8-6-2009

Obtaining Parsimonious Hydraulic Conductivity Fields Using Head and Transport Observations: A Bayesian Geostatistical Parameter Estimation Approach

Michael Fienen
U.S. Geological Survey

R. Hunt
U.S. Geological Survey

D. Krabbenhoft
U.S. Geological Survey

Tom Clemo
Boise State University

 DOI: 10.1029/2008WR007431

Obtaining parsimonious hydraulic conductivity fields using head and transport observations: A Bayesian geostatistical parameter estimation approach

M. Fienen,[1] R. Hunt,[1] D. Krabbenhoft,[1] and T. Clemo[2,3]

Received 8 September 2008; revised 23 April 2009; accepted 4 May 2009; published 6 August 2009.

[1] Flow path delineation is a valuable tool for interpreting the subsurface hydrogeochemical environment. Different types of data, such as groundwater flow and transport, inform different aspects of hydrogeologic parameter values (hydraulic conductivity in this case) which, in turn, determine flow paths. This work combines flow and transport information to estimate a unified set of hydrogeologic parameters using the Bayesian geostatistical inverse approach. Parameter flexibility is allowed by using a highly parameterized approach with the level of complexity informed by the data. Despite the effort to adhere to the ideal of minimal a priori structure imposed on the problem, extreme contrasts in parameters can result in the need to censor correlation across hydrostratigraphic bounding surfaces. These partitions segregate parameters into facies associations. With an iterative approach in which partitions are based on inspection of initial estimates, flow path interpretation is progressively refined through the inclusion of more types of data. Head observations, stable oxygen isotopes ($^{18}O/^{16}O$ ratios), and tritium are all used to progressively refine flow path delineation on an isthmus between two lakes in the Trout Lake watershed, northern Wisconsin, United States. Despite allowing significant parameter freedom by estimating many distributed parameter values, a smooth field is obtained.

Citation: Fienen, M., R. Hunt, D. Krabbenhoft, and T. Clemo (2009), Obtaining parsimonious hydraulic conductivity fields using head and transport observations: A Bayesian geostatistical parameter estimation approach, *Water Resour. Res.*, *45*, W08405, doi:10.1029/2008WR007431.

1. Introduction and Background

[2] The value of delineating three-dimensional groundwater flow paths in groundwater systems is well accepted in hydrogeology. Often they are a necessary construct for interpreting geochemical data and estimating capture zones and related risk assessment. However, delineation of a flow path in the field is not trivial given the complexities of natural hydrogeologic parameter heterogeneity and added challenges associated with transient conditions of field settings. Because of these complexities, commonly collected hydraulic data, such as heads and flows, cannot uniquely constrain the set of possible hydraulic parameters leading to accurate delineation of flow paths. Transport information from plume characterization in the field can help narrow the range of possible flow paths, but often better reflects the bulk flow of the plume (center of mass) without providing the smaller-scale insight needed for assessment of geochemical evolution or mass balance, or well capture/stagnation point identification. Thus, a quantitative framework or realization of the natural system (i.e., a model) is used to relate the direct field information collected from an area of interest (heads, flows, concentration) to output of high-resolution flow paths. A modeling approach has been employed for flow path delineation in several iterations at the U.S. Geological Survey (USGS) Water, Energy and Biogeochemical Budgets Trout Lake site in northern Wisconsin. We revisit flow path delineation at this site using updated data and techniques both to refine the conceptual flow path delineation models at the site, and to evaluate the applicability of Bayesian geostatistical parameter estimation in a field setting.

[3] Using numerical models for transport simulations such as flow path delineation requires more detailed characterization of subsurface properties such as hydraulic conductivity, than models addressing water budgets that rely more on bulk, spatially averaged aquifer parameters [see, e.g., *Haitjema* [1995], *Moore and Doherty* [2005], and *Hunt et al.* [2007]. Our ability to directly characterize the three-dimensional hydraulic conductivity distributions in the field is limited, so models cannot be expected to represent all heterogeneity that occurs at a site. Thus, models used for flow path delineation are challenged by the fact that the world is complex both in space and time because of hydraulic conductivity variation and environmental transience, the complexity impacts simulated flow paths and can never be fully realized. The question then becomes representing a level of complexity high enough to

[1] U.S. Geological Survey, Middleton, Wisconsin, USA.
[2] Center for Geophysical Investigation of the Shallow Subsurface, Boise State University, Boise, Idaho, USA.
[3] Now at Intera, Inc., Richland, Washington, USA.

0043-1397/09/2008WR007431$09.00

accurately delineate flow paths given our state of knowledge about the field setting without overextending the information available in our field observations. Answering these questions allows us to obtain a parsimonious answer, one that does not suffer from unnecessary or unsupported complexity, without unreasonably restricting the level of complexity prior to collecting the data.

[4] The concepts underlying the appropriate level of complexity are current research topics [e.g., *Doherty*, 2003; *Moore and Doherty*, 2006; *Hunt et al.*, 2007; *Tonkin and Doherty*, 2009; *Fienen et al.*, 2009], especially given increases in computational capability. In the case of adding parameters to simulate hydraulic conductivity variations, the ability to create highly complex models with hundreds of thousands or millions of cells has become commonplace. Populating these large numbers of cells with parameter values is problematic because information provided by measurements and prior knowledge about hydraulic parameters often is limited relative to the number of parameters that can be included. Given the flexibility of a large number of parameters, overfitting is a danger in which heterogeneity of a parameter field may be introduced by an algorithm in response to observation noise. The appropriate level of complexity sought in this work is therefore a level of complexity that takes advantage of the flexibility afforded by a large number of parameters without overfitting the observations by fitting to noise. Two end-member approaches have evolved to address variability in the hydraulic conductivity field and can be broadly described as: a priori simplicity usually through lumping into homogeneous zones [e.g., *Hill*, 2006], and multiple stochastic representations of the possible hydraulic conductivity complexity [e.g., *Gomez-Hernandez*, 2006].

[5] A problematic characteristic of the first method is that simple models with few free parameters at the outset limit the ability of the models to provide insight about the uncaptured real-world complexity that is known to exist. This leads to the question raised by *Hunt et al.* [2007, p. 255]: ''Does the model adequately reflect the response of the physical system, or have bias and/or uncertainty been introduced into model predictions as a result of 'handcuffing' to a simplified parameter structure?'' Moreover, when the simplification is specified a priori there is a danger that a conceptual model might be chosen because it is the first to be considered reasonable. This potential pitfall was recognized by *Chamberlin* [1890] when he described a ''paternal affection'' for the first explanation made that is in some way consistent with our observations. The affection can blind us from exploring other equally (or more) likely explanations. To overcome this potential bias, he urged us to employ the method of multiple working hypotheses in which many possible conceptualizations are evaluated. While it is possible to test multiple lumped model candidates, the geostatistical inverse approach discussed in this work is inherently well suited to consider multiple hypotheses as part of its natural application.

[6] The second method, the ''complex'' formulation, which provides multiple realizations that are consistent with the stochastic characteristics of the aquifer properties, honors the concept of multiple working hypotheses. This methodology acknowledges the nonuniqueness inherent in parameter estimation. It can be argued that because the true subsurface complexity can never be completely known, the idea of a single calibrated model is not useful; given this precept, an alternative approach would then use a large number of model runs to explore the range of predictive possibilities rather than making a discrete prediction.

[7] This approach has not been widely adopted in practical applications to date, perhaps because a single, calibrated model commonly forms the basis of environmental decision making. A discrete prediction of flow path is often the most expedient for geochemical interpretation, capture zone delineation, and risk assessment. However, recent United States government mandates indicate an awakening on the part of decision makers to the need for representing uncertainty around results from a single calibrated model. For example, *Office of Management and Budget* [2003, p. 38] stated ''The important uncertainties connected with your regulatory decisions need to be analyzed and presented...A good analysis provides specific references to...uncertainty analyses.'' Furthermore, *National Science and Technology Council Committee on Environment and Natural Resources* [2007, p. 12] stated that ''The effectiveness of (today's) decisions depends on the quality of information and on incorporating knowledge about the reliability (or conversely, the uncertainty) associated with predictive management tools.'' Environmental resource managers and decision makers have recognized the perils of assuming adequacy of a single explanation of water resources phenomena. This drives the need for practitioners to provide more than a single model on which to base decisions.

[8] As an alternative between the two end-member approaches described above, *Hunt et al.* [2007] advocate an intermediate approach of regularized inversion where flexibility is maintained through a high number of parameters, but is constrained through mathematical regularization. In this context, ''flexibility'' refers to the freedom of parameters to vary at small scales rather than be partitioned into a small number of homogeneous zones. They define regularization as ''any process that makes a function more regular or smooth; it can be broadly interpreted as any method that helps provide an approximate and meaningful answer to an ill-posed problem'' (such as having more parameters than field observations available to constrain the parameters) [*Hunt et al.*, 2007, p. 256]. Myriad regularized inversion examples are available from diverse fields such as geophysics, medical imaging, and others. *Hunt et al.* [2007] and *Gallagher and Doherty* [2007] note that the simplicity (zoned) approach described above is a type of informal (but less transparent and powerful) regularization strategy that reduces a complex real world to a small number of model parameters. Other more flexible and sophisticated regularization strategies are suggested therein, including Tikhonov regularization [*Tikhonov and Arsenin*, 1977; *Ory and Pratt*, 1995], truncated singular value decomposition [*Aster et al.*, 2005], and a hybrid combination of the two [*Tonkin and Doherty*, 2005]. They maintain that a more useful way to think of the regularized inversion approach is that it does not

necessarily produce an accurate representation of real-world complexity but rather reflects the level of complexity that is supported by the data. These approaches also provide a framework to explore the implications of complexity that is not supported by the data.

[9] One underlying problem with the intermediate approaches of regularized inversion for model calibration is that small changes in regularization constraints can result in far different but equally well calibrated models. Although for any given set of regularization constraints there is a unique calibrated model, there is still a family of possible calibrated models that could be considered reasonable. Thus, model calibration performed using regularized inversion may yield a best fit between simulated and observed targets, but it may not reflect the most likely set of parameters given what is known about a site.

[10] In this work, flow paths are used to characterize the hydraulic conductivity field at a field site in northern Wisconsin. The field site has a long history of attempts to delineate subsurface flow paths using various multidisciplinary approaches. This work presents the next step in that evolution. Rather than the regularization approaches of *Hunt et al.* [2007], parameter estimation stability is obtained through the Bayesian geostatistical inverse method [*Kitanidis and Vomvoris*, 1983; *Hoeksema and Kitanidis*, 1984]. In this implementation, the appropriate level of complexity is guided by the data following (in spirit) the principle of maximum entropy in which "...the least biased estimate possible on the given information..." is the goal [*Jaynes*, 1957, p. 620]. This method is attractive because it retains flexibility through a large number of free parameters while explicitly considering prior information regarding system properties and the strength of assumptions used to frame the problem. We ultimately seek a simple answer in the sense that it has "...freedom from useless accessories" (from the definition of "simplicity") [*Simpson and Weiner*, 1989]. The level of complexity is informed by the data through systematic tradeoff between simplicity and the ability of the model outputs to match observations. This allows unbiased selection from among the family of possible calibrated solutions contingent upon prior information. In that way, the calibration results represent not only the best fit, but also the pointwise, statistically most likely model given what is known about a site. When choosing the "most likely" solution, we accept a level of smoothing that is somewhat higher than reality. However, conditional realizations can be made from the distribution of parameters that more fully characterize the variability in the answer [*Kitanidis*, 1995; *Tonkin and Doherty*, 2009].

[11] The Bayesian geostatistical inverse method is powerful but has seen few applications to field data [*Michalak and Kitanidis*, 2002, 2003; *Fienen et al.*, 2004, 2006; *Li et al.*, 2007, 2008; *Mueller et al.*, 2008; *Gourdji et al.*, 2008]. The goals of this work are therefore twofold: first, to apply the Bayesian geostatistical inverse method to the problem of flow path delineation through hydraulic conductivity characterization using field flow and transport data and second, to use available common use models with necessary method and model modifications in such a way that the method could be systematically applied on a broader scale to practical problems. The remainder of this paper starts with a review of the inverse methodology, including innovations necessary for the application motivating this work, and follows with discussion of the actual field project and results.

2. Methodology

[12] The Bayesian geostatistical inverse method at the foundation of this work was developed by *Kitanidis and Vomvoris* [1983] and *Hoeksema and Kitanidis* [1984] for linear problems and extended in an iterative quasi-linear approach by *Kitanidis* [1995]. Incorporation of diffuse information about the mean and implementation of a modified Levenberg-Marquardt algorithm for numerical stability was introduced by *Nowak and Cirpka* [2004]. A summary of the method is presented here with details available in the references above and in section S1 of Text S1 in the auxiliary material.[1]

[13] The application in this paper is to a flow and transport groundwater problem in which hydraulic conductivity parameters are estimated. The data available are piezometric head, stable oxygen isotope ratios, and tritium concentrations. The data are added progressively starting with using head-only values, and finally adding tritium and oxygen isotopes. The hydraulic conductivity fields estimated from these various data sources are then used to delineate flow paths using particle tracking. The flow paths represent an integrated, conceptual representation of the site for comparison with previous modeling work.

2.1. Bayesian Geostatistical Inverse Method

[14] The core of the Bayesian geostatistical inverse method is Bayes' theorem, which states

$$p(\mathbf{s}|\mathbf{y}) \propto L(\mathbf{y}|\mathbf{s})p(\mathbf{s}), \tag{1}$$

where $\mathbf{y}$ are the measured data, $\mathbf{s}$ are the unknown parameters, $p(\mathbf{s}|\mathbf{y})$ is the posterior probability density function (pdf) of $\mathbf{s}$ given $\mathbf{y}$, $L(\mathbf{y}|\mathbf{s})$ is the likelihood function, and $p(\mathbf{s})$ is the prior pdf of $\mathbf{s}$. Details of these pdf's are explained in the following paragraphs with more mathematical details in section S1 of Text S1 in the auxiliary material.

[15] Figure 1 depicts one-dimensional distributions graphically illustrating equation (1). In this example, the prior distribution ($p(\mathbf{s})$) is diffuse, meaning the variance is relatively high and, correspondingly, commitment to a particular value is low. The likelihood function ($L(\mathbf{y}|\mathbf{s})$), on the other hand, has lower variance, suggesting a process that brings a higher level of certainty to the estimation of the parameters ($\mathbf{s}$) than is indicated by the prior distribution only. The resulting posterior distribution ($p(\mathbf{s}|\mathbf{y})$) is a convolution of the prior and likelihood functions. The peak is shifted significantly from the prior toward the likelihood and is higher. The posterior distribution also has lower variance than the prior as illustrated by the spread in Figure 1. Since the likelihood function is more focused (less spread) and higher than the prior, the attributes of the likelihood contribute more to the posterior than the prior does.

[1] Auxiliary materials are available in the HTML. doi:10.1029/2008WR007431.

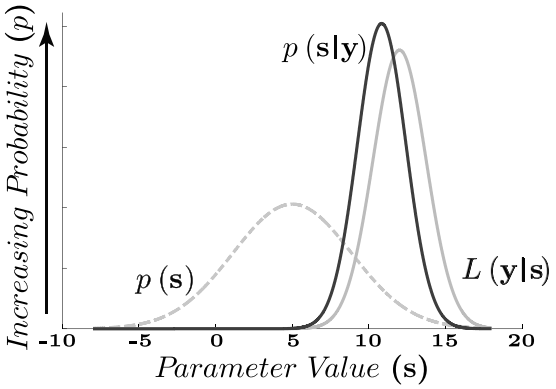

Figure 1. Graphical illustration of Bayes' theorem.

