Journal of Hydrology 365 (2009) 261–274

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Ensemble Kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems

H.J. Hendricks Franssen *, W. Kinzelbach

Institute of Environmental Engineering, ETH Zurich, Wolfgang-Pauli-Strasse 15, 8093 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 22 April 2008
Received in revised form 25 November 2008
Accepted 27 November 2008

Keywords:
Groundwater hydrology
Data assimilation
Inverse modelling
Sequential self-calibration
Ensemble Kalman Filtering
Transient groundwater flow

SUMMARY

Monte-Carlo (MC) type inverse modelling techniques like the sequential self-calibration (SSC) method, can be used as a benchmark to test other inverse modelling procedures. In comparison studies MC type inverse modelling methods outperformed other inverse parameter estimation methods, but the large amount of CPU time needed can become prohibitive to use such methods. Therefore, an interest exists to develop alternative methods that perform nearly as good and need much less CPU time. In this paper Ensemble Kalman Filtering (EnKF) is promoted for the off-line calibration of transient groundwater flow models with many nodes. An augmented state vector approach is used to calibrate parameters together with the updating of the states. EnKF and SSC are compared in two calibration experiments, for a mildly heterogeneous case ($\sigma^2_{\ln T} = 1.0$) and a strongly heterogeneous case ($\sigma^2_{\ln T} = 2.7$). For the mildly heterogeneous case, EnKF gives very similar results to SSC, also in terms of calibrated log-transmissivity fields. For the strongly heterogeneous case, EnKF gives still similar results as SSC, although the characterisation of the log$T$ field improves less compared to SSC. On the other hand, EnKF needed in both cases around a factor of 80 less CPU time than SSC. In addition, the performance of EnKF and SSC was compared in two prediction experiments: (1) the prediction of groundwater flow in a different flow situation (without pumping and with a different time series of recharge rate), (2) the prediction of solute transport towards a pumping well. Both in the mildly and strongly heterogeneous cases the quality of the predictions with EnKF was as good as for SSC. Given the good performance of EnKF, the strength of EnKF to include multiple sources of uncertainty (e.g., related to external forcing) and the reduced CPU time needed compared to MC type inverse modelling, EnKF seems to be an interesting candidate for the stochastic calibration of large subsurface hydrological models.

Introduction

In the geosciences different strategies exist to calibrate predictive models. In groundwater hydrology, inverse modelling is used to calibrate model parameters with the help of historical time series of hydraulic head data. Often it is assumed that the uncertainty with respect to the model parameters (in particular, transmissivity) is the dominant source of uncertainty and therefore methods are strongly focused on the estimation of the unknown parameters. Normally, only few measurement data are available, which makes the estimation of the unknown model parameters non-trivial given the problems of non-uniqueness, identifiability and instability (e.g., Carrera and Neuman, 1986b). Geostatistical approaches (e.g., de Marsily, 1978; Kitanidis and Vomvoris, 1983) and the zonation approach (Carrera and Neuman, 1986a), that limit the number of parameters to be estimated, provide a solution to obtain stable estimates. The conditional mean estimates of these deterministic methods result in overly smoothed parameter fields, which are not suited to predict early arrival and tailing of contaminant breakthrough curves (Alcolea et al., 2008). Monte-Carlo (MC) type inverse modelling does not seek a unique solution to the inverse problem, but a large set of conditional realizations, which are all conditioned on transmissivity and hydraulic head measurements. Sequential self-calibration (SSC) is an inverse type MC method, and was first proposed for 2D steady-state groundwater flow (Sahuquillo et al., 1992; Gómez-Hernández et al., 1997), later extended to transient groundwater flow and the joint calibration of transmissivities and storativities (Hendricks Franssen et al., 1999), 3D fractured media (Gómez-Hernández et al., 2001), the coupled inverse modelling of groundwater flow and mass transport (Hendricks Franssen et al., 2003) and the integration of remote sensing data in stochastic groundwater flow models (Hendricks Franssen et al., 2008a). Another technique for MC type inverse modelling is the Pilot Points method (de Marsily, 1978). LaVenue et al. (1995) proposed a MC type variant of the Pilot Points method, which was extended for transient groundwater flow in 3D fractured media (LaVenue and de Marsily, 2001). Alcolea et al.

* Corresponding author. Tel.: +41 44 633 30 74; fax: +41 44 633 10 61.
E-mail address: hendricks@ifu.baug.ethz.ch (H.J.H. Franssen).

0022-1694/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2008.11.033

(2006) proposed the Regularized Pilot Points method, which includes a plausibility term in the definition of the objective function, besides the term which measures the mismatch of simulated and measured hydraulic heads. Also the Representer method (Bennett, 1992; Valstar, 2001) was adapted to handle the calibration of multiple equally likely stochastic realizations. A comparison study (Hendricks Franssen et al., submitted) tested different MC type inverse modelling methods for the characterisation of groundwater flow and mass transport around a well and found that the MC type inverse methods yielded very similar results. As geological media often exhibit complex structures which are not well described with a Multi-Gaussian Model, recently the inverse estimation of hydraulic conductivities for Non-Multi-Gaussian media has become an important research topic. One example is the conditional probabilities method (Capilla et al., 1999), which is a modification of SSC based on indicator geostatistics and the perturbation of conditional probabilities. Other examples are the combination of truncated pluri-Gaussian simulation and the gradual deformation approach (Hu et al., 2001), the inverse modelling of multimodal hydraulic conductivity distributions with the representer method (Janssen et al., 2006), or the probability perturbation method combined with multiple-point statistics (Caers and Hoffman, 2006). However, a major handicap of MC type methods is the needed CPU time. Therefore, alternative inverse methods were developed that aim at reducing strongly the needed CPU time, by reducing the physical model using eigenvector techniques (Vermeulen et al., 2004), by conditional estimation that can resolve the details of the parameter fields that are relevant for a given problem (Tonkin et al., 2007) or by the Moment Equations method (Hernandez et al., 2003, 2006), for instance.

In groundwater hydrology, model improvement by conditioning to data usually focuses on parameter estimation. However, most methods are also able to handle measurement uncertainty in the inverse estimation procedures. Conceptual model uncertainty is rarely handled in groundwater hydrological studies, and only addressed in some studies, where different conceptual models are evaluated and compared on the basis of statistical selection criteria (e.g., Carrera and Neuman, 1986c; Foglia et al., 2007). Here, we consider that conceptual model uncertainty consists of uncertainty with regard to the exact mathematical description of physical and chemical processes (an example could be the validity of the Darcy law for high groundwater flow velocities close to wells in gravel aquifers), uncertainty with regard to the statistical model adopted for the uncertain parameters (e.g., transmissivity) and uncertainty with regard to external forcing (recharge rate and lateral inflow, for example). Therefore, the calibration of geostatistical parameters (range, nugget and sill) aims at handling part of the conceptual uncertainty, and some inversion methods allow to treat this kind of uncertainty explicitly (e.g., Hoeksema and Kitanidis, 1984; Woodbury and Ulrych, 2000). Nevertheless, in other studies it was found that it is difficult to correct an erroneous conceptual model of $\log T$ with hydraulic head data (Kerrou et al., 2008). Uncertainty with respect to forcing is sometimes handled in groundwater hydrological studies by treating the uncertain external forcing as additional parameters to be calibrated in a statistical framework (Hendricks Franssen et al., 2004, 2008a). However, such an approach is complicated and hardly practicable for time-varying recharge rates. Altogether, nowadays the calibration of large (groundwater) hydrological models, with multiple important sources of uncertainty like interaction with surface water bodies and recharge rate, and with a reliable characterisation of uncertainty, is still an open issue.

In the atmospheric sciences for numerical weather prediction (NWP), or for land surface models, the improvement of model predictions is normally centred on improving the initial conditions of the simulation model, by assimilating measurement data in real-time. For NWP commonly variational data assimilation is used (for an overview see e.g., Kalnay, 2003), like for instance 4D variational data assimilation (4DVAR) which is implemented in the European Model for Numerical Weather Forecasting ECMWF. Ensemble Kalman Filtering (EnKF) (Evensen, 1994) is a method for the sequential assimilation of measurement data and the updating of the model states based on optimally combining the ensemble of model predictions and the measurements. EnKF is increasingly considered an alternative for the updating of the states of atmospheric circulation models. Houtekamer et al. (2005) found that EnKF underestimated the growth rates of perturbations, compared with reality. Kalnay et al. (2007) report that simulation experiments with EnKF become competitive with 4DVAR. Wang et al. (2007) investigated the performance of a hybrid data assimilation scheme, using Ensemble Transform Kalman Filtering and optimal interpolation. Such an approach combines CPU efficiency with accuracy. Besides the classical stochastic EnKF (Burgers et al., 1998), in the atmospheric science literature different variants of a deterministic EnKF were proposed, like the Ensemble Square Root Filter (Whitaker and Hamill, 2002), the Ensemble Adjustment Kalman Filter (Anderson, 2001) and the Ensemble Transform Kalman Filter (Bishop et al., 2001). The deterministic EnKF formulation avoids the need to perturb the observed values and eliminates therefore an additional error source from the data assimilation framework. In the atmospheric sciences literature the uncertainty is centred on the unknown states (which should be able to include uncertainty with regard to the physical equations), and these are updated. Parameter uncertainty is seldom considered, an exception being the work of Aksoy et al. (2006), where parameter uncertainty of a 2D mesoscale meteorological model for sea breeze was considered. Also in oceanography data assimilation methods like EnKF have been applied frequently.

The combined estimation of states and parameters was addressed in the petroleum engineering literature, starting with the work of Naevdal et al. (2003). Naevdal et al. (2003) estimated unknown parameters together with the states by an augmented state vector approach. Wen and Chen (2006) proposed two-step iterative EnKF, which reduces the inconsistency due to the joint updating of states and parameters. In their approach, for a certain time step, first the parameters are updated (on the basis of pressure measurements), and next, with the new parameter values, also the states are updated (assimilating the same hydraulic head data). The work in the petroleum engineering literature is oriented on CPU efficient estimation of parameters, and the ability to handle real-time information from sensors. However, the work in the petroleum engineering literature was not focused on considering other sources of uncertainty besides parameter uncertainty and measurement uncertainty. Hendricks Franssen and Kinzelbach (2008) compared iterative and non-iterative EnKF for a groundwater flow problem, and considered also uncertainty with respect to the forcings (recharge rate). In non-iterative EnKF, states and parameters are updated jointly in one step. Iterative EnKF on the contrary, first updates the parameters, and repeats the calculations for that particular time step with the new parameters to update also the states. In the surface hydrology literature, there is an increasing concern about the joint handling of measurement, parameter and conceptual uncertainty, and uncertainty with regard to the forcings (e.g., Liu and Gupta, 2007). Vrugt et al. (2005) used sequential data assimilation with the Ensemble Kalman Filter, together with global optimization, to estimate parameters and states of a surface hydrological model. Moradkhani et al. (2005) also handled the uncertainty with regard to states and parameters jointly, by adopting an augmented state vector approach with the EnKF, and using a dual approach, where first parameters, and later states are updated for a certain time step.

