Journal of Hydrology 545 (2017) 42–54

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Preconditioning an ensemble Kalman filter for groundwater flow using environmental-tracer observations

Daniel Erdal *, Olaf A. Cirpka

Center for Applied Geoscience, University of Tübingen, Germany

ARTICLE INFO

Article history:
Received 5 August 2016
Received in revised form 28 November 2016
Accepted 30 November 2016
Available online 18 December 2016
This manuscript was handled by P. Kitanidis, Editor-in-Chief, with the assistance of Wolfgang Nowak, Associate Editor

Keywords:
Data assimilation of hydraulic heads
Ensemble Kalman filter
Kalman ensemble generator
Groundwater-age tracers
Phreatic aquifer
Groundwater recharge
Hydraulic conductivity
Specific yield

ABSTRACT

Groundwater resources management requires operational, regional-scale groundwater models accounting for dominant spatial variability of aquifer properties and spatiotemporal variability of groundwater recharge. We test the Ensemble Kalman filter (EnKF) to estimate transient hydraulic heads and groundwater recharge, as well as the hydraulic conductivity and specific-yield distributions of a virtual phreatic aquifer. To speed up computation time, we use a coarsened spatial grid in the filter simulations, and reconstruct head measurements at observation points by a local model in the vicinity of the piezometer as part of the observation operator. We show that the EnKF can adequately estimate both the mean and spatial patterns of hydraulic conductivity when assimilating daily values of hydraulic heads from a highly variable initial sample. The filter can also estimate temporally variable recharge to a satisfactory level, as long as the ensemble size is large enough. Constraining the parameters on concentrations of groundwater-age tracers (here: tritium) and transient hydraulic-head observations cannot reasonably be done by the EnKF because the concentrations depend on the recharge history over longer times while the head observations have much shorter temporal support. We thus use a different method, the Kalman Ensemble Generator (KEG), to precondition the initial ensemble of the EnKF on the groundwater-age tracer data and time-averaged hydraulic-head values. The preconditioned initial ensemble exhibits a smaller spread as well as improved means and spatial patterns. The preconditioning improves the EnKF particularly for smaller ensemble sizes, allowing operational data assimilation with reduced computational effort. In a validation scenario of delineating groundwater protection zones, the preconditioned filter performs clearly better than the filter using the original initial ensemble.

1. Introduction

Numerical modeling of flow and transport in the subsurface is an indispensable tool for groundwater resources management. While the governing equations for physically based models are well agreed upon, obtaining the associated parameters remains a practical and scientific challenge. Direct observations of hydraulic conductivity hardly exist, and if so, the support volume of such measurements is smaller than the typical size of the computational grid cells (e.g., Zlotnik et al., 2000). Classical pumping-test analysis leads to effective transmissivity values of aquifers on larger scales (Kruseman and de Ridder, 1990), but in a heterogeneous formation different tests lead to different values even though the model assumes spatial uniformity (e.g., Sanchez-Vila et al., 2006; Li et al., 2007). Therefore, the estimation of aquifer parameters typically relies on observations of dependent quantities, such as hydraulic heads. Automatic calibration of groundwater models by fitting sparse hydraulic-head data is common in practice, usually assuming a limited number of zones with uniform hydraulic parameters (Hill and Tiedeman, 2007). In research, various inverse methods have been developed to estimate hydraulic conductivity as a continuous field, typically introducing a smoothness constraint or geostatistical prior knowledge (see the reviews of Sanchez-Vila et al., 2006; Vrugt et al., 2008; Zhou et al., 2014). While many studies have concentrated on estimating the hydraulic-conductivity field, a groundwater flow model also requires boundary conditions and specific-yield values, which need to be estimated as well. Transport applications additionally require porosity. Among the boundary conditions, groundwater recharge has a particularly strong impact on the system behavior. It varies in space and time, but may be difficult to quantify with certainty (de Vries and Simmers, 2002; Scanlon et al., 2002).

In the synthetic test case of a preceding study (Erdal and Cirpka, 2016), we used the Ensemble Kalman Filter (EnKF) to estimate both hydraulic conductivity and recharge as spatially variable

* Corresponding author.
E-mail address: daniel.erdal@uni-tuebingen.de (D. Erdal).

http://dx.doi.org/10.1016/j.jhydrol.2016.11.064
0022-1694/

fields based on hydraulic-head observations. We could show that the problem is inherently ill posed because unresolved variability in hydraulic conductivity can be compensated by an erroneous recharge pattern. Therefore, the assumed prior knowledge of the two fields to be estimated determines the predictive power of the calibrated model. The key problem is that hydraulic-head values themselves are not indicative of groundwater fluxes. Essentially the same hydraulic-head field is computed by a model if both hydraulic conductivity and recharge are multiplied by the same uniform factor. Hence, additional measurements, which relate to groundwater fluxes, should help disentangling effects of conductivity and recharge variations.

A specific measurement type, often used to quantify groundwater fluxes, are environmental tracers (e.g., Cook and Solomon, 1997). Here we restrict the analysis to groundwater-age tracers. They include natural and ubiquitous anthropogenic compounds with known concentration in the atmosphere or in precipitation, such as tritium (^{3}H) (e.g., Cartwright and Morgenstern, 2012; Gusyev et al., 2013), the ratio of tritium to tritiogenic helium (^{3}H/^{3}He) (e.g., Poreda et al., 1988; Portniaguine and Solomon, 1998; Åkesson et al., 2014), chloroflourocarbons (CFCs) (e.g., Busenberg and Plummer, 1992), sulfur hexafluoride (SF_6) (e.g., Busenberg and Plummer, 2000; Busenberg and Plummer, 2008), ^{85}Kr (e.g., Held et al., 1992), and ^{222}Rn (e.g., Hoehn and von Gunten, 1989). The basic assumption is that a water parcel is disconnected from the atmosphere upon groundwater recharge. The concentrations of CFCs, SF_6, and ^{85}Kr are conserved so that the time of recharge relates to the equilibrium concentration of these compounds to the atmospheric concentrations at that time. In case of tritium, which replaces ^{1}H in water molecules, radioactive decay leads to a decrease of ^{3}H and an increase of tritiogenic ^{3}H, whereas ^{222}Rn increases in groundwater due to its emanation from the matrix. In practice, multiple environmental tracers are often used, for example to calibrate groundwater models (e.g., Bauer et al., 2001), constrain recharge (e.g., Lin et al., 2013; Rueedi et al., 2005) or distinguish between different flow modes (e.g., Darling et al., 2012).

Zuber et al. (2011) divided the use of groundwater-age tracer data in two categories. In the first category, the observed tracer concentration is converted to a groundwater age using analytical expressions that either neglect mixing of water parcels of different age altogether, or make simple assumptions of mixing. Subsequently, the groundwater age is used to infer aquifer properties. The problem of this approach is that the relationship between the groundwater age and the environmental-tracer concentrations is nonlinear, so that the apparent groundwater age is not the true mean age. To overcome this shortcoming, the second category of inverse methods directly takes the tracer concentration as observable. Here, a groundwater flow-and-transport model is used to compute the concentrations, and aquifer parameters are inferred by fitting the concentration values, leading to a calibrated model that consistently accounts for the effects of groundwater mixing on the environmental-tracer concentrations (Turnadge and Smerdon, 2014; Suckow, 2014; McCallum et al., 2015).

The process of model calibration, that is, inferring model parameters from indirect observations, can be divided into batch approaches, that make use of all observations simultaneously, and sequential approaches that repeatedly recalibrate the model using one set of data in a given step. For transient parameters, such as groundwater recharge, sequential methods are the method of choice. They are also particularly suited when considering online observations of hydraulic head (Huber et al., 2011; Hendricks Franssen et al., 2011). A particularly popular sequential method in groundwater modeling is the Kalman filter (e.g., Ferraresi et al., 1996; Eppstein and Dougherty, 1996; Hantush and Mariño, 1997). Among the different variants, the Ensemble Kalman Filter (EnKF) (Evensen, 1994) has the particular advantage that it can be used in conjunction with any model relating model inputs to observations, without the necessity to compute the Jacobian of that relationship. Originally developed for data assimilation only, EnKF can also be used to estimate parameters, simply by extending the state vector with the vector of parameters. A typical groundwater example is the estimation of hydraulic conductivity (e.g., Hendricks Franssen and Kinzelbach, 2008; Drécourt et al., 2006; Tong et al., 2010; Tong et al., 2011; Xu et al., 2013a; Xu et al., 2013b; Crestani et al., 2013; Tong et al., 2013, among others). A few studies also analyzed the joint estimation of conductivity and recharge (Hendricks Franssen et al., 2004; Erdal and Cirpka, 2016).

Using environmental-tracer observations in a sequential model-calibration framework is complicated because the tracers carry the cumulative information of the velocity field since their introduction into groundwater, which is typically years before their measurement. A filter approach performs an update of states and parameters using current data and cannot correct the past. Hence, combining environmental tracer data, which depend on the past velocity field over a long period of time, with frequently measured hydraulic-head observations in a sequential assimilation is problematic. A smoother can make use of data over a larger time scale, but storing all fields (in particular the heads throughout the domain) over years of simulated time and updating them in order to meet environmental-tracer observations appears unrealistic if the dynamics are on the order of days to months. Oppositely, as pointed out by Massoudieh (2013), environmental tracers, even multiple tracers, taken at one time point cannot provide much information of transient groundwater flow. They merely reflect the flow conditions averaged over longer times. Hence, the question is open as how to include environmental-tracer information into a transient model where both the mean behavior and the temporal fluctuations are of interest.

An alternative way of conditioning an EnKF framework on additional information is by pre-calibration/preconditioning of the model prior to the filter simulation. Huber et al. (2011) used estimates of hydraulic conductivity and leakage coefficients to pre-calibrate a real-word groundwater model using pilot points. Subsequently, the parameters of the pre-calibrated model were updated by data assimilation using continuous hydraulic-head measurements. The authors showed that the pre-calibration of the model was helpful. In the present study, we test how environmental-tracer data can be used to precondition a groundwater model without performing the full transient calculations. The preconditioned model consists of a parameter ensemble that is consistent with the tracer measurements under time-averaged flow conditions. In the subsequent EnKF application, the groundwater-age data are no more used, and state/parameter updates are based on transient head measurements. That is, the environmental-tracer data are accounted for in the initial sample of the Ensemble Kalman filter.

