Sensitivity of temporal moments calculated by the adjoint-state method and joint inversing of head and tracer data

Olaf A. Cirpka *, Peter K. Kitanidis

Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020, USA

Received 27 August 1999; received in revised form 7 January 2000; accepted 19 January 2000

Abstract

Including tracer data into geostatistically based methods of inverse modeling is computationally very costly when all concentration measurements are used and the sensitivities of many observations are calculated by the direct differentiation approach. Harvey and Gorelick (Water Resour Res 1995;31(7):1615–26) have suggested the use of the first temporal moment instead of the complete concentration record at a point. We derive a computationally efficient adjoint-state method for the sensitivities of the temporal moments that require the solution of the steady-state flow equation and two steady-state transport equations for the forward problem and the same number of equations for each first-moment measurement. The efficiency of the method makes it feasible to evaluate the sensitivity matrix many times in large domains. We incorporate our approach for the calculation of sensitivities in the quasi-linear geostatistical method of inversing (“iterative cokriging”). The application to an artificial example of a tracer introduced into an injection well shows good convergence behavior when both head and first-moment data are used for inversing, whereas inversing of arrival times alone is less stable. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Inverse modeling; Sensitivity; Adjoint state; Temporal moments

1. Introduction

In the inverse problem of groundwater hydrology, we infer the hydraulic conductivity distribution within a particular aquifer from measured hydraulic heads and other information, such as tracer data collected at the site. Due to the non-uniqueness of the solution and measurement errors, the inverse problem is best solved in a statistical framework [22,29]. A common distinction is made between regression-like techniques [7,12] in which parameters, much fewer in number than the observations, of a deterministic model are fitted by minimizing a least-square criterion and methods based on a geostatistical description of the aquifer variability [10,15–17,19,25,26,30,32,36,37]. Requiring a known distribution of zones or trends, the regression techniques are more restrictive than the geostatistically based models in which the a priori knowledge is limited to the spatial correlation of the estimated property. The parameters that describe the geostatistical structure may themselves be inferred from the measurements [16,17,19].

The dependency of the heads and concentrations on the hydraulic conductivity is nonlinear. For small variations of the conductivities, linearization about the mean conductivity (or its logarithm) is valid [10,15,19,26,30,32]. If the (log-)conductivity varies more strongly, linearization may be done about the estimate itself. This approach is referred to as quasi-linear geostatistical inversing [16] or iterative cokriging [36]. The present paper describes a quasi-linear method that is applicable to highly variable conductivity fields.

In the geostatistical method of inversing, calculating the sensitivities of the measurements with respect to the parameters is often the computationally most costly step. The sensitivity matrix, or Jacobian, multiplied with the covariance matrix of the parameters yields the cross-covariance matrix connecting the variations of the parameters with those of the measurements. From the standpoint of code-development, the direct differentiation method is the simplest approach: the forward problem (flow and transport with given parameters) is repeated with a small variation of a single parameter in each run. Dividing the variations of all measured...
quantities by the variation of each parameter yields the sensitivity matrix. Thus, for $m$ parameters and $n$ measurements, $m + 1$ problems must be solved. Since we like to approximate the log-conductivity on a nodal or element-wise basis, the number of parameters, in this case element-related log-conductivities, is usually larger by orders of magnitudes than the number of measurements. Nonetheless, the approach has been used by various authors since it is easy to implement and applicable to complicated coupled problems [10,15].

The adjoint-state method for the determination of the sensitivity matrix requires the solution of only one forward problem and $n$ adjoint problems which are similar to the forward problem. The adjoint-state method has become therefore the common approach for evaluating the sensitivity of head data with respect to the log-conductivity distribution [22,29,33,34]. For inverting concentration data, however, the approach is more complicated because concentrations are not directly dependent on the hydraulic conductivity. It is also prone to numerical errors because of instabilities arising from advection domination. In the present study, we modify the adjoint-state method of Sun and Yeh [30] for concentration data. Particularly, our modification in the development of the adjoint-state equations leads to a more stable behavior, so that satisfactory sensitivity matrices can be evaluated under the application of linear Finite Element methods.

Most inverse methods in groundwater hydrology have been restricted to hydraulic-head measurements. Sun and Yeh [30,31] developed a continuous adjoint-state method and inversing procedure for concentration data. Skaggs and Barry [28] presented a discrete adjoint-state method for transient transport solved by the Laplace transform Galerkin method. Mayer and Huang [21] presented an anisotropic approach for heads and tracer data based on a genetic algorithm, including the fitting of geostatistical parameters. However, the geostatistical parameters were used for kriging of log-conductivity measurements only, neglecting any direct spatial correlation of heads and concentrations with conductivities. Harvey and Gorelick [10] showed that including the first temporal moment in inversing of head-data leads to a significant improvement of the estimate. In that study, they evaluated the moments by solving the transient transport equation and integrating the concentration history over time; then they calculated the sensitivities by the direct differentiation approach. The same authors [11] developed moment-generating equations for the direct solution of the temporal moments without solving for the concentration history. James et al. [15] used these equations in order to evaluate the distribution of retardation factors from conservative and partitioning tracer data. They used the direct differentiation approach for the calculation of the sensitivity matrix. All these studies were strictly linear and were therefore limited to cases of a small log-conductivity variance.

The advantages of using travel-time data rather than actual concentrations for the purpose of inversing have been shown by Ezzedine and Rubin [6]. The same authors analyzed the travel times of the Cape Cod data set [27]. Their aim, however, was to infer the structural parameters of the log-conductivity field rather than to solve the complete inverse problem and obtain the actual distribution. Varni and Carrera [35] developed moment-generating equations for transient flow fields. They demonstrated that neglecting dispersion in the evaluation of groundwater ages leads to discontinuous distributions causing instabilities in inversing.

In the present paper, we present a computationally efficient method for the determination of the log-conductivity distribution inferred from measurements of the hydraulic heads and tracer data. The method is efficient because we use temporal moments rather than concentration data themselves and we apply an improved adjoint-state method for the evaluation of the sensitivity matrix. The focus of the paper is on the derivation of the adjoint-state method which is a modification of the continuous approach developed by Sun and Yeh [30] and explained in detail by Sun [29]. As we will show, our modification of the method leads to a significant improvement of its numerical stability. We integrate the adjoint-state method into a quasi-linear approach of inversing based on iterative cokriging. To our knowledge, this is the first application of the adjoint-state method to temporal moments of tracer data. Due to the efficiency of the method, we can incorporate tracer data into a non-linear geostatistical method of inversing that is not restricted to the case of a small variance of the log-conductivity. We complete the paper by applying the method to an example of a tracer introduced into an injection well.

