Hydrol. Earth Syst. Sci. Discuss., 11, 4163–4208, 2014
www.hydrol-earth-syst-sci-discuss.net/11/4163/2014/
doi:10.5194/hessd-11-4163-2014

Hydrology and Earth System Sciences
Open Access
Discussions

Estimation of heterogeneous aquifer parameters using centralized and decentralized fusion of hydraulic tomography data from multiple pumping tests

A. H. Alzraiee, D. Baú, and A. Elhaddad

Department of Civil and Environmental Engineering (1372), Colorado State University, Fort Collins, CO, 80523, USA

Received: 10 March 2014 – Accepted: 28 March 2014 – Published: 11 April 2014

Correspondence to: A. H. Alzraiee (ayman.alzraiee@gmail.com)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Characterization of spatial variability of hydraulic properties of groundwater systems at high resolution is essential to simulate flow and transport phenomena. This paper investigates two schemes to invert transient hydraulic head data resulting from multiple pumping tests for the purpose of estimating the spatial distributions of the hydraulic conductivity, K, and the specific storage, S_s, of an aquifer. The two methods are centralized fusion and decentralized fusion. The centralized fusion of transient data is achieved when data from all pumping tests are processed concurrently using a central inversion processor, whereas the decentralized fusion inverts data from each pumping test separately to obtain optimal local estimates of hydraulic parameter, which are consequently fused using the Generalized Millman Formula, an algorithm for merging multiple correlated or uncorrelated local estimates. For both data fusion schemes, the basic inversion processor employed is the Ensemble Kalman Filter, which is employed to assimilate the temporal moments of the transient hydraulic head measurements resulting from multiple pumping tests. Assimilating the temporal moments instead of the hydraulic head transient data themselves is shown to provide a significant improvement in computational efficiency. Additionally, different assimilation strategies to improve the estimation of S_s are investigated. Results show that estimation of the K and S_s distributions using temporal moment analysis is fairly good; however, the centralized inversion scheme consistently outperforms the decentralized inversion scheme. Investigations on the sensitivity of the inversion estimates to errors in geostatistical parameters of the random fields of K and S_s reveal that the estimates are not sensitive to errors in the correlation length and the variance of hydraulic properties, but are noticeably sensitive to errors in the stationary mean. The proposed inversion schemes are expanded to estimate the geostatistical parameters of the K and S_s fields. The results show that the estimation of the true stationary mean of the K field and, to a lesser degree, the stationary mean of the S_s field can be successfully achieved, while

the estimation of correlation length and standard deviation for both the K and S_s fields are not as effective.

1 Introduction

A detailed description of hydraulic properties, such as hydraulic conductivity, K, and specific elastic storage, S_s, of groundwater systems is essential to predict flow and solute transport in porous media. Typically, these properties are inherently heterogeneous, and cannot be determined uniquely using a finite set of sparse measurements. A direct method to map the spatial variability of these properties is to collect a large number of core samples, which are then analyzed in the laboratory to obtain conductivity and storage properties. These methods, however, are laborious, expensive, and time consuming (Butler Jr. et al., 1999). In general, sampling of groundwater system states, such as hydraulic head or solute concentrations, is relatively easier and more cost-effective. Therefore, characterization of the aquifer parameters using system states can be achieved by solving an inverse problem (Sun, 1994; Tarantola, 2004).

Pumping tests (Theis, 1935; Cooper and Jacob, 1946) and slug tests (Butler Jr., 1998) are classical examples of inverse methods used to infer hydraulic properties of porous media using hydraulic head data. In a pumping test, an aquifer is stressed at a well and the response of the hydraulic head field is monitored at a number of observation wells. The resulting data are processed using an analytical solution to obtain a lumped estimate of the transmissivity and the storativity of the aquifer at a scale equal to the radius of the developed cone of depression. While these estimates are useful to guide future groundwater development of an aquifer at a regional scale, they provide little or no information about the local spatial variability of parameters, which is essential, for example, to model solute transport processes. In addition, the estimates obtained by pumping tests are shown to be affected by the location of the

HESSD
11, 4163–4208, 2014

pumping well and the degree of heterogeneity within the cone of depression (Wu et al., 2005).

A relatively recent alternative method for estimating the spatial distribution of aquifer parameters at a high resolution is hydraulic tomography (HT) (Gottlieb and Dietrich, 1995; Butler Jr. et al., 1999; Yeh and Liu, 2000; Berg and Illman, 2011). In HT, an aquifer is stressed at different locations and the responses to these stresses at a network of observation wells are inverted to map the parameters spatially.

During the last decade, HT has been intensively studied to assess its performance both numerically and experimentally, with few field applications (Straface et al., 2007; Bohling et al., 2007). HT studies have covered several flow conditions, ranging from steady-state flow (Yeh and Liu, 2000) to transient flow (Zhu and Yeh, 2005) in both confined and unconfined aquifers (Cardiff and Barrash, 2011). HT has been applied to joint unconfined and vadose zone flow problems (Mao et al., 2013) and for both 2-D (Yeh and Zhu, 2007) and 3-D settings (Cardiff et al., 2012). A number of sandbox laboratory experiments have been performed to validate HT methods off-site (Liu, 2002; Liu et al., 2007; Illman et al., 2010), which have deemed HT a promising technique for characterizing aquifer properties at high resolution. For instance, Illman et al. (2010) compared various approaches to characterize the K field using a sandbox and found that HT consistently outperformed kriging interpolation of small scale K measurements. A comprehensive list of previous HT studies is provided by Cardiff and Barrash (2011).

In HT studies, hydraulic head transient data have been inverted using different algorithms, such as the sequential successive linear estimator (SSLE) (Yeh and Liu, 2000), the quasi-linear approach (Kitanidis, 1995; Liu and Kitanidis, 2011), the Bayesian Maximum a Posteriori (MAP) approach (Castagna and Bellin, 2009), and the Ensemble Kalman Filter (EnKF) (Schöniger et al., 2012).

Despite the success in verifying its estimates numerically and experimentally, HT faces two major challenges related to the heavy computational burden associated with the inversion process (Zhu and Yeh, 2005) and the non-uniqueness of the solution of the inverse problem, a situation where infinite possible combinations of input parameters and model structures produce the same model output (Moore and Doherty, 2006). With respect to the latter, Bohling and Butler (2010) caution practicing hydrologists against "overselling" the reliability of HT estimates, and argue that some form of regularization is typically necessary to reduce uncertainties associated with the non-uniqueness effect. In this work, HT data are inverted using the EnKF. While not resolving the non-uniqueness issue completetely, inversion algorithms based on the EnKF constitute an ideal framework to handle the problem of non-uniqueness resulting from parameter uncertainty only, as opposed to non-uniqueness resulting from uncertainty in conceptual models and process assumptions.

With its roots in Bayesian analysis, the EnKF updates a prior ensemble of possible realizations of system states and parameters based upon collected state measurements, so that the posterior state-parameters ensemble resembles a non-unique set of possible solutions. Therefore, the ensemble mean of the posterior ensemble provides an unbiased estimate of the system parameters. The EnKF offers several other advantages, such as computational efficiency (Franssen and Kinzelbach, 2009), avoiding sensitivity computations, such as those required by the SSLE (Yeh and Liu, 2000), and improved accuracy when using ensemble-based covariance estimations instead of sensitivity-based covariance estimations (Schöniger et al., 2012).

A possible effective approach to improving parameter estimations for ill-posed problems is by integrating data from independent sources, which may be related to different physical processes, such as hydraulic, geophysical, geomechanical, chemical processes (Bohling and Butler, 2010). In this situation, different physical processes (models) are utilized to relate measured responses to aquifer properties. The inversion of such multi-source data may take two general avenues: centralized fusion (CF) and decentralized fusion (DF). In this work, we investigate and compare the two approaches, one based on CF and another based on DF, to assimilating transient hydraulic head HT data for the characterization of the K and S_s fields of a confined

HESSD
11, 4163–4208, 2014

aquifer. With the CF method, all data resulting from all experiments are inverted simultaneously using a single "global" EnKF. The DF method, on the other hand, assimilates each data set resulting from a single experiment separately using a "local" EnKF to obtain a local estimate of parameters. The multiple local estimates are then "fused" using the Generalized Millman Formula (GMF) algorithm (Bar-Shalom and Campo, 1986; Shin et al., 2006), which is an unbiased linear estimator of multiple correlated or uncorrelated estimates. The two inversion schemes are implemented to assimilate the responses resulting from five pumping tests. However, the methodology can be generalized to merge multiple parameter estimations resulting from inverting different physical processes.

As mentioned earlier, computational cost constitutes an issue for the application of HT methods for aquifer characterization. Typically, HT-based algorithms require inverting a large amount of transient data resulting from multiple experiments and at multiple observation wells, which produces the so-called "data-overload" problem (Zhu and Yeh, 2005). Assimilation of transient data with the EnKF is computationally intensive for two reasons. First, the computation of the forecast ensemble of states and parameters requires simulating transient flow for a large number of realizations, which typically involves a considerable computational effort. Second, the resulting spatio-temporal cross-covariance matrix is typically large and difficult to manipulate. In this study, we propose to assimilate temporal moments of transient hydraulic head data (Harvey and Gorelick, 1995; Von Asmuth and Maas, 2001; Li et al., 2005; Bakker et al., 2008; Olsthoorn, 2008; Von Asmuth et al., 2008), rather than the hydraulic head data themselves.

In the temporal moment analysis, the original parabolic partial differential equation (PDE) governing groundwater flow is transformed into two simpler and easier to solve Poisson-type PDEs (Zhu and Yeh, 2006; Li et al., 2005). Although it has been shown that inversion of temporal moments provides a drastic reduction in central processing unit (CPU) time and a reliable estimate of the K field, it has also been found to produce an unreliable characterization of the S_s field (Yin and Illman, 2009). In this work, we

HESSD
11, 4163–4208, 2014

devise a strategy that can optimize the estimation of the S_s field, while still benefiting from the reduced problem complexity achieved with the temporal moment formulation.

The article is organized as follows. The methodologies of the two inversion schemes are presented in Sect. 2. A description of the numerical experiments used to investigate the inversion approaches is provided in Sect. 3. In Sect. 4, the obtained results are presented and discussed.

2 Methodology

In the following, we provide an overall description of the proposed HT approaches, followed by a detailed description of each component of the methodology. For the purpose of estimating the hydraulic parameters K and S_s, we assume that a series of separate pumping tests is conducted from N_p wells installed at different locations within a confined aquifer. In each pumping test, the pumping well i ($i \in \{1,2,...,N_p\}$) is operated at the flow rate Q_i. The resulting transient hydraulic head data, $h_{ij}(t)$ are recorded at number N_o of observation wells ($j \in \{1,2,...,N_o\}$).

The size of such measurement data sets is typically quite large. To reduce the computational requirement associated with the inversion of large amount of temporal data, the hydraulic head hydrographs are used to compute the temporal moments of the hydraulic head at each observation well j, in particular, the zeroth-temporal, m_0^j, and the first-temporal moment, m_1^j. Procedures followed to calculate the temporal moments of the measured hydraulic head are discussed in Sect. 2.1. These temporal moments are treated as observations.

