



# Is high-resolution inverse characterization of heterogeneous river bed hydraulic conductivities needed and possible?

W. Kurtz<sup>1</sup>, H.-J. Hendricks Franssen<sup>1</sup>, P. Brunner<sup>2</sup>, and H. Vereecken<sup>1</sup>

<sup>1</sup>Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences: Agrosphere (IBG-3), 52425 Jülich, Germany

<sup>2</sup>University of Neuchâtel, Centre for Hydrogeology and Geothermics, 2000 Neuchâtel, Switzerland

Correspondence to: W. Kurtz (w.kurtz@fz-juelich.de)

Received: 4 April 2013 – Published in Hydrol. Earth Syst. Sci. Discuss.: 7 May 2013

Revised: 14 August 2013 – Accepted: 18 August 2013 – Published: 7 October 2013

**Abstract.** River–aquifer exchange fluxes influence local and regional water balances and affect groundwater and river water quality and quantity. Unfortunately, river–aquifer exchange fluxes tend to be strongly spatially variable, and it is an open research question to which degree river bed heterogeneity has to be represented in a model in order to achieve reliable estimates of river–aquifer exchange fluxes. This research question is addressed in this paper with the help of synthetic simulation experiments, which mimic the Limmat aquifer in Zurich (Switzerland), where river–aquifer exchange fluxes and groundwater management activities play an important role. The solution of the unsaturated–saturated subsurface hydrological flow problem including river–aquifer interaction is calculated for ten different synthetic realities where the strongly heterogeneous river bed hydraulic conductivities ( $L$ ) are perfectly known. Hydraulic head data (100 in the default scenario) are sampled from the synthetic realities. In subsequent data assimilation experiments, where  $L$  is unknown now, the hydraulic head data are used as conditioning information, with the help of the ensemble Kalman filter (EnKF). For each of the ten synthetic realities, four different ensembles of  $L$  are tested in the experiments with EnKF; one ensemble estimates high-resolution  $L$  fields with different  $L$  values for each element, and the other three ensembles estimate effective  $L$  values for 5, 3 or 2 zones. The calibration of higher-resolution  $L$  fields (i.e. fully heterogeneous or 5 zones) gives better results than the calibration of  $L$  for only 3 or 2 zones in terms of reproduction of states, stream–aquifer exchange fluxes and parameters. Effective  $L$  for a limited number of zones cannot always reproduce the true states and fluxes well and results in biased estimates of net exchange fluxes between aquifer and stream.

Also in case only 10 head data are used for conditioning, the high-resolution characterization of  $L$  fields with EnKF is still feasible. For less heterogeneous river bed hydraulic conductivities, a high-resolution characterization of  $L$  is less important. When uncertainties in the hydraulic parameters of the aquifer are also regarded in the assimilation, the errors in state and flux predictions increase, but the ensemble with a high spatial resolution for  $L$  still outperforms the ensembles with effective  $L$  values. We conclude that for strongly heterogeneous river beds the commonly applied simplified representation of the streambed, with spatially homogeneous parameters or constant parameters for a few zones, might yield significant biases in the characterization of the water balance. For strongly heterogeneous river beds, we suggest adopting a stochastic field approach to model the spatially heterogeneous river beds geostatistically. The paper illustrates that EnKF is able to calibrate such heterogeneous streambeds on the basis of hydraulic head measurements, outperforming zonation approaches.

## 1 Introduction

It is now well known that rivers and streams closely interact with the adjacent groundwater body (Bouwer and Madock, 1997; Winter, 1999; Sophocleous, 2002). These interactions have a number of consequences on the hydrological, chemical and biological environment around streams. For example, the resulting exchange fluxes between these two compartments influence the regional water balance and groundwater flow (Woessner, 2000) and thus also affect the yield of management activities close to streams, such as river

bank filtration (Zhang et al., 2011; Schubert, 2002). Additionally, the different chemical composition of river water and groundwater also has implications on chemical and ecological processes around streams (Brunke and Gonser, 1997; Sophocleous, 2002).

