations are satisfied. This constraint can be incorporated into the minimization procedure if the state equations are multiplied by unknown Lagrange multipliers and added (or adjoined) to the objective function. The necessary condition for a constrained local minimum requires that the derivatives of this adjoined objective, taken with respect to the unknown states (h and c), Lagrange multipliers (λ_c and λ_h), parameters (α) and model errors (w_h and ϵ^t), must all be zero. When the individual derivatives are computed and set equal to zero the result is a set of nonlinear coupled equations called the Euler-Lagrange equations. A short derivation is given in Appendix A; a complete derivation is provided by *Valstar* [2001].

[14] The Euler-Lagrange equations naturally group into two adjoint (or Lagrange multiplier) equations, one for transport and one for flow, a parameter equation, and two forward equations, one for transport and one for flow. The forward equations incorporate estimates for the model errors. The complete set may be written in indicial notation as follows:

$$B_{ji}\lambda_{c_i}^t = D_{ji}\lambda_{c_i}^{t+1} + \epsilon_m^{t+1}\frac{\partial Y_{mi}(c^t)}{\partial c_j^t}\lambda_{c_i}^{t+1} + \frac{\partial M_p(h,c)}{\partial c_j^t}\left[P_v^{-1}\right]_{pn} \cdot [z_n - M_n(h,c)]; \quad t = N_t - 1, N_t - 2, \ldots, 1 \quad (8)$$

$$B_{ji}\lambda_{c_i}^{N_t} = \frac{\partial M_p(h,c)}{\partial c_j^{N_t}}\left[P_v^{-1}\right]_{pn}[z_n - M_n(h,c)] \quad (9)$$

$$A_{gf}\lambda_{h_f} = \frac{\partial M_p(h,c)}{\partial h_g}\left[P_v^{-1}\right]_{pn}[z_n - M_n(h,c)] - \sum_{t=1}^{N_t}\left[c_j^t\frac{\partial B_{ji}}{\partial h_g} - c_j^{t-1}\frac{\partial D_{ji}}{\partial h_g}r\right]\lambda_{c_i}^t \quad (10)$$

$$\alpha_l = \bar{\alpha}_l - P_{\alpha_{lk}}\left[h_g\frac{\partial A_{gf}}{\partial \alpha_k}\lambda_{h_f} + \sum_{t=1}^{N_t}\left(c_j^t\frac{\partial B_{ji}}{\partial \alpha_k} - c_j^{t-1}\frac{\partial D_{ji}}{\partial \alpha_k}\right)\lambda_{c_i}^t\right] \quad (11)$$

$$A_{fg}h_g = q_f + w_{h_f} \quad (12)$$

$$w_{h_f} = P_{w_{h_{fd}}}\lambda_{h_d} \quad (13)$$

$$B_{ij}c_j^t = D_{ij}c_j^{t-1} + u_i^t + Y_{ir}\left(c^{t-1}\right)\epsilon_r^t; \quad t = 1, 2, \ldots, N_t \quad (14)$$

$$c_j^0 = C_{init_j} \quad (15)$$

$$\epsilon_r^t = P_{\epsilon_{rm}}Y_{ms}\left(c^{t-1}\right)\lambda_{c_s}^t; \quad t = 1, 2, \ldots, N_t \quad (16)$$

where d, f, and g range from 1 to the number of head state variables; i, j, m, r, and s range from 1 to the number of concentration state variables; k and l range from 1 to the number of uncertain parameters; and n and p range from 1 to the number of measurements. Indices repeated within a

single product term are assumed to be summed over appropriate ranges. Different indices used for the same variables (such as the indices n and p in $M(h, c)$ in equation (8)) only denote how matrix-matrix or matrix-vector multiplications should be performed.

[15] Note that equations (12) and (14) have the same form as the flow and transport state equations (3) and (4), with the true model errors replaced by estimates that depend on the head and concentration adjoint variables, respectively. The adjoint variables determine how the model error estimates are distributed over space and/or time while the model error covariances control their magnitude. When the error covariances or adjoint variables are zero the corresponding model error estimates are also zero. To simplify notation in the remainder of this paper we use the symbols h, c, λ_h, λ_c, α, w_h and ϵ to represent the best available approximate solution to the Euler-Lagrange equations, rather than the unknown true values.

[16] The Euler-Lagrange equations form a two-point boundary value problem, with the boundary values given by the initial condition of the transport equation and the terminal condition of the transport adjoint equation. The solution may not be unique since it only defines a local (rather than global) minimum of the least squares objective. The likelihood of multiple solutions decreases when the prior covariances of the parameters and model errors are small and the quadratic prior terms in the least squares function that contain the parameters and model errors, are given more weight. Techniques for solving two-point boundary value problems include the sweep method [*Gelfand and Fomin*, 1963; *Bryson and Ho*, 1975; *Bennett*, 1992] and the representer method [*Bennett*, 1992]. In this paper, we use an extension of the representer method of Bennett.

3.2. Representer Solution

[17] The representer method writes each of the unknowns in the Euler-Lagrange equations as an expansion in a set of unknown basis functions called representers. The number of terms in each representer expansion is the total number of measurements. For our problem the representer expansions have the following form:

$$\lambda_{c_i}^t = \sum_{p=1}^{N_z} \Phi_{ip}^t b_p \tag{17}$$

$$\lambda_{h_f} = \sum_{p=1}^{N_z} \Gamma_{fp} b_p \tag{18}$$

$$\alpha_l = \bar{\alpha}_l + \sum_{p=1}^{N_z} \Psi_{lp} b_p \tag{19}$$

$$h_g = h_{F_g} + h_{corr_g} + \sum_{p=1}^{N_z} \Xi_{gp} b_p \tag{20}$$

$$c_j^t = c_{F_j}^t + c_{corr_j}^t + \sum_{p=1}^{N_z} \Omega_{jp}^t b_p \tag{21}$$

where b_p is an unknown representer coefficient that determines the weight given to the representer associated with measurement z_p, Φ_{ip}^t is the concentration adjoint representer, Γ_{fp} is the head adjoint representer, Ψ_{lp} is the parameter representer, Ξ_{gp} is the head representer, Ω_{jp}^t is the concentration representer, h_{F_g} and $c_{F_j}^t$ are prior head and concentration estimates, h_{corr_g} and $c_{corr_j}^t$ are correction terms which allow the head and concentration expansions to be taken about the latest estimates, and N_z is the total number of measurements. The prior estimates h_F and c_F^t are the solutions obtained by solving equations (1) and (2) with $\alpha = \bar{\alpha}$.

[18] The representer expansions can be shown to be exact solutions to the Euler-Lagrange equations when the original optimization problem has constraints that are linear in all decision variables. In this case the representers are simply the cross covariances between each unknown variable and each measurement, the prior head and concentration estimates are the mean values, and the correction terms are zero. So, for example, Ξ_{gp} is the cross covariance between the head h_g and the measurement z_p. In our problem the linear constraint requirement is not met and the representer expansions are only approximations. However, a representer such as Ξ_{gp} can still be viewed as an influence function that specifies how the unknown head h_g should be modified when the measurement z_p differs from the nominal model predictions. In this sense, the representer approach provides a very flexible parameterization of the inverse problem. The spatial variation of each unknown depends on the governing equation, its forcing terms, and specified prior information. This is in contrast to methods that assume each unknown has a fixed spatial structure, such as a piecewise constant function.