The effect of the spatial variability of the aquifer hydraulic parameters, namely the hydraulic conductivity, K, and specific elastic storage, S_s, on the spatial distribution of the temporal moments of the hydraulic head are achieved by means of moment-generating PDEs, which are discussed in Sect. 2.2. The numerical solution of these equations is also discussed in the same section. Two numerical models, one to predict the zeroth-temporal moment, m_0, and another to predict the first-temporal moment m_1,

HESSD
11, 4163–4208, 2014

are employed to simulate an ensemble of randomly generated realizations of the K and S_s fields. At this point, the forecast temporal moments, obtained by solving the moment-generating PDEs numerically, and the observed moments, computed from transient hydraulic head measurements, are available and can be subsequently utilized by the EnKF to update both K and S_s fields. Finally, the overall inversion algorithm is applied either through a CF scheme or a DF scheme, as discussed in Sect. 2.3.

2.1 Estimation of temporal moments of measured hydraulic head

In pumping tests, data may be recorded with high temporal frequency of measurements or even continuously in time. Assimilating such a large amount of transient data using a Kalman filter (Kalman, 1960) scheme is computationally prohibitive and impractical (Evensen, 2009). Time series analyses allow us to shrink the hydrographs of hydraulic head data into low order temporal moments, which are related to aquifer hydraulic properties through moment-generating partial differential equations. To illustrate, assume that an aquifer system is stressed by a well with a time dependent flow rate $Q(t)$ resulting in transient change in hydraulic head $h(\boldsymbol{x};t)$, where the vector $\boldsymbol{x}$ includes the coordinates of the location of an observation well, and t represents time. For linear systems, $h(\boldsymbol{x};t)$ can be expressed as a function of $Q(t)$ through a convolution integral (Von Asmuth and Maas, 2001; Li et al., 2005; Bakker et al., 2008; Olsthoorn, 2008; Von Asmuth et al., 2008):

$$h(\boldsymbol{x};t) = \int_0^t Q(\tau)\theta(\boldsymbol{x};t-\tau)\mathrm{d}\tau \tag{1}$$

where $\theta(\boldsymbol{x};t-\tau)$ is the Impulse Response Function (IRF), which is the response of the aquifer at location $\boldsymbol{x}$ and time t to a unit flow rate at the well at time τ. Accordingly, the objective of time series analysis is to obtain the IRF for every stress source and at each observation well. A possible approach to achieve this is by fitting a parametric

HESSD
11, 4163–4208, 2014

function to represent the IRF for each stress source at each observation well (Von-Asmuth et al., 2002; Bakker et al., 2008). Consequently, the obtained IRF function can be used to calculate the kth temporal moment as follows:

$$m_k(\boldsymbol{x}) = \int_0^\infty t^k \theta(\boldsymbol{x};t)\mathrm{d}t \tag{2}$$

Alternatively, Li et al. (2005) proposed the following equations for calculating the zeroth moment, m_0, and the first moment, m_1, of the hydraulic head resulting from a constant continuous extraction rate Q:

$$m_0(\boldsymbol{x}) = \frac{h(\boldsymbol{x};0) - h(\boldsymbol{x};\infty)}{Q} \tag{3}$$

$$m_1(\boldsymbol{x}) = \frac{\int_0^\infty [h(\boldsymbol{x};t) - h(\boldsymbol{x};\infty)]\mathrm{d}t}{Q} \tag{4}$$

where $h(\boldsymbol{x};0)$ and $h(\boldsymbol{x};\infty)$ represent, respectively, the initial and the steady state hydraulic heads at location $\boldsymbol{x}$. Using Eqs. (3) and (4), the observed zeroth-temporal moment and the first-temporal moment are computed at all observation wells and for each pumping test. In symbolic form, the observed moments from each pumping test can be denoted as $m_{0,ij}$ and $m_{1,ij}$ ($i \in \{1,2,...,N_\mathrm{p}\}$; $j \in \{1,2,...,N_\mathrm{o}\}$). At this point, the transient-hydraulic head large dataset at each observation well is shrunk into the two values $m_{0,ij}$ and $m_{1,ij}$. In the following sections, the numerical simulation of temporal moments is presented.

2.2 Moment generating equations

Transient groundwater flow in a saturated heterogeneous porous medium is governed by the PDE:

$$\nabla[\boldsymbol{K}(\boldsymbol{x})\nabla h] + f(\boldsymbol{x};t) = S_\mathrm{s}(\boldsymbol{x})\frac{\partial h}{\partial t} \tag{5}$$

HESSD
11, 4163–4208, 2014

where ∇ is the differential operator, $\boldsymbol{K}$ is the hydraulic conductivity tensor, S_s is the specific elastic storage, and f represents generic source/sink terms. PDE (Eq. 5) may be solved by imposing Dirichlet boundary conditions $h(\boldsymbol{x};t) = h_D(\boldsymbol{x};t)$ at a prescribed portion of the domain boundary Γ_D, Neumann boundary conditions $K(\boldsymbol{x})\nabla h(\boldsymbol{x};t) = q_N(\boldsymbol{x};t)$ at another portion of the domain boundary Γ_N, and initial boundary conditions $h(\boldsymbol{x};0) = h_0(\boldsymbol{x})$ throughout the domain.

For a unit impulse extraction, the kth temporal moment, m_k, of the hydraulic head drawdown, $s(\boldsymbol{x};t) = h(\boldsymbol{x};0) - h(\boldsymbol{x};t)$, might be computed by multiplying Eq. (5) by t^k and integrating over the time interval $[0, +\infty)$. The resulting moment-generating equation is (Li et al., 2005; Yin and Illman, 2009) the following:

$$\nabla[\boldsymbol{K}(\boldsymbol{x})\nabla m_k] + \delta_k(\boldsymbol{x}_w) + kS_s(\boldsymbol{x})m_{k-1} = 0 \tag{6}$$

where $\delta_k(x_w)$ is equal to unity if $k = 0$ and equal to zero if $k > 0$. Similarly, the boundary conditions of the temporal-moment equation are expressed as $m_k(\boldsymbol{x}) = 0$ for the Dirichlet boundary Γ_D and $K(\boldsymbol{x})\nabla m_k = 0$ for the Neumann boundary Γ_N.

Because the observations (Sect. 2.1) consist of the zeroth and the first temporal moments, the simulation of only m_0 and m_1 is sought. These moments can be obtained by solving numerically the following two PDEs:

$$\nabla[\boldsymbol{K}(\boldsymbol{x})\nabla m_0] + 1(\boldsymbol{x}_w)=0 \tag{7}$$

$$\nabla[\boldsymbol{K}(\boldsymbol{x})\nabla m_1] + S_s(\boldsymbol{x})m_0=0 \tag{8}$$

Equation (7) is equivalent to a steady-state flow problem characterized by a unit extraction rate, denoted as $1(\boldsymbol{x}_w)$, at well location $\boldsymbol{x}_w$. Equation (8) is equivalent to a steady-state flow problem with a forcing term constituted by a spatially variable recharge equal to $S_s(\boldsymbol{x})m_0$. Both Eqs. (7) and (8) can be solved using a common groundwater flow simulator, such as the well-known finite-difference model MODFLOW2000 (Harbaugh et al., 2000).

HESSD
11, 4163–4208, 2014

Hydraulic tomography data fusion

A. H. Alzraiee et. al

2.3 Inversion approaches

This section presents the approaches adopted here to invert the temporal moments in order to characterize the spatial distributions of K and S_s. Using a Bayesian framework to pose the inversion problem, the vector of system parameters, $\boldsymbol{\phi}$, can be updated in light of newly collected data $\boldsymbol{m}$ as follows:

$$p(\boldsymbol{\phi}|\boldsymbol{m},\boldsymbol{I}) = \frac{p(\boldsymbol{m}|\boldsymbol{\phi})p(\boldsymbol{\phi},\boldsymbol{I})}{p(\boldsymbol{m}|\boldsymbol{I})} \tag{9}$$

where $p(\boldsymbol{\phi}|\boldsymbol{m},\boldsymbol{I})$ is the posterior probability distribution function (PDF) of $\boldsymbol{\phi}$ given the measurements $\boldsymbol{m}$ and the generic "prior" information $\boldsymbol{I}$; $p(\boldsymbol{m}|\boldsymbol{\phi})$ is the likelihood PDF, that is, the probability of the measurements $\boldsymbol{m}$ conditional to the parameters $\boldsymbol{\phi}$; $p(\boldsymbol{\phi},\boldsymbol{I})$ is the prior PDF of $\boldsymbol{\phi}$; and $p(\boldsymbol{m}|\boldsymbol{I})$ is a normalization term. An exact solution to Eq. (9) can be obtained if the measurements $\boldsymbol{m}$ are related to the parameters $\boldsymbol{\phi}$ through a linear relationship, and when all PDFs in Eq. (9) are Gaussian. This solution is widely known as the Kalman Filter (KF) (Kalman, 1960). With the KF, the data assimilation of state follows a two-stage forecast-update process. In the forecast stage, a forward in time prediction of the current state, along with its error covariance is first made. The forecast state is then updated as field measurements become available.

In addition to being limited to Gaussian linear systems, the KF is computationally expensive when applied to large scale problems. Evensen (1994) expanded the applicability of the KF to nonlinear systems through the EnKF. Within the EnKF, the prior PDFs of the system states are approximated using an ensemble of realizations that characterize the prior uncertainty in the system parameters and states.

2.3.1 Forecast of parameters and system states

From the perspective of subsurface flow, the major parameters that typically characterize a groundwater system are the hydraulic conductivity, K, and the specific storage, S_s. These parameters are inherently heterogeneous and cannot be

HESSD
11, 4163–4208, 2014

determined uniquely using a finite set of measurements. Therefore, it is convenient to describe these parameters using a geostatistical conceptual model (Matheron, 1962; Isaaks and Srivastava, 1990; Cressie, 1993; Diggle and Ribeiro, 2007), according to which an heterogeneous field is modelled as a spatially-distributed random process, characterized by a trend model and a covariance model. In this study, we assume the log-transformed parameters $Y = \ln(K)$ and $Z = \ln(S_s)$ to fit to two independent isotropic and stationary (with no trend) Gaussian processes (de Marsily, 1986), with prescribed covariance models $C_{YY}(d;\sigma_Y^2;\lambda_Y)$ and $C_{ZZ}(d;\sigma_Z^2;\lambda_Z)$, respectively. The scalar d represents the distance between any two points. The parameters σ^2 and λ represent the correlation length and the variance of each random process. The stationary means of the two fields are denoted as μ_Y and μ_Z. A spherical covariance function is assumed for both C_{YY} and C_{ZZ} (Deutsch and Journel, 1997).

Using these geostatistical models, it is possible to generate an ensemble of N_{ens} equally likely realizations for both Y and Z. The ensemble of the natural logarithm of K is obtained as $\mathbf{Y} = [\mathbf{Y}_1, ..., \mathbf{Y}_{N_{\text{ens}}}]$, where $\mathbf{Y}_k \in \mathrm{R}^{n\times 1}$ $(k \in \{1,2,\ldots,N_{\text{ens}}\})$ is a realization of Y, and n is the number of cells of the finite-difference grid adopted to discretize the aquifer domain. The ensemble of the natural logarithm of S_s, $\mathbf{Z} \in \mathrm{R}^{n\times N_{\text{ens}}}$ is generated in a similar fashion. The resulting ensembles can be seen as discrete approximations of the forecast, or prior, joint PDFs of Y or Z. The sequential Gaussian simulation algorithm SGSIM (Deutsch and Journel, 1997) is used herein to generate the random fields of Y and Z.

