Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 281 (2003) 251–264
Journal of Hydrology
www.elsevier.com/locate/jhydrol

# Geostatistical inversion of coupled problems: dealing with computational burden and different types of data

Agustín Medina, Jesús Carrera*

*School of Civil Engineering, Technical University of Catalonia (UPC), Campus NOrd D-2, Barcelona 08034, Spain*

Received 1 September 2002; revised 15 February 2003; accepted 8 May 2003

**Abstract**

Practical application of geostatistical inversion to coupled problems is hampered by a number of difficulties. In this paper, we address two of them: first, the computational cost of sensitivity (Jacobian) matrices and, second, the evaluation of the relative weights of different types of data. Regarding the first, we revise the adjoint state equations to propose a form whose cost is independent of the number of unknown parameters and only grows with the number of observation wells. Regarding the second, we derive expressions for the relative weights of different types of data. These expressions are based on minimizing the expected likelihood, rather than the likelihood itself. The efficiency of both improvements is tested on a synthetic example. The example analyzes a wide range of groundwater flow and solute transport conditions. Yet, the expected likelihood consistently yields the optimal weights. The proposed form of the adjoint state equations leads to one order of magnitude reduction in CPU time with respect to the conventional sensitivity equations.


*Keywords:* Geostatistical inversion; Jacobian matrices; Computational burden

## 1. Introduction

Modeling groundwater flow or transport is difficult because of uncertainties in characterizing hydraulic conductivity, defining boundary conditions and estimating internal sink and sources. As a result, computed concentrations and/or heads often fail to resemble field measurements. It is then natural to try to use these measurements to reduce those uncertainties. This is what has motivated the large body of research in groundwater inverse models (see reviews by Yeh, 1986; Carrera, 1987; McLaughlin and Townley, 1996). Also, it is widely recognized that heterogeneity plays an important role when modeling flow and transport. Heterogeneity has led naturally to the so called geostatistical inversion methods, in which some parameters (usually the log-transmissivity, $Y$) are viewed as regionalized variables.

Geostatistical inversion methods can be classified in two groups, which we will call estimation and simulation inversions. The first group seeks an optimum identification of the regionalized variables (i.e. conditional expectation, minimum error estimations or similar) given all available information. If

* Corresponding author. Tel.: +34-934016890; fax: +34-934016504.
*E-mail address:* carrera@etseccpb.upc.es (J. Carrera).

0022-1694/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0022-1694(03)00190-2

the estimated fields are used for predicting linear or slightly non-linear problems, the resulting prediction will also be optimal. However, these fields may yield very poor predictions in highly non-linear problems because they (the estimated fields) are much smoother than reality. In these cases, rather than a single 'optimal' estimation, one may seek a number of equally likely simulations of the unknown fields, all of them conditioned upon all available information. This is the objective of simulation inversions.

Estimation inversion methods may be linear (Kitanidis and Vomvoris, 1983; Hoeksema and Kitanidis, 1984, 1985; Dagan, 1985; Rubin and Dagan, 1987) or non-linear (Clifton and Neuman, 1982; Carrera and Neuman, 1986a; Sadeghipur and Yeh, 1984; Hill et al., 1998; Cooley, 1982). In this context, linearization refers to the process of expressing heads as a linear function of perturbations of $Y$ around its mean. This avoids the need for iterating, thus leading to simple and computationally efficient methods. However, Carrera and Glorioso (1991) showed that iterating is profitable when $Y$ variances are large, or when the estimation relies heavily on heads and/or concentrations. As a result, non-linear methods tend to yield superior results (Zimmerman et al., 1998). It is not surprising that they have become somewhat standard (see e.g. Kitanidis, 1995; Yeh et al., 1996).

Simulation inversion is much more recent. It can be traced back to the work of Sahuquillo et al. (1992). The standard approach is the one described by Gómez-Hernández et al. (1997) and by Capilla et al. (1997). In essence, it consists of generating a random field, $Y(x)$, conditioned upon direct measurements of $Y$ and, in fact, of any other variable linearly dependent on $Y$. This random field is then perturbed so as to minimize the distance between measured and computed heads (and concentrations) while maintaining the conditioning on direct data. The resulting fields can then be considered as equally likely simulations of $Y$ fields conditioned on direct measurements of $Y$ and heads. It should be stressed that the main difference between simulation and estimation inversions relies on the ability to accommodate unusual statistical assumptions. This is true both in the input data (a fact that was exploited by Capilla et al. (1999) to invert fields with arbitrary distributions) and on the output results (by generating a number of equally probable fields, one does not have to assume that they are normal or belong to any other distribution). Aside from these differences, simulation and estimation inversions are very similar. Certainly, both are computationally demanding and both require a proper estimation of the statistical properties of the input fields.

Of all practical difficulties shared by all geostatistical inversion methods, we will stress two of them: computational burden and estimation of statistical parameters. Geostatistical inversion is computationally demanding because it involves optimization with respect to a large number of parameters. Carrera and Neuman (1986b) advocated the use of gradient search optimization methods because these only require computing the gradient at each iteration. This can be achieved at a cost independent of the number of unknowns by using the adjoint state method. However, Cooley (1985) proved that Gauss–Newton methods converge much faster. Unfortunately, these require computing the Jacobian matrix (derivatives of heads w.r.t. parameters), which implies solving a problem equivalent to that of the direct problem for each parameter. The growth of geostatistical inversion, which demands a large number of parameters, has brought gradient search methods back into fashion (Gómez-Hernández et al., 1997; Capilla et al., 1999), but at the cost of a severe loss in convergence rate and robustness. One of the objectives of this paper is to revise the adjoint state methodology of Carrera and Medina (1994) to compute the Jacobian matrices at a cost that grows with the number of wells, but is independent of the number of parameters.

Regarding statistical parameters, they have been traditionally estimated by maximum likelihood (Kitanidis and Vomvoris, 1983; Carrera and Neuman, 1986a). However, likelihood estimation is often ambiguous. For example, Carrera (1994) shows that the optimum weight between head and transmissivity data may occur not at the maximum of the likelihood function, but at an inflection point. To overcome these difficulties, Akaike (1978) had advocated the use of the expected value of the likelihood, rather than the likelihood itself. Both Honjo et al. (1994) and Kitanidis (1995) support the use of this expected value. The second objective of this work is to explore the use of the expected likelihood for the purpose of estimating the relative weight among different sources of information.

In order to achieve these two objectives (namely, revised adjoint state equations and revised objective function), we start by reviewing the methodology, so as to introduce the basic notation. We then present the revised adjoint state equations and computation of the Jacobian. Third, we present the methodology for estimating the relative weights. We conclude with the application to one example.

## 2. Brief review of the methodology

### 2.1. Direct problem

We outline here the approach of Medina and Carrera (1996). Let us consider, without loss of generality, a 2D transient transport problem with steady state flow:

$$\nabla(\mathbf{T}\nabla h) + q = 0 \text{ in } \Omega \tag{1}$$

$$b\phi R\frac{\partial c}{\partial t} = \nabla(\mathbf{D}\nabla c) - \mathbf{q}\nabla c + q(c - c') \text{ in } \Omega \tag{2}$$

where $h$ is head, $\mathbf{T}$ is transmissivity tensor, $q$ is an areally distributed sink/source (recharge/extraction), $\phi$ is porosity, $b$ is thickness, $c$ is solute concentration, $\mathbf{D}$ is dispersion tensor, $\mathbf{q} = -\mathbf{T}\nabla h$ (notice that in 2D $\mathbf{q}$ is not Darcy flux, but flow rate per unit width), $c'$ is external solute concentration in areally distributed sinks or sources, $\Omega$ is the problem domain and $t$ is time. Eqs. (1) and (2) are solved with appropriate initial and boundary conditions. These equations are solved using a semidiscretization finite element method (Dautray and Lions, 1988), where spatial and temporal variabilities are considered separately. This leads to:

$$\boldsymbol{\psi}_0 = \mathbf{A}\mathbf{h}_0 - \mathbf{b}_0 = 0 \tag{3}$$

$$\boldsymbol{\psi}_j \equiv \left(\theta\mathbf{E} + \frac{\mathbf{F}}{\Delta t}\right)\mathbf{c}_j - \left[(\theta - 1)\mathbf{E} + \frac{\mathbf{F}}{\Delta t}\right]\mathbf{c}_{j-1} - \mathbf{g}_{j-1+\theta} = 0 \tag{4}$$

where $\mathbf{A}$ is the matrix of the discretized flow equation, $\mathbf{E}$ and $\mathbf{F}$ are the matrices of the discretized transport equation, $\mathbf{h}_0$ is the vector of steady state nodal heads, $\mathbf{c}$ is the vector of nodal concentrations, $\mathbf{b}$ and $\mathbf{g}$ are the right hand sides of the algebraic flow and transport equations after discretization, respectively, $\theta$ is the time weight ($\theta = 0$ is the explicit scheme, $\theta = 1$ is implicit and $\theta = 0.5$ is Crank–Nicholson scheme) and $j$ goes from 1 to $N_t$ (number of time steps).