[19] In a linear constraint problem the representer expansions can be substituted into the Euler-Lagrange equations and closed-form expressions obtained for the representers and their coefficients. In the nonlinear problem of interest here an iterative solution approach must be used since obtaining the closed-form solutions is not possible. The basic concept is to approximate each of the nonlinear functions $A(\alpha)$, $B(h, \alpha)$, $D(h, \alpha)$, and $Y(c)$ by a first-order Taylor series expansion taken round the most recently computed estimates of h, c, and α. The resulting linearized (and approximate) Euler-Lagrange equations can be solved, the unknowns are updated, and the linearization is repeated until the iteration procedure converges. After convergence the Euler-Lagrange equations are solved exactly and the parameter and model error estimates are the solution of the original optimization problem.

[20] When the representer definitions are substituted into the linearized Euler-Lagrange equations they decouple the equation set, making it possible to derive a sequential set of expressions for (1) the representers, (2) the correction terms, and (3) the representer coefficients (a short derivation is given in Appendix B; an extensive derivation is given by *Valstar* [2001]). In any given iteration the values of h, c, α, λ_h, λ_c and ϵ appearing in an update equation are the ones computed in the previous iteration.

[21] The algorithm is initialized with the "initial guess" $h = h_F$, $c^t = c_F^t$, $\alpha = \bar{\alpha}$, $\lambda_h = 0$, and $\lambda_c^t = 0$. Then, the

updated representers are obtained using the following formula:

$$B_{ji}\Phi_{ip}^{t} = D_{ji}\Phi_{ip}^{t+1} + \frac{\partial M_p(h,c)}{\partial c_j^t} + \epsilon_m^{t+1}\frac{\partial Y_{mi}(c^t)}{\partial c(t)}\Phi_{ip}^{t+1}; \quad t = N_t - 1, N_t - 2, \ldots, 1 \tag{22}$$

$$B_{ji}\Phi_{ip}^{N_t} = \frac{\partial M_p(h,c)}{\partial c_j^{N_t}} \tag{23}$$

$$A_{gf}\Gamma_{fp} = \frac{\partial M_p(h,c)}{\partial h_g} - \sum_{t=1}^{N_t}\left[c_j^t\frac{\partial B_{ji}}{\partial h_g} - c_j^{t-1}\frac{\partial D_{ji}}{\partial h_g}\right]\Phi_{ip}^t \tag{24}$$

$$\Psi_{lp} = -P_{\alpha_l k}\left[h_g\frac{\partial A_{gf}}{\partial \alpha_k}\Gamma_{fp} + \sum_{t=1}^{N_t}\left(c_j^t\frac{\partial B_{ji}}{\partial \alpha_k} - c_j^{t-1}\frac{\partial D_{ji}}{\partial \alpha_k}r\right)\Phi_{ip}^t\right] \tag{25}$$

$$A_{fg}\Xi_{gp} = -\frac{\partial A_{fg}}{\partial \alpha_k}\Psi_{kp}h_g + [P_{w_h}]_{fd}\Gamma_{dp} \tag{26}$$

$$B_{ij}\Omega_{jp}^t = D_{ij}\Omega_{jp}^{t-1} + Y_{ir}\left(c^{t-1}\right)[P_\epsilon]_{rm}Y_{ms}\left(c^{t-1}\right)\Phi_{sp}^t + \frac{\partial Y_{ir}(c^{t-1})}{\partial c_j^{t-1}}\Omega_{jp}^{t-1}\epsilon_r^t - \left[\frac{\partial B_{ij}}{\partial \alpha_k}\Psi_{kp} + \frac{\partial B_{ij}}{\partial h_g}\Xi_{gp}\right]c_j^t + \left[\frac{\partial D_{ij}}{\partial \alpha_k}\Psi_{kp} + \frac{\partial D_{ij}}{\partial h_g}\Xi_{gp}\right]c_j^{t-1}; \quad t = 1, 2, \ldots, N_t \tag{27}$$

$$\Omega_{jp}^0 = 0 \tag{28}$$

[22] Then the updated head and concentration corrections are obtained from the following equations:

$$A_{fg}h_{corr_g} = q_f + \frac{\partial A_{fg}}{\partial \alpha_k}(\alpha_k - \bar{\alpha}_k)h_g - A_{fg}h_{F_g} \tag{29}$$

$$B_{ij}c_{corr_j}^t = D_{ij}c_{corr_j}^{t-1} + u_i^t + \left[\frac{\partial B_{ij}}{\partial \alpha_k}(\alpha_k - \bar{\alpha}_k) + \frac{\partial B_{ij}}{\partial h_g}\left(h_g - h_{F_g} - h_{corr_g}\right)\right]c_j^t - \left[\frac{\partial D_{ij}}{\partial \alpha_k}(\alpha_k - \bar{\alpha}_k) + \frac{\partial D_{ij}}{\partial h_g}\left(h_g - h_{F_g} - h_{corr_g}\right)\right]c_j^{t-1} - B_{ij}c_{F_j}^t + D_{ij}c_{F_j}^{t-1} - \frac{\partial Y_{ir}(c^{t-1})}{\partial c_j^{t-1}}\left(c_j^{t-1} - c_{F_j}^{t-1} - c_{corr_j}^{t-1}\right)\epsilon_r^t; \quad t = 1, 2 \ldots, N_t \tag{30}$$

$$c_{corr_j}^0 = 0 \tag{31}$$

[23] Then the updated representer coefficient vector is obtained by solving the following equation:

$$\left[P_{v_{np}} + M_n\left(\Xi_p, \Omega_p\right)\right]b_p = [z_n - M_n(h_F + h_{corr}, c_F + c_{corr})] \tag{32}$$

[24] After the representers, corrections, and representer coefficient vector are found they may be substituted into equations (17)–(21) to give updated values for the adjoint variables and the parameter vector. Alternatively, the adjoint updates may be obtained by substituting equation (32) into equations (8)–(10). The parameter updates can then be derived from equation (11). The latter approach uses less memory but requires somewhat more computational effort. In either case, the updated head and concentration are computed directly from equations (12)–(16). The A, B, D, and Y derivatives required for the next iteration are evaluated at these updated values. The computational effort is dominated by the computation of the representers. For each measurement, the effort is approximately equal to solving two flow and transport models. For a large number of measurements, the computational effort becomes considerable.

4. Computational Issues

4.1. Eigenvalue Decomposition

[25] In the algorithm outlined above the coefficient vector b is obtained by solving the set of linear equations given in equation (32). When measurements are highly correlated this matrix equation is ill-conditioned and the solution is very sensitive to small changes in the measurement residuals on the right hand side. We resolve this problem by replacing the original matrix equation by a reduced rank approximation based on the leading eigenvectors of the original coefficient matrix. However, in groundwater inverse problems concentration measurements near zero often have small variances. A straightforward reduced rank approximation tends to ignore such small variance measurements, even though they convey important information about the boundaries of the solute plume. The influence of small variance measurements can be retained if the rows and columns of the measurement residual covariance matrix ($P_v + M(\Xi, \Omega)$) are rescaled before the eigenvalue decomposition is performed. The scale factors are selected to give values of unity on the main diagonal of this matrix. In a synthetic example described by *Valstar* [2001] rescaling gives much better predictions near the low-concentration edges of the solute plume.