[16] In the Bayesian geostatistical method, the posterior pdf is calculated as

$$p(\mathbf{s}|\mathbf{y}) \propto \underbrace{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{h}(\mathbf{s}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{s}))\right)}_{\mathrm{L}(\mathbf{y}|\mathbf{s})} \underbrace{\exp\left(-\frac{1}{2}(\mathbf{s}-\mathbf{X}\beta^*)^T\mathbf{G}_{\mathbf{ss}}^{-1}(\mathbf{s}-\mathbf{X}\beta^*)\right)}_{\mathrm{p}(\mathbf{s})}, \tag{2}$$

where $\mathbf{s}$ is the vector of parameter values at distributed spatial locations in the model, $\mathbf{X}\beta^*$ is the prior mean, $\mathbf{G_{ss}}$ is the prior covariance of $(\mathbf{s} - \mathbf{X}\beta^*)$, $\mathbf{h}(\mathbf{s})$ is the modeled outputs collocated with observations $(\mathbf{y})$, and $\mathbf{R}$ is the epistemic uncertainty covariance, modeled as $\sigma_R^2\mathbf{I}$ where σ_R^2 represents epistemic uncertainty, and $\mathbf{I}$ is an identity matrix. In general terms, the likelihood function ($L(\mathbf{y}|\mathbf{s})$) characterizes the misfit between model outputs and observations while the prior pdf ($p(\mathbf{s})$) defines a characteristic (such as smoothness or continuity) that is assumed to apply to the parameter field. The prior pdf also serves the role of regularization.

[17] The best estimate of $\mathbf{s}$ maximizes the posterior pdf. A computationally efficient method to find the best estimates of $\mathbf{s}$ and β ($\hat{\mathbf{s}}$ and $\hat{\beta}$, respectively) is through

$$\hat{\mathbf{s}} = \mathbf{X}\hat{\beta} + \mathbf{Q_{ss}}\mathbf{H^T}\xi, \tag{3}$$

which is the superposition of the prior mean (first term) and an innovation term which considers deviations of the predictions from the observations (second term). $\mathbf{H}$ in the second term (often referred to as the Jacobian, sensitivity, or susceptibility matrix) is the sensitivity of observation values to parameter values where $H_{ij} = \frac{\partial \mathbf{h}(\mathbf{s})_i}{\partial \mathbf{s}_j}$ calculated using adjoint state methods.

[18] The values for $\hat{\beta}$ and ξ are found by solving the $(n + p) \times (n + p)$ linear system of cokriging equations

$$\begin{bmatrix} Q_{\mathbf{yy}} & \mathbf{HX} \\ \mathbf{X}^T\mathbf{H}^T & -\mathbf{Q}_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} \xi \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ -\mathbf{Q}_{\beta\beta}^{-1}\beta^* \end{bmatrix}, \tag{4}$$

where $\mathbf{Q_{yy}}$ is the autocovariance matrix of the observations, defined as $\mathbf{HQ_{ss}H^T} + \mathbf{R}$.

[19] In hydrogeologic applications, the numerical forward model is typically nonlinear. Provided that the nonlinearities introduced are not too extreme, a solution can be obtained through successive linearizations following the quasi-linear extension [*Kitanidis*, 1995]. The forward model, $\mathbf{h}(\mathbf{s})$ is expanded about the current best estimate of the parameters $\tilde{\mathbf{s}}$,

$$\mathbf{h}(\mathbf{s}) \approx \mathbf{h}(\tilde{\mathbf{s}}) + \tilde{\mathbf{H}}(\mathbf{s} - \tilde{\mathbf{s}}), \tag{5}$$

where $\tilde{\mathbf{H}}$, as a function of $\tilde{\mathbf{s}}$, is evaluated at each linearization. Linearizations are performed until the change in $\tilde{\mathbf{s}}$ in successive iterations does not change over a specified tolerance. We assign the subscript $_k$ to indicate iteration number, and correct the measurements for the kth linearization as

$$\mathbf{y}'_k = \mathbf{y} - \mathbf{h}(\tilde{\mathbf{s}}_k) + \tilde{\mathbf{H}}_k\tilde{\mathbf{s}}_k. \tag{6}$$

Then the cokriging equations (equation (4)) are updated:

$$\begin{bmatrix} \tilde{\mathbf{Q}}_{\mathbf{yy},k} & \tilde{\mathbf{H}}_k\mathbf{X} \\ \mathbf{X}^T\tilde{\mathbf{H}}_k^T & -\mathbf{Q}_{\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} \xi_k \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \mathbf{y}'_k \\ -\mathbf{Q}_{\beta\beta}^{-1}\beta^* \end{bmatrix}, \tag{7}$$

where $\tilde{\mathbf{Q}}_{\mathbf{yy},k} = \tilde{\mathbf{H}}_k\mathbf{Q_{ss}}\tilde{\mathbf{H}}_k^T + \mathbf{R}$. From this set of equations, the next estimate of $\mathbf{s}$ is

$$\tilde{\mathbf{s}}_{k+1} = \mathbf{X}\hat{\beta}_k + \mathbf{Q_{ss}}\tilde{\mathbf{H}}_k^T\xi_k. \tag{8}$$

[20] This can be iterated until there is no difference in the parameter estimates, or when there is no further improvement in the objective function. In many cases, numerical instability makes convergence difficult. A modified Levenberg-Marquardt approach is adopted for this work which splits the system in equation (7) into a projection step and an innovation step [*Nowak and Cirpka*, 2004]. The objective function to minimize is $-\ln p(\mathbf{s}|\mathbf{y})$ which is equivalent to maximizing equation (2):

$$\mathcal{L} = -\ln p(\mathbf{s}|\mathbf{y}) = (\mathbf{s} - \mathbf{X}\beta^*)^T\mathbf{G}_{\mathbf{ss}}^{-1}(\mathbf{s} - \mathbf{X}\beta^*) + (\mathbf{y} - \mathbf{h}(\mathbf{s}))^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{s})). \tag{9}$$

Further detail about this equation and its variables is provided in section S1 of Text S1 in the auxiliary material.

[21] We adopt an empirical Bayes [*Robbins*, 1956; *Casella*, 1985] approach to inference in this method meaning that the general characteristics of the prior and epistemic covariances introduced above are provided in the model setup, but the values of structural parameters that control the balance between smoothness and misfit are estimated from the data. In equation (2) the structural parameter for $L(\mathbf{y}|\mathbf{s})$ is the epistemic uncertainty parameter (σ_R^2) and the structural parameters for $p(\mathbf{s})$ are covariance parameters (θ). Estimating the structural parameters (σ_R^2 and θ) from the data is

performed using restricted maximum likelihood as discussed in section S1 of Text S1 in the auxiliary material. This approach is consistent with the principle of maximum entropy such that the smoothest solution is chosen on the basis of the structural parameters estimated from the data. For a discussion of subtle formal differences from the principle of maximum entropy, see *Rubin* [2003, pp. 333–342].

[22] An extension to this approach is the inclusion of information about the prior mean [*Nowak and Cirpka*, 2004]. The mean is estimated in the solution, but a prior value and covariance can be supplied to constrain the estimate. Typically, a relatively high covariance magnitude is used so the information about the mean is "diffuse" and principally serves the role of providing numerical stability.

[23] The forward model is constructed to provide calculations of values collocated in space and time with the measured observations. The likelihood function brings the observed data into the calculation by quantifying the difference (misfit) between the model outputs and collocated observations. In all modeling, we must acknowledge that perfect correspondence between model outputs and observations is neither attainable nor desirable. The observations themselves are corrupted by measurement errors and there is usually a lack of perfect correspondence between the exact nature of the measurements and the simulated counterparts. This corruption is due to uncertainty from sources including the paucity of observations, imperfections in the conceptual model, and approximations made to codify the physics of the phenomena into a numerical model framework. All of these sources of uncertainty are described by the term "epistemic uncertainty" [*Rubin*, 2003, p. 4]. This epistemic uncertainty characterizes the expected misfit between forecasts and observations, and is expressed through a covariance function. As a result, the likelihood function can be characterized by a Gaussian distribution with zero mean and covariance defined by the epistemic uncertainty.

[24] With both the prior pdf and likelihood function expressed as Gaussian distributions, the posterior pdf (found through equation (S1) in section S1 of Text S1 in the auxiliary material) is also Gaussian. The values of the parameters **s** that result in the maximum value of the posterior pdf are therefore the most likely solution on a point-by-point basis. The solution as a whole is always a somewhat smoothed version of reality, but the small-scale variability can be characterized through conditional realizations. The balance between the strength of smoothing and the closeness of correspondence between model outputs and observations is found through calculation of optimal values for the structural parameters, including a value to quantify the epistemic uncertainty.

2.2. Facies Associations

[25] In an idealized problem, a single covariance model is flexible enough to encompass the entire variability of the hydraulic parameters. However, in many hydrogeologic applications, lithologic contacts and unconformities can create discontinuities in parameter values that a single covariance model cannot characterize. Partitioning the field on the basis of either the data [e.g., *Fienen et al.*, 2004] or through interrogation of preliminary solutions [e.g., *Fienen et al.*, 2008; *Cardiff and Kitanidis*, 2008] can greatly improve results. This partitioning is implemented by imposing discontinuities in the stochastic field that censor correlation between all cells that do not occur in the same partition. In this context, "stochastic" refers to the entity being partitioned (namely the correlation structure of the parameter field) but we emphasize here that the locations of the imposed discontinuities are themselves considered deterministic and certain. This partitioning is consistent with zonal boundaries in models made up of homogeneous zones.

[26] We borrow the term "facies association" from the facies architecture field to describe these partitions. The term "facies association" typically refers to descriptive properties of a subset of a medium in the field or at least for a specific project. "Architectural elements" is used in the broader case where the characteristics are more formally defined [see *Collinson*, 1969; *Walker*, 1984, 1992; *Swift et al.*, 2003]. It is appropriate to use the less restrictive and less transferable term "facies association" in this work because when we subdivide the correlation structure of the medium, we often base the discontinuities (bounding surfaces, or contacts) on perceived hydraulic properties. These properties will often coincide with differences in age, provenance, or depositional environment, but such coincidence is not required for or by their use. In all cases, partitioning into facies associations must be based on readily observable hydrologic or lithologic attributes.

[27] To clarify our terminology in this work, partitions delineated by discontinuities within a distributed parameter field are referred to as "facies associations," whereas zones of piecewise continuity are referred to as "homogeneous zones." The facies associations delineate subregions of the model domain that share correlation characteristics and are uncorrelated from neighboring facies associations; they are usually delineated by features that are easily identified in measured data or geologic conceptualizations of a given site area. In facies associations, variability of parameter values within each cell is allowed and constrained by the a priori covariance structure, whereas in homogeneous zones, a single parameter value represents the property for the entire zone. Details about the implementation of facies associations is included in section S1 of Text S1 in the auxiliary material.

2.3. Calculation and Weighting of Parameter Sensitivities

[28] All gradient-based parameter estimations schemes (including PEST [*Doherty*, 2008] and UCODE [*Poeter et al.*, 2005] among others) require the calculation of a Jacobian matrix of parameter sensitivities. The Bayesian geostatistical approach which, in this implementation, uses a Gauss-Newton method for its solution, is not an exception. Parameter sensitivities quantify the incremental change in an observation value resulting from an incremental perturbation of a parameter value. In this section we discuss both the method used to efficiently calculate the Jacobian matrix and the weighting scheme used in this work.

2.3.1. Calculation of Sensitivities Using Adjoint States

[29] In underdetermined problems where the number of parameters exceeds, often greatly, the number of observations, it is computationally advantageous to calculate the parameter sensitivity or Jacobian matrix **H** using adjoint state equations. In quasi-linear problems such as this one, the **H** matrix must be recalculated many times, so the computational expense of a single evaluation of **H** is compounded. Traditional calculation of **H** using finite

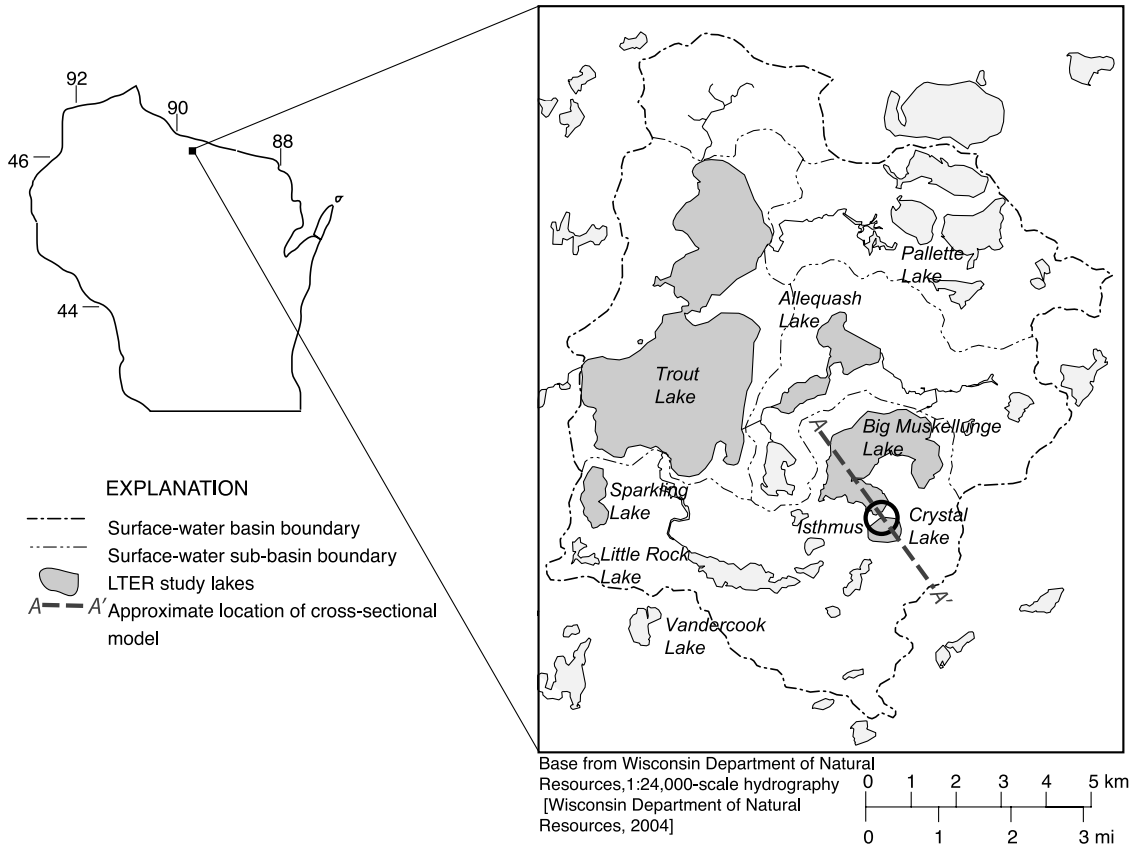

Figure 2. Location of Trout Lake WEBB site in Vilas County, Wisconsin. The entire Trout Lake basin is depicted, and the circle indicates the location of the isthmus between Crystal and Big Muskellunge lakes.

differences [e.g., *Aster et al.*, 2005, p. 181] or the sensitivity equation [e.g., *Yeh*, 1986] requires on the order of m (the number of parameters) forward model runs [e.g., *Hill and Tiedeman*, 2007, p. 47] and suffers from inaccuracy due to selection of an appropriate increment. An adjoint state formulation is both more accurate, as it is closed form, and much faster, requiring on the order of n (the number of observations) forward model runs. Further details of the adjoint state implementation are in section S2 of Text S1 in the auxiliary material.

2.3.2. Jacobian Transformations and Their Uses

[30] Numerical problems often occur when highly disparate parameter sensitivities are encountered. For example, a single measurement, or a single data type, can appear extremely sensitive because of a very high parameter sensitivity value in the Jacobian matrix (**H**). The result, somewhat counterintuitively, is that a parameter with a very high sensitivity may oscillate wildly as the algorithm is compelled to promote large changes. Various strategies have addressed the issue of equalizing contributions of sensitivity to **H** [e.g., *Li and Oldenburg*, 1996; *Doherty*, 2008; *Hill and Tiedeman*, 2007, chapter 4]. *Li and Oldenburg* [1996] addressed the disparately high resolution in magnetic data at shallow depths which overwhelms the contribution of deeper measurements. They implemented a weighting scheme directly on **H** to equalize the contributions of measurements throughout the domain. *Weiss and Smith* [1998] also discuss methods for determining the weight of different data types which, in turn, also impacts that values in **H** corresponding to each type.

[31] In section S1 of Text S1 in the auxiliary material we discuss the transformations of observations and parameters and the corrections both transformations require on the Jacobian. When a few high observation values are responsible for too broad a range of sensitivities, transforming **H** without transforming the observations may improve performance without underweighting the contribution to the objective function of the high observation values because the objective function is not dependent upon the current value of **H**. This is illustrated in the example in section 3.