### 2.2. Statistical formulation and objective function

The estimation problem is formulated as maximizing the likelihood function. To build this function, we assume that residuals (differences between measured and computed heads and/or concentrations) and prior estimation errors follow a normal distribution and that errors on heads, concentrations and model parameters are independent. The likelihood function is:

$$L = (2\pi)^{-n/2}\left(|\mathbf{C}_h||\mathbf{C}_c|\prod_j|\mathbf{C}_j|\right)^{-1/2}\exp\left[-\frac{1}{2}\left((\mathbf{h}-\mathbf{h}^*)^t \times\mathbf{C}_h^{-1}(\mathbf{h}-\mathbf{h}^*)+(\mathbf{c}-\mathbf{c}^*)^t\mathbf{C}_c^{-1}(\mathbf{c}-\mathbf{c}^*) +\sum_j(\mathbf{p}_j-\mathbf{p}_j^*)^t\mathbf{C}_j^{-1}(\mathbf{p}_j-\mathbf{p}_j^*)\right)\right] \tag{5}$$

where $\mathbf{C}_h$, $\mathbf{C}_c$ and $\mathbf{C}_j$, are the covariance matrices of errors in heads, concentrations and type $j$ parameters, respectively; $j$ represents type of parameter ($j=1$ for transmissivity, 2 for storativity, etc.); $\mathbf{h}^*$ is the vector of $n_h$ head measurements; $\mathbf{c}^*$ is the vector of $n_c$ concentration measurements; $\mathbf{p}_j^*$ is the vector of $n_j$ prior information data of the $j$-th type parameters; $\mathbf{h}$ is the vector of computed heads; $\mathbf{c}$ is the vector of computed concentrations and $\mathbf{p}_j$ is the vector of computed values of the $j$-th type parameters. Maximizing this function is equivalent to minimizing the support (minus two times the log-likelihood) function:

$$S = \tau_h^{-1}F_h + \tau_c^{-1}F_h + \sum_j\tau_j^{-1}F_j + n_h\ln(\tau_h) + n_c\ln(\tau_c) +\sum_j k_j\ln(\tau_j) + \ln|\mathbf{V}_h| + \ln|\mathbf{V}_c| +\sum_j\ln|\mathbf{V}_j| + n\ln(2\pi) \tag{6}$$

where $F_h=(\mathbf{h}-\mathbf{h}^*)^t\mathbf{V}_h^{-1}(\mathbf{h}-\mathbf{h}^*)$, $F_c=(\mathbf{c}-\mathbf{c}^*)^t\mathbf{V}_c^{-1}(\mathbf{c}-\mathbf{c}^*)$, $F_j=(\mathbf{p}_j-\mathbf{p}_j^*)^t\mathbf{V}_j^{-1}(\mathbf{p}_j-\mathbf{p}_j^*)$. The total number of unknown parameters is $n_p=\sum k_j$, where $k_j$ is the number of unknown parameters of type $j$. The total number of data (heads and concentrations plus prior information of unknown parameters) is $N=n_h+n_c+n_p$. Parameters $\tau_h$, $\tau_c$, and $\tau_j$ are unknown

and multiply matrices $V_h$, $V_c$ and $V_j$ to obtain the true but unknown covariance matrices ($\mathbf{C}_h = \tau_h \mathbf{V}_h$, $\mathbf{C}_c = \tau_c \mathbf{V}_c$ and $\mathbf{C}_j = \tau_h \mathbf{V}_j$, see e.g. Neuman and Yakowitz, 1979). Later we will generalize this formulation for the case of arbitrary number of types of variables, not just heads and concentrations.

Neglecting terms in Eq. (6) that depend only on statistical parameters, $\tau_h$, $\tau_c$, $\tau_j$ and possibly others (required to define $\mathbf{V}_h$, $\mathbf{V}_c$ and $\mathbf{V}_j$), and multiplying Eq. (6) by $\tau_h$, the objective function to be minimized becomes:

$$F = F_h + \lambda_c F_c + \sum_j \lambda_j F_j \tag{7}$$

where $\lambda_c$ and $\lambda_j$ are weights between the different pieces of information (heads, concentrations and prior information). Observe that weights in Eq. (6) are related to statistical parameters in Eq. (5) ($\lambda_c = \tau_h/\tau_c$, $\lambda_j = \tau_h/\tau_j$).

One of the problems related to multiple objective functions, such as Eq. (6) is the definition of weights, $\lambda$. These weights control the relative importance assigned to the different types of information (heads, concentrations, prior estimates, etc.). The actual relevance of those weights is not always analyzed, (see Neuman, 1973; Neuman and Yakowitz, 1979). It is evident, however, that giving too much importance to some data may bias the estimation. In fact, the main motivation behind the use of the likelihood function, in the first place, was the estimation of statistical parameters such as the $\lambda$'s. However, this is not always straight-forward. For one thing, no minimum exists for $S$ (6) as a function of $\lambda$ in many geostatistical inversion problems (Carrera, 1994). It is to overcome these difficulties that the expected likelihood function will be introduced in Section 4.

### 2.3. Minimization

Optimum parameters are the ones minimizing (7). A very robust estimation can be obtained using Marquardt's method. This method consists of updating iteratively the set of all parameters, $\mathbf{p}^m$, $m$ being iteration number, by means of:

$$(\mathbf{H}(p^m) + \mu_m \mathbf{I})(\mathbf{p}^{m+1} - \mathbf{p}^m) = -\nabla F(\mathbf{p}^m) \tag{8}$$

where $\mu_m$ is Marquardt's parameter, updated iteratively, $\mathbf{I}$ is the identity matrix, and $\nabla F$ and $\mathbf{H}$ are the gradient and first order approximation to the Hessian of $F$, respectively, given by:

$$\nabla F = 2\mathbf{J}_h^t \mathbf{V}_h^{-1}(\mathbf{h} - \mathbf{h}^*) + 2\lambda_c \mathbf{J}_c^t \mathbf{V}_c^{-1}(\mathbf{c} - \mathbf{c}^*) + 2\sum_j \lambda_j \mathbf{V}_j^{-1}(\mathbf{p}_j - \mathbf{p}_j^*) \tag{9}$$

$$\mathbf{H} = 2\mathbf{J}_h^t \mathbf{V}_h^{-1} \mathbf{J}_h + 2\lambda_c \mathbf{J}_c^t \mathbf{V}_c^{-1} \mathbf{J}_c + 2\sum_j \lambda_j \mathbf{V}_j^{-1} \tag{10}$$

where $\mathbf{J}_h$ and $\mathbf{J}_c$ are the Jacobian matrices, i.e. derivatives of heads and concentrations w.r.t. model parameters.

Once the Jacobian matrices $\mathbf{J}_h$ and $\mathbf{J}_c$ are computed, evaluating Eqs. (9) or (10) or solving Eq. (8) is neither computationally intensive nor conceptually difficult. The computational problem is associated with the evaluation of the Jacobian matrices. As mentioned in the introduction, this is usually achieved by means of the so called sensitivity equations. They are obtained by direct derivation of discretized state Eqs. (3) and (4), which leads to a set of $n_p$ systems of equations very similar to Eqs. (3) and (4). This becomes a very heavy burden for realistic geostatistical inversion problems, in which $n_p$, the number of estimated parameters can be equal to several hundreds. An alternative is discussed below.

## 3. Computation of Jacobian matrices using the adjoint state

The adjoint state method is a common technique to obtain derivatives of an objective function, depending on heads and/or concentrations, with respect to model parameters (Sun and Yeh, 1992; Townley and Wilson, 1985 among others). For this reason, only an outline is presented in this section for the sake of completeness, as well as some hints about its practical implementation following the guidelines of Carrera and Medina (1994).