In the forecast stage, Eqs. (7) and (8) are solved numerically to predict the system states, that is, the temporal moments, in each pumping test. Each realization $\boldsymbol{Y}_k$ in the ensemble $\mathbf{Y}$ is numerically simulated using Eq. (7) to obtain $\boldsymbol{m}_{0,k} \in \mathrm{R}^{n\times 1}$, a vector including the spatial distribution of the zeroth moment at the cells of the finite-difference grid. Next, $\boldsymbol{m}_{0,k}$ and the parameters $\mathbf{Y}_k$ and Z_k are used to compute the first moment vector $\boldsymbol{m}_{1,k} \in \mathrm{R}^{n\times 1}$ by solving Eq. (8). Therefore, all realizations of states $\boldsymbol{m}_{0,k}$ and $\boldsymbol{m}_{1,k}$ $(k \in \{1,2,\ldots,N_{\text{ens}}\})$ can be assembled into the $n \times N_{\text{ens}}$ matrices $\mathbf{M_0} = [\boldsymbol{m}_{0,1}, ..., \boldsymbol{m}_{0,N_{\text{ens}}}]$ and $\mathbf{M_1} = [\boldsymbol{m}_{1,1}, ..., \boldsymbol{m}_{1,N_{\text{ens}}}]$, respectively. To proceed to the update stage, we propose two alternatives: CF and DF. Schematic diagrams of the two methods are provided in Fig. 1.

2.3.2 Parameter estimation by centralized fusion

In the CF scheme (Fig. 1b), forecast ensembles obtained from simulating independent pumping tests are augmented into a single global forecast matrix $\mathbf{X}_f^Y = [\mathbf{Y}, \mathbf{M}_0^1, \ldots, \mathbf{M}_0^{N_p}]^T$, where $\mathbf{M}_0^i$ represents the zeroth-moment ensemble for the ith pumping test ($i \in \{1, 2, \ldots, N_p\}$). Note that the matrix $\mathbf{X}_f^Y$ has size $(N_p + 1)n \times N_{ens}$. As a matter of fact, there are several possibilities to assemble the forecast matrix, some of which are listed in Table 1. Formulations A, B, and C provide alternatives for forming $\mathbf{X}_f^Y$ in order to estimate the Y field, whereas formulations D and E address possible alternatives for estimating the Z field. In Sect. 4 we investigate the implications of employing different formulations of the forecast matrix. Here, we focus exclusively on formulation A in Table 1 to illustrate the CF procedure.

From the augmented state-parameter forecast matrix $\mathbf{X}_f^Y$, the global prior covariance matrix $\mathbf{P}_f^Y \in \mathrm{R}^{(N_p+1)n \times (N_p+1)n}$ can be approximated as

$$\mathbf{P}_f^Y = \frac{\left(\mathbf{X}_f^y - \bar{\mathbf{X}}_f^y\right) \times \left(\mathbf{X}_\mathrm{f}^y - \bar{\mathbf{X}}_f^y\right)^T}{N_{ens} - 1} \tag{10}$$

where $\bar{\mathbf{X}}_f^Y$ is the prior ensemble mean matrix, calculated as $\bar{\mathbf{X}}_f^Y = \mathbf{X}_f^\mathrm{Y} \cdot \mathbf{1}_{N_{ens}}$ and $\mathbf{1}_{N_{ens}} \in \mathrm{R}^{N_{ens} \times (N_p+1)n}$ is a matrix with all elements equal to $1/N_{ens}$.

To facilitate the assimilation procedure, measurements collected from N_0 observations wells and N_p pumping tests are vertically concatenated in a single vector. Therefore, the vector of measurements for the zeroth-moment can be denoted as $\boldsymbol{d}_0 = [m_{0,i,j}] \in \mathrm{R}^{N_p N_0 \times 1}$, where i is the pumping test index, and j is the observation well index.

HESSD
11, 4163–4208, 2014
Hydraulic tomography data fusion
A. H. Alzraiee et. al

Following an EnKF-like procedure, the measurements $\boldsymbol{d}_0$ are assimilated to update both systems states and parameters. Therefore, the update state-parameter matrix, $\mathbf{X}_u^Y \in \mathrm{R}^{(N_p+1)n \times N_{ens}}$ and the update covariance matrix, $\mathbf{P}_u^Y \in \mathrm{R}^{(N_p+1)n \times (N_p+1)n}$ can be expressed as follows:

$$\mathbf{X}_u^Y = \mathbf{X}_f^Y + \boldsymbol{K} \cdot \left(\mathbf{D}_0 - \mathbf{H} \cdot \mathbf{X}_f^Y\right) \tag{11}$$

$$\mathbf{P}_u^Y = (\mathbf{I} - \boldsymbol{K} \cdot \mathbf{H}) \cdot \mathbf{P}_f^Y \cdot (\mathbf{I} - \boldsymbol{K} \cdot \mathbf{H})^T + \boldsymbol{K} \cdot \mathbf{R} \cdot \boldsymbol{K}^T \tag{12}$$

where $\mathbf{D}_0 \in \mathrm{R}^{N_p N_0 \times N_{ens}}$ is a matrix whose columns are obtained by perturbing the measurement vector $\boldsymbol{d}_0$ with a Gaussian zero mean noise, characterized by the error covariance matrix $\mathbf{R} \in \mathrm{R}^{N_p N_0 \times N_p N_0}$; $\mathbf{H} \in \mathrm{R}^{N_p N_o \times (N_p+1)n}$ is a matrix that maps each measurement to its location in the finite-difference grid and to its corresponding pumping test. The matrix $\boldsymbol{K} \in \mathrm{R}^{(N_p+1)n \times N_p N_o}$ is called "Kalman gain", and is computed as:

$$\boldsymbol{K} = \mathbf{P}_f^Y \cdot \mathbf{H}^T \cdot \left(\mathbf{H} \cdot \mathbf{P}_f^Y \cdot \mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{13}$$

In the context of parameter estimation, we are interested exclusively in updated parameters. Consequently, the ensemble of log-transformed hydraulic conductivity fields is extracted from the updated state-parameter matrix (Eq. 11) as $\mathbf{Y}_u = \mathbf{X}_u^Y(1:n, 1:N_{ens})$. The posterior ensemble mean of the hydraulic conductivity is thus computed as $\hat{\mathbf{Y}} = \mathbf{Y}_u \cdot \hat{\mathbf{1}}_{N_{ens}}$, where $\hat{\mathbf{1}}_{N_{ens}}$ is a $N_{ens} \times 1$-vector in which all elements are equal to $1/N_{ens}$.

A procedure similar to that described above to obtain the ensemble $\mathbf{Y}_u$ by assimilating the zeroth moment of hydraulic head measurements (Eq. 3) may be devised to derive the specific elastic storage ensemble $\mathbf{Z}_u$, using observations of the first moment of hydraulic head (Eq. 4). The formulations D and E, presented in Table 1, provide two possible methods for assembling the forecast matrix in order to estimate the Z field. Since the first-temporal moment m_1 (Eq. 3) depends on the zeroth temporal moment m_0, as well as the K and S_s fields, the uncertainty on K might affect the

HESSD
11, 4163–4208, 2014

estimation of S_s. To reduce the influence of the uncertainty on K on the estimation of S_s, it is possible, for example, to use the posterior ensemble mean $\hat{\mathbf{Y}}$ to solve Eqs. (7) and (8).

This assimilation strategy is denoted as E in Table 1. In this case, the forecast matrix is expressed as $\mathbf{X}_f^Z = [\mathbf{Z}, \mathbf{M}_1^1, \ldots, \mathbf{M}_1^{N_p}]^T$, where $\mathbf{M}_1^i$ represents the first-moment ensembles for the ith pumping test. $\mathbf{X}_f^Z$ is updated by assimilating the observations of the hydraulic head first-moment, $\boldsymbol{d}_1 = [m_{1,i,j}] \in \mathrm{R}^{N_p N_0 \times 1}$ where i is the pumping test index, and j is the observation well index, and applying equations similar to Eqs. (11) and (12). The ensemble mean of the updated Z is thus computed as $\hat{\mathbf{Z}} = \mathbf{Z}_u \mathbf{1}_{N_{ens}}$. This mean represents the best unbiased estimate of the unknown true parameter. In Sect. 4.1, we show that this approach significantly improves the estimation of S_s.

2.3.3 Parameter estimation by decentralized fusion

For conciseness, this section describes the DF algorithm to estimate the K field only. The estimation of S_s field is achieved by applying an analogous procedure.

In the DF approach (Fig. 1b), the data from each pumping test are assimilated separately using a "local" EnKF. The application of the EnKF to each of the N_p pumping tests produces multiple estimates of the hydraulic properties of the aquifer, which are characterized by the means of the posterior ensembles, $\hat{\mathbf{Y}}_u^1, \ldots, \hat{\mathbf{Y}}_u^{N_p}$, and their corresponding posterior covariances, $\mathbf{P}_u^{Y,1}, \ldots, \mathbf{P}_u^{Y,N_p}$. The objective of the DF algorithm is to merge these estimates and produce an integrated global estimate $\tilde{\mathbf{Y}}$ of the parameters. The multiple estimates are fused using the GMF (Bar-Shalom and Campo, 1986; Shin et al., 2006):

$$\tilde{\mathbf{Y}} = \mathbf{W}^{\mathbf{T}} \cdot \hat{\mathbf{Y}}_u^{1:N_p} \tag{14}$$

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

11, 4163–4208, 2014

Hydraulic tomography data fusion

A. H. Alzraiee et. al

where the matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{N_P}]^T$, of size $nN_P \times n$, includes the $n \times n$ weight matrices $\mathbf{w}_i$ $(i = 1, 2, ..., N_\mathrm{p})$ and the $N_\mathrm{p} n \times 1$ vector $\hat{\mathbf{Y}}_\mathrm{u}^{1:N_\mathrm{p}}$ is assembled by vertical concatenation of the means of the posterior ensembles $\hat{\mathbf{Y}}_\mathrm{u}^1, \ldots, \hat{\mathbf{Y}}_\mathrm{u}^{N_\mathrm{p}}$.

The weight matrices in Eq. (14) are given by the solution of the optimization problem:

$$\mathbf{W} = \min_{\mathbf{W}} \|\mathbf{Y} - \tilde{\mathbf{Y}}\|_2 \tag{15}$$

where $\|\cdot\|_2$ represents the Euclidean norm operator. In addition, Eq. (15) is subject to a constraint required to obtain a "best linear unbiased estimate" (BLUE) of Y, which is expressed by the following set of linear equations:

$$\mathbf{I}_{n,N_\mathrm{p}} \cdot \mathbf{W} = \mathbf{I}_n \tag{16}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, and $\mathbf{I}_{n,N_\mathrm{p}}$ is the $n \times nN_\mathrm{p}$ matrix formed by horizontal concatenation of $\mathbf{I}_n$ for N_p times.