When considering a well defined or preconditioned prior, it could be beneficial if some parameters slowly return to given values such as the initial mean if no informative data is available. Deardorff (1978) implemented this approach under the name of a force-restore term, and a similar one is considered in this work for the temporal variations in groundwater recharge.

The objective of the present study is to show that an appropriate approach to including environmental tracer information in data assimilation of a transient groundwater flow model can be to precondition the initial sample on the tracer data.

The rest of the paper is organized as follows: Section 2 outlines the methods and models used in the present study. Section 3 describes the application to a synthetic test case. Section 4 shows

the results obtained for this test case. Section 5 discusses some issues of the present study and draws general conclusions.

2. Methods

2.1. Governing equations

2.1.1. Groundwater flow

For modeling of groundwater flow on large scales, we commonly assume the validity of the Dupuit assumption, which states that the relevant variations of hydraulic head are in the horizontal direction only. A phreatic aquifer can then be described by:

$$S_y \frac{\partial h}{\partial t} - \nabla \cdot (K(h - z_0)\nabla h) = R \tag{1}$$

in which $S_y(\mathbf{x})$ [–] denotes the specific yield (here the drainage-effective porosity), $K(\mathbf{x})$ [$\mathrm{L\,T^{-1}}$] is the depth-averaged hydraulic-conductivity, $R(\mathbf{x},t)$ [$\mathrm{L\,T^{-1}}$] is the groundwater recharge, $z_0(\mathbf{x})$ [L] is the geodetic height of the aquifer bottom, $h(\mathbf{x},t)$ [L] is the hydraulic-head field to be simulated, t [T] is time, $\mathbf{x}$ [L] is the vector of horizontal spatial coordinates. Eq. (1) is subject to suitable boundary conditions.

2.1.2. Transport of environmental tracers

We consider an atmospheric radioactive tracer (tritium) that undergoes advective transport in the aquifer. Under these conditions, the tritium concentration at a point within the aquifer is the concentration of the water parcel at the time of recharge, corrected for the radioactive decay which uniquely depends on the groundwater age. We assume that concentrations are measured in observation wells with finite sampling screens, which implies averaging of the local concentrations over depth:

$$c_{meas}(x,y,t) = \frac{1}{z_2 - z_1} \int_{z_1}^{z_2} c_{in}(t - \tau(x,y,\xi,t)) \exp(-\lambda\tau(x,y,\xi,t)) d\xi \tag{2}$$

in which c_{meas} [tritium units] is the measured concentration, x [L] and y [L] are the horizontal spatial coordinates, t [T] is the time of observation, z_1 [L] and z_2 [L] are the vertical coordinates of the bottom and top of the vertical sampling screen. Further, c_{in} [tritium units] is the input concentration of the tracer at the time of introduction into groundwater, τ [T] is the travel time from the recharge point to the observation well and λ [$\mathrm{T^{-1}}$] is the tracer-specific decay constant. The travel time is simulated by backward particle tracking using the semi-analytical approach of Pollock (1988). The velocity fields required to calculate the travel times are calculated as $\mathbf{v} = \mathbf{q}/\theta$, where $q(x,y,t)$ is the specific discharge acquired by solving the groundwater flow equation and $\theta(x,y)$ is the (potentially spatially variable) porosity of the medium.

It should be noted that a random-walk component could easily have been included in the simulations to account for dispersion. However, it has been observed that neglecting local-scale dispersion is acceptable when considering the probability density function of travel times averaged over a large cross-section (or depth) (e.g., Sanz-Prat et al., 2015; Loschko et al., 2016). Since including a random-walk component would also come with an increase in computational demand, we chose not to include it in this study as we believe it would not change the result to any large extent.

2.2. Ensemble Kalman filter (EnKF) for state and parameter estimation

In this study, we use the standard Ensemble Kalman filter (EnKF) (Evensen, 1994) with an augmented state vector and a damping factor applied to the Kalman gain. The EnKF can be split into two steps. The first step is the forecast of each ensemble member:

$$\mathbf{x}_t^f = \mathbf{f}(\mathbf{x}_{t-1}^a) \tag{3}$$

$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t). \tag{4}$$

in which $\mathbf{x} = [\mathbf{h}, \mathbf{\Phi}]^T$ is the $(n_s + n_p) \times 1$ augmented state vector consisting of the flow model states (here the $n_s \times 1$ vector of hydraulic-head values at all grid points, $\mathbf{h}$) and the $n_p \times 1$ vector of parameters to be estimated ($\mathbf{\Phi}$), t is the time index, $\mathbf{f}$ is the numerical groundwater model, $\mathbf{y}$ is the $n_o \times 1$ vector of predicted observations, and $\mathbf{g}$ is the observation operator that relates the observations to the states. The superscripts a and f stand for analysis (see below) and forecast. The forecast model in Eq. (3) is run for each ensemble member (n_e in total) resulting in an $(n_s + n_p) \times n_e$ state matrix $\mathbf{X} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n_e)}]$ and an $n_o \times n_e$ observation matrix $\mathbf{Y}$. Note that the forecast model f propagates the hydraulic heads from one time level to the next, but does not alter the parameters.

The second step is the analysis, in which the state matrix is updated to better fit the current observations:

$$\mathbf{X}_t^a = \mathbf{X}_t^f + b(\mathbf{P}_{xy}(\mathbf{P}_{yy} + \mathbf{R})^{-1})\left(\mathbf{Y}_t^{(obs)} - \mathbf{Y}_t\right), \tag{5}$$

with

$$\mathbf{P}_{xy} = \left(\mathbf{X}_t^f - \overline{\mathbf{X}}_t^f\right)(\mathbf{Y}_t - \overline{\mathbf{Y}}_t)^T \tag{6}$$

$$\mathbf{P}_{yy} = (\mathbf{Y}_t - \overline{\mathbf{Y}}_t)(\mathbf{Y}_t - \overline{\mathbf{Y}}_t)^T \tag{7}$$

where $\mathbf{P}_{xy}$ is the $(n_s + n_p) \times n_o$ cross-covariance matrix between the augmented states and the predicted observations and $\mathbf{P}_{yy}$ is the $n_o \times n_o$ covariance matrix of the predicted observations. Both matrices are computed from the ensemble. $\overline{\mathbf{X}}_t^f$ is an $(n_s + n_p) \times n_e$ matrix in which each column is the ensemble average of $\mathbf{x}_t^f$, and $\overline{\mathbf{Y}}_t$ is an $n_o \times n_e$ matrix repeating the ensemble average of $\mathbf{y}_t$ as columns. $\mathbf{R}$ is the $n_o \times n_o$ covariance matrix expressing observation errors, typically assumed to be a diagonal matrix, and $\mathbf{Y}^{(obs)}$ is the $n_o \times n_e$ matrix of observation, consisting of the true observations plus random observation noise corresponding to the uncertainty of the measurement, generated for each observation, each ensemble member and at each observation time. The superscript T denotes a matrix transpose.

The coefficient b is the so-called damping factor (e.g., Hendricks Franssen and Kinzelbach, 2008) which serves to slow down the update of the augmented state matrix. In this work, the time step size of the groundwater model is one day (see below), and the damping factor is set to 0.5 for all stationary variables and 1.0 for the transient variables. The choice of the damping factor can largely impact the performance of the filter. In this work it is set to a slight damping such that the aim of an operational stable estimate is reached after one year. For a more in-depth description of the filter algorithm, please refer to Evensen (2003) or Burgers et al. (1998).

In the applications presented in this work, we use the EnKF to estimate spatially distributed hydraulic conductivity, a single homogeneous porosity value, mean annual recharge values for 4 predefined land-use zones as well as a temporal recharge trend parameter. More details on this will be given in Section 3.

2.3. Preconditioning using the Kalman Ensemble Generator (KEG)

Observation of environmental tracers that have stayed years in the aquifer are difficult to include in a sequential assimilation with daily updates. The concentration values are affected by the past groundwater-flow conditions over the entire travel time of the water parcels. Thus, updating parameters and states at the time

of concentration observations makes no sense. In this work, we therefore use the concentration information for preconditioning of the highly variable initial sample (see Fig. 3) before starting the sequential filter. To this end, we use the iterative Quasi-Linear Kalman Ensemble Generator (KEG) of Nowak (2009), which makes use of all data at the same time. In contrast to the filter approach, we neglect the transient behavior of groundwater flow in the preconditioning. This implies that hydraulic-head observations are time averaged, and recharge values to be estimated are time averages, too. In practice, this means that the transient flow model is replaced by a steady-state model and that the particle tracking is performed using the resulting steady-state velocity field. In the following, we briefly describe the general outline of the KEG implementation used in this work, while a detailed description can be found in the original work (Nowak, 2009).

Like the EnKF, the KEG is based on a state matrix containing all parameters to be updated, here denoted $\mathbf{Z}$ to differentiate it from the augmented state matrix of the EnKF. $\mathbf{Z}$ contains only parameters, and no states. In difference to the EnKF, the KEG makes synchronously use of all observations and updates all parameters at once. Because the underlying parameter-to-observation relationship is nonlinear, a single update cannot lead to an acceptable model performance, and the update is repeated using the same data in an iterative manner. To avoid over-fitting by repeatedly using the same data, the iterations must be stopped such that the residuals meet the prescribed error statistics. For an arbitrary iteration step j, the KEG can be described in seven steps:

1. Calculate the covariance matrices $\mathbf{P}_{zy}$ and $\mathbf{P}_{yy}$ based on the previous iteration states ($\mathbf{Z}^{(j-1)}$) and observations ($\mathbf{Y}^{(j-1)}$) as in Eq. (5).
2. For each ensemble member i, calculate a vector of residuals $\mathbf{r}_i = \mathbf{y}^{(true)} - \mathbf{y}_i^{(j-1)}$ and a sum of normalized squared errors $L_i^{(j-1)} = \mathbf{r}_i^T \mathbf{R}^{-1} \mathbf{r}_i$ for the previous iteration.
3. Update the state vector:

$$\mathbf{z}_i^{(j)} = \mathbf{z}_i^{(j-1)} + a(\mathbf{P}_{zy}(\mathbf{P}_{yy} + \mathbf{R})^{-1})\mathbf{r}_i \tag{8}$$

   where a is a scaling factor initially set to 1.
4. Run the flow model using the new parameters $\mathbf{z}_i^{(j)}$.
5. Calculate the new sum of normalized squared errors $L_i^{(j)}$ using the new observations.
6. If the new sum of normalized squared errors of ensemble member i is smaller than the previous one, this ensemble member is accepted for this iteration.
7. If not, the scaling factor a is reduced by a factor of two and the steps 3–7 are repeated until acceptance or until the update is smaller than a predefined limit.