Let us assume we wish to compute the derivatives w.r.t. $\mathbf{p}$ of a linear combination (with coefficients $\mathbf{a}$) of concentrations $\mathbf{c}$ at time $k$, that is, $f^k = \mathbf{a}^t \mathbf{c}^k$. This derivative is constrained by the fact that state variables ($\mathbf{h}$ and $\mathbf{c}$) and parameters must satisfy the state

equations (3) and (4). Therefore, a natural choice would be to build the Lagrangian:

$$\mathscr{L}^k = f^k + \sum_{j=0}^{N_t} \boldsymbol{\mu}_j^{k^t} \boldsymbol{\psi}_j \tag{11}$$

where $\boldsymbol{\mu}_j^k$ is the adjoint state associated with function $f^k$ and the $j$-th state equation. If we take total derivatives with respect to $\mathbf{p}$ in Eq. (11), we will have terms depending on $\partial \mathbf{c}_j/\partial \mathbf{p}$. As discussed in Section 2, evaluating them is extremely expensive. The adjoint states $\boldsymbol{\mu}_j^k$ are chosen precisely so as to eliminate these derivatives from the gradient of $f_k$. This can be achieved by, first, factoring $\partial \mathbf{c}_j/\partial \mathbf{p}$ out in the derivatives of Eq. (11) w.r.t. parameters, and, then, setting the factor to zero. This leads to the adjoint state equations:

$$\mathbf{a}^t \delta_j^k + \boldsymbol{\mu}_j^{k^t}\left(\theta \mathbf{E} + \frac{\mathbf{F}}{\Delta t}\right) - \boldsymbol{\mu}_{j+1}^{k^t}\left((\theta - 1)\mathbf{E} + \frac{\mathbf{F}}{\Delta t}\right) = 0 \tag{12}$$

where $\delta_j^k = 1$ when $j = k$ and zero otherwise. Eq. (11) is solved backwards in time, starting with $\boldsymbol{\mu}_N^k$ ($\boldsymbol{\mu}_{N+1}^k = 0$). Steady state adjoint vector (associated with steady state heads) is given by (Carrera and Medina, 1994):

$$\boldsymbol{\mu}_0^{k^t}\mathbf{A} + \mathbf{a}^t \delta_0^k + \sum_{j=1}^{k} \boldsymbol{\mu}_j^{k^t}\left(\frac{\partial \mathbf{E}}{\partial \mathbf{h}_0}\mathbf{c}_{j+\theta-1} + \frac{\partial \mathbf{F}}{\partial \mathbf{h}_0}\frac{\mathbf{c}_j - \mathbf{c}_{j-1}}{\Delta t} + \frac{\partial \mathbf{g}_{j+\theta-1}}{\partial \mathbf{h}_0}\right) = 0 \tag{13}$$

where $\mathbf{g}_{j+\theta-1} = \theta \mathbf{g}_j + (1-\theta)\mathbf{g}_{j-1}$ and $\mathbf{c}_{j+\theta-1} = \theta \mathbf{c}_j + (1-\theta)\mathbf{c}_{j-1}$. The computation of $\partial \mathbf{E}/\partial \mathbf{h}_0$ is quite expensive, but it should be noticed that it needs to be computed only once, because it depends neither on parameters nor on time. It depends on $\mathbf{h}_0$ through Darcy velocity and boundary flows, that are time independent because the flow equation is in steady state. For this reason, $\partial \mathbf{E}/\partial \mathbf{h}_0$ is computed once and stored. It should be noticed that the amount of storage is not as large as it might look (number of elements times nine for triangular linear elements).

Once the different adjoint states are computed, we obtain the gradient of $f^k$ from the derivative of $\mathscr{L}^k$, where now all terms containing derivatives of state variables have been eliminated. This leads to:

$$\frac{\mathrm{d}f^k}{\mathrm{d}\mathbf{p}} = \frac{\partial f^k}{\partial \mathbf{p}} + \sum_{j=0}^{N_t} \boldsymbol{\mu}_j^{k^t} \frac{\partial \boldsymbol{\psi}_j}{\partial \mathbf{p}} \tag{14}$$

It should be noticed that due to the linear dependence of $\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_n$ on model parameters, their derivatives with respect to model parameters are constant through points and time, which allows us to make some computations only once by using some storage (it is not usually a problem in current computers).

What we have described up to here is nothing but the standard adjoint state method. Yet, contrary to the standard adjoint state method, one does not need to compute a state per measurement because Eq. (12) for $\boldsymbol{\mu}_j^k$ does not depend separately on $k$ and $j$ (if it did, one would need $N_t$ adjoint states). As it is, the equation only depends on $k - j$, but not on $k$ or $j$ separately. Indeed, Eq. (12) is homogeneous for $k < j$, because $\boldsymbol{\mu}_{N+1}^k = 0$, for all $k$, and the forcing term $\mathbf{a}^t \delta_j^k$ is null for $k \neq j$. As a result, if $\mathbf{E}$ and $\mathbf{F}$ are time independent, $\boldsymbol{\mu}_{N_t}^{N_t}$ is equal to $\boldsymbol{\mu}_{N_t-1}^{N_t-1}$, $\boldsymbol{\mu}_{N_t-2}^{N_t-2}$, etc. Applying again Eq. (12), we would get that $\boldsymbol{\mu}_{N_t-1}^{N_t-1} = \boldsymbol{\mu}_{N_t-2}^{N_t-1} = \cdots = \boldsymbol{\mu}_1^2$ and so on. In short, for any $k$ and $j$,

$$\boldsymbol{\mu}_j^k = \boldsymbol{\mu}_{N-k+j}^N \tag{15}$$

This is very important, because it shows that one can use a single adjoint state per observation point to compute the derivatives of all concentrations at each observation point with respect to all parameters. That is, one needs to solve $N_t$ systems of equations, such as Eq. (12), per observation point. Actually, if flow parameters are uncertain, so that one needs to compute derivatives of heads and take into account the uncertainty of flow into transport, then one would also need to solve Eq. (13) at every observation time (at most, $N_t$ for the transport equations, $j \neq 0$, and one for the flow equation, $j = 0$). That is, at most, one needs to solve $2N_t + 1$ linear systems per observation point. This is a dramatic reduction with respect to the $(N_t + 2)(N_t + 1)/2$ linear systems required by the conventional formulation of the adjoint state. A full comparison is made in Table 1 for varying number of times, wells, and parameters. As one can see, the proposed alternative is always superior to the conventional adjoint state. The direct derivation method should only be preferred when the number of parameters is smaller than twice the number of time

Table 1
Computational cost of Jacobian matrices (measured in number of linear systems like Eq. (4) to be solved)

| | | | | | |
|---|---|---|---|---|---|
| | Number of times $N_t$ | 10 | 100 | 1000 | 100 |
| | Number of obs. points $N_w$ | 10 | 10 | 10 | 100 |
| | Number of parameters $N_p$ | 1000 | 1000 | 1000 | 100 |
| Direct derivation (sensitivity equations) | $N_t N_p$ | $10^4$ | $10^5$ | $10^6$ | $10^4$ |
| Adjoint state (conventional) | $N_w \frac{(N_t+2)(N_t+1)}{2}$ | $6.6\times10^2$ | $5\times10^4$ | $5\times10^6$ | $5\times10^5$ |
| Adjoint state (proposed) | $N_w(2N_t+1)$ | $2\times10^2$ | $2\times10^3$ | $2\times10^4$ | $2\times10^4$ |

steps. Actually, the adjoint state method suffers some overhead due to the fact that it has to be solved backwards in time, so the advantage is not as big as one might expect. Still, as we shall see later, it is very advantageous.