4.2. Convergence

[26] We assume that convergence is achieved when the residuals between the state variables, obtained from equations (12) and (14), and the linear representer expansions, obtained from equations (20) and (21), are smaller than a threshold value for each measurement. For this threshold value we generally use 1% of the measurement error standard deviation. There is no guarantee that the iterative representer solution to the Euler-Lagrange equations converges. In practice, it is helpful to introduce a relaxation factor so that each new parameter and model error estimate is a weighted combination of the estimate from the previous iteration and the estimate derived from the representer expansion. The optimal relaxation factor is adjusted to minimize the mean squared scaled difference between the head and concentration estimates obtained from the state equations (12)–(16) and the corresponding relaxed estimates obtained from the representer expansions (20) and (21). Each difference is scaled by its standard deviation. The relaxation adjustment procedure can be viewed as a line search along a search direction determined by the representer approximation. We cannot guarantee that the method always converges for highly nonlinear estimation problems.

In our experience so far, we did not encounter any convergence problems.

4.3. Accuracy Estimates

[27] After convergence the accuracy of the parameter and state estimates can be inferred from approximate posterior (conditional) covariances, which are based on linearizations of the state equations. These are the covariances of the parameters and states, conditioned on the measurements used in the inverse procedure. The posterior variances are never larger than the prior (unconditional) variances and should be significantly smaller if the measurements provide useful information about the estimated variables. The approximate posterior covariances of the parameters are given by:

$$P^{post}_{\alpha_{kl}} = P_{\alpha_{kl}} - \Psi_{kp}\left((M(\Omega,\Xi)+P_v)^{-1}\right)_{pq}\Psi_{lq} \qquad (33)$$

The approximate posterior covariances of the states are

$$P^{post}_{h_{fg}} = P_{h_{fg}} - \Xi_{fp}\left((M(\Omega,\Xi)+P_v)^{-1}\right)_{pq}\Xi_{gq} \qquad (34)$$

$$P^{post,t_1,t_2}_{c_{ij}} = P^{t_1,t_2}_{c_{ij}} - \Omega^{t_1}_{ip}\left((M(\Omega,\Xi)+P_v)^{-1}\right)_{pq}\Omega^{t_2}_{jq} \qquad (35)$$

The posterior parameter variances (the diagonals of the covariance matrix $P^{post}_{\alpha_{kl}}$) can be readily calculated since the prior parameter covariance $P_{\alpha_{kl}}$ is known. However, the posterior variances of the states depend on the prior covariances $P_{h_{fg}}$ and $P^{t_1,t_2}_{c_{ij}}$, which can be readily calculated only at the measurement locations. A significant amount of additional computational effort is required to obtain the complete prior covariances of the states. In our example we only consider the posterior variances of the parameters and states. The latter required considerable additional computations.

4.4. Comparison to Other Nonlinear Inverse Methods

[28] The primary factors that influence the computational requirements of nonlinear least squares inverse methods are the number of forward model simulations needed to determine the minimization search step and the number of search steps required to converge to an acceptable solution. The number of forward simulations per step in the representer algorithm is proportional to the number of scalar measurements used in the inversion. This is comparable to a Gauss-Newton or iterative cokriging algorithm that relies on measurement sensitivity derivatives derived from coupled forward and adjoint model solutions. By contrast, a conjugate gradient search algorithm using adjoint-based derivatives requires only one forward and one adjoint simulation per search step. The number of steps required by any of these search algorithms generally increases with the number of unknown parameters (i.e., the dimension of the search space). Gauss-Newton and iterative cokriging algorithms must treat model errors at each time step as unknown parameters. Consequently, the total number of parameters (and the number of iterations required for convergence) increases as the length of the simulation interval increases. The representer method does not require the model errors to be included in the parameter vector. Instead, they are derived directly from the adjoint vector associated with the corresponding model state. In our experience, the number of iterations required for the representer method to converge is no greater when time-dependent model errors are included than when it is neglected. This is a significant benefit in practical groundwater applications, where recharge rates and/or contaminant source histories are uncertain.

4.5. Applicability of the Inverse Method for Various Measurement Sets

[29] The representer method is applicable for both densely and sparsely filled measurement sets. For densely filled measurements sets, some measurements are likely to be strongly correlated (short distance in space and time). This correlation shows up during the eigenvalue decomposition of the measurement residual covariance matrix (P_v + $M(\Xi, \Omega)$). For sparsely filled measurements sets, some measurements may have only a weak correlation with all other measurements. The representer coefficient b of that measurement is obtained almost independently of the other measurement values. For regions with no measurements at all its parameters are hardly updated, but their uncertainty does not reduce either.

5. Example

[30] We can now illustrate the performance of our representer-based inverse method with a synthetic example that includes time-dependent model errors. Such an example is useful for initial tests since the true answer is known and errors can be evaluated unambiguously. Of course, additional full-scale tests with real field data are needed to adequately assess the practical applicability of any inverse procedure.

[31] Our test problem considers mobile and immobile solute transport in a vertical two-dimensional section with recharge occurring on the top boundary (surface), as shown in Figure 1. A spatially distributed mass of solute enters a source area on the surface at time $t = 0$ and is subsequently partitioned into mobile and immobile fractions. Mobile solute moves with the prevailing steady state flow field (from right to left) and out the left boundary. Mass transfer between the mobile and immobile fractions at any time and location is proportional to the difference between mobile and immobile concentrations [*Brusseau and Rao*, 1989; *Harvey and Gorelick*, 2000]. There are no flow model errors but mobile and immobile transport model errors account for deviations from the idealized linear mass transfer model.

[32] The flow state equation (3) is obtained by discretizing the following partial differential equation:

$$\frac{\partial}{\partial x_i}\left[K_{ij}\frac{\partial h(x)}{\partial x_j}\right] = 0 \qquad \mathrm{x} \in D \qquad (36)$$

with boundary conditions:

$$h(x) = 0 \qquad \mathrm{x} \in \partial D_h$$

$$K_{ij}(x)\frac{\partial h(x)}{\partial x_i}n_i(x) = q_b(x) \qquad \mathrm{x} \in \partial D_q$$

where: $h(x)$ is the hydraulic head at location x in the computational domain D, $q_b(x)$ is the specified boundary

