Estimation of various flow and mass transport parameters can be seen as the inverse problem of groundwater modeling. In this article, current methodologies for parameter estimation are classified according to whether they consider the errors in observation data, in parameter structure, and in model applications. The structures of hydraulic parameters are usually very complex and unknown. This article gives an overview on adaptive parameterization methods for simultaneously identifying the parameter values, the complexity level, and the pattern of parameter structure. Methods for model reliability analysis are described. Finally, a generalized inverse problem is introduced that can find a representative parameter for a given model application.

INTRODUCTION

The Forward Problem

Mathematical models of groundwater flow and groundwater quality (Chapter 149, Hydrodynamics of Groundwater, Volume 4, Chapter 152, Modeling Solute Transport Phenomena, Volume 4, Chapter 155, Numerical Models of Groundwater Flow and Transport, Volume 4, Chapter 150, Unsaturated Zone Flow Processes, Volume 4 and Chapter 157, Sea Water Intrusion Into Coastal Aquifers, Volume 4) can be represented in the following general form:

\[ L(u; q; p; b; x) = 0 \quad (1) \]

where \( L \) is a set of partial differential operators, \( u \) a set of state variables, \( q \) a set of control variables, \( p \) a set of model parameters, \( b \) a set of initial and boundary conditions, and \( x \) a set of spatial and time variables. When \((q; p; b)\) are given, the problem of solving the state variables \( u \) from (1) is called the forward problem (FP). The general form of FP solution can be represented by

\[ u = M(q; p; b; x) \quad (2) \]

Various analytical methods (Chapter 151, Hydraulics of Wells and Well Testing, Volume 4) and numerical methods (Chapter 155, Numerical Models of Groundwater Flow and Transport, Volume 4) have been developed for solving FP. Presently, the solution of FP is rather established in geohydrology. In the real world, however, \((q; p; b)\) cannot be determined completely by geological information and direct measurements. When incorrect \((q; p; b)\) are used in a simulation model, the result of model prediction is unreliable and the model may become useless, no matter how accurate the FP solution is. The determination of correct \((q; p; b)\) is, therefore, the key to successful groundwater modeling. Bear (1979) pointed out that there are three basic problems in groundwater modeling: the prediction problem, the calibration problem, and the management problem. A model, including its structure and parameters, must be carefully calibrated by observations of state variables and other correlated information. We must make sure that a model can give reliable prediction results before it is used, for example, in a management problem, for finding the optimal decisions. This article describes how model parameters are calibrated and how the reliability of model prediction can be guaranteed.

The Inverse Problem

For a model given by (2), the observed values of state variables, \( u^{obs} \), at observation locations and times \( x^{obs} \) can
be expressed by the following observation equation:

\[ u^{\text{obs}} = M(q; p; b; x^{\text{obs}}) + \varepsilon \]  

where \( \varepsilon \) contains both observation and model errors. It is well known that a useful model must be very carefully calibrated with all available observations, \( \{u^{\text{obs}}\} \). The problem of model calibration can be seen, in a certain sense, as the inverse problem (IP) of partial differential equations (Isakov, 1998). IP seeks model parameters \( (q, p, b) \) when state variables \( (u) \) are measured, while FP predicts state variables \( (u) \) when model parameters \( (q, p, b) \) are given. In their original meaning, FP finds “results” from “causes”, whereas IP finds “causes” on the basis of “results”. Because the same “results” may be caused by different “causes”, IP is usually ill-posed, that is, its solution (the identified parameters) may be nonunique and discontinuously dependent on data. Examples of ill-posed IPs in groundwater modeling can be found in Sun (1994).

IP can also be considered in statistical frameworks as the parameter estimation problem. With different assumptions on the probability distribution of observation errors, different criteria of parameter estimation can be derived. The statistical method gives not only the estimated parameters, but also the reliability of the estimation.

**The Study of IP in Groundwater Modeling**

In groundwater modeling, the history of studying IP can be traced all the way back to the beginning. For example, in 1950s and 1960s, analytical solutions were used to identify the hydraulic conductivity and storage coefficient around a well through fitting a curve to aquifer testing data (Chapter 151, Hydraulics of Wells and Well Testing, Volume 4). In 1970s, the “history matching” methods developed in petroleum engineering were used to identify hydraulic parameters by numerical optimization (Chavent et al., 1975). Hydrogeologists then developed their own methods for calibrating groundwater models (e.g. Carrera and Neuman, 1986; Yeh, 1986; Sun, 1994; McLaughlin and Townley, 1996; etc.). Currently, many groundwater packages include modules for solving IP. Hydrogeologists solve IP to identify hydraulic parameters, boundary conditions, pollution sources, dispersivities, adsorption kinetics, and reaction coefficients (Chapter 149, Hydrodynamics of Groundwater, Volume 4, Chapter 152, Modeling Solute Transport Phenomena, Volume 4, Chapter 155, Numerical Models of Groundwater Flow and Transport, Volume 4, Chapter 150, Unsaturated Zone Flow Processes, Volume 4). The IP-based sensitivity analysis and reliability analysis play very important roles in optimal management of groundwater resources, in optimal design of observation networks, and in feasibility study of remediation plans (Chapter 153, Groundwater Pollution and Remediation, Volume 4).

There are two inherent difficulties in solving IP for groundwater modeling: (i) the geological structure of a real aquifer is usually very complex and unknown, and (ii) data that can be used for model calibration are usually very limited, in both quantity and quality. These two difficulties are intertwined in the inverse solution process. A simple model structure (or parameter structure when the model structure is parameterized) may not fit the observed data well. On the other hand, a complicated model structure may cause over-parameterization when data are limited. Attempting to find the real structure of an aquifer and accurate parameter values by solving IP is cost-ineffective and may lead to a useless model. We must systematically consider the relationship between the complexity of model structure, the identifiability of model parameter, the sufficiency of data, and the reliability of model application. A new methodology for parameter structure reduction and robust design (Sun, 2005) may help us to find a way to construct cost-effective and reliable groundwater models.

**Parameterization**

Hydraulic parameters are usually functions of location and/or time. In this article, we use \( \theta(x) \) to represent a distributed parameter (function). Obviously, it is impossible to identify a distributed parameter with infinite degrees of freedom by a finite set of observation data. **Parameterization** refers to approximately representing a distributed parameter by a function with low degrees of freedom. A general representation of parameterization is

\[ \theta(x) \approx \sum_{j=1}^{m} \theta_j \phi_j(x, v) \]  

where \( m \) is called the dimension of parameterization, \( \theta_j \) \((j = 1, 2, \ldots, m)\) are weighting coefficients, \( \{\phi_j(x, v)\} \) is a set of basis functions with a set of shape parameters \( v \). Various parameterization methods, such as the zonation method, the finite element method, the spline method, the geostatistical method, the wavelet method, and the natural neighboring method (Sun, 1994; Liu, 1994; Sambridge, 2001), can be seen as special cases of (4). We will use \((S, \theta)\) to denote a parameterization representation (PR) of a distributed parameter \( \theta(x) \), where \( S \) represents a parameter structure determined by \( m \) basis functions, and the vector

\[ \theta = (\theta_1, \theta_2, \ldots, \theta_m)^T \]
consists of parameter values associated with the structure. The same distributed parameter may have different PRs when it is approximated by different structures.

In Sun and Sun (2002), three kinds of inverse problem are identified, namely, the classical inverse problem (CIP), the extended inverse problem (EIP), and the generalized inverse problem (GIP). In CIP, the structure $S$ is assumed to be known and only the weighting coefficients (5) need to be identified. In EIP, all components in (4) $(m, \theta, \varphi_j(x,v))$, i.e., both $S$ and $\theta$ are identified simultaneously by data. In GIP, these components are identified on the basis of the reliability of model application.

In the next section, CIP is defined in both deterministic and statistical frameworks. Various criteria solution methods for optimal parameter estimation are described. The Section “Parameter structure identification” considers the function identification problem or EIP. All components of PR are optimized by data. Statistical criteria for complexity selection and various algorithms for pattern identification are reviewed. The geostatistical inverse solution is discussed in this section as a special method of function identification. The Section “Structure reduction and experimental design” is dedicated to the solution of the GIP, through which the simplest PR is found and is used to replace the true parameter for model applications. Concepts and algorithms on robust design are discussed. The Section “Software and applications” is a short review on available software packages and application of parameter estimation in groundwater modeling.