The solution to Eq. (15) is obtained by least-square minimization, which, together with Eq. (16), yields the following linear sets of equations:

$$\mathbf{C} \cdot \mathbf{W} = \mathbf{B} \tag{17}$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_{1,1} & \ldots & \mathbf{c}_{1,N_\mathrm{p}} \\ & \vdots & \\ \mathbf{c}_{N_\mathrm{p}-1,1} & \ldots & \mathbf{c}_{N_\mathrm{p}-1,N_\mathrm{p}} \\ \mathbf{I}_n & \ldots & \mathbf{I}_n \end{bmatrix} \tag{18a}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{0}_n \\ \vdots \\ \mathbf{0}_n \\ \mathbf{I}_n \end{bmatrix} \tag{18b}$$

HESSD
11, 4163–4208, 2014

Hydraulic tomography data fusion

A. H. Alzraiee et. al

Matrix $\mathbf{C}$ has size $nN_\mathrm{p} \times nN_\mathrm{p}$, whereas matrix $\mathbf{B}$ has size $nN_\mathrm{p} \times n$. In matrix $\mathbf{B}$ (Eq. (18a), $\mathbf{0}_1$ is a zero matrix with size $n \times n$. The generic term $\mathbf{c}_{i,j}$ in matrix $\mathbf{C}$ (Eq. (18a) is given by:

$$\mathbf{c}_{i,j} = \mathbf{P}_\mathrm{u}^{Y,i,j} - \mathbf{P}_\mathrm{u}^{Y,i,N_\mathrm{p}} \tag{19}$$

where $\mathbf{P}_\mathrm{u}^{Y,i,j}$ is the updated cross-covariance matrix for the Y fields estimated from the assimilation of data corresponding to pumping tests i and j, which is calculated as:

$$\mathbf{P}_\mathrm{u}^{Y,i,j} = \left(\mathbf{Y}_u^i - \hat{\mathbf{Y}}_\mathrm{u}^i\right)\left(\mathbf{Y}_u^j - \hat{\mathbf{Y}}_\mathrm{u}^j\right)^T / (N_\mathrm{ens} - 1) \tag{20}$$

From Eq. (17), $\mathbf{W}$ is obtained as:

$$\mathbf{W} = \mathbf{C}^{-1} \cdot \mathbf{B} \tag{21}$$

Once the weight matrix $\mathbf{W}$ is calculated, it is substituted in Eq. (14) to provide the estimate $\tilde{\mathbf{Y}}$. The posterior covariance of $\tilde{\mathbf{Y}}$ can be computed as (Shin et al., 2006):

$$\tilde{\mathbf{P}} = \mathbf{W}^T \cdot \mathbf{P} \cdot \mathbf{W} \tag{22}$$

where $\mathbf{P}$ is a $nN_\mathrm{p} \times nN_\mathrm{p}$ matrix formed by the covariance matrices $\mathbf{P}_\mathrm{u}^{Y,i,j}$ $(i,j = 1,\ldots,N_\mathrm{p})$.

The inversion of the matrix $\mathbf{C}$ in Eq. (21) constitutes the most intensive part of the GMF. In HT, it is typically required to estimate hydrogeological parameters at high resolution, which often renders the GMF approach computationally very intensive. To circumvent this obstacle, we propose a novel localized fusion algorithm. In essence, instead of computing Eq. (21) for all the cells of the domain at once, the fused estimate at any given cell is computed by considering only those cells within a specified radius around it. The implicit assumption behind this method is that neighboring cells will have the majority of influence on the estimation. Indicating as n' $(< n)$ the number of grid cells within a specified distance from the cell of interest, the resulting size for the "local" matrices in Eq. (21) is: $n' \times n'N_\mathrm{p}$ for $\mathbf{B}$, and $n'N_\mathrm{p} \times n'N_\mathrm{p}$ for $\mathbf{C}$. The GMF localization is meant to improve the computational efficiency in two ways. First, the inversion of matrices $\mathbf{C}$ of smaller size is less CPU intensive; second, the fusion algorithm can be directly parallelized on multi-core processors.

2.4 Options for data fusion formulation

The forecast matrix $\mathbf{X}_f$ can be assembled according to different formulations of the data fusion problem. Table 1 shows a list of the formulations investigated herein.

Formulations A, B, and C seek the estimation of Y field. Formulation A consists of assimilating measurements of the zeroth temporal moment m_0 (Eq. 3), with the forecast model given by numerical solution of the PDE (Eq. 7). Formulation B consists of assimilating measurements of the first temporal moment m_1 (Eq. 4), with the forecast model given by numerical solution of the PDE (Eq. 8), in which the K forecast ensemble and its corresponding m_0 forecast ensemble, in turn obtained from the numerical solution of PDE (Eq. 7), are used. In Formulation C, measurements of both m_0 and m_1 are assimilated, and the forecast model is obtained by solving Eqs. (7) and (8) combined.

Formulations D and E aim at estimating the Z field. In the estimation of S_s, it is possible to find a one-to-one correspondence between S_s and m_1 based on Eq. (8) if the K field and its corresponding m_0 field are known. However, since the K field is unknown, one can choose instead to represent it using, for example, its forecast ensemble $\mathbf{Y}$, or a best unbiased estimate, calculated as the mean of the posterior ensemble $\hat{\mathbf{Y}}$ obtained in formulation A. These alternatives are investigated in Formulations D and E. In both instances, measurements of m_1 are assimilated, and the forecast model consists of the numerical solution of PDE (Eq. 8). In Formulation D, the K forecast ensemble and its corresponding m_0 forecast ensemble, obtained from the numerical solution of PDE (Eq. 7), are used. Instead, in Formulations E, the posterior mean of Y, as estimated in Formulation A, and its corresponding m_0 distribution, obtained from the numerical solution PDE (Eq. 7), are used.

2.5 Estimation of geostatistical parameters by fusion of HT data

The geostatistical parameters characterizing the spatial distributions of Y and Z are rarely known before field data are collected and analyzed. Here, we propose a modification to the inversion schemes presented above in order to estimate not only the Y and Z fields, but also their geostatistical parameters, denoted respectively as $\mu_Y, \sigma_Y, \lambda_Y$ and $\mu_Z, \sigma_Z, \lambda_Z$ in Sect. 2.3.1. The methodology stems from the following procedure:

1. An ensemble of geostatistical parameters for the Y field is generated from three hypothetical prior non-informative uniform PDFs: $\mu_Y \in U(\mu_{Y\,\text{min}}, \mu_{Y\,\text{max}})$, $\sigma_Y \in U(\sigma_{Y\,\text{min}}, \sigma_{Y\,\text{max}})$, and $\lambda_Y \in U(\lambda_{Y\,\text{min}}, \lambda_{Y\,\text{max}})$. The three geostatistical parameters are assumed to be uncorrelated and thus are generated independently. The ensemble of all generated realizations is denoted as $\mathbf{\Theta}_f^Y = [\boldsymbol{\theta}_{Y,1}, \ldots, \boldsymbol{\theta}_{Y,N_\text{ens}}]^T$, where the generic vector $\boldsymbol{\theta}_{Y,k}$ represents the realization $[\mu_{Y,k}, \sigma_{Y,k}, \lambda_{Y,k}]^T$ $(k \in \{1, \ldots, N_\text{ens}\})$.
2. Each realization $\boldsymbol{\theta}_{Y,k}$ is used to generate a single realization of the Y field, which is in turn used to forecast the zeroth moment m_0 spatial distribution by solving PDE (Eq. 7). Next, the state-parameter forecast matrix of formulation A (see Sect. 2.3.2) is augmented as follows: $\mathbf{X}_f^Y = [\mathbf{Y}, \mathbf{M}_0^1, \ldots, \mathbf{M}_0^{N_\text{p}}, \mathbf{\Theta}_f^Y]^T$.
3. The forecast matrix $\mathbf{X}_f^Y$ is updated (Eqs. 11–13) by assimilating observation of the zeroth moment m_0, which yields an updated ensemble of geostatistical parameters, denoted as $\mathbf{\Theta}_\text{u}^Y$.
4. The updated geostatistical parameter ensemble $\mathbf{\Theta}_\text{u}^Y$ is used to generate a new improved prior ensemble of realizations $\mathbf{Y}$. Finally, formulation A (Sect. 2.4, Table 1) is applied to estimate the Y field by assimilating measurements of the zeroth moment m_0.

The uncertainty on the geostatistical parameters of the Z field can be reduced following a procedure similar to that outlined above for the Y field. In this case,

HESSD
11, 4163–4208, 2014

however, the initial ensemble of geostatistical parameters $\boldsymbol{\Theta}_f^Z$ is generated from three prior non-informative uniform PDFs: $\mu_Z \in U(\mu_{Z\,\min}, \mu_{Z\,\max})$, $\sigma_Z \in U(\sigma_{Z\,\min}, \sigma_{Z\,\max})$, and $\lambda_Z \in U(\lambda_{Z\,\min}, \lambda_{Z\,\max})$. The update of $\boldsymbol{\Theta}_f^Z$ may attained by assembling the augmented forecast matrix $\mathbf{X}_f^Z = [\mathbf{Z}, \mathbf{M}_1^1, \ldots, \mathbf{M}_1^{N_\mathrm{p}}, \boldsymbol{\Theta}_f^Z]^T$ and assimilating observations of the first moment m_1 as in formulation E (Sect. 2.4, Table 1).

3 Numerical experiments

3.1 Model setup

The testing of the inversion schemes proposed in this work is based on a number of hypothetical two-dimensional cases. The method is, however, directly applicable to three-dimensional problems. We consider a two-dimensional horizontal 1 km × 1 km, 10 m thick confined aquifer, discretized into 10 000 cells (100 gridblocks along the x–y coordinate directions, and a single gridblock along the z direction). Table 2 provides a detailed description of data regarding the aquifer model.

The aquifer is subject to constant-head boundary conditions on the left and right edges of the domain, at which the hydraulic head h is set equal to 45 m. Anywhere else, no-flow boundary conditions are imposed. The "true" K and S_s fields in the aquifer are assumed to fit to the geostatistical models introduced in Sect. 2.3.1 and generated synthetically using the sequential Gaussian simulation algorithm SGSIM (Deutsch and Journel, 1997), with the geostatistical parameters $\mu_Y = 1.5$ ln m day^{-1}), $\sigma_Y = 1$ ln m day^{-1}, $\lambda_Y = 350$ m, $\mu_Z = -10$ ln m^{-1}, $\sigma_Z = 1$ ln m^{-1}, and $\lambda_Z = 350$ m (Table 2). These two fields are used in five MODFLOW2000 (Harbaugh et al., 2000) simulations to reproduce the aquifer response to five separate pumping tests, conducted from the locations and with the pumping rates specified in Table 2.

The duration of these hypothetical pumping tests is 10 days. The output of each simulation provides the reference system from which collection of hydraulic head data is simulated. Hydraulic head observations are recorded from a network of 36 monitoring wells, whose locations are depicted in Fig. 2.

Table 3 describes the numerical experiments carried out to evaluate and compare the performance of the CF and DF schemes. Experiment set I investigates the performance of different formulations of the forecast matrix, as listed in Table 1, using the CF approach. Experiment set II is similar to experiment set I, but the DF approach is used instead. In both experiment sets I and II (Table 3), the parameters characterizing the geostatistical models of Y and Z are assumed to be known as prior information and equal to those of the "true" fields given in Table 2. The prior ensembles of Y and Z realizations are generated using the Sequential Gaussian Simulation algorithm SGSIM (Deutsch and Journel, 1997). These ensembles are assumed to be uncorrelated.

In experiment set III, we consider the case in which the geostatistical model is unknown, and investigate the capability of the CF approach to reduce the uncertainty in the geostatistical parameters of Y and Z. In addition, we address the impact of the geostatistical model uncertainty on the quality of the Y field estimation using CF. Finally, in experiment set IV we apply the CF approach in a general case to jointly estimate the geostatistical parameters and the Y and Z fields.