Blasone et al. (2008a, 2008b) increased the efficiency of the joint estimation of states and parameters with a Markov Chain Monte-Carlo method.

There are three main reasons why Ensemble Kalman Filtering could be attractive to condition groundwater flow models to state information:

(1) It is ideally suited to assimilate data available from sensors in real-time.
(2) It is more CPU efficient than MC type inverse modelling.
(3) It is a flexible framework to consider different sources of uncertainty in an integrated fashion.

However, for applications of Ensemble Kalman Filtering in groundwater flow problems there are also draw backs. The performance of EnKF is optimal for normally distributed state variables and linear dynamics. Non-normally distributed state variables and non-linear dynamics can result in a worse performance of the Ensemble Kalman Filter, with an increasing underestimation of variance over time (filter inbreeding), which even could result in filter divergence. The ensemble approach reduces the limitations of the Kalman Filter concerning normality and linear dynamics, but a large number of realizations might be needed to avoid problems with filter inbreeding. Adaptive covariance matrices, that use information from innovation statistics, were found to reduce filter inbreeding problems (e.g., Wang et al., 2007; Anderson, 2007; Reichle et al., 2008). Also the selection of realizations on the basis of resampling was found to reduce the efficiency and the problems with filter inbreeding (Wen and Chen, 2007; Hendricks Franssen and Kinzelbach, 2008). When resampling is applied, the simulations start with a large number of stochastic realizations, from which a subset is sampled, for example on the basis of the reproduction of univariate and bivariate cumulative probability density functions (as compared with the large set of stochastic realizations) of the state variable. Problems with filter inbreeding can also be reduced if the particle filter is used for sequential data assimilation. Particle filters approximate the random variable by a large number of "particles", each of which carries a probability mass (Doucet et al., 2001). The main advantage of the particle filter, as compared to EnKF, is that the particle filter also gives unbiased estimates of the posterior distribution for non-Gaussian distributions, and is less affected by non-linear model dynamics. However, a main disadvantage as compared with EnKF is that the method needs much more CPU time. The method was recently found to be potentially applicable to large-scale problems (e.g., Xiong et al., 2006). The efficiency of EnKF can also be improved by sampling stochastic realizations of input parameters along the leading eigenvectors of the covariance matrix (Verlaan and Heemink, 1997; Evensen, 2004; Zhang et al., 2007). However, this approach does not always reduce the problems with filter inbreeding (Madsen and Sörensen, 2007).

In order to investigate to what extend the mentioned problems limit the application of EnKF in groundwater hydrology studies, we compare its performance with SSC, a MC type method for the inverse modelling of groundwater flow. The comparison is limited to cases with parameter uncertainty and measurement uncertainty, and is evaluated both for a mildly and a strongly heterogeneous transmissivity field. In addition, the two methods are also compared, for both the mildly and strongly heterogeneous case, in two prediction cases. In the first case, the groundwater flow situation is changed due to different spatio-temporally variable recharge rates and the shutting down of the pumping wells. In the second case, only one well pumps, and the advective transport towards that well is evaluated for both methods. The comparison of the performance of the two methods also includes a comparison of the CPU time needed by the two methods, in order to find out to what extend EnKF can reduce the needed CPU time compared to SSC.

Theory

We consider 2D saturated transient groundwater flow:

$$\nabla \cdot (\mathbf{T}\nabla h) = S\frac{\partial h}{\partial t} + q \tag{1}$$

where $\mathbf{T}$ is the transmissivity tensor [L^2T^{-1}], h the hydraulic head [L], S the storativity [−], t is time [T] and q represents sinks (abstractions) and sources (recharge) [LT^{-1}].

This equation is solved with a finite difference method (details will be provided later). Time series of hydraulic head measurements are available and are sequentially assimilated to update the transmissivities and the model states (in case of EnKF) or assimilated in batch to calibrate transmissivities (in case of SSC).

EnKF with updating of parameters

For an ensemble of stochastic realizations, the transient groundwater flow equation is solved. Input information includes the initial conditions, boundary conditions, forcings and parameter values. There might be different sources of uncertainty, like parameter uncertainty, forcing uncertainty and uncertain initial conditions. Here the only parameter that is uncertain is log-transmissivity Y, but also the initial conditions and forcings might be uncertain. The equation is solved for each stochastic realization according to:

$$\mathbf{x}^0_{i,\mathrm{h}} = M(\mathbf{x}^-_{i,\mathrm{h}}) \tag{2}$$

where i refers to the stochastic realization ($i = 1, \ldots, P$), M to the groundwater flow model, $\mathbf{x}_\mathrm{h}$ is the state vector from the previous time step (superscript −) or the actual time step (superscript 0). The vector $\mathbf{x}_\mathrm{h}$ is of dimension N, where N is the number of grid cells of a finite difference groundwater flow model.

If at a certain time step hydraulic head data are available, the Ensemble Kalman Filter (EnKF), based on the formulation of Burgers et al. (1998) is applied to update the model states and the parameters, by combining the model prediction and the measurement data in an optimal manner. For this an augmented state vector approach is used, which means that the state vector contains both the states (hydraulic heads) and the parameters (log-transmissivities) according to

$$\mathbf{x}_i = \begin{Bmatrix} \mathbf{x}_{i,\mathrm{h}} \\ \mathbf{x}_{i,\mathrm{Y}} \end{Bmatrix} \tag{3}$$

where the subscript h refers to the part associated with the states and the subscript Y to the part associated with the log-transmissivities. The vector $\mathbf{x}_i$ is of dimension $2N$, because it is assumed that the groundwater flow equation is solved with a finite differences formulation. Note, however, that the number of states to be updated is not necessarily the same as the number of parameters. From the ensemble of vectors $\mathbf{x}_i$ the covariance matrix ($2N \times 2N$) is estimated. Therefore, the covariance matrix does not have to be calculated explicitly. The covariance matrix is

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{\mathrm{hh}} & \mathbf{C}_{\mathrm{Yh}} \\ \mathbf{C}_{\mathrm{hY}} & \mathbf{C}_{\mathrm{YY}} \end{pmatrix} \tag{4}$$

where the subscript hh refers to covariances between hydraulic heads at two locations (grid cells), the subscript hY to cross covariances between a hydraulic head value at one location and a log-transmissivity value at another location and YY to covariances between log-transmissivities at two locations. These covariances are estimated from the ensemble of augmented state vectors $\mathbf{x}_i$. Finally,

the measurement data for the actual time step are stored in the vector $\mathbf{y}^0$ of dimension n. The data are perturbed (as motivated in Burgers et al. (1998)) according to

$$\mathbf{y}_i^0 = \mathbf{y}^0 + \boldsymbol{\varepsilon}_i^0 \tag{5}$$

where $\boldsymbol{\varepsilon}$ is a vector of random numbers, drawn from a normal distribution with expectation zero and standard deviation equal to the expected measurement standard deviation. For each stochastic realization a different vector of perturbed observations is used. Here it is assumed that only state measurements are available.

The model predictions (2) and the measurement data (5) are combined to yield an updated ensemble of states (hydraulic heads) and parameters (log-transmissvities). The updating of the states and parameters is done according to

$$\mathbf{x}_i^+ = \mathbf{x}_i^0 + \alpha\mathbf{K}(\mathbf{y}_i^0 - \mathbf{H}\mathbf{x}_i^0) \tag{6}$$

where $\mathbf{x}^+$ is the updated vector containing updated states and parameters, α is a damping factor that takes a value between 0 and 1, $\mathbf{H}$ is a linear operator that maps the simulated heads and transmissivities at the grid nodes on the measurements ($n \times 2N$). In the examples that will be presented later in this paper, the measurements always coincide with a grid node and therefore matrix H only consists of 1's and 0's. However, Eq. (6) is general and also could account for measurements that do not coincide with grid nodes. The damping factor reduces the perturbation of the log-transmissivity field Y. Smoothing the perturbation limits the error associated with, on one side, the linearization of the relation between log-transmissivities and hydraulic heads, and, on the other side, erroneous numerical covariances due to sampling fluctuations (a limited number of stochastic realizations). $\mathbf{K}$ is the Kalman gain matrix ($2N \times n$)

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_\mathrm{h} \\ \mathbf{K}_\mathrm{Y} \end{bmatrix} \tag{7}$$

where $\mathbf{K}_\mathrm{h}$ is related to the states (hydraulic heads) and $\mathbf{K}_\mathrm{Y}$ to the parameters (log-transmissivities) and $\mathbf{K}$ is calculated from:

$$\mathbf{K} = \mathbf{C}^0\mathbf{H}^\mathbf{T}(\mathbf{H}\mathbf{C}^0\mathbf{H}^\mathbf{T} + \mathbf{R}^0)^{-1} \tag{8}$$

$\mathbf{C}^0$ is the covariance matrix for the actual time step (which was estimated from the ensemble of stochastic realizations) and $\mathbf{R}^0$ ($n \times n$) is the measurement error covariance matrix for the actual time step, which is estimated a priori. If it is assumed that measurement errors at different measurement locations are uncorrelated $\mathbf{R}^0$ is a diagonal matrix with the variances of the measurement errors on the diagonal. Eq. (8) is solved taking profit of the fact that only the entries of matrix $\mathbf{C}^0$ which correspond to nonzero elements in $\mathbf{H}$ (the measurement operator) are needed for calculating $\mathbf{K}$.

The ensemble of updated vectors $\mathbf{x}_\mathrm{i}$ (hydraulic heads and log-transmissivities) is the input into the groundwater flow model for the next time step. When new measurement data become available and will be assimilated, Eqs. (5)–(8) will be applied again to update the hydraulic heads and the log-transmissivities together. However, it is also possible to update only the hydraulic heads (without the transmissivities), excluding the transmissivities from the Eqs. (3)–(8).

The procedure was implemented in the C software EnKF3d.

Monte-Carlo type inverse modelling with sequential self-calibration

Also for sequential self-calibration (SSC) the transient groundwater flow equation is solved for an ensemble of stochastic realizations, and the same input information (initial conditions, boundary conditions, forcings and parameter values) is used as for EnKF. Also SSC is able to handle multiple sources of uncertainty. A difference with EnKF is that different sources of uncertainty, like forcing uncertainty (recharges) and uncertain boundary conditions, are handled by calibrating them. In the formulation below we assume that only the log-transmissivity Y is uncertain, as is the case in the synthetic experiments that will be presented later. The first step of SSC differs from EnKF; the transient groundwater flow equation is not evaluated for only one time step, but for the whole period simulated (all time steps)

$$\mathbf{x}_{i,\mathrm{h}}^{t+1} = M(\mathbf{x}_{i,\mathrm{h}}^t) \quad 0 \le t \le NT - 1 \tag{9}$$

where t indicates the time step and NT is the number of simulation time steps.