The exchange mechanisms between river systems and aquifers are complex and mainly depend on the pressure difference between stream and aquifer (Sophocleous, 2002), the form of the river bed (Cardenas et al., 2004; Boano et al., 2006), hydraulic properties of the river bed and the adjacent aquifer (Genereux et al., 2008) and the state of hydraulic connection between river and groundwater (Brunner et al., 2009). Exchange fluxes can exhibit a high degree of spatial and temporal variability, which is often related to the spatial heterogeneity of hydraulic parameters of the river bed and the adjacent aquifer (Conant, 2004; Rosenberry and Pitlick, 2009; Genereux et al., 2008), making the model-based prediction of exchange fluxes challenging. Calver (2001) compared literature data on river bed conductivities which ranged from  $10^{-9}$  to  $10^{-2}$   $\text{ms}^{-1}$  with a concentration of values in the range of  $10^{-7}$  to  $10^{-3}$   $\text{ms}^{-1}$ . A striking feature in this data compilation is that estimated river bed permeabilities can also vary considerably for a single measurement site. This variability of hydraulic river bed properties can be found at different scales along a river reach. Genereux et al. (2008) determined river bed conductivities with permeameter tests for 46 locations along a 262 m-long river reach. They observed a spatial variation in hydraulic conductivity of nearly four orders of magnitude ranging from approximately  $1 \times 10^{-7}$  to  $7.5 \times 10^{-4}$   $\text{ms}^{-1}$ . They also found that measured river bed conductivities had a bimodal distribution and tended to be higher in the middle of the stream. Hatch et al. (2010) estimated river bed conductivities along a 11 km-long river reach of the Pajaro River and determined values ranging from  $10^{-6}$  to  $10^{-4}$   $\text{ms}^{-1}$ . Springer et al. (1999) determined hydraulic conductivities for five reattachment bars of the Colorado River over a range of 200 miles. Measured hydraulic conductivities varied over 2 orders of magnitude within the reattachment bars, and differences between the medians of the five reattachment bars were up to one log unit.

Different modelling studies have already tried to assess the consequences of river bed and aquifer heterogeneity on the prediction of exchange fluxes between streams and groundwater. For example, Bruen and Osman (2004) investigated the impact of heterogeneous aquifer hydraulic conductivities on river–aquifer exchange fluxes with a synthetic 2-D stream–aquifer model. They compared Monte Carlo simulations using heterogeneous fields of hydraulic conductivity with simulations using homogeneous fields. This comparison was made for different geostatistical parameters and connection regimes between river and aquifer. They found that the uncertainty in fluxes increases with an increasing degree of heterogeneity. They also found that a homogeneous and an ensemble of heterogeneous  $K$  fields gave similar results under connected conditions but different results for a

disconnected regime and unsaturated conditions below the river bed. Fleckenstein et al. (2006) compared simulation results for six realizations of geostatistically simulated facies distributions with a homogeneous aquifer model and found comparable net seepage fluxes for the different models. However, they also identified that the different facies distributions show considerable variability in the spatial distribution of seepage fluxes and in the state of connection between stream and aquifer. Kalbus et al. (2009) investigated the effect of heterogeneous conductivities within the streambed and the adjacent aquifer by simulating 2-D groundwater flow and heat transport using the leakage concept. They found that the heterogeneity of aquifer properties has more impact on river–aquifer exchange than that of the streambed. However, they also mention that homogeneous streambeds lead to an unrealistic homogenization of water fluxes between river and aquifer. Frei et al. (2009) simulated the spatio-temporal distribution of seepage fluxes for a losing river reach in a Monte Carlo framework. They applied a hydrofacies model for the distribution of hydraulic conductivities and found that highly permeable parts of the river reach ( $\approx 50\%$  of total length) make up 98% of total seepage within their simulations. They argue that heterogeneity at the hydrofacies scale dominates the spatial pattern of river–aquifer interactions and that within-facies heterogeneity is of minor importance.

These different studies emphasize that the incorporation of heterogeneity in models for river–aquifer exchange can be important for a reliable prediction of exchange fluxes. In practical applications river bed conductivities are mostly estimated through calibration but heterogeneity is often neglected in the calibration procedure. One reason is that measurements of river bed conductivities are usually scarce and an estimation of the corresponding heterogeneity would require intensive field measurements. Especially for larger streams, in situ measurements are difficult to perform because of the higher discharge. As a consequence, in most cases there is only limited prior knowledge on the heterogeneity of hydraulic parameters for a certain site. Another reason is that the computational demand for inversions with gradient-based methods and also the complexity of the inversion increases with a higher-resolution representation of heterogeneity in the model. Due to a lack of prior knowledge on the magnitude and variability of river bed properties and in order to ease the inversion procedure, leakage parameters are often lumped together in leakage zones; i.e. the inversion is only done for very few parts of the river and the underlying heterogeneity is reduced to the different leakage zones. Irvine et al. (2012) carried out a systematic analysis on the simplification of heterogeneity to quantify its implications on the prediction of infiltration fluxes. They simulated infiltration curves (i.e. the relationship between water table depth and infiltration flux) for a variety of heterogeneous distributions of river bed conductivities. Different data points from these synthetic infiltration curves were then used to calibrate models with a homogeneous distribution of river bed

conductivities. Forward simulations with the derived homogeneous values of river bed conductivities were subsequently used to compare the simulated infiltration curves with the equivalent ones for the fully heterogeneous medium. They found that the calibrated homogeneous models reproduced exchange fluxes well when the state of connection between stream and aquifer was equal for calibration and prediction of the homogeneous models. However, when the state of connection was transitional or differed between calibration and prediction, the homogeneous models could not adequately reproduce the infiltration fluxes of the corresponding heterogeneous references.