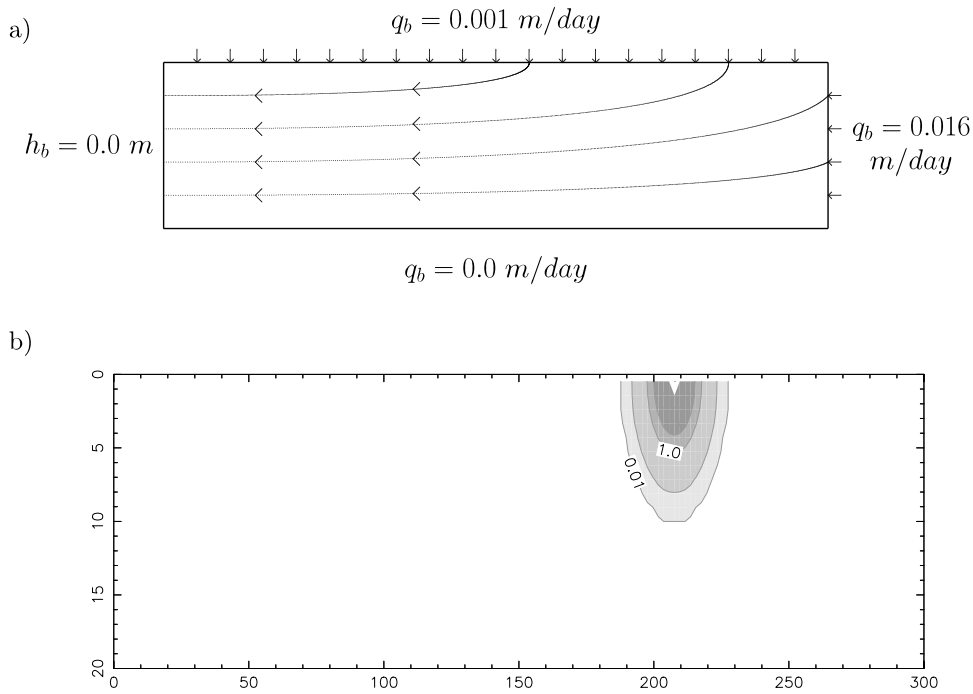

Figure 1. (a) Model area and flow boundary conditions and (b) initial concentration contours of 0.01, 0.1, 1.0, 2.0 and 5.0 g L^{-1}.

flux discharge rate per unit area, $K_{ij}(x)$ is the hydraulic conductivity (assumed to be isotropic in our example), ∂D_h and ∂D_q are the specified head and specified flow portions of the domain boundary ∂D, and $n_i(x)$ is the outward unit vector normal to the boundary ∂D_q.

[33] The concentration state equation (4) is obtained by discretizing the following partial differential equations, which describe transport in the mobile and immobile portions of the porous medium, respectively:

$$\frac{\partial c_m(x,t)}{\partial t} + u_i(x)\frac{\partial c_m(x,t)}{\partial x_i} - \frac{\partial}{\partial x_i}\left[D_{ij}\frac{\partial c_m(x,t)}{\partial x_j}\right] + \frac{f_r(x)}{\theta_m} \cdot [c_m(x,t) - c_{im}(x,t)] = c_m(x,t)\epsilon_m(x,t) \quad x \in D \tag{37}$$

$$\frac{\partial c_{im}(x,t)}{\partial t} - \frac{f_r(x)}{\theta_{im,e}}[c_m(x,t) - c_{im}(x,t)] = c_{im}(x,t)\epsilon_{im}(x,t) \quad x \in D \tag{38}$$

with initial conditions:

$$c_m(x,t) = c_{m_0}(x) \quad t = 0 \quad x \in D$$

$$c_{im}(x,t) = c_{im_0}(x) \quad t = 0 \quad x \in D$$

and boundary conditions:

$$c_m(x,t) = 0 \qquad \text{on } \partial D \text{ for inflow boundaries}$$

$$\frac{\partial c_m(x,t)}{\partial x_i} n_i(x) = 0 \qquad \text{on } \partial D \text{ for outflow boundaries}$$

where $c_m(x, t)$ is the mobile solute concentration at location x and time t, $c_{im}(x, t)$ is the immobile solute concentration, $u_i(x, t)$ is the pore water velocity (obtained from Darcy's law). $D_{ij}(x)$ is the dispersion coefficient, which is computed from the Scheidegger relationship using constant dispersivities, $f_r(x)$ is the exchange rate between the mobile and immobile fractions. θ_m is the mobile porosity, $\epsilon_m(x, t)$ is the mobile fraction of the model errors, $\theta_{im,e}$ is the effective immobile porosity and $\epsilon_{im}(x, t)$ is the immobile fraction of the model errors.

[34] The spatial domain used in our example has 151 by 41 nodes and uses finite rectangular elements of 2.0 by 0.5 m each. In the inverse procedure, the parameters estimated are the hydraulic conductivity and exchange rate coefficient (which vary over space) and the effective immobile porosity (which is assumed to be spatially uniform). The groundwater flow boundary conditions for the example are shown in Figure 1a. These conditions create a nonuniform flow moving generally from right to left, with a downward component near the upper right corner of the domain. Hydraulic heads are measured in m and solute concentrations are measured in g L^{-1}. The initial mobile concentration distribution is shown in Figure 1b. The initial immobile concentration is zero everywhere. The simulation consists of 120 time steps of 5 days each.

[35] The spatially variable ln conductivity and ln exchange rate coefficient used in the synthetic experiment have mean values of ln(1) and ln(0.01) and variances of 2.0 and 1.0, respectively, where the conductivity is measured in m day^{-1} and the exchange rate is measured in day^{-1}. Both parameters have exponential covariance functions with correlation lengths of 50 m and 2 m in the horizontal and vertical directions, respectively. The simulated ln conductivity field is shown in Figure 2a. The natural logarithm of the effective immobile porosity is a unitless spatially uniform random parameter with mean ln(0.1) and variance of 1. The synthetic realization of the effective

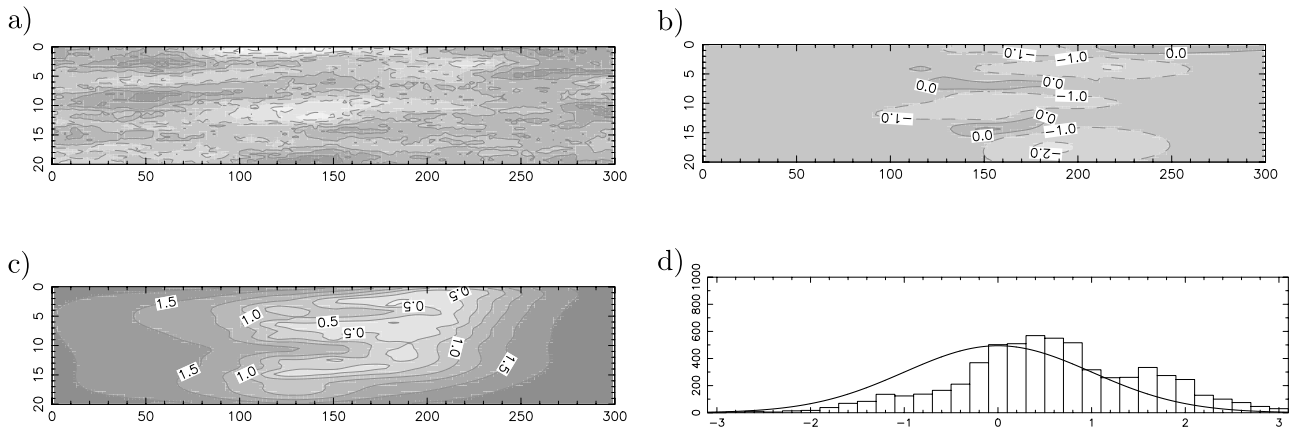

Figure 2. (a) Generated (''true'') hydraulic conductivities, (b) posterior conductivities, (c) approximate posterior conductivity variance, and (d) histogram of the normalized conductivity residuals. Figures 2a and 2b have the same legend.

immobile porosity is 0.13. The other model parameters are assumed to be perfectly known. The horizontal and transversal dispersivities are 0.5 m and 0.05 m. The effective mobile porosity is 0.2.