In the experiment sets I–III the size N_{ens} of the ensemble is 200, whereas in the experiment IV N_{ens} is set to 1000. In all tests, the observed hydraulic heads associated with each pumping test are obtained at the 36 observation wells shown in Fig. 2. The temporal moments at each observation well are estimated using Eqs. (3) and (4). Since the temporal moments are assumed to be the measured quantities, their measurement error is assumed to fit to a normal distribution with zero mean and standard deviation equal to the corresponding forecast standard deviation multiplied by 0.01.

3.2 Performance metrics

The performances of the fusion methods may be evaluated qualitatively by visual comparison of the maps of the estimated hydraulic parameters, represented by the average distributions $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ (Sect. 2.3.2), with the corresponding maps of the "true" reference fields. In addition, a quantitative evaluation of these performances is achieved using the following statistics: the mean absolute error L_1, the root mean square error L_2, the mean error μ_e, and the correlation coefficient r. L_1 is computed as:

$$L_1 = \frac{1}{n}\sum_{i=1}^{n} |\phi_{\text{true}}(i) - \hat{\phi}(i)| \tag{23}$$

where $\phi_{\text{true}}(i)$ is the value of "true" parameter at the grid cell i and $\hat{\phi}(i)$ is the corresponding value of estimated parameter. L_2 is computed as:

$$L_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n} [\phi_{\text{true}}(i) - \hat{\phi}(i)]^2} \tag{24}$$

The coefficient r quantifies the correlation between the "true" and the estimated fields and is calculated as:

$$r = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} [\phi_{\text{true}}(i) - \bar{\phi}_{\text{true}}][\hat{\phi}(j) - \bar{\phi}]}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} [\phi_{\text{true}}(i) - \bar{\phi}_{\text{true}}]^2[\hat{\phi}(j) - \bar{\phi}]^2}} \tag{25}$$

where $\bar{\phi}_{\text{true}}$ and $\bar{\phi}$ are the overall means of the true and the estimated parameter fields, respectively. Values of r range between 1 and -1, with $r = 1$ indicating perfect positive correlation (i.e., the estimated and the "true" fields are identical), $r = 0$ indicating no correlation, and $r = -1$ indicating perfect negative correlation. Finally, the error mean μ_e is obtained as:

$$\mu_e = \frac{1}{n}\sum_{i=1}^{n} [\phi_{\text{true}}(i) - \hat{\phi}(i)] \tag{26}$$

HESSD
11, 4163–4208, 2014

and is meant to provide a measure of the biasedness of the estimate. Values of μ_e close to zero indicate an unbiased estimate.

4 Results and discussions

4.1 Centralized fusion of HT data

In this section, the performance of each of the forecast formulations given in Table 1 is evaluated using the CF scheme (Fig. 1). The tests presented constitute the experiment set I described in Table 3. The results of these tests are summarized in Table 4, which reports values of the four performance statistics, L_1, L_2, r and μ_e (Eqs. 23–26) for the formulation schemes A–E. As explained in Sect. 2.4, these formulations seek the estimation of the Y field. The comparison of the metrics L_1, L_2, and r reported in Table 4 reveals that the CF scheme performs significantly better under formulation A than under formulation C. In turn, formulation B is slightly less effective than formulation C.

These results find an explanation in that with formulation A the Y field is estimated by assimilating m_0 data only, whereas with formulation C the Y field is estimated by assimilating both m_0 and m_1 data. While in formulation A the heterogeneity of the Y field affects directly the spatial variability of m_0 via PDE (Eq. 7), in formulation C such heterogeneity influences both the m_0 and m_1 spatial distributions via PDEs (Eqs. 7 and 8). In addition, the spatial variability of m_1 depends not only on Y but also on Z. This makes the estimation of Y using PDEs (Eqs. 7 and 8) less effective given the added uncertainty in Z.

In the case of formulation B, the performance of the CF scheme is even lower than with formulation C since only m_1 data are assimilated and thus the impact of the added uncertainty in Z is inevitably more pronounced. In Table 4, it is worth observing that for all formulations A–C, the mean error μ_e is very low, on the order of 10^{-5}, which provides substantial evidence of the unbiasedness of the estimates obtained by CF.

HESSD
11, 4163–4208, 2014

Figure 3a and c present the maps of the "true" reference field $\mathbf{Y}_{\text{true}}$ and the average of the update ensemble $\hat{\mathbf{Y}}$ obtained using the forecast formulation A, respectively. The similarity between the two maps is remarkable. Figure 3e shows a scatter plot obtained using the components of $\mathbf{Y}_{\text{true}}$ on the x axis, and the corresponding components of $\hat{\mathbf{Y}}$ on the y axis. The data points in this plot tend to gather along the identity line, which provides a further visual proof of the satisfactory performance of the CF scheme.

In formulations D and E (Table 1), the estimation of Z field is sought using the CF approach. The values of the metrics L_1, L_2, and r given in Table 4 indicate that with formulation D the CF scheme performs significantly worse than with formulation E. Indeed, estimating the Z field based exclusively on m_1 data through PDE (Eq. 8) is inevitably affected by the uncertainty on the Y and the m_0 fields, in a fashion very similar to that highlighted above for formulation B. A similar outcome has been observed by other researchers (Yin and Illman, 2009). It is interesting to observe that formulations B and D are substantially the same, although they attempt to estimate different parameters. Thus it is not coincidental that they exhibit the two lowest estimation performances.

Based on the results of formulation B, the estimation of Z may be improved if the uncertainty on the Y and m_0 fields can be reduced. Formulation E (Table 1) stems from the idea of using the best unbiased estimate $\hat{\mathbf{Y}}$ obtained with formulation A, and the corresponding m_0 field calculated by solving the PDE (Eq. 7), within the the forecast model based on PDE (Eq. 8) and assimilate m_1 measurements only, as in formulation D. The values of L_1, L_2, and r shown in Table 4 reveal that this solution allows for a significant improvement in the estimation of the Z field, and the performance of the CF approach becomes comparable with that observed in formulations A-C, when estimating the Y field. Note in Table 4 that with both formulations D and E the CF approach produces negligible values of μ_e (10^{-6}), which demonstrate that the estimates of Z are substantially unbiased.

Figure 3b and d depicts the "true" Z field and that estimated by CF using formulation E, respectively. A comparison between the two maps shows that the CF scheme is

HESSD
11, 4163–4208, 2014

able to capture fairly well the spatial heterogeneity of Z. Figure 3f shows a scatter plot of $\mathbf{Z}_{\text{true}}$ against $\hat{\mathbf{Z}}$. Similar to Fig. 3e, the data are distributed along the identity line, that is, a general agreement between "true" and the estimated Z can be observed. However, Fig. 3f shows that higher and lower values of Z, located on the "tails" of the distribution, are not well identified, which highlights the tendency of the CF scheme to produce smoothed estimates of the Z field.

In the tests presented above, the average CPU time required to calculate the spatial distributions of temporal moments, that is, to solve either of the PDEs (Eqs. 7 and 8) using MODFLOW2000 (Harbaugh et al., 2000) is about 2 s per run. In practice, a forecast simulation with an ensemble size N_{ens} of 200 or 1000 (see Table 1) requires a CPU time of the order of minutes. This is because the moment-generating PDEs (Eqs. 7 and 8) are Poisson-type equations, which are computationally much less intensive to solve than the parabolic PDE (Eq. 5) from which they are derived. In this regard, note the PDE (Eq. 5) is time-dependent, whereas in PDEs (Eqs. 7 and 8) the time variable is eliminated by integration (Eq. 2).

Considering the temporal moments of hydraulic head measurements allows also for a significant reduction of the CPU requirements of data assimilation. In the numerical experiments conducted here, 36 observation wells are used to monitor the hydraulic head during each pump test. In each observation well, 100 temporal measurements are recorded, resulting in 3600 measurements per single pumping test. Since we assumed that five pumping tests are performed to characterize the aquifer, the total number of available hydraulic head data is ($36 \times 100 \times 5$) 18 000. Direct assimilation of transient hydraulic head data using either the EnKF or the Ensemble Smoother (Evensen, 2009) would require, respectively, the inversion (Eq. 13) of a 180×180 matrix for each of the 100 measurement times, or the inversion all-at-once of a $18\,000 \times 18\,000$ matrix. In either situation the computational effort would not be trivial. Instead, by introducing temporal moments, for example when estimating the Y field with formulation A, the data assimilation step involves the inversion of a 180×180 matrix only once.

HESSD
11, 4163–4208, 2014

4.2 Decentralized fusion of HT data

In this section, the DF scheme based on the GMF (Sect. 2.3.3) is employed to estimate aquifer parameters based on the same HT data used in the previous section with the CF scheme. These tests form the experiment set II introduced in Table 3. In this case, the formulations A and E are adopted for the estimation of the Y and the Z fields, respectively, since these have been shown to perform best among those tested in Sect. 4.1. Following the approach outlined in Sect. 2.3.3 to reduce computational intensity, in the calculation of the weight coefficients $\mathbf{W}$ (Eq. 21), for each grid cell, only cells within a radius of 50 m are used in the inversion of the matrix $\mathbf{C}$. The results of numerical tests (not shown here) have suggested that, in this problem, no significant improvement in accuracy is achieved if this radius is increased beyond 50 m, while a significant increase in computational cost is observed.

Figure 4a–e in show the "local" estimates of the Y field obtained using formulation A of the EnKF to assimilate HT data collected separately in each of the five pumping tests (Fig. 2). Figure 4f shows the global estimate of the Y field produced by Eq. (14), and Fig. 4g shows the "true" reference Y field. The similarity between the two maps in Fig. 4f and g indicates that the DF scheme is able to estimate fairly well the spatial distribution of hydraulic conductivity. In Fig. 4h, the scatter plot of $\mathbf{Y}_{\text{true}}$ vs. $\tilde{\mathbf{Y}}$ provides further proof of the good performance of the DF scheme. The resulting correlation coefficient, r, between the two distributions is equal to 0.723, which is less than that obtained by using the CF scheme with formulation A ($r = 0.825$, see formulation A in Table 4).

Figure 5a–e show the estimations of the Z field obtained using formulation E (Table 1) and applying the EnKF separately to each of the five pumping tests (Fig. 2). Figure 5f and g shows the DF global estimate of the Z field and the "true" reference field, respectively. The comparison of the two maps in subpanels f and g indicates the DF scheme is able to capture the main features of the heterogeneity of the Z field. The scatter plot of $\mathbf{Z}_{\text{true}}$ vs. $\tilde{\mathbf{Z}}$ shows that the correlation coefficient r is equal to 0.645,

HESSD
11, 4163–4208, 2014

Hydraulic tomography data fusion

A. H. Alzraiee et. al

which is smaller than that produced by the CF scheme with formulation E ($r = 0.759$, see formulation E in Table 4).

The results presented in Figs. 4 and 5 indicate that, in the joint estimation of Y and Z, the CF scheme consistently outperforms the DF scheme. This can be explained by observing that all of the r coefficients obtained with "local estimations", that is, the application of the EnKF separately to the five pumping tests (panels a–e in Figs. 4 and 5), are smaller than the corresponding coefficients produced by the CF scheme (see formulations A and E in Table 4), which applies the EnKF "globally", that is, to the five pumping tests altogether. Since the GMF (Eq. 14) constitutes in essence a weighted average of the "local" estimates of the Y and Z fields, with weights (Eq. 21) that are inversely related to the corresponding "local" covariances (Shin et al., 2006), it produces fused estimates with a coefficient r that cannot be larger than those associated with the best "local" estimate and, consequently, those obtained with the "global" CF estimate.