## 4. Finding the optimal weights

As discussed in the introduction, we seek the minimization of the expected likelihood, given by:

$$\bar{L}=\int(2\pi)^{-n/2}\left(|\tau_h\mathbf{V}_h||\tau_c\mathbf{V}_c|\prod_j|\tau_j\mathbf{V}_j|\right)^{-1/2} \times\exp\left[-\frac{1}{2}\left(\tau_h^{-1}F_h+\tau_c^{-1}F_c+\sum_j\tau_j^{-1}F_j\right)\right]\mathrm{d}\mathbf{p} \quad (16)$$

As this integral cannot be solved analytically, heads and concentrations are linearized around the estimate of $\mathbf{p}(\hat{\mathbf{p}})$, i.e.

$$\mathbf{h}(\mathbf{p})=\mathbf{h}(\hat{\mathbf{p}})+\mathbf{J}_h(\mathbf{p}-\hat{\mathbf{p}}) \quad \text{and} \quad \mathbf{c}(\mathbf{p})=\mathbf{c}(\hat{\mathbf{p}})+\mathbf{J}_c(\mathbf{p}-\hat{\mathbf{p}}) \quad (17)$$

where $\mathbf{J}_h$ and $\mathbf{J}_c$ are Jacobian matrices of heads and concentrations at $\mathbf{p}=\hat{\mathbf{p}}$, respectively, and $\hat{\mathbf{p}}$ are estimated parameters. Substituting Eq. (17) into Eq. (16), integrating, taking natural logarithm, multiplying by $-2$ and neglecting constant terms, one obtains:

$$\bar{S}=\sum_i\alpha_i^{-1}F_{u_i}+\sum_j\tau_j^{-1}F_j +\ln\left|\sum_i\alpha_i^{-1}\mathbf{J}_{u_i}^t\mathbf{V}_{u_i}^{-1}\mathbf{J}_{u_i}+\sum_j\tau_j^{-1}\mathbf{V}_j^{-1}\right| -\sum_i n_i\ln(\alpha_i^{-1})+\sum_j k_j\ln(\tau_j^{-1}) \quad (18)$$

we have generalized Eq. (18) for the case in which one has got an arbitrary number of types of measurements $u_i$ (i.e. $\alpha_1=\tau_h$, $\alpha_2=\tau_c$, $F_{u_1}=F_h$, $F_{u_2}=F_c$, in previous equations). In order to compute $\bar{S}$, one has to compute the statistical parameters $\alpha_i$, $\tau_j$. Observe that different combinations of $\alpha_i$, $\tau_j$ can be found for given weights ($\lambda$ in Eq. (7)). We look for statistical parameters ($\alpha_i$, $\tau_j$) that minimize Eq. (18). In Appendix A, we show that minimization of $\bar{S}$ with respect to the statistical parameters leads to:

$$\alpha_1=\frac{F_{u_1}}{n_1-tr(\mathbf{H}^{-1}\mathbf{J}_{u_1}^t\mathbf{V}_{u_1}^{-1}\mathbf{J}_{u_1})} \quad (19)$$

$$\alpha_i=\frac{F_{u_i}+\alpha_1 tr(\mathbf{H}^{-1}\mathbf{J}_{u_i}^t\mathbf{V}_{u_i}^{-1}\mathbf{J}_{u_i})}{n_i} \quad (20)$$

$$\tau_j=\frac{F_j+\alpha_1 tr(\mathbf{H}^{-1}\mathbf{V}_j^{-1})}{k_j} \quad (21)$$

where $\mathbf{H}$ is as given by Eq. (10). It is interesting to notice that these values coincide with those derived by Carrera and Neuman (1986a) for the case of only head data and neglecting the traces of the matrices appearing in Eqs. (19)–(21).

On occasions, the user sets the values of the relative weights ($\lambda_i$ and $\mu_j$) on the basis of his (her) own feeling. In such cases, the estimation of $\alpha_i$ and $\tau_j$ must be made by minimizing $\bar{S}$ constrained to $\lambda_i$ and $\mu_j$. We show in the appendix that this leads to:

$$\alpha_1=\frac{F}{\sum_i n_i+\sum_j k_j}=\frac{F}{N} \quad (22)$$

where $F$ is the total objective function (generalization of Eq. (7) with an arbitrary number of types of measurements and $N$ is the total number of data, including both measurements and prior estimation of

model parameters). Again, using this equation together with the restrictions, one obtains

$$\alpha_i = \frac{\alpha_1}{\lambda_i} \tag{23}$$

$$\tau_j = \frac{\alpha_1}{\mu_j} \tag{24}$$

It is interesting to notice that Eqs. (22)–(24) are identical to the equations derived by Carrera and Neuman (1986a), who used the likelihood function (but not its expected value). Inserting these values into Eq. (18) leads to

$$\bar{S}_1 = \sum_i \alpha_i^{-1} F_{u_i} + \sum_j \tau_j^{-1} F_j + \ln|\mathbf{H}| - \sum_i n_i \ln(\alpha_i^{-1}) - \sum_j k_j \ln(\alpha_1 \tau_j^{-1}) \tag{25}$$

$$\bar{S}_2 = N + \ln|\mathbf{H}| + N \ln(\alpha_1) - \sum_i n_i \ln(\lambda_i) - \sum_j k_j \ln(\mu_j) \tag{26}$$

The robustness of these two functions is studied below.

## 5. Synthetic example

### 5.1. Methodology

We generate four examples (two flow and two transport) on a single domain. In essence, the examples consist of generating data from a synthetic problem and then use this data to test the effectiveness of the approaches described above. Specifically, our work consists of the following steps:

1. Define the problem domain.
2. Generate the 'true' transmissivity fields.
3. Generate measurements.
4. Solve the different problems.
   steady state flow
   transient flow
   transient transport,

Following is a brief description of each step and the results.

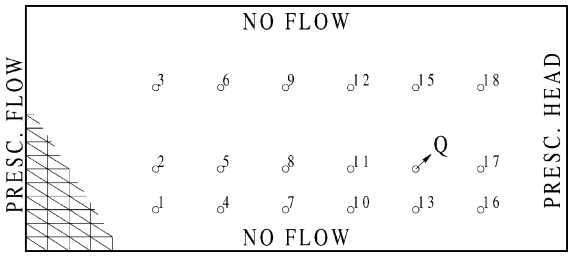


Fig. 1. Schematic description of the flow domain. Water flows from left to right. The first flow problem (Example 1) consists of the steady-state conditions. A constant solute in flux is simulated with this flow field to obtain the solute invasion transport problem (Example 3). The steady state flow plus the response to pumping defines the transient flow problem (Example 2). Finally, the Example 4 consists of simulating four tracer tests, injecting at points 10, 12, 16 and 18 and steady-state pumping at point 14.

### 5.2. Problem domain

The base example is a rectangular domain with an extension of $200 \times 90\ m^2$, containing 18 control points (Fig. 1). No flow is imposed at the top and bottom boundaries. A flow rate of $9\ m^3$/day is prescribed at the left boundary (uniformly distributed) and a head of 0 m is prescribed at the right side. Observation point 14 is an extraction well with a rate of $10\ m^3$/day. All flow parameters, but transmissivity, are assumed known. Transmissivity is treated as a lognormal random field. Two transport problems are simulated. The first one consists of the invasion of a solute from the left boundary; this is simulated by assigning a concentration of 100 ppm to the incoming water. The second example consists of the simulation of four tracer tests.

### 5.3. Generation of 'true' transmissivity field

Generation of the transmissivity field consists of three steps.

1. A log-transmissivity field is generated with an exponential variogram with a sill of 5 and a range of 50 m. We use sequential simulation with the GCOSIM code (Gómez-Hernández, 1991). This simulation was made with a resolution of $1 \times 1\ m^2$.
2. Six transmissivity values are taken from the above field at observation points (3, 4, 8, 9, 13, 15). We use them as point data to obtain kriging estimates on a grid of $25 \times 18$ regular blocks. In this way we

obtain a kriged field and their covariance matrix. One transmissivity value will be estimated at each one of these blocks.

3. Four random fields are generated such that their means and covariance matrices are equal to the ones of the kriged field. Those simulations will be our 'true' fields (see Fig. 2), and the computed field will be compared with it.

### 5.4. Generation of 'measurements'

Heads and concentrations are computed at every grid node of Fig. 1 by performing a flow and solute transport simulation. These are assumed to be the 'real' heads and concentrations of the system. Measurements are obtained by adding a white noise to these 'real' values. The variance of this noise is included in the $\mathbf{V}_h$ and $\mathbf{V}_c$ matrices. As a result, $\alpha_h$ and $\alpha_c$ should be equal to 1.

### 5.5. Results for Example1: steady state flow

Fig. 3 contains the transmissivity fields calibrated against steady state heads at the 17 points (pumping well was excluded). The 'true' transmissivity field is also included for comparison. As one may see, the fields are a simplified version of reality (independently of the $\lambda_T$ weight), although a little better than the kriging field (prior information for the calibration process). It should be stated that we have found some convergence difficulties for very small $\lambda_T$ (0.001) due to the large number of estimated parameters (450 $T$ values). Regarding the estimation of $\lambda_T$, we know exactly the statistical variables (that is, the true covariance matrices), therefore the optimal value of $\lambda_T$ should be 1. Table 2 contains the values of $\lambda_T$ for which the estimation statistics reach their optimum value. For instance, the minimum value of SSE is attained at $\lambda_T = 5$ in simulation #2. The displayed statistics are as follows:

SSE—Sum of square errors of estimated log-$T$'s. It is the basic raw criterion to evaluate the goodness of the estimated field.

$F_h$—Head fit criterion. See Eqs. (6) and (7). It is not a good estimation criterion, because it should grow monotically with $\lambda$ (reducing $\lambda$ unconstrains the problem).