3. Application to Trout Lake Isthmus

[32] This work was performed as part of an ongoing project at the U.S. Geological Survey's (USGS) Trout Lake Water, Energy and Biogeochemical Budgets (WEBB) watershed research site in Vilas County, northern Wisconsin (Figure 2). The Trout Lake basin has been studied extensively since the founding limnological research of Birge and Juday in the early 1900s [*Juday et al.*, 1938] and remains an active site of interdisciplinary research by the USGS [*Walker and Bullen*, 2000] and the Long-Term Ecological Research (LTER) program [*Magnuson et al.*, 1984]. The present work focuses on a narrow isthmus between Big Muskellunge Lake and Crystal Lake (Figure 2) where paired flow path delineation and geochemical evolution studies were initiated over 20 years ago [*Kenoyer*, 1986].

The two lakes and intervening isthmus have been intensively instrumented and extensively studied. Some of the first research relating lakes and terrestrial sources of water and nutrients were hypothesized at or near this isthmus [*Juday et al.*, 1938; *Hurley et al.*, 1985; *Kenoyer*, 1986; *Kenoyer and Bowser*, 1992a, 1992b; *Bullen et al.*, 1996] Advancements in the understanding of hydrogeochemical evolution of groundwater originating as recharge from upgradient Crystal Lake to downgradient Big Muskellunge Lake were also achieved here [*Kenoyer and Anderson*, 1989; *Bullen et al.*, 1996; *Bowser and Jones*, 2002]. In order to relate observed groundwater chemistry trends to geochemical processes, sources of water, and time of travel, the hydrogeochemical evolution studies estimated flow paths using flownet analyses [*Kenoyer*, 1986; *Kenoyer and Bowser*, 1992a, 1992b], head and water isotope sampling near [*Schindler and Krabbenhoft*, 1998] and at [*Bullen et al.*, 1996] the isthmus, and numerical modeling of flow [*Kim et al.*, 1999] and isotope transport [*Kim et al.*, 2000]. In general, the flow paths from the various studies were consistent from the upgradient lake to midway along the isthmus. Thereafter, various inferred flow paths had less agreement after encountering a significant change in lithology. Estimated flow paths after the lithology change were reported as being mostly horizontal [*Kenoyer and Bowser*, 1992a; *Bowser and Jones*, 2002] or horizontal to slightly plunging [*Bullen et al.*, 1996; *Kim et al.*, 1999]. However, additional instrumentation and sampling on and near the isthmus (especially at greater depths) suggest the previously published flow paths may not be fully accurate. Thus, the objective of this work was to develop a detailed flow path conceptual model that incorporates all the data types and understanding of the site. This insight, in turn, could then be used to update the geochemical evolution understanding and in combined flow and transport modeling. Details about the type and use of these data follow in the remainder of this section.

3.1. Conceptual Framework and Previous Modeling

[33] The hydrogeology of the region around Crystal and Big Muskellunge lakes is typical of glacial outwash deposits resulting from multiple advances of continental glaciers. The aquifer consists of 40–60 m of unconsolidated Pleistocene hydraulically continuous glacial outwash sand and gravel. Surface water features (lakes, streams, and wetlands) are weakly connected via a poorly integrated drainage network but well connected through the subsurface groundwater system. The deepest outwash sediment overlies relatively impermeable Precambrian crystalline bedrock. The aquifer sediment consists of sand and gravel, with localized silt deposits; the variability in sediment reflects conditions during and between glacial advances [*Attig*, 1985]. Movement of groundwater within the watershed has been previously simulated using a number of models, including an analytic element screening model and a three-dimensional, finite difference model. *Walker and Bullen* [2000] and *Hunt et al.* [2006] provide a more detailed description of the hydrogeological framework and previous watershed-scale groundwater modeling efforts.

[34] On the isthmus, a series of nested piezometers is installed along a cross-sectional transect coincident with the perceived flow direction (based on ground and surface water head observations) from Crystal to Big Muskellunge lakes (Figure 3). Drilling logs and head observations from piezometers reveal an aquifer section consisting of shallow sand and gravel, a relatively thin (less than one m thick) silt layer, and a deeper sand and gravel unit (Figure 3). The effect of the silt is easily identified by a severe drop in hydraulic head over a very short distance, indicating the presence of a laterally continuous low hydraulic conductivity feature near the surface near Big Muskellunge Lake, dipping to about 14m depth near Crystal Lake. Because of the high water table and sediment size and sorting, deeper sediments have not been recovered using standard methods of coring (e.g., split spoon sampling). The shallowmost sand and gravel, however, has been extensively characterized including slug testing [*Kenoyer*, 1986] and rhodamine tracer testing [*Kenoyer*, 1988]. The head data and drilling observations have been used in the past to delineate homogeneous zones and in this work are used to delineate boundaries among facies associations.

[35] *Kim et al.* [1999] developed a two-dimensional cross-sectional numerical groundwater model focused on the isthmus. Regional flow and lake-aquifer interactions were simulated through a priori specification of general head boundary conditions [*Harbaugh*, 2005, pp. 6–4] and the upstream boundary was near the lake-isthmus interface. In order to obtain reasonable calibration results with a reasonable vertical to horizontal anisotropy ratio, *Kim et al.* [1999] divided the silt into multiple layers each with a low anisotropy ratio that, in aggregate, provided sufficient anisotropy to match calibration targets. *Kim et al.* [1999] stated however, that this construction was a feature not intended to represent the actual lithology at the site. The model calibration was attained by trial-and-error calibration of piecewise-constant/homogeneous zones and the calibration flow field was compared to the stable water isotopes that represented the terrestrial recharge on the isthmus and the lake plume originating as recharge from Crystal Lake. *Hunt et al.* [2008] expanded the modeling to simulate transience.

[36] The present model extends the previous modeling efforts of this field site in several ways. First, to minimize boundary condition artifacts, the perimeter boundaries of the cross section model were extended upgradient to the Trout Lake watershed divide simulated by regional modeling and downgradient beyond the northwest shore of Big Muskellunge lake (Figure 2). Extending boundaries should ensure that misspecification of boundaries does not impact behavior of water or solutes in the isthmus. The geometry of these "far-field" areas distant from the isthmus is derived from aquifer thickness and lake bathymetry information and starting aquifer hydraulic conductivity values were set to regional parameter values from *Pint et al.* [2003]. Second, the groundwater–surface water interface representation was also extended and refined so the entire lake was included along the cross section and simulated using constant head boundary conditions for the lake and lower-conductivity lake sediments. Third, the bottom elevation was refined to reflect information gained from recent emplacement of piezometers at depths more than twice those of the previous piezometer network. Finally, improved vertical and horizontal resolution was attained by using a finer grid resolution. The minimum lateral cell discretization in the *Kim et al.* [1999] model was 5 m whereas we use 2.4 m uniformly

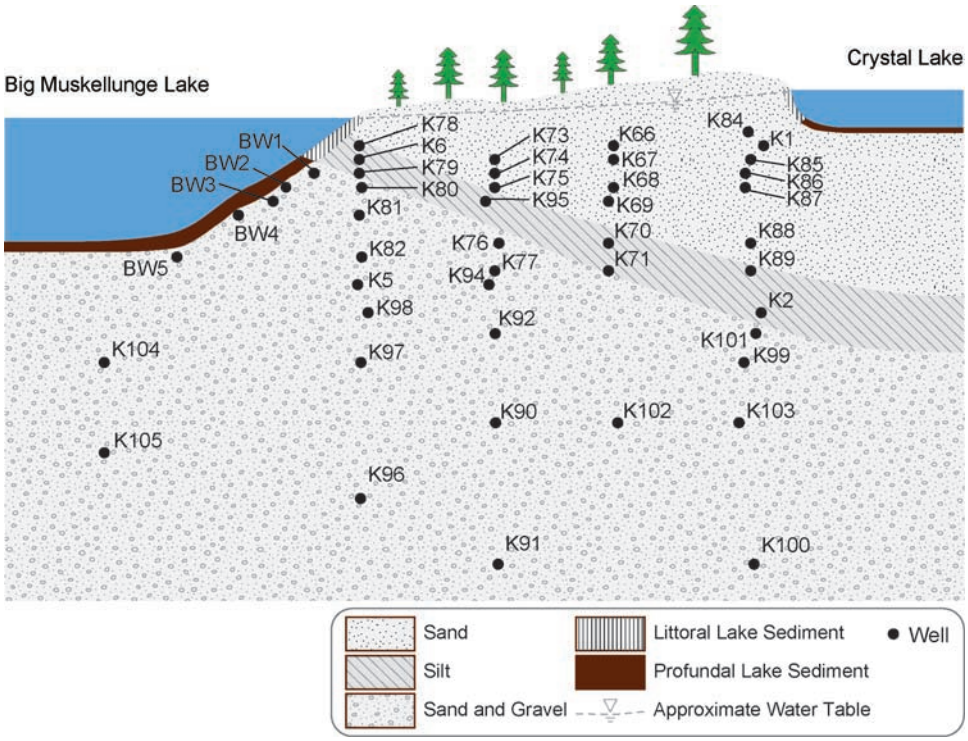

Figure 3. Cartoon representation of the general lithology on the isthmus with well names and locations in the isthmus cross section shown in model coordinates. Figure not to scale. Note that the distance from lakeshore to lakeshore is approximately 115 m, and the depth to the deepest wells is approximately 11 m from the land surface.

throughout the isthmus. The maximum lateral cell discretization in the far field is 12.5 m. A preliminary parameter estimation based on this model with discrete homogeneous zones representing the lithology of the isthmus was performed using PEST [*Doherty*, 2008], and the resulting values were used as initial values for the present parameter estimation. The homogeneous zones used in the preliminary model are illustrated in Figure 4.

3.2. Delineating Facies Associations

[37] Although a primary goal of the work is to interject the maximum amount of flexibility into the problem as practical, two areas required further constraint. The first involved distant far-field aquifer and lake sediment areas where no data were available to inform the parameterization. Only the bulk properties of these areas significantly impact areas of the isthmus interior where calibration data exist. Second, the head data on the isthmus clearly indicate distinct differences with a higher head in the shallow sediments, a sloping, thin, laterally continuous transition zone, and a bottom aquifer of lower head. Thus, rather than piecewise constant zones such as those used by *Kim et al.* [1999], the initial lumped model was adapted such that the homogeneous zones within the isthmus were converted to a

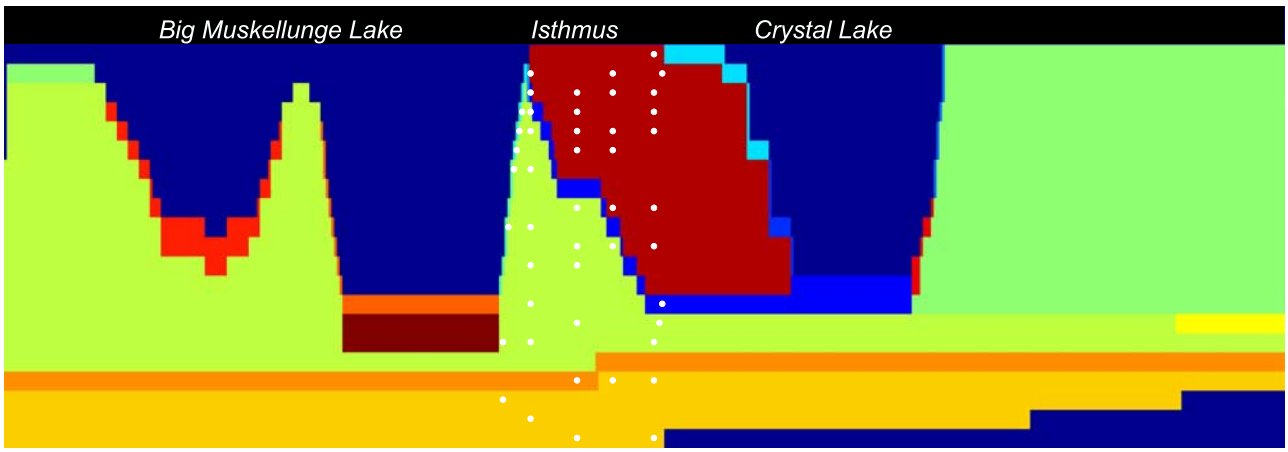

Figure 4. Original proposed model zones. White circles indicate locations of wells within the isthmus. These zones were used as homogeneous parameters with PEST. Figure not to scale.

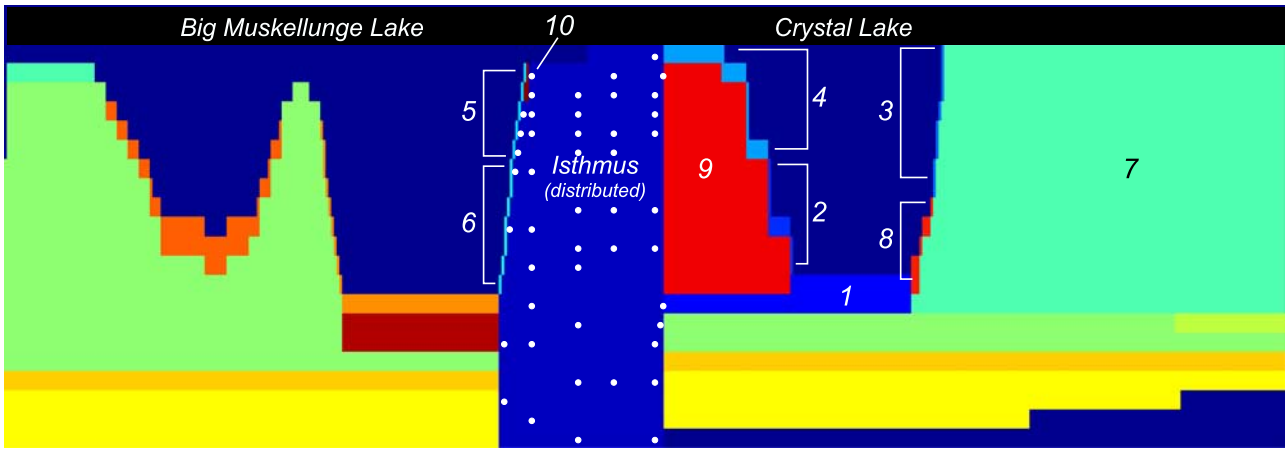

Figure 5. Final major model zones, including outer homogeneous zones and the inner distributed core within the isthmus. White circles indicate well locations, and numbered zones indicate homogeneous zones for which parameters are estimated by importance, particularly in influencing transport. Unnumbered zones are held at values estimated in the preliminary parameter estimation phase performed with PEST. Within the zone labeled "Isthmus (distributed)" each model cell is treated as a parameter. Figure not to scale.

distributed parameter zone with the freedom for each model cell to vary, but this distributed zone was subdivided into three facies associations. Both horizontal and vertical hydraulic conductivity were estimated separately, resulting in two parameter values for each of the 1162 model cells; 2324 hydraulic conductivity values in total for the focused region of the isthmus. Furthermore, ten homogeneous zones representing the far field and lake sediment areas (Figure 5) were estimated bringing the total number of parameters to 2344. The remaining 4614 active model cells were set at constant values lumped into eight homogeneous zones in the far field.

[38] A distinction was made between hydraulic conductivity of course-grained littoral (nearshore) and fine-grained profundal (deep water) lake sediments. This generalization is commonly used in three-dimensional groundwater models including lakes [e.g., *Winter*, 1976; *Krabbenhoft et al.*, 1990a] and is necessary to create the significant head drop between the lake and groundwater system in the midlake region. On the basis of previous studies, the region between the watershed divide and Crystal Lake was delineated as zone 7 and the shallowest portion of Crystal Lake as zone 9 (Figure 5). Figure 5 also shows zone 10 which was added after initial parameter estimation was performed on the basis of head data. This zone corresponds to the shallowest point in the silt in an extremely sensitive region that allows hydraulic communication through the otherwise continuous silt, and reflects the potential for wave and near surface processes that may affect the physical properties in this location compared to the deeper and more protected silt sediments. More details about the addition of zone 10 are in section 3.5.2.