THE CLASSICAL PARAMETER ESTIMATION PROBLEM

Inverse Problems as Statistics

In the statistical framework, the estimated parameter, $\theta$, is regarded as a random vector. The problem of parameter estimation can be considered as a procedure that transfers information from observations ($u^{\text{obs}}$) to the unknown parameter ($\theta$) through a model ($u = M(\theta;x)$), and thus reduces the uncertainty of the estimated parameter. The best method of parameter estimation should extract information from data as much as possible, and thus reduce the uncertainty of the estimated parameter as much as possible.

Let $p(\theta)$ be the joint probability density distribution (pdf) of a parameter vector $\theta$. The uncertainty associated with $p(\theta)$ is measured by its entropy (Bard, 1974):

$$H(p) = -E(\log p) = -\int_{(\Omega)} p(\theta) \log p(\theta) d\theta$$

where $E(\log p)$ is the mathematical expectation of $\log p(\theta)$, and $(\Omega)$ is the whole distribution space. The negative value of $H(p)$, that is, $-H(p) = E(\log p)$, is defined as the information content of the distribution $p(\theta)$. The prior information on parameters $\theta$ can be described by a pdf, $p_0(\theta)$, which is called the prior distribution of $\theta$. After transferring the information from the observation data to the estimated parameter(s), we obtain a new pdf, $p_*(\theta)$, which is called the posterior distribution of $\theta$. The information contents contained in the prior and posterior distributions are $-H(p_0)$ and $-H(p_*)$, respectively. Their difference, $I(p_0, p_*) = H(p_0) - H(p_*)$, measures the information content transferred from the observation data.

The posterior distribution $p_*(\theta)$ is the pdf of $\theta$ conditioned by observation data $u^{\text{obs}}$, that is, the conditional pdf, $p(\theta|u^{\text{obs}})$. On the other hand, the conditional pdf, $p(u^{\text{obs}}|\theta)$, is the pdf of $u^{\text{obs}}$ conditioned by parameter $\theta$. If there is no model error, $p(u^{\text{obs}}|\theta)$ is equal to $p(\varepsilon|\theta) = \text{the pdf of observation errors when } \theta \text{ is given}$. The conditional pdf, $p(u^{\text{obs}}|\theta)$, is usually called the likelihood function of observations and is denoted by $L(\theta)$. According to Bayes’ theorem, we have

$$p_*(\theta) = cL(\theta)p_0(\theta), \text{ where constant }$$

$$c = \left(\int_{(\Omega)} p(u^{\text{obs}}|\theta)p_0(\theta) d\theta\right)^{-1}$$

Equation (7) facilitates the transfer of information from observation data to the estimated parameter. The maximum a posterior estimation that maximizes the information transferred from observation data and prior information is given by

$$\hat{\theta} = \arg \min_{\theta} \{ -p_*(\theta) \}$$

(8)

Substituting (7) into (8) and using $\ln p_*(\theta)$ to replace $p_*(\theta)$, we have

$$\hat{\theta} = \arg \min_{\theta} \{ -\ln L(\theta) - \ln p_*(\theta) \}$$

(9)

The estimated $\hat{\theta}$ thus depends on both the likelihood function and the prior distribution.

Criteria for Optimal Parameter Estimation

Different criteria for optimal parameter estimation can be obtained on the basis of different assumptions on the prior distribution and the distribution of observation error. When $p_0(\theta)$ is a uniform distribution and the admissible region is a multidimensional box $\Theta_{ad} = (\theta_U, \theta_L)$, where $\theta_U$ and $\theta_L$ are the upper and lower bounds of the estimated parameter vector, we have the following maximum likelihood estimator (MLE):

$$\hat{\theta} = \arg \min_{\theta} \{ -\ln L(\theta) \}, \text{ s.t. } \theta \in \Theta_{ad}.$$
When the observation error vector is normally distributed with zero mean and constant covariance matrix $V_e$ (an $n \times n$ matrix, where $n$ is the number of observation data), the likelihood function can be specified as

$$L(\theta) = (2\pi)^{-n/2} |V_e|^{-1/2} \exp \left\{ -\frac{1}{2} (u^{\text{obs}} - u^{\text{cal}}(\theta))^T V_e^{-1} (u^{\text{obs}} - u^{\text{cal}}(\theta)) \right\}$$

where $u^{\text{cal}}(\theta) = M(\theta, \chi^{\text{obs}})$, and the MLE (10) reduces to the generalized least squares estimator (GLSE):

$$\hat{\theta} = \arg \min_{\theta} (u^{\text{obs}} - u^{\text{cal}}(\theta))^T V_e^{-1} (u^{\text{obs}} - u^{\text{cal}}(\theta)), \text{ s.t. } \theta \in \Theta_{ad}$$

Furthermore, if all components of $\varepsilon$ are independent from each other, $V_e^{-1}$ reduces to a diagonal matrix, $W$, and GLSE becomes the weighted least squares estimator (WLSE):

$$\hat{\theta} = \arg \min_{\theta} (u^{\text{obs}} - u^{\text{cal}}(\theta))^T W (u^{\text{obs}} - u^{\text{cal}}(\theta)), \text{ s.t. } \theta \in \Theta_{ad}$$

When all observation errors have the same variance, $W$ reduces to a unit matrix.

When $p_0(\theta)$ is normally distributed with mean $\theta_0$ and covariance matrix $V_\theta$ (an $m \times m$ matrix, where $m$ is the dimension of $\theta$), the above estimators must be changed according to (9), that is, adding $-\ln p_0(\theta)$ to the objective function of minimization. For example, the GLSE becomes

$$\hat{\theta} = \arg \min_{\theta} ((u^{\text{obs}} - u^{\text{cal}}(\theta))^T V_\theta^{-1} (u^{\text{obs}} - u^{\text{cal}}(\theta))$$

$$+ [\theta - \theta_0]^T V_\theta^{-1} [\theta - \theta_0])$$

When $V_e = \sigma^2 I$ and $V_\theta = \tau^2 I$, where $I$ denotes the unit matrix, (14) reduces to the following regularized least squares estimator (RLSE) with $\lambda = \tau^2/\sigma^2$:

$$\hat{\theta} = \arg \min_{\theta} ((u^{\text{obs}} - u^{\text{cal}}(\theta))^T V_\theta^{-1} [u^{\text{obs}} - u^{\text{cal}}(\theta)]$$

$$+ \lambda [\theta - \theta_0]^T [\theta - \theta_0])$$

In the deterministic framework, when observation error exists, we can only find an approximate inverse solution. IP is thus transformed into an optimization problem. After parameterization, the unknown parameter vector $\theta$ can be estimated by minimizing the “distance” (or a misfit function), $d_D$, between the observed data and the model output measured in the observation space $D$, that is, let

$$\hat{\theta} = \arg \min_{\theta} d_D(u^{\text{obs}}, u^{\text{cal}}(\theta)), \text{ s.t. } \theta \in \Theta_{ad}$$

where $\Theta_{ad}$ is the admissible region of the unknown parameter that can be estimated by prior information (Chapter 151, Hydraulics of Wells and Well Testing, Volume 4, Chapter 148, Aquifer Characterization by Geophysical Methods, Volume 4). A norm defined in the observation space can be used to measure the “distance” $d_D$. When the weighted L-2 norm is used, (16) is the same as the WLSE given in (13) but the weights may have different meanings. For example, we may assign larger weights to those observation wells that are more important for model prediction. Generally, the statistical criteria given above can still be used in the deterministic framework regardless of the statistical assumptions on observation error and prior information. When the observation error is not normally distributed, it is better to use L-1 norm to measure the “distance” in (16):

$$d_D(u^{\text{obs}}, u^{\text{cal}}) = \sum_{i=1}^{n} |u_i^{\text{obs}} - u_i^{\text{cal}}|$$

where $\{u_i\}$ is a set of weighting coefficients. L-1 norm is less sensitive to observation error than L-2 norm and is thus more robust. The following Kullback–Leibler (KL) misfit function provides another robust criterion for parameter identification:

$$d_D(u^{\text{obs}}, u^{\text{cal}}) = \sum_{i=1}^{n} u_i^{\text{obs}} [\ln u_i^{\text{obs}} - \ln u_i^{\text{cal}}]$$

No matter what misfit functions or criteria are used for parameter estimation, the formulated optimization problem might have multiple local minima and its solution might be unstable due to the ill-posed nature of IP and the nonlinearity of the misfit function. The regularization theory (Tikhonov and Arsenin, 1977) may significantly improve the stability of the inverse solution with

$$\hat{\theta} = \arg \min_{\theta} \left\{ d_D(u^{\text{obs}}, u^{\text{cal}}(\theta)) + \lambda R(\theta) \right\}$$