[36] The concentration model errors $\epsilon_m(x, t)$ and $\epsilon_{im}(x, t)$ are independent random zero mean fields uncorrelated over time. Their spatial covariance functions are exponential with a variance of 10^{-4} and correlation lengths of 50 m in the horizontal direction and 2 m in the vertical direction. Note that these errors are multiplied by their respective concentrations in the transport state equations.

[37] The reference (or ''true'') head and mobile concentration distributions after 120 time steps are shown in Figures 3a and 4a. Prior estimates for the hydraulic conductivity, exchange rate coefficients and immobile porosity are set equal to the specified mean values. Values of the prior heads and mobile concentrations after 120 time steps are shown in Figures 3b and 4b, respectively.

[38] Estimates of all uncertain parameters are derived from 4 noisy head measurements and 185 noisy mobile concentration measurements. The immobile concentrations are not measured. The locations of the head measurements are shown in Figure 3c, and the locations of the concentration measurements are shown in Figure 4c. The measurement errors are zero mean, with a standard deviation of 10^{-2} m for the head measurements and 10% of the real concentration value for the concentration measurements. After 10, 20 and 30 time steps, only the 10 most upstream wells are monitored; after 40, 50, 60, 70 and 80 time steps the 15 most upstream are monitored, and after 90, 100, 110 and 120 time steps all wells are monitored.

[39] The estimated ln conductivity field is shown in Figure 2b. The associated posterior (conditional) variance is shown in Figure 2c. The variance plot indicates that the ln conductivity variance is reduced from 2.0 to less than 0.5 in some locations. The normalized differences (difference between real and estimated ln conductivity divided by the square root of the posterior variance) are shown in Figure 2d. If the ln conductivity estimates are optimal (i.e., the posterior variance values are consistent with reality) these normalized differences should have a mean near 0.0, a

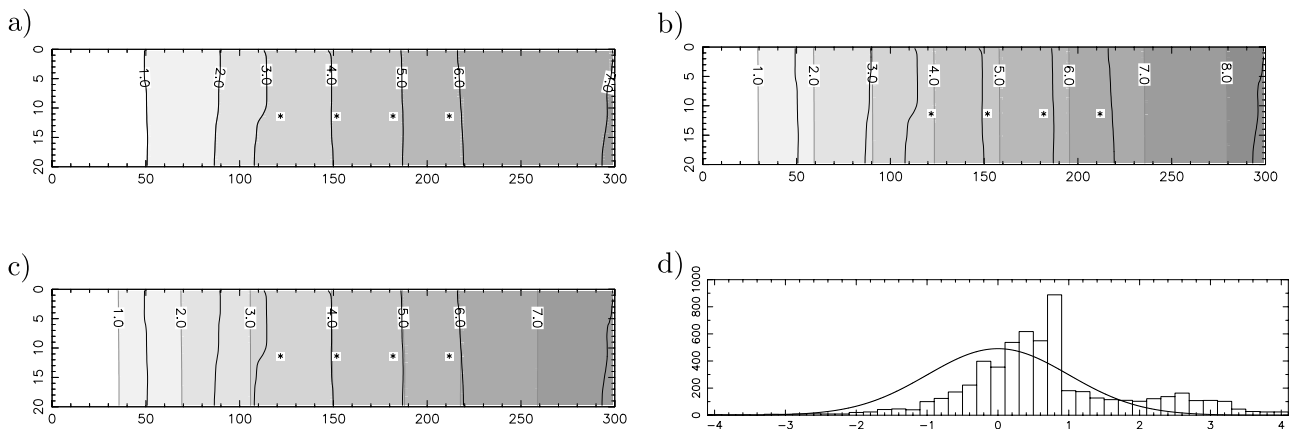

Figure 3. (a) Generated (''true'') hydraulic heads, (b) prior hydraulic heads, (c) posterior hydraulic heads, and (d) histogram of the normalized hydraulic head residuals. Hydraulic head measurement locations are denoted by asterisks in Figures 3a–3c. Solid lines in Figures 3b and 3c denote the true hydraulic head contours from Figure 3a.

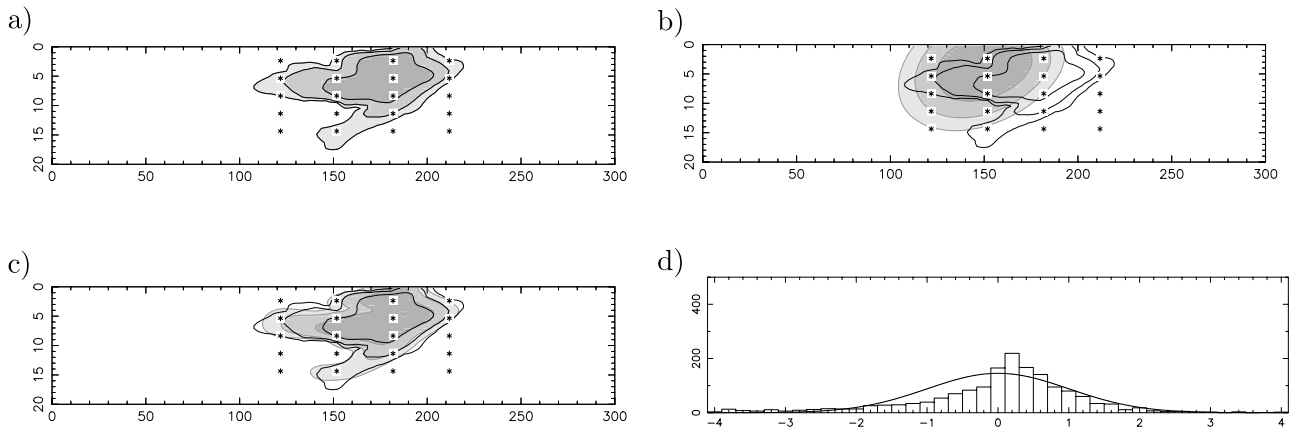

Figure 4. (a) Generated (“true”) solute concentrations, (b) prior concentrations, (c) posterior concentrations, and (d) histogram of the normalized concentration residuals after 120 time steps. Concentration measurement locations are denoted by asterisks in Figures 4a–4c. Solid lines in Figures 4b and 4c denote the true concentration contours from Figure 4a. Contours in Figures 4a–4c denote 0.001, 0.01, and 0.1 g L^{-1}.

variance near 1.0 and a normal probability distribution. In this example, the distribution of the normalized residuals of the posterior ln conductivity field shows a bias. The ln exchange rate coefficients showed minor changes for the posterior estimate and relatively little variance reduction after estimation, suggesting that the available measurements do not contain much information about this particular parameter. The immobile porosity estimate improved from 0.1 to 0.109.