If no acceptance could be reached, ensemble member i is re-sampled from the full ensemble at iteration step $j-1$ using a random combination of all members. The algorithm iterates over j until manually terminated or, formally correct, all ensembles are finally accepted. The final acceptance is based on calculating a critical cumulative probability value of the χ^2-distribution.

In our application, the observation types and parameters in the preconditioning differ from those in the subsequent filter step. In preconditioning, the observations include time-averaged hydraulic-head values and concentrations of environmental tracers. The parameters include the depth-averaged hydraulic-conductivity field, and the time-averaged recharge field. In the sequential filter step, the observations are transient hydraulic-head measurements and the parameters include hydraulic conductivity and transient groundwater recharge for a set of recharge zones (see below). We are fully aware that the time-averaged hydraulic-head field may not exactly meet the steady-state groundwater flow equation applying the same hydraulic-conductivity field and time-averaged groundwater recharge. This implies that preconditioning in the described way may introduce a slight bias in the prior ensembles of hydraulic conductivity and mean recharge. Conversely, we do not see that accounting for the environmental-tracer data in a transient model is reasonable, and the slight inconsistencies between the steady-state and transient groundwater-flow simulations also ensure that we do not use the same information (the hydraulic-head measurements) twice in an identical manner.

3. Application to a synthetic test case

For investigating how the information included in environmental tracers could be incorporated into a sequential filter, we define the synthetic test case of a depth-averaged phreatic aquifer.

3.1. Hypothetical application scenario

As a hypothetical application, we conceptualize a local regulator or water supplier wanting a real-time operational groundwater model, in which the current state of the aquifer as well as the current recharge and recharge trends can be monitored. Besides providing a current estimate of the system state, the model should help the stake holders in evaluating risks and governing strategies for the management of the aquifer. To achieve this goal, the model is primarily driven and updated with daily observations of heads from a limited number (12) of observation wells. Ideally, the model should be operational (that is, stationary parameters have become stable after repeated updates,) after one year of sequential calibration. The scenario concept is similar to the operational EnKF application of Hendricks Franssen et al. (2011).

As a validation scenario, we assume that the groundwater manager wants to explore the effects of 4 new drinking-water wells. The operational model is used to evaluate the required protection zones needed for these wells. In particular, we are interested in delineating the zones of one and three years of advective travel time towards the proposed wells. The practical implementation of the validation scenario is detailed further below.

3.2. Model domain and flow model

The scenario is inspired by the valley of River Neckar between the cities of Rottenburg and Tübingen in southwest Germany. Fig. 1 shows the model domain and the parameters of the virtual truth. The model has a spatial resolution of 10 m, leading to a total number of 92,679 unknowns. The true fields of hydraulic conductivity and porosity are generated as 3-D fields consisting of discrete zones using the software TPROGS. The 3-D fields are populated with values, a trend from east to west is added and to achieve the 2-D fields, the 3-D fields are averaged across the vertical direction which results in the fields shown in Fig. 1A and B.

The true groundwater recharge is constructed by subdividing the domain by a fictive land use map (see Fig. 1C). We compute reference evapotranspiration using observed meteorological data from the weather station of the German Meteorological Survey (DWD) at the airport Stuttgart-Echterdingen and applied the Food and Agriculture Organization (FAO) Penman-Monteith equation (Allen et al., 1998). Each land-use class was assigned a different crop factor and a different intercept value to create 4 sets of meteorological forcings. Finally, we used a simple one-dimensional model of unsaturated flow driven with these forcings to obtain one time series of groundwater recharge for each land-use class.

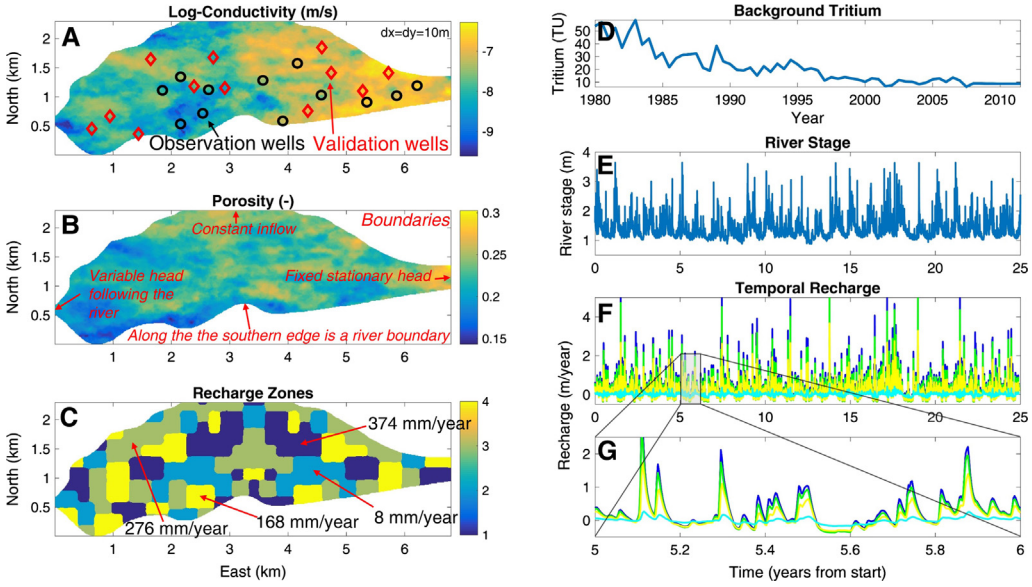

Fig. 1. Setup of the synthetic test case. (A) Natural logarithm of hydraulic conductivity; (B) specific yield/porosity field and hydraulic boundary conditions; (C) spatial pattern of land use and mean groundwater recharge for each land use class; (D) time series of input tritium concentrations; (E) time series of river stages; (F) time series of groundwater recharge; (G) detail of the recharge time series.

The resulting patterns in space and time can be seen in Fig. 1F and G, where also mean values of each land-use class are given.

The river (southern boundary) is simulated as leakage boundary with a given thickness and conductivity of the river bed and driven by observations from a gaging station in River Neckar (Fig. 1E). The model is run for 25 years with a temporal resolution of 6 h. 12 observation wells are placed in the domain, distributed such that each land use has 3 observation points and that the domain is reasonably well covered. 4 pumping wells, operational between year 1 and 3, are also placed within the close proximity of 4 observation points (see Table 1).

The transient groundwater-flow equation is solved using a cell-centered Finite Volume scheme for spatial discretization and a fully implicit Euler scheme for temporal integration. For linearization we use a standard Newton-Raphson scheme aided by a line search algorithm based on a bisectional search, and the linearized system of equations is solved using the built-in solvers of Matlab.

3.3. Transport of environmental tracers

In this study, we consider tritium (^{3}H) as environmental tracer. Tritium replaces ^{1}H in water molecules and thus behaves like an ideal tracer except for the radioactive decay with a half-life of about 12.4 years. To simulate the tracer concentration, we assume that the water in groundwater recharge is in equilibrium with the temporally variable atmospheric background concentration shown in Fig. 1D (IAEA/WMO, 2012). In this study, we do not consider the fate of the daughter compound, ^{3}He.

To simulate the transport to an observation well, we restrict the analysis to advective-reactive transport using the semi-analytical particle tracking approach of Pollock (1988). In the virtual truth, transient flow in the horizontal direction is taken from the flow model described above, while for vertical transport we assume that the vertical Darcy velocity scales linearly between the groundwater table, where it matches the groundwater recharge, and the aquifer bottom, where it is zero. In the virtual truth, all head fields are saved throughout the 25 years of simulated groundwater flow. Afterwards, particles are introduced at the final time step at each well and tracked backwards in space and time until they reach the groundwater surface. For each particle, the time of introduction to the groundwater is stored. The particle tracker is implemented as Matlab code and parallelized on the Graphics Processing Unit (GPU), which leads to a thousand-fold speed-up in comparison to calculations on the Central Processing Unit (CPU). GPU parallelization is done using the parallel computing toolbox of Matlab and hardly requires any adaptations of the code.

We assume that the virtual tracer-measurement has taken place following a single sampling campaign in the year 2011, considering 12 observation wells (red[1] circles in Fig. 1A). We further assume depth-oriented sampling over 1 m screening intervals ranging from the groundwater table to 5 m below. The concentration of tritium in the sample is evaluated by averaging the concentration carried by each particle, which is obtained by multiplying the inflow concentration at the injection time with the exponential decay function for the travel time, see Eq. (2). We do not consider dispersive groundwater mixing within the aquifer, assuming that mixing upon sampling is more important than dispersive mixing in the aquifer. The maximum time of particle tracking is 25 years, which is considered sufficient because only the top 5 m of the groundwater body are sampled.

In the following analysis, we do not consider the true age of the sampled particles. Instead, we analyze the concentration of the environmental tracer, as suggested by Turnadge and Smerdon (2014), Suckow (2014), and McCallum et al. (2015). This implies that the same procedure of computing tritium concentrations in the water samples is applied to every ensemble member in every iteration of the KEG used for preconditioning.