In (19), $\lambda R(\theta)$ is called a regularization term, $R(\theta)$ the regularization function and $\lambda$ the regularization factor. For example, we may choose $R(\theta) = d_P(\theta, \theta_0)$, where $d_P(\theta, \theta_0)$ is a norm defined in the parameter space that measures the “distance” between the estimated parameter $\theta$ and its prior estimation $\theta_0$. When L-2 norm is used, (19) has the same form as the statistical criterion (16). But, regularization is more general and robust. The value of $\lambda$ represents an appropriate compromise between the misfit of observations and the misfit of prior information. From this point of view, the regularization term can be considered as a penalty function, (19) can be seen as a vector optimization problem, and $\lambda$ as the weighting coefficient. Finding the optimal regularization factor is still an open problem in
real case studies (Vogel, 2002). Besides the identified parameter, the regularization term may contain also the state variable and its derivatives. In general, we may add more regularization terms to (19) if other kinds of information or measurement are available. When we have the measurements of \( k \) state variables \( v_1, v_2, \ldots, v_k \), the coupled inverse problem (Sun, 1994) can be solved by:

\[
\hat{\theta} = \arg \min_\theta \left\{ d_p[u^{\text{obs}}, u^{\text{cal}}(\theta)] 
+ \sum_{j=1}^k \lambda_j d_{D,j}[v_j^{\text{obs}}, v_j^{\text{cal}}(\theta)] + \lambda_{k+1} R(\theta) \right\} \tag{20}
\]

In the field of groundwater modeling, (20) has been used to identify the hydraulic conductivity by coupling the measurements of head, concentration, water temperature, water content, water age, flow velocity, as well as geophysical measurements of head, concentration, water temperature, water content, water age, flow velocity, as well as geophysical data (e.g. Sun and Yeh, 1990; Portniaguine and Solomon, 1998; Bravo et al., 2002; Lin and Zhang, 2004).

### Optimization Algorithms

The criteria for parameter estimation, in both statistical and deterministic frameworks, can be seen as a constrained optimization problem

\[
\hat{\theta} = \arg \min_\theta E(\theta), \theta \in \Theta_{ad} \tag{21}
\]

where the objective \( E(\theta) \) is a misfit function. Various numerical methods have been developed for solving (21) (Kelley, 1999). An iterative procedure for solving (21) includes: (1) choose a starting point \( \theta_0 \); (2) designate a way to generate a search sequence: \( \theta_0, \theta_1, \theta_2, \ldots, \theta_k, \theta_{k+1}, \ldots \); and (3) specify a termination criterion. The search sequence has the following general form \( \theta_{k+1} = \theta_k + \lambda_k d_k \), where vector \( d_k \) is called a displacement direction, and \( \lambda_k \) is a step size along the direction. Different optimization methods use different algorithms to generate \( d_k \) and \( \lambda_k \) in each iteration. This iteration process usually leads to or terminates at a local minimum. When L-2 norm is used for parameter identification consists of the following steps:

1. Define an initial population. According to the prior distribution, we choose a set of possible solutions \( \{\theta_0^0, \theta_0^1, \ldots, \theta_0^n\} \) from the admissible set \( \Theta_{ad} \) as the initial population. For a homogeneous prior distribution, the initial population can be selected randomly.
2. Evaluate the fitness of the initial population. The simulation model is used to calculate the values of \( \{E(\theta_0^0), E(\theta_0^1), \ldots, E(\theta_0^n)\} \).
3. Use genetic operators to create newer and “fitter” populations. The most common genetic operators are the selection, crossover, and mutation.
4. Evaluate the fitness of the new generation and test if a near-optimal solution has been found. If not, return to step 3 to create the next generation.

In the above procedure, all units in a generation must be encoded for genetic operations and decoded for fitness evaluation. The fitness of a parameter unit \( \theta \) is evaluated by the value of \( E(\theta) \). When significant computation effort is needed for solving FP, as in the case of groundwater modeling, GA could become very inefficient. An artificial neural network (ANN), after being trained, can approximately represent a complicated input–output relationship. Using ANN to replace the solution of the FP for fitness evaluation will significantly increase the effectiveness of GA (Morshed and Kaluarachchi, 1998). Solomatine et al. (1999) compared several global optimization algorithms for groundwater model calibration. Giacobbo et al. (2002) used GA to estimate the dispersivity parameters in a mass transport model. Shigidi and Garcia (2003) explained how to determine the optimal node number and training patterns of a neural network for transmissivity estimation.

### Sensitivity Coefficients

When a gradient-based optimization method is used to solve the inverse problem, the gradient \( g = [\partial E / \partial \theta] \) must be calculated in each iteration. When Levenberg–Marquardt method is used, the Jacobian, \( \mathbf{J} = [\partial u^{\text{cal}} / \partial \theta] \), must be
calculated. In mathematical modeling, the first-order derivatives of dependent variables with respect to parameters, such as the elements of \( g \) and \( J \), are called sensitivity coefficients. Sensitivity coefficients are important not only for parameter estimation, but also for model reliability analysis and experimental design. From the analysis of sensitivity coefficients, we can understand the model behavior. There are three calculation methods.

The perturbation method uses the finite difference approximation of derivatives. For each component \( \theta_j \) of \( \theta \), we assign an increment (perturbation) \( h_j = \lambda \theta_j \), where \( \lambda = 1\% - 10\% \), and run the simulation model to find the corresponding increment of objective function \( E \) (or the increments of the dependent variables \( u \)). The sensitivity coefficients are then calculated approximately as the proportion of increments. If the perturbation method is used to obtain the gradient \( g \) or the Jacobian \( J \) for an updated \( \theta \), we need to run the simulation model \( m + 1 \) times when the one-side finite difference approximation is used, and \( 2m + 1 \) times when the two-side finite difference approximation is used. The perturbation method is the simplest one for sensitivity analysis and has been extensively used in groundwater modeling software, but its accuracy is low and is dependent on the value of perturbation.

The sensitivity equation method differentiates the governing equation as well as its subsidiary conditions with respect to \( m \) components of \( \theta \) to obtain \( m \) sensitivity equations and then solves these equations to obtain sensitivity coefficients. For the general model given in (1), that is, \( L(u, \theta) = 0 \), its sensitivity equations with respect to \( \theta_j \) are given by

\[
\nabla_u L \frac{\partial u}{\partial \theta_j} + \nabla_\theta L \frac{\partial \theta}{\partial \theta_j} = 0, \quad j = 1, 2, \ldots, m
\]  

(23)

The gradient operator \( \nabla_u L = [\partial L/\partial u] \) generally has the same structure as the original operator \( L \), and, thus, (23) can be solved the same as the original problem. If the sensitivity equation method is used to obtain the gradient \( g \) or the Jacobian \( J \) for an updated \( \theta \), we need to solve the \( m \) sensitivity equations in (23), in addition to the original equation. The total computation effort is equal to \( m + 1 \) simulation runs.

The adjoint state method uses the variational form of the original model and solves the adjoint state equations to obtain the sensitivity coefficients. By defining the scalar product of two vector functions \( u \) and \( v \) as \( (u, v)_\Omega = \int_\Omega (u \cdot v) \, \text{d}\Omega \), where \( \Omega \) is the time/space definition domain, for a vector operator matrix \( A \), we may find a transposed adjoint operator matrix \( A^+ \), such that the following identity stands:

\[
(A u, v)_\Omega = (u, A^+ v)_\Omega
\]  

(24)

Now consider a general performance criterion \( E = \int_\Omega f(u, \theta) \, \text{d}\Omega \), where \( f(u, \theta) \) is a user-chosen function. The variation of \( E \) can be represented by

\[
\delta E = \left( \delta u, \frac{\partial f}{\partial u} \right)_\Omega + \left( \delta \theta, \frac{\partial f}{\partial \theta} \right)_\Omega
\]  

(25)

Using (24) and the variational form \( \nabla_u L \delta u + \nabla_\theta L \delta \theta = 0 \) of the original problem (1), for any vector function \( \psi \), we can rewrite (25) as

\[
\delta E = \left( \delta u, \nabla_u^+ L \psi + \frac{\partial f}{\partial u} \right)_\Omega + \left( \delta \theta, \nabla_\theta^+ L \psi + \frac{\partial f}{\partial \theta} \right)_\Omega
\]  

(26)

Let \( \psi \) be the adjoint state of the original state variable \( u \), that is, the solution of the following adjoint equation

\[
\nabla_u^+ L \psi + \frac{\partial f}{\partial u} = 0
\]  

(27)

with zero-boundary condition and zero-final condition, the first term on the right-hand side of (26) vanishes. For any component \( \theta_j \), we then have:

\[
\frac{\partial E}{\partial \theta_j} = \int_\Omega \left[ \nabla_\theta^+ L \psi + \frac{\partial f}{\partial \theta} \right] \, \text{d}\Omega, \quad j = 1, 2, \ldots, m
\]  

(28)

In the above equation, \([\cdot]\) represents the \( j \)-th component of vector \([\cdot]\). By selecting different \( f(u, \theta) \), \( E \) can be a misfit function of parameter estimation or a model output corresponding to an observation. Using the adjoint state method to calculate the gradient \( g = [\partial E/\partial \theta] \), we only need to solve the original problem (1) once and the adjoint problem (27) once. To calculate the Jacobian \( J = [\partial u^{cal}/\partial \theta] \), \( n \) adjoint problems need to be solved, where \( n \) is the number of observations. In Sun (1994), adjoint problems are derived for various groundwater flow, mass transport, and coupled problems.