However, the DF scheme has an operational advantage over the CF scheme, in that the "raw" transient data are not required to apply fusion. Indeed, only estimates of the field and the covariance are required. Note also when applying the DF scheme to the considered problem, the inversion of the matrix **C** (Eq. 21) would be computationally overwhelming since its size ($nN_\mathrm{p} \times nN_\mathrm{p}$) is equal to 50 000 by 50 000. This application is made possible only by implementing the localized DF described at the end of Sect. 2.3.3. By doing so, the algorithm requires about 40 CPU minutes to complete the calculation without parallelization of the computation. Using a multicore computer would further reduce this time by a factor equal to the number of processors available.

4.3 Effects of uncertainty on geostatistical parameters

In the numerical experiments presented above, the geostatistical parameters of the Y and Z fields are assumed to be priorly known. Since this is not the case in most practical situations, an evaluation of the sensitivity of the HT-based estimates to the uncertainty on geostatistical parameters is necessary. In this section, such

HESSD
11, 4163–4208, 2014

sensitivity is assessed by repeating the estimation using values of the geostatistical parameters obtained by multiplying the "true" parameters by a factor α, where $\alpha \in \{0.4, 0.6, 0.8, 1.2, 1.4, 1.6\}$. When one of the three geostatistical parameters, $\mu_Y, \lambda_Y, \sigma_Y$ or $\mu_Z, \lambda_Z, \sigma_Z$, is varied, the other two are kept constant and equal to their "true" reference value.

These tests constitute the experiment set III described in Table 3. The CF scheme is implemented using formulation A for the estimation of Y, and formulation E for the estimation of Z. Results are analyzed in terms of the correlation coefficients r between the "true" and the estimated parameter distributions as a function of α, as illustrated in Fig. 6a for the Y field, and in Fig. 6b for the Z field.

The profiles in Fig. 6a show that the correlation coefficient r, and thus the quality of the estimation of the Y field are significantly reduced by errors in the selection of the stationary mean μ_Y. On the other hand, errors in the standard deviation σ_Y and in the correlation scale λ_Y have a relatively low impact on the effectiveness of the fusion scheme. Substantially similar observations can be made in Fig. 6b for the geostatistical parameters characterizing the Z field. Note, however, that in the estimation of Z, the correlation coefficient r drops rather quickly for α greater than or less than one, indicating a substantial failure of the inversion scheme. A comparison of the results depicted in Fig. 6a and b indicate that, in general terms, errors in the selection of the stationary mean influence more the estimation of the Z field than the estimation of the Y field.

4.4 Inversion of geostatistical parameters

The sensitivity of the estimation of the Y and Z field to their corresponding means μ_Y and μ_Z motivates the extension of the CF to estimating these geostatistical parameters. For this purpose, the procedure introduced in Sect. 2.5 is adopted. The ensemble size N_{ens} used in these experiments is equal to 1000 (see experiment set IV in Table 3). For the Y field, geostatistical parameters are sampled from the three independent uniform PDFs: $U(\mu_{Y\,\text{min}}, \mu_{Y\,\text{max}}) = U(0,2)$ (ln m day^{-1}); $U(\sigma_{Y\,\text{min}}, \sigma_{Y\,\text{max}}) = U(0.001,2)$ (ln

HESSD
11, 4163–4208, 2014

m day^{-1}); and $U(\lambda_{Y\min}, \lambda_{Y\max}) = U(0, 700)$ (m). Likewise, for the Z field, geostatistical parameters are generated from the independent uniform PDFs: $U(\mu_{Z\min}, \mu_{Z\max}) = U(-10, -5)$ (ln m^{-1}); $U(\sigma_{Z\min}, \sigma_{Z\max}) = U(0.001, 2)$ (ln m^{-1}); and $U(\lambda_{Z\min}, \lambda_{Z\max}) = U(0, 700)$ (m).

Figure 7a–c and d–f show the prior and posterior cumulative distribution functions (CDF) of the geostatistical parameters $\mu_Y, \sigma_Y, \lambda_Y$ and $\mu_Z, \sigma_Z, \lambda_Z$, respectively. In Fig. 7a, the reduction in the uncertainty of the posterior CDF of μ_Y is noticeably greater than that observed for σ_Y in Fig. 7b, and λ_Y in Fig. 7c. This is not surprising since it has been shown in Sect. 4.3 that the CF results are more sensitive to errors in μ_Y than errors in σ_Y and λ_Y.

Similar observations can be made for the CDFs of geostatistical parameters of Z in Fig. 7d–f. However, the reduction in the uncertainty on μ_Z is significantly smaller than the reduction in uncertainty of μ_Y. This is due the propagation of errors in the estimation Y and its geostatistical parameters into the estimation of Z and its geostatistical parameters, as in formulation E.

It is finally important to point out that the assumption of stationary of the Y and Z fields might not be met in practical applications. While this assumption cannot be generally validated, but can be disproved using existing field data (Rubin, 2003), in cases where the Y and Z fields are characterized by unknown spatial trends, it is possible to extend the methods presented above by including the trend function coefficients in the unknowns of the inversion problem.

5 Conclusions

Determination of the spatial variability of aquifers' hydraulic properties at high resolution is vital for studying flow and transport processes. In this work, two approaches were developed and implemented to achieve this goal: centralized fusion (CF) and decentralized fusion (DF). CF utilizes a global EnKF scheme to simultaneously invert data obtained from multiple pumping tests. DF uses the generalized Millman formula

HESSD
11, 4163–4208, 2014

(GMF) to merge together estimates obtained from "local" EnKF applications to each of the pumping tests. The proposed inversion methods assimilated the zeroth and first temporal moments of hydraulic head data collected in monitoring wells, which significantly expedites the stochastic simulation procedures.

The performance of the fusion schemes, measured as the deviation of the estimated field from the "true" reference field, are promising for both inversion schemes. The numerical tests presented in this work show that the CF scheme using the global EnKF consistently outperforms the DF scheme based on the GMF. To optimize the inversion procedures, different formulations of the forecast matrix were investigated, and results indicate that the estimation of the aquifer parameters is significantly affected by the chosen formulation. For instance, the estimation of the specific elastic storage field was significantly improved by using a specific formulation of the forecast matrix based on the assimilation of measurements of the first temporal moment of hydraulic head data, with the posterior mean of hydraulic conductivity obtained with the assimilation of measurements of the zeroth temporal moment.

The analysis of the effect of errors in the geostatistical parameters suggests that the estimation of the spatial distributions of hydraulic conductivity and specific elastic storage are particularly sensitive to errors in the assumed stationary means. The expansion of the proposed inversion methods to estimate the geostatistical parameters of the hydraulic conductivity and specific elastic storage showed that the CF scheme can effectively identify the stationary means of the hydraulic conductivity field and, to a lesser degree, of the specific elastic storage field.

Acknowledgements. Authors would like to express their great appreciation to Colorado State University Libraries Open Access Research and Scholarship Fund for their valuable support to publish this article.

HESSD
11, 4163–4208, 2014

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀◀ ▶▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

References

Bakker, M., Maas, K., and Von-Asmuth, J. R.: Calibration of transient groundwater models using time series analysis and moment matching, Water Resour. Res., 44, W04420, doi:10.1029/2007WR006239, 2008. 4168, 4170, 4171

Bar-Shalom, Y. and Campo, L.: The effect of the common process noise on the two-sensor fused-track covariance, IEEE T. Aero. Elec. Sys., AES-22, 803–805, doi:10.1109/TAES.1986.310815, 1986. 4168, 4177

Berg, S. J. and Illman, W. A.: Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciofluvial aquifer-aquitard system, Water Resour. Res., 47, W10507, doi:10.1029/2011WR010616, 2011. 4166

Bohling, G. C. and Butler, J. J.: Inherent limitations of hydraulic tomography, Ground Water, 48, 809–24, doi:10.1111/j.1745-6584.2010.00757.x, 2010. 4167

Bohling, G. C., Butler, J. J., Zhan, X., and Knoll, M. D.: A field assessment of the value of steady shape hydraulic tomography for characterization of aquifer heterogeneities, Water Resour. Res., 43, W05430, doi:10.1029/2006WR004932, 2007. 4166

Butler Jr., J. J.: The design, performance, and analysis of slug tests, CRC Press. Boca Raton, FL, USA, 1998. 4165

Butler Jr., J. J., McElwee, C. D., and Bohling, G. C.: Pumping tests in networks of multilevel sampling wells: motivation and methodology, Water Resour. Res., 35, 3553, doi:10.1029/1999WR900231, 1999. 4165, 4166

Cardiff, M. and Barrash, W.: 3-D transient hydraulic tomography in unconfined aquifers with fast drainage response, Water Resour. Res., 47, W12518, doi:10.1029/2010WR010367, 2011. 4166

Cardiff, M., Barrash, W., and Kitanidis, P. K.: A field proof-of-concept of aquifer imaging using 3-D transient hydraulic tomography with modular, temporarily-emplaced equipment, Water Resour. Res., 48, W05531, doi:10.1029/2011WR011704, 2012. 4166

Castagna, M. and Bellin, A.: A Bayesian approach for inversion of hydraulic tomographic data, Water Resour. Res., 45, W04410, doi:10.1029/2008WR007078, 2009. 4166

Cooper, H. H. and Jacob, C. E.: A generalized graphical method of evaluating formation constants and summarizing well-field history, US Dept. of the Interior, Geological Survey, Water Resources Division, Ground Water Branch; Washington DC, USA, 1946. 4165

Cressie, N.: Statistics for Spatial Data, in: Wiley Series in Probability and Statistics, Wiley-Interscience, Hoboken, New Jersey, USA, 1993. 4174

de Marsily, G.: Quantitative Hydrogeology: Groundwater Hydrology for Engineers, 1st Edn., Academic Press, Waltham, Massachusetts, USA, 1986. 4174

Deutsch, C. V. and Journel, A. G.: GSLIB, Oxford University Press, Oxford, UK, 1997. 4174, 4182, 4183

Diggle, P. J. and Ribeiro, P. J.: Model-based Geostatistics, Springer Series in Statistics, Springer, New York, USA, 2007. 4174

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10143–10162, doi:10.1029/94JC00572, 1994. 4173

Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, Springer, New York, USA 2009.. 4170, 4187

Franssen, H. H. and Kinzelbach, W.: Ensemble Kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems, J. Hydrol., 365, 261–274, doi:10.1016/j.jhydrol.2008.11.033, 2009. 4167

Gottlieb, J. and Dietrich, P.: Identification of the permeability distribution in soil by hydraulic tomography, Inverse Probl., 11, 353–360, doi:10.1088/0266-5611/11/2/005, 1995. 4166

Harbaugh, A. W., Banta, E. R., Hill, M. C., and McDonald, M. G.: MODFLOW-2000, The US Geological Survey Modular Ground-Water Model-User Guide to Modularization Concepts and the Ground-Water Flow Process, Tech. rep., United States Geological Survey, Reston, Virginia, USA, 2000. 4172, 4182, 4187

Harvey, C. F. and Gorelick, S. M.: Temporal moment-generating equations: modeling transport and mass transfer in heterogeneous aquifers, Water Resour. Res., 31, 1895–1911, doi:10.1029/95WR01231, 1995. 4168

Illman, W. A., Zhu, J., Craig, A. J., and Yin, D.: Comparison of aquifer characterization approaches through steady state groundwater model validation: a controlled laboratory sandbox study, Water Resour. Res., 46, W04502, doi:10.1029/2009WR007745, 2010. 4166