3.4. Model implementation

3.4.1. Grid coarsening in EnKF runs

Running a model with the resolution of the synthetic truth is computationally demanding, especially if an ensemble of models is considered. To avoid a runtime of several weeks for a single EnKF run, a coarser prediction model (CPM) is used in the ensemble predictions. It has a resolution of 100 m, leading to 968 unknowns.

[1] For interpretation of color in Figs. 1, 4 and 6, the reader is referred to the web version of this article.

Table 1
Data of the five pumping wells: position, start and stop times (in days since start of the model simulation), and pumping rates.

Well	x (m)	y (m)	Start (day)	Stop (day)	Rate (m^3/h)
# 1	4350	1160	100	300	3.6
# 2	3350	1260	200	400	3.6
# 3	1840	1160	300	500	3.6
# 4	1150	350	400	600	3.6
# 5	3250	1860	500	700	3.6

Coarsening the model has obvious drawbacks. For example, hydraulic-head observations close to rivers or pumping wells are strongly affected by their exact location, with a 100 m resolution being too coarse. To mitigate some of these issues, a simple 4-step downscaling approach is applied:

1. For each coarse cell including an observation, we set up a fine-resolution model (FRM) covering only the area of the coarse grid cell.
2. As boundaries at the current time step, the fine-resolution model takes the calculated fluxes across the boundaries of the coarse grid cell.
3. For each time step of the flow model, the coarse prediction model is run, followed by all fine-resolution models.
4. The observations are taken from one grid cell within each of the fine-resolution models.

This downscaling approach saves computer time, particularly if the number of observations is low. However, if the fine-resolution model is close to a feature causing strong curvature of the hydraulic-head field, but the feature is situated outside of the refined model, significant differences to a model run resolving the entire domain at high resolution can be obtained. Such problems could be mitigated by an adaptive grid, which would come at the cost of a higher overhead in the model evaluations and has not been attempted in the current study. The approach presented here is fast, easy to implement, and applicable to many groundwater flow models. The accuracy of using a coarse model in conjunction with a downscaled observation operator of course depends on the validity of using a coarse-grid representation to begin with. In many practical applications, this can be evaluated before applying data assimilation. If, for example, using a coarse grid breaks the connectivity of an aquifer body, the approach is prohibited. A recent study by Li et al. (2012) designed a more rigorous approach to upscale conductivity fields for an EnKF assimilation study. Their approach showed good results, but requires observations of hydraulic-conductivity to be applicable.

When considering the coarse prediction model in a sequential filter, the relation between its fine-resolution models and the filter has to be explicitly considered. There are primarily two ways of viewing a fine-resolution model: (A) as part of the state-propagation model to be included into the augmented state vector $\mathbf{x}$, or (B) as a non-linear observation operator (g in Eq. (3)) that is only used to improve the observations. When testing approach A, we observed that the fine-resolution models were well estimated. The same was true for the coarse prediction model, as long as we were close to the observation locations. As the fine-resolution models were updated, the error between the predicted observations and the true observations rapidly decreased. However, the coarse model cells further away from an observation location remained practically unaltered. This led to a rather poor prediction of heads throughout the coarse model domain and the approach of taking the fine-resolved states and parameters into the state vector has therefore been abandoned in this work.

When considering the fine-resolution model as part of the observation operator (B), we need to enforce the connection between the coarse-grid cell and its fine-resolution model when the coarse prediction model undergoes the analysis step. We tested two approaches: The first (B.1) is to re-sample all fine-resolution models after each analysis step, for example by interpolating them directly from the updated coarse prediction model. The second (B.2) is to alter the mean of each fine-resolution model such that it follows the value of its updated coarse-grid cell. Using approach B.1 is likely recommended when the difference between the time step used for updating the system and the flow model time step is large, so that the fine-resolution models get enough time to evolve away from the smooth interpolation. For the present work, however, the time step difference is rather small (daily updates and 6 h time steps) so that approach B.2 is followed.

3.4.2. Operational EnKF

The basic EnKF simulations are run for a total of 10 years using an ensemble of 480 members and an observation noise of 2 cm. In difference to the synthetic truth, in which each land use has its own temporal recharge, the EnKF flow model has only one temporal trend factor and each land use has its own annual mean recharge around which the trend is centered. The force-restore term introduced in the introduction is applied such that the mean of the temporal recharge trend is pulled towards unity. If the observations are not informative with respect to recharge, the mean approaches unity after approximately one year. It should be noted that the force-restore term likely has minor influence on the parameter update in our application because the strong seasonality of groundwater recharge has a clear impact on hydraulic heads, leading to a continuous update of the trend parameter.

The coarse land-use map and the mean of the initial sample of conductivities are shown in Fig. 2. The initial sample of recharge and porosity are randomly drawn from a uniform distribution between the minimum and maximum values listed in Table 2. The log-conductivity fields are generated using a multi-Gaussian random-field generator in which the mean value, variance, rotation angle, and correlation lengths are randomly drawn from uniform distributions (Table 2). The conductivity fields are generated with a resolution of 10 m. For the EnKF, they are upscaled to the coarse resolution of 100 m by using the value of the corresponding center cell of the fine-scale grid. It should be noted that more rigorous approaches for upscaling the initial fields exist (e.g., coarse-graining as studied by Attinger (2003)). Most likely, the cell-center approach probably is suboptimal. An example of 4 conductivity fields is shown in Fig. 3. The initial ensemble members of the temporal trend parameter of groundwater recharge are sampled from a normal distribution with mean 1.0 and standard deviation of 2.0.

There are two different versions of the operational EnKF: without and with preconditioning. In the version without preconditioning, we update 968 hydraulic head values (one per coarse grid cell), 968 spatially distributed hydraulic-conductivity values, a single value of mean porosity, four values of long-term recharge related to the four land-use types, and one parameter representing the spatially uniform transient trend of groundwater recharge. Hence, in this version of the operational EnKF the augmented state vector has in total 1942 entries to be updated.

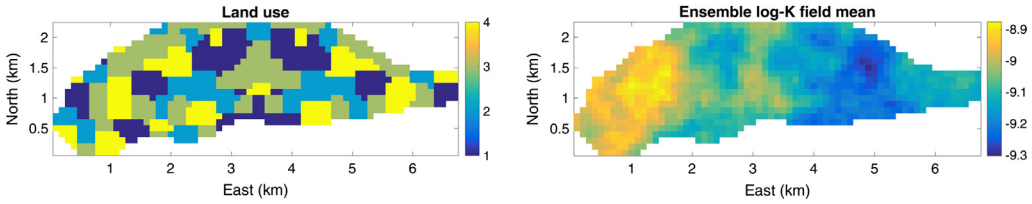

Fig. 2. Land use (see Fig. 1) and the ensemble mean of the initial conductivity fields for the Coarse Prediction Model. Please note the difference in plotting scale of the conductivity field compared to Fig. 1.

Table 2
Parameters used for initial uniform sampling. Upper section contains parameters that are directly sampled to the initial ensemble, while the lower section contains parameters that are used as input to a random-field generator which generates the initial ensemble of conductivity fields. R is recharge, p porosity and l correlation length.

	Min	Max
R (mm/year)	30	1000
p (–)	0.05	0.9
$\log(\overline{K})$ (m/s)	−6.9	−11.5
$\sigma_{\log(K)}$ (–)	0.5	4.0
ang (rad)	$-\pi/2$	$\pi/2$
l_x (m)	100	3000
l_y (m)	100	3000

In the second version of the operational EnKF, the preconditioning has already reduced the sample spread of the hydraulic-conductivity field, porosity, and the long-term groundwater recharge values in the four zones. Here, we do not update the porosity and the four long-term groundwater recharge values during the operational EnKF. This reduces the augmented state vector to 1927 entries to be updated.

3.4.3. Implementation of the KEG preconditioner

For the KEG calibration, we considered the same ensemble of 480 members as for the basic standard EnKF. We used the full 10 m resolution model because the coarse-resolution model led to significant errors in the transport simulations. The calibration is driven with two observation types: the concentrations of tritium in the 12 observation wells at 5 depth intervals, and the hydraulic-head measurements averaged over the full simulation time. The observation errors were set to 0.3 TU (tritium units) and 0.3 m respectively. The parameters to be conditioned were the spatially distributed hydraulic-conductivity field, the spatially uniform porosity, and the mean recharge values for the four land-use types.

We did not achieve formal acceptance of the KEG iterations according to the χ^2-criterion applied to the calculated sum of normalized squared errors and decided to apply a fixed number of 19 iterations instead. If ran longer, the resampling (see Section 2.3) occurred too frequently and, hence, the ensemble spread became too small, essentially collapsing the entire ensemble to a single realization. It should be noted that in the original ensemble of 480 members, not all flow setups could be solved for steady-state. These non-solvable members were removed from the KEG simulation, leading to an effective ensemble size of 458 members for both the KEG and the preconditioned EnKF.

3.5. Validation scenario

As outlined above, the validation scenario considers the zones of one and three year travel time towards four wells. As discussed in the previous section, particle tracking on the coarse-resolution model is suboptimal, so that we interpolated all parameters of the EnKF back to the original 10 m resolution and ran a full fine-scale model for predictions. The validation scenario is simulated by starting the flow model from the updated ensemble on the first day of year 6 and running the model for 3 years with 6 h temporal resolution. All parameters apart from the transient groundwater recharge are kept fixed. To reduce the computational time, we did not consider the full ensemble and performed a single run using the ensemble-mean parameters. This approximation is likely having only minor influence on the result, since the spread of the ensembles for the time of interest is quite limited. Subsequently, the particle tracker is run in backward mode for the same time period. The particles are initialized over an area of $150 \times 150\ m^2$ to make sure that we cover a larger area than one grid cell of the coarse-prediction model. After 1 and 3 years of back-tracking, the outer contour of the particle plume is delineated and these outlines are compared to the same simulations using the synthetic truth.