When we use the sensitivity equation method (or the adjoint state method), different sensitivity equations (or different adjoint state equations) must be derived for different problems under consideration, the corresponding codes must be developed, and the correctness of each code must be validated. To avoid this manual work, a new approach called automatic differentiation (AD) is being developed that can directly differentiate a Fortran or a C++ code of forward solution to obtain the code for calculating sensitivity coefficients (Elizondo et al., 2002). A complete list on AD software packages can be found from http://www-unix.mcs.anl.gov/autodiff/AD_Tools. The AD-based sensitivity analysis has been used in parameter estimation, uncertainty analysis, and experimental design (for example, Barhen and Reister, 2003).

The Reliability of Parameter Estimation

Linearization is the simplest method to estimate the reliability of the estimated parameter. During the estimation procedure, model output \( u^{cal}(\theta) \) can be approximately regarded
as a linear function. In k-th iteration, we have $u_{cal}^{\ast}(\theta) \approx J_{(\theta_{k})}(\theta - \theta_{k}) + u_{cal}^{\ast}(\theta_{k})$, where $J = [\partial u_{cal}/\partial \theta]$. From the theory of linear regression, when GLSE (12) is used, the covariance matrix of the estimated parameter is given by

$$Cov(\theta, \theta) = (J^{T}V^{-1}_eJ + V^{-1}_\theta)^{-1}$$

(29)

Since the diagonal of the matrix gives the variances of estimation, the confidence region can then be calculated. Finding the confidence region for the estimated parameter, however, is not the sole purpose of reliability study. What we need to know is the reliability of model application when the identified parameter is used in the model. Let $g(\theta)$ be a set of given model application. There are several methods that can be used to assess the uncertainty of $g(\theta)$ on the basis of the uncertainty of the identified parameter $\theta$.

When $g(\theta)$ is close to a linear function, we may use the first-order approximation:

$$\hat{g} = g(\hat{\theta}), Cov(\hat{g}, \hat{g}) = J_{e}Cov(\hat{\theta}, \hat{\theta})J_{e}^{T}$$

(30)

where $Cov(\theta, \theta)$ is given by (29) and $J_{e} = [\partial g/\partial \theta]$ can be obtained by the methods of sensitivity analysis. A recent example of using (30) is the work of Kunstmann et al. (2002), in which the reliability of groundwater head and concentration predictions is assessed. When $g(\theta)$ is not close to a linear function, we may use the perturbation method to find the variation $\delta g$ from the variation $\delta \theta$ by solving both mean and perturbation equations (Chapter 154, Stochastic Modeling of Flow and Transport in Porous and Fractured Media, Volume 4). For example, the uncertainty associated with the model-predicted concentration distribution $C$ is measured by its variance, $\sigma_{C}^{2}$, which can be directly solved from the perturbation equation (Tang and Pinder, 1979).

In Section “Inverse problems as statistics”, we only find the maximum of the posterior $p_{\ast}(\theta)$ for parameter estimation. If we can find the whole distribution $p_{\ast}(\theta)$, the mean and variance of any model application $g(\theta)$ can be estimated by the following integrals:

$$E(g) = \int g(\theta)p_{\ast}(\theta)\ d\theta$$

and

$$Var(g) = \int \{g(\theta) - E(g)\}^{2}p_{\ast}(\theta)\ d\theta$$

(31)

The above equation is generally applicable for any nonlinear model and the unknown parameter may have any type of distribution. Letting $g(\theta) = \theta$ in (31), we can obtain the mean and variance of the estimated parameter. Moreover, with the whole distribution $p_{\ast}(\theta)$, we can search for the global optima for parameter estimation rather than a local one. To obtain $p_{\ast}(\theta)$, we can use a sampling method to find the values of $p(u|\theta)$ and then use Bayes’ theorem to find the values of $p_{\ast}(\theta) = p(\theta|u)$. The Markov Chain Monte Carlo (MCMC)-based sampling approaches, such as the Metropolis algorithm and Gibbs sampler algorithm, use a constructed random walk procedure to sample the posterior distribution. Readers may refer Gilks et al. (1996) for a detailed discussion on MCMC. The sampling method is applicable for solving IP and for uncertainty analysis when the dimension $m$ of $\theta$ is not high. Sohn et al. (2000) gave an example of using the sampling method in groundwater modeling for uncertainty assessment.

All of the above-mentioned methods of uncertainty analysis are for CIP, and thus they assume there is no parameter structure error. In practice, however, the structure error often dominates the parameter value error and is the cause of model failure. We will discuss this problem in the following sections.

PARAMETER STRUCTURE IDENTIFICATION

Parameter Structure Complexity

The statistical approach can be used not only for parameter estimation and reliability analysis, but also for determining the model complexity level. The classical model selection criteria, such as the Akaike’s Information Criterion (AIC) and the Bayesian Information Criterion (BIC), tend to select the simplest model if it can fit the observations in the same extent as complex models do (Akaike, 1974; Schwarz, 1978). For example, the Schwarz criterion is given by

$$d = d_{P}(u_{\text{obs}}, u_{\text{cal}}) + m/n \left[\ln n/2\right]$$

(32)

where $n$ is the number of data, $m$ the dimension of the unknown parameter. On the right-hand side of (32), the first term measures the fitting error, the second term is a penalty to the complexity of parameter structure. Another alternative of model complexity selection is resampling (Lahiri, 2003). In this approach, the prediction risk is estimated via cross-validation, and the model providing lowest estimated risk is chosen.

In the statistical learning theory, the following “$\varepsilon$-insensitive misfit function” is used:

$$d_{D,\varepsilon}(u_{\text{obs}}, u_{\text{cal}}) = \sum_{i=1}^{n} w_{i}d_{i,\varepsilon}^{2}$$

(33)

where $d_{i,\varepsilon} = 0$, when $|u_{i,\text{obs}} - u_{i,\text{cal}}| \leq \varepsilon$; otherwise, $d_{i,\varepsilon} = |u_{i,\text{obs}} - u_{i,\text{cal}}| - \varepsilon$. This definition means that the difference between observed and model calculated values is replaced by zero if it is less than $\varepsilon$ (the upper bound of observation error, for example). The use of (33) (or its regularization form) in groundwater modeling is strongly suggested because it can avoid fitting noised observations with an overly complex structure. The structure risk minimization...
(SRM) criterion provides a very general framework for complexity control (Vapnik, 1998). Under SRM, a set of possible models is ordered according to their complexity, the theoretical upper bounds of prediction risk can be found under certain assumptions, and a structure that provides the minimal prediction risk is chosen. Cherkassky and Ma (2003) showed that SRM outperforms AIC and BIC for their testing data sets. In groundwater modeling, we need to identify not only the complexity level but also the pattern (the pattern of hydraulic conductivity, for example). No analytical expression for the upper bound of prediction risk is available in this case.