2. Governing equations

Consider flow and transport in an aquifer with spatially varying but locally isotropic hydraulic conductivity $K$. The steady-state groundwater flow equation without internal sinks and sources is

$$\frac{\partial}{\partial x_i} \left( K \frac{\partial \phi}{\partial x_i} \right) = 0 \quad (1)$$

subject to the boundary conditions:

$$\phi = \hat{\phi} \quad \text{on} \quad \Gamma_1, \quad (2)$$

$$n_i \frac{\partial \phi}{\partial x_i} K = \hat{q} \quad \text{on} \quad \Gamma_2 \quad (3)$$

in which $\phi$ is the hydraulic head and $K$ the hydraulic conductivity, $n_i$ is the unit vector normal to the
boundary \( \Gamma \) and pointing outwards. \( \phi \) and \( q \) are prescribed values of the head at the boundary and the specific-discharge component normal to it. The specific discharge \( q \) is given by Darcy’s law

\[
q_i = -K \frac{\partial \phi}{\partial x_i}.
\]  

We consider transport of a conservative tracer, introduced via an inflow boundary, described by the well-known advection–dispersion equation

\[
\frac{\partial c}{\partial t} + \frac{\partial}{\partial x} (v_i c - D_{ij} \frac{\partial c}{\partial x_j}) = 0
\]  

in which \( v_i \) is the seepage velocity and \( D_{ij} \) is the dispersion tensor given by:

\[
D_{ij} = \frac{v_i v_j}{|v|} (x_l - x_i) + \delta_{ij} (x_l v + D_m),
\]  

\[
v_i = \frac{q_i}{\theta},
\]

where \( x_l \) and \( x_i \) are the longitudinal and transverse dispersivities, \( D_m \) the molecular diffusion coefficient, \( \delta_{ij} \) the Kronecker delta-function which equals unity if the indices are identical and zero otherwise, and \( \theta \) is the porosity. We apply the following boundary conditions:

\[
n_i D_{ij} \frac{\partial c}{\partial x_j} = 0 \quad \text{on } \Gamma \setminus \Gamma_{in},
\]

\[
n_i \left( v_i c - D_{ij} \frac{\partial c}{\partial x_j} \right) = n_i v_i \hat{c}(x_r, t) \quad \text{on } \Gamma_{in}
\]

in which \( \Gamma_{in} \) is the inflow boundary, with \( n_i v_i < 0 \), and \( \Gamma \setminus \Gamma_{in} \) denotes the part of the boundary that does not belong to \( \Gamma_{in} \); \( \hat{c}(x_r, t) \) is a prescribed concentration in the inflow. In the most general case, \( \hat{c}(x_r, t) \) varies with time \( t \) and \( x_r \) denoting the spatial coordinates along the inflow boundary. In the following analysis, we consider that the tracer is introduced as a pulse with a spatially homogeneous inflow concentration:

\[
\hat{c}(x_r, t) = \hat{c} \delta(t - t_0).
\]

We may characterize the concentration history \( c(x, t) \) at any point within the domain by its temporal moments \( m_k(x) \)

\[
m_k(x) = \int_{t_0}^{\infty} \hat{c}(x, t) \, dt
\]

in which the subscript \( k \) determines the order of the temporal moment. Multiplying Eq. (5) by \( \hat{r}^k \), integrating over time and applying integration by parts leads to the moment-generating equations [11]

\[
\frac{\partial}{\partial x_j} \left( v_i m_k - D_{ij} \frac{\partial m_k}{\partial x_j} \right) = k m_{k-1}
\]

subject to the boundary conditions:

\[
n_i D_{ij} \frac{\partial m_k}{\partial x_j} = 0 \quad \text{on } \Gamma \setminus \Gamma_{in},
\]

\[
n_i \left( v_i m_k - D_{ij} \frac{\partial m_k}{\partial x_j} \right) = n_i v_i \hat{m}_k(x_r, t) \quad \text{on } \Gamma_{in}
\]

in which \( \hat{m}_k(x_r, t) \) is the \( k \)th moment of \( \hat{c}(x_r, t) \).

In the case of the Dirac-pulse inflow concentration, \( \hat{m}_k(x_r, t) \) simplifies to a constant for the zeroth moment and to zero for all higher moments. Since the transport equation is linear with respect to \( c \), the temporal development of the concentration distribution can be evaluated from the solution of a Dirac-pulse boundary-value problem by convolution. The moments of a function derived by convolution are the sums of moments of the input function and the transfer function. Hence, for homogeneously distributed inflow concentrations, we can determine the temporal moments related to the Dirac-pulse inflow concentration by subtracting the moments of the inflow concentration from that of the measured breakthrough curve.

In the following, we will restrict our analysis to the zeroth and first moment. Eq. (11) considers the resident concentration and its temporal moments which, in most field applications, hardly differ from the respective flux concentration and its moments [20]. The zeroth moment \( m_0 \) of the flux concentration is the solute mass divided by the discharge passing the observation point, whereas the first moment \( m_1 \) is the zeroth moment times the mean arrival time.

The moment-generating equations are steady-state transport equations and can be solved much faster than the transient transport equation (5). The right-hand side of Eq. (12) is zero for the zeroth moment. Assuming small transverse dispersivities, the distribution of the zeroth moment is almost binary: Inside the plume, the zeroth moment is about its value at the inflow boundary; outside the plume it is zero. The transition at the edge of the plume is rather abrupt. Obviously, if we want to apply inverse methods that use gradient-based optimization, matching the zeroth-moment will be difficult. By contrast, the first moment is monotonically increasing with the travel distance. The gradient in the direction of flow is inversely proportional to the velocity and therefore sensitive to the hydraulic conductivity.

### 3. Stochastic partial-differential equations

Stochastic partial-differential equations relate fluctuations of independent parameters, here the log-conductivity field, with those of dependent variables, here hydraulic heads and temporal moments of tracers. By introducing a continuous trial function, referred to as adjoint state, and applying Green’s theorem, we can reformulate the stochastic pde’s such that no differential
operators are applied to the state variables of interest [29]. In Section 4, we will derive sensitivities of heads and temporal moments using the adjoint states. The approach of continuous adjoint states has been applied to coupled flow and transport problems (using concentration measurements) by Sun and Yeh [30]. The following description follows that of Sun [29], but we introduce a modification that leads to higher stability in the numerical evaluation.

3.1. Groundwater flow

In the following, we will use the log-conductivity \( Y = \ln(\hat{K}) \) rather than the conductivity \( \hat{K} \). Let us assume that \( Y \) is our estimate of \( Y \) and \( \phi \) the solution of the hydraulic head associated with \( Y \). We are interested in the change of \( \phi \) if we change \( Y \) by a small perturbation \( Y' \)

\[
Y = \hat{Y} + Y',
\]

(15)

\[
\phi = \hat{\phi} + \phi'.
\]

(16)

From the perturbation of the log-conductivity \( Y \) we derive that of the conductivity \( K \) by linearization

\[
K = \hat{K}(1 + Y')
\]

(17)

in which \( \hat{K} = \exp(\hat{Y}) \). Substituting Eqs. (16) and (17) into Eq. (1) and dropping products of perturbations yields

\[
\frac{\partial}{\partial x_i} \left( \hat{K} \frac{\partial \hat{\phi}}{\partial x_i} \right) + \frac{\partial}{\partial x_i} \left( \hat{K} \frac{\partial \phi'}{\partial x_i} + \hat{K} Y' \frac{\partial \hat{\phi}}{\partial x_i} \right) = 0.
\]

(18)