The simulation results are compared with the measurement data by an objective function:

$$J = \sum_{t=1}^{NT} \left[(\mathbf{y}^t - \mathbf{H}\mathbf{x}_{i,\mathrm{h}}^t)^T \mathbf{C}_{\mathrm{hh}}^{-1} (\mathbf{y}^t - \mathbf{H}\mathbf{x}_{i,\mathrm{h}}^t) \right] + \Psi_Y (\mathbf{Y}^{prior} - \mathbf{Y}^{sim})^T \mathbf{C}_{\mathbf{YY}}^{-1} (\mathbf{Y}^{prior} - \mathbf{Y}^{sim}) \tag{10}$$

where the first term penalizes the mismatch between the simulated and measured hydraulic heads and the second term penalizes the modification of the input parameters (the log-transmissivities). The superscripts t in the first term indicate that measurements and simulated values are time dependent. The vector $\mathbf{Y}$ of dimension N^{prior} contains the log-transmissivities at the log$T$ measurement locations and master blocks (see later). The superscript *prior* indicates the prior log-transmissivity values at the mentioned locations, and the superscript *sim* the updated values at those same locations. The prior log-transmissivity values are the kriged log$T$ values. $\mathbf{C_{YY}}$ is two times the kriging covariance matrix ($N^{prior} \times N^{prior}$); which contains the covariances between the log-transmissivities at the master blocks (for a proof see Alcolea et al. (2008)) and measurement locations. Finally, Ψ_Y is a weighting parameter.

Eq. (10) is minimized with gradient-based techniques. The gradient of the objective function is calculated with adjoint state equations, and the updating direction is calculated from the gradient with help of non-linear optimization algorithms like steepest descent, Flecher-Reeves conjugate gradients, Hestenes-Stiefel conjugate gradients or quasi-Newton methods. The optimal step size along the updating direction is determined with linear search methods. When the optimal step size along the updating direction is found, all parameters (log-transmissivities) are updated. Details can be found in Hendricks Franssen (2001) and are not repeated here. The updated ensemble of log-transmissivity values (and possibly other parameters, if they would have been considered) is used as input to solve again (9) and (10). The optimization is finished if J drops below a user-defined value $\boldsymbol{\varepsilon}$. Measurement errors of hydraulic head have to be taken into account in setting $\boldsymbol{\varepsilon}$ (see Hendricks Franssen and Gómez-Hernández, 2003).

The gradient calculation, and the subsequent vector with the updating direction, are not calculated for all grid cells N, but only at the so-called master blocks, in order to reduce the required CPU time. The gradient and updating vector have therefore dimension N^{prior} ($N^{prior} << N$). The master blocks cover the aquifer regularly, with around two master blocks per log$T$ correlation length. The master blocks have different positions in the different stochastic realizations, and also change during the optimization process. Although the gradient and updating direction are calculated at the master blocks, the mathematical expressions for the gradient elements take into account that the optimized perturbation of the log-transmissivities at the master blocks are interpolated by ordinary kriging to the rest of the grid cells (i.e., the log-transmissivities at all the grid cells are updated). The detailed mathematical expressions can be found in Hendricks Franssen (2001) and other papers, and are not repeated here.

SSC solves in fact for each of the stochastic realizations a smoothing problem and is, apart from some peculiarities, very similar to 4DVAR. Therefore, in theory it is expected that SSC is superior to EnKF, especially at the beginning of an assimilation period.

Synthetic set-up of calibration experiments

In a first set of synthetic experiments historical data were used to calibrate a groundwater flow problem. Two test cases were considered, one for a mildly heterogeneous aquifer and one for a strongly heterogeneous aquifer. Both EnKF and SSC assimilate two years of time series of hydraulic head data. EnKF assimilates hydraulic head data in real-time and updates at each time step the spatial distribution of hydraulic head data. The transmissivities are updated together with the states. SSC incorporates the complete time series of hydraulic head data at once, and with transient inverse modelling the spatially distributed transmissivities are calibrated. These inverse conditional simulations are compared for both cases with unconditional simulations.

The synthetic case consists of a 2D transient saturated groundwater flow problem in an area of 6 km × 6 km for a period of 2 years. The transient groundwater flow equation was solved using block-centred finite differences in space and implicit finite differences in time. The area is discretized into 60 × 60 grid cells of 100 m × 100 m. The simulation period is discretized into 120 time steps. The Northern boundary is assigned a prescribed head value, with values increasing from 400 m in the North-western corner to 412 m in the North-eastern corner. The other boundaries are impermeable. Three wells are present, pumping at the grid cells (20,20), (20,40) and (40,40) with rates of respectively 0.10 m^3/s, 0.15 m^3/s and 0.15 m^3/s. Fig. 1 depicts the situation.

The flow situation and forcing were tested for two different log-transmissivity fields. Both have a spherical anisotropic variogram with a mean transmissivity of −2.0 $\log(m^2/s)$, a range equal to 600 m in the *y*-direction and 2 km in the *x*-direction. One field is moderately heterogeneous with a variance of 0.1886 $\log_{10}(m^2/s)^2$ (variance in terms of ln*T* is 1.0). The strongly heterogeneous field has a variance of 0.509 $\log_{10}(m^2/s)^2$ (variance in terms of ln*T* is 2.7). The reference log-transmissivity field was generated with sequential Gaussian simulation using the code GCOSIM3D (Gómez-Hernández and Journel, 1993). Fig. 1 shows the two reference log*T* fields.

Both the case with a mildly heterogeneous log*T* field and a strongly heterogeneous log*T* field have the same spatio-temporally variable recharge rate field. The generation of the spatio-temporally variable recharge rates consisted of two steps. First, a time series of monthly average recharge rates was estimated. Second, spatially variable monthly recharge rate fields were created by

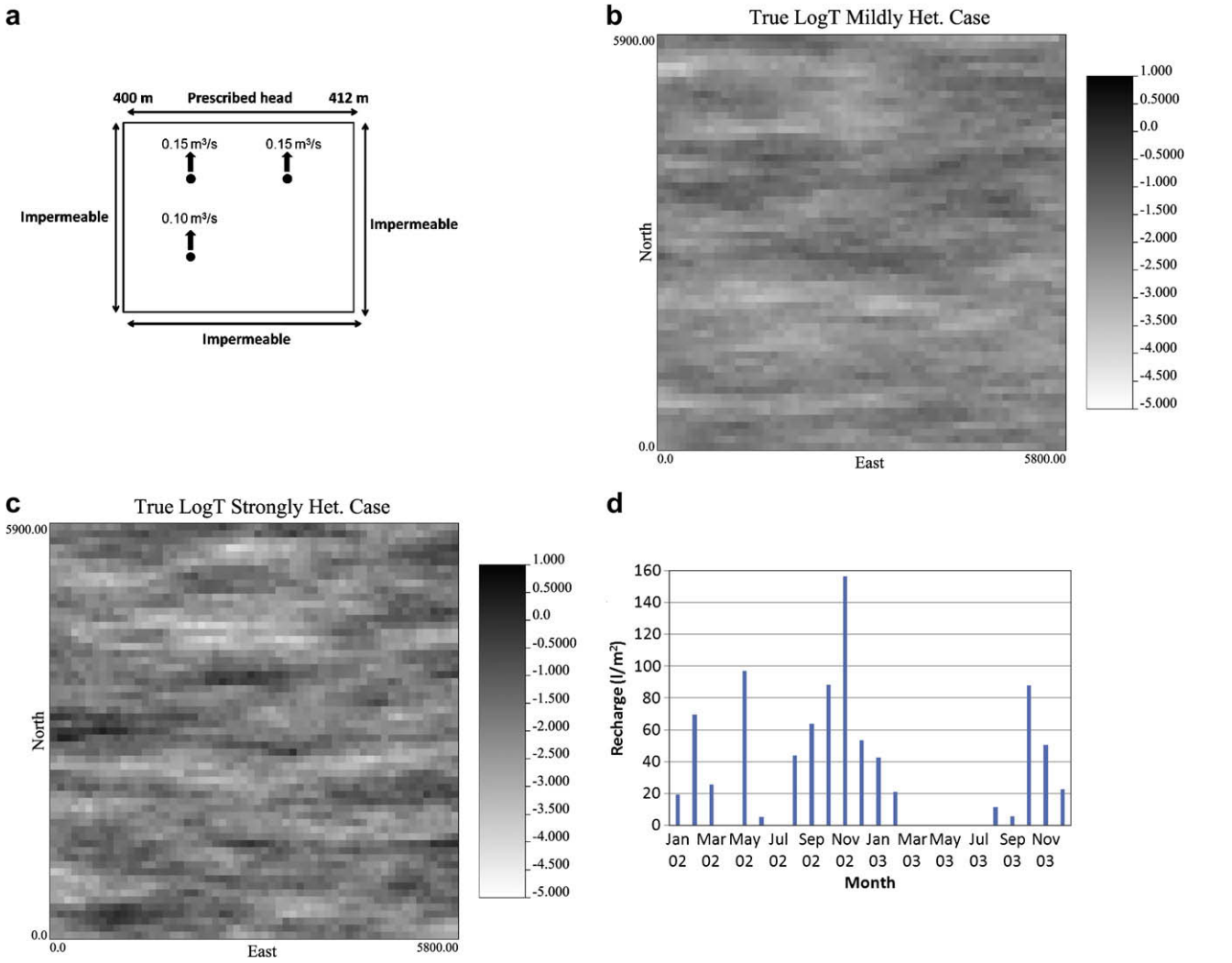

Fig. 1. Set-up for synthetic calibration experiments with the reference log-transmissivity field for the mildly and strongly heterogeneous case and the input average monthly recharge rates.

applying a spatially correlated perturbation to the mean (time-dependent) recharge rate. The monthly average recharge rates were estimated on the basis of observed data in the years 2002 and 2003 at the meteorological station of Zurich-Reckenholz, Switzerland, in order to generate realistic time series of the recharge rate. Potential evapotranspiration according to Penman-Monteith was calculated using as input measured global radiation, air temperature, wind speed and vapour pressure. These observational data were measured, quality checked and stored by the Swiss Federal Office of Meteorology and Climatology MeteoSwiss. Actual evapotranspiration was calculated from precipitation measurements in Zurich-Reckenholz, the calculated potential evapotranspiration, and a simple soil water balance model that takes into account transpiration by vegetation (in this case grass), bare soil evaporation and soil properties (in this case sandy loam). From the precipitation measurements and actual evapotranspiration, recharge rates were obtained, which were grouped into monthly values. The calculated monthly mean recharge rate values, and a spherical variogram for recharge rate with a range of 600 m, were used to generate spatially variable recharge rate fields for each of the 24 months (January 2002–December 2003), with help of GCOSIM3D (Gómez-Hernández and Journel, 1993). The sill for the recharge rate variogram differed for each month and was larger for months with more recharge and lower for months with less recharge. The created spatio-temporally variable recharge rate fields were assumed to be known for the calibration experiments. Fig. 1 shows the 24 monthly recharge values for each of the months.