Isaaks, E. H. and Srivastava, R. M.: An Introduction to Applied Geostatistics, Oxford University Press, USA, 1990. 4174

Kalman, R.: A New Approach to Linear Filtering and Prediction Problems, Transactions of the J. Basic Eng.-T. ASME, 35–45, available at: http://www.cs.unc.edu/~welch/kalman/media/pdf/Kalman1960.pdf, last access: 4 March 2014, 1960. 4170, 4173

HESSD
11, 4163–4208, 2014

Kitanidis, P. K.: Quasi-linear geostatistical theory for inversing, Water Resour. Res., 31, 2411, doi:10.1029/95WR01945, 1995. 4166

Li, W., Nowak, W., and Cirpka, O. A.: Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown, Water Resour. Res., 41, W08403, doi:10.1029/2004WR003874, 2005. 4168, 4170, 4171, 4172

Liu, S.: Effectiveness of hydraulic tomography: sandbox experiments, Water Resour. Res., 38, 1034, doi:10.1029/2001WR000338, 2002. 4166

Liu, X. and Kitanidis, P. K.: Large-scale inverse modeling with an application in hydraulic tomography, Water Resour. Res., 47, W02501, doi:10.1029/2010WR009144, 2011. 4166

Liu, X., Illman, W. A., Craig, A. J., Zhu, J., and Yeh, T.-C. J.: Laboratory sandbox validation of transient hydraulic tomography, Water Resour. Res., 43, W05404, doi:10.1029/2006WR005144, 2007. 4166

Mao, D., Yeh, T.-C. J., Wan, L., Wen, J.-C., Lu, W., Lee, C.-H., and Hsu, K.-C.: Joint interpretation of sequential pumping tests in unconfined aquifers, Water Resour. Res., 49, 1782–1796, doi:10.1002/wrcr.20129, 2013. 4166

Matheron, G.: Traité de géostatistique appliquée, Mém. Bur. Rech. Géol. Miniéres, Paris, France, 333 pp., 1962. 4174

Moore, C. and Doherty, J.: The cost of uniqueness in groundwater model calibration, Adv. Water Resour., 29, 605–623, doi:10.1016/j.advwatres.2005.07.003, 2006. 4167

Olsthoorn, T. N.: Do a bit more with convolution, Ground Water, 46, 13–22, doi:10.1111/j.1745-6584.2007.00342.x, 2008. 4168, 4170

Rubin, Y.: Applied Stochastic Hydrogeology, Oxford University Press, New York, USA, 2003. 4191

Schöniger, A., Nowak, W., and Hendricks Franssen, H.-J.: Parameter estimation by ensemble Kalman filters with transformed data: approach and application to hydraulic tomography, Water Resour. Res., 48, W04502, doi:10.1029/2011WR010462, 2012. 4166, 4167

Shin, V., Lee, Y., and Choi, T.-S.: Generalized Millman's formula and its application for estimation problems, Signal Process., 86, 257–266, doi:10.1016/j.sigpro.2005.05.015, 2006. 4168, 4177, 4179, 4189

Straface, S., Yeh, T.-C. J., Zhu, J., Troisi, S., and Lee, C. H.: Sequential aquifer tests at a well field, Montalto Uffugo Scalo, Italy, Water Resour. Res., 43, W07432, doi:10.1029/2006WR005287, 2007. 4166

Sun, N.-Z.: Inverse Problems in Groundwater Modeling (Theory and Applications of Transport in Porous Media), Springer, New York, USA, 1994. 4165

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM: Society for Industrial and Applied Mathematics, Philadelphia, PA, USA 2004. 4165

Theis, C. V.: The Relation Between the Lowering of the Piezometric Surface and the Rate and Duration of Discharge of a Well Using Ground Water Storage, US Department of the Interior, Geological Survey, Water Resources Division, Ground Water Branch, Washington DC, USA, 1935. 4165

Von Asmuth, J. and Maas, K.: The method of impulse response moments: a new method integrating time series-, groundwater-and eco-hydrological modelling, in: Impact of Human Activity on Groundwater Dynamics. Proceedings of a Symposium Held During the Sixth IAHS Scientific Assembly, Maastricht, Netherlands, 18–27 July 2001, 51–58, 2001. 4168, 4170

Von Asmuth, J. R., Bierkens, M. F. P., and Maas, K.: Transfer function-noise modeling in continuous time using predefined impulse response functions, Water Resour. Res., 38, 1287, doi:10.1029/2001WR001136, 2002. 4171

Von Asmuth, J. R., Maas, K., Bakker, M., and Petersen, J.: Modeling time series of ground water head fluctuations subjected to multiple stresses, Ground Water, 46, 30–40, doi:10.1111/j.1745-6584.2007.00382.x, 2008. 4168, 4170

Wu, C.-M., Yeh, T.-C. J., Zhu, J., Lee, T. H., Hsu, N.-S., Chen, C.-H., and Sancho, A. F.: Traditional analysis of aquifer tests: comparing apples to oranges?, Water Resour. Res., 41, W09402, doi:10.1029/2004WR003717, 2005. 4166

Yeh, T.-C. J. and Liu, S.: Hydraulic tomography: development of a new aquifer test method, Water Resour. Res., 36, 2095, doi:10.1029/2000WR900114, 2000. 4166, 4167

Yeh, T.-C. J. and Zhu, J.: Hydraulic/partitioning tracer tomography for characterization of dense nonaqueous phase liquid source zones, Water Resour. Res., 43, W06435, doi:10.1029/2006WR004877, 2007. 4166

Yin, D. and Illman, W. A.: Hydraulic tomography using temporal moments of drawdown recovery data: a laboratory sandbox study, Water Resour. Res., 45, W01502, doi:10.1029/2007WR006623, 2009. 4168, 4172, 4186

Zhu, J. and Yeh, T.-C. J.: Characterization of aquifer heterogeneity using transient hydraulic tomography, Water Resour. Res., 41, W07028, doi:10.1029/2004WR003790, 2005. 4166, 4168

Zhu, J. and Yeh, T.-C. J.: Analysis of hydraulic tomography using temporal moments of drawdown recovery data, Water Resour. Res., 42, W02403, doi:10.1029/2005WR004309, 2006. 4168

HESSD
11, 4163–4208, 2014

Hydraulic tomography data fusion

A. H. Alzraiee et. al

Table 1. Alternative formulations of the forecast matrix investigated in the numerical experiments.

Formulation	Description	Forecast matrix
A	Estimate K field by assimilating m_0 measurements only, with PDE (Eq. 7) as forecast model.	$\mathbf{X}_f^Y = [\mathbf{Y};\mathbf{M}_0^1;\ldots;\mathbf{M}_0^{N_p}]$
B	Estimate K field by assimilating m_1 measurements only, with PDE (Eq. 8) as forecast model, in which the Y forecast ensemble and its corresponding m_0 forecast ensemble (obtained from PDE, Eq. 7) are used.	$\mathbf{X}_f^Y = [\mathbf{Y};\mathbf{M}_1^1;\ldots;\mathbf{M}_1^{N_p}]$
C	Estimate K field by joint assimilation of m_0 and m_1 measurements, with PDEs (Eqs. 7 and 8) as forecast model.	$\mathbf{X}_f^Y = [\mathbf{Y};\mathbf{M}_0^1;\ldots;\mathbf{M}_0^{N_p};\mathbf{M}_1^1;\ldots;\mathbf{M}_1^{N_p}]$
D	Estimate S_s field by assimilating m_1 measurements only, with PDE (Eq. 8) as forecast model, in which the Y forecast ensemble and its corresponding m_0 forecast ensemble (obtained from PDE, Eq. 7) are used.	$\mathbf{X}_f^Z = [\mathbf{Z};\mathbf{M}_1^1;\ldots;\mathbf{M}_1^{N_p}]$
E	Estimate S_s field by assimilating m_1 measurements only, with PDE (Eq. 8) as forecast model, in which the posterior mean of Y, as estimated in **A**, and its corresponding m_0 distribution (obtained from PDE, Eq. 7) are used.	$\mathbf{X}_f^Z = [\mathbf{Z};\mathbf{M}_1^1;\ldots;\mathbf{M}_1^{N_p}]$

HESSD
11, 4163–4208, 2014
Hydraulic tomography data fusion
A. H. Alzraiee et. al

Table 2. Model setting for the numerical experiments.

Finite-Difference Grid Properties	
Domain Dimensions [x,y,z] (m, m, m)	[1000, 1000, 50]
Cell Size [x,y,z] (m, m, m)	[10, 10, 10]
Total number of Cells	10 000
Boundary Conditions	
Dirichlet boundary conditions at:	
$x = 0$ m	$h = 45$ m
$x = 1000$ m	$h = 45$ m
Neumann boundary conditions at:	
$y = 0$ m	no-flow
$y = 1000$ m	no-flow
Geostatistical Parameters	
$[\mu_Y, \sigma_Y, \lambda_Y]$ (ln m day^{-1}, ln m day^{-1}, m)	[1.5, 1, 350]
$[\mu_Z, \sigma_Z, \lambda_Z]$ (ln m^{-1}, ln m^{-1}, m)	[−10, 1, 350]
Pumping Tests	
Well #1 [$x, y; Q$] (m, m; m^3 day^{-1})	[500, 500; 500]
Well #2 [$x, y; Q$] (m, m; m^3 day^{-1})	[200, 500; 500]
Well #3 [$x, y; Q$] (m, m; m^3 day^{-1})	[800, 500; 500]
Well #4 [$x, y; Q$] (m, m; m^3 day^{-1})	[500, 200; 500]
Well #5 [$x, y; Q$] (m, m; m^3 day^{-1})	[500, 800; 500]
Observation Wells	See layout in Fig. 2

HESSD
11, 4163–4208, 2014

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Title Page
Abstract
Introduction
Conclusions
References
Tables
Figures
◀|
|▶
◀
▶
Back
Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

HESSD

11, 4163–4208, 2014

Hydraulic tomography data fusion

A. H. Alzraiee et. al

Table 3. Numerical experiments carried out to analyze the performance of the data fusion schemes.

Experiment set	Purpose	Ensemble size
I	Investigate the CF performance based on the formulations listed in Table 1	200
II	Investigate the DF performance based on the formulations listed in Table 1	200
III	Investigate the CF performance under uncertain geostatistical parameters	200
IV	Investigates the CF performance in relation tojoint estimation of geostatistical parameters and Y and Z fields	1000

Table 4. Performance statistics for the formulations of Table 1 using CF.

Performance Statistics	Formulation				
	$Y = \ln K$			$Z = \ln S_s$	
	A	B	C	D	E
Mean Absolute Error: L_1	0.318	0.353	0.343	0.596	0.363
Root Mean Square Error: L_2	0.408	0.446	0.438	0.730	0.460
Correlation Coefficient: r	0.825	0.787	0.803	0.292	0.759
Mean Error: μ_e	1.40×10^{-5}	1.54×10^{-5}	1.01×10^{-5}	-6.13×10^{-6}	-5.31×10^{-6}

HESSD
11, 4163–4208, 2014

HESSD
11, 4163–4208, 2014

Hydraulic tomography data fusion

A. H. Alzraiee et. al

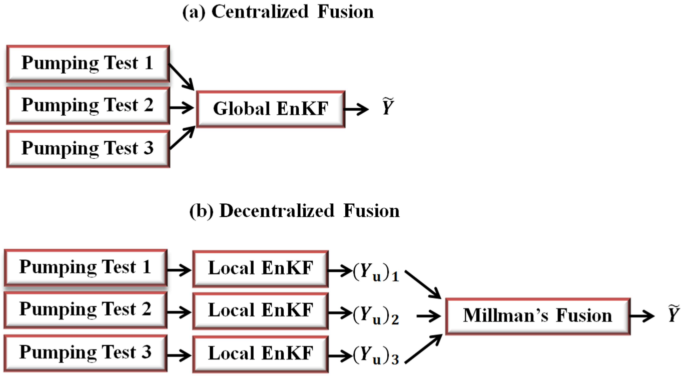

Fig. 1. Flowcharts illustrating the structure of **(a)** the CF approach and **(b)** the DF approach.