One way to account for the underlying heterogeneity of river–aquifer systems in the calibration of groundwater models in a stochastic framework is the ensemble Kalman filter (EnKF) (Evensen, 1994). EnKF and its variants have already been applied successfully for the characterization of heterogeneous subsurface properties in groundwater modelling. Examples are the work of Chen and Zhang (2006), Hendricks Franssen and Kinzelbach (2008), and Nowak (2009), who assimilated piezometric heads with EnKF to improve the estimation of hydraulic conductivity fields. Sun et al. (2009) assessed the value of different deterministic ensemble filters for subsurface characterization. Liu et al. (2008) and Li et al. (2012) used tracer data, and Camporese et al. (2011) assimilated time-lapse data of electrical resistivity tomography to infer hydraulic conductivity fields. The studies of Jafarpour and Tarrahi (2011) and Huber et al. (2011) addressed the influence of variogram uncertainty and prior information in the initial parameter ensemble on the estimation of hydraulic conductivity fields with EnKF. Zhou et al. (2011) and Schoeniger et al. (2012) investigated techniques to account for non-Gaussianity in the assimilation with EnKF. A general advantage of ensemble-based data assimilation with EnKF and its variants is that they are able to calibrate model parameters on the basis of the forward integration of an ensemble of different parameter fields and therefore explicitly account for the high variability of hydraulic parameters in natural settings. Therefore, this methodology should also be well suited for the characterization of highly variable river bed properties. In Kurtz et al. (2012) EnKF has been applied specifically to river–aquifer systems in order to identify the temporal change of river bed conductivities. In this study we concentrate on the question whether the estimation of a few effective values for river bed hydraulic conductivity can reproduce spatially and temporally strongly variable river–aquifer exchange fluxes with the use of data assimilation. For this purpose we compare different ensembles of leakage parameters that either resemble the fully heterogeneous structure of different synthetic reference fields or where different degrees of spatial aggregation are used. Our specific research questions are as follows:

- To which degree does a zonation of river bed properties change predicted exchange fluxes between river and

aquifer compared to a full representation of river bed heterogeneity?

- Is EnKF able to identify the main structural features of a fully heterogeneous field of river bed conductivities through assimilation of hydraulic head measurements?
- How does EnKF perform for different parameterization approaches (i.e. detailed representation of heterogeneity versus few zones) under conditions with different amounts of observation data and additional sources of uncertainty?

## 2 Data assimilation with the ensemble Kalman filter

### 2.1 General description of the data assimilation algorithm

The ensemble Kalman filter (EnKF) is a sequential data assimilation approach which can be utilized to improve the prediction capability of a particular model in a Monte Carlo framework. The methodology was originally applied for atmospheric and oceanographic models (e.g. Evensen, 1994; Houtekamer and Mitchell, 1998) and later used in modified variants, which include parameter estimation in surface hydrology (e.g. Moradkhani et al., 2005) and subsurface hydrology (e.g. Chen and Zhang, 2006; Liu et al., 2008; Hendricks Franssen and Kinzelbach, 2008; Nowak, 2009). The basic idea of EnKF is that different model realizations (e.g. with different forcings and/or parameters) are propagated forward in time until state measurements become available. The predicted states are then improved by optimally combining the ensemble of model predictions and measurement data. EnKF can also be used to jointly estimate model states and parameters. In this case the model states  $\psi_i$  and the model parameters  $\phi_i$  for realization  $i$  are combined in the state–parameter vector  $\Psi_i$  (e.g. Chen and Zhang, 2006; Hendricks Franssen and Kinzelbach, 2008):

$$\Psi_i = \begin{pmatrix} \psi_i \\ \phi_i \end{pmatrix}. \quad (1)$$

For each assimilation cycle (i.e. at times when measurements become available) EnKF then performs a Bayesian update on the ensemble of the state–parameter vector  $\Psi$ .

$$p(\Psi | \mathbf{y}^0) \propto p(\mathbf{y}^0 | \Psi) p(\Psi), \quad (2)$$

where  $p(\Psi | \mathbf{y}^0)$  is the posterior (updated) distribution of the state–parameter vector  $\Psi$  given the observations  $\mathbf{y}^0$ ,  $p(\mathbf{y}^0 | \Psi)$  is the likelihood of measurements  $\mathbf{y}^0$  given  $\Psi$  and  $p(\Psi)$  is the prior distribution of  $\Psi$ . The prior distribution of the states  $\psi$  (as part of  $\Psi$ ) is usually obtained by advancing each realization  $i$  of the ensemble with a model  $M$ :

$$\psi_i^t = M\left(\psi_i^{t-1}, \phi_i, \zeta_i\right), \quad (3)$$

where  $\psi_i^t$  is the predicted state vector for time step  $t$ ,  $\psi_i^{t-1}$  is the predicted state vector of the preceding time step  $t-1$ ,  $\phi_i$  are model parameters and  $\zeta_i$  are model forcings.