[40] The posterior head and mobile concentration distributions are shown in Figures 3c and 4c. They show a clear improvement when compared to the prior results. The normalized head and concentration residuals are shown in Figures 3d and 4d. In order to calculate these residuals the posterior variances of the heads and concentrations were calculated. For the concentration residuals only the nodes with a prior concentration variance larger than 10^{-10} g^2 L^{-2} was taken into account to prevent numerical inaccuracy effects. The normalized concentration residuals seem normally distributed, but the head residuals show a clear positive bias. We believe that this bias is due to the strong spatial correlation between the head residuals. Figure 3c shows that in both the left and the right side of the domain the head estimates are consistently higher than the true values.

[41] Considering that the posterior variance of the ln conductivity in the plume area is relatively large (0.5–1.0), it is remarkable that the predicted concentration estimates are so close to the true values. This suggests that the concentration values in this problem are relatively insensitive to small-scale fluctuations in conductivity.

[42] It is difficult to obtain accurate estimates of model errors from the relatively small number of measurements used in this example. However, the recognition that such errors are present reduces the estimator’s dependence on the model and gives the measurements more influence than they would otherwise have. In Figure 5 the posterior estimate of the mobile concentration after 120 time steps is shown in case model errors are neglected in the inverse model. The results are inferior in the region (150, 10) when compared to the posterior estimate when model errors are taken into account; see Figure 4c. So there is a definite benefit to acknowledge the presence of model errors even when available measurements are too sparse to allow such errors to be identified.

6. Conclusions

[43] The representer-based inverse method introduced in this paper provides an efficient way to solve inverse problems that include uncertain model errors. The method’s efficiency is achieved by converting the original inverse problem to an equivalent problem where the number of

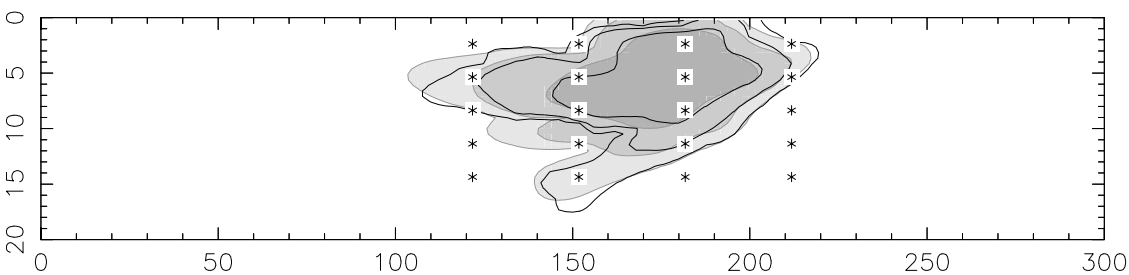

Figure 5. Posterior concentrations after 120 time steps in case model errors are neglected. Concentration measurement locations are denoted by asterisks. Solid lines denote the true concentration contours. Contours denote 0.001, 0.01, and 0.1 g L^{-1}.

independent unknowns is proportional to the number of measurements. The solution obtained with the representer-based inverse method is equal to the original large-scale inverse problem which has a number of unknown parameters proportional to the number of grid nodes. This is especially advantageous in groundwater problems, where the total number of measurements is often relatively small.

[44] The explicit incorporation of model errors allows the algorithm to properly balance information from measurements and model predictions. If the measurement error variances are small compared to the model error variances the measurements will be given more weight. If the opposite is true the model predictions will be given more weight. The balanced approach to errors taken in the representer formulation highlights the need to quantify the accuracy of both model and data in inverse applications. In practice it is unrealistic to assume that either model predictions or measurements are perfect. Other inversion procedures commonly lack the flexibility to balance the uncertainty of parameters, model errors, and measurement errors.

[45] Although our representer-based inverse algorithm is complicated to derive it is relatively easy to implement. The representers and adjoint variables are obtained by solving equations similar in structure to the original flow and transport equations. In most cases the matrix derivatives can be derived in closed form and the iterative solution algorithm converges quickly. This may not always be the case, especially if the original problem is highly nonlinear. Overall, the algorithm's performance and computational efficiency are good enough to encourage applications to field problems of realistic size. Such applications will undoubtedly provide better understanding of the method's advantages and limitations.

Appendix A: Derivation of the Euler-Lagrange Equations

[46] The Euler-Lagrange equations are derived by multiplying the flow equation (3) with 2 times the head adjoint vector λ_h and multiply the transport equation (4) with 2 times the concentration adjoint vector λ_c^t and take the summation of the last product over all time steps and add them to the objective function (7):

$$J = [z - M(h,c)]^T P_v^{-1}[z - M(h,c)] + (\alpha - \bar{\alpha})^T P_\alpha^{-1}(\alpha - \bar{\alpha}) + w_h^T P_{w_h}^{-1} w_h + \sum_{t=1}^{N_t} \epsilon^{tT} P_\epsilon^{-1} \epsilon^t + 2\lambda_h^T [Ah - q - w_h] + 2\sum_{t=1}^{N_t} \lambda_c^{tT} \left[Bc^t - Dc^{t-1} - u^t - Y\left(c^{t-1}\right)\epsilon^t\right] \quad \text{(A1)}$$

[47] Taking the variation of this equation yields:

$$\begin{aligned}\frac{1}{2}\Delta J = \sum_{t=1}^{N_t-1} \Bigg[& -[z_n - M_n(h,c)]\left[P_v^{-1}\right]_{np} \frac{\partial M_p(h,c)}{\partial c_j^t} + \lambda_{c_i}^t B_{ij} - \lambda_{c_i}^{t+1} D_{ij} - \lambda_{c_i}^{t+1} \frac{\partial Y_{im}(c^t)}{\partial c_j^t} \epsilon_m^{t+1} \Bigg] \Delta c_j^t \\ & + \left[-[z_n - M_n(h,c)]\left[P_v^{-1}\right]_{np} \cdot \frac{\partial M_p(h,c)}{\partial c_j^{N_t}} + \lambda_{c_i}^{N_t} B_{ij} \right] \Delta c_j^{N_t} \\ & + \Bigg[\lambda_{h_f} A_{fg} + \sum_{t=1}^{N_t} \lambda_{c_i}^t \left(\frac{\partial B_{ij}}{\partial h_g} c_j^t - \frac{\partial D_{ij}}{\partial h_g} c_j^{t-1} - [z_n - M_n(h,c)] \cdot \left[P_v^{-1}\right]_{np} \frac{\partial M_p(h,c)}{\partial h_g} \right) \Bigg] \Delta h_g + \Bigg[(\alpha_l - \bar{\alpha}_l)\left[P_\alpha^{-1}\right]_{lk} \\ & + \sum_{t=1}^{N_t} \lambda_{c_i}^t \left(\frac{\partial B_{ij}}{\partial \alpha_k} c_j^t - \frac{\partial D_{ij}}{\partial \alpha_k} c_j^{t-1} \right) + \lambda_{h_f} \frac{\partial A_{fg}}{\partial \alpha_k} h_g \Bigg] \Delta \alpha_k \\ & + \left[w_{h_f} \left[P_{w_h}^{-1}\right]_{fd} - \lambda_{h_d} \right] \Delta w_{h_d} \\ & + \sum_{t=1}^{N_t} \left[\epsilon_r^t \left[P_\epsilon^{-1}\right]_{rm} - \lambda_{c_i}^t Y_{im}\left(c^{t-1}\right) \right] \Delta \epsilon_m^t \end{aligned} \quad \text{(A2)}$$