The reference recharge rate and transmissivity fields were used to generate the reference hydraulic head fields over the two years period, for both the moderately and strongly heterogeneous transmissivity fields. The storativity coefficient for both the mildly and strongly heterogeneous $\log T$ field is equal to 0.1. The hydraulic head field is sampled at 28 locations which serve as monitoring locations. The monitoring locations are laid out on a regular grid that covers the aquifer. The measurement data that are sampled from the reference fields are error free, but in the data assimilation procedure it is assumed that they have a normally distributed measurement error with expectation zero and a standard deviation of 4 cm. Hydraulic head data are assimilated each 6 days with EnKF, and SSC uses the same hydraulic head data to inversely calibrate the model.

Results for calibration experiments

SSC was applied to calibrate 100 equally likely log-transmissivity fields to the 2 years time series of hydraulic head data at 28 locations, for the two different cases (mildly and strongly heterogeneous $\log T$ field). First, 100 equally likely log-transmissivity fields were generated with GCOSIM3D, using the correct variograms (the same variograms as used for the generation of the reference transmissivity fields). In addition, the 100 equally likely $\log T$ fields were calibrated to the hydraulic head data. With INVERTO (version 2g (2008), earlier versions with less features are documented in Hendricks Franssen (2001)), in total 200 master blocks are laid on a regular grid, with a random starting point, that covers the aquifer. The master blocks have different positions in each of the stochastic realizations and also change position during the inverse conditioning of a stochastic realization. In order to accommodate to the anisotropy, the density of master blocks in the y-direction is larger than in the x-direction (one master block per three grid cells versus one master block per six grid cells). The calibration was terminated if the total objective function value decreased below 10 m^2. This takes into account that the hydraulic head data have a measurement error.

EnKF was used to assimilate sequentially the hydraulic head data. In this case, 500 stochastic realizations were used, in order to avoid filter inbreeding. If less stochastic realizations are used, filter inbreeding occurs, which can be observed on the basis of the innovation statistics. In addition, $\log T$ was updated each month (each five time steps; 24 times in total) with help of the augmented state vector approach ($\alpha = 0.1$). Larger α values also resulted in an underestimation of the posterior variance. The fact that values for the number of stochastic realizations, the updating frequency of $\log T$ and α have to be chosen, results in additional CPU time, which in fact also should be taken into account in assessing the overall needed CPU time. Therefore, the issue of determining the values for these parameters in relation to the CPU time will be addressed in the discussion.

The results of EnKF and SSC are also compared with unconditional simulations. In that case, 500 unconditional simulations, generated by GCOSIM3D, are used as input for the transient groundwater flow equation.

The results of SSC, EnKF and the unconditional simulations are analyzed with the help of two criteria: the average absolute error (*AAE*) and the average ensemble standard deviation (*AESD*). The average absolute error is the absolute difference between the ensemble average of hydraulic head or log-transmissivity at a certain grid cell and certain time step and the true value of hydraulic head or log-transmissivity at that grid cell and time step, averaged over all grid cells and time steps. The average ensemble standard deviation is the result of evaluating the ensemble standard deviation (of hydraulic head or log-transmissivity) at all grid cells and all time steps, and averaging it

$$AAE(X) = \frac{1}{NMT}\sum_{j=1}^{M}\sum_{t=1}^{T}\sum_{i=1}^{N}|X_{i,j,t} - X_{i,t}^{ref}| \tag{11}$$

$$AESD(X) = \sqrt{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(X_{i,t} - \overline{X}_{i,t})^2} \tag{12}$$

where N is the number of grid cells, T is the number of time steps, M is the number of stochastic realizations, X is hydraulic head or log-transmissivity, the superscript *ref* refers to the reference field and the overbar indicates ensemble averaging.

Table 1 summarizes the main results of the calibration. For the mildly heterogeneous case, SSC and EnKF yield very similar results. Compared with unconditional simulations, SSC reduces the average absolute error of the log-transmissivity field by 22.4%, EnKF by 20.2%. On the other hand, EnKF has a slightly better characterisation of the hydraulic head field (76.2% reduction of *AAE*) than SSC (75.2% reduction of *AAE*). The uncertainty estimation does not show large differences between the two methods, taking into account that the estimated variances are more influenced by

Table 1
Evaluation of the calibration experiments. The evaluation measures are given for unconditional realizations, inverse conditioning with SSC and EnKF, for both the mildly and strongly heterogeneous cases. For details on the evaluation measures see "Results for calibration experiments".

	Unconditional	SSC	EnKF
Mildly heterogeneous case			
AAE(Y)	0.361	0.280	0.288
AESD(Y)	0.435	0.375	0.358
AAE(h)	0.727	0.180	0.172
AESD(h)	0.592	0.157	0.210
Strongly heterogeneous case			
AAE(Y)	0.542	0.469	0.495
AESD(Y)	0.707	0.593	0.583
AAE(h)	0.664	0.200	0.228
AESD(h)	1.023	0.305	0.334

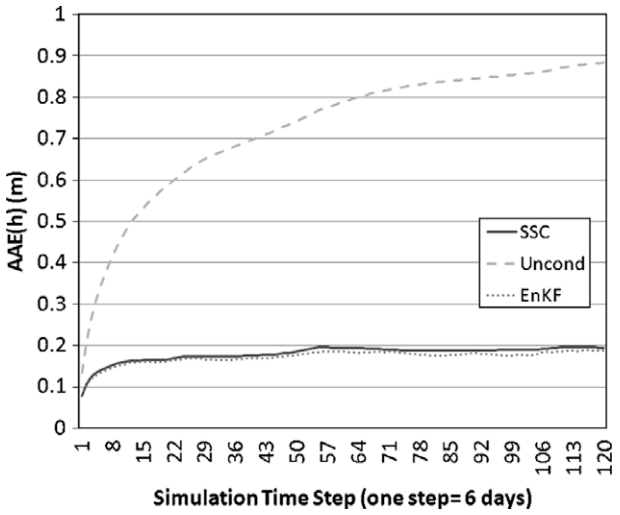

Fig. 2. Temporal evolution of $AAE(h)$ for the mildly heterogeneous calibration case. Shown are results for the unconditional simulations, inversely conditioned realizations with SSC and conditioned realizations with EnKF.

sampling fluctuations. For transmissivity, EnKF has a slightly lower standard deviation than SSC, whereas for hydraulic head it has a somewhat higher standard deviation. Fig. 2 shows the evolution of $AAE(h)$ for the unconditional simulations, SSC and EnKF. As these are the results of a calibration experiment, it means that for SSC all measurement data were used simultaneously to update the transmissivities, and the calculated heads are the result of this updating on the basis of all measured head data. The very similar performance of SSC and EnKF, can be noticed in the figure. For the strongly heterogeneous case, somewhat larger differences are observed. The average absolute error of the log-transmissivity field is reduced to a lower value, compared with the unconditional simulations, when SSC is used (−13.5%) than with EnKF (−8.7%). For the characterisation of the hydraulic head field the differences between SSC and EnKF are smaller; SSC reduces the average absolute error of hydraulic head by 69.9%, EnKF by 65.7%. Note that for hydraulic head the *AAE* is averaged over multiple time steps. As the characterisation of the log$T$ field improves in the case of EnKF over time, especially in the beginning of the assimilation process the log$T$ realizations are worse than the inverse conditioned log$T$ realizations by SSC. As a consequence, in spite of assimilating hydraulic head data and adapting the states, these poorer log$T$ fields result also in a worse characterisation of the hydraulic head field. As later in the assimilation the log$T$ fields conditioned by SSC and EnKF differ less in quality (as measured by $AAE(Y)$), also the hydraulic head fields are less different. SSC and EnKF differ only

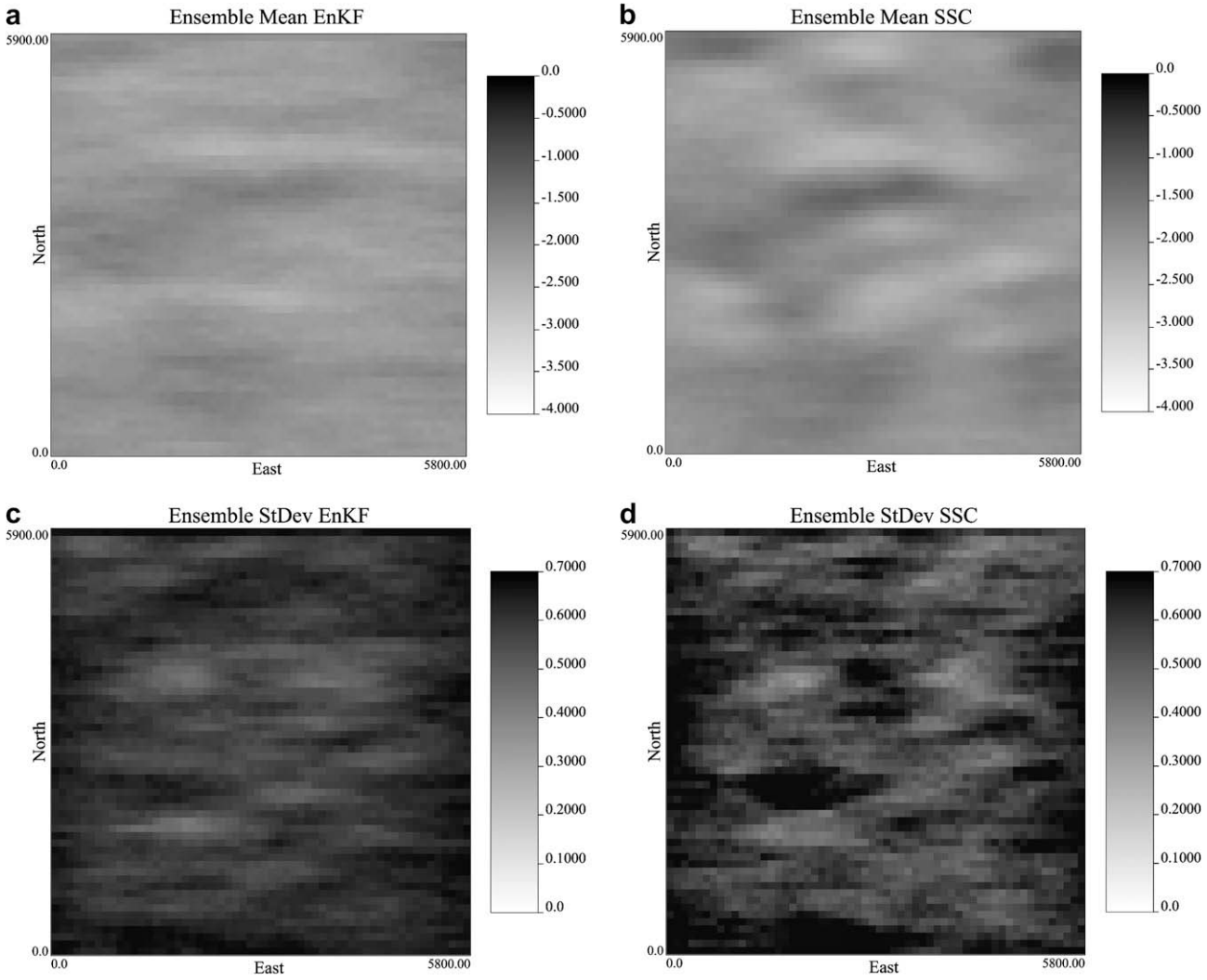

Fig. 3. Estimated ensemble average log$T$ fields and ensemble log$T$ standard deviation fields for the strongly heterogeneous case. The results for the inverse conditioned simulations with SSC and the conditional simulations with EnKF are shown.