$S_2$—(Eq. (3.11) of Carrera, 1994). Support function (6) evaluated after unconstrained minimization with respect to $\tau_h$ and $\tau_i$.

$S_3$—(Eq. (3.12) of Carrera, 1994). Support function (6) evaluated after minimization with respect to $\tau_h$ and $\tau_i$, but constrained by $\lambda_T = \tau_h/\tau_T$.

$\bar{S}_1$—(Eq. (25)). Expected support evaluated after unconstrained minimization of (18) with respect to $\tau_h$ and $\tau_i$.

$\bar{S}_2$—(Eq. (26)). Expected support evaluated after minimization of Eq. (18) with respect to $\tau_h$ and $\tau_i$, but constrained by $\lambda_T = \tau_h/\tau_T$.

In theory, the minimum value of $\bar{S}$ should be attained at $\lambda_T = 1$, and this happens in two of the four cases.

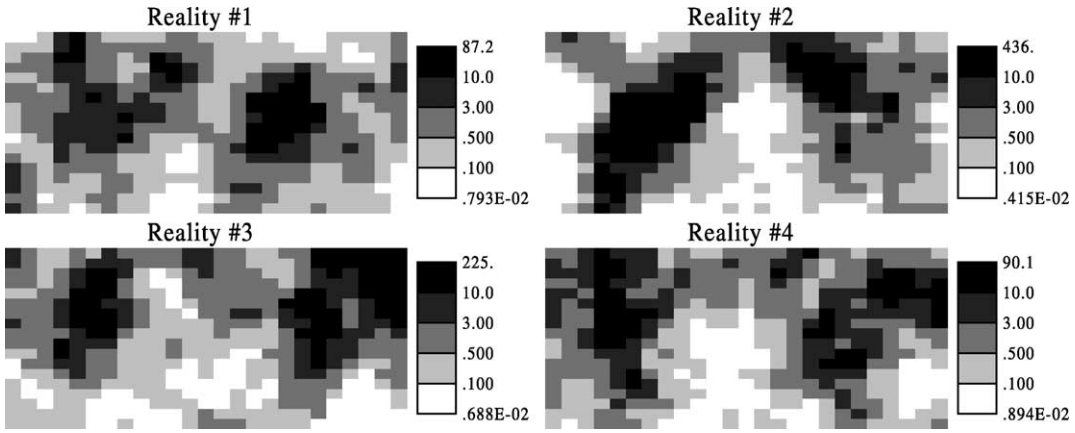


Fig. 2. Four conditional simulations used as true fields.

### 5.6. Results for Example 2: transient flow

Initial conditions are obtained as a steady state without pumping and transient state is generated by pumping at well #14. Measurements are taken at the 17 observation points in 9 times (0.25, 0.5, 0.75, 1.0, 2.0, 3.0, 5.0, 10.0 and 20.0 days), that is, 153 measurements are used. The estimated fields obtained with $\lambda_T = 0.001$ and 1.0 for true field #2 are displayed in Fig. 3. As a general statement, the new fields are

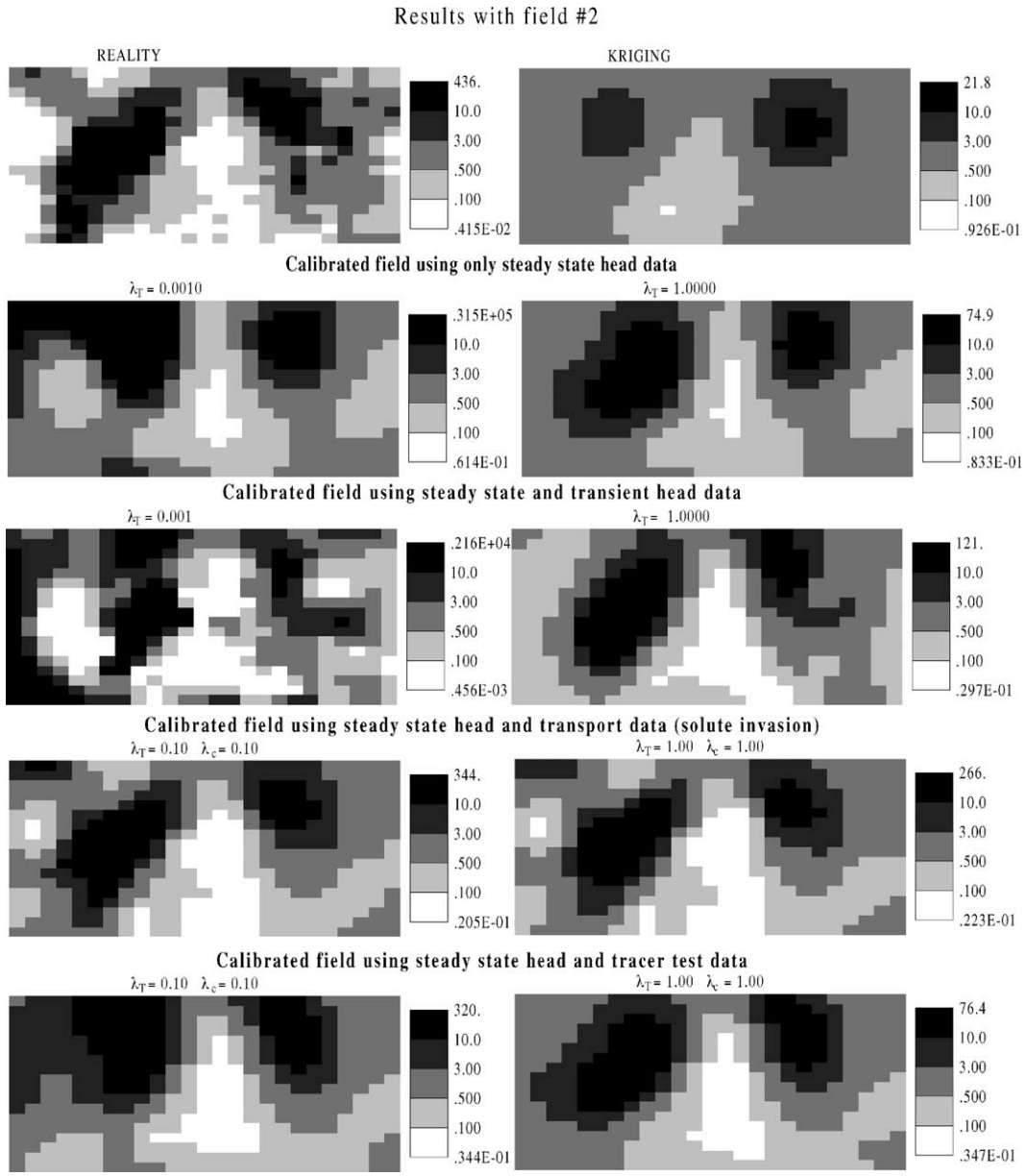


Fig. 3. Transmissivity fields obtained with "reality" #2 (recall Fig. 2) for the four problems. Notice also that the essential features of the true field are captured by all problems when the optimum weight ($\lambda_T = 1, \lambda_c = 1$) is used. However, these features are smeared when $\lambda_T$ grows (the kriging field is the limit when $\lambda_T$ tends to infinity for all cases). Spurious fluctuations on transmissivities are computed on the left side in all cases when the weights are too small.

Table 2
Values of $\lambda_T$ for which the shown statistics are minima (tested $\lambda_T$ values are 0.001, 0.01, 0.03, 0.1, 0.3, 1.0, 5.0, 20.0) in Example 1 (steady-state flow problem). Notice that $\bar{S}_2$ identifies $\lambda_T = 1$ as optimum in two of the cases, although estimation errors happen to be smaller for larger $\lambda_T$'s in three cases

| | Simulation 1 $\lambda_T$ | Simulation 2 $\lambda_T$ | Simulation 3 $\lambda_T$ | Simulation 4 $\lambda_T$ |
|---|---|---|---|---|
| SSE | 1 | 5 | 20 | 5 |
| $F_h$ | 0.001 | 0.001 | 0.001 | 0.001 |
| $S_2$ | 20 | 20 | 0.001 | 20 |
| $S_3$ | 20 | 20 | 0.001 | 20 |
| $\bar{S}_1$ | 0.01 | 1 | 0.001 | 0.001 |
| $\bar{S}_2$ | 1 | 1 | 0.001 | 0.001 |

closer to reality than those obtained using steady state flow data only. It should be noticed that the kriged field is much smoother than the reality or the computed field with head information. However, the objective function is only about double for the kriged field than for the estimated one. Transmissivities estimated with $\lambda_T = 0.001$ are somewhat unstable (see the maximum and minimum values in the gray scale for each field).