Note that in the linear approach of inverting \( \hat{Y} \) is the prior mean, which is usually a constant, whereas in the quasi-linear theory of inverting \( Y \) is the best estimate of \( Y \) based on the last iteration. Taking expected values of (18) yields

\[
\frac{\partial}{\partial x_i} \left( \hat{K} \frac{\partial \hat{\phi}}{\partial x_i} \right) = 0
\]

(19)

subject to the boundary conditions:

\[
\hat{\phi} = \hat{\phi} \quad \text{on } \Gamma_1,
\]

(20)

\[
n_i \frac{\partial \hat{\phi}}{\partial x_i} \hat{K} = \hat{q} \quad \text{on } \Gamma_2.
\]

(21)

Subtracting Eq. (19) from Eq. (18) yields the linearized stochastic partial differential equation of the hydraulic heads

\[
\frac{\partial}{\partial x_i} \left( \hat{K} \frac{\partial \phi'}{\partial x_i} + \hat{K} Y' \frac{\partial \hat{\phi}}{\partial x_i} \right) = 0.
\]

(22)

The boundary conditions are:

\[
\phi' = 0 \quad \text{on } \Gamma_1,
\]

(23)

\[
n_i \frac{\partial \phi'}{\partial x_i} \hat{K} + n_i \frac{\partial \hat{\phi}}{\partial x_i} \hat{K} Y' = 0 \quad \text{on } \Gamma_2.
\]

(24)

Eq. (22) relates the head-perturbations to the log-conductivity perturbations. A weak form of Eq. (22) may be derived by introduction of the trial function \( \psi_\phi \)

\[
\int_V \psi_\phi \left( \frac{\partial}{\partial x_i} \left( \hat{K} \frac{\partial \phi'}{\partial x_i} + \hat{K} Y' \frac{\partial \hat{\phi}}{\partial x_i} \right) \right) dV = 0.
\]

(25)

Application of Green’s theorem yields

\[
- \int_V \frac{\partial \psi_\phi}{\partial x_i} \hat{K} \frac{\partial \phi'}{\partial x_i} Y' dV + \int_V \frac{\partial}{\partial x_i} \left( \hat{K} \frac{\partial \psi_\phi}{\partial x_i} \phi'\right) dV
\]

\[
+ \int_f \left[ n_i \psi_\phi \hat{K} \frac{\partial \phi'}{\partial x_i} Y' + n_i \psi_\phi \hat{K} \frac{\partial \phi'}{\partial x_i} \right] d\Gamma = 0.
\]

(26)

Considering the boundary conditions (Eqs. (23) and (24) related to Eq. (22), the boundary integrals vanish if the following boundary conditions are chosen for \( \psi_\phi \):

\[
\psi_\phi = 0 \quad \text{on } \Gamma_1,
\]

(27)

\[
n_i \hat{K} \frac{\partial \psi_\phi}{\partial x_i} = 0 \quad \text{on } \Gamma_2.
\]

(28)

If we do not apply Eq. (28) as boundary condition on \( \Gamma_2 \), the following boundary integral remains on the left-hand side of Eq. (26):

\[
\int_{\Gamma_2} -n_i \hat{K} \frac{\partial \psi_\phi}{\partial x_i} \phi' d\Gamma.
\]

(29)

It should be noted that some boundary condition for \( \psi_\phi \) on \( \Gamma_2 \) is needed in all cases. As we will see later, for coupled problems, Eq. (28) may be replaced with a more expedient choice.

Substituting Eqs. (16) and (17) into Darcy’s law yields

\[
q_i = -\hat{K}(1 + Y') \left( \frac{\partial \hat{\phi}}{\partial x_i} + \frac{\partial \phi'}{\partial x_i} \right)
\]

(30)

in which \( q_i \) is the specific discharge. The linearized mean specific discharge is given by

\[
\bar{q}_i = -\hat{K} \frac{\partial \hat{\phi}}{\partial x_i}.
\]

(31)

Subtracting Eq. (31) from Eq. (30) gives the linearized governing equation for the velocity perturbations

\[
q_i' = -\hat{K} Y' \frac{\partial \hat{\phi}}{\partial x_i} - \hat{K} \frac{\partial \phi'}{\partial x_i}
\]

(32)

3.2. Temporal moments

As in the case of groundwater flow, we introduce perturbations of the seepage velocity and the temporal moments:
subject to the boundary conditions:

\[ v_i = \bar{v}_i + v_i^\prime, \]
\[ m_0 = \bar{m}_k + m_k^\prime, \]
\[ m_{k-1} = \bar{m}_{k-1} + m_{k-1}^\prime. \]

The effects of the spatial variability of the pore-scale dispersion tensor on the mass flux are small in comparison to the velocity fluctuations. Therefore we follow the common praxis of the stochastic linear theory, and neglect the perturbations of the dispersion tensor \([9]\).

Again neglecting products of perturbations and taking expected values yields

\[ \frac{\partial \bar{m}_k}{\partial x_i} \bar{v}_i - \frac{\partial}{\partial x_i} \left( D_{ij} \frac{\partial \bar{m}_k}{\partial x_j} \right) = k \bar{m}_{k-1} \]  

subject to the boundary conditions:

\[ n_i D_{ij} \frac{\partial \bar{m}_k}{\partial x_j} = 0 \quad \text{on} \quad \Gamma \setminus \Gamma_{in}, \]
\[ n_i \left( \bar{v}_i \bar{m}_k - D_{ij} \frac{\partial \bar{m}_k}{\partial x_j} \right) = n_i \bar{v}_i \bar{m}_k \quad \text{on} \quad \Gamma_{in}. \]

Subtracting Eq. (36) from Eq. (12) gives the linearized governing equation for the \(k\)th moment perturbations

\[ \frac{\partial}{\partial x_i} \left( \bar{m}_k v_i^\prime \right) + \frac{\partial m_k^\prime}{\partial x_i} \bar{v}_i - \frac{\partial}{\partial x_i} \left( D_{ij} \frac{\partial m_k^\prime}{\partial x_j} \right) = k \bar{m}_{k-1}. \]

In the derivation of Eq. (36), it was considered that, due to the continuity of the volumetric fluxes, the expression \((\partial \bar{m}_k / \partial x_i) v_i^\prime\) is identical to \(\partial (\bar{m}_k v_i^\prime) / \partial x_i\). The boundary conditions are:

\[ n_i D_{ij} \frac{\partial m_k^\prime}{\partial x_j} = 0 \quad \text{on} \quad \Gamma \setminus \Gamma_{in}, \]
\[ n_i \left( \bar{v}_i m_k^\prime + D_{ij} \frac{\partial m_k^\prime}{\partial x_j} \right) = n_i \bar{v}_i m_k \quad \text{on} \quad \Gamma_{in}. \]

A weak form of Eq. (39) may be derived by introduction of a trial function \(\psi_k\)

\[ \int_{\gamma} \psi_k \left[ \frac{\partial m_k^\prime}{\partial x_i} v_i^\prime + \frac{\partial m_k^\prime}{\partial x_i} \bar{v}_i - \frac{\partial}{\partial x_i} \left( D_{ij} \frac{\partial m_k^\prime}{\partial x_j} \right) - k m_{k-1}^\prime \right] \, dV = 0. \]