[39] While the individual parameter values are allowed to vary, they are constrained by a functional relationship to each other through a prior covariance function which enforces continuity of neighboring parameters within the field; effectively this is a smoothness constraint. Initially, a single covariance function with a single mean value parameter was assigned to estimate the parameter field within the distributed isthmus zone. The discontinuity in hydraulic parameters between the silt layer and the surrounding sand layers is clearly observed, however, and a single covariance function assuming continuity in parameter values cannot be expected to accommodate such sharp contrasts. The covariance matrix was therefore subdivided through the imposition of discontinuities which censor correlation among the partitions, resulting in facies associations. The facies association boundaries are shown in Figure 6. We stress that the

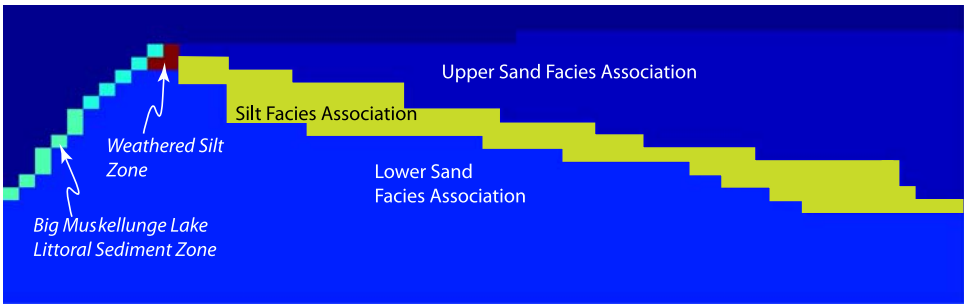

Figure 6. Facies associations as defined by stochastic discontinuities in the distributed isthmus parameter field. Figure not to scale.

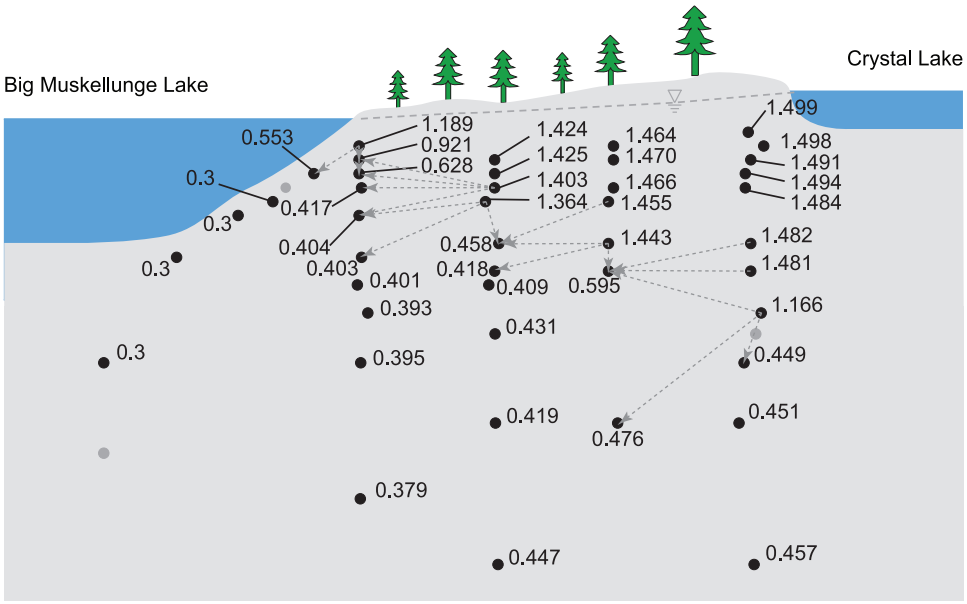

Figure 7. Hydraulic head measurements, in meters relative to a consistent but arbitrary datum, measured on 24–26 May 1999 which are characteristic of steady state conditions between Crystal and Big Muskellunge lakes. The Crystal Lake water level was 1.52, and the Big Muskellunge Lake water level was 0.3. Black circles indicate measured wells, while grey circles indicate wells for which data are not available. Dashed grey lines with arrows indicate head difference targets used as observation data in the inverse problem. Figure not to scale.

subdivision into facies associations does not result in a single homogeneous value representing the hydraulic property for the entire zone. The ten homogeneous zones in Figure 5 correspond to areas that are important for controlling the upstream migration of terrestrial and lacustrine recharge with different isotopic concentrations as discussed in sections 3.6 and 3.7.

[40] To minimize the bias imposed through this grouping, the widest boundaries possible between wells which represent the upper and lower head values across the silt were selected to delineate the contacts between the silt layer and the sands above and sand/gravel below. This resulted in three facies associations; upper sand, silt, and lower sand/gravel. The prior information was incorporated through the use of an exponential covariance function with integral scale equal to ten times the maximum domain length in the isthmus. The behavior of this covariance function is like a (nonstationary) linear variogram but is stationary which has important computational properties. The strength of smoothing imposed by the covariance function is controlled by a single structural parameter (θ). One value of θ is estimated in each facies association. The same covariance structure was assumed to apply to both horizontal and vertical hydraulic conductivity so three prior distribution structural parameters were estimated; one per facies association. These structural parameters were estimated using restricted maximum likelihood as described in section S1.6 of Text S1 in the auxiliary material.

[41] A single epistemic uncertainty term (σ_R^2 in equation (2)) is the final structural parameter estimated and was distributed among the data types using weights to reflect not only measurement uncertainty, but other sources of uncertainty including incorrect conceptual model and data insufficiency. Nonetheless, the first attempt to assign weights was based on the principal of using the inverse of measurement error to normalize the contribution of each data type [e.g., *Hill and Tiedeman*, 2007, p. 294].

[42] A simpler approach that yielded reasonable results was to normalize each data type by the median of all measurements available of that type. This results in a relative contribution of each observation on the order of unity and was a practical decision which enhances the stability of the algorithm. The interpretation of the epistemic uncertainty structural parameter requires multiplication by the normalization weight prior to comparison with mean squared error misfit of the observations. Tying the epistemic uncertainty deterministically by data type limits the interpretation of epistemic uncertainty in comparison with measurement error. However, measurement error is only a single source of epistemic uncertainty which may be eclipsed by other sources of model uncertainty such as the error from model structure and zonation [*Gaganis and Smith*, 2001; *Moore and Doherty*, 2005; *Gallagher and Doherty*, 2007], so directly tying epistemic uncertainty to measurement error is often inappropriate.

3.3. Anisotropy and Prior Information

[43] An important constraint on hydraulic conductivity in sedimentary deposits is anisotropy. While anisotropy is, in part, a function of scale, horizontal hydraulic conductivity should generally be greater than or equal to vertical hydraulic conductivity [e.g., *Freeze and Cherry*, 1979, pp. 148–154]. *Kenoyer* [1988] calculated a site-specific anisotropy ratio (horizontal to vertical) ranging from 2.4 to 8.3 using a tracer test on the scale of about 1 m in the shallow sediments on the isthmus. Initial calibration held the anisotropy ratio constant at several values within this range for all sediments in the domain. Extremely low vertical gradients in the deeper

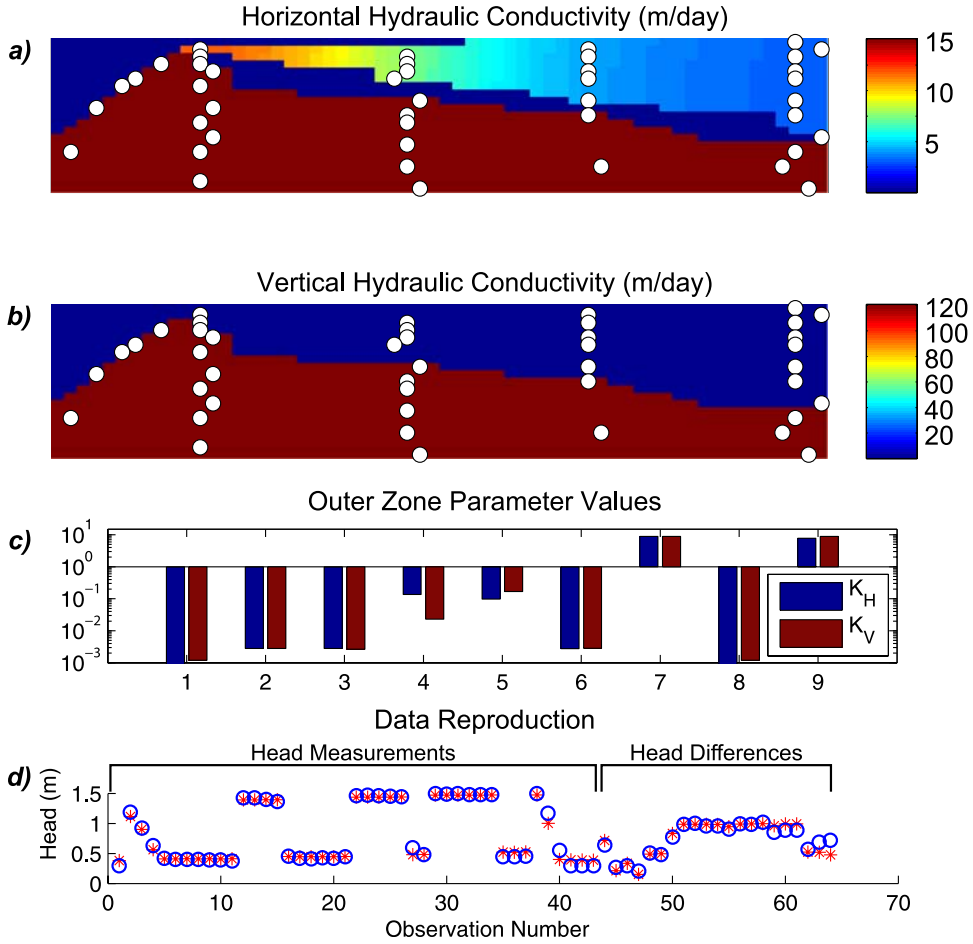

Figure 8. (a) Initial horizontal and (b) vertical hydraulic conductivity results for the isthmus core and (c) outlying homogeneous zones based only on head observations and a flow model (MODFLOW). (d) Reproduction of head and head difference measurements. Note that the highest level in Figure 8a is ~64, but the scale is truncated to show detail of variation in the shallow sand facies association. Figure not to scale.

zone below the silt (Figure 7) indicate that, in the absence of other information, the head data alone may indicate an anisotropy ratio less than unity. Horizontal and vertical hydraulic conductivity were, therefore, decoupled and estimated independently. The results from *Kenoyer* [1988] for the shallowest zone, and general understanding that anisotropy in water-worked sediments associated with glacial outwash material should be greater or equal to unity served as a qualitative check on performance of the inversion.

[44] The prior values estimated using PEST on the lumped zonal first model were used both as starting values for modeling and as diffuse prior information. These values are treated as prior information for the present inversion and were assumed uncorrelated. The values for covariance of prior information were selected to make the importance of the prespecification minimal while stabilizing the inversion. The starting values for structural parameters θ and σ_R^2 were selected assuming slightly more variability in the shallowest sand unit and smoother solutions in the deeper sand. These initial structural parameters favored a simple, flat solution from which complexity in the form of heterogeneity could be introduced by the algorithm if supported by the data, in keeping with the principle of parsimony. The covariance of the structural parameters was selected to keep the structural parameters from varying wildly. Further description of the estimation of structural parameters and the use of diffuse prior information are discussed in section S2 of Text S1 in the auxiliary material.

3.4. Staged Inversion

[45] The model development in the Trout Lake watershed provided an opportunity to test the parameter estimation process with three distinct hydrogeochemical data sets: head, oxygen (δ^{18}O), and tritium. The inclusion of these three independent data sets adds confidence to our model interpretations because of their specific strengths for providing inferences about parameter estimation and subsequent flow path delineation. Head measurements, although instantaneous, provide a check on the general direction and magnitude of groundwater flow, whereas the water isotopes (δ^{18}O and tritium) are powerful tracers of the water molecule itself (i.e., not a reactive solute) and provide strong evidence of water origination locations and time of entry into the groundwater system.

[46] In the following sections, we first discuss the three types of data available and then the implementation and

Table 1. Lookup Table to Reference the Observation Index Value in the Results in Figures 9, 12, and 16 to the Well Locations in Figure 3 and the Raw Data in Figures 7, 11, and 14[a]

Index	δ^{18}O	Head	^{3}H	Head Difference
1	BW1	K104	K79_2006	
2	BW2	K78	K80_2006	
3	BW3	K6	K81_1998	
4	BW4	K79	K81_2006	
5	BW5	K80	K5_1992	
6	K1	K81	K5_1998	
7	K100	K82	K5_2006	
8	K101	K5	K98_2006	
9	K102	K98	K96_1998	
10	K103	K97	K96_2006	
11	K104	K96	K95_1998	
12	K105	K73	K76_1998	
13	K2	K74	K76_2006	
14	K5	K75	K77_2006	
15	K6	K95	K92_1998	
16	K66	K76	K92_2006	
17	K67	K77	K90_1998	
18	K68	K94	K90_2006	
19	K69	K92	K91_1998	
20	K70	K90	K91_2006	
21	K71	K91	K67_1992	
22	K73	K66	K70_1992	
23	K74	K67	K70_1998	
24	K75	K68	K70_2006	
25	K76	K69	K71_1992	
26	K77	K70	K101_2006	
27	K78	K71	K102_2006	
28	K79	K102	K87_1992	
29	K80	K84	K87_1998	
30	K81	K85	K87_2006	
31	K82	K86	K88_1992	
32	K85	K87	K89_1992	
33	K86	K88	K99_1998	
34	K87	K89	K99_2006	
35	K88	K99	K1_1992	
36	K89	K103	K2_1992	
37	K90	K100		
38	K91	K1		
39	K92	K2		
40	K94	BW1		
41	K95	BW3		
42	K96	BW4		
43	K97	BW5		
44	K98			K78 − BW1
45	K99			K78 − K6
46				K6 − K79
47				K79 − K80
48				K6 − K80
49				K75 − K6
50				K75 − K79
51				K75 − K80
52				K75 − K81
53				K95 − K81
54				K95 − K82
55				K95 − K76
56				K69 − K76
57				K70 − K76
58				K70 − K77
59				K70 − K71
60				K88 − K71
61				K89 − K71
62				K2 − K71
63				K2 − K102
64				K2 − K99

[a]Multiple observation indices for ^{3}H correspond to discrete time samples, identified by year following the underscore.

results for the three data types cumulatively. First we examine parameter estimation only using head data and a flow model. Initial results based only on heads motivated revision of the zone setup. Following this revision, we add δ^{18}O data requiring the addition of a pseudo steady state transport model, and finally we add tritium data and an associated transient transport model.

3.5. Head Inversion

[47] In most hydrogeologic investigations, observations of head are the easiest and least expensive to obtain. However, groundwater flow is modeled as a diffusive process, so point measurement head data alone are not particularly sensitive to heterogeneity in the subsurface [see, e.g., *Haitjema*, 1995, p. 274]. The first attempt to estimate hydraulic conductivity used only head data with a forward flow model MODFLOW-2005 [*Harbaugh*, 2005]. Figure 7 shows head data collected on 24–26 May 1999 along with lake levels in Crystal and Big Muskellunge lakes. The distribution of head values in the isthmus is treated as steady state as the upgradient and downgradient lakes fluctuate in concert [*Kim et al.*, 1999] with variability damped out in the long term.

[48] The steep drop in head over a short distance seen in Figure 7 is indicative of a laterally extensive low hydraulic conductivity feature restricting flow between the upper and lower aquifers. Even without a model, this observation is highly informative, especially in the geologic context of a site in which glacial outwash sediments are punctuated by the deposition of fine-grained lacustrine materials associated with buried ice blocks and temporary lakes [e.g., *Attig*, 1985]. It is desirable to provide the inversion with the subjective knowledge that the difference in head over the silt is important.

[49] To calculate the sensitivity of the difference between two observations the corresponding rows of **H** are subtracted and the result is appended to the end of the **H** matrix, while the difference in head values is appended to the end of the observation vector. However, the new rows of **H** are linear combinations of previously existing rows and the relative rank (rank compared to dimension) of **H** decreases. At the limit, if all subtractions were made, the linearly dependent rows of **H** would completely eclipse the informative rows. Another impact of appending row differences onto **H** is the amplification of certain sensitivity eigenvalues. This, combined with double accounting in the observation vector and, therefore, in the epistemic uncertainty covariance matrix, enhances the importance placed on the wells between which head differences are evaluated. The quantitative amplification is consistent with the qualitative desire to inform the inverse process of the importance of head differences. The relative rank degradation discussed above indicates that a limited number of head difference targets should be included. A deeper analysis based on eigenvalue decomposition of the sensitivity matrix **H** is ongoing and beyond the scope of this research.