The posterior distribution of  $\Psi$  in Eq. (2) is then calculated with the EnKF analysis scheme which proceeds in the following steps: first, the forecasted values of  $\psi$  at observation points have to be extracted from the ensemble:

$$\hat{\psi}_i^t = \mathbf{H}\psi_i^t, \quad (4)$$

where  $\hat{\psi}_i^t$  is the state vector at observation points and  $\mathbf{H}$  is a matrix that maps or interpolates the whole state vector  $\psi_i^t$  on the observation points. Next, the measured states  $y^0$  have to be perturbed with values from a normal distribution for each realization. This perturbation is necessary to derive the correct posterior variance of  $\Psi$  after the assimilation step (Burgers et al., 1998) and the magnitude of this perturbation is usually derived from the measurement error of the state measurement:

$$y_i = y^0 + \epsilon_i, \quad (5)$$

where  $y_i$  are the perturbed observations for realization  $i$ ,  $y^0$  is the measurement vector and  $\epsilon_i$  is the corresponding perturbation vector which is sampled from a normal distribution  $\mathcal{N}(0, R)$  with a mean of zero and a covariance that corresponds to the covariance matrix of observation data  $\mathbf{R}$ .

Finally, the posterior distribution of  $\Psi$  is found by applying the following equation on each ensemble member  $i$ :

$$\Psi_i^a = \Psi_i^t + \alpha \mathbf{G} (y_i - \hat{\psi}_i^t), \quad (6)$$

where  $\Psi_i^a$  is the analysed (updated) state–parameter vector for realization  $i$ ,  $\Psi_i^t$  is the forecasted state–parameter vector (with  $\psi_i^t$  from Eq. 3),  $\hat{\psi}_i^t$  is the simulated state vector at observation points and  $y_i$  is the perturbed measurement vector.  $\alpha$  is a damping factor which is used to decrease the effect of filter inbreeding for parameter updates (see Hendricks Franssen and Kinzelbach, 2008).  $\mathbf{G}$  is the Kalman gain which is calculated as follows:

$$\mathbf{G} = \mathbf{C}\mathbf{H}^T (\mathbf{H}\mathbf{C}\mathbf{H}^T + \mathbf{R})^{-1}, \quad (7)$$

where  $\mathbf{C}$  is the covariance matrix of  $\Psi^t$  and  $\mathbf{R}$  is the covariance matrix of observation data  $y^0$ . The Kalman gain  $\mathbf{G}$  is calculated once for all ensemble members and weights the uncertainties in the prediction of  $\Psi$  from the forward simulations with the measurement error of  $y^0$ . In the analysis step (Eq. 6) the weighted factors of  $\mathbf{G}$  are used to correct each member of  $\Psi_i^t$  with the residuals at observation points (bracketed term in Eq. 6). In order to derive  $\mathbf{G}$  it is not necessary to calculate the full covariance matrix of  $\Psi^t$ . Instead, it is sufficient to only calculate the covariances between  $\Psi^t$  and  $\hat{\psi}^t$ :

$$\mathbf{C}\mathbf{H}^T = \begin{pmatrix} \mathbf{C}_{\psi^t \hat{\psi}^t} \\ \mathbf{C}_{\phi \hat{\psi}^t} \end{pmatrix}. \quad (8)$$

## 2.2 Specific usage of EnKF for river–aquifer interactions

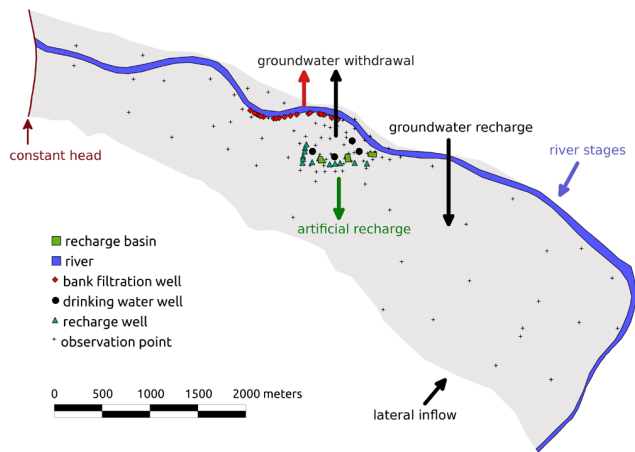
In this study the focus is on the investigation of river–aquifer exchange fluxes. Thus the model states of interest are piezometric heads  $h$ , and the most relevant model parameters are hydraulic conductivities of the river bed, which are implemented in the model as leakage coefficients  $L$ . Therefore, the state–parameter vector  $\Psi$  which was introduced in the previous section is composed of  $h$  and  $\log_{10}(L)$ . For a subset of simulations also the hydraulic conductivities of the aquifer ( $K$ ) are updated in the assimilation process. For these simulations  $\Psi$  consists of  $h$ ,  $\log_{10}(K)$  and  $\log_{10}(L)$ . The model  $M(\psi_i^{t-1}, \phi_i, \zeta_i)$  that is used to advance  $h$  in time is a groundwater model that is capable of simulating variably saturated flow and that includes a parameterization to simulate river–aquifer exchange fluxes. The observation data  $y^0$  consist of measurements of  $h$  in the aquifer. The covariance matrix  $\mathbf{R}$  includes the measurement errors of observation data on the diagonal, but covariances between observation points are assumed to be zero (i.e. measurement errors are assumed to be independent).