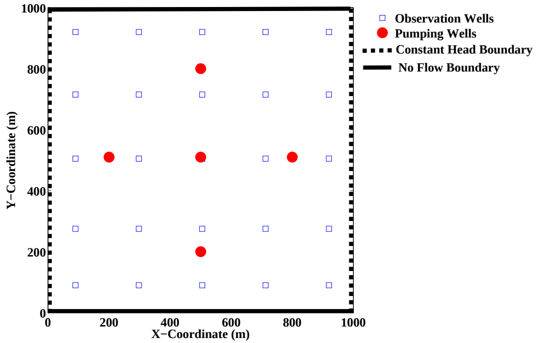

Fig. 2. Locations of pumping wells, observation wells, and boundary conditions.

HESSD
11, 4163–4208, 2014
Hydraulic tomography data fusion
A. H. Alzraiee et. al

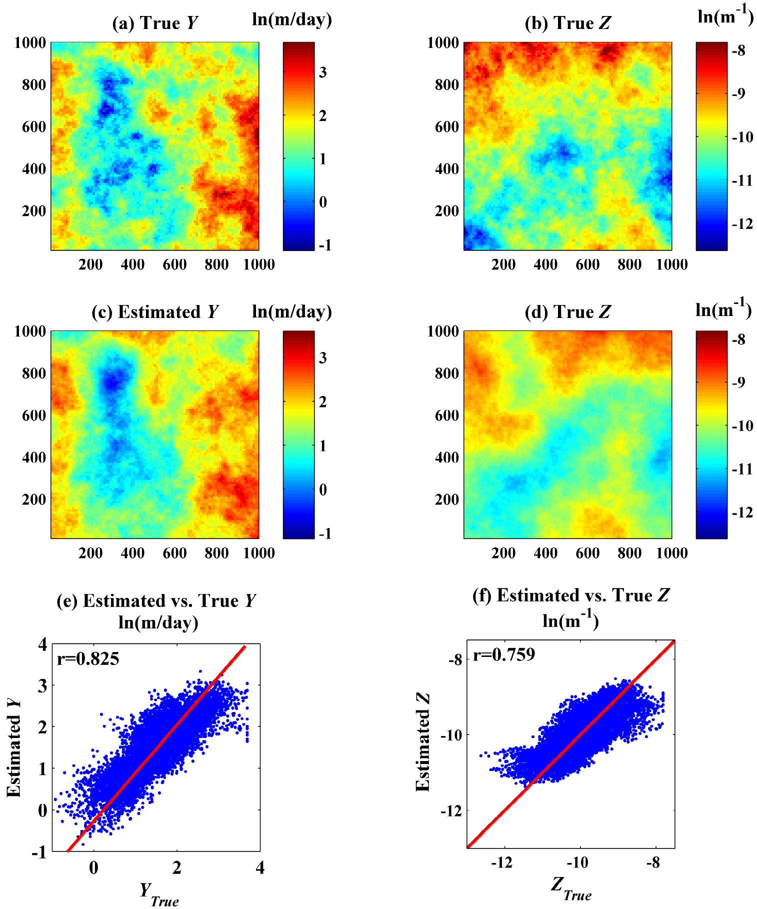

Fig. 3. Maps of (**a** and **b**) the "true" reference Y and Z fields, and (**c** and **d**) the Y and Z fields estimated using the CF scheme with formulations A and E, respectively. (**e** and **f**) Scatter plots of $\mathbf{Y}_{\mathrm{true}}$ vs. $\hat{\mathbf{Y}}$ and $\mathbf{Z}_{\mathrm{true}}$ vs. $\hat{\mathbf{Z}}$.

HESSD
11, 4163–4208, 2014

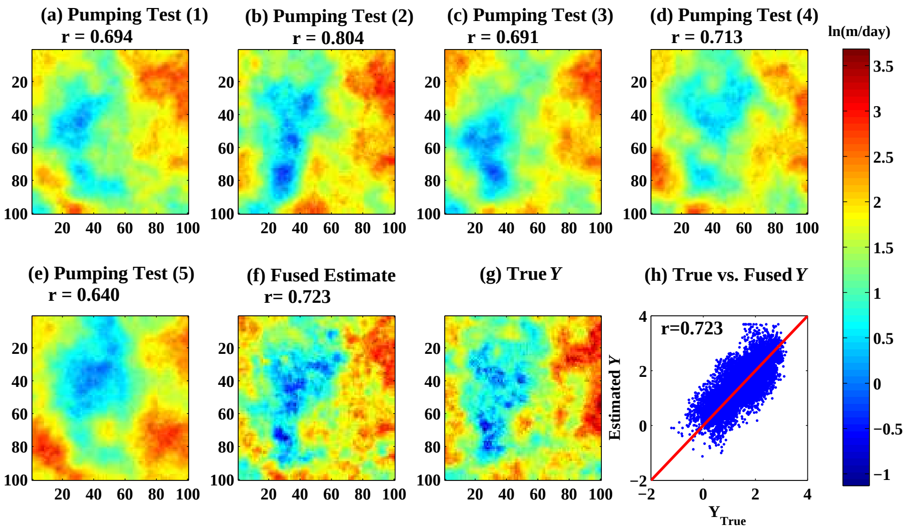

Fig. 4. Maps of the Y field obtained with **(a–e)** local EnKF estimates for each of the five hypothesized pumping tests, and **(f)** the application of the DF scheme. The "true" reference field is given in subpanel **(g)**. Subpanel **(h)** shows the scatter plot of $\mathbf{Y}_{\text{true}}$ vs. $\tilde{\mathbf{Y}}$.

HESSD
11, 4163–4208, 2014

Hydraulic tomography data fusion

A. H. Alzraiee et. al

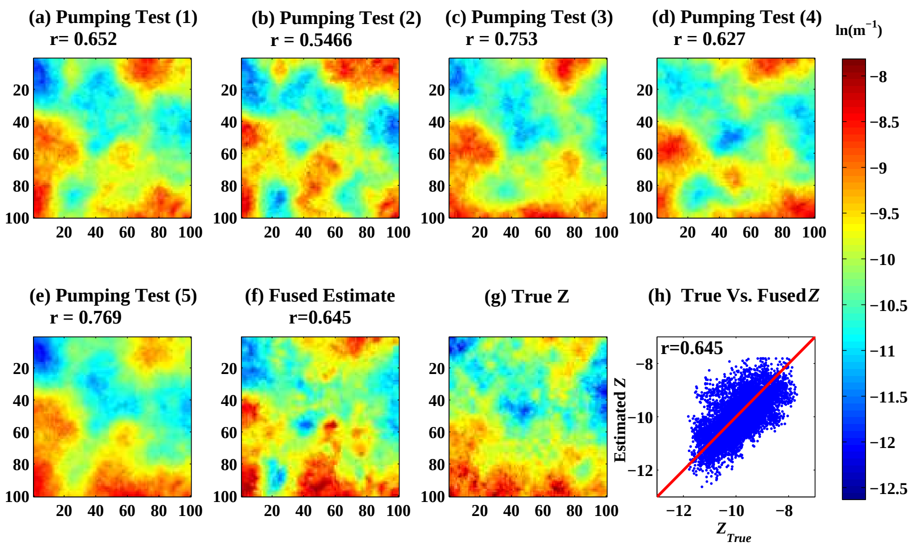

Fig. 5. Maps of the Z field obtained with **(a–e)** local EnKF estimates for each of the five hypothesized pumping tests, and **(f)** the application of the DF scheme. The "true" reference field is given in subpanel **(g)**. Subpanel **(h)** shows the scatter plot of $\mathbf{Z}_{\text{true}}$ vs. $\tilde{\mathbf{Z}}$.

HESSD
11, 4163–4208, 2014

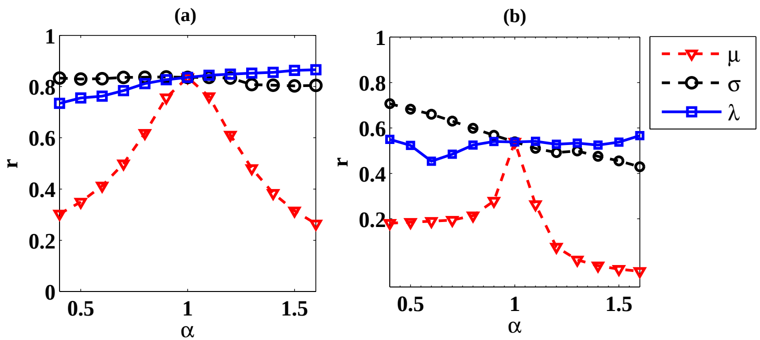

Fig. 6. Sensitivity of the coefficient of regression r between the "true" and the estimated parameter distributions to errors in the selection of the geostatistical parameters for **(a)** the Y field and **(b)** the Z field. α represents the ratio between the selected geostatistical parameter and the "true" geostatistical parameter.

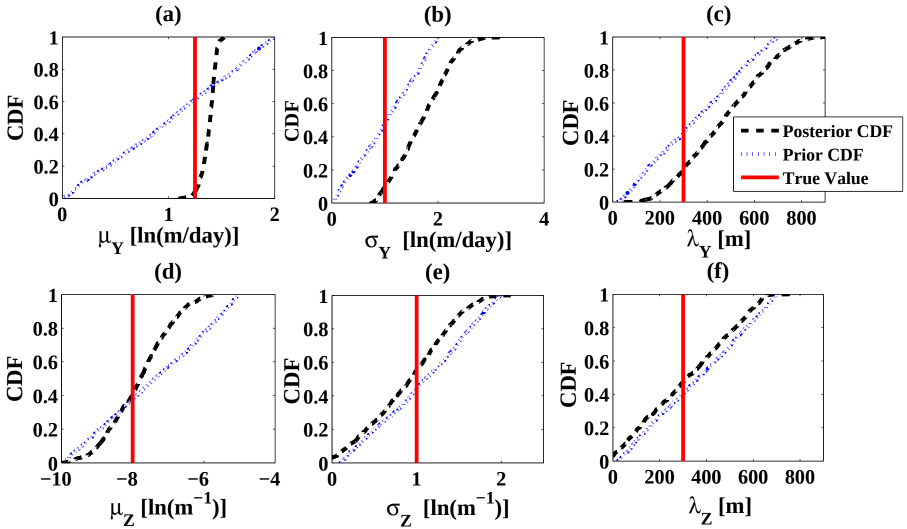

Fig. 7. Prior and posterior CDFs of the geostatistical parameters **(a)** μ_Y, **(b)** σ_Y, **(c)** λ_Y, **(d)** μ_Z, **(e)** σ_Z, and **(f)** λ_Z estimated in Experiment set IV. Each subpanels shows the corresponding "true" reference value of the corresponding geostatistical parameter.

HESSD
11, 4163–4208, 2014