Application of Green’s theorem yields

\[ \int_{\gamma} \left[ \psi_k \frac{\partial m_k^\prime}{\partial x_i} v_i^\prime - \frac{\partial \psi_k}{\partial x_i} v_i m_k^\prime - \frac{\partial}{\partial x_i} \left( \frac{\partial \psi_k}{\partial x_i} D_{ij} \right) m_k^\prime \right] \, dV + \int_{\gamma} n_i \left[ \psi_k \bar{v}_i m_k^\prime - \psi_k D_{ij} \frac{\partial m_k^\prime}{\partial x_j} \right] \, dV = 0. \]

In their approach of inverting transport data, Sun and Yeh [30] applied Green’s theorem also to the first term in Eq. (43). This simplifies the treatment of total-flux boundary conditions. As we will discuss in the follow-
normal to the boundary. As a consequence \( \partial \tilde{m}_i / \partial x_i = 0 \) at this boundary and the condition for \( \psi_\phi \) becomes
\[
n_i \frac{\partial \psi_\phi}{\partial x_i} = 0.
\]

(49)

4. Calculation of sensitivities

We want to find the sensitivities of the hydraulic head \( \phi \) as well as the zeroth and first temporal moments of a tracer \( m_0 \) and \( m_1 \) at a given location \( x_i \) on the log-hydraulic conductivity \( Y_i \) at multiple locations \( x_i \)
\[
\frac{\partial \phi_i}{\partial Y_i}, \quad \frac{\partial m_{0i}}{\partial Y_i}, \quad \frac{\partial m_{1i}}{\partial Y_i}.
\]

We use the continuous adjoint-state method as described by Sun and Yeh [30] and Sun [29]. These authors derived the equation on the basis of a general performance function. Since we are interested in the specific case of calculating the sensitivity matrix, we can simplify the derivation.

Consider the simulated value of the measured quantity (hydraulic head or temporal moment) denoted by \( J \) at location \( x_i \)
\[
J(\phi, m_0, m_1, x_i) = \begin{cases} 
\phi(x_i) & \text{if sensitivity of } \phi, \\
m_0(x_i) & \text{if sensitivity of } m_0, \\
m_1(x_i) & \text{if sensitivity of } m_1.
\end{cases}
\]

(50)

This is formally equivalent to
\[
J(\phi, m_0, m_1, x_i) = \int_V R(\phi, m_0, m_1, x) \, dV,
\]

(51)

where
\[
R = \begin{cases} 
\phi(x) \delta(x - x_i) & \text{if sensitivity of } \phi, \\
m_0(x) \delta(x - x_i) & \text{if sensitivity of } m_0, \\
m_1(x) \delta(x - x_i) & \text{if sensitivity of } m_1.
\end{cases}
\]

(52)

The variation \( J' \) can be derived by
\[
J' = \int_V \left[ \left( \frac{\partial R}{\partial \phi} \phi' \right) + \left( \frac{\partial R}{\partial m_0} m_0' \right) + \left( \frac{\partial R}{\partial m_1} m_1' \right) \right] dV
\]
\[
= \int_V \delta(x - x_i) \phi' \, dV \quad \text{if sensitivity of } \phi,
\]
\[
= \int_V \delta(x - x_i) m_0' \, dV \quad \text{if sensitivity of } m_0,
\]
\[
= \int_V \delta(x - x_i) m_1' \, dV \quad \text{if sensitivity of } m_1.
\]

(53)

We are interested in finding \( J' / Y_i \) for all locations \( x_i \). However, the dependency of \( J' \) on \( Y_i \) is not given in (53). Therefore we sum up Eqs. (26) and (47) (for \( k = 1 \)), (47) (for \( k = 0 \)) and (53) leading to
\[
J' = \int_V \left[ -\frac{\partial \tilde{m}_1}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \psi_1 - \frac{\partial \tilde{m}_0}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \psi_0 \right. \\
\left. - \frac{\partial \psi_\phi}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \right] Y' \, dV + \int_V \left[ \frac{\partial}{\partial x_i} \left( \frac{\partial \tilde{m}_1}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \psi_1 \right) \right. \\
\left. + \frac{\partial}{\partial x_i} \left( \frac{\partial \tilde{m}_0}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \psi_0 \right) + \frac{\partial}{\partial x_i} \left( \frac{\partial \psi_\phi}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \right) \right] \, dV
\]
\[
+ \int_V \left[ \frac{\partial \psi_\phi}{\partial x_i} - \frac{\partial}{\partial x_i} \left( D_{ij} \frac{\partial \psi_1}{\partial x_j} \right) + \frac{\partial R}{\partial \psi_0} \right] m'_i \, dV.
\]

(54)

Now we choose \( \psi_1, \psi_0 \), and \( \psi_\phi \) within the interior of the domain such that:
\[
-\frac{\partial}{\partial x_i} \left( \psi_1 \tilde{v}_i \right) - \frac{\partial}{\partial x_i} \left( D_{ij} \frac{\partial \psi_1}{\partial x_j} \right) = \frac{\partial R}{\partial \psi_1},
\]

(55)

\[
-\frac{\partial}{\partial x_i} \left( \psi_0 \tilde{v}_i \right) - \frac{\partial}{\partial x_i} \left( D_{ij} \frac{\partial \psi_0}{\partial x_j} \right) = \frac{\partial R}{\partial \psi_0} + \psi_1,
\]

(56)

\[
\frac{\partial}{\partial x_i} \left( K \frac{\partial \psi_\phi}{\partial x_i} \right) = -\frac{\partial}{\partial x_i} \left( \frac{\partial \tilde{m}_1}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \psi_1 \right) - \frac{\partial}{\partial x_i} \left( \frac{\partial \tilde{m}_0}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \psi_0 \right) - \frac{\partial R}{\partial \psi_\phi}
\]

(57)

subject to the boundary conditions:
\[
n_i D_{ij} \frac{\partial \psi_k}{\partial x_j} = 0 \quad \text{on } \Gamma \setminus \Gamma_{out},
\]

(58)

\[
n_i \left( \psi_1 \tilde{v}_i + D_{ij} \frac{\partial \psi_1}{\partial x_j} \right) = 0 \quad \text{on } \Gamma_{out},
\]

(59)

\[
\psi_0 = 0 \quad \text{on } \Gamma_1,
\]

(60)

\[
- n_i K \frac{\partial \psi_\phi}{\partial x_i} = n_i \frac{\partial \tilde{m}_1}{\partial x_i} \tilde{\psi}_\phi + n_i \frac{\partial \tilde{m}_0}{\partial x_i} \tilde{\psi}_\phi \quad \text{on } \Gamma_2.
\]

(61)