Table 2
CPU time needed for conditioning to two years of hydraulic head data, both for SSC (100 realizations) and EnKF (500 realizations)

	SSC	EnKF
Mildly heterogeneous case	Ca. 14 days (MacOSX)	Ca. 4 h (MacOSX)
Strongly heterogeneous case	Ca. 25 days (Linux)	Ca. 7.5 h (Linux)

marginally when looking to the average ensemble standard deviation. This illustrates that 500 stochastic realizations were enough to avoid filter inbreeding, in case of EnKF. Fig. 3 shows the estimated ensemble average log$T$ fields for SSC and EnKF. In case of EnKF, this is the average log$T$ field from the last time step. The average log$T$ field converges in the EnKF procedure first fast, and later slowly to the final result. After ten updates, the average log$T$ field shows only very small changes, in this case. The improved resemblance with the true log$T$ field can be noticed, as well as the fact that SSC and EnKF detect the same high and low transmissivity zones. Nevertheless, also clear differences between the log$T$ fields are obvious from the figures, especially for the ensemble standard deviations.

In order to evaluate the performance of EnKF and SSC, it is important to consider also the CPU time required. SSC was found to perform very well in strongly heterogeneous formations compared to other inverse methods (Zimmerman et al., 1998; Hendricks Franssen et al., submitted), but also requires a lot of CPU time, which makes its use prohibitive for a larger number of nodes. Table 2 summarizes the CPU time needed for the inverse conditioning in the synthetic study. The differences in CPU time needed are remarkable. For the mildly heterogeneous case, where both methods give very similar results, it is a factor of 84 (the calculations were performed using a single processor of a MacOSX machine from the year 2005). For the strongly heterogeneous case, where SSC only gives slightly superior results compared with EnKF, it is a factor of 80 (the calculations were performed using a single processor Linux PC from the year 2002). For mildly heterogeneous formations, alternative inverse methods which need a very limited amount of CPU time, might give results comparable to SSC as well. However, this is merely the case for steady-state flow conditions and simple aquifer configurations and boundary conditions.

Synthetic set-up of two prediction experiments

For the two different cases (mildly and strongly heterogeneous reference transmissivity field), two additional synthetic prediction experiments were carried out. The calibrated log$T$ fields were used as input for those prediction experiments.

In experiment 1 transient groundwater flow was again simulated for two years, starting with perfectly known initial conditions, and known boundary conditions. Initial and boundary conditions were the same as for the calibration experiment. As compared with the calibration experiments, all wells are shut down and the groundwater levels in the area are subjected to a perfectly known spatio-temporally variable recharge rate. However, for these two years the recharge rates are different as compared with the calibration experiments. The spatio-temporally variable recharge rates were now estimated on the basis of observed data in the years 2004 and 2005. This estimation consisted again of two steps. First, average monthly recharge rate values were calculated for January 2004–December 2005, using data observed at the meteorological station of Zurich-Reckenholz, Switzerland with the help of the same procedure explained before in "Synthetic set-up of calibration experiments". These calculated monthly mean recharge rate values were perturbed to yield for each month a spatially variable recharge rate field. The applied perturbation shows spatial autocorrelation, which is modelled with a spherical variogram with a range of 600 m. The sill for the recharge rate variogram differed for each month and was larger for months with more recharge and lower for months with less recharge. The spatially autocorrelated perturbations were again simulated with GCOSIM3D (Gómez-Hernández and Journel, 1993). Fig. 4 shows the 24 monthly recharge values for each of the months. The transient groundwater flow equation was solved with INVERTO (applied in its forward mode) using a finite differences solution as detailed in the beginning of "Results for calibration experiments". The response in terms of hydraulic head time series was observed for each of the two cases (mildly heterogeneous log$T$ field and strongly heterogeneous log$T$ field) for the three sets of generated log$T$ realizations: unconditional realizations, realizations conditioned with SSC and realizations conditioned with EnKF. These simulated hydraulic head time series are compared with the "true" hydraulic head time series for the reference fields, according to Eqs. (11) and (12).

In experiment 2 steady-state groundwater flow was simulated with perfectly known, and unchanged, boundary conditions and a mean steady-state recharge rate of 360 mm/year. The recharge rate was spatially variable with the same geostatistical model as used in the first prediction experiment. The spatially variable recharge rate field was simulated with GCOSIM3D. Compared with the calibration experiment two of the three wells were shut down, whereas the well at (20,40) was pumping with a rate of 0.15 m^3/s. Steady-state groundwater flow was solved with a block-centred finite differences scheme, using INVERTO to solve the forward problem. For this well the catchment was calculated with help of a particle tracking algorithm as implemented in the software 3DTRANSP (Hendricks Franssen, 2001). Advective transport was solved with particle tracking using bilinear interpolation of the velocity field. At the centre of each of the grid cells of the model a particle was released and subjected to advective transport. For each of the grid cells it was recorded whether the particle reached the pumping well (and formed part of the well catchment) or left the system through a boundary (and did not form part of the well catchment). The well catchments for the two reference fields (mildly and strongly heterogeneous log$T$ fields) were calculated. Also for each of the two cases (mildly heterogeneous log$T$ field and strongly heterogeneous log$T$ field) and the three sets of generated log$T$ realizations (unconditional realizations, realizations conditioned with SSC and realizations conditioned with EnKF) the ensemble of well catchments was calculated. At each of the grid cells the average capture probability was calculated for these different cases. The true well catch-

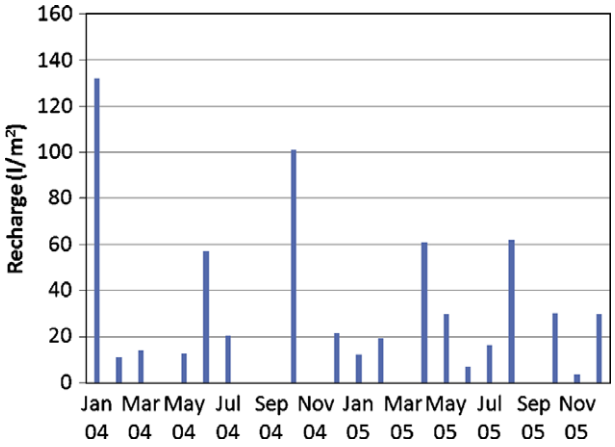

Fig. 4. Average monthly recharge rates for the first prediction experiment with new transient flow conditions.

ments and the calculated stochastic well catchment were compared according to:

$$AAE(CP) = \frac{1}{NM}\sum_{j=1}^{M}\sum_{i=1}^{N}|CP_{i,j} - CP_i^{ref}| \quad (13)$$

where CP is the capture probability, which for the reference fields is 0 or 1 at a location i. Fig. 5 shows the reference well catchments for both the mildly and strongly heterogeneous case.

Results for prediction experiments

Transient groundwater flow

For the two cases (mildly and strongly heterogeneous log$T$) and the three sets of equally likely stochastic realizations (unconditional, inverse with SSC, inverse with EnKF) time series of hydraulic head were calculated, as response to the new flow conditions. The time series of hydraulic head were obtained by solving the forward problem. The simulated hydraulic head time series were compared with the reference ones, according to the Eqs. (11) and (12). Table 3 summarizes the results. The fields obtained from inverse conditioning reproduce the hydraulic head time series in the prediction case better than the unconditional realizations do. For the mildly heterogeneous case the improvement is 46–54% (in terms of reduction of average absolute error of hydraulic head), for the strongly heterogeneous case 36–39%. Recall that for the calibration experiments the $AAE(h)$ reductions are stronger (66–76% compared with unconditional realizations), which is typical and illustrates the challenge of predicting the flow for a new situation, on the basis of calibrated transmissivity distributions. The most important observation is that the differences between EnKF and SSC for these prediction experiments are small. Whereas for the mildly heterogeneous case SSC gives a slightly better result in terms of $AAE(h)$ reduction compared with unconditional simulations (54% versus

Table 3
Evaluation measures for the first prediction problem (prediction of transient groundwater flow with new flow conditions). Compared are unconditional stochastic realizations and the sets of inverse conditioned realizations with SSC or EnKF, for both the mildly and strongly heterogeneous case. For details on the evaluation measures see "Results for calibration experiments".

	Unconditional	SSC	EnKF
Mildly heterogeneous case			
$AAE(h)$	0.433	0.200	0.233
$AESD(h)$	0.367	0.478	0.219
Strongly heterogeneous case			
$AAE(h)$	0.346	0.223	0.210
$AESD(h)$	0.574	0.535	0.328