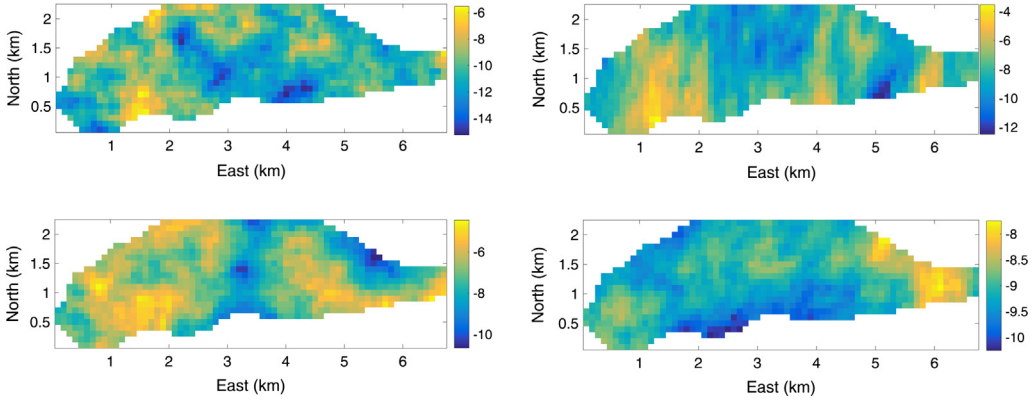

Fig. 3. Four examples from the ensemble of initial conductivity fields. Please note the different scalings on each plot.

3.6. Effects of ensemble size

A crucial question when dealing with ensemble-based methods is what ensemble size to use. Although theoretically a big ensemble is always better, the computational cost can easily be overwhelming for larger models, and in practice small ensembles may often be used. In this work we consider a base ensemble of 480 members. This is the ensemble that is used for the standard EnKF results as well as for the preconditioning. To test the impact of smaller ensembles, we also use ensembles of 240, 120 and 60 members. The smaller ensembles are randomly drawn subsamples. In case of the standard EnKF, the subsample is drawn from the original initial ensemble and in case of preconditioned EnKF the subsample is drawn from the preconditioned ensemble. Each of the EnKF runs with reduced ensemble size are repeated 5 times using a different random subsample.

4. Results

4.1. Preconditioning using environmental tracers

The results of the preconditioning using the KEG algorithm are shown in Fig. 4. The conductivity field has improved, as can be seen in the spatial pattern (cf. Figs. 1 and 2). Quantitatively, the Root Mean Square Error (RMSE) of the mean log-conductivity fields has been reduced from 1.6 for the initial sample to 0.7 for the final KEG estimate. Similarly, the spatial mean of the log-conductivity field has also improved, from −9.1 in the initial sample, to −7.7 in the final KEG estimate, which is much closer to the true value of −7.5.

When considering the long-term average groundwater recharge, all four land-use class types have a distinct mean and a limited spread after preconditioning. In most cases, these distributions correspond reasonably with the true mean value and, important for our EnKF application, with the relative magnitudes being very reasonable. The cyan and yellow histograms show, however, a bias. Fig. 4C shows that the mean porosity is very well estimated by the calibration algorithm. Overall the results of the preconditioning of the stationary parameters of the initial ensemble is considered successful, and the result is considered useful as input to the EnKF simulations.

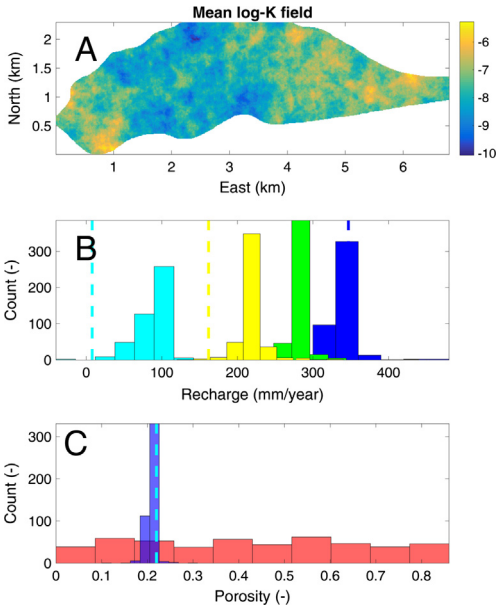

Fig. 4. Results from preconditioning by the Kalman Ensemble Generator (KEG). (A) Estimated ensemble-mean natural logarithm of conductivity, (B) histogram of estimate recharge by land-use type, (C) histogram of prior (red) and estimated (blue) porosity. Dashed lines indicates the true value. Please note that the dashed green line cannot be seen as it overlays the green histogram. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Sequential calibration with EnKF

The EnKF is run for 10 simulation years with daily assimilation of hydraulic heads. Two versions are compared: In the first, we run the filter with the highly uncertain initial ensemble, whereas in the second we use the preconditioned initial ensemble meeting the environmental-tracer and mean hydraulic-head observations.

Fig. 5 shows the stationary parameters resulting from the filter assimilation using the original initial ensemble. Similar to the preconditioning discussed above, the sequential filter improves the conductivity fields, both with respect to the spatial pattern (Fig. 5) and mean value (−7.7 compared to the true −7.5 and initial −9.1). However, the resulting mean conductivity field still holds spatial features that do not correspond to the truth, such as the low conductivity zone found on the left of Fig. 5A. As the correctness of the spatial patterns can be difficult to assess in a printed article, we also consider the evolution of the error of the ensemble mean log-conductivity field when the field in question is smoothed using a variable smoothing window size. If the overall pattern is correct, the error should decrease as the smoothing window size increases. Table 3 shows this result for window sizes between 3×3 and 21×21 cells. From the Table 3 it is clear that the error is reduced during smoothing for the standard EnKF, while for the

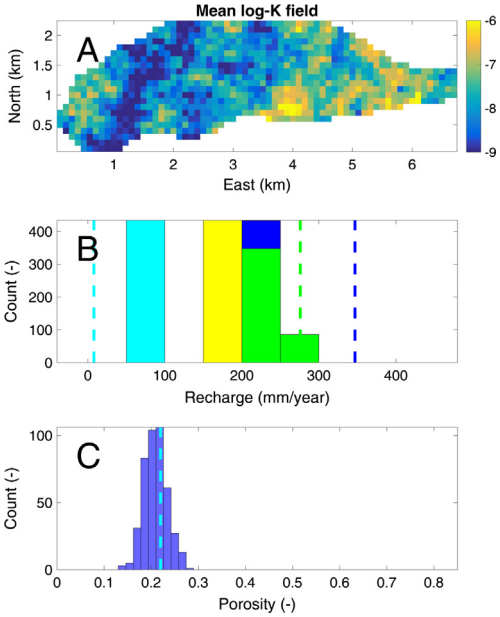

Fig. 5. Results from standard EnFK simulation. (A) Estimated ensemble-mean natural logarithm of conductivity (color range limited between −9 and −6), (B) histogram of estimate recharge by land-use type, (C) histogram of estimated porosity. Dashed lines indicates the true value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

initial sample this is not the case. This shows that the pattern in the initial sample is wrong, while after filtering using the standard EnKF it improves.

The filter results of the long-term groundwater recharge approach stable parameter values, mainly corresponding to the truths but showing a poor internal order (“green” land use showing

Table 3
Root mean square error of the smoothed ensemble-mean log-hydraulic conductivity field for different window sizes.

Window size	RMSE		
	Initial sample	Standard EnKF	Preconditioned EnKF
1×1	1.6	0.81	0.79
3×3	1.6	0.68	0.65
7×7	1.6	0.62	0.58
11×11	1.6	0.58	0.53
21×21	1.6	0.51	0.49

larger values than "blue" land use). The final estimate of the porosity is well centered about the true mean porosity.

When considering the dynamics of groundwater recharge, shown as blue lines in Fig. 6 for the case using the original initial ensemble, the filter performs well. The trend of the weakly averaged recharge is often captured. This is beneficial for the dynamic behavior of the model, since it displays more dynamic behavior than e.g. the trend of the monthly averaged recharge. Obviously, the daily fluctuations used to drive the synthetic truth are not captured by the filter. This is expected, as the groundwater-flow equation is a diffusion equation, and high-frequency components of the forcing are strongly dampened in the hydraulic-head response.

The hydraulic-head values are mainly matched in a satisfactory way (blue lines in Fig. 7). As enforced by the filter, the observations (left column of Fig. 7) are well matched at all times. The right column of Fig. 7 shows hydraulic-head time series at 3 validation points, where the agreement between filter estimate and the true values is worse, although in large the dynamics of the groundwater table are captured well, subject to a bias that stabilizes with time.

The key question of the present study is whether the preconditioning of the initial ensemble on the environmental-tracer data and the time-averaged hydraulic heads can improve the performance of the filter. Since the values for mean recharge and porosity were estimated with a rather small ensemble spread, these parameters are no more updated by the filter. Hence, the filter using the preconditioned initial ensemble has to update conductivities, the spatially uniform temporal trend coefficient of groundwater recharge, as well as the hydraulic-head field.

Since the pattern of the ensemble-mean conductivity did not change considerably from what is shown in Fig. 4, we do not show a new plot of it. The evaluation of the pattern (Table 3, last column) shows a decreasing trend of the error with increasing smoothing window, indicating that the overall pattern is appropriate. Further, when comparing the two last columns of Table 3, it can be seen that the preconditioned EnKF is slightly better at estimating the conductivity field than the standard EnKF. In the preconditioned EnKF, the mean value of the hydraulic conductivity kept improving during filtering, with a final estimate at -7.5, which corresponds well to the true one. The temporal groundwater recharges, shown in Fig. 6 as green lines, are also very well estimated compared to the weekly mean truths. The performance is also slightly better than the already very reasonable estimate obtained by starting the EnKF with the original initial sample, especially when considering the negative recharges where the standard EnKF tends to overestimate the extraction.

When evaluating the hydraulic heads at points that are not observation points (green lines in right column of Fig. 7), the EnKF based on preconditioning shows the right dynamics, and a smaller bias than the EnKF using the original initial ensemble. This is also quantified in Table 4, where the RMSE of the two filters are compared as means of the 12 validation wells shown in Fig. 1. As can be seen, both approaches have a very similar dynamic behavior (Table 4, bottom row), while the preconditioned EnKF has a notably smaller error in the direct head comparison (Table 4, top row). This shows clearly that the standard filter has a much stronger problem with bias in the predicted heads. A possible reason for this is that the preconditioned EnKF starts with a much better conductivity field and, thereby, also starts with a better hydraulic-head field since the models are started at their respective steady state. It is then possible that the standard EnKF does not fully manage to correct the bias at locations that are far away from any observation.