Eqs. (55)–(57), are known as the adjoint partial differential equations to Eqs. (12) and (1). \( \psi_1, \psi_0 \), and \( \psi_\phi \) are the adjoint states to \( m_1, m_0 \), and \( \phi \). Due to the dependencies, Eqs. (55)–(57) must be solved in the order of the equations’ numbers. If \( \psi_1, \psi_0 \), and \( \psi_\phi \) meet the adjoint pde’s, Eq. (54) simplifies to:
\[
J' = \int_V \left[ -\frac{\partial \tilde{m}_1}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \psi_1 - \frac{\partial \tilde{m}_0}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \psi_0 \right. \\
\left. - \frac{\partial \psi_\phi}{\partial x_i} \frac{\tilde{\psi}_\phi}{\tilde{\psi}_\phi} \right] \, dV.
\]

(62)

Since \( Y_i \) is the constant conductivity within the sub-volume \( V_i \), the sensitivity of \( J \) with respect to \( Y_i \) is given by
\[
\frac{\partial J}{\partial Y_j} = \int_{V_1} \left[ -\frac{\partial \tilde{m}_1}{\partial x_i} \tilde{K} \frac{\partial \phi}{\partial x_i} \psi_1 - \frac{\partial \tilde{m}_0}{\partial x_i} \tilde{K} \frac{\partial \psi_0}{\partial x_i} - \frac{\partial \psi_0}{\partial x_i} \tilde{K} \frac{\partial \phi}{\partial x_i} \right] dV.
\]

From the equations given above, we can derive the sensitivities of hydraulic heads, zeroth and first temporal moments by:

- In case of a head measurement, both \( \psi_0 \) and \( \psi_1 \) are zero throughout the domain. The only adjoint pde to be solved is Eq. (57) in which a point-source is considered at the location of the measurement \( x_c \). The boundary conditions simplify to a fixed value of zero at boundaries on \( \Gamma_1 \) (Eq. (60)) and no-flux boundaries of \( \psi_0 \) on \( \Gamma_2 \) (Eq. (61)).

- In case of a zeroth-moment measurement, only \( \psi_1 \) is zero throughout the domain. Two adjoint pde’s are to be solved: Eq. (56) for the adjoint state of the zeroth moment \( \psi_0 \) with a point-source at the location of the measurement \( x_c \), succeeded by Eq. (57) for the adjoint state of the head \( \psi_0 \) in which only the zeroth-moment related term is considered as source.

- In case of a first-moment measurement, all three adjoint states must be calculated. First we solve for the adjoint state of the first moment \( \psi_1 \) by Eq. (55) with a point-source at the location of the measurement \( x_c \). Then we solve for the adjoint state of the zeroth moment by Eq. (56) with the value of \( \psi_1 \) as distributed source. Finally we solve for the adjoint state of the heads \( \psi_0 \) by Eq. (57) with the distributed source related to both moments.

The improvement of our formulation in comparison to that of Sun and Yeh [30] may be illustrated by a zeroth-moment measurement. If we had applied Green’s theorem to the first term of Eq. (43) as supposed by Sun and Yeh [30], the term \( \int_{V_1} (\partial \tilde{m}_0 / \partial x_i) \psi_0 dV \) would be replaced by \( (\partial \psi_0 / \partial x_i) \tilde{m}_0 \). Since the distribution of \( \psi_0 \) is dominated by the point-source at \( x_c \), it varies much stronger than \( \tilde{m}_0 \), and its gradient is more prone to numerical errors. Therefore, it is computationally advantageous to consider the source-term

\[
- \frac{\partial}{\partial x_i} \left( \frac{\partial \tilde{m}_0}{\partial x_i} \tilde{K} \frac{\partial \psi_0}{\partial x_i} \right)
\]

rather than

\[
\frac{\partial}{\partial x_i} \left( \frac{\partial \psi_0}{\partial x_i} \tilde{K} \frac{\partial \phi}{\partial x_i} \tilde{m}_0 \right)
\]

in the governing equation for \( \psi_0 \). This may illustrated for the case of a homogeneous-flux boundary condition of \( \tilde{m}_0 \) at all inflow boundaries. In this case, \( \tilde{m}_0 \) is constant throughout the domain regardless of the log-conductivity distribution. Consequently, the sensitivity of \( \tilde{m}_0 \) with respect to \( Y \) is zero everywhere. Using the expression derived in the present paper, this behavior is revealed: with \( \partial \tilde{m}_0 / \partial x_i = 0 \) there is no source-term for \( \psi_0 \), and the latter is zero everywhere. Also the sensitivities \( \partial \psi_0 / \partial Y \) are zero because \( \partial \tilde{m}_0 / \partial x_i = 0 \) and \( \partial \psi_0 / \partial x_i = 0 \). By contrast \( \partial \psi_0 / \partial x_i \neq 0 \) so that there is a source-term for \( \psi_0 \) if one uses the method of Sun and Yeh [30]. Only if the numerical scheme is extremely accurate, the effects of \( \partial \psi_0 / \partial x_i \) and \( \partial \psi_1 / \partial x_i \) will cancel out in the evaluation of the sensitivities. The same arguments apply to the sensitivities of first-moment measurements. Sun [29] recommended the application of higher-order FEM for the calculation of the forward and adjoint transport problems. This is no more necessary with our approach.

5. Numerical implementation

The numerical implementation described in the following is restricted to the two-dimensional case. Extension to a third spatial dimension is straightforward. We use the standard FEM for the discretization of the groundwater flow equation and its adjoint. \( \phi \) and \( \psi_0 \) are approximated by bilinear functions \( N \) within rectangular elements of constant size and the hydraulic conductivity is assumed constant within an element.

The system of equations is assembled from the element-related matrices \( M^0_{\psi} \) that are identical for all elements except for the element-wise constant conductivity. The same matrices are used for the evaluation of the sensitivities of head-measurements

\[
M^0_{\psi} = \int_{V_\psi} \nabla N^T \nabla N dV,
\]

\[
\frac{\partial \psi_0}{\partial Y_j} = - \int_{V_1} \frac{\partial \psi_0}{\partial x_i} \tilde{K} \frac{\partial \phi}{\partial x_i} dV = - \tilde{K} \tilde{\psi}_0 M^0_{\psi} \psi_0
\]

in which \( \tilde{\psi}_0 \) and \( \tilde{\phi} \) are the vectors of \( \psi_0 \)- and \( \phi \)-values at the vertices of the element related to \( Y_j \). For the solution of advective-dispersive equations, we use the Streamline-Upwind Petrov–Galerkin (SUPG) method [2] which is identical to the Galerkin Least-Square (GLS) method [13] when using bilinear elements. Starting point is the steady-state advection–dispersion equation

\[
\frac{\partial m}{\partial t} + \frac{\partial}{\partial x_i} \left( D_{ij} \frac{\partial m}{\partial x_j} \right) = r
\]

in which the index of the moment has been omitted and possible source-terms are expressed by \( r \). For the adjoint states, the principal structure of the governing equation is identical, but the velocity vector points into the opposite direction. Let \( (\cdot, \cdot)_V \) denote the inner product in the entire domain and \( (\cdot, \cdot)_{\psi V} \) the corresponding
boundary integral, then the SUPG formulation of Eq. (66) is [2]

\[
( - u_m + D\nabla m, \nabla N)_v + (u_m - D\nabla m, N)_v
+ (v - \nabla \nabla N)_v \\
= (r, N)_v + (r, v \cdot \nabla N)_v \\
\]