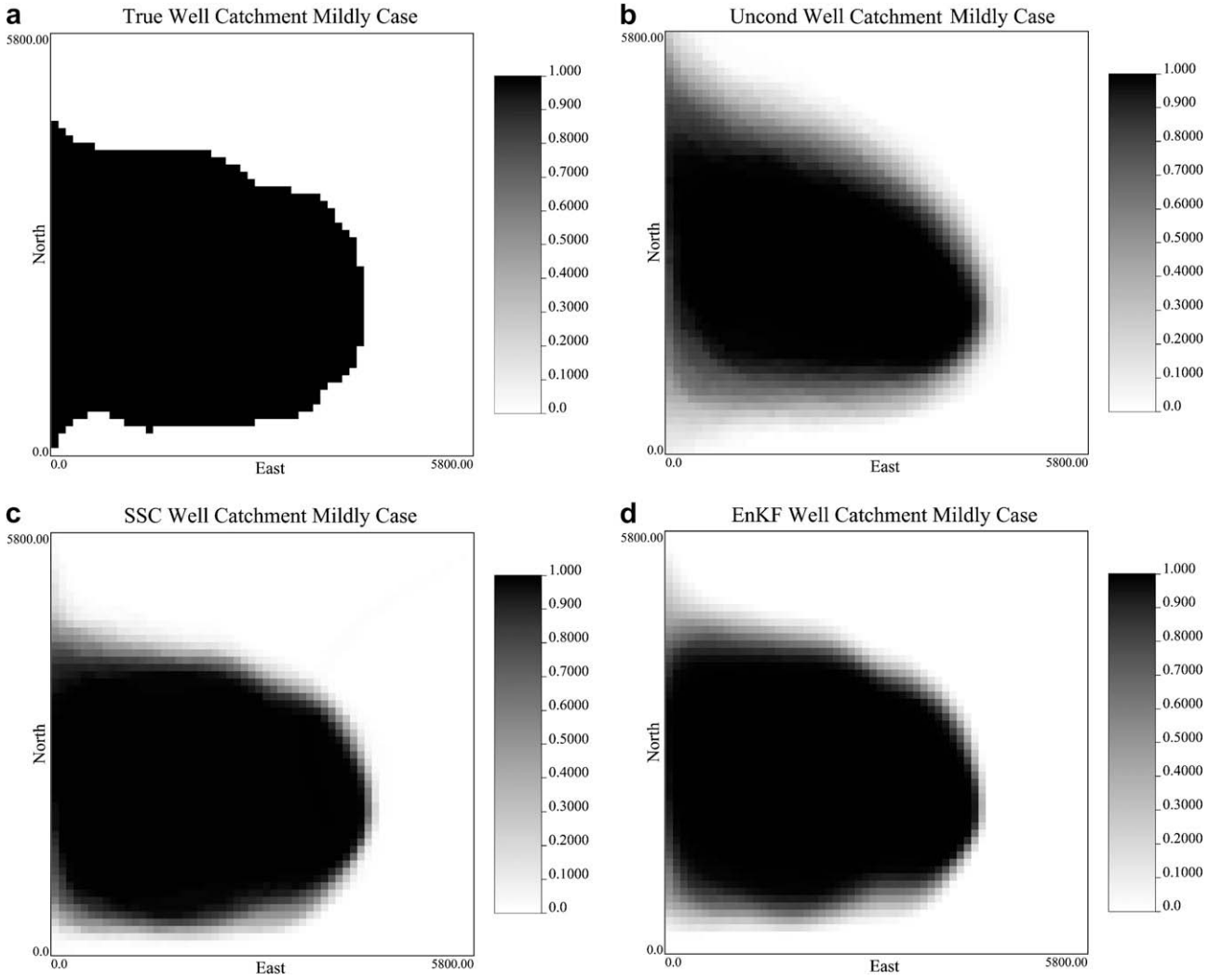

Fig. 5. True well catchment and predicted well catchments on the basis of unconditional simulations, inverse simulations with SSC and conditional simulations with EnKF. Results are given both for the mildly and strongly heterogeneous case.

46% for EnKF), somewhat surprisingly the opposite is the case for the strongly heterogeneous case (36% versus 39% for EnKF). We attribute these variations to sampling fluctuations, and conclude that in these prediction experiments EnKF performs as well as SSC, also for the strongly heterogeneous field. The obtained ensemble standard deviations ($AESD(h)$) are somewhat difficult to interpret. Both for the mildly and strongly heterogeneous log$T$ fields EnKF yields a considerably smaller ensemble standard deviation than SSC. For the mildly heterogeneous case EnKF yields a $AESD(h)$ more than 50% lower than SSC, whereas for the strongly heterogeneous case this difference is also around 40%. Fig. 6 gives the results for the strongly heterogeneous case, and shows that EnKF gives especially at the end of the simulation period better predictions than SSC. Fig. 6 also illustrates that during the whole simulation period SSC has a higher prediction uncertainty than EnKF. However, EnKF does not underestimate the posterior variance related to filter inbreeding:

(1) With an underlying multi-Gaussian transmissivity field, the hydraulic head distribution often does not deviate strongly from the normal distribution. For normally distributed variables the ratio between $AESD(h)/AAE(h)$ is expected to be 1.25. For EnKF, there is no systematic deviation from this ratio, which is 0.94 for the mildly heterogeneous case and 1.56 for the strongly heterogeneous case. On the other hand, for SSC these ratios are much larger (2.39 for the mildly heterogeneous case and 2.40 for the strongly heterogeneous case).
(2) The ratios $AESD(h)/AAE(h)$ for EnKF are very similar to the ratios for unconditional simulations. For unconditional realizations the ratio is 0.85 for the mildly heterogeneous case and 1.66 for the strongly heterogeneous case. This shows that the underlying Multi-Gaussian log$T$ distributions together with the imposed flow conditions give local probability distributions for hydraulic head which, altogether, give ratios for $AESD(h)/AAE(h)$ close to the expected ratios for the normal distribution. There is no reason to expect that this should be very different for inversely conditioned realizations. Therefore, the high ratios for SSC are surprisingly large.
(3) For the mildly heterogeneous case, the SSC conditioned realizations have a larger $AESD(h)$ than the unconditional realizations. For the strongly heterogeneous case the SSC inverse conditioned realizations have only a slightly smaller $AESD(h)$ than the unconditional ones. In both cases the central prediction by SSC is much better than the unconditional prediction. This again points to the fact that the predicted ensemble standard deviations by EnKF are not too small, but the predicted ensemble standard deviations by SSC are too large.

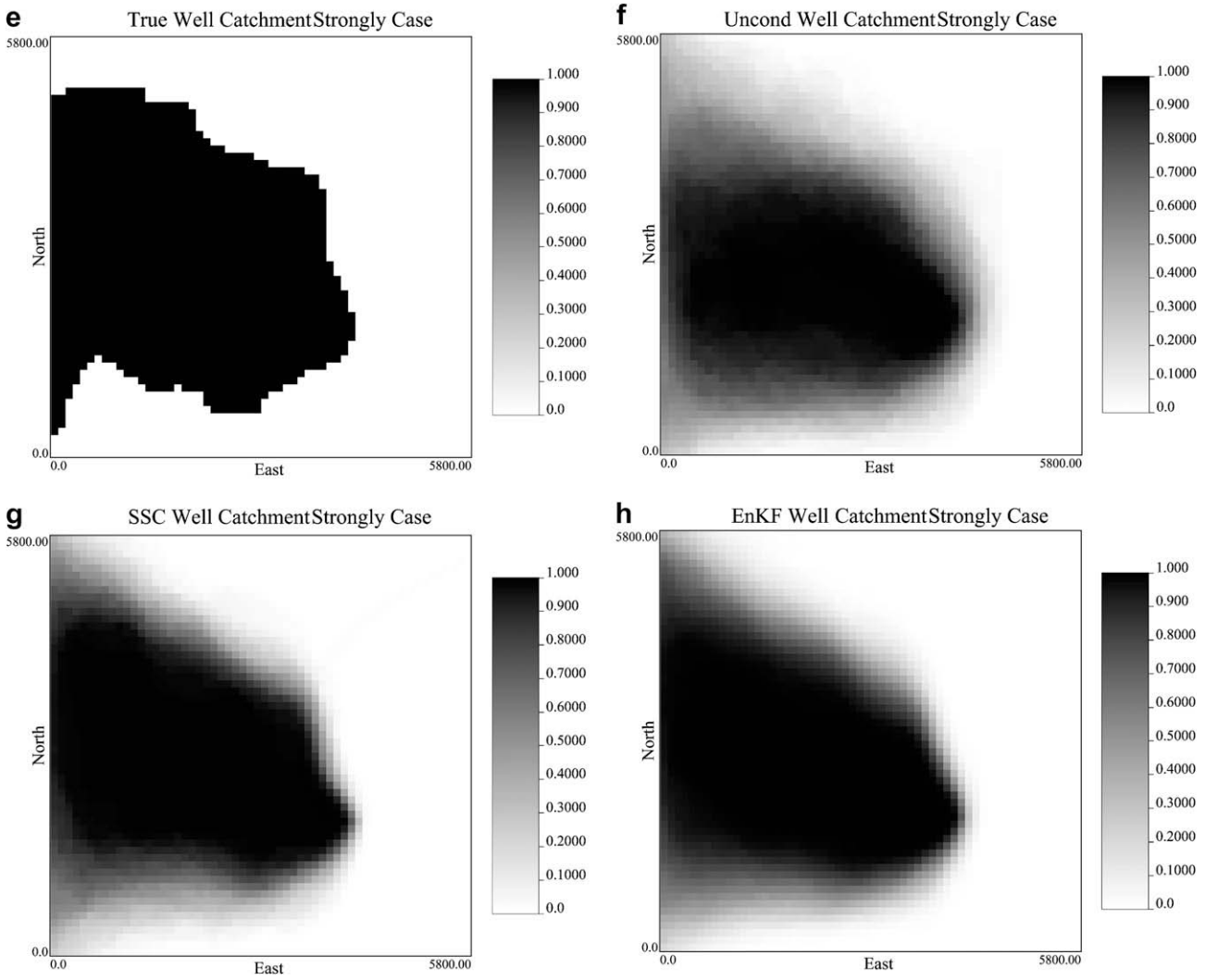

Fig. 5. *(continued)*

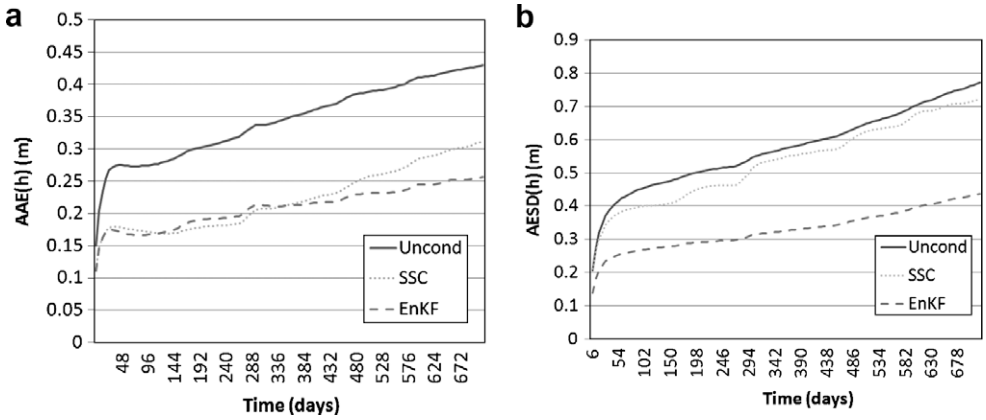

Fig. 6. Temporal evolution of $AAE(h)$ and $AESD(h)$ for the first prediction experiment (strongly heterogeneous case) with new transient flow conditions.

It is a somewhat surprising observation that SSC gives a too large prediction uncertainty, as for the calibration experiments a strong variance reduction was observed, and also in many other synthetic experiments in the past, no overestimation of the variance was found. An explanation seems to be that no $\log T$ data are used for conditioning. As a consequence, locally the $\log T$ field was perturbed heavily to match the hydraulic head data. For the recharge conditions in the calibration experiment this did not give an increased hydraulic head variance, but for the changed recharge conditions in the prediction experiment, for some realizations at some locations, very high hydraulic head values were calculated at locations with low transmissivity values. This contributed to an increased ensemble standard deviation for hydraulic head. However, the same could have happened for EnKF, so that currently our best explanation for the overestimation of the SSC variance is related with sampling fluctuations, due to the inclusion of a very few extreme realizations in the ensemble.