In summary, both filters show a good behavior with reasonable precision for the dynamic processes of interest. Generally, the preconditioned filter shows a slightly improved performance, particularly with respect to bias in the heads and the spatial pattern of the conductivity field.

It should be noted that due to the setup used in this study, the simulation using the original initial ensemble runs with less information than the preconditioned simulation, since no tracer data is used in the former. It would have been possible to include the tracer data also in the standard EnKF (although not so straight forward, as discussed above) and do a one-time update based on these data. This would correspond to doing one step with the KEG without considering the acceptance/rejectance criterion. The improvement after such a step is very minor. By contrast, the strength of the preconditioning lies in its repeated use of the data until convergence (i.e. a classic batch calibration approach). Hence, we did not complicate the EnKF simulations by including tracer observations, believing that the change in performance of the EnKF would have been very minor.

4.3. Validation scenario

As both filters performed reasonably well with respect to predicting the dynamics of the groundwater table, we considered the delineation of the well protection zone according to the validation scenario also for both filters. Fig. 8 shows the corresponding three-year capture zones. Here we can see differences between the two filters, especially for the wells situated most eastern and most southern, where the original-sample filter completely fails to reproduce the correct travel times. This can possibly be attributed to low conductivity zones in the standard filter results, as well as the over-predicted dynamics of the low-recharge fields. The preconditioned filter shows for most of the wells reasonably satisfying outlines. As outlined in Section 3.5, we downscale the mean conductivity fields before performing the particle tracking, which may have a notable smoothing effect on the presented outlines.

4.4. Effect of ensemble size

To further distinguish between the two filters, we now consider the effect of the ensemble size. We exemplify this by considering the estimated annual mean recharge, calculated as the mean of the estimated recharge times the recharge trend for the last 5 years of the EnKF simulations, as well as the Root Mean Square Error (RMSE) of the validation points (cyan crosses in Fig. 1A). The result is shown in Figs. 9 and 10. It should be noted that we do not resample the full ensemble of size 480, therefore there is only one box at this size in the figures. All smaller samples are drawn as a sub-set of these 480 original members. In this section we are not mainly interested in the errors or their magnitudes, but of the stability of the estimations when the ensemble size changes. As can clearly be seen in Fig. 9, the spread of estimated recharge values is large when considering the standard EnKF, while the preconditioned EnKF shows a much more stable result. Of course, during the preconditioned EnKF runs only the temporal trend is estimated and therefore the relation between the different recharge zones is fixed, limiting the possible deviations somewhat. The same trend is, however, also seen Fig. 10, from which it becomes obvious that

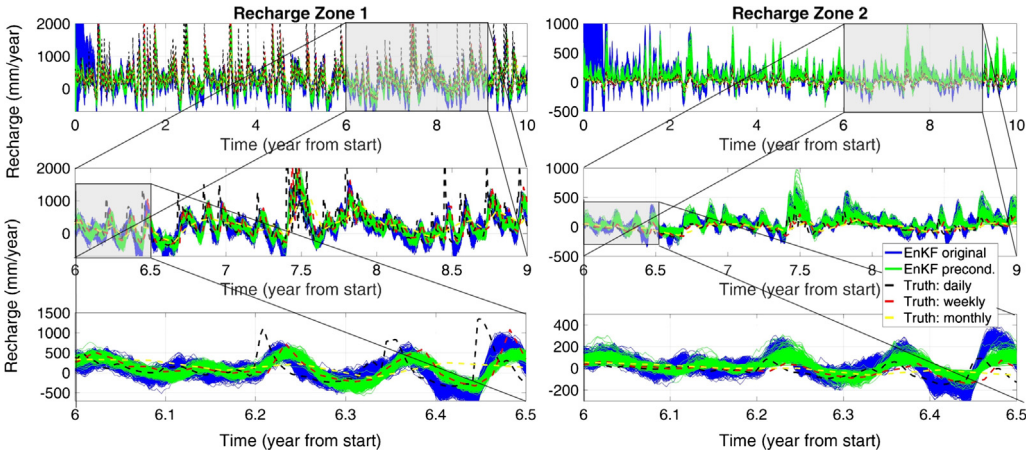

Fig. 6. Temporally variable recharge for recharge zones 1 and 2 (blue and cyan in Fig. 1), shown over different time intervals. True values correspond to the mean over the time period mentioned in the legend.

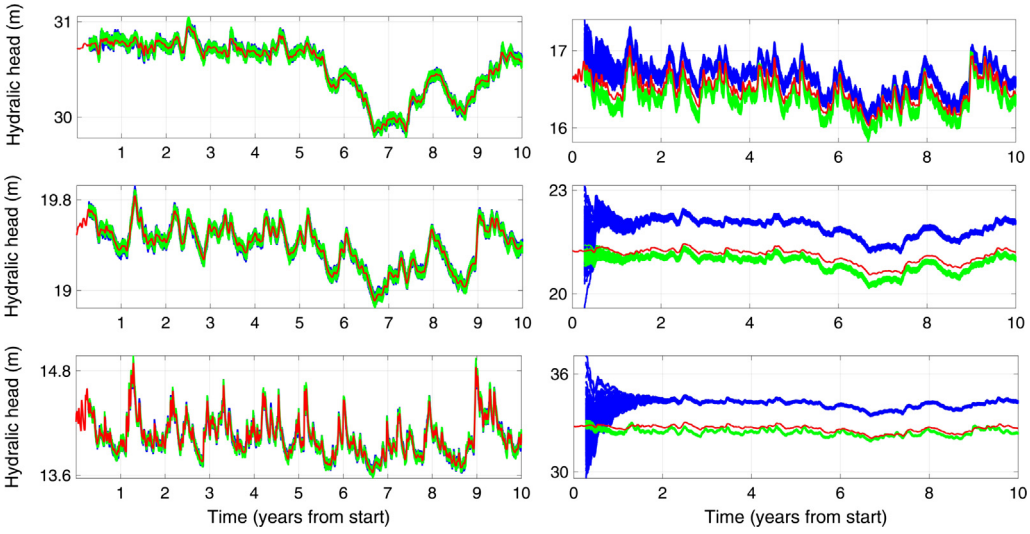

Fig. 7. Hydraulic head at 6 example locations during the filter operation. Left column shows heads that are used as observations, while the right column shows validation locations. Blue is original filter, green is preconditioned filter and red is the truth. Please note that in the left column the green lines are overlaying the blue lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Mean of root mean square error for the 12 validation hydraulic-head observations for the last 5 years of simulation. Top row compares hydraulic-heads directly while the bottom row compares the dynamic behavior.

	Standard EnKF	Precond. EnKF
h (m)	0.56	0.20
$\partial h/\partial t$ (m/day)	0.016	0.016

the result using the standard filter highly depends on the initial draw, while the preconditioned ensemble show stable result also for 60 ensemble members. This suggests that when using the preconditioned EnKF, it is possible to run an operative model using a much smaller ensemble size (here 60) and still get a satisfying result.

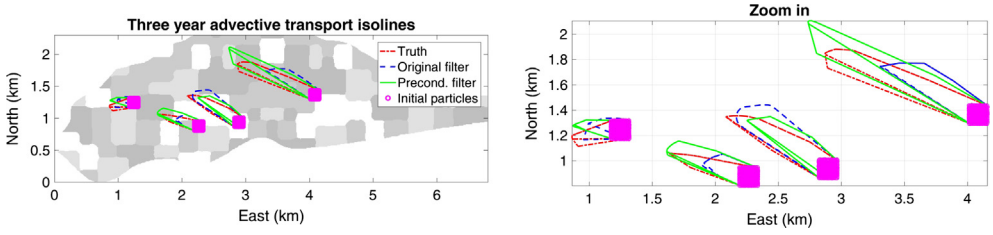

Fig. 8. Predictions of well capturing zones for the four proposed wells.

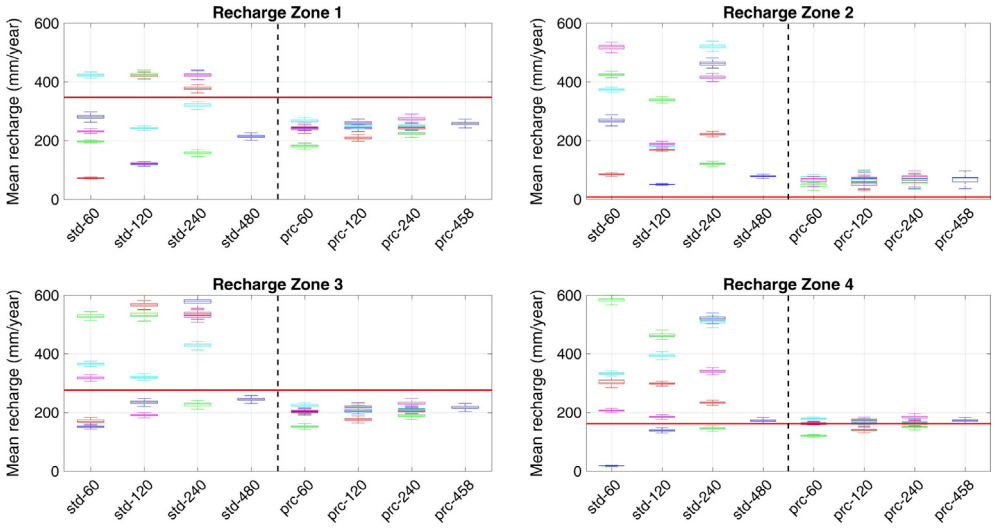

Fig. 9. Estimated mean annual recharge for the 4 recharge zones using the standard EnKF (std) and the preconditioned EnKF (prc) with different ensemble sizes (60, 120, 240, and 480/458). The boxes show the ensemble spread of the mean over the last 5 years. Different colors mark different initial samples and the red lines mark the true values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