(67)
in which Green’s theorem has been applied to both transport terms. At the inflow boundaries of the domain, the advective part of the boundary integral may be multiplied by the vector of nodal inflow moments \( \mathbf{m}_n \). At outflow boundaries of the domain, the boundary integral leads to extra terms on the main diagonal of the system of equations. The stabilization term acts like streamline-diffusion but is treated consistently. The stabilization factor \( \tau \) is determined by [2,8,13]:

\[
\tau = \frac{h_d}{2v} \zeta(P_{el}), \\
\]

(68)

\[
\zeta(P_{el}) = \coth \left( \frac{P_{el}}{2} \right) - \frac{2}{P_{el}} \approx \sqrt{\frac{P_{el}^2}{36 + P_{el}^2}}, \\
\]

(69)

\[
h_d = \frac{hk\sqrt{2}}{\sqrt{h^2 + k^2}}, \\
P_{el} = \frac{\|v\|}{\zeta\|v\| + D_m}, \\
\]

(70)

where \( h_d \) is an effective grid-spacing, \( P_{el} \) is the element-related Péclet number, \( \zeta(P_{el}) \) an upwind coefficient, \( \|\| \) the Euclidean norm, \( h \) and \( k \) are the dimensions of the element in the \( x \)- and \( y \)-directions. It is important to notice that the SUPG stabilization does not only influence the advective transport term but also the source term. This is of particular interest for the adjoint-state equations. Boundary conditions, however, are not affected. Thus we need to evaluate the element-matrices for advective-dispersive transport \( \mathbf{M}_{el}^{ad} \) and for the source term \( \mathbf{M}_{el}^s \):

\[
\mathbf{M}_{el}^{ad} = \int_{V_{el}} \left( - \nabla N^T \mathbf{v} N + \nabla N^T D \nabla N \right) \\
+ \tau(\mathbf{v}^T \nabla N)^T \nabla N \right) \, dV, \\
\]

(72)

\[
\mathbf{M}_{el}^s = \int_{V_{el}} \left[ N^T \mathbf{N} + \tau(\mathbf{v}^T \nabla N)^T \mathbf{N} \right] \, dV. \\
\]

(73)
The integration is done by Gaussian quadrature using three points per direction. The element-matrices are assembled to the matrices of the entire system.

We restrict the boundary conditions to advective-flux conditions. That is, all diffusive fluxes are set to zero at the boundaries of the domain. This simplification leads only to minor errors, since we consider the advection-dominated case. Also we do not consider fixed temporal moments (although the implementation would be straightforward). Then we have to distinguish between inflow and outflow boundaries. At inflow boundaries, the temporal moments of the flux concentrations must be prescribed. At outflow boundaries, the mass-flux leaving the domain equals the volumetric flux times the concentration:

\[
\mathbf{M}_{mod}^{ad} |_{for}(i, i) = \mathbf{M}_{mod}^{ad} |_{for}(i, i) + \frac{q_i}{n} \forall i | q_i > 0, \\
\mathbf{M}_{mod}^{ad} |_{ad}(i, i) = \mathbf{M}_{mod}^{ad} |_{ad}(i, i) - \frac{q_i}{n} \forall i | q_i < 0, \\
\mathbf{M}_{mod}^{ad} |_{for} = \mathbf{M}_{mod}^{ad} |_{for} + \frac{q_i}{n} \hat{m}(i) \forall i | q_i < 0. \\
\]

Calculating the sensitivity matrix requires solving for one steady-state flow equation and two steady-state transport equations for the forward problem, one flow equation for each head measurement, and two transport plus one flow equation for each first-moment measurement. If enough memory is available, all adjoint states can be stored so that the solution of the last iteration can be used as initial guess for the solver. If memory is limited, the adjoint states may be kept only until the sensitivity of the specific measurement is calculated.

6. Quasi-linear method of inversing

The adjoint-state method for the calculation of sensitivities may be applied to any gradient-based inversion scheme. However, it is particularly attractive in geostatistical models in which the log-conductivity is modeled as a random spatial function discretized at the same level as the heads and the concentration moments. In this framework, we consider that the spatial distribution of log-conductivity has the following linear structure:

\[
Y(x) = \sum_{k=1}^p f_k(x) \beta_k + \epsilon(x) \]

(77)
in which the first term is the deterministic part with known spatial functions \( f_k(x) \) and unknown linear coefficients \( \beta_k \). The simplest case is that of an unknown constant mean, \( p = 1, f_1(x) \) equals unity throughout the domain, and \( \beta_k \) is the mean. \( \epsilon(x) \) is a random function with zero mean and characterized through a generalized covariance function. After discretization, \( Y \) is represented by an \( m \times 1 \) vector \( \mathbf{s} \) in which \( m \) is the number of elements. The mean vector and covariance matrix of \( \mathbf{s} \) are:

\[
E[\mathbf{s}] = X \beta, \\
E[(\mathbf{s} - X \beta)(\mathbf{s} - X \beta)^T] = \mathbf{Q},
\]

(78)

(79)

where \( X \) is a known \( m \times p \) and \( \beta \) are \( p \) unknown drift coefficients. \( \mathbf{Q} \) is a known \( m \times m \) matrix. Let \( \mathbf{y} \) be the \( n \times 1 \) vector of observations (both heads and temporal moments) consisting of part \( \mathbf{h}(\mathbf{s}) \) predicted by \( \mathbf{s} \) and an error component \( \mathbf{v} \):

\[
\mathbf{y} = \mathbf{h}(\mathbf{s}) + \mathbf{v}. \\
\]

(80)
The observation error is assumed to be a random, normally distributed function with zero mean and known covariance matrix \( R \). The most common model for \( R \) is based on uncorrelated measurement errors so that \( R = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2) \) with the vector of measurement errors \( \sigma \). In the quasi-linear method of inversing, we linearize Eq. (80) about the last estimate of \( s \) denoted by \( \hat{s} \):

\[
\mathbf{h}(s) = \mathbf{h}(\hat{s}) + \mathbf{H}(s - \hat{s})
\]

in which \( \mathbf{H} \) is the Jacobian or sensitivity matrix the calculation of which has been explained in the previous sections

\[
\mathbf{H} = \left. \frac{\partial \mathbf{h}}{\partial s} \right|_{s=\hat{s}}.
\]

We estimate \( s \) by the continuous function \( \hat{s} \) [18]

\[
\hat{s} = \mathbf{Xb} + \mathbf{QH}^T \hat{\xi}
\]

in which \( \mathbf{b} \) is the \( 1 \times p \) vector of the \( \beta \)-estimate and \( \hat{\xi} \) is a \( 1 \times n \) vector of weights associated with the measurements. We find \( \mathbf{b} \) and \( \hat{\xi} \) by solving the function-estimate form of the linearized cokriging system [18]

\[
\begin{bmatrix}
\Sigma \\
\mathbf{H} \mathbf{X}^T \\
0
\end{bmatrix}
\begin{bmatrix}
\hat{\xi} \\
\mathbf{b}
\end{bmatrix}
= 
\begin{bmatrix}
\mathbf{y} - \mathbf{h}(\hat{s}) - \mathbf{H} \hat{s}
\end{bmatrix}
\]

with

\[
\Sigma = \mathbf{H} \mathbf{Q} \mathbf{H}^T + \mathbf{R}.
\]