[50] *Kim et al.* [1999] also used head difference targets as supplemental data in their parameter estimation. The head differences in their work were distributed throughout the shallower part of the isthmus where small gradients are observed rather than focused on the larger head differences in the vicinity of the silt layer.

Table 2. Mean Squared Error Compared With Epistemic Uncertainty Term σ_R^2 Calculated by the Algorithm for Each of the Scenarios[a]

Data Set	MSE	σ_R^2	θ_1	θ_2	θ_3
Head-only original	-	-	7.43×10^{-3}	4.39×10^{-5}	6.44×10^{-22}
Head	0.0030	0.0042	-	-	-
Head difference	0.0074	0.0064	-	-	-
Head-only revised	-	-	3.32×10^{-9}	7.06×10^{-10}	3.62×10^{-11}
Head	0.0030	0.0038	-	-	-
Head difference	0.0072	0.0059	-	-	-
Head and $\delta^{18}O$	-	-	8.87×10^{-10}	1.15×10^{-8}	3.92×10^{-13}
Head	0.0351	0.2446	-	-	-
Head difference	0.0714	0.3749	-	-	-
$\delta^{18}O$	4.7760	2.300	-	-	-
Head, $\delta^{18}O$, and ^3H	-	-	8.33×10^{-36}	2.07×10^{-29}	1.50×10^{-16}
Head	0.0274	1.5241	-	-	-
Head difference	0.0641	2.3363	-	-	-
$\delta^{18}O$	3.9648	14.3321	-	-	-
^3H	202.4076	56.9149	-	-	-

[a]Structural parameters related to the prior information variogram are also presented for each scenario; θ_1, θ_2, and θ_3 correspond to the upper facies association (FA), the middle silt FA, and the deepest FA, respectively. MSE, mean squared error.

3.5.1. Using Preliminary Results to Inform Zonation

[51] Figure 8 shows the hydraulic conductivity field based only on the head and head difference data. Figure 8c shows the correspondence between observations and collocated model predictions and Table 1 indicates the well number each observation/prediction pair in Figure 8 corresponds to. The mean squared error for heads and head differences indicate correspondence with weight-adjusted epistemic uncertainty terms in Table 2 (see values for "Head-only original").

[52] The hydraulic conductivity field is generally homogeneous with the exception of the upper sand facies association where systematic heterogeneity is observed increasing from the Crystal Lake edge to the Big Muskellunge edge of the isthmus. This contradicts previous hydrogeologic field investigations inferred that this upper unit was likely to be homogeneous. Previous calibration of the lumped model was improved by adding a small zone of high hydraulic conductivity at the shallowest portion of the silt layer. Three model cells were therefore removed from their facies associations and incorporated into a small homogeneous zone (zone 10, Figure 5) to reflect the altered shallowest silt.

3.5.2. Results Based on Head-Only Parameter Estimation

[53] Addition of zone 10 discussed above in section 3.5.1 eliminated the need for heterogeneity in the upper sand facies association (Figure 9). Figure 9c and Table 2 (see values for "Head-only revised") show this change did not significantly impact misfit. However, this process highlights the importance of keeping the underlying hydrogeologic interpretation at the forefront of any parameter estimation project. For the remainder of this paper, zone 10 is included and results are based on the revised problem setup described here.

[54] The flow path was delineated using particle tracking with MODPATH [*Pollock*, 1994] on the basis of the best estimate of the head-only hydraulic conductivity field and representative particle trajectories shown in Figure 10. The upward trajectories toward Big Muskellunge Lake are driven by vertical hydraulic conductivity estimated greater than horizontal in response to extremely low gradients in the raw head data. However, although appearing well simulated, these estimated parameters were inconsistent with other knowledge of the site. The $\frac{K_H}{K_V}$ anisotropy ratio in the upper zone is 14, while in the deeper sand beneath the silt it is about 0.5. The deep zone anisotropy ratio is not geologically realistic, nor does the shallow anisotropy agree with the tracer test results [*Kenoyer*, 1986]. The excellent fit to the data and parsimonious result both highlight the possibility of nonunique calibrations when only head data are used [see also *Poeter and Hill*, 1997]. Despite the flexibility allowed by having over 2,000 free parameters, the algorithm favors a parsimonious, homogeneous answer when such an answer is consistent with the data. A rougher answer would result in overfitting, but the algorithm preempts such a mistake.

[55] The poor simulation of anisotropy indicates the need for more information to further constrain the solution beyond what head data alone provided. Rather than explicitly enforcing an anisotropy ratio, however, in section 3.6 we supplement the head data with isotope transport data seeking improved results.

3.6. Head and $\delta^{18}O$ Inversion

[56] Fractionation of water isotopes (e.g., $H_2^{18}O$ versus $H_2^{16}O$) only occurs in surface water because of evaporation processes that preferentially remove the lighter isotope ($H_2^{16}O$) resulting in $H_2^{18}O$ enrichment of the residual water. In the isotopic literature it is common to report the relative abundance of water isotopes in "delta" (δ) notation relative to a known standard, in this case VSMOW [*Coplen*, 1994] in units of per mil (‰) as

$$\delta^{18}O‰ = \frac{(^{18}O/^{16}O)_{sample} - (^{18}O/^{16}O)_{VSMOW}}{(^{18}O/^{16}O)_{VSMOW}} \times 1,000. \quad (10)$$

For example, an enriched Crystal Lake surface water sample would have a $^{18}O/^{16}O$ ratio ($\delta^{18}O$) of −3.4‰ relative to VSMOW which is 0.0‰ by definition. Little, if any, additional fractionation of water can occur at the temperature and pressure encountered in the subsurface, so a water parcel that infiltrates rapidly (i.e., terrestrial recharge rather than infiltrating after residence in a surface water body) will retain the $\delta^{18}O$ signature of the precipitation source.

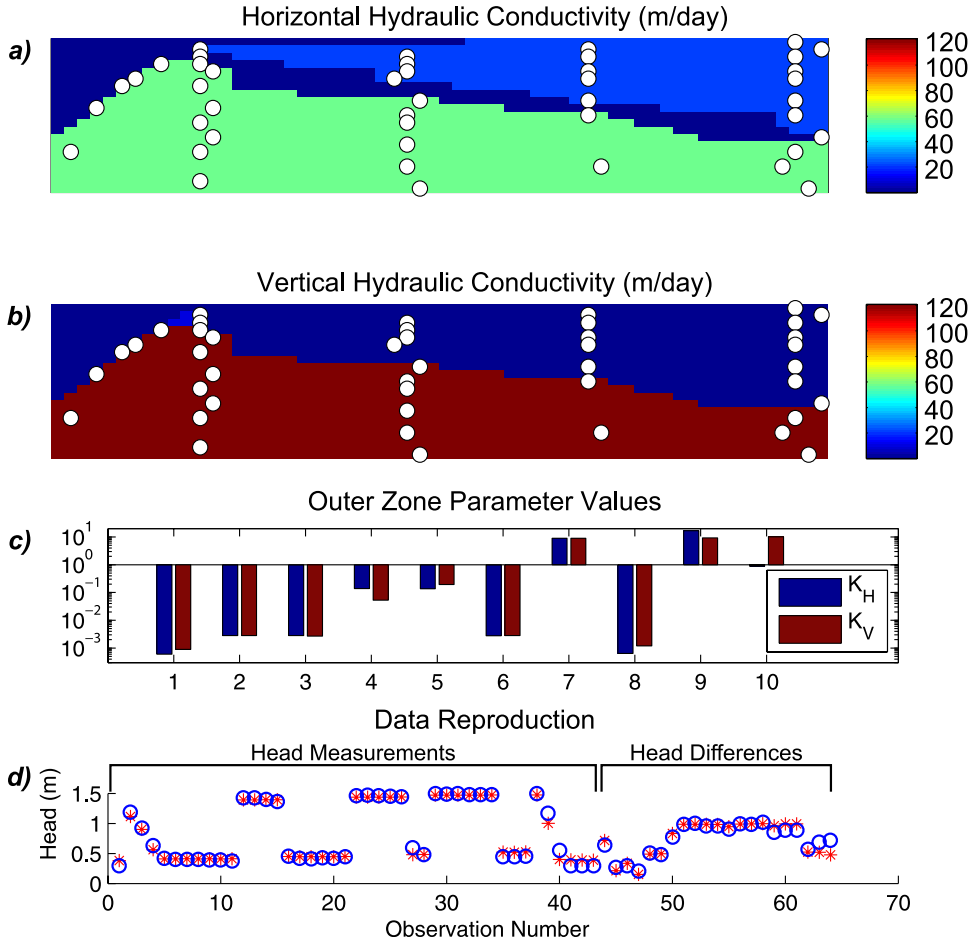

Figure 9. (a) Horizontal and (b) vertical hydraulic conductivity results for the isthmus core and (c) outlying homogeneous zones based only on head observations and a flow model (MODFLOW). (d) Reproduction of head and head difference measurements. Figure not to scale.

Considerable variability is present in the $\delta^{18}O$ signatures of precipitation (e.g., rain versus snow or frontal versus convective storm). However, because of mixing of these pulses in the unsaturated zone, an $\delta^{18}O$ signature for terrestrial recharge is distinct from that of lake recharge making $\delta^{18}O$ an excellent traceable mark of the water (i.e., a flow path indicator) [*Krabbenhoft et al.*, 1990b].

[57] Precipitation (and therefore terrestrial recharge) in northern Wisconsin typically has a $\delta^{18}O$ signature of about −11.7‰ whereas water recharged from Crystal Lake has a $\delta^{18}O$ signature of about −3.3‰ to −3.5‰ [*Krabbenhoft et al.*, 1994]. Crystal Lake water is, in fact, at or near the theoretical maximum fractionation level of surface water for this location [*Krabbenhoft et al.*, 1994]. This difference in water source signature has been effectively used in other groundwater/lake studies in the Trout Lake watershed [*Krabbenhoft et al.*, 1990b; *Bullen et al.*, 1996; *Walker and Krabbenhoft*, 1998; *Kim et al.*, 1999]. The present work augments these previous studies with additional sampling at much greater depths in the isthmus and below the downgradient lake bed.

[58] Figure 11 shows the result of this expanded sampling, illustrating a plume of water recharging through Crystal Lake and flowing through the isthmus, mixing with terrestrially recharged water in the shallowest part of the isthmus and with terrestrial water at depth from upgradient of Crystal Lake flowing under the upgradient lake. On inspection of the data, boundaries can be inferred delineating the upper and lower mixing zones around the lake water plume. Incorporating stable water isotope data into the inverse problem was valuable in regional modeling for the site [*Hunt et al.*, 2006]; inclusion into the isthmus modeling also provides information to better simulate these boundaries.

[59] The $^{18}O/^{16}O$ ratios were treated as a conservative tracer and modeled using MT3DMS [*Zheng*, 1990] in addition to MODFLOW. Because terrestrial and lake end-member water isotope compositions do not vary in time, the transport was modeled as quasi steady state by running the model to 80 years to reduce the effect of specification of initial concentrations and only recording the final concentration. Adding MT3DMS runs greatly increased computational expense from several seconds per model run using only MODFLOW to several minutes using both MT3DMS and MODFLOW. For a single forward run, this increase is not significant, but multiplied by over 2,000 to calculate sensitivities using finite difference, the increase is formidable. Thus an adjoint state formulation of MT3DMS was adopted to calculate the sensitivity matrix **H** with much lower computational effort. The adjoint formulation was

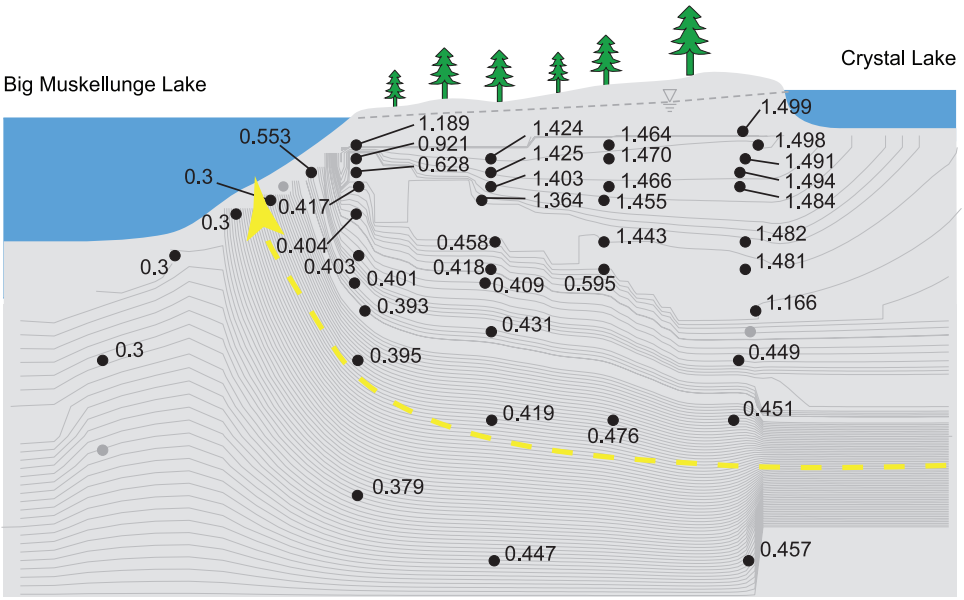

Figure 10. Pathlines based on hydraulic conductivity field estimated using only head data. The yellow dashed line is representative of the general flow path at depth. The black dots represent wells for which head data are available, the numbers represent hydraulic head in meters (referring to an arbitrary datum), and grey dots represent wells for which head data are not available. Figure not to scale.

based on work by *Samper and Neuman* [1986] as implemented by T. M. Clemo (Joint adjoint state sensitivity calculations for MT3DMS and MODFLOW 2005, manuscript in preparation, 2009) and is discussed in section S2 in Text S1 of the auxiliary material.

[60] Similarly to the head calibration, the results using both head and $\delta^{18}O$ in Figure 12 were homogeneous within the facies associations which absorb the variance required to match the data. By adding another set of data, a small degradation in fitting the head data was required to obtain the most probable best fit of both the head and water isotope data. Table 1 indicates the observation/prediction pair index that corresponds to each well. While a rougher solution might enable a marginally better fit, the algorithm has

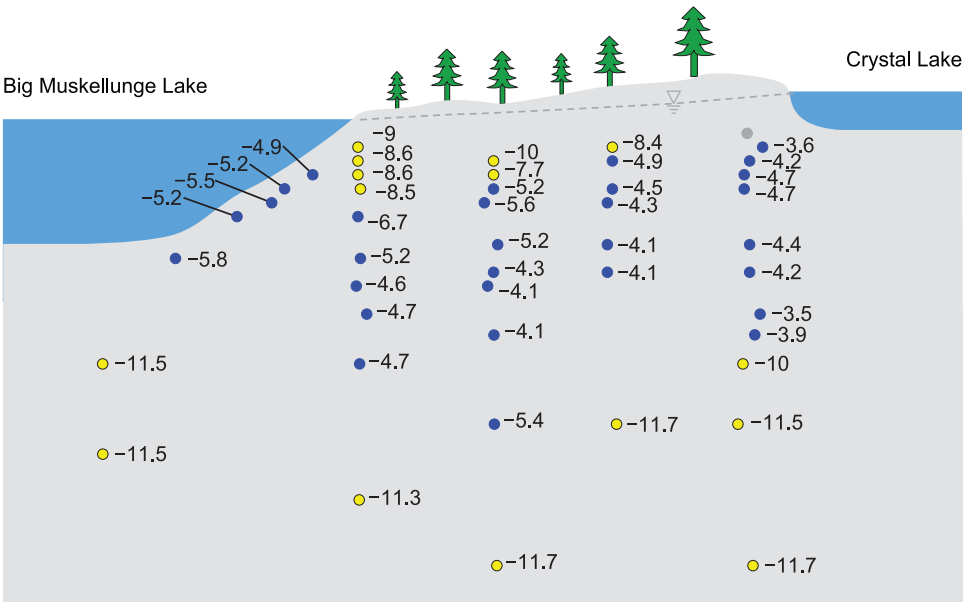

Figure 11. The $\delta^{18}O$ in ‰ Vienna Standard Mean Ocean Water (VSMOW) from groundwater samples collected in the isthmus on 26 and 27 October 1998. Background $\delta^{18}O$ is −11.7‰ VSMOW in portions of the region unimpacted by recent recharge. Blue circles indicate sampled wells with concentrations likely to be of lacustrine provenance, while yellow circles indicate wells with concentrations likely to be of terrestrial provenance. Grey circles indicate wells for which data are not available. Figure not to scale.