Identification of the Optimal Parameterization

With different parameter structures for parameter estimation, we will obtain different parameter values and different fitting residuals. When the dimension of parameterization $m$ is too small (under-parameterized), the fitting residual may have a large value. On the other hand, when $m$ is too large (over-parameterized), the model prediction becomes unreliable. Moreover, even if we can find an appropriate $m$ through a model selection criterion, the identified parameter may be still very different from the real one if the structure pattern of the unknown parameter is not correctly assigned. In EIP, structure $S$ and parameter values $\theta$ are identified simultaneously by solving the following optimization problem

$$
(S, \theta) = \arg \min_{(S, \theta)} [d_{D}[u^{\text{obs}}, u^{\text{cal}}(S, \theta)] + \lambda d_{F}[(S, \theta) - (S_{0}, \theta_{0})]]
$$

(34)

where $u^{\text{cal}}(S, \theta) = M(S, \theta; \xi^{\text{obs}})$. The second term on the right-hand side of the above equation is a regularization term, in which $(S_{0}, \theta_{0})$ is a prior guess of the estimated parameter representation $(S, \theta)$. This term may be replaced by a constraint $(S, \theta) \in \Theta_{\text{ad}}$. (39) is a combinatorial optimization problem and is very difficult to solve because the dimension of the shape vector may be high. In practice, we can only identify an approximation of the true structure because of data limitation.

Sun and Yeh (1985) was the first to solve problem (34) to identify the hydraulic conductivity of a heterogeneous aquifer without knowing its structure. The shape vector that determines the parameter structure consists of the coordinates of a set of moving basis points and two variable parameters used for determining the type of basis functions, from discrete to continuous. Problem (34) is solved as a min–min optimization problem. The number of basis points is increased one by one. For each increase in basis points, the outer minimization loop finds the optimal structure (shape vector); and for each structure, the inner minimization loop finds the optimal parameter values associated with the structure. The same problem (34) was solved by Zheng and Wang (1996) with simulated annealing to find the global optimization of structure parameters. Heredia et al. (2000) presented the following criterion to replace the objective function given in (34):

$$
d_{D}[u^{\text{obs}}, u^{\text{cal}}(S, \theta)] + m \ln \left( \frac{n}{2\pi} \right) + \ln \left| \det \mathbf{F} \right| - 2 \ln p_{0}(S, \theta)
$$

(35)

where $m$ is the dimension of $\theta$, $n$ the number of data, $\det \mathbf{F}$ is the determinant of the information matrix, $\mathbf{F} = (\mathbf{J}^{T}\mathbf{J})^{-1}$, $p_{0}(S, \theta)$ contains the prior information transferred from geology or geophysics on both parameter structure and parameter values. The first term of (35) measures the data fitting error, the second term is the penalty to the structure complexity, the third term is the penalty to the parameter uncertainty, and the final term is the prize to the prior information. In Heredia et al. (2000), the identified zonation pattern is obtained by gradually modifying an initially guessed pattern from the available prior information. Recently, Amur et al. (2002) presented an adaptive procedure for pattern identification. In each iteration, a refinement indicator and a coarsening indicator are calculated to determine if a zone should be refined or coarsened. In Tsai et al. (2003), a genetic algorithm (GA), a grid search method, and a quasi-Newton algorithm are combined to find the optimal pattern of parameter structure in three dimensions. These optimization processes need huge computational effort.

In order to make the solution of EIP practical, Sun and Sun (2002) presented a simple tree-regression procedure that can find a nearly optimal parameter structure with less computational effort. In this procedure, the following nested structure sequence is formed by gradually increasing the complexity:

$$
S_{1} \subset S_{2} \subset S_{3} \subset \cdots \subset S_{m} \subset S_{m+1} \subset \cdots
$$

(36)

In the above equation, $S_{1}$ is a homogeneous structure, $S_{m+1}$ is generated by dividing a selected zone of $S_{m}$ into two zones with a linear boundary. The parameter value associated with the selected zone is the most sensitive one to the misfit function. The location of the linear boundary and the parameter values associated with all $m + 1$ zones are then determined by the method of Sun and Yeh (1985). This sequence is terminated when one of the following criteria is satisfied: (1) the value of $E(\hat{\theta}_{m})$ becomes small compared with the observation error; (2) the difference between $E(\hat{\theta}_{m})$ and $E(\hat{\theta}_{m-1})$ becomes small; (3) there are no significant changes in parameter values when one zone is divided into two zones; and (4) $E(\hat{\theta}_{m})$ becomes insensitive to all zones. In this case, no more information can be extracted from the observation data. With this tree-regression algorithm, the over-parameterization problem
can be avoided, the dimension of the shape vector is minimized, and all parameters involved in the optimization procedure are sensitive to the observation data.

The Geostatistical Method for Parameter Estimation

We treat the unknown parameter as a random variable because it is identified from the observation equation that contains uncontrollable observation and model errors. From the point of view of geostatistics (Wackernagel, 2003), a distributed physical parameter itself, such as the hydraulic conductivity, may be regarded as a spatial random field because of the variability of natural formations at different scales (Chapter 154, Stochastic Modeling of Flow and Transport in Porous and Fractured Media, Volume 4). Usually, a random field is described approximately by its first two moments. A geostatistical inverse method finds the mean and covariance functions for the estimated random field \( \theta(x) \) using the measurements of the fields (kriging) and the measurements of correlated fields (co-kriging).

In groundwater modeling, for example, the estimated random field is often the log-hydraulic conductivity \( \theta(x) = \ln K(x) \), and the correlated fields may be hydraulic head, concentration, water temperature, water content, and so on. The log-hydraulic conductivity is often assumed to be normally distributed with an exponential type of covariance function.

For any two points \( x_i \) and \( x_j \), we assume

\[
Cov_{\theta\theta}(x_i, x_j) = \psi_1 \delta_{ij} + \psi_2 \exp \left( -\frac{d_{ij}}{\psi_3} \right) \tag{37}
\]

where \( \delta_{ij} \) is the Kronecker delta, \( d_{ij} \) is the distance between \( x_i \) and \( x_j \), and \( \psi = (\psi_1, \psi_2, \psi_3) \) are statistical parameters to be identified. The value of the distributed parameter estimated by the geostatistical method at any unmeasured location \( x \) is a linear combination of all available measurements:

\[
\hat{\theta}(x) = \sum_{i=1}^{m} \lambda_i(x) \theta^*(x_i) + \sum_{k=1}^{K} \sum_{l=1}^{nk} \mu_{kj}(x) u_{kj}^*(x_{kj}) \tag{38}
\]

where \( \{\theta^*(x_i)|i = 1, 2, \ldots, m\} \) is a set of parameter measurements, \( K \) is the number of correlated fields, \( \{u_{kj}^*(x_{kj})|j = 1, 2, \ldots, nk\}(k = 1, 2, \ldots, K) \) are \( K \) sets of measurements of the correlated fields, \( \{\lambda_i(x)\} \) and \( \{\mu_{kj}(x)\} \) are kriging and co-kriging coefficients associated with \( x \), respectively. These coefficients can be obtained by solving the kriging and co-kriging equations. The coefficient matrices of these equations depend on the auto-covariance and cross-covariance functions between the parameter field and the correlated fields. The solution of a geostatistical inverse problem consists of two stages. In the first stage, the unknown mean function (characterized by parameters \( \beta \)) and the covariance function (characterized by parameters \( \psi \)) are estimated by the MLE on the basis of all sets of measurement data. In the second stage, the same sets of measurement data are used to obtain the co-kriging estimation (38) as the most possible realization. The geostatistical inverse solution can provide distributed values of the unknown parameter for describing the spatial variability. Moreover, it can also give the variance of estimation:

\[
Var[\hat{\theta}(x) - \theta(x)] = \sigma_0^2 = \sum_{i=1}^{m} \lambda_i(x) Cov_{\theta\theta}(x, x_i) - \sum_{k=1}^{K} \sum_{j=1}^{nk} \mu_{kj}(x) Cov_{\theta\theta}(x, x_{kj}) \tag{39}
\]

On the right-hand side of (39), the first term is the unconditional uncertainty, that is, \( \psi_2 \) in (37), the second term is the reduction in uncertainty after conditioned by the parameter measurements, and the third term is the reduction in uncertainty after conditioned by the measurements of correlation fields.