The evolution of $\bar{S}_2$, $F_h$, $S_1$, and SSE versus $\lambda_T$ are shown in Fig. 4. One can see that $\bar{S}_2$ correctly identifies 1.0 as the optimum value of $\lambda_T$. In this case, estimation errors (as measured by SSE) are also minimum (or nearly so) for the optimal $\lambda_T$.

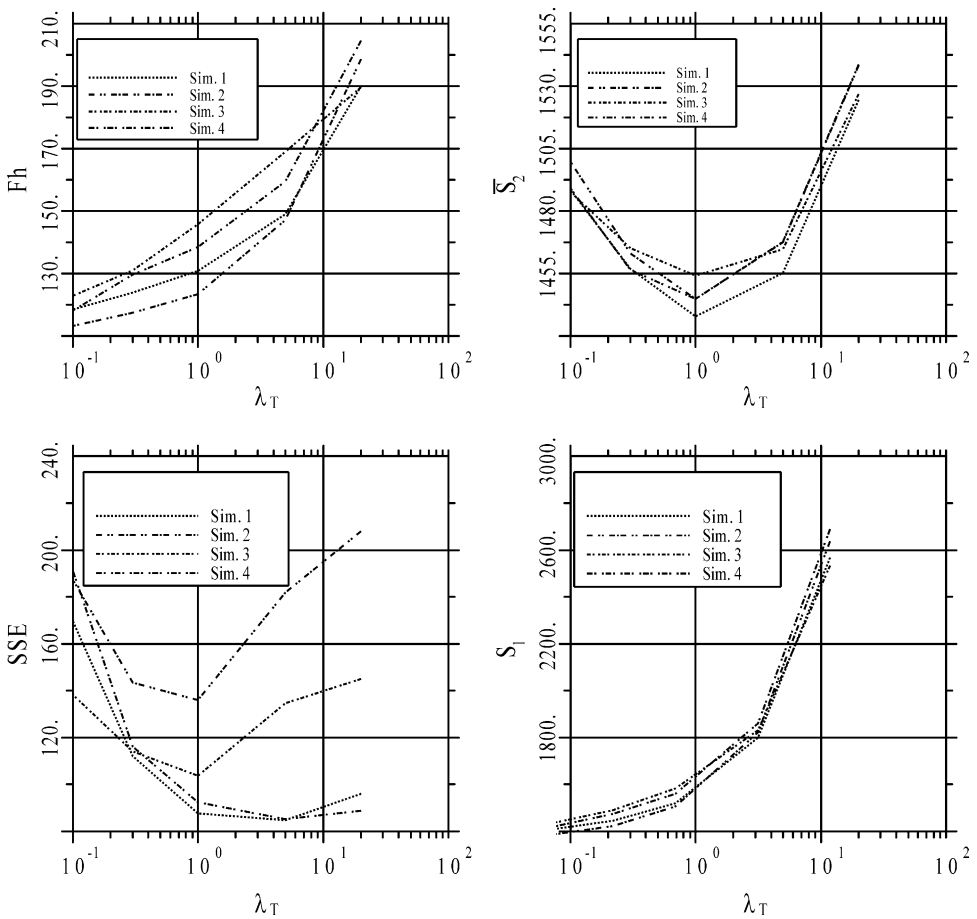


Fig. 4. Dependence of $S_1$, $\bar{S}_2$, SSE and $F_h$ on $\lambda_T$ for the second flow problem. Several things should be noticed. First, the second expected likelihood (Eq. (26)) consistently identifies the optimum $\lambda_T$ ($\bar{S}_2$). The first expected likelihood (Eq. (25)), ($\bar{S}_1$), underestimates $\lambda_T$. $F_h$ cannot be used to estimate $\lambda_T$ because it grows monotonically with $\lambda_T$. Finally, the smallest estimation errors (SSE) tend to be obtained with the optimal weights.

Table 3 displays the values of $\lambda_T$ that lead to the minimum values of the estimation statistics (*SSE*, $F_h$, $S_2$, $S_3$, $\bar{S}_1$, and $\bar{S}_2$). In this case, the optimum $\lambda_T$ is found in the four realizations. The remaining statistics tend to miss the mark. The use of the proposed adjoint state method was very advantageous, a ratio of 10.6 was obtained (CPU time using direct derivation divided by CPU time using adjoint state method).

### 5.7. Results for Example 3: solute invasion transport problem

Two transport problems are considered. The first transport problem is created by introducing a mass flux at the left boundary in Fig. 1. This is the one we have termed invasion problem. Concentrations are sampled at the 18 observation points. In addition, steady state heads are available at these points. Concentrations are sampled 10 times at the 18 observation points, so that we have a total of 180 concentration data and 17 head data. The transmissivity field is shown in Fig. 3 for two pairs of $\lambda_T$ and $\lambda_c$. Again, the fields estimated with the optimal weights are a smooth version of the true field, but capture the overall trends of variability. On the other hand, these trends are missed by the kriged field and by the field with too small $\lambda_T$ and $\lambda_c$, for which some spurious variability trends show up.

The values of $\lambda_T$ and $\lambda_c$ yielding the minimum values of several estimation statistics are shown in Table 4. It can be noticed that $\bar{S}_2$ identifies the theoretical optimum values, $\lambda_T = \lambda_c = 1$. On the other hand, the minimum value of SSE is not always reached at this theoretical minimum. In this example, a ratio of 5.4 was obtained (CPU time using direct derivation divided by CPU time using adjoint state method).

Table 3
Values of $\lambda_T$ for which the shown statistics are minima (tested $\lambda_T$ values are 0.001, 0.01, 0.03, 0.1, 0.3, 1, 5, 20) in Example 2 (transient flow problem). Notice that $\bar{S}_2$ identifies the optimum weights in all simulations. In this case, such weights lead to optimal (or nearly so) estimation errors

| | Simulation 1 $\lambda_T$ | Simulation 2 $\lambda_T$ | Simulation 3 $\lambda_T$ | Simulation 4 $\lambda_T$ |
|---|---|---|---|---|
| SSE | 5 | 1 | 1 | 5 |
| $F_h$ | 0.001 | 0.001 | 0.001 | 0.001 |
| $S_2$ | 20 | 20 | 20 | 20 |
| $S_3$ | 20 | 20 | 20 | 20 |
| $\bar{S}_1$ | 0.3 | 0.3 | 0.3 | 0.3 |
| $\bar{S}_2$ | 1 | 1 | 1 | 1 |

Table 4
Values of $\lambda_T$ and $\lambda_c$ leading to the minima of the shown statistics for Example 3 (solute invasion transport problem, tested pairs are: $\lambda_T = 0.1$, 1, 5 and 20 and $\lambda_c = 0.1$, 0.5 and 1.0). Notice that $\bar{S}_2$ consistently identifies the optimum $\lambda$'s in all but one case. These $\lambda$'s also tend to yield optimal (or nearly so) parameters

| | Simulation 1 | | Simulation 2 | | Simulation 3 | | Simulation 4 | |
|---|---|---|---|---|---|---|---|---|
| | $\lambda_T$ | $\lambda_c$ | $\lambda_T$ | $\lambda_c$ | $\lambda_T$ | $\lambda_c$ | $\lambda_T$ | $\lambda_c$ |
| SSE | 5 | 0.1 | 1 | 1 | 1 | 0.5 | 5 | 1 |
| $F_h$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| $F_c$ | 0.1 | 1 | 0.1 | 1 | 0.1 | 1 | 0.1 | 1 |
| $S_2$ | 20 | 0.5 | 5 | 0.5 | 20 | 0.5 | 20 | 0.5 |
| $S_3$ | 20 | 0.5 | 20 | 0.5 | 20 | 0.5 | 20 | 0.5 |
| $\bar{S}_1$ | 0.1 | 0.1 | 1 | 0.5 | 0.1 | 0.1 | 0.1 | 0.1 |
| $\bar{S}_2$ | 1 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 |

### 5.8. Results for Example 4: four tracer tests transport problem

The second transport problem consists of the analysis of four tracer tests injecting at 4 different locations (boreholes 10, 12, 16 and 18) and pumping at point 14. The steady state heads are also used as data. Total number of data is 53 (17 steady state heads