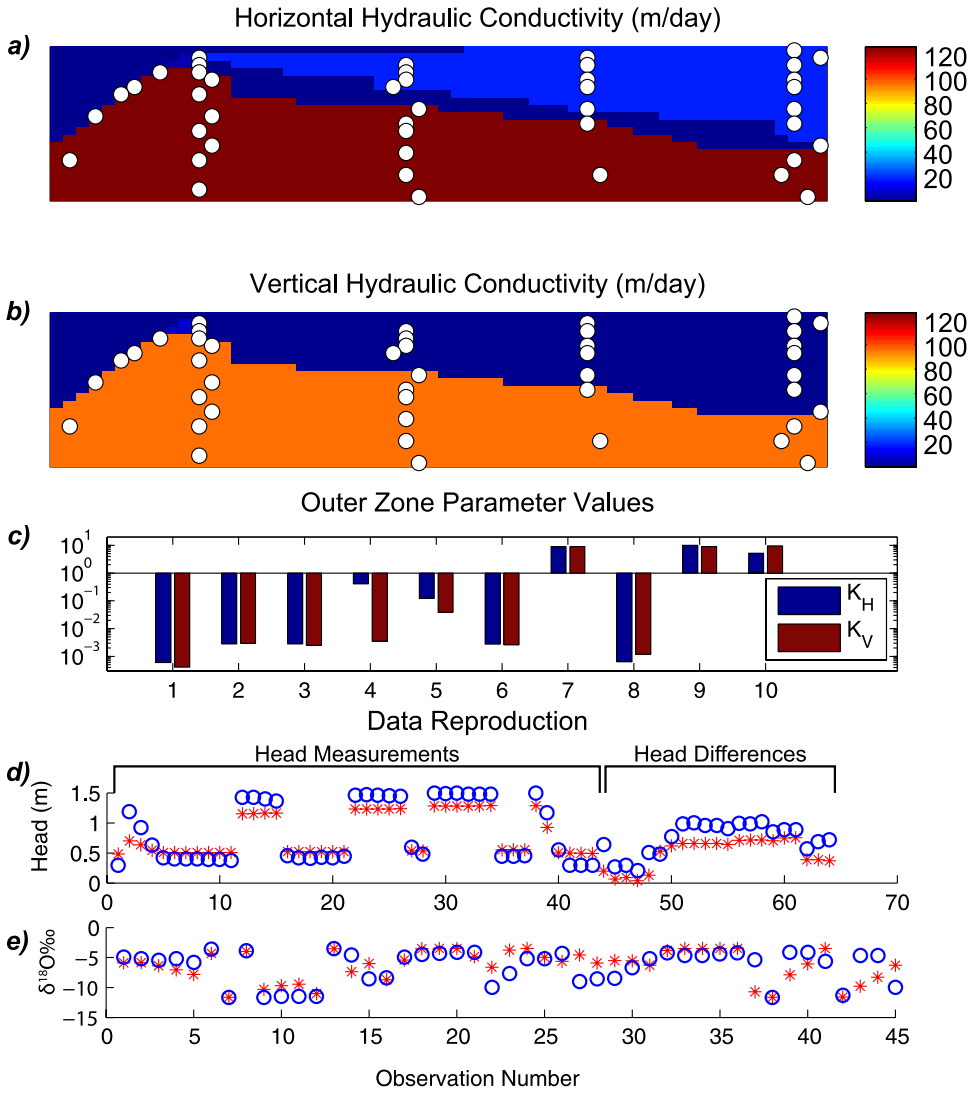

Figure 12. (a) Horizontal and (b) vertical hydraulic conductivity results for the isthmus core and (c) outlying homogeneous zones based on head observations and a flow model (MODFLOW) and $\delta^{18}O$ data and a steady state transport model (MT3D). (d) Reproduction of head and head difference measurements and (e) reproduction of $\delta^{18}O$ measurements. Figure not to scale.

successfully balanced a reasonable fit with a parsimoniously simple field.

[61] Degradation in the head fit occurs predominantly at the Big Muskellunge Lake edge of the silt in wells K78 and K6 (observation indices 2 and 3). These wells are also used in several head difference targets so the looser fit propagates to head differences. The overall head difference across the silt was also reduced somewhat with the addition of $\delta^{18}O$ data. The fit for $\delta^{18}O$ was reasonable although $\delta^{18}O$ values were underestimated at the Big Muskellunge Lake edge of the upper unit near the top of the silt. The underestimation of $\delta^{18}O$ values in K5, K98, and K97 (observation indices 8–10) is likely due to the consistent observation of upward flow from the head data conflicting with more horizontal flow indicated by the $\delta^{18}O$ data.

[62] The $\frac{K_H}{K_V}$ anisotropy ratio in the lower zone becomes geologically reasonable (1.3) in the lower aquifer, but the upper aquifer had an estimated anisotropy of about 18, not within the measured values of *Kenoyer* [1988]. Thus, the head and $\delta^{18}O$ parameter estimation results are more geologically reasonable than those obtained using only head data, but still are not consistent with all that is known about the site. Particle trajectories calculated with MODPATH (Figure 13) reflect a very different flow regime due in large part to the increase in horizontal hydraulic conductivity relative to the result based only on head data.

3.7. Head, ^{18}O, and Tritium Inversion

[63] Tritium (^3H) is the heaviest isotope of hydrogen and is abbreviated "T." Like $H_2^{18}O$, tritiated water is a specific isotope of water, but is unstable, with a half-life of 12.32 years [*Lucas and Unterweger*, 2000]. Thus, although it does not exhibit properties of a reactive solute, it decays once formed. Through nuclear proliferation and above-ground testing in the 1950s and 1960s, a distinct peak in tritium concentration in atmospheric fallout was produced around 1963 and atmospheric abundance has been decreasing ever since [*Michel*, 2005]. This maximum tritium

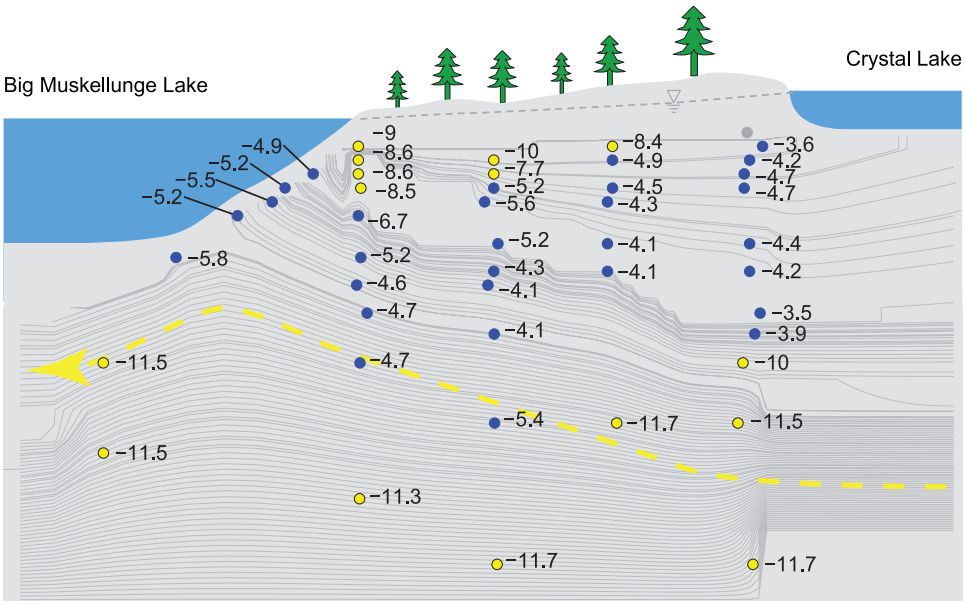

Figure 13. Pathlines based on hydraulic conductivity field estimated using head and δ^{18}O data. The yellow dashed line is representative of the general flow path at depth. The numbers next to the dots are δ^{18}O in ‰ Vienna Standard Mean Ocean Water (VSMOW) from groundwater samples collected in the isthmus on 26 and 27 October 1998. Blue circles indicate sampled wells with concentrations likely to be of lacustrine provenance, while yellow circles indicate wells with concentrations likely to be of terrestrial provenance. Grey circles indicate wells for which data are not available. Figure not to scale.

content in precipitation has been used widely by groundwater researchers as a means to estimate ages of relatively young (20th century) infiltration. The tritium concentration history for fallout can be used to hindcast travel times within flow paths and, when combined with the δ^{18}O data, it can be used to define ages of groundwater from the different sources (e.g., lacustrine versus terrestrial recharge).

[64] Tritium concentration were measured at Trout Lake in three sampling events in 1992, 1998, and 2006 (Figure 14). Tritium simulation requires two transient input functions: the precipitation function reflecting the change in tritium in the atmosphere; and the tritium concentration in the lake water reflecting mixing of the precipitation signal precipitation over multiple years (Appendix A). The two input functions are depicted in Figure 15. Because the input functions vary over time, the combined head-δ^{18}O-tritium parameter estimation requires a transient model. Annual stress periods were used with time steps controlled by MT3DMS to meet a stability criterion specified through a Courant number set at 0.5. Porosity was set to 0.3 on the basis of field measurements at the isthmus [*Kenoyer*, 1988]. The reaction rate for radioactive decay was also included, and once again, the adjoint state versions of MT3DMS and MODFLOW appreciably reduced the overall run time of the parameter estimation. For each Jacobian matrix calculation, the run time was reduced from 423 h to 5 h.

[65] On the basis of initial runs, inversion of the three different data sets and their associated forward models was difficult to stabilize. In particular, the tritium concentration values vary over 2 orders of magnitude with only two measurements exceeding 50 TU (tritium units). The difference in concentrations results in potentially unduly high weighting to the higher concentrations, and negatively impacts sensitivity calculations. The elements of the Jacobian matrix (**H**) corresponding to the higher concentration values are also elevated relative to other observations. To remove the singularities arising from disparate sensitivity values, each element of **H** was transformed by dividing the calculated sensitivity by the corresponding observation value. This transformation normalizes disparate magnitude of concentrations and results in a sensitivity matrix more appropriate to the problem solution. This alteration is commonly observed when observations are transformed, but in this case the direct adjustment to sensitivities through the **H** matrix was made for numerical stability. This follows an approach similar to that of *Li and Oldenburg* [1996] as discussed in section 2.3.

[66] The hydraulic conductivity field estimated for the combined model considering head, δ^{18}O and tritium data also resulted in good agreement between model outputs and observations (Figure 16). Table 1 provides a key to interpret the indices in the data reproduction charts in Figures 16d–16f. Even though inclusion of transport data often suggests the need for subsurface heterogeneity, the most probable parameter fields continue to be homogeneous: the simplest model consistent with the data is selected by the algorithm.

[67] The simulation of anisotropy also improved with the additional transport data. Anisotropy values for $\frac{K_H}{K_V}$ are 6.6 in the upper zone and 1.2 in the lower zone. The upper aquifer value is within the range of anisotropy measured by *Kenoyer* [1988] in the upper aquifer and the lower aquifer value is geologically reasonable. The head data reproduction was similar to the previous combined head and δ^{18}O inversion. The δ^{18}O data reproduction is nearly the same as in the previous results with some improvement in cases where the tritium data reinforce the behavior of the δ^{18}O data, and degraded where the two data types conflict. Reproduction of tritium concentrations is generally good

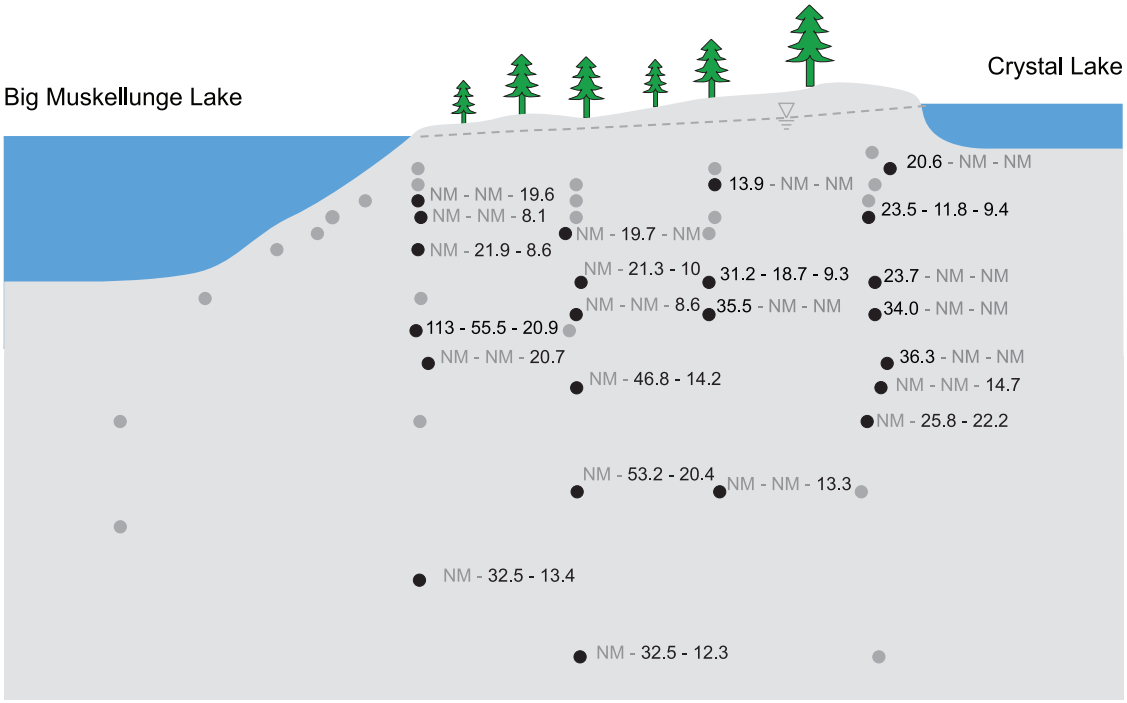

Figure 14. Tritium in tritium units (TU) from groundwater samples collected in the isthmus in 1992, 1998, and 2006. Black circles indicate sampled wells, while grey circles indicate wells for which data are not available. The three numbers adjacent to measured wells indicate the tritium concentrations measured in (from left to right) 1992, 1998, and 2006. "NM" indicates "not measured" and is applied only to wells in which at least one sample was collected. Figure not to scale.

with the exception of the high concentrations in K5 (observation indices 5–7), and the 1998 higher concentrations in K92 (observation index 15) and K90 (observation index 17). As indicated in the particle tracking results in Figures 13 and 17, these three wells are located in a region of divergent flow and may reflect low transport velocities associated with stagnation points that are not well simulated by the finite difference grid used. Furthermore, while the steady state assumption is valid here, no system is at perfect steady state and subtle variations in the flow field may confound the correct representation of stagnation points. The prominence of the diverging flow is a feature of the flow system not included in previous simulations of flow paths on the isthmus.

4. Discussion and Conclusions

[68] The flow paths delineated on the basis of coupled flow and transport simulations in this work are improved over previous work in three main ways. First, new information known about the site was included in the parameter estimation process that was not available in previous studies. This included new understanding gained from tritium sampling from all wells and additional deep field piezometer installation and sampling for $\delta^{18}O$ and tritium. Second, potential artifacts on the simulated flow paths from perimeter boundary specification were minimized by expanding the model. Thus, the flow paths calculated for the isthmus are more representative of the properties of the system in the vicinity of the isthmus. Finally, the refined general flow path at depth shown in Figure 17 reflects a conceptualization of flow not previously reported for the site, yet maintains agreement with what is known about properties of the sediments on the isthmus, and the degree of heterogeneity expected in glacial outwash. It is unlikely the refined flow path would have been obtained using only the head calibration data set regardless of parameter estimation approach.