## 3 Model description

Synthetic data assimilation experiments were performed with a 3-D finite element model of the Limmat valley aquifer in Zurich (Switzerland). The boundary conditions for this model are given schematically in Fig. 1. Groundwater recharge is imposed as a flux boundary condition on the first model layer. Groundwater inflows from the surrounding hills on the northern and southern boundary of the model were also treated as flux boundary conditions. Within the model domain also management activities take place: groundwater is withdrawn close to the river through several bank filtration and drinking water wells. The pumped water from the bank filtration wells is recharged to the aquifer through several recharge wells and three recharge basins which are located south of the drinking water wells. At the western side of the model domain a constant head boundary condition is imposed. The two rivers in the model (Limmat and Sihl) reside at the northern and eastern boundary of the model respectively, and river stages are imposed on each river node of the model.

All model forcings (recharge, lateral inflows, river stages, pumping rates) are transient and based on real-world measurements for that site. A more detailed description on the calculation of forcing data is given in Hendricks Franssen et al. (2011).

The groundwater modelling software SPRING (Delta h Ingenieurgesellschaft mbH, 2006) was used to numerically solve the transient saturated/unsaturated flow equation for each realization of the ensemble. In SPRING, river–aquifer exchange fluxes for each leakage node  $i$  are calculated



**Fig. 1.** Schematic representation of model domain and boundary conditions.

according to a Cauchy-type boundary condition (leakage principle):

$$Q_i = L_i A_i (h_i^{\text{river}} - h_i^{\text{gw}}), \quad (9)$$

where  $Q_i$  is the volumetric flux between river and aquifer [ $L^3 T^{-1}$ ],  $L_i$  is leakage coefficient [ $T^{-1}$ ],  $A_i$  is the predefined area for each leakage node [ $L^2$ ],  $h_i^{\text{river}}$  is river stage [ $L$ ] and  $h_i^{\text{gw}}$  is groundwater level underneath the river [ $L$ ].

Internally, SPRING does not directly make use of the leakage coefficient  $L$  (as in Eq. 9) as an input parameter but instead requests leakage coefficients combined with areal information. In this study, the leakage coefficient  $L$  multiplied with half of the river width is used as an input parameter (denoted  $L^*$ ). The usage of  $L^*$  instead of  $L$  does not have a significant influence on the comparability of different simulation results because the river width is almost constant in the utilized model. Hence, the term leakage coefficient is used synonymously for  $L$  and  $L^*$  in the following.

Porosities in the model domain were set to a constant value of 0.15 for all simulations. Hydraulic conductivities of the aquifer ( $K$ ) were taken from a prior calibration of the model with a modified pilot point method (Alcolea et al., 2006) using 87 piezometric head measurements (for further details on the calibration see Hendricks Franssen et al., 2011; Huber et al., 2011). The calibrated  $K$  field was used for the majority of simulations performed in this study (for exception see below). Van Genuchten parameters for unsaturated flow were taken from the standard lookup table that is provided in the utilized groundwater modelling software SPRING (summarized in Table 1) and were the same for all conducted simulations.

**Table 1.** Lookup table for van Genuchten parameters in SPRING.  $S_{\text{res}}$  is residual saturation,  $\alpha_{\text{vg}}$  is the parameter related to the inverse of air entry suction and  $n$  is the parameter related to pore size distribution.

saturated hydraulic conductivity $K$ [ $\text{ms}^{-1}$ ]	van Genuchten parameters		
	$S_{\text{res}}$ [-]	$\alpha_{\text{vg}}$ [ $\text{m}^{-1}$ ]	$n$ [-]
$K \geq 9.81 \times 10^{-4}$	0.2	$1.37 \times 10^{-2}$	1.5
$9.81 \times 10^{-4} \geq K \geq 9.81 \times 10^{-7}$	0.4	$2.24 \times 10^{-3}$	1.35
$K \leq 9.81 \times 10^{-7}$	0.9	$1.23 \times 10^{-3}$	1.3

#### 4 Synthetic experiments

The general setup of the synthetic experiments consists of the following steps:

1. Generation of ten spatially heterogeneous distributions of  $\log_{10}(L^*)$  with sequential Gaussian simulation (SGS).
2. Finite elements solution of the transient variably saturated flow equation (for 609 days) using SPRING. A solution is calculated for each of the ten generated  $\log_{10}(L^*)$  fields of step 1. The calculated hydraulic heads and river–aquifer exchange fluxes for these ten simulations serve as the ten “true” reference solutions.
3. Generation of  $\log_{10}(L^*)$  ensembles (100 realizations) for a fully heterogeneous case and three different zonations (5, 3 and 2 leakage zones).
4. For each of the ten references (step 1 and 2): assimilation of hydraulic head data from step 2 with EnKF for the four  $\log_{10}(L^*)$  ensembles of step 3.