There are several problems associated with the geostatistical inverse method described above. First, the estimation variance (39) is not reliable because it does not take account of the structure error. The form of the covariance structure may not be correctly assigned and the statistical parameters in the mean and covariance functions may not be well calibrated by the measured data. Second, although the identified parameter is distributed, its degree of freedom is low. Only a few statistical parameters (\( \beta \) and \( \psi \)) are adjusted to fit all measured data. As a result, the fitting residual of the MLE may be significant. To increase the degree of freedom of the parameter structure, a so-called “pivot point” method was presented (Ahmed and de Marsily, 1987). In this method, a term \( \sum_{p=1}^{np} \lambda_p(x) \theta(x_p) \) is added to the right-hand side of (38), where \( \{x_p|p = 1, 2, \ldots, np\} \) is a set of pivot points, \( \{\lambda_p(x)\} \) the kriging coefficients. Parameter values at these pivot points, \( \{\theta(x_p)\} \), are identified together with the statistical parameters by the MLE. Lavenue and Pickens (1992) included optimizing the locations of pivot points in the parameter identification procedure. This can be considered as an alternative method of adaptive parameterization.

Gomez–Hernandez et al. (1997) presented a similar methodology called “the self-calibrating approach”. The unconditioned random field is first conditioned by \( m \) measurements \( \{\theta^*(x_i)|i = 1, 2, \ldots, m\} \) to obtain a seed field. After a set of master locations \( \{x_p|p = 1, 2, \ldots, np\} \) is assigned, a method for solving CIP is then used to estimate the unknown statistical parameters as well as the parameter values at the master locations. Assume that the best fitting parameter values at the master locations are \( \hat{\theta}(x_p) = \theta^*(x_p) + \Delta \theta(x_p) \), where \( \theta^*(x_p) \) is the value of the original seed field. The seed field is then updated by kriging with all \( m \) measurement points \( \{x_i\} \) and \( np \) master locations.
\( \{x_p\} \) as the basis points, that is, let

\[
\theta^*(x) = \sum_{i=1}^{m} \lambda_i(x) \theta^*(x_i) + \sum_{j=1}^{np} \lambda_j(x) \tilde{\theta}(x_p)
\] (40)

There is no essential difference between the revised pivot point approach and the self-calibrating approach. When the unknown parameter is hydraulic conductivity, its measured values (obtained by fitting pumping test data with an analytical solution) may contain significant error. Because the parameter measurement values are regarded as “hard” data, their errors will be finally transferred into the estimated parameter structure. Moreover, when measurement data are limited and the number of pivot points is not appropriately controlled, the over-parameterization problem may occur. Kitanidis (1999) suggested using the generalized covariance functions for structure description that can produce the flattest estimate of a distributed parameter. The geostatistical method was presented in the 1980s for steady-state flow condition. After 20 years of development, it has been extended to transient flow and unsaturated flow conditions. Besides using head measurements, velocity, concentration, arriving time, water age, geophysics, and geology measurements have been also used as correlated data to condition the log-K field (Rubin et al., 1992; Harvey and Gorelick, 1995; Yeh and Zhang, 1996; Hubbard and Rubin, 2000; Medina and Carrera, 2003).

**Uncertainty Caused by Structure Error**

The error of an identified PR \((S, \theta)\) of an unknown parameter to be close to its true structure and its true values is generally impossible, cost-ineffective, and unnecessary. Instead of the trueness, we should emphasize on the reliability of model application and the cost-effectiveness of model construction. The GIP defined in Sun and Sun (2002) aims at finding a representative parameter (or a model after it is parameterized) that has the simplest structure, and yet can produce reliable results for given model applications. The basic idea is to offset the error in parameter structure by the error in parameter values. Let \(g_E(M)\) be a set of given model applications (predictions, managements, or decisions), where \(M\) represents a model or a PR \((S, \theta)\). The reliability of model application can be represented by

\[
\|g_E(M) - g_E(M_t)\|_E < \varepsilon
\] (42)

where \(\|\|_E\) is a norm defined in the model application space, \(M_t\) is the true parameter, \(\varepsilon\) is an accuracy requirement. The GIP requires finding the simplest parameter structure \(S^*\) and its associated model parameter \(\theta^*\) in the admissible region such that (42) is satisfied when \(PR M^* = (S^*, \theta^*)\) is used as \(M = (S, \theta)\) in that equation. Therefore, the parameter identified by solving GIP is a representative one for the specified model application. The so-defined GIP has the following advantages: first, the reliability of model application is incorporated into the identification procedure; second, the parameter identification problem is replaced by a weak requirement (42). This requirement may be satisfied by parameters that are not close to the true parameter in the posterior distributions with respect to all models (Hoeting et al., 1999):

\[
p(\Delta|D) = \sum_{i=1}^{K} p(\Delta|S_k, D) p(S_k|D)
\] (41)

From Bayesian theorem, the weights \(p(S_k|D) = p(D|S_k) p(S_k)/\sum_{i=1}^{K} p(D|S_i) p(S_i)\), in which \(p(D|S_k)\) is the likelihood of structure \(S_k\). Once the posterior distribution (41) is obtained, the posterior mean \(E(\Delta|D)\) and the posterior variance \(Var(\Delta|D)\) can be calculated, and the latter gives the reliability estimation of model prediction. Neuman (2003) presented effective methods to calculate \(p(\Delta|S_k, D)\) and \(p(D|S_k)\) through the maximum likelihood estimation. When the number \(K\) is large or infinite as often seen in the parameter structure identification, the calculation of (41) will be infeasible.

**STRUCTURE REDUCTION AND EXPERIMENTAL DESIGN**

**Identification of Representative Parameters**

From a practical point of view, making the identified PR \((S, \theta)\) of an unknown parameter to be close to its true structure and its true values is generally impossible, cost-ineffective, and unnecessary. Instead of the trueness, we should emphasize on the reliability of model application and the cost-effectiveness of model construction. The GIP defined in Sun and Sun (2002) aims at finding a representative parameter (or a model after it is parameterized) that has the simplest structure, and yet can produce reliable results for given model applications. The basic idea is to offset the error in parameter structure by the error in parameter values. Let \(g_E(M)\) be a set of given model applications (predictions, managements, or decisions), where \(M\) represents a model or a PR \((S, \theta)\). The reliability of model application can be represented by

\[
\|g_E(M) - g_E(M_t)\|_E < \varepsilon
\] (42)

where \(\|\|_E\) is a norm defined in the model application space, \(M_t\) is the true parameter, \(\varepsilon\) is an accuracy requirement. The GIP requires finding the simplest parameter structure \(S^*\) and its associated model parameter \(\theta^*\) in the admissible region such that (42) is satisfied when \(PR M^* = (S^*, \theta^*)\) is used as \(M = (S, \theta)\) in that equation. Therefore, the parameter identified by solving GIP is a representative one for the specified model application. The so-defined GIP has the following advantages: first, the reliability of model application is incorporated into the identification procedure; second, the parameter identification problem is replaced by a weak requirement (42). This requirement may be satisfied by parameters that are not close to the true parameter in the posterior distributions with respect to all models (Hoeting et al., 1999):

\[
p(\Delta|D) = \sum_{i=1}^{K} p(\Delta|S_k, D) p(S_k|D)
\] (41)

From Bayesian theorem, the weights \(p(S_k|D) = p(D|S_k) p(S_k)/\sum_{i=1}^{K} p(D|S_i) p(S_i)\), in which \(p(D|S_k)\) is the likelihood of structure \(S_k\). Once the posterior distribution (41) is obtained, the posterior mean \(E(\Delta|D)\) and the posterior variance \(Var(\Delta|D)\) can be calculated, and the latter gives the reliability estimation of model prediction. Neuman (2003) presented effective methods to calculate \(p(\Delta|S_k, D)\) and \(p(D|S_k)\) through the maximum likelihood estimation. When the number \(K\) is large or infinite as often seen in the parameter structure identification, the calculation of (41) will be infeasible.