[48] In the minimum of the objective function the variation of the objective function is zero for any variation of the random variables. Forcing this constraint, equation (A2) yields the Euler-Lagrange equations:

[49] For the variations of the concentrations Δc, it yields the adjoint system for concentrations:

$$B_{ji}\lambda_{c_i}^t = D_{ji}\lambda_{c_i}^{t+1} + \frac{\partial M_p(h,c)}{\partial c_j^t}\left[P_v^{-1}\right]_{pn}[z_n - M_n(h,c)] + \epsilon_m^{t+1}\frac{\partial Y_{mi}(c^t)}{\partial c^t}\lambda_{c_i}^{t+1}; \quad t = N_t - 1, \ldots, 1 \quad \text{(A3)}$$

$$B_{ji}\lambda_{c_i}^{N_t} = \frac{\partial M_p(h,c)}{\partial c_j^{N_t}}\left[P_v^{-1}\right]_{pn}[z_n - M_n(h,c)] \quad \text{(A4)}$$

[50] For the variations of the heads Δh, it yields the adjoint system for heads:

$$A_{gf}\lambda_{h_f} = \frac{\partial M_p(h,c)}{\partial h_g}\left[P_v^{-1}\right]_{pn}[z_n - M_n(h,c)] - \sum_{t=1}^{N_t}\left[c_j^t \frac{\partial B_{ji}}{\partial h_g} - c_j^{t-1}\frac{\partial D_{ji}}{\partial h_g}\right]\lambda_{c_i}^t \quad \text{(A5)}$$

[51] For the variations of the parameters $\Delta\alpha_k$, it yields the parameter equation:

$$\alpha_l = \bar{\alpha}_l - P_{\alpha_{lk}}\left[h_g \frac{\partial A_{gf}}{\partial \alpha_k}\lambda_{h_f} + \sum_{t=1}^{N_t}\left(c_j^t \frac{\partial B_{ji}}{\partial \alpha_k} - c_j^{t-1}\frac{\partial D_{ji}}{\partial \alpha_k}\right)\lambda_{c_i}^t\right] \quad \text{(A6)}$$

[52] For the variation of the model errors of the transport equation $\Delta\epsilon^t$, it yields:

$$\epsilon_r^t = [P_\epsilon]_{rm} Y_{mi}\left(c^{t-1}\right)\lambda_{c_i}^t; \; t = N_t - 1, \ldots, 1 \quad \text{(A7)}$$

[53] For the variation of the model errors of the flow equation $\Delta\ w_{hd}$, it yields:

$$w_{h_f} = [P_{w_h}]_{fd}\lambda_{h_d} \quad \text{(A8)}$$

[54] The flow and transport equations are equal to the original equations (3) and (4).

Appendix B: Derivation of Representer Equations

[55] In equations (17)–(21), the representer definitions were introduced in order to solve the Euler-Lagrange

equations (8)–(16). By inserting these representer definitions in the Euler-Lagrange equations, explicit expressions for all representers and their coefficients and correction terms are obtained. However, these expressions still depend on the optimal estimates for the parameters and state variables, which are unknown initially and have to be found iteratively.

B1. Concentration Adjoint Representer

[56] Inserting the representer definitions (17)–(21) in the concentration adjoint equations (8) and (9) yields:

$$B_{ji}\Phi_{ip}^{t}b_p = D_{ji}\Phi_{ip}^{t+1}b_p + \frac{\partial M_p(h,c)}{\partial c_j^t}\left[P_v^{-1}\right]_{pn} \cdot [z_n - M_n(h_F + h_{corr} + \Xi b, c_F + c_{corr} + \Omega b)] + \epsilon_m^{t+1}\frac{\partial Y_{mi}(c^t)}{\partial c^t}\Phi_{ip}^{t+1}b_p; \quad t = N_t - 1, \ldots, 1 \tag{B1}$$

$$B_{ji}\Phi_{ip}^{N_t}b_p = \frac{\partial M_p(h,c)}{\partial c_j^{N_t}}\left[P_v^{-1}\right]_{pn} \cdot [z_n - M_n(h_F + h_{corr} + \Xi b, c_F + c_{corr} + \Omega b)] \tag{B2}$$

[57] So far the adjoint representers Φ and the coefficients b are both unknown. In order to get a unique expression for the adjoint representer Φ, the coefficients b for all p measurements are defined as:

$$b_p = \left[P_v^{-1}\right]_{pn}[z_n - M_n(h_F + h_{corr} + \Xi b, c_F + c_{corr} + \Omega b)] \tag{B3}$$

[58] Inserting this definition in equations (B1) and (B2), these equations can only be fulfilled for nonzero b when:

$$B_{ji}\Phi_{ip}^{t} = D_{ji}\Phi_{ip}^{t+1} + \frac{\partial M_p(h,c)}{\partial c_j^t} + \epsilon_m^{t+1}\frac{\partial Y_{mi}(c^t)}{\partial c^t}\Phi_{ip}^{t+1}; \quad t = N_t - 1, \ldots, 1 \tag{B4}$$

$$B_{ji}\Phi_{ip}^{N_t} = \frac{\partial M_p(h,c)}{\partial c_j^{N_t}} \tag{B5}$$

B2. Head Adjoint Representer

[59] Inserting the representer definitions (17)–(21) in the head adjoint equation (10) yields an expression for the head adjoint representer:

$$A_{gf}\Gamma_{fp}b_p = \frac{\partial M_p(h,c)}{\partial h_g}\left[P_v^{-1}\right]_{pn}[z_n - M_n(h_F + h_{corr} + \Xi b, c_F + c_{corr} + \Omega b)] - \sum_{t=1}^{N_t}\left[c_j^t\frac{\partial B_{ji}}{\partial h_g} - c_j^{t-1}\frac{\partial D_{ji}}{\partial h_g}\right]\Phi_{ip}^{t}b_p \tag{B6}$$

[60] Inserting the definition of representer coefficients (B3), equation (B6) can only be fulfilled for nonzero b if:

$$A_{gf}\Gamma_{fp} = \frac{\partial M_p(h,c)}{\partial h_g} - \sum_{t=1}^{N_t}\left[c_j^t\frac{\partial B_{ji}}{\partial h_g} - c_j^{t-1}\frac{\partial D_{ji}}{\partial h_g}\right]\Phi_{ip}^{t} \tag{B7}$$