This procedure is used for four scenarios which differ in the degree of heterogeneity of  $\log_{10}(L^*)$  reference fields, the information content of the initial ensemble and the sources of parameter uncertainty:

- A Strongly heterogeneous  $\log_{10}(L^*)$  fields for references and initial ensembles (no prior information)
- B Moderately heterogeneous  $\log_{10}(L^*)$  fields for references and initial ensembles (no prior information)
- C Strongly heterogeneous  $\log_{10}(L^*)$  fields with prior information in the initial ensembles
- D Same as A but with uncertain  $\log_{10}(K)$  fields

Scenario A serves as a base scenario for which the relevant results of the simulation experiments are presented in detail. The results for scenarios B, C and D are then discussed with respect to deviations from scenario A. In this sense, scenario B is utilized to assess the performance of data assimilation for different degrees of heterogeneity and scenario C should

give insight into the value of prior information for the data assimilation with EnKF. Scenario D is used to investigate the effect of uncertainty in the aquifer hydraulic parameters on the assimilation process and parameter identifiability.

#### 4.1 Reference fields

The reference fields of leakage coefficients for scenario A were generated by adding perturbation fields to a predefined mean value of  $\log_{10}(L^*)$  of  $-2.78 \log_{10}(\text{m s}^{-1})$ . The perturbation fields were generated by SGS with the code GCOSIM3D (Gómez-Hernández and Journel, 1993) for the rivers Sihl and Limmat. These perturbation fields had a grid size of 50 m and a spatial extend of  $7000 \times 1000$  m for the river Limmat and  $2500 \times 500$  m for the river Sihl. A spherical variogram was chosen for geostatistical simulations of the ten perturbation fields. The nugget was set to  $0 \log_{10}(\text{m s}^{-1})$  for all simulations. The range of the variograms was sampled from a uniform distribution with values ranging from 1000 to 2000 m for each reference field. Values for the sill were also sampled from a uniform distribution ranging from 1 to  $2 \log_{10}(\text{m}^2 \text{s}^{-2})$ . The simulated fields were directed onto the main axis of the rivers, and the leakage coefficient  $\log_{10}(L^*)$  for each river node was determined by the overlying grid block of the geostatistically simulated perturbation field plus the predefined mean value of  $-2.78 \log_{10}(\text{m s}^{-1})$ . The different reference fields of leakage coefficients along the river reach are shown in Fig. 2.

For the creation of reference fields for scenario B a similar methodology as for scenario A was applied. The only difference between these two scenarios is that for scenario B a sill between 0.1 and  $0.5 \log_{10}(\text{m}^2 \text{s}^{-2})$  was used, which results in a lower degree of variability for these references.

For scenario C the reference fields of  $\log_{10}(L^*)$  have a predefined zonation with five leakage zones whose spatial location corresponds to the one of ensemble  $Z_5$  (see below). The predefined zonal values for  $\log_{10}(L^*)$  are summarized in Table 2. Similar to scenario A, perturbation fields were added to these predefined zonal values. These perturbation fields were created in a similar way as for scenario A with a nugget of  $0 \log_{10}(\text{m s}^{-1})$ , a sill of  $1 \log_{10}(\text{m}^2 \text{s}^{-2})$  and a range of 600 m. Compared to scenario A these reference fields include a higher contrast between different parts of the river reach (realized through the predefined zonation) and an additional component of intrazonal variability (realized through the perturbation fields).

For all reference runs of the different scenarios the  $K$  field from the prior calibration of the model (see Sect. 3) was used.