**STRUCTURE REDUCTION AND EXPERIMENTAL DESIGN**

**Identification of Representative Parameters**

From a practical point of view, making the identified PR \((S, \theta)\) of an unknown parameter to be close to its true structure and its true values is generally impossible, cost-ineffective, and unnecessary. Instead of the trueness, we should emphasize on the reliability of model application and the cost-effectiveness of model construction. The GIP defined in Sun and Sun (2002) aims at finding a representative parameter (or a model after it is parameterized) that has the simplest structure, and yet can produce reliable results for given model applications. The basic idea is to offset the error in parameter structure by the error in parameter values. Let \(g_E(M)\) be a set of given model applications (predictions, managements, or decisions), where \(M\) represents a model or a PR \((S, \theta)\). The reliability of model application can be represented by

\[
\|g_E(M) - g_E(M_t)\|_E < \varepsilon
\] (42)

where \(\|\|_E\) is a norm defined in the model application space, \(M_t\) is the true parameter, \(\varepsilon\) is an accuracy requirement. The GIP requires finding the simplest parameter structure \(S^*\) and its associated model parameter \(\theta^*\) in the admissible region such that (42) is satisfied when \(PR M^* = (S^*, \theta^*)\) is used as \(M = (S, \theta)\) in that equation. Therefore, the parameter identified by solving GIP is a representative one for the specified model application. The so-defined GIP has the following advantages: first, the reliability of model application is incorporated into the identification procedure; second, the parameter identification problem is replaced by a weak requirement (42). This requirement may be satisfied by parameters that are not close to the true parameter in the posterior distributions with respect to all models (Hoeting et al., 1999):

\[
p(\Delta|D) = \sum_{i=1}^{K} p(\Delta|S_k, D) p(S_k|D)
\] (41)

From Bayesian theorem, the weights \(p(S_k|D) = p(D|S_k) p(S_k)/\sum_{i=1}^{K} p(D|S_i) p(S_i)\), in which \(p(D|S_k)\) is the likelihood of structure \(S_k\). Once the posterior distribution (41) is obtained, the posterior mean \(E(\Delta|D)\) and the posterior variance \(Var(\Delta|D)\) can be calculated, and the latter gives the reliability estimation of model prediction. Neuman (2003) presented effective methods to calculate \(p(\Delta|S_k, D)\) and \(p(D|S_k)\) through the maximum likelihood estimation. When the number \(K\) is large or infinite as often seen in the parameter structure identification, the calculation of (41) will be infeasible.
parameter space; third, the data requirement is minimized because GIP attempts to find the simplest representative parameter structure. The complexity of parameter structure is determined by the requirement of model application. Once the complexity is determined, the sufficiency of existing data can be assessed.

The true parameter is data-independent, but the solutions of CIP and EIP discussed before are data-dependent, that is, different parameter structures and values may be obtained for the same parameter when different data sets are used for parameter identification. To make the identified parameter to be independent of data, the cross-validation method is often used: using one set of data for parameter identification and another set of data for validation. This method is often ineffective in groundwater modeling because of data limitation. As a result, most of groundwater models used in real case studies are data-dependent models. Their reliability for different model applications is not guaranteed. The solutions of GIP, on the other hand, are application dependent. The following is a brief description of the stepwise regression method presented by Sun et al. (1998) for solving the GIP. Using the nested structure sequence defined in (36), for each complexity level $m$, we first solve the EIP to find the optimal PR and its fitting residual:

$$RE_m = \min_{\theta_m} \{ \|u_D^{\text{obs}} - u_D^\text{cal}(S_m, \theta_m)\|_D + \lambda \| (S_m, \theta_m) - (S^0_m, \theta^0_m)\|_F \}$$

(43)

where $u_D^{\text{obs}}$ denotes the observed state values based on an experimental design $D$. We then calculate the maximum model application error introduced by replacing $S_m$ with $S_{m-1}$, which is defined by

$$AE_m = \max_{\theta_m} \min_{\theta_{m-1}} \| g_E(S_m, \theta_m) - g_E(S_{m-1}, \theta_{m-1})\|_E,$$

s.t. $\theta_{m-1} \in \Theta(S_{m-1})$, $\theta_m \in \Theta(S_m)$

(44)

Let us consider three cases. (a) If both $AE_m$ and $RE_m$ are large, we increase $m$ to $m+1$. $S_{m+1}$ is obtained by dividing a zone of $S_m$, which is the most sensitive one to the specified model applications, into two zones. The boundary between the two zones is determined by minimizing the fitting residual $RE_{m+1}$. (b) If $AE_m$ is small, stop and use the EIP solution $(S_m, \theta_m)$ as the identified parameter. (c) If $AE_m$ is large but $RE_m$ is small, new data need to be collected.

The maximum application error defined in (44) is a special case of the structure error defined in Sun (1994) and Sun et al. (1998). Letting $(S_A, \theta_A)$ and $(S_B, \theta_B)$ be two different PRs of a distributed parameter $\theta(x)$ in its admissible region $\Theta$, the distance between them can be measured in parameter, observation, and model application spaces by $d_p$, $d_p$, and $d_E$, respectively. The overall distance $d$ between the two PRs is defined by $d = d_E + \mu d_D + \lambda d_p$, where $\mu$ and $\lambda$ are weighting coefficients. A PR $(S_B, \theta_{AB})$ is called a projection of $(S_A, \theta_A)$ onto the structure $S_B$, when

$$\theta_{AB} = \arg \min_{\theta_B} d(S_A, \theta_A; S_B, \theta_B), \text{s.t. } \theta_B \in \Theta(S_B)$$

(45)

Finding $\theta_{AB}$ from (45) is equivalent to solving a classical inverse problem. The structure error $SE(S_A, S_B)$ resulted from using parameter structure $S_B$ to replace parameter structure $S_A$ is defined by the following max-min problem:

$$SE(S_A, S_B) = \max_{\theta_B} \min_{\theta_A} d(S_A, \theta_A; S_B, \theta_B), \text{s.t. } \theta_A \in \Theta(S_A), \theta_B \in \Theta(S_B)$$

(46)

If we take $\lambda = 0$ and $\mu = 0$ in the definition of distance, $SE(S_m, S_{m-1})$ reduces to the $AE_m$ in (44). Let us introduce the following concept (Sun, 2005): A PR $(S_A, \theta_A)$ is called the worst-case parameter (WCP) for simplifying a structure $S_A$ to a structure $S_B$, if

$$SE(S_A, S_B) = \min_{\theta_B} d(S_A, \theta_A; S_B, \theta_B), \text{s.t. } \theta_B \in \Theta(S_B)$$

(47)

When we know WCP, the structure error can be obtained by solving a min problem (47) rather than a max-min problem (46). WCP is a parameter that is associated with the maximal error when the model structure is reduced. It cannot be located in any inner point of the admissible region $\Theta(S_A)$. If $\Theta(S_A)$ is a multidimensional box defined by the upper and lower bounds of each parameter component, then WCP must be located at a vertex of the multidimensional box (Sun, 2005). With this proposition, the genetic algorithm (GA) becomes very suitable for searching the WCP because each vertex can be encoded as a binary string with 0 corresponding to the lower bound and 1 to the upper bound of a parameter component. Note that WCP depends on the flow conditions and may not be unique. Once we have an effective method for calculating the structure error, the stepwise regression procedure for solving the GIP becomes more effective.

Robust Experimental Design

The objective of experimental design is either to provide the maximum information with a certain budget, or to provide certain information with the minimum budget. For groundwater modeling, the decision variables of design include (i) how to excite the system (pumping locations and rates), and (ii) how to observe the responses of the system (observation locations and frequencies). In statistics, the theory of experimental design has been well developed for linear models for which the observation equation can be represented by $u_D^{\text{obs}} = A_D \theta + \varepsilon_D$, where $A_D$ is an $n \times m$ matrix. If the observation error $\varepsilon_D$ is
normally distributed with zero mean and variance $\sigma^2$, the covariance matrix of the estimated parameter $\hat{\theta}$ is given by $\text{Cov}(\hat{\theta}) = \sigma^2 (A_D^T A_D)^{-1}$. The determinant of the covariance matrix is a measurement of the uncertainty of the estimated parameter. The matrix $A_D^T A_D$ is called the information matrix. To maximize a norm of the information matrix means to minimize the uncertainty of the estimated parameter. For example, the $D$-optimal design criterion finds a design $D^*$ from all admissible designs $\{D_{ad}\}$ such that the determinant of $A_D^T A_D$ is maximized (or the volume of the corresponding confidence ellipsoid is minimized). Therefore, for a linear model, the optimal design decisions variables can be obtained by solving an optimization problem.

A nonlinear model can be linearized by first-order approximation in the neighborhood of the unknown parameter. The sensitivity matrix $J_D = [\partial y^o_D/\partial \theta]$ is used to approximate $A_D$ and the matrix $J_D^T J_D$ is used to approximate the information matrix. All the elements of the sensitivity matrix should be evaluated at the true parameter. As a result, for nonlinear problems, the $D$-optimal design depends on the parameter to be estimated. In this case, a sequential design process is needed. The initial design is based on the initial guess of the unknown parameter, after the design is conducted in the field and the data are collected, a new design is obtained on the basis of the updated parameter. Note that a design obtained by the $D$-optimal criterion is neither necessary nor sufficient for the purpose of model application.