Since \( \mathbf{H} \) depends on values chosen for \( s \) estimated by Eq. (83), the procedure has to be repeated until \( \hat{s} \) is practically equal to \( \hat{s} \). We use a weighted root mean-square criterion for quantifying convergence

\[
\varepsilon = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \hat{s}_i - \hat{s}_i^* \right)^2 / P(i,i)}
\]

in which \( \hat{V} \) is the \( a \) posteriori covariance matrix of estimation. Typically, we require \( \varepsilon \) to be smaller than 0.1 or 0.25. \( \hat{V} \) can be calculated from the inverse of the cokriging matrix:

\[
\begin{bmatrix}
\mathbf{P}_{xy} & \mathbf{P}_{xb} \\
\mathbf{P}_{yb}^T & \mathbf{P}_{bb}
\end{bmatrix}
= 
\begin{bmatrix}
\Sigma \\
\mathbf{H} \mathbf{X}^T \\
0
\end{bmatrix}
^{-1}
\]

where \( \mathbf{P}_{xy} \) is \( n \times n \), \( \mathbf{P}_{xb} \) is \( n \times p \) and \( \mathbf{P}_{bb} \) is \( p \times p \). Then

\[
\hat{V} = \mathbf{Q} - \mathbf{QH}^T \mathbf{P}_{xy} \mathbf{H}^T - \mathbf{X}^T \mathbf{P}_{bb} \mathbf{X}^T

= \mathbf{QH}^T \mathbf{P}_{xy} \mathbf{X}^T
\]

Typically, the values of the first moment measurements are several orders of magnitude larger than those of the head measurements, leading to poorly conditioned cokriging matrices. This problem can be overcome by multiplication of the measured first moments and the corresponding measurement errors with a constant leading to values in the same order of magnitude as that of the heads. The same factor must be applied to the sensitivities of the first-moment measurements.

While we refer to the method above as \textit{iterative cokriging}, it is identical to the Gauss–Newton method for maximizing the restricted likelihood function of the \( Y \)-field [16]. Therefore the \textit{a posteriori} covariance matrix of estimation \( \hat{V} \) is unbiased. The method can easily be extended to include the estimation of the structural parameters as described by Kitanidis [16].

7. Application

In order to illustrate the method, we have created an artificial two-dimensional test case. All parameters are listed in Table 1. \( L_1 \) and \( L_2 \) are the dimensions of the aquifer. An injection well of specific strength \( Q_w \) is located at \( x_w, y_w \). No flow is allowed across the top and bottom boundaries, whereas the heads are fixed at the left-hand and right-hand side boundaries. The “true” log-conductivity distribution was generated on a rectangular grid using the spectral approach of Dykaar and Kitanidis [5]. An anisotropic exponential covariance function for the log-conductivities was chosen. This conductivity field is shown in Fig. 1(a).

Based on the “true” log-conductivity distribution, we simulated the hydraulic heads, the zeroth and the first temporal moments of a tracer introduced into the injection well. The zeroth moment of the inflow concentration in the well \( m_0 \) was set to unity and the corresponding first moment \( m_1 \) to zero. A random error with zero mean and standard deviation \( \sigma_w \) was added to the head-values at the grid-points of \( x = 5, 10, 15, 20, 25, 30, 35 \) m and \( y = 4, 8, 10, 12, 16 \) m. These values were taken as head measurements. The true head-distribution and the locations of the head “measurements” are shown in Fig. 3(a). The zeroth and first moments were calculated for a tracer test in which the tracer is injected as a pulse into the injection well. A random error with zero mean and standard deviation \( \sigma_m \) of 10% of the actual value was added to the first-moment values at the grid-points of \( x = 10, 15, 20, 25, 30, 35 \) m and \( y = 8, 10, 12, 16 \) m except the well-location itself. These values were taken as first-moment measurements. The true zeroth-moment distribution is shown in Fig. 4(a) and the first-moment distribution in Fig. 5(a).

From the artificially created measurements we inferred the best estimate of the log-conductivity field by the quasi-linear method of inversing described above. We compared best estimates using only the head measurements, only the first-moment measurements and both types of measurements. The boundary conditions and the structural parameters of the
Table 1
Parameters of the test case

<table>
<thead>
<tr>
<th>Geometric parameters and discretization</th>
<th>$L_x = 20$ m</th>
<th>$\Delta x = 0.2$ m</th>
<th>$\Delta y = 0.2$ m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geostatistical parameters of the $\ln(K)$ field ($K$ in m/s)</td>
<td>$\sigma_y^2 = 1$</td>
<td>$\lambda_y = 4$ m</td>
<td>$\lambda_y = 2$ m</td>
</tr>
<tr>
<td>Boundary conditions for flow</td>
<td>$x_w = 10$ m</td>
<td>$y_w = 10$ m</td>
<td>$Q_w = 1.5 \times 10^{-4}$ m$^2$/s</td>
</tr>
<tr>
<td>Transport parameters</td>
<td>$x_t = 0.05$ m</td>
<td>$z_t = 0.01$ m</td>
<td>$D_m = 10^{-9}$ m$^2$/s</td>
</tr>
<tr>
<td>Measurement error</td>
<td>$\sigma_y = 5 \times 10^{-4}$ m</td>
<td>$\sigma_m, m_1 = 0.1$</td>
<td></td>
</tr>
</tbody>
</table>

*Fig. 1. Distribution of log-conductivity $Y = \ln(K)$ ($K$ in m/s). (a) True distribution. (b)–(d) Best estimates using indicated type of data for inversing. (○) Locations of head measurements; (⋆) locations of first-moment measurements.*
covariance matrix were identical to the true solution. The measurement errors were assumed uncorrelated with the actual values of the standard deviations used for the creation of the “measurements”. As initial guess, the log-conductivity was assumed to be $-7$ throughout the domain, however, the method converged also with different initial values. Using both types of measurements, the method converged to the convergence criterion $\varepsilon = 0.25$ as defined in Eq. (86) in four iterations and to $\varepsilon = 0.1$ in six iterations. Using only the head measurements, $\varepsilon = 0.1$ was reached within three iterations, whereas this criterion could not be met within twenty iterations when only the first-moment data were used. The results shown for inverting only the first-moment data refer to $\varepsilon = 0.12$ which was reached after seven iterations.

The best estimates of the log-conductivity distribution are shown in Fig. 1(b)–(d) and the corresponding variance of estimation $\hat{\sigma}^2 = \text{diag}(\hat{V})$ in Fig. 2. The best estimate using only the head data shown in Fig. 1(b) already includes most of the large-scale features. As one anticipates from a best estimate, it is a smoother field than the true log-conductivity field shown in Fig. 1(a). It should be mentioned that the assumed value of 0.5 mm for the standard deviation of head measurements is extremely optimistic for field measurements. Such accuracy in the measurement of hydraulic heads can be achieved normally only under well-controlled laboratory conditions. Had we used a larger error criterion, less features would have been detected.