Table 5
Values of $\lambda_T$ and $\lambda_c$ leading to the minima of the shown statistics for Example 4 (tracer test transport problem, tested pairs are: $\lambda_T = 0.1$, 1, 5 and 20 and $\lambda_c = 0.1$, 0.5 and 1.0). Notice that $\bar{S}_2$ consistently identifies the optimum $\lambda$'s in all but one case. These $\lambda$'s also tend to yield optimal (or nearly so) parameters

| | Simulation 1 | | Simulation 2 | | Simulation 3 | | Simulation 4 | |
|---|---|---|---|---|---|---|---|---|
| | $\lambda_T$ | $\lambda_c$ | $\lambda_T$ | $\lambda_c$ | $\lambda_T$ | $\lambda_c$ | $\lambda_T$ | $\lambda_c$ |
| SSE | 1 | 0.1 | 5 | 1 | 5 | 1 | 1 | 1 |
| $F_h$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| $F_c$ | 0.1 | 1 | 0.1 | 1 | 0.1 | 1 | 0.1 | 1 |
| $S_2$ | 20 | 0.5 | 20 | 0.5 | 20 | 0.5 | 20 | 0.5 |
| $S_3$ | 20 | 0.5 | 20 | 0.5 | 20 | 0.1 | 20 | 0.1 |
| $\bar{S}_1$ | 0.1 | 0.1 | 1 | 1 | 0.1 | 1 | 1 | 1 |
| $\bar{S}_2$ | 1 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 |

and 36 concentration measurements, 9 for each test). The estimated transmissivity field is also shown in Fig. 3. Values of $\lambda_T$ and $\lambda_c$ yielding the minimum values of several estimation statistics are shown in Table 5. The analysis is similar to that of the previous example. It is interesting to point that it is also valid for a rather different kind of problem.

## 6. Concluding remarks

We have presented two methodological contributions to the solution of geo-statistical inverse problems. First, we have revised the formulation of the adjoint state method for computing Jacobian matrices. Second, we have used the expected value of the likelihood function to derive the relative weights of different types of information.

The proposed form of the adjoint state equation had been introduced by Carrera and Medina (1994), but had gone unnoticed, probably because we had not tested it. We have presented a set of tests illustrating that it can reduce computation cost by an order of magnitude in realistic problems. The main advantage of the method is that it yields the Jacobian matrix at a cost that is independent of the number of parameters. Therefore, the relative savings of the method grow with this number.

The expected value of the likelihood function had been reported to be more robust than the likelihood function itself for the purpose of estimating statistical parameters. We have used it to derive the weights to be assigned to transmissivity and concentration data, relative to head data. The results are indeed robust in the sense that the theoretical optimum was correctly identified in most cases. In fact, only the steady state flow case showed poor results (the optimal weights were identified in only two out of four simulations). It should be noticed that in this case, the optimal weights did not yield very good estimations of tranmissivity (as measured by the SSE). In all other examples, optimal weights were consistently identified and they led to optimal, or nearly so, transmissivities.

## Acknowledgements

This project was funded by ENRESA (Spanish Agency for Nuclear Waste Disposal), CICYT and CIRIT (Spanish and Catalonian, respectively, Research Funding Agencies).

## Appendix A

We need to minimize $\bar{S}$ (18) w.r.t. $\alpha_i$ and $\tau_j$. We will do it under two different assumptions. First, assuming that there are no constraints. Second, assuming that relative weights $\lambda_i = \alpha_1 \alpha_i^{-1}$ and $\mu_j = \alpha_1 \tau_j^{-1}$ are known.

### A.1. Unconstrained minimization

Differentiating Eq. (18) with respect to $\alpha_i^{-1}$ and $\tau_j^{-1}$ leads to

$$\frac{\partial \bar{S}}{\partial \alpha_i^{-1}} = F_{u_i} + \alpha_1 tr(\mathbf{H}^{-1} \mathbf{J}_{u_i}^t \mathbf{V}_{u_i}^{-1} \mathbf{J}_{u_i}) - \frac{n_i}{\alpha_i^{-1}} \tag{A.1}$$

$$\frac{\partial \bar{S}}{\partial \tau_j^{-1}} = F_j + \alpha_1 tr(\mathbf{H}^{-1} \mathbf{V}_j^{-1}) - \frac{k_j}{\tau_j^{-1}} \tag{A.2}$$

where $\mathbf{H}$ is defined by Eq. (10).

Setting Eqs. (A.1) and (A.2) to zero leads to Eqs. (19)–(21) for $\alpha_1$, $\alpha_i$ and $\tau_j$.

### A.2. Constrained minimization

In order to impose the constraints $\lambda_i = \alpha_1 \alpha_i^{-1}$ and $\mu_j = \alpha_1 \tau_j^{-1}$ we build the Lagrange function:

$$\mathscr{L} = \bar{S} + \sum_i \nu_i (\lambda_i \alpha_1^{-1} - \alpha_i^{-1}) + \sum_j \rho_j (\mu_j \alpha_1^{-1} - \tau_j^{-1}) \tag{A.3}$$

We now take derivatives w.r.t. $\alpha_1^{-1}$, $\alpha_i^{-1}$ and $\tau_j^{-1}$

$$\frac{\partial \mathscr{L}}{\partial \alpha_1^{-1}} = \frac{\partial \bar{S}}{\partial \alpha_1^{-1}} + \sum_{i \geq 2} \nu_i \lambda_i + \sum_j \rho_j \mu_j \tag{A.4}$$

$$\frac{\partial \mathscr{L}}{\partial \alpha_i^{-1}} = \frac{\partial \bar{S}}{\partial \alpha_i^{-1}} - \nu_i \qquad i \geq 2 \tag{A.5}$$

$$\frac{\partial \mathscr{L}}{\partial \tau_j^{-1}} = \frac{\partial \bar{S}}{\partial \tau_j^{-1}} - \rho_j \qquad \forall j \tag{A.6}$$

Finally, we set these equations to zero. Multiplying Eqs. (A.5) by $\lambda_i$ and Eq. (A.6) by $\mu_j$ and adding all the equations together, allows us to get $\alpha_1 : \alpha_i$ and $\tau_j$ are

derived directly from the constraints. This leads to Eqs. (22)–(24).

### A.3. Derivatives of $\bar{S}$

$\bar{S}$ is a function of weights ($\lambda_i$ and $\mu_j$), not only through the explicit dependencies of Eq. (18) but also through the parameters minimizing Eq. (7). In this section, we would like to find this dependence. More precisely, let us find the derivative of $\bar{S}$ with respect to $\lambda_i$ and $\mu_j$. We will find it only for $\bar{S}_2$, which seems to be slightly more robust than $\bar{S}_1$ in finding the optimal weights, as seen in the examples section. Differentiating Eq. (26) with respect to $\lambda_i$ leads to:

$$\frac{\partial \bar{S}}{\partial \lambda_i} = \frac{\partial \ln|\mathbf{H}|}{\partial \lambda_i} + \frac{N}{\alpha_1}\frac{\partial \alpha_1}{\partial \lambda_i} - \frac{n_i}{\lambda_i} \tag{A.7}$$

Neglecting dependence of the jacobians on $\lambda$ (first order approximation):

$$\frac{\partial \ln|\mathbf{H}|}{\partial \lambda_i} \simeq tr(\mathbf{H}^{-1}\mathbf{J}_{u_i}^t\mathbf{V}_{u_i}^{-1}\mathbf{J}_{u_i}) \tag{A.8}$$

Assuming that we have found the optimum, $\nabla F(\mathbf{p}) = 0$, so

$$\frac{\partial F}{\partial \lambda_i} = \nabla F(\mathbf{p})\frac{\partial p}{\partial \lambda_i} + F_{u_i} = F_{u_i} \tag{A.9}$$

Substituting Eqs. (A.8) and (A.9) into Eq. (A.7), we obtain

$$\frac{\partial \bar{S}}{\partial \lambda_i} = tr(\mathbf{H}^{-1}\mathbf{J}_{u_i}^t\mathbf{V}_{u_i}^{-1}\mathbf{J}_{u_i}) + \frac{F_{u_i}}{\alpha_1} - \frac{n_i}{\lambda_i} \tag{A.10}$$

$(i > 1)$

In a similar way,

$$\frac{\partial \bar{S}}{\partial \mu_j} = tr(\mathbf{H}^{-1}\mathbf{V}_j^{-1}) + \frac{F_j}{\alpha_1} - \frac{k_j}{\mu_j} \tag{A.11}$$

Notice the formal similarity between Eqs. (A.10) and (A.11). In fact, we can write them in a unique expression. Let $w_i$ be one of the weights ($\lambda_i$ or $\mu_i$). Observe that both equations may be written as

$$\frac{\partial \bar{S}}{\partial w_i} = tr(\mathbf{H}^{-1}\mathbf{A}_i) + \frac{F_i}{\alpha_1} - \frac{m_i}{w_i} \tag{A.12}$$

where the meaning of $\mathbf{A}_i$, $F_i$ and $m_i$ depends on whether $w_i$ is $\lambda_i$ or $\mu_i$.