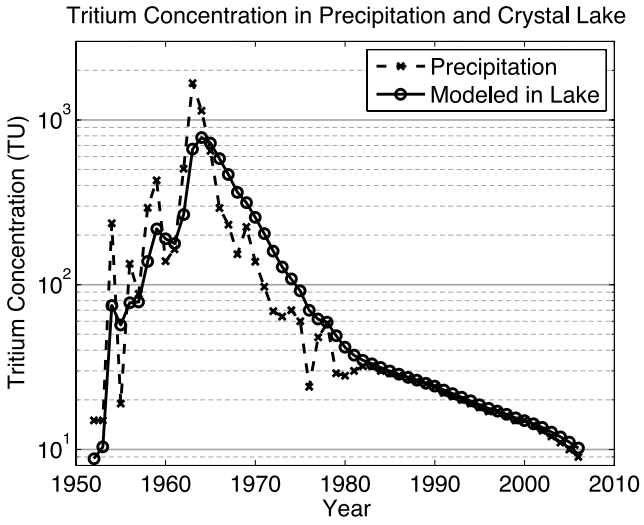

Figure 15. Tritium concentration measured in precipitation for 1953–2006 and modeled tritium concentrations in Crystal Lake based on the mixing model of *Michel and Kraemer* [1995].

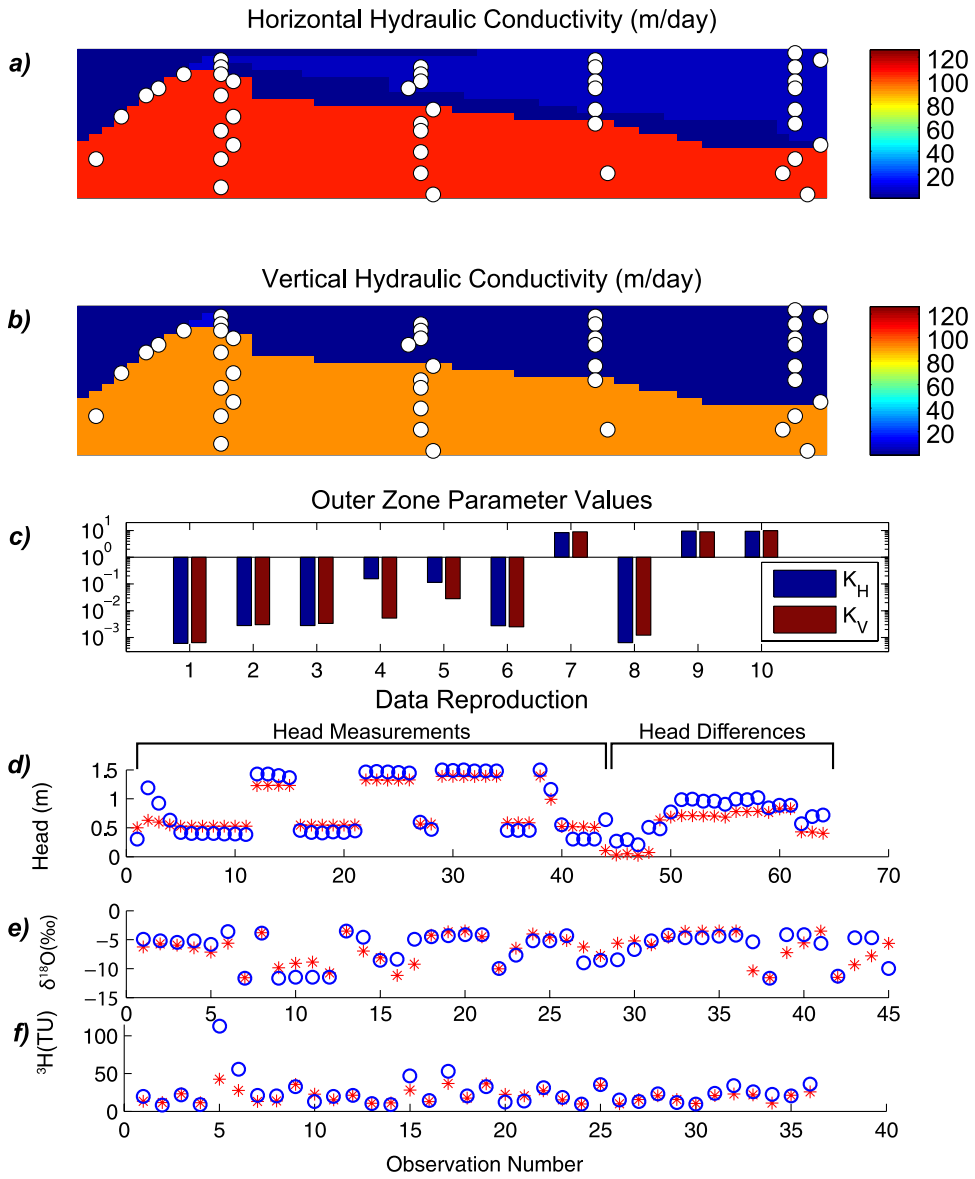

Figure 16. (a) Horizontal and (b) vertical hydraulic conductivity results for the isthmus core and (c) outlying homogeneous zones based on head observations and a flow model (MODFLOW), $\delta^{18}O$ data and a steady state transport model (MT3D), and ^3H data and a transient 54-year transport model (MT3D). (d) Reproduction of head and head difference measurements, (e) reproduction of $\delta^{18}O$ measurements, and (f) reproduction of ^3H measurements. Figure not to scale.

[69] This coupled flow and transport inversion illustrated an approach to obtain a parsimonious solution within a flexible framework using a Bayesian geostatistical approach. The approach neither insists that the hydraulic conductivity field must rigidly conform to a lumped set of homogeneous zones defined a priori, nor must include a presupposed highly heterogeneous field. Instead, the algorithm flexibly allows heterogeneity while seeking the simplest (in the sense of "freedom from useless accessories") solution that is consistent with the field data. That is, although extremely heterogeneous solutions are possible, the algorithm employed collapsed the family of possible complex parameter sets to a more homogeneous conceptualization without user intervention. This is accomplished by allowing the freedom in both a spatially distributed parameter set and the maximum likelihood calculation of structural parameters. The structural parameters drive the smoothness in the solution as an expression of prior information in the Bayesian context. Intermediate solutions were more heterogeneous but, given a similar level of fit to observations, the algorithm chooses the simpler model. This is in keeping with the principal of parsimony empowered by the data specific to the application rather than relying on preconceived notions of zonation and aquifer heterogeneity. The level of homogeneity in the parameter estimates is consistent with expectations based on both a general understanding of hydrostratigraphic conditions at the site, and the isotope data. The sharp divisions between water derived from terrestrial versus lacustrine water recharge indicates a

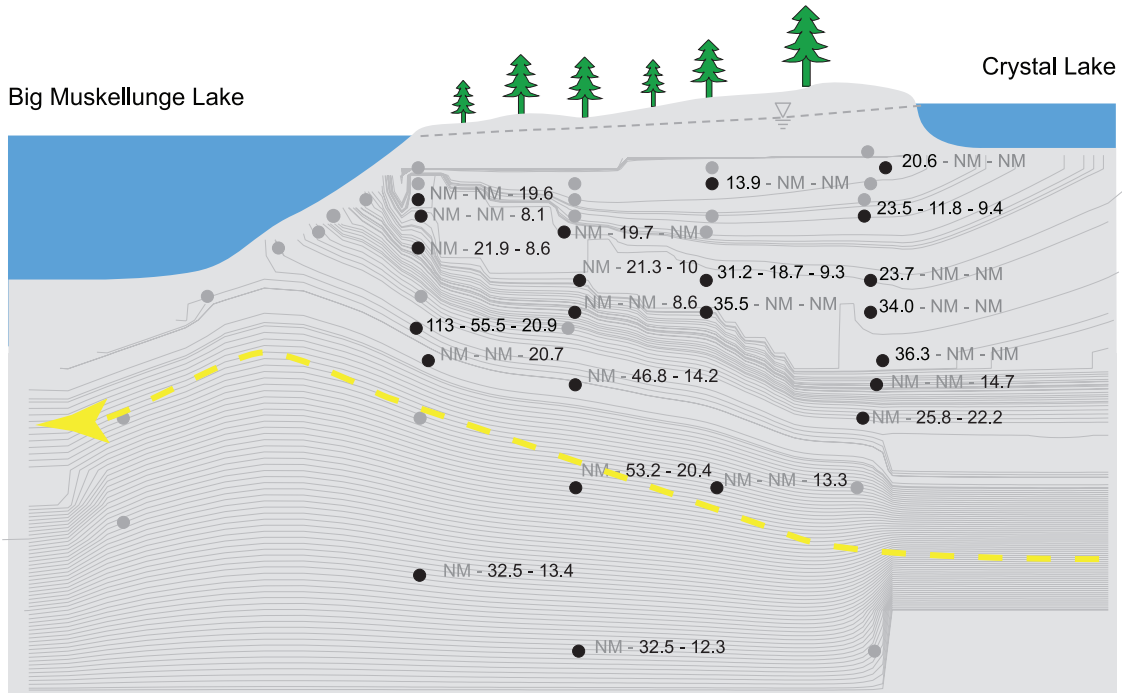

Figure 17. Pathlines based on hydraulic conductivity field estimated using head, $\delta^{18}O$, and tritium data. The yellow dashed line is representative of the general flow path at depth. The numbers next to the dots indicate tritium concentrations in tritium units (TU). The labeling scheme is further described in Figure 14. Figure not to scale.

low level of mechanical mixing of the plumes, consistent with a largely homogeneous hydraulic conductivity field.

[70] Calculation of a parameter field based on the flexible, distributed model has two additional advantages. First, the structural noise induced through lumping [*Gallagher and Doherty*, 2007] is reduced because the resulting zonation is driven by information contained in the calibration data rather than modeler preconceptions. Thus, inasmuch as a better calibration reduces the uncertainty in a model prediction, the model's predictive abilities are increased. Second, the approach used here provided an unbiased estimate of parameters which considers a wide range of potential solutions. Popular approaches to considering various conceptual models based on information theory [see, e.g., *Burnham and Anderson*, 2002; *Poeter and Anderson*, 2005] are appropriate for overdetermined regression problems (number of observations greater than number of parameters estimated). These approaches are less useful when the candidate lumped parameter sets include large numbers of parameters and high degrees of model flexibility.

[71] Ideally, parameter flexibility through allowing a fully distributed parameter field allows any zonation to be derived from information contained in the observations. This ideal was not tractable in this problem because of inversion instability caused by the laterally extensive and confining silt layer. The geologic discontinuity caused the geostatistical model of prior information used to characterize the parameter field to fail. Partitioning the problem domain (and covariance used therein) into facies associations based on these discontinuities was essential to stabilize the inversion and provide solutions. The location of stochastic discontinuities used to partition into facies associations was not ambiguous or arbitrarily determined; rather, the locations of the facies associations were easily identified in measured head data and drilling logs which confirmed the presence of the silt and estimated the location [*Kenoyer*, 1986]. The number and extent of facies association boundaries were informed by feedback from initial parameter estimation results. This highlights the interactive nature of parameter estimation: it is not a black box. The end result, then, is a set of parameter estimates including relevant complexity to the information provided by the observations used in the estimation process and consistent with other information available about the hydrogeology of the site.

[72] The flexibility afforded by high number of parameters and coupled flow and transport simulations comes at significant computational expense using traditional methods for the calculation of sensitivities. Adjoint state formulations of the two forward models, MODFLOW for flow [*Clemo*, 2007] and MT3DMS for transport (Clemo, manuscript in preparation, 2009), were powerful tools to alleviate this computational burden, which in turn makes this approach feasible. This increase in computational expense was more than offset with improved simulation capability. Parameter estimates based only on the short run time head calibration resulted in a very different flow field than longer run time calibrations that included isotope transport. The transport of $\delta^{18}O$ provides information on source of water, discriminating between water derived from lacustrine and terrestrial recharge. Tritium adds information on travel times and also contributes information on provenance due to lag and attenuation of precipitation-derived tritium after being mixed in Crystal Lake. At the limit, the flow paths near the Big Muskellunge Lake discharge area are concluded to be very different when based only on heads (Figure 10) as compared to heads and transport (Figures 13 and 17). The

importance of transport information for calibration is consistent with previous work at Trout Lake [*Anderman and Hill*, 1999; *Hunt et al.*, 2006]. Moreover, the refined flow path information added by considering isotope transport is likely to have significant implications for identification of geochemical processes operating on the isthmus. Such inclusion of transport information will likely have application in ground and surface water interaction investigations by quantifying the nature of exchange of water between the lacustrine and aquifer systems.

[73] A significant impediment to more widespread adoption of the Bayesian geostatistical inverse method in practice is the lack of a practical, publicly available tool. Part of this research effort has involved the development of such a tool that will be a module compatible with PEST [*Doherty*, 2008] and the JUPITER framework [*Banta et al.*, 2006].

Appendix A: Tritium Input Function

[74] At the Trout Lake site, terrestrial recharge on the isthmus is assumed to contain the same concentration as rainwater, while recharge from the lake is modeled with a simple mixing model [*Michel and Kraemer*, 1995]. The mixing model reflects attenuation and smearing of the signal over the residence time of water within the lake, resulting in the input function shown in Figure 15.

[75] The lake mixing model is a simple mass balance

$$\frac{dC_L}{dt} = k_I C_I - \lambda C_L - k_O C_L - k_E C_E, \qquad \text{(A1)}$$

where C_L is the average concentration in the lake, t is time, k_I is the rate constant for water gained through runoff and precipitation, C_I is the concentration in precipitation, λ is the tritium decay constant (0.05625 a^{-1}), k_O is the rate constant for water lost through outflow, k_E is the rate constant for water leaving through evaporation, and C_E is the concentration in water lost through evaporation and molecular exchange calculated following *Imboden et al.* [1977].

[76] The tritium in precipitation is based on latitudinal correlation with the Vienna record. The background initial concentration for tritium is 8.8 [*Clark and Fritz*, 1997, p. 175].

[77] **Acknowledgments.** This work was funded by the National Academies Research Associateship Program, which provided postdoctoral research funding for the first author. The USGS Office of Groundwater and the USGS Water, Energy and Biogeochemical Budgets (WEBB) program provided further funding. The authors wish to thank Bob Michel, Daniel Feinstein, Dave Saad, John Walker, John Doherty, and Wolfgang Nowak for their support and consultation. The personnel of the NSF North Temperate Lakes Long-Term Ecological Research Trout Lake Site also provided valuable support. Critical reviews by John Walker, Anna Michalak, and three anonymous reviewers also greatly improved this manuscript.

References

Anderman, E. R., and M. C. Hill (1999), A new multistage groundwater transport inverse method: Presentation, evaluation, and implications, *Water Resour. Res.*, *35*(4), 1053–1063, doi:10.1029/1998WR900114.

Aster, R. C., B. Borchers, and C. H. Thurber (2005), *Parameter Estimation and Inverse Problems*, *Int. Geophys. Ser.*, vol. 90, Elsevier Acad., Amsterdam.

Attig, J. W. (1985), Pleistocene geology of Vilas County, Wisconsin, technical report, Wis. Geol. and Nat. Hist. Surv., Madison, Wis.

Banta, E., E. Poeter, J. Doherty, and M. Hill (2006), Jupiter: Joint universal parameter identification and evaluation of reliability—An application programming interface (API) for model analysis, *U.S. Geol. Surv. Tech. Methods, 6-E1*.

Bowser, C. J., and B. F. Jones (2002), Mineralogic controls on the composition of natural waters dominated by silicate hydrolysis, *Am. J. Sci.*, *302*, 582–662, doi:10.2475/ajs.302.7.582.

Bullen, T. D., D. P. Krabbenhoft, and C. Kendall (1996), Kinetic and mineralogic controls on the evolution of groundwater chemistry and Sr-87/Sr-86 in a sandy silicate aquifer, northern Wisconsin, USA, *Geochim. Cosmochim. Acta*, *60*(10), 1807–1821, doi:10.1016/0016-7037(96)00052-X.

Burnham, K. P., and D. R. Anderson (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed., Springer, New York.

Cardiff, M. and P. Kitanidis (2008), A Bayesian level-set inversion protocol for structural zonation, *Eos Trans. AGU*, *89*(53), Fall Meet. Suppl., Abstract H41A-0834.

Casella, G. (1985), An introduction to empirical Bayes data analysis, *Am. Stat.*, *39*(2), 83–87, doi:10.2307/2682801.

Chamberlin, T. C. (1890), The method of multiple working hypotheses, *Science*, *15*, 92–96.

Clark, I. D., and P. Fritz (1997), *Environmental Iisotopes in Hydrogeology*, 175 pp., CRC Press, Boca Raton, Fla.

Clemo, T. (2007), MODFLOW-2005 ground water model—User guide to the adjoint state based sensitivity process (ADJ), *Tech. Rep. BSU CGISS 07-01*, Boise State Univ., Boise, Idaho. (Available at http://cgiss.boisestate.edu/pubs/CGISS_Techreports.html)

Collinson, J. D. (1969), Sedimentology of grindslow shales and kinderscout grit—A deltaic complex in Namurian of northern England, *J. Sediment. Petrol.*, *39*(1), 194–221.

Coplen, T. B. (1994), Reporting of stable hydrogen, carbon, and oxygen isotopic abundances, *Pure Appl. Chem.*, *66*(2), 273–276, doi:10.1351/pac199466020273.

Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170–177, doi:10.1111/j.1745-6584.2003.tb02580.x.

Doherty, J. (2008), PEST, model independent parameter estimation, User manual: 5th edition, S. S. Papadopulos, Bethesda, Md. (Available at http://www.sspa.com/PEST/)

Fienen, M., P. Kitanidis, D. Watson, and P. Jardine (2004), An application of Bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at Oak Ridge National Laboratory, *Math. Geol.*, *36*(1), 101–126, doi:10.1023/B:MATG.0000016232.71993.bd.

Fienen, M. N., J. Luo, and P. K. Kitanidis (2006), A Bayesian geostatistical transfer function approach to tracer test analysis, *Water Resour. Res.*, *42*, W07426, doi:10.1029/2005WR004576.

Fienen, M. N., T. Clemo, and P. K. Kitanidis (2008), An interactive Bayesian geostatistical inverse protocol for hydraulic tomography, *Water Resour. Res.*, *44*, W00B01, doi:10.1029/2007WR006730.

Fienen, M., C. Muffels, and R. Hunt (2009), On constraining pilot point calibration with regularization in PEST, *Ground Water*, doi:10.1111/j.1745-6584.2009.00579.x, in press.

Freeze, R. A., and J. A. Cherry (1979), *Groundwater*, Prentice-Hall, Englewood Cliffs, N. J.

Gaganis, P., and L. Smith (2001), A Bayesian approach to the quantification of the effect of model error on the predictions of groundwater models, *Water Resour. Res.*, *37*(9), 2309–2322, doi:10.1029/2000WR000001.

Gallagher, M. R., and J. Doherty (2007), Parameter interdependence and uncertainty induced by lumping in a hydrologic model, *Water Resour. Res.*, *43*, W05421, doi:10.1029/2006WR005347.

Gomez-Hernandez, J. J. (2006), Complexity, *Ground Water*, *44*(6), 782–785, doi:10.1111/j.1745-6584.2006.00222.x.

Gourdji, S. M., K. L. Mueller, K. Schaefer, and A. M. Michalak (2008), Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, *J. Geophys. Res.*, *113*, D21115, doi:10.1029/2007JD009733.

Haitjema, H. M. (1995), *Analytic Element Modeling of Groundwater Flow*, Academic, San Diego, Calif.

Harbaugh, A. W. (2005), MODFLOW-2005, the U.S. Geological Survey modular ground-water model—The ground-water flow process, *U.S. Geol. Surv. Tech. Methods, 6-A16*.

Hill, M. C. (2006), The practical use of simplicity in developing ground water models, *Ground Water*, *44*(6), 775–781, doi:10.1111/j.1745-6584.2006.00227.x.

Hill, M. C., and C. R. Tiedeman (2007), *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*, Wiley Intersci., Hoboken, N. J.

Hoeksema, R. J., and P. K. Kitanidis (1984), An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling, *Water Resour. Res.*, *20*(7), 1003–1020, doi:10.1029/WR020i007p01003.

Hunt, R. J., D. T. Feinstein, C. D. Pint, and M. P. Anderson (2006), The importance of diverse data types to calibrate a watershed model of the Trout Lake Basin, northern Wisconsin, USA, *J. Hydrol.*, *321*(1–4), 286–296, doi:10.1016/j.jhydrol.2005.08.005.

Hunt, R. J., J. Doherty, and M. J. Tonkin (2007), Are models too simple? Arguments for increased parameterization, *Ground Water*, *45*(3), 254–262, doi:10.1111/j.1745-6584.2007.00316.x.

Hunt, R. J., D. E. Prudic, J. F. Walker, and M. P. Anderson (2008), Importance of unsaturated zone flow for simulating recharge in a humid climate, *Ground Water*, *46*(4), 551–560, doi:10.1111/j.1745-6584.2007.00427.x.

Hurley, J. P., D. E. Armstrong, G. J. Kenoyer, and C. J. Bowser (1985), Ground water as a silica source for diatom production in a precipitation-dominated lake, *Science*, *227*(4694), 1576–1578, doi:10.1126/science.227.4694.1576.

Imboden, D. M., R. F. Weiss, H. Craig, R. L. Michel, and C. R. Goldman (1977), Lake Tahoe geochemical study. 1. Lake chemistry and tritium mixing study, *Limnol. Oceanogr.*, *22*(6), 1039–1051.

Jaynes, E. T. (1957), Information theory and statistical mechanics, *Phys. Rev.*, *106*(4), 620–630, doi:10.1103/PhysRev.106.620.

Juday, C., E. A. Birge, and V. W. Meloche (1938), Mineral content of the lake waters of northeastern Wisconsin, *Trans. Wis. Acad. Sci. Arts Lett.*, *31*, 223–276.

Kenoyer, G. I. (1986), Groundwater/lake dynamics and chemical evolution in a sandy silicate aquifer in northern Wisconsin, Ph.D. dissertation, Univ. of Wis.-Madison, Madison.

Kenoyer, G. J. (1988), Tracer test analysis of anisotropy in hydraulic conductivity of granular aquifers, *Ground Water Monit. Rem.*, *8*(3), 67–70, doi:10.1111/j.1745-6592.1988.tb01086.x.

Kenoyer, G. J., and M. P. Anderson (1989), Groundwater's dynamic role in regulating acidity and chemistry in a precipitation-dominated lake, *J. Hydrol.*, *109*(3–4), 287–306, doi:10.1016/0022-1694(89)90020-6.

Kenoyer, G. J., and C. J. Bowser (1992a), Groundwater chemical evolution in a sandy silicate aquifer in northern Wisconsin: 1. Patterns and rates of change, *Water Resour. Res.*, *28*(2), 579–589, doi:10.1029/91WR02302.

Kenoyer, G. J., and C. J. Bowser (1992b), Groundwater chemical evolution in a sandy silicate aquifer in northern Wisconsin: 2. Reaction modeling, *Water Resour. Res.*, *28*(2), 591–600, doi:10.1029/91WR02303.

Kim, K., M. P. Anderson, and C. J. Bowser (1999), Model calibration with multiple targets: A case study, *Ground Water*, *37*(3), 345–351, doi:10.1111/j.1745-6584.1999.tb01110.x.

Kim, K., M. P. Anderson, and C. J. Bowser (2000), Enhanced dispersion in groundwater caused by temporal changes in recharge rate and lake levels, *Adv. Water Resour.*, *23*(6), 625–635, doi:10.1016/S0309-1708(99)00050-0.

Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, *31*(10), 2411–2419, doi:10.1029/95WR01945.

Kitanidis, P. K., and E. G. Vomvoris (1983), A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water Resour. Res.*, *19*(3), 677–690, doi:10.1029/WR019i003p00677.

Krabbenhoft, D. P., M. P. Anderson, and C. J. Bowser (1990a), Estimating groundwater exchange with lakes: 2. Calibration of a three-dimensional, solute transport model to a stable isotope plume, *Water Resour. Res.*, *26*(10), 2455–2462, doi:10.1029/WR026i010p02455.

Krabbenhoft, D. P., C. J. Bowser, M. P. Anderson, and J. W. Valley (1990b), Estimating groundwater exchange with lakes: 1. The stable isotope mass balance method, *Water Resour. Res.*, *26*(10), 2445–2453, doi:10.1029/WR026i010p02445.

Krabbenhoft, D. P., C. J. Bowser, C. Kendall, and J. R. Gat (1994), Use of O-18 and deuterium to assess the hydrology of groundwater-lake systems, in *Environmental Chemistry of Lakes and Reservoirs*, *Adv. Chem. Ser.*, vol. 237, edited by L. A. Baker, pp. 67–90, Am. Chem. Soc., Washington, D. C.

Li, W., A. Englert, O. A. Cirpka, J. Vanderborght, and H. Vereecken (2007), Two-dimensional characterization of hydraulic heterogeneity by multiple pumping tests, *Water Resour. Res.*, *43*, W04433, doi:10.1029/2006WR005333.

Li, W., A. Englert, O. A. Cirpka, and H. Vereecken (2008), Three-dimensional geostatistical inversion of flowmeter and pumping test data, *Ground Water*, *46*(2), 193–201, doi:10.1111/j.1745-6584.2007.00419.x.

Li, Y. G., and D. W. Oldenburg (1996), 3-D inversion of magnetic data, *Geophysics*, *61*(2), 394–408, doi:10.1190/1.1443968.

Lucas, L. L., and M. P. Unterweger (2000), Comprehensive review and critical evaluation of the half-life of tritium, *J. Res. Natl. Inst. Stand. Technol.*, *105*(4), 541–549.

Magnuson, J. J., C. J. Bowser, and T. K. Kratz (1984), Long-term ecological research (LTER) on north temperate lakes of the United States, *Int. Ver. Theor. Angew. Limnol. Verh.*, *22*(1), 533–535.

Michalak, A. M., and P. K. Kitanidis (2002), Application of Bayesian inference methods to inverse modeling for contaminant source identification at Gloucester Landfill, Canada, in *Computational Methods in Water Resources XIV*, vol. 2, edited by S. Hassanizadeh et al., pp. 1259–1266, Elsevier, Amsterdam.

Michalak, A. M., and P. K. Kitanidis (2003), A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification, *Water Resour. Res.*, *39*(2), 1033, doi:10.1029/2002WR001480.

Michel, R. L. (2005), Tritium in the hydrologic cycle, in *Isotopes in the Water Cycle: Past, Present, and Future of a Developing Science*, edited by P. K. Aggarwal, J. R. Gat, and K. F. O. Froehlich, pp. 53–66, Springer, Amsterdam.

Michel, R. L., and T. F. Kraemer (1995), Use of isotopic data to estimate water residence times of the Finger Lakes, New York, *J. Hydrol.*, *164*, 1–18, doi:10.1016/0022-1694(94)02586-Z.

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

Moore, C., and J. Doherty (2006), The cost of uniqueness in groundwater model calibration, *Adv. Water Resour.*, *29*(4), 605–623, doi:10.1016/j.advwatres.2005.07.003.

Mueller, K. L., S. M. Gourdji, and A. M. Michalak (2008), Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements, *J. Geophys. Res.*, *113*, D21114, doi:10.1029/2007JD009734.

National Science and Technology Council Committee on Environment and Natural Resources (2007), A strategy for federal science and technology to support water availability and quality in the United States, technical report, Subcomm. on Water Avail. and Qual., XXXXXXXX.

Nowak, W., and O. A. Cirpka (2004), A modified Levenberg-Marquardt algorithm for quasi-linear geostatistical inversing, *Adv. Water Resour.*, *27*(7), 737–750, doi:10.1016/j.advwatres.2004.03.004.

Office of Management and Budget (2003), Circular A-4, technical report, Washington, D. C.

Ory, J., and R. G. Pratt (1995), Are our parameter estimators biased? The significance of finite-difference regularization operators, *Inverse Probl.*, *11*(2), 397–424, doi:10.1088/0266-5611/11/2/009.

Pint, C. D., R. J. Hunt, and M. P. Anderson (2003), Flowpath delineation and ground water age, Allequash Basin, Wisconsin, *Ground Water*, *41*(7), 895–902, doi:10.1111/j.1745-6584.2003.tb02432.x.

Poeter, E., and D. Anderson (2005), Multimodel ranking and inference in ground water modeling, *Ground Water*, *43*(4), 597–605, doi:10.1111/j.1745-6584.2005.0061.x.

Poeter, E. P., and M. C. Hill (1997), Inverse models: A necessary next step in ground-water modeling, *Ground Water*, *35*(2), 250–260, doi:10.1111/j.1745-6584.1997.tb00082.x.

Poeter, E., M. Hill, E. Banta, S. Mehl, and S. Christensen (2005), UCODE_2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation, *U.S. Geol. Surv. Tech. Methods, 6-A11*.

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, 94-464*.

Robbins, H. (1956), An empirical Bayes approach to statistics, in *Proceeding of the Third Berkeley Symposium on Mathematical Statistics*, vol. 1, edited by J. Neyman, pp. 157–163, Univ. of Calif. Press, Berkeley, Calif.

Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford Univ. Press, Oxford, U. K.

Samper, F. J., and S. Neuman (1986), Adjoint state equations for advective-dispersive transport, in *VI International Conference on Finite Elements in Water Resources*, edited by A. S. da Costa, pp. 423–437, Springer, Berlin.

Schindler, J. E., and D. P. Krabbenhoft (1998), The hyporheic zone as a source of dissolved organic carbon and carbon gases to a temperate forested stream, *Biogeochemistry*, *43*(2), 157–174, doi:10.1023/A:1006005311257.

Simpson, J. A. and E. S. C. Weiner (Eds.) (1989), *Oxford English Dictionary*, 2nd ed., Oxford University Press, Oxford, U. K.

Swift, D. J. P., B. S. Parsons, A. Foyle, and G. F. Oertel (2003), Between beds and sequences: Stratigraphic organization at intermediate scales in the Quaternary of the Virginia coast, USA, *Sedimentology*, *50*(1), 81–111, doi:10.1046/j.1365-3091.2003.00540.x.

Tikhonov, A. N., and V. I. A. Arsenin (1977), *Solutions of Ill-Posed Problems*, Halsted, Washington, D. C.

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.

Tonkin, M., and J. Doherty (2009), Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, *Water Resour. Res.*, *45*, W00B10, doi:10.1029/2007WR006678.

Walker, J. F., and T. D. Bullen (2000), Trout Lake, Wisconsin: A water, energy, and biogeochemical budgets program site, *U.S. Geol. Surv. Fact Sheet, 161-99*.

Walker, J. F., and D. P. Krabbenhoft (1998), Groundwater and surface-water interactions in riparian and lake-dominated systems, in *Isotope Tracers in Catchment Hydrology*, edited by J. J. McDonnell and C. Kendall, pp. 467–488, Elsevier, Amsterdam.

Walker, R. G. (1984), General introduction: Facies, facies sequences and facies models, in *Facies Models*, edited by R. G. Walker, 2nd ed., pp. 1–9, Geol. Assoc. of Can., Toronto, Ont., Canada.

Walker, R. G. (1992), Facies, facies models and modern stratigraphic concepts, in *Facies Models: Response to Sea Level Change*, edited by R. G. Walker and N. P. James, pp. 1–14, Geol. Assoc. of Can., St. John's, Newfoundland, Canda.

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

Winter, T. C. (1976), Numerical simulation analysis of the interaction of lakes and groundwater, *U.S. Geol. Surv. Prof., 1001*.

Yeh, W. W.-G. (1986), Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resour. Res.*, *22*(2), 95–108.

Zheng, C. (1990), MT3D, a modular three-dimensional transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater system, technical report, U.S. Environ. Prot. Agency, Washington, D. C.

T. Clemo, Intera, Inc. 1933 Jadwin Avenue, Suite 130, Richland, WA 99354, USA.

M. Fienen, R. Hunt, and D. Krabbenhoft, U.S. Geological Survey, 8505 Research Way, Middleton, WI 53562-3586, USA. (mnfienen@usgs.gov)