The best estimate based on first-moment data alone shown in Fig. 1(c) does not recover any features outside of the plume since there is hardly any sensitivity of concentration data to these conductivity values. The sensitivity of first-moment data to a narrow stripe of log-conductivity values makes the inversion procedure less stable and the best estimates less reliable than that of head data. These findings are in

---

**Fig. 2.** Estimation variance of log-conductivity $\hat{\sigma}^2$ (m$^2$/s$^2$). (a) Based on head data. (b) Based on first-moment data. (c) Based on head and first-moment data. (o) Locations of head measurements; (+) locations of first-moment measurements.
agreement with the study of Varni and Carrera [35] on groundwater ages. Combing head and first-moment data in inversing as shown in Fig. 1(d) leads to an improvement of the estimate based on head data alone (Fig. 1(b)). In addition to the large-scale features that had been already detected by the head inversion, details at a smaller scale are recovered within the plume. The improvement is also evident from the variance of estimation shown in Fig. 2. In comparison to the variance from head-inversion (Fig. 2(a)), the additional information of mean solute arrival times leads to tighter error bounds in the interior of the plume (Fig. 2(c)).

For quantitative analysis, we calculated the difference $\Delta Y$ between the estimated log-conductivity and the actual value for all inverse solutions. Table 2 lists the root mean square of $\Delta Y$, the standard deviation of estimation $\sigma_Y$ and the weighted error $\Delta Y/\sigma_Y$. Including first-moment data improves the absolute error RMS($\Delta Y$) by

<table>
<thead>
<tr>
<th>Data used</th>
<th>RMS ($\Delta Y$)</th>
<th>RMS ($\sigma_Y$)</th>
<th>RMS ($\Delta Y/\sigma_Y$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only heads</td>
<td>0.982</td>
<td>0.831</td>
<td>1.176</td>
</tr>
<tr>
<td>Only first moments</td>
<td>1.128</td>
<td>0.928</td>
<td>1.216</td>
</tr>
<tr>
<td>Heads and first moments</td>
<td>0.896</td>
<td>0.782</td>
<td>1.153</td>
</tr>
</tbody>
</table>
10%. It should be noted, however, that this quantity refers to the log-conductivity values in the entire domain, whereas the first-moment measurements affect essentially only the conductivities in the interior of the plume.

Figs. 3(b), 4(b) and 5(b) show the distribution of the hydraulic heads, zeroth and first moment based on the best estimate of the log-conductivity distribution using both types of data. The distributions are smoother than the true values shown in Figs. 3(a), 4(a) and 5(a), but they are very similar, and the measurements are reproduced within the prescribed measurement error.

8. Discussion and conclusions

Tracer data contain valuable information that should be included in inverse methods whenever they are available. So far, however, this has been done mainly in the context of regression-like techniques [1,23]. Including tracer information into geostatistically based models has been discussed by various authors [10,15,29,30], but the majority of the studies on geostatistically based inversing have been restricted to head-measurements. A particular reason may be the high computational effort of the direct differentiation approach for the evaluation of sensitivities. The direct differentiation approach is appropriate for regression-like techniques because they have only a few parameters. This is not the case in geostatistical inverse problems in which the conductivities values in all elements or at all nodes are to be estimated. For such problems, adjoint-state methods are much more efficient. However, the continuous adjoint-state method for coupled problems of Sun and Yeh [30] is sensitive to numerical errors, requiring high-order Finite Element methods [29].

A second obstacle in using tracer data for inversing is that individual concentrations are not very informative [6]. Concentrations at each observation point must be measured at many times and then processed in order to identify the conductivity distribution. Using temporal moments rather than directly concentrations does not only decrease the computational effort and the memory requirements, it also leads to more sensitive information [6,10]. The moment-generating equations derived by Harvey and Gorelick [11] and applied in the present study, allow to compute temporal moments by solving steady-state transport equations. Together with the improved adjoint-state method presented here, this makes it feasible to perform multiple evaluations of the sensitivity matrix in large domains.

With the improved efficiency of our method we can apply tracer data to the quasi-linear method of inversing. The linear method used so far for inversing temporal moments [10,15] is restricted to the case of a small variance. In our application, the estimate of the first iteration that is identical to the solution of the linear method was considerably different from the final result.

Our modification of the adjoint-state method of Sun and Yeh [30] seems minor but its effect on the numerical accuracy is remarkable. In the approach of Sun and Yeh [30], the source-term for the adjoint state of the heads is evaluated by differentiating the product of the concentration and the gradient of its adjoint state. Since the adjoint state of the concentration (in our case of the temporal moments) is the solution of a point-source problem, its gradient can be calculated accurately only with high-order methods. By contrast, the gradients of the temporal moments are approximated more accurately, since the temporal moments are relatively smooth functions. By reversing the order of differentiation, we can therefore use linear function spaces for the discretization of heads and temporal moments without introducing major numerical errors.

For the stabilization of advection-dominated transport, we use the SUPG method [2]. It is well known that the standard Galerkin method or central differentiation in Finite Differences leads to excessive oscillations in the vicinity of discontinuities. The point-source problem associated with the adjoint states of the temporal moments is exactly the type for which the standard Galerkin method would fail. It should be mentioned, however, that the SUPG method is not monotonic, that is, spurious oscillations remain. In the application shown, this is the case in the forward-problem near the stagnation point, and in the adjoint-state problem parallel to the adjoint-state “plume”. These small oscillations appear also in the sensitivity fields. Fortunately, they do not cause stability problems, since the multiplication of the sensitivity matrix with the covariance matrix of the log-conductivities smooths small-scale fluctuations. This would have been different if the integral scale of the conductivity had been of the size of the grid spacing. Oscillations could be suppressed entirely by the introduction of a nonlinear discontinuity-capturing operator [3,14,24]. However, we decided against these techniques in order to avoid additional linearization.

We did not use zeroth-moment information in the inversing procedure since the sensitivities of the zeroth moment were essentially zero. The almost binary information content of the zeroth moment (within the plume versus outside the plume) may be used in statistical procedures not relying on gradients. Further research will be necessary to combine these type of data with the gradually varying heads and arrival times.

Our application indicates that inversing only arrival times without using head data might lead to
convergence problems. Arrival-time data are very sensitive to the log-conductivities along a narrow stripe upstream of the measurement location, whereas hydraulic heads are influenced from conductivities in all directions. We recommend therefore to combine the more stable head data and the more specific arrival-time data in inversing efforts.

Our method can easily be extended to higher-order temporal moments. For the second central moment, however, the dispersion parameters will have a great impact. The small-scale variability of the log-conductivity field enhances dilution of the plume and therefore increases the growth of the second central moment [4]. The best estimate of the log-conductivity field is smoother than the actual distribution so that results of the second-central moment will be biased. Effective dispersion parameters may be estimated jointly with the log-conductivity distribution using hydraulic heads as well as first and second temporal moments. The meaning of these parameters for dilution and reactive mixing may be investigated in follow-up studies.

References