### A.4. Second order derivatives

Second order derivatives are obtained by differentiating Eq. (18) with respect to the weights. The starting point here is the assumption that the estimate of parameters is a minimum of Eq. (7), i.e. $\nabla F(\mathbf{p}) = 0$. By differentiation of this expression with respect to $\lambda_i$ and $\mu_j$, we get:

$$\mathbf{H}\frac{\partial \mathbf{p}}{\partial \lambda_i} = -\nabla F_{u_i} \tag{A.13}$$

where we have implicitly approximated the hessian of $F$ by its own first order approximation (10). Neglecting again the dependence of jacobians on the parameters, we get:

$$\frac{\partial F_i}{\partial w_j} = \nabla F_i^t\frac{\partial \mathbf{p}}{\partial w_j} = -\nabla F_i^t\mathbf{H}^{-1}\nabla F_j \tag{A.14}$$

$$\frac{\partial tr(\mathbf{H}^{-1}\mathbf{A}_i)}{\partial w_j} \simeq tr(\mathbf{H}^{-1}\mathbf{A}_j\mathbf{H}^{-1}\mathbf{A}_i) \tag{A.15}$$

Finally, we differentiate Eq. (A.12) w.r.t. $w_j$, using the previous equations. This leads to:

$$\frac{\partial^2 \bar{S}}{\partial w_j \partial w_i} = tr(\mathbf{H}^{-1}\mathbf{A}_j\mathbf{H}^{-1}\mathbf{A}_i) - \frac{\nabla F_i^t\mathbf{H}^{-1}\nabla F_j}{\alpha_1} - \frac{F_iF_j}{N\alpha_1^2} + \frac{m_i}{w_i^2}\delta_{ij} \tag{A.16}$$

## References


Akaike, H., 1978. A new look at the Bayes procedure. Biometrika 65 (1), 53–59.

Capilla, J.E., Gómez-Hernández, J., Sahuquillo, A., 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data 2. Demonstration on a synthetic aquifer. Journal of hydrology 203 (1–4), 175–188.

Capilla, J.E., Rodrigo, J., Gómez-Hernández, J.J., 1999. Simulation of non-Gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information. Mathematical Geology 31 (7), 907–927.

Carrera, J., 1987. State of the art of inverse problem applied to the flow and solute transport equations. In: Custodio, E., Gurguí, A., Lobo Ferreira, J.P. (Eds.), Groundwater Flow and Quality Modeling, Dordrecht-Reidel, pp. 273–283.

Carrera, J., 1994. INVERT-4. A Fortran program for solving the groundwater flow inverse problem. User's guide. CIMNE. Technical Report, 160pp. + appendices, Barcelona, Spain.

Carrera, J., Glorioso, L., 1991. On geostatistical formulations of the groundwater flow inverse problem. Advances in Water Resources 14 (5), 273–283.

Carrera, J., Medina, A., 1994. An improved form of adjoint-state equations for transient problems. In: Peters, A., Wittun G., Herrling, B., Meissner, U., Brebbia, C.A., Gray, W.G., Pinder, G.F. (Eds.), X International Conference on Computational Methods in Water Resources, Kluwer Academic.

Carrera, J., Neuman, S.P., 1986a. Estimation of aquifer parameters under steady-state and transient conditions I. Background and statistical framework. Water Resources Research 22 (2), 199–210.

Carrera, J., Neuman, S.P., 1986b. Estimation of aquifer parameters under steady-state and transient conditions II. Uniqueness, stability and solution algorithms. Water Resources Research 22 (2), 211–227.

Clifton, P.M., Neuman, S.P., 1982. Effects of kriging and inverse modeling on conditional simulation of the Avra Valley aquifer in southern Arizona. Water Resources Research 18 (4), 1215–1234.

Cooley, R.L., 1982. Incorporation of prior information on parameters into non-linear regression groundwater-flow models. 1. Theory. Water Resources Research 18 (4), 965–976.

Cooley, R.L., 1985. A comparison of several methods of solving nonlinear-regression groundwater-flow problems. Water Resources Research 21 (10), 1525–1538.

Dagan, G., 1985. Stochastic modeling of groundwater flow by unconditional and conditional probabilities: the inverse problem. Water Resources Research 21 (1), 65–72.

Dautray, R., Lions, J.L., 1988. Analyse mathematique et calcul numérique pour les sciences et les techniques. Ed. Masson.

Gómez-Hernández, J.J., 1991. A Stochastic approach to the simulation of block conductivity fields conditioned upon data measured at a smaller scale. PhD Dissertation, Department of Applied Sciences, Stanford University, California, USA.

Gómez-Hernández, J.J., Sahuquillo, A., Capilla, J., 1997. Stochastic simulation of transmissivity fields conditional to both transmisivity and piezometric data. I Theory. Journal of Hydrology 203 (1–4), 162–174.

Hill, M.C., Cooley, R.L., Pollock, D.W., 1998. A controlled experiment in ground water flow model calibration. Ground Water 36 (3), 520–535.

Hoeksema, R.J., Kitanidis, P.K., 1984. An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. Water Resources Research 20 (7), 1003–1020.

Hoeksema, R.J., Kitanidis, P.K., 1985. Comparison of Gaussian conditional mean and kriging estimation in the geostatistical solution of the inverse problem. Water Resources Research 21 (6), 825–836.

Honjo, Y., Wen-Tsung, L., Sakajo, S., 1994. Aplication of Akaike information criterion statistics to geotechnical inverse analysis: the extended Bayesian method. Structural Safety 14, 5–29.

Kitanidis, P.K., 1995. Quasi-linear geostatistical theory for inversing. Water Resources Research 31 (10), 2411–2419.

Kitanidis, P.K., Vomvoris, E.G., 1983. A geostatistical approach to the inverse problem in groundwater modeling (steady-state) and one dimensional simulations. Water Resources Research 19 (3), 677–690.

McLaughlin, D., Townley, L.R., 1996. A reassesment of the groundwater inverse problem. Water Resources Research 32 (5), 1131–1161.

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

Neuman, S.P., 1973. Calibration of distributed parameter groundwater flow model viewed as a multiple objective decision process under uncertainty. Water Resources Research 9 (4), 1006–1021.

Neuman, S.P., Yakowitz, S., 1979. A statistical approach to the inverse problems of aquifer hydrology. 1. Theory. Water Resources Research 15 (4), 845–860.

Rubin, Y., Dagan, G., 1987. Stochastic identification of transmissivity and effective recharge in steady groundwater flow: I. Theory. Water Resources Research 23 (7), 1185–1192.

Sadeghipur, J., Yeh, W.W.-G., 1984. Parameter identification of groundwater aquifer models. A generalized least squares approach. Water Resources Research 20 (7), 971–979.

Sahuquillo, A., Capilla, J., Gómez-Hernández, J.J., Andreu, J., 1992. Conditional simulation of transmissivity fields honouring piezometric data. In: Blain, W.R., Cabrera, E. (Eds.), Fluid Flow Modeling, Comput Mech, Billerica, MA, pp. 201–212.

Sun, N.Z., Yeh, W.W.G., 1992. A stochastic inverse solution for transient groundwater flow: parameter identification and reliability analysis. Water Resources Research 28 (12), 3269–3280.

Townley, L.R., Wilson, J.L., 1985. Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models. Water Resources Research 21 (12), 1851–1860.

Yeh, W.W.G., 1986. Review of parameter identification procedures in groundwater hydrology. The inverse problem. Water Resources Research 22 (2), 95–108.

Yeh, T.C.J., Jin, M., Hanna, S., 1996. An iterative stochastic inverse method: conditional effective transmissivity and hydraulic head fields. Water Resources Research 32 (1), 85–92.

Zimmerman, D.A., de Marsily, G., Gotway, C.A., Marrietta, M.G., Axness, C.L., Beauheim, R.L., Bras, R.L., Carrera, J., Dagan, G., Davies, P.B., Gallegos, D.P., Galli, A., Gómez-Hornández, J., Grindrod, P., Gutjahr, A.L., Kitanidis, P.K., Lavenue, A.M., Mc Laughlin, D., Neuman, S.P., Ramarao, B.S., Ravenne, C., Rubin, Y., 1998. Comparison of seven geostatistically-based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resources Research 34 (6), 1373–1413