B3. Parameter Representers

[61] Inserting the representer definitions (17)–(21) in the parameter equation (11) yields:

$$\bar{\alpha}_l + \Psi_{lp}b_p = \bar{\alpha}_l - P_{\alpha_{lk}}\left[h_g\frac{\partial A_{gf}}{\partial \alpha_k}\Gamma_{fp}b_p + \sum_{t=1}^{N_t}\left(c_j^t\frac{\partial B_{ji}}{\partial \alpha_k} - c_j^{t-1}\frac{\partial D_{ji}}{\partial \alpha_k}\right)\Phi_{ip}^{t}b_p\right] \tag{B8}$$

[62] For any nonzero b, this equation can only be fulfilled if for each p:

$$\Psi_{lp} = -P_{\alpha_{lk}}\left[h_g\frac{\partial A_{gf}}{\partial \alpha_k}\Gamma_{fp} + \sum_{t=1}^{N_t}\left[c_j^t\frac{\partial B_{ji}}{\partial \alpha_k} - c_j^{t-1}\frac{\partial D_{ji}}{\partial \alpha_k}\right]\Phi_{ip}^{t}\right] \tag{B9}$$

B4. Head Representers

[63] We would like to define the head representers in such a way that during each iteration the head representers are the exact linearization around the head estimate of the previous iteration. Afterward, the head correction term will be chosen in such a way that the flow equation (12) will be fulfilled. First, we perturb the flow equation around the estimate of the previous iteration:

$$A_{fg}\left(\alpha + \Psi_p\delta b_p\right)\left(h_g + \Xi_{gp}\delta b_p\right) = q_f + w_{h_f} + [P_{w_h}]_{fd}\Gamma_{dp}\delta b_p \tag{B10}$$

where: $\Psi_p\delta b_p$ perturbation of parameters
$\Xi_{gp}\delta_{\ p}$ perturbation of heads
$[P_{w_h}]_{fd}\Gamma_{dp}\delta b_p$ perturbation of model errors for flow equation.

[64] Linearizing this equation yields:

$$A_{fg}\Xi_{gp}\delta b_p = q_f - \left[A_{fg} + \frac{\partial A_{fg}}{\partial \alpha_k}\Psi_{kp}\delta b_p\right]h_g + w_{h_f} + [P_{w_h}]_{fd}\Gamma_{dp}\delta b_p \tag{B11}$$

[65] When we take the expectation of this equation and subtract it from the same equation, it yields:

$$A_{fg}\Xi_{gp}\delta b_p = -\frac{\partial A_{fg}}{\partial \alpha_k}\Psi_{kp}\delta b_p h_g + [P_{w_h}]_{fd}\Gamma_{dp}\delta b_p \tag{B12}$$

[66] Divide this equation by δb_p yields the head representer equation we wanted to acquire:

$$A_{fg}\Xi_{gp} = -\frac{\partial A_{fg}}{\partial \alpha_k}\Psi_{kp}h_g + [P_{w_h}]_{fd}\Gamma_{dp} \tag{B13}$$

[67] Now we will choose the head correction term in such a way that the forward flow equation (12) will be fulfilled. First, we insert the representer definitions (17)–(21) in the flow equation (12). This yields:

$$A_{fg}\left(h_{F_g} + h_{corr_g} + \Xi_{gp}b_p\right) = q_f + [P_{w_h}]_{fd}\Gamma_{dp}b_p \tag{B14}$$

[68] Now we multiply equation (B13) by b_p, sum it over all measurements, subtract it from equation (B14) and use representer definition (19). It yields:

$$A_{fg}h_{corr_g} = q_f + \frac{\partial A_{fg}}{\partial \alpha_k}(\alpha_k - \bar{\alpha}_k)h_g - A_{fg}h_{F_g} \tag{B15}$$

B5. Concentration Representer and Correction Term

[69] The derivation of the concentration representers and concentration correction term are done by the same steps as we used in the derivation of the head representers and the head correction term.

B6. Determination of Representer Coefficients

[70] The coefficient equation was defined in equation (B3). Rearranging yields:

$$\left[P_{v_{np}} + M_n\left(\Xi_p, \Omega_p\right)\right]b_p = \left[z_n - M_n(h_F + h_{corr}, c_F + c_{corr})\right] \tag{B16}$$

References

Bennett, A. F. (1992), *Inverse Methods in Physical Oceanography*, Cambridge Univ. Press, New York.

Brusseau, M. L., and P. S. C. Rao (1989), Sorption nonideality during organic contaminant transport in porous media, *Crit. Rev. Environ. Control*, *19*, 33–99.

Bryson, A. E., and Y. C. Ho (1975), *Applied Optimal Control*, Taylor and Francis,Philadelphia, Pa.

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*, 199–210.

Cooley, R. L. (1977), A method of estimating parameters and assessing reliability for models of steady state groundwater flow: 1. Theory and numerical properties, *Water Resour. Res.*, *13*, 318–324.

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

Eknes, M., and G. Evensen (1995), Parameter estimation solving a weak constraint variational problem, paper presented at Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, World Meteorol. Organ., Tokyo.

Gelfand, I., and S. V. Fomin (1963), *Calculus of Variation*, Prentice-Hall, Old Tappan, N. J.

Gelhar, L. W., and C. L. Axness (1983), Three-dimensional stochastic analysis of macrodispersion in aquifers, *Water Resour. Res.*, *19*, 161–180.

Harvey, C., and S. M. Gorelick (2000), Rate limited mass transfer or macrodispersion: Which dominates plume evolution at the Macrodispersion Experiment (MADE) site, *Water Resour. Res.*, *36*, 637–650.

Pinder, G. F., and W. G. Gray (1977), *Finite Element Simulation in Surface and Subsurface Hydrology*, Academic, San Diego, Calif.

Reichle, R., D. McLaughlin, and D. Entekhabi (2001), Variational data assimilation of microwave radiobrightnes observations for land surface hydrologic applications, *IEEE Trans. Geosci. Remote Sens.*, *39*(8), 1708–1718.

Reid, L. B. (1996), A Functional inverse approach for three-dimensional characterization of subsurface contamination, Ph.D. thesis, Mass. Inst. of Technol., Cambridge.

Schweppe, F. C. (1973), *Uncertain Dynamic Systems*, Prentice-Hall, Old Tappan, N. J.

Sun, C. C. (1998), A stochastic approach for characterizing soil and groundwater contamination at heterogeneous field sites, Ph.D. thesis, Mass. Inst. of Technol., Cambridge.

Sun, N.-Z. (1994), *Inverse Problems in Groundwater Modeling*, Kluwer Acad., Norwell, Mass.

Valstar, J. R. (2001), Inverse modeling of groundwater flow and transport, Ph.D. thesis, Delft Univ. of Technol., Delft, Netherlands.

D. B. McLaughlin, Massachusetts Institute of Technology, Ralph M. Parsons Laboratory, 15 Vassar Street, Cambridge, MA 02139, USA.

C. B. M. te Stroet, J. R. Valstar, and F. C. van Geer, Netherlands Institute of Applied Geoscience TNO–National Geological Survey, PO Box 80015, 3508 TA Utrecht, Netherlands. (j.valstar@nitg.tno.nl)