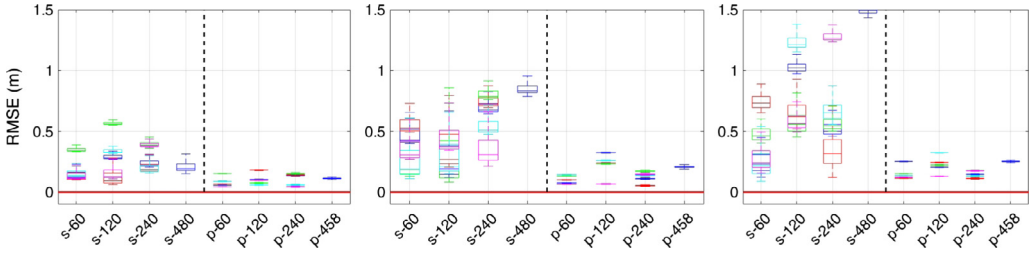

Fig. 10. Root mean square error of the validation points using the standard EnKF (s) and the preconditioned EnKF (p) with different ensemble sizes (60, 120, 240, and 480/458). The boxes show the ensemble spread of the mean over the last 5 years. Different colors mark different initial samples and the red lines mark zero. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Discussion and conclusions

We have presented a method for operational data assimilation in transient regional groundwater flow using the Ensemble Kalman Filter in different variants. In the following we would like to address seven important points related to the presented study:

1. In practical applications of data assimilation, computer power is typically limited. With given resources, one thus has to decide whether to spend them on a larger ensemble or a finer resolved grid. In Section 3.4.1 we have introduced an approach in which we have combined a domain-covering, coarse groundwater flow model with observation operators that include a small, highly resolved model each to obtain hydraulic-head values at specific measurement locations. This method seems to provide a sufficiently good performance of the EnKF. Obviously, since the model used for data assimilation is now on a coarse grid, the resulting fields are not as detailed as they could be in a model with finer resolution. Conversely, given the spatial limitations in the available observations, the coarse model might be a more realistic representation of what can be achieved. The concept of coarse and fine models has allowed us to run a reasonably large ensemble of 480 members within acceptable simulation times (here: 34 h on 8 CPU cores). Would we have opted for using the fully resolved model in the filter simulations, it is likely that the ensemble size would have been much smaller, bringing along the issues outlined in Section 4.4.
2. Overall, both variants of the EnKF were able to track the temporal behavior of groundwater recharge (see Fig. 6). When updating the filter with daily observations, it is natural that the filter cannot, and should not, track daily trends. The weekly mean, however, was well estimated. For the scenario setup used in this paper, this is important since one of the main objectives was to have an operational model system that can simulate groundwater recharge with a temporal resolution that is adequate for management purposes. With the current model, this requirement has clearly been fulfilled. The preconditioned filter showed slightly better tracking results than the original filter, but the differences were small.
3. The choice of damping factors (here 50% on all state variables apart from the temporal trend) can be crucial to the performance of the EnKF. Although the use of the damping factor is rather common, it remains an ad hoc parameter and few guidelines exist on how to select it. In this work, a rather weak damping is used with the aim of reaching stable parameters after one year. If, instead, the goal had been to find only parameters and the full 10 years of transient simulations had been used for the parameter estimation, the damping could have been set much stronger. This would have resulted in much blurrier and mostly useless system predictions during most of the EnKF run time, but it could potentially lead to better parameter estimates at the very end. If this is the pure interest, however, a proper calibration, such as the Kalman Ensemble Generator used in this work to precondition the initial ensemble, may be a better

choice altogether. Similarly, it could be considered to reduce the temporal observation density if one is only interested in the model behavior on large time scales.

4. The entire study is based on experiments using a randomized synthetic truth. This comes with a number of benefits, but also some inherent limitations. Synthetic truths are commonly used in parameter estimation studies because there is a controllable truth to compare any estimation to. However, the existence of a perfect model (the so called identical twin) is also the largest issue of the approach. In any real situation concerning groundwater, we would never be able to model all processes and spatial structures, and there will always be model errors of varying significance and correlations. A key when considering synthetic test cases is, hence, to introduce realistic model errors. In this work the consideration of the coarse prediction model, the homogeneous porosity, and the assumption that the temporal trend in recharge is identical for all land-use classes, introduced model error into the EnKF in a reasonably realistic way.
5. Throughout this study the recharge is considered variable, but on long time scales it is stable. When using environmental-tracer data in a particular catchment this assumption has to be tested. Longer periods of droughts, drastic change of land-use or changes in management policies for agriculture may all influence groundwater recharge on time scales of several years to decades. Under such conditions, the use of a steady-state model may not be appropriate. It would, however, be possible to use a transient model in the preconditioning step. Since the interesting time scales are still large, such a transient model could be run with temporally strongly smoothed forcing, considering only inter-annual variability, using large time steps to capture the long-time trends in recharge, which are relevant for the environmental tracer. By this, the model would remain computationally efficient, but of course the costs would be notably higher than under the steady-state flow assumption applied in the current study.
6. The good result shown for the preconditioning KEG-simulations in this study is, of course, partly related to the good (and synthetic) environmental-tracer data used. After the preconditioning we have already captured the essential parts of the parameter fields, leaving little work to be done by the EnKF on these aspects. This is likely the reason why we can reduce the ensemble size without losing performance. If the worth of the tracer data would have been smaller, it is likely that the preconditioning would not by itself have been enough to capture the important trends of the parameter fields, and repeatability of the parameter estimates by the EnKF would have required a larger ensemble size. However, as long as the preconditioning helps reducing variability of the initial sample in a constructive way, it is reasonable to believe that also the ensemble size of the EnKF could be reduced without loss of stability. Alas, likely not as much as in this study.
7. When considering the transport of environmental tracers in the preconditioning step, we introduced a model error by replacing transient flow model with a prediction of steady-state flow and by assuming a homogeneous porosity. Still, background concentrations are assumed perfectly known and the influence of the unsaturated zone is completely ignored. As has been shown by Schwientek et al. (2009), among others, the unsaturated zone can have large impacts on the concentration of environmental tracers reaching the groundwater table, especially when considering tritium. Also the background concentration of tritium varies in space and time and is therefore rarely exactly known at the entry location. The choice of tritium as a tracer is though rather unconventional, as it is nowadays commonly replaced by the ratio of tritium to titriogenic ^{3}He or CFC (e.g., Scanlon et al., 2002), where background concentrations are less problematic. That tritium alone is only rarely considered in field studies, depends on its decreasing background concentrations. Since the nuclear bomb testings in the sixties, which led to a large increase in atmospheric tritium concentrations, the values have dropped until present times when we are again approaching natural background values. For a tracer observation, this means that a given concentration observed in groundwater could be related to multiple advective ages, which creates ambiguity. This, however, is also partly why we consider tritium a good tracer for our synthetic test case. As pointed out above, it is important to include realistic model errors in a synthetic experiment. If we would have considered a state-of-the-art multi-tracer setup, we could easily have produced much better results in the parameter estimation, simply by having better and information-richer data. However, since all experiments are based on the same simplified virtual truth, it would likely result in unrealistically good estimates and rather show a flaw of using synthetic setups. Instead using an ambiguous tracer gives us more certainty that the result can be reasonably realistic. As also pointed out by Stewart and Morgenstern (2016), tritium may be coming back as a tracer in the next decades as we now approach steady background concentrations, though so far primarily in the southern hemisphere.

Overall, the presented study has shown that the performance of a transient data assimilation for groundwater flow can be improved by preconditioning the initial ensemble using environmental-tracer data. We would not recommend accounting for the groundwater-age related data within the normal filter procedure because groundwater age is a cumulative property depending on the velocity and thus hydraulic-head distribution over long times prior to the observation. Apart from a notably better performance in delineating well-capture zones when using the preconditioned method over the standard EnKF of head observations, the main advantage of the preconditioning is that the on-line filter predictions can be performed with stable results using much smaller ensemble sizes.

Acknowledgements

Financial support from the Deutsche Forschungsgemeinschaft (DFG) under CI 26/13-1 in the framework of research unit FOR 2131 is gratefully acknowledged.

References

Åkesson, M., Bendz, D., Carlsson, C., Sparrenbom, C.J., Kreuger, J., 2014. Modelling pesticide transport in a shallow groundwater catchment using tritium and helium-3 data. Appl. Geochem. 50, 231–239.
Allen, R., Pereira, L., Raes, D., Smith, M., 1998. Crop Evapotranspiration (Guidelines for Computing Crop Water Requirements) Tech. Rep.. FAO – Food and Agriculture Organization of the United Nations, Rome.
Attinger, S., 2003. Generalized coarse graining procedures for flow in porous media. Comput. Geosci. 7 (4), 253–273.
Bauer, S., Fulda, C., Schäfer, W., 2001. A multi-tracer study in a shallow aquifer using age dating tracers 3H, 85Kr, CFC-113 and SF6 – indication for retarded transport of CFC-113. J. Hydrol. 248 (1–4), 14–34.
Burgers, G., van Leeuwen, P.V., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126, 1719–1724.
Busenberg, E., Plummer, L.N., 1992. Use of chlorofluorocarbons (CCl3F and CCl2F2) as hydrologic tracers and age-dating tools: the alluvium and terrace system of central Oklahoma. Water Resour. Res. 28 (9), 2257–2283.
Busenberg, E., Plummer, L.N., 2000. Dating young groundwater with sulfur hexafluoride: natural and anthropogenic sources of sulfur hexafluoride. Water Resour. Res. 36 (10), 3011–3030.
Busenberg, E., Plummer, L.N., 2008. Dating groundwater with trifluoromethyl sulfurpentafluoride (SF 5CF3), sulfur hexafluoride (SF6), CF 3Cl (CFC-13), and CF2Cl2 (CFC-12). Water Resour. Res. 44, WR006150.