Advective transport towards a well

For the two cases (mildly and strongly heterogeneous $\log T$) and the three sets of equally likely stochastic realizations (unconditional, inverse with SSC and with EnKF) the capture zone probability distribution was estimated. The simulated capture zone probability distribution was compared with the reference well catchment, according to the Eq. (13). Table 4 summarizes the results. Fig. 5 shows the average well capture zones for the mildly and strongly heterogeneous case and the three different sets of equally likely stochastic realizations, together with the reference ("true") well catchments. Again, it is found that the differences between SSC and EnKF are limited, although SSC performs in this case somewhat better than EnKF. For the mildly heterogeneous case, SSC improves the characterisation of the well catchment with 55% (compared with unconditional simulations, in terms of $AAE(CP)$), EnKF with 43%. For the strongly heterogeneous case the differences are very small: 36% improvement for SSC, 35% improvement for EnKF. Fig. 5 also illustrates that EnKF and SSC yield very similar results. The ensemble well catchments are more like the true catchment than the unconditional realizations.

Table 4
Evaluation measures for the second prediction problem (prediction of advective transport towards a well). Compared are unconditional stochastic realizations and the sets of inverse conditioned realizations with SSC or EnKF, for both the mildly and strongly heterogeneous case. For details on the evaluation measures see "Synthetic set-up of two prediction experiments".

	Unconditional	SSC	EnKF
Mildly heterogeneous case			
$AAE(CP)$	0.103	0.046	0.059
Strongly heterogeneous case			
$AAE(CP)$	0.141	0.091	0.092

Discussion

We would like to re-stress that EnKF is an interesting option for the conditioning of groundwater flow models to state information. First, because the methodology is very suited to incorporate data in real-time, which are increasingly available from on-line sensors. Second, because it is more CPU efficient than MC type inverse modelling. In this study, we found that EnKF reduced the CPU time, compared with SSC, by a factor of 80. Third, because EnKF is a flexible framework to incorporate different sources of uncertainty, like measurement uncertainty, parameter uncertainty, forcing uncertainty and uncertainty with regard to initial conditions. The comparison of EnKF with MC type inverse modelling focused on parameter uncertainty, which is a source of uncertainty that EnKF traditionally did not handle. Nevertheless, in this synthetic study very similar results for EnKF and SSC were obtained. It could be expected that in case of multiple sources of uncertainty, including forcing uncertainty, EnKF could outperform SSC. Nevertheless, EnKF has also drawbacks, which were already addressed in the introduction (see "Introduction") and which will be discussed into more detail below.

For non-normal state distributions and non-linear dynamics EnKF is expected to give sub-optimal results. In this synthetic study it was found that for a strongly heterogeneous $\log T$ field EnKF still yielded good results, and only the characterisation of the unknown $\log T$ field was worse as compared with SSC. However, there is potential to further improve the performance of EnKF:

- The use of an increasing number of stochastic realizations allows to reduce the problems with filter inbreeding, however, at the cost of an increased CPU time. Therefore, it is important to sample the high dimensional space of stochastic variables efficiently, and avoid, for instance, the random drawing of stochastic realizations that are clustered (close together) in this high dimensional space. Eigenvector techniques (e.g., Zhang et al. (2007)) and resampling strategies (e.g., Wen and Chen, 2006) are suited for this. In this synthetic study, 500 stochastic realizations were sampled randomly, without using for instance resampling. Hendricks Franssen and Kinzelbach (2008) showed that with resampling techniques 100 realizations could be as efficient as 200

stochastic realizations sampled randomly. Therefore, it is likely that the performance of 500, selectively sampled, stochastic realizations would be better than 500 randomly sampled stochastic realizations.
- Adaptive covariance matrices allow to correct the underestimation of the variance, which can be observed by the innovation statistics. This reduces the problems with filter inbreeding and might reduce the number of needed realizations. Currently we have implemented an adaptation of the covariance matrices on the basis of innovation statistics, but more sophisticated methods (e.g., Reichle et al., 2008) might result in better results and reduced filter inbreeding.
- Until now, the parameters were updated with EnKF on the basis of one fixed value for α. There is potential to optimize the value of α each time that $\log T$ is updated. Probably this improves the estimated $\log T$ parameters. However, this will also result in extra CPU costs. Linear search techniques could probably limit the required number of additional evaluations of the flow equation, for optimizing α.

Here EnKF and SSC were compared for a 2D confined groundwater flow problem with Multi-Gaussian log-transmissivity random fields and a known geostatistical model. We expect that EnKF and SSC still give very similar results for 3D phreatic groundwater flow problems. Already for the examples presented in this paper, hydraulic head showed a strongly skewed distribution close to pumping wells. The groundwater levels will most probably not show larger deviations from normality for 3D phreatic problems. On the other hand, it might make an important difference if log-transmissivity has a Non-Multi-Gaussian distribution instead of a Multi-Gaussian distribution. It is expected that EnKF could perform increasingly worse (as compared to SSC) for the updating of an underlying log-hydraulic conductivity field that exhibits strongly Non-Multi-Gaussian properties. The performance of EnKF under such conditions should be the topic of future investigations. Finally, in reality the geostatistical model is not known and the sensitivity of the methods to an erroneous geostatistical model would be important to know. Hendricks Franssen (2001) found that the results obtained by SSC are not strongly influenced by erroneous variogram parameters, but a mistake in the conceptual geological model (for instance a Multi-Gaussian model for log-transmissivity instead of a Non-Multi-Gaussian model) has an important negative impact on the results (Kerrou et al., 2008). It is speculative to predict whether for EnKF the impact of an erroneous geostatistical model is less or more severe than for SSC. Probably the negative impact would be the same for SSC and EnKF, but extended experiments are needed to investigate this.

Altogether we found very promising results: in the prediction experiments EnKF gave similarly good results as SSC. At the same time, the CPU time needed for the calibration was a factor 80 lower. However, this was for a certain parameter set (number of $\log T$ realizations, α, updating frequency of $\log T$). We selected this parameter set on the basis of more experiments, and therefore a criticism could be, that in reality the CPU time needed for EnKF would be larger. However, we believe that this is not the case for the following reasons:

- For SSC also some parameter values need to be defined, like the minimum step size to start the optimization along the updating direction vector and the number of master blocks. With experience it was possible to obtain default values/rules for the values of those parameters so that in practice SSC only needs to be executed once for a certain problem. In the same way, this knowledge can also become available if more experience with EnKF is acquired.
- With respect to the number of stochastic realizations: it is found that with 500 stochastic realizations also in the strongly heterogeneous case no filter inbreeding occurs. Therefore it is likely that in many applications 500 stochastic realizations suffice. The reduction of CPU time with a factor of 80 is based on the conditioning of 500 stochastic realizations for EnKF, versus 100 stochastic realizations for SSC. On the other hand, filter inbreeding can easily be detected on the basis of innovation statistics, for instance by the ratio $AESD/AAE$. It can often already be observed in the beginning of the assimilation experiments. Therefore, the CPU costs of an additional test with a smaller number of stochastic realizations (e.g., 250) can be very limited if the execution is stopped if filter inbreeding becomes apparent. SSC, on the other hand, is not affected by filter inbreeding and although sampling fluctuations are reduced if a larger number of stochastic realizations is used, results are not principally worse for a smaller amount of realizations. The needed CPU time by SSC increases linearly as function of the number of stochastic realizations.
- With respect to the damping parameter α: also this parameter can be evaluated by means of innovation statistics; a too large α might yield filter inbreeding. On the basis of our current experiences, we advice to use $\alpha = 0.1$ as the default value. For mildly heterogeneous $\log T$ formations the optimum α value could be larger. Additional experience or automatic procedures should give in the future better insight in the optimum α values. Nevertheless, the a priori value of $\alpha = 0.1$ together with the check for the occurrence of filter inbreeding with help of innovation statistics, are not expected to cost much more CPU time as reported in this study.
- With respect to the updating frequency of log-transmissivities: it is recommended to frequently update $\log T$. However, after a certain time, if the same stresses apply to the system, no additional information can be gained from the head measurements. This can be observed by monitoring the changes of the $\log T$ field.
- The simulations with EnKF were still based on the random sampling of stochastic realizations. As the sampling can be more efficient (with little additional CPU costs), the same accuracy could be obtained with less realizations.

Another point is whether the observed reduction of needed CPU time also holds for other settings. One factor of influence is the number of observation locations. In this application, with 25 measurements, the matrix inversion of Eq. (8) (the matrix has sizes of $n \times n$ where n is the number of measurement data) required only a very limited amount of CPU time, as compared to the forward runs for the stochastic realizations. We found that also for 100 measurement data this is still the case. However, for a much larger number of measurement data the matrix inversion might suppose a larger amount of needed CPU time than the forward runs. Therefore, for a very large amount of measurement data, the CPU advantage of EnKF would reduce, as the needed CPU time by SSC hardly depends on the amount of measurement data. On the other hand, for a very large amount of measurement data, the data from a limited area around the point of interest could be assimilated, without considering all measurement data. A second factor of influence is the problem size. It is expected that for much larger problems (here the problem only consisted of 3600 nodes) EnKF will have approximately the same advantage over SSC as for this smaller problem. The main CPU cost for both EnKF and SSC is related to solving the forward problem, and SSC has, independent of the problem size, to solve it many more times than EnKF, because of its iterative approach.

In summary, the CPU time reduction (when comparing EnKF with SSC) in this synthetic study is also expected to hold in other applications.

For the reasons given before, EnKF is interesting for the conditioning of transient groundwater flow models with many nodes. Inverse type MC methods become extremely CPU intensive for groundwater flow models with many nodes and time steps. Currently, we already developed a new code, EnKF-SPRING, based on finite elements and able to handle unsaturated flow and interactions with rivers. EnKF-SPRING is used to assimilate daily piezometric head data to update a 3D unsaturated–saturated groundwater flow including river–aquifer interaction. In that case study, 100 stochastic realizations are processed through a groundwater flow model that contains nearly 200,000 elements and assimilates daily piezometric head data for more than 600 days. This conditioning takes 10 days on a single processor Linux PC (from the year 2007). This further stresses the relevance and potential of EnKF for conditioning large integrated hydrological models to state information.