#### 4.2 Zonation

For the assimilation experiments four ensembles of leakage coefficients were generated which differed in their spatial representation of heterogeneity. The first ensemble  $Z_{\text{het}}$  represents the full heterogeneity of the reference fields, and the

**Table 2.** Predefined zonal leakage values for generation of reference fields for scenario C. Leakage zones are numbered from west to east.

leakage zone	$\log_{10}(L^*_{\text{zonal}})$ [ $\log_{10}(\text{m s}^{-1})$ ]
I	-5.51
II	-1.96
III	-3.88
IV	-5.37
V	-6.44

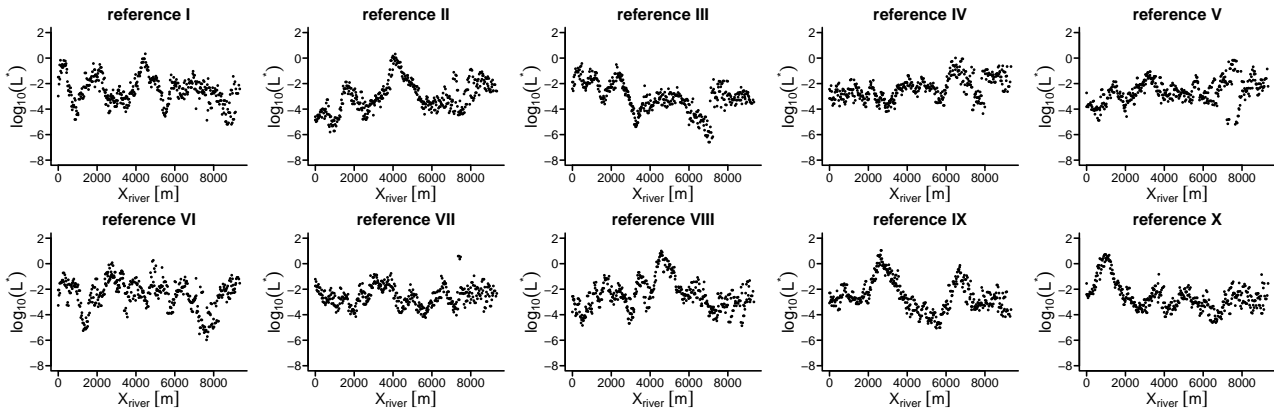
number of zones is equal to the number of river nodes (i.e. 457). The second ensemble  $Z_5$  only represents 5 leakage zones, which were positioned according to the main hydrological features of the river reach (i.e. position of two weirs, confluence of the rivers Sihl and Limmat), and the groundwater management activities at this site, which resulted in four leakage zones for the river Limmat and one leakage zone for the river Sihl. For the third ensemble  $Z_3$  the river Limmat is divided into two leakage zones and the river Sihl is the third leakage zone. For the fourth ensemble  $Z_2$  the river Limmat is aggregated to one leakage zone and again the river Sihl serves as a separate leakage zone. The spatial arrangement of leakage parameters for the fully heterogeneous case ( $Z_{\text{het}}$ ) and the three zonation approaches ( $Z_5$ ,  $Z_3$  and  $Z_2$ ) is depicted in Fig. 3.

#### 4.3 Ensemble generation

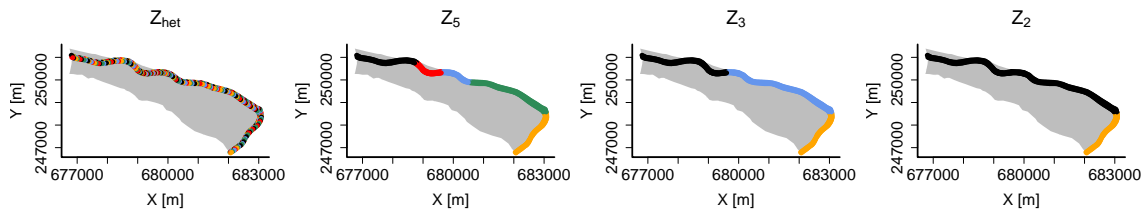
The generation of the ensembles for  $Z_{\text{het}}$  for the different scenarios corresponded closely to the generation of the respective reference fields. However, a higher degree of uncertainty with respect to geostatistical parameters was used for the ensembles. For scenario A the range parameter varied between 50 and 5000 m and the sill value between 0.1 and  $3.0 \log_{10}(\text{m}^2 \text{s}^{-2})$ . For scenario B the sill for  $Z_{\text{het}}$  had values between 0.1 and  $1.5 \log_{10}(\text{m}^2 \text{s}^{-2})$  and the range varied between 50 and 5000 m. For scenario C the ensemble for  $Z_{\text{het}}$  was generated with the same geostatistical parameters as for scenario A.

For each scenario, the initial  $\log_{10}(L^*)$  ensembles of  $Z_5$ ,  $Z_3$  and  $Z_2$  were derived from the corresponding initial  $\log_{10}(L^*)$  ensemble of  $Z_{\text{het}}$ . This was done by calculating the arithmetic average of  $\log_{10}(L^*)$  values for each realization of  $Z_{\text{het}}$  according to the respective zonation scheme. For example, the value of  $\log_{10}(L^*)$  for one of the five leakage zones of  $Z_5$  for a single realization is calculated from the corresponding realization of  $Z_{\text{het}}$  by averaging the  $\log_{10}(L^*)$  of  $Z_{\text{het}}$  that are within the respective zone of  $Z_5$ . This procedure is then repeated for all leakage zones and all realizations of  $Z_5$ . In Fig. 4 the generated ensembles for scenario A are compared to reference field I.

For scenarios A, B and C the utilized  $K$  fields were identical to the ones of the reference field; i.e. the only source



**Fig. 2.** Reference fields of leakage coefficients for scenario A along the river reach. Distance  $X_{\text{river}}$  is calculated starting from the two most western river nodes.  $L^*$  has units  $\text{m s}^{-1}$ .



**Fig. 3.** Spatial representation of heterogeneity for four parameter ensembles. In the fully heterogeneous case ( $Z_{\text{het}}$ ) an individual leakage coefficient is assigned to each river node. For the zoned ensembles ( $Z_5$ ,  $Z_3$  and  $Z_2$ ) each colour corresponds to a separate leakage zone. The river nodes within a leakage zone share the same leakage coefficient.

of uncertainty are the unknown leakage parameters. For scenario D also the initial  $K$  fields of the different ensembles were made uncertain and  $K$  was jointly updated together with  $h$  and  $L$ . The initial  $K$  fields were the same for all four ensembles and were generated by adding perturbation fields to the calibrated  $K$  field of the reference runs. These perturbation fields were generated by SGS with the code GCOSIM3D (Gómez-Hernández and Journel, 1993) on a very fine grid ( $1 \text{ m} \times 1 \text{ m} \times 1 \text{ cm}$ ) and then upscaled to the model grid through simplified renormalization (Renard et al., 2000). The geostatistical parameters for SGS (nugget:  $0 \log_{10} (\text{m s}^{-1})$ ; sill:  $0.584 \log_{10} (\text{m}^2 \text{ s}^{-2})$ ; range in horizontal direction: 99 m; range in vertical direction: 3.2 m) were derived from real-world measurements of hydraulic conductivities for the Hardhof site (approx. 800  $K$  measurements from flow meter surveys and pumping tests).

**4.4 Settings for data assimilation with EnKF**

The meta parameters for data assimilation experiments for scenario A, B, C and D with EnKF are summarized in Table 3. The value for the damping factor  $\alpha$  of 0.1 was based on recommendations from Hendricks Franssen and Kinzelbach (2008) and was the same for all calculations. The updating frequency of 10 days was chosen in order to allow the model states to synchronize to the updated model parameters

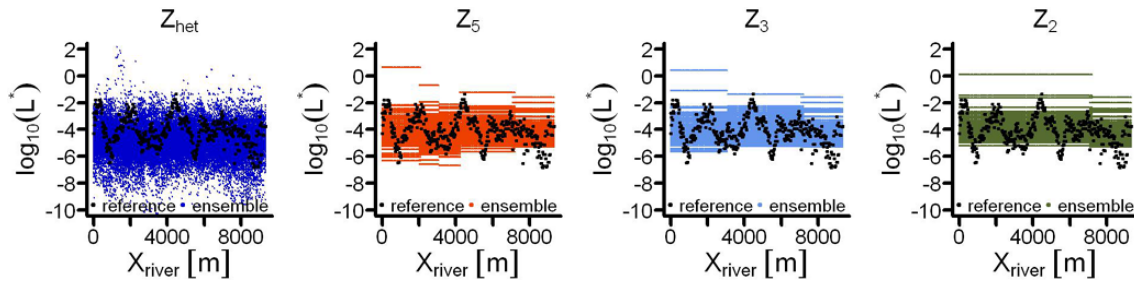
between the updating cycles. This setting was used to reduce the problem of inconsistency between model states and parameters after the update with EnKF. For all four scenarios 100 observation points were used as input data for EnKF. Scenario A was additionally simulated with a lower amount of observations (10 measurements). The other settings for assimilation with EnKF were held constant for all scenarios.

**4.5 Performance assessment of simulations**

The performance of the data assimilation experiments is assessed by the prediction error of hydraulic heads throughout the model domain, the prediction error of river–aquifer exchange fluxes and the correction of leakage coefficients during the update. For the prediction error of hydraulic heads the root mean square error ( $\text{RMSE}_h$ ) between the predicted mean hydraulic head and the hydraulic head of the reference is calculated using

$$\text{RMSE}_h(t) = \sqrt{\frac{1}{N_{\text{nodes}}} \sum_i^{N_{\text{nodes}}} (\bar{h}_i(t) - h_i^{\text{ref}}(t))^2}, \quad (10)$$

where  $\bar{h}_i$  is the mean hydraulic head for model node  $i$  [L],  $h_i^{\text{ref}}$  is the hydraulic head of the reference simulation for node  $i$  [L],  $N_{\text{nodes}}$  is the total number of model nodes and  $t$  is time step [T].



**Fig. 4.** Initial ensembles of leakage coefficients (coloured) and reference field I (black) for scenario A for all river nodes (rivers Limmat and Sihl) along the river reach. Distance  $X_{\text{river}}$  is calculated starting from the two most western river nodes.  $L