Cartwright, I., Morgenstern, U., 2012. Constraining groundwater recharge and the rate of geochemical processes using tritium and major ion geochemistry: ovens catchment, southeast Australia. J. Hydrol. 475, 137–149.
Cook, P.G., Solomon, D.K., 1997. Recent advances in dating young groundwater: chlorofluorocarbons, 3H/3He and 85Kr. J. Hydrol. 191 (1–4), 245–265.
Crestani, E., Camporese, M., Baú, D., Salandin, P., 2013. Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation. Hydrol. Earth Syst. Sci. 17 (4), 1517–1531.
Darling, W.G., Gooddy, D.C., MacDonald, A.M., Morris, B.L., 2012. The practicalities of using CFCs and SF 6 for groundwater dating and tracing. Appl. Geochem. 27 (9), 1688–1697.
de Vries, J.J., Simmers, I., 2002. Groundwater recharge: an overview of processes and challenges. Hydrogeol. J. 10 (1), 5–17.
Deardorff, J.W., 1978. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. J. Geophys. Res. 83, 1889–1903.
Drécourt, J.-P., Madsen, H., Rosbjerg, D., 2006. Calibration framework for a Kalman filter applied to a groundwater model. Adv. Water Resour. 29 (5), 719–734.
Eppstein, M.J., Dougherty, D.E., 1996. Simultaneous estimation of transmissivity values and zonation. Water Resour. Res. 32 (11), 3321–3336.
Erdal, D., Cirpka, O.A., 2016. Joint inference of groundwater-recharge and hydraulic-conductivity fields from head data using the Ensemble-Kalman filter. Hydrol. Earth Syst. Sci. Discuss. 20, 555–569.
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99 (10), 143–162.
Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dyn. 53 (4), 343–367.
Ferraresi, M., Todini, E., Vignoli, R., 1996. A solution to the inverse problem in groundwater hydrology based on Kalman filtering. J. Hydrol. 175 (1–4), 567–581.
Gusyev, M.A., Toews, M., Morgenstern, U., Stewart, M., White, P., Daughney, C., Hadfield, J., 2013. Calibration of a transient transport model to tritium data in streams and simulation of groundwater ages in the western Lake Taupo catchment, New Zealand. Hydrol. Earth Syst. Sci. 17 (3), 1217–1227.
Hantush, M.M., Mariño, M.A., 1997. Estimation of spatially variable aquifer hydraulic properties using Kalman filtering. J. Hydraul. Eng. 123 (11), 1027–1035.
Held, J., Schuhbeck, S., Rauert, W., 1992. A simplified method of 85Kr measurement for dating young groundwaters. Int. J. Radiat. Appl. Instrument. Part 43 (7), 939–942.
Hendricks Franssen, H.J., Kaiser, H.P., Kuhlmann, U., Bauser, G., Stauffer, F., Müller, R., Kinzelbach, W., 2011. Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. Water Resour. Res. 47 (2), W02532.
Hendricks Franssen, H.J., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the Ensemble Kalman Filter: joint estimation of states and parameters and the filter inbreeding problem. Water Resour. Res. 44 (9), W09408.
Hendricks Franssen, H.J., Stauffer, F., Kinzelbach, W., 2004. Joint estimation of transmissivities and recharges – application: stochastic characterization of well capture zones. J. Hydrol. 294, 87–102.
Hill, M.C., Tiedeman, C.R., 2007. Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty. John Wiley & Sons.
Hoehn, E., von Gunten, H.R., 1989. Radon in groundwater – a tool to assess infiltration from surface waters to aquifers. Water Resour. Res. 25 (8), 1795–1803.
Huber, E., Hendricks-Franssen, H.J., Kaiser, H.P., Stauffer, F., 2011. The role of prior model calibration on predictions with ensemble Kalman filter. Ground Water 49 (6), 845–858.
International Atomic Energy Agency/World Meteorological Organization (IAEA/WMO), 2012. Global Network of Isotopes in Precipitation. The GNIP Database. Accessible at: <http://www.iaea.org/water>.
Kruseman, G., de Ridder, N., 1990. Analysis and Evaluation of Pumping Test Data, vol. 47. Publication. Internat. Inst. Land Reclam. Improv., Wageningen, NL.
Li, W., Englert, A., Cirpka, O.A., Vanderborght, J., Vereecken, H., 2007. 2-d characterization of hydraulic heterogeneity by multiple pumping tests. Water Resour. Res. 43 (4), W04433.
Li, L., Zhou, H., Hendricks Franssen, H.J., Gómez-Hernndez, J.J., 2012. Modeling transient groundwater flow by coupling ensemble Kalman filtering and upscaling. Water Resour. Res. 48, W01537.
Lin, D., Jin, M., Liang, X., Zhan, H., 2013. Estimating groundwater recharge beneath irrigated farmland using environmental tracers fluoride, chloride and sulfate. Hydrogeol. J. 21 (7), 1469–1480.
Loschko, M., Wöhling, T., Rudolph, D.L., Cirpka, O.A., 2016. Cumulative relative reactivity: a concept for modeling aquifer-scale reactive transport. Water Resour. Res. 52, 8117–8137.
Massoudieh, A., 2013. Inference of long-term groundwater flow transience using environmental tracers: a theoretical approach. Water Resour. Res. 49 (12), 8039–8052.
McCallum, J.L., Cook, P.G., Simmons, C.T., 2015. Limitations of the use of environmental tracers to infer groundwater age. Groundwater 53, 56–70.
Nowak, W., 2009. Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator. Water Resour. Res. 45, W04431.
Pollock, D.W., 1988. Semianalytical computation of path lines for finite-difference models. Ground Water 26 (6), 743–750.
Poreda, R.J., Cerling, T.E., Salomon, D.K., 1988. Tritium and helium isotopes as hydrologic tracers in a shallow unconfined aquifer. J. Hydrol. 103 (1–2), 1–9.
Portniaguine, O., Solomon, D.K., 1998. Parameter estimation using groundwater age and head data, Cape Cod, Massachusetts. Water Resour. Res. 34 (4), 637–645.
Rueedi, J., Brennwald, M.S., Purtschert, R., Beyerle, U., Hofer, M., Kipfer, R., 2005. Estimating amount and spatial distribution of groundwater recharge in the Lullemmeden basin (Niger) based on 3H, 3He and CFC-11 measurements. Hydrol. Process. 19 (17), 3285–3298.
Sanchez-Vila, X., Guadagnini, A., Carrera, J., 2006. Representative hydraulic conductivities in saturated groundwater flow. Rev. Geophys. 44 (2005), 1–46.
Sanz-Prat, A., Lu, C., Finkel, M., Cirpka, O.A., 2015. On the validity of travel-time based nonlinear bioreactive transport models in steady-state flow. J. Contam. Hydrol., 175–176 (26–43)
Scanlon, B., Healy, R., Cook, P., 2002. Choosing appropriate techniques for quantifying groundwater recharge. Hydrogeol. J. 10, 18–39.
Schwientek, M., Maloszewski, P., Einsiedl, F., 2009. Effect of the unsaturated zone thickness on the distribution of water mean transit times in a porous aquifer. J. Hydrol. 373 (3–4), 516–526.
Stewart, M.K., Morgenstern, U., 2016. Importance of tritium-based transit times in hydrological systems. Wiley Interdiscip. Rev. Water 3 (2), 145–154.
Suckow, A., 2014. The age of groundwater – definitions, models and why we do not need this term. Appl. Geochem. 50, 222–230.
Tong, J., Hu, B.X., Yang, J., 2010. Using data assimilation method to calibrate a heterogeneous conductivity field conditioning on transient flow test data. Stoch. Environ. Res. Risk Assess. 24 (8), 1211–1223.
Tong, J., Hu, B.X., Yang, J., 2011. Assimilating transient groundwater flow data via a localized ensemble Kalman filter to calibrate a heterogeneous conductivity field. Stoch. Environ. Res. Risk Assess. 26 (3), 467–478.
Tong, J., Hu, B.X., Yang, J., 2013. Data assimilation methods for estimating a heterogeneous conductivity field by assimilating transient solute transport data via ensemble Kalman filter. Hydrol. Process. 27 (26), 3873–3884.
Turnadge, C., Smerdon, B.D., 2014. A review of methods for modelling environmental tracers in groundwater: advantages of tracer concentration simulation. J. Hydrol. 519, 3674–3689.
Vrugt, J.A., Stauffer, P.H., Wöhling, T., Robinson, B.A., Vesselinov, V.V., 2008. Inverse modeling of subsurface flow and transport properties: a review with new developments. Vadose Zo. J. 7 (2), 843–864.
Xu, T., Jaime Gómez-Hernández, J., Li, L., Zhou, H., 2013a. Parallelized ensemble Kalman filter for hydraulic conductivity characterization. Comput. Geosci. 52, 42–49.
Xu, T., Jaime Gómez-Hernández, J., Zhou, H., Li, L., 2013b. The power of transient piezometric head data in inverse modeling: an application of the localized normal-score EnKF with covariance inflation in a heterogenous bimodal hydraulic conductivity field. Adv. Water Resour. 54, 100–118.
Zhou, H., Gómez-Hernández, J.J., Li, L., 2014. Inverse methods in hydrogeology: evolution and recent trends. Adv. Water Resour. 63, 22–37.
Zlotnik, V., Zurbuchen, B., Ptak, T., Teutsch, G., 2000. Support volume and scale effect in hydraulic conductivity: experimental aspects. In: Zhang, D., Winter, C.L. (Ed.). Theory, Modelling, and Field Investigation in Hydrogeology: A Special Volume in Honor of Shlomo P. Neuman's 60th Birthday. No. 348 in Geological Society of America Special Papers. pp. 215–231, Symposium on Theory, Modeling, and Field Investigation in Hydrogeology in Honor of Shlomo P Neumans 60th Birthday, Tuscon, AZ, OCT 17, 1998.
Zuber, A., Rózański, K., Kania, J., Purtschert, R., 2011. On some methodological problems in the use of environmental tracers to estimate hydrogeologic parameters and to calibrate flow and transport models. Hydrogeol. J. 19 (1), 53–69.