Conclusions

EnKF is interesting for the conditioning of groundwater flow models to state information. The three main reasons are that (1) it is very flexible for real-time modelling and incorporation of data from on-line sensors, (2) CPU efficient as compared with MC type inverse modelling techniques and (3) flexible with regard to the handling of multiple sources of uncertainty jointly. However, EnKF is expected to perform worse than MC type inverse modelling for more strongly non-linear problem and non-normal state distributions. Therefore, in a number of synthetic experiments, both for mildly and strongly heterogeneous log$T$ fields, the performance of EnKF and sequential self-calibration (SSC) were compared. It was found that for the calibration experiments EnKF and SSC yielded very similar results, both in terms of average absolute errors and average ensemble standard deviations. For the strongly heterogeneous log$T$ field SSC gave somewhat better log$T$ estimates than EnKF. On the other hand, EnKF needed around a factor of 80 less CPU time than SSC in both cases. In addition, the calibration results were used in two prediction experiments, one on the prediction of transient groundwater flow with a different time variable forcing, and a second one on the prediction of advective transport towards a well. For the two prediction experiments, which were carried out both for the mildly and strongly heterogeneous log$T$ field, EnKF and SSC yielded very similar results. Also for the strongly heterogeneous case SSC did not give better prediction results than EnKF. These results stress the interesting properties of EnKF for the conditioning of transient groundwater flow models, not only on-line and in real-time, but also off-line with historical time series.

Acknowledgments

The study was performed within the project "Real-time control of a well-field using a groundwater model", a cooperation between ETH Zurich, Water Supply Zurich and TK Consult Zurich. This project is funded by the Swiss Innovation Promotion Agency CTI under Contract No. 7608.2 EPRP-IW.

References

Aksoy, A., Zhang, F., Nielsen-Gammon, J.W., 2006. Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. Monthly Weather Review 134, 2951–2970.

Alcolea, A., Carrera, J., Medina, A., 2006. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. Advances in Water Resources 29 (11), 1678–1689.

Alcolea, A., Carrera, J., Medina, A., 2008. Regularized pilot points method for reproducing the effect of small scale variability, application to simulations of contaminant transport. Journal of Hydrology. doi:10.1016/j.jhydrol.2008.03.004.

Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimilation. Monthly Weather Review 129, 2884–2903.

Anderson, J.L., 2007. An adaptive covariance inflation error correction algorithm for ensemble filters. Tellus 59A, 210–224.

Bennett, A.F., 1992. Inverse Methods in Physical Oceanography. Cambridge University Press, New York.

Bishop, C.H., Etherton, B.J., Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. Monthly Weather Review 129, 420–436.

Blasone, R.S., Vrugt, J.A., Madsen, H., Rosbjerg, D., Robinson, B.A., Zyvoloski, G.A., 2008a. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling. Advances in Water Resources. doi:10.1016/j.advwatres.2007.12.003.

Blasone, R.S., Madsen, H., Rosbjerg, D. 2008b. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. Journal of Hydrology. doi:10.1016/j.hydrol.2007.12.026.

Burgers, G., van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the Ensemble Kalman Filter. Monthly Weather Review 126, 1719–1724.

Caers, J., Hoffman, T., 2006. The probability perturbation method: a new look at bayesian inverse modeling. Mathematical Geology 38 (1), 81–100.

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

Carrera, J., Neuman, S.P., 1986a. Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. Water Resources Research 22 (2), 199–210.

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

Carrera, J., Neuman, S.P., 1986c. Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data. Water Resources Research 22 (2), 228–242.

de Marsily G. 1978. De l'identification des systèmes hydrogéologiques. Thèse, Univ. Paris VI.

Doucet, A., de Freitas, N., Gordon, N.J., 2001. Sequential Monte Carlo Methods in Practice. Springer Verlag, New York.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods in forecast error studies. Journal of Geophysical Research 99 (C5), 10143–10162.

Evensen, G., 2004. Sampling strategies and square root analysis schemes for the EnKF. Ocean Dynamics 54, 539–560.

Foglia, L., Mehl, S.W., Hill, M.C., Perona, P., Burlando, P., 2007. Testing alternative ground water models using cross validation and other methods. Ground Water 45 (5), 627–641. doi:10.1111/j.1745-6584.2007.00341.x.

Gómez-Hernández, J.J., Journel, A.G., 1993. Joint sequential simulation of multi-Gaussian fields. In: Soares, A. (Ed.), Geostatistics Tróia'92, vol. 1. Kluwer Academic Publishers, pp. 85–94.

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

Gómez-Hernández, J.J., Hendricks Franssen, H.J., Cassiraga, E.F., 2001. Stochastic analysis of flow response in a three-dimensional fractured block. International Journal of Rock Mechanics & Mining Sciences 38, 31–44.

Hendricks Franssen, H.J. 2001. Inverse stochastic modelling of groundwater flow and mass transport. Ph.D. Thesis, Technical University of Valencia.

Hendricks Franssen, H.J., Gómez-Hernández, J.J., 2003. Impact of measurement errors in the sequential self-calibrated approach. Advances in Water Resources 26 (5), 501–511.

Hendricks Franssen, H.J., Kinzelbach, W., 2008. Real-time groundwater flow modelling with the Ensemble Kalman Filter: joint estimation of states and parameters and the filter inbreeding problem. Water Resources Research 44, W09408. doi:10.1029/2007WR006505.

Hendricks Franssen, H.J., Gómez-Hernández, J.J., Sahuquillo, A., Capilla, J.E., 1999. Joint simulation of transmissivity and storativity fields conditional to hydraulic head data. Advances in Water Resources 23 (1), 1–13.

Hendricks Franssen, H.J., Gómez-Hernández, J.J., Sahuquillo, A., 2003. Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data. Journal of Hydrology 281 (4), 281–295.

Hendricks Franssen, H.J., Stauffer, F., Kinzelbach, W., 2004. Joint estimation of transmissivities and recharges-application: stochastic characterization of well capture zones. Journal of Hydrology 294, 87–102.

Hendricks Franssen, H.J., Brunner, P., Makobo, P., Kinzelbach, W., 2008a. Equally likely inverse solutions to a groundwater flow problem including pattern information from remote sensing images. Water Resources Research 44, W01419. doi:10.1029/2007WR006097.

Hendricks Franssen, H.J., Alcolea, A., Riva, M., Bakr, M., van de Wiel, N., Stauffer, F., Guadagnini, A. submitted. A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments. Advances in Water Resources.

Hernandez, A.F., Neuman, S.P., Guadagnini, A., Carrera, J., 2003. Conditioning mean steady state flow on hydraulic head and conductivity through geostatistical inversion. Stochastic Environmental Research and Risk Assessment 7, 329–338. doi:10.1007/s00477-003-0154-4.

Hernandez, A.F., Neuman, S.P., Guadagnini, A., Carrera, J., 2006. Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media. Water Resources Research 42, W05425. doi:10.1029/2005WR004449.

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

Houtekamer, P.L., Michell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., Hansen, B., 2005. Atmospheric data assimilation with an Ensemble Kalman filter: results with real observations. Monthly Weather Review 133, 604–620.

Hu, L.Y., Blanc, G., Noetinger, B., 2001. Gradual deformation and iterative calibration of sequential stochastic simulations. Mathematical Geology 33 (4), 475–489.

Janssen, G.M.C.M., Valstar, J.R., van der Zee, S.E.A.T.M., 2006. Inverse modeling of multimodal conductivity distributions. Water Resources Research 42, (W03410). doi:10.1029/ 2005WR004356.

Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, United Kingdom.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., Ballabrera-Poy, J., 2007. 4-D-Var or ensemble Kalman filter? Tellus 59A, 758–773.

Kerrou, J., Renard, P., Hendricks Franssen, H.J., Lunati, I., 2008. Issues in characterizing connectivity and heterogeneity in non-multi-Gaussian media. Advances in Water Resources 31, 147–159. doi:10.1016/j.advwatres.2007.07.002.

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

Lavenue, A.M., de Marsily, G., 2001. Three-dimensional interference test interpretation in a fractured aquifer using the pilot point inverse method. Water Resources Research 37 (11), 2659–2675.

LaVenue, A.M., RamaRao, B.S., de Marsily, G., Marietta, M.G., 1995. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. 1. Theory and computational experiments. Water Resources Research 31 (3), 475–494.

Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. Water Resources Research 43, W07401. doi:10.1029/2006WR005756.

Madsen, H., Sörensen, J.T. 2007. Recent advances in data assimilation in large scale hydrodynamic and hydrological forecasting models, International workshop on Advances in Hydroinformatics 4–7 June 2007, Niagara Falls Canada.

Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources 28, 135–147.

Naevdal, G., Johnsen, L.V., Aanonsen, S.I., Vefring, E.H. 2003. Reservoir monitoring and continuous model updating using ensemble Kalman filter. Society of Petroleum Engineers 84372.

Reichle, R.H., Crow, W.T., Keppenne, C.L., 2008. An adaptive ensemble Kalman filter for soil moisture data assimilation. Water Resources Research 44, W03423. doi:10.1029/2007WR006357.

Sahuquillo, A., Capilla, J.E., Gómez-Hernández, J.J., Andreu, J. 1992. Conditional simulation of transmissivity fields honoring piezometric data. In: Blain, W. R., & Cabrera, E. (Eds.), Hydraulic Engineering Software IV, Fluid Flow Modeling, pp. 201–214.

Tonkin, M., Doherty, J., Moore, C., 2007. Efficient nonlinear predictive error variance for highly parameterized models. Water Resources Research 43, W07429. doi:10.1029/2006WR005348.

Valstar, J.R. 2001. Inverse modeling of groundwater flow and transport, Ph.D. Thesis, Delft University of Technology, the Netherlands.

Verlaan, M., Heemink, A.W., 1997. Tidal flow forecasting using reduced rank square root filters. Stochastic Hydrology and Hydraulics 11, 349–368.

Vermeulen, P.T.M., Heemink, A.W., Valstar, J.R., 2004. Inverse modelling of groundwater flow using model reduction. Water Resources Research 41, W06003. doi:10.1029/2004WR003698.

Vrugt, J., Dirks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. Water Resources Research 41, W01017. doi:10.1029/2004WR003059.

Wang, X., Hamill, T.M., Whitaker, J.S., Bishop, C.H., 2007. A comparison of hybrid ensemble transform Kalman filter-optimum interpolation and ensemble square root filter analysis schemes. Monthly Weather Review 135, 1055–1076.

Wen, X.-H., Chen, W.H., 2006. Real-time reservoir updating using ensemble Kalman filter: the confirming approach. SPEJ 11 (4), 431–442.

Wen, X.-H., Chen, W.H., 2007. Practical issues on real-time reservoir updating using ensemble Kalman filter. SPEJ 12 (2), 156–166.

Whitaker, J.S., Hamill, T.M., 2002. Ensemble data assimilation without perturbed observations. Monthly Weather Review 130, 1913–1924.

Woodbury, A., Ulrych, T., 2000. A full-Bayesian approach to the groundwater inverse problem for steady state flow. Water Resources Research 36 (8), 2081–2093.

Xiong, X., Navon, I.M., Uzunoglu, B., 2006. A note on the particle filter with posterior Gaussian resampling. Tellus 58A, 456–460.

Zhang, D., Lu, Z., Chen, Y., 2007. Dynamic reservoir data assimilation with an efficient, dimension-reduced Kalman filter. SPEJ 12 (1), 108–117.

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