We have learned that the problem of parameter identification actually is a problem of how to transfer information from observation data to the estimated parameter. No identification method is effective if the quantity and quality of data are not sufficient. But how sufficient is sufficient? For a distributed parameter, the parameter structure error caused by parameterization can never be avoided. A parameter with more complicated structure needs more data to identify and vice versa. Therefore, the identification of parameter structure must be considered in the design stage. The problem of observation design for parameter structure identification is a new and interesting topic. Without knowing the complexity of the estimated parameter, we cannot decide whether the data provided by a design is sufficient. If we do not know how to determine the sufficiency of a design, how can we say a design is the optimal one? The solution of GIP provides important insight into experimental design because it can determine whether the existing data are sufficient for identifying a useful representative model. If the existing data are insufficient, we have to collect more data. A more challenging problem is thus presented: can we determine the sufficiency of an experimental design before it is actually conducted in the field? An experimental design that can provide sufficient information for identifying all parameters in the admissible region is called a robust design. We can prove that if a design is sufficient for a WCP, it must be a robust one. Once a WCP is found by GA, a robust design can be obtained through a heuristic procedure, in which the pumping rates, numbers of observation wells, and frequencies are increased gradually until the design becomes sufficient for identifying a representative parameter of the WCP. A numerical example of robust design can be found in Sun (2005).

SOFTWARE AND APPLICATIONS

Available Software Packages

The Levenberg–Marquardt algorithm is used in most parameter identification codes of groundwater modeling, such as PEST developed by Doherty (2000), UCODE developed by USGS (Poeter and Hill, 1999), and the coupled inverse solution code developed by Sun (1994). PEST has been incorporated into several packages of groundwater modeling (GMS, PMWIN, VMF, and others, see www.scisotware.com), in which MODFLOW developed by USGS is used to solve the FP. UCODE has been incorporated into MODFLOW2000 and PMWIN. Parameters used in the flow model, such as hydraulic conductivity, storage coefficient, and recharge rates can be identified separately or simultaneously. In PEST, the sensitivity coefficients are calculated by the finite difference approximation, while in UCODE, the sensitivity equation method is used. In all the above-mentioned software packages, values of objective function, fitting curves, Jacobian, and correlation between parameters can be displayed during the inverse solution procedure. The code developed by Sun (1994) can identify both flow and mass transport parameters simultaneously for groundwater quality modeling. All of these packages can only solve the CIP, that is, the unknown parameter is parameterized by the zonation method with a given pattern. Changing the zonation pattern by hand is a very time-consuming work. Therefore, codes that can identify the parameter pattern automatically are badly needed.

Applications

Most studies of inverse solution in groundwater modeling and all the methods reviewed in this article are aimed at identification of hydraulic conductivity in confined and unconfined aquifers. After 40 years of study, this is still an open problem for highly heterogeneous formations. The methodologies for solving EIP and GIP have not been extensively used in real case studies. During the last decade, using software to solve CIP for model calibration, however, has become more and more popular. Besides hydraulic conductivity, we can identify storage coefficient and specific yield when transient head measurements are
available. Boundary inflow, infiltration, and recharge rates, leaky conductance, and other sink and source terms in the flow equations can also be identified. Generally speaking, any property or any part of a model can be identified provided (i) the property is appropriately parameterized, and (ii) there are data available that are sensitive to the property.

In an unsaturated zone, the hydraulic conductivity is a function of pressure head $\psi$ or a function of water content $\theta$. It is generally expressed by an empirical formula other than parameterized by a set of basis functions. The most often used expression is

$$K(\theta) = K_s S_c^2 [1 - (S_c^{1/m})^m]^2$$

(48)

where $S_c = (\theta - \theta_r)/(\theta_s - \theta_r)$, $\theta(\psi) = \theta_r + (\theta_s - \theta_r)[1 + |x\psi|^m]^{-m}$, and $m = 1 - 1/n$ (Chapter 150, Unsat urated Zone Flow Processes, Volume 4). The function identification problem is then reduced to identifying six constant parameters: $p = (\theta_s, \theta_r, \alpha, n, K_s, \lambda)$. When concentration measurements are available, the dispersion and reaction parameters can be identified simultaneously with the hydraulic parameters. Abbasi et al. (2003) shows that $n$ and $\theta_r$ in $p$ are the most sensitive parameters to water content, while $\theta_s$ and the longitudinal dispersion coefficient $D_l$ are the most sensitive parameters to concentration.

Because of the nonlinearity of the unsaturated flow model, the least squares criterion may have many local minima, and it is better to use global optimization algorithms for inverse solution. Altmann–Dieses et al. (2002) presented an optimal experimental design method for identifying the parameters in unsaturated zones. To study groundwater contamination caused by organic compounds, we need to develop models to simulate various reactive transport processes in multiphase flow. The governing equations for this case are a set of partial differential equations that describe the mass balance of each component in each phase. There are a lot of physical, chemical, and biological parameters that must be identified before a model can be used for prediction and design purposes. In recent years, many authors used the least squares criterion to find the best fitting between model outputs and concentration data obtained from experiments (Gramling et al., 2002). The fitting procedure usually contains two steps: first, using tracer test data to identify flow and dispersion parameters, and then fitting the reactive transport data to identify mass exchange, reaction, and decay term. If the model outputs cannot fit the observed data well, more complicated expressions containing more unknown parameters may be used in the reaction model. For a nonlinear kinetic process, mass exchange, reaction, and decay terms are unknown functions of concentrations. The solution of EIP for identifying the functional structure, thus, should be considered. When the number of unknown parameters increases, we must consider the identifiability problem caused by possible correlation between parameters. For example, Sun et al. (2001) shows that only two of four parameters characterizing the attachment/detachment process of colloids are identifiable. The study for identifying complex chemical and biological parameters is being developed.

Locating contaminant sources and recovering their release history from the measurements of contaminant plumes are critical for remediation design and environmental litigation support. Three problems were considered in groundwater references: (i) find the release history when a source location is known; (ii) find the location of a source when its release history is known; and (iii) find both source locations and release history. Pollution source identification is a kind of inverse problem. Therefore, all the criteria and methodologies reviewed in the previous sections can be used once the identified source is parameterized. The unknown parameter vector $\theta$ associated with a source may consist of its coordinates $(x_s, y_s)$ and the released concentrations $\{C_s(t_j)\}_{j=1}^{m}$, where $m$ is the number of times or time periods. When the weighted least squares criterion is used, the objective function of optimization is:

$$E(\theta) = \sum_{k=1}^{K} \sum_{l=1}^{L} w_{kl} [C_{kl}^{cal}(\theta) - C_{kl}^{obs}]^2$$

where $\{C_{kl}^{obs}\}$ are observed values of the plume at $K$ observation times and $L$ locations, $\{C_{kl}^{cal}(\theta)\}$ are the corresponding model output. The objective function based on WLSE criterion is highly nonlinear and usually nonconvex. Many authors have designed different optimization algorithms, including evolution algorithms, to find the minimum of $E(\theta)$ (Mahar and Datta, 2001; Aral et al., 2001). By considering the ill-posed nature of the source identification problem, Skaggs and Kabala (1994) added a regularization term to the objective function to stabilize the inverse solution. Atmadja and Bagtzoglou (2001) used the backward beam equation for source identification, in which the advection transport is reversed in time but the dispersion transport is kept positive because that cannot be reversed. This method can assess the relative importance of each potential source. Neupauer and Wilson (1999) obtained the backward-in-time distribution of a potential source by solving the adjoint state equation that, actually, is equivalent to the solution of the backward beam equation. Snodgrass and Kitanidis (1997) and Michalak and Kitanidis (2003) used the geostatistical method for contaminant history recovery that can give also the variance of estimation error. Up to date, development of accurate and effective methods for source identification is still an open problem.
CONCLUSION

Inverse problems in groundwater modeling can be defined in both deterministic and statistical frameworks. Bayesian inference provides the most general method for parameter estimation. It can incorporate not only prior information but also observation and model errors into the estimation process. Several software packages are available for solving the classical inverse problems in groundwater modeling. Methods that can automatically identify both parameter structure and values from prior information are needed. Because the structure of a real aquifer is usually very complex and unknown, and the available data are always very limited in both quantity and quality, finding a representative model is the only feasible way in groundwater modeling. The generalized inverse problem does not require the uniqueness of inverse solution. Instead, it requires the reliability of model application. Collecting sufficient data is the key of successfully solving the inverse problem. The problem of identifying hydraulic parameters of highly heterogeneous aquifers is still not well resolved. The identification of chemical and biological reaction functions and the identification of pollution sources are being studied.

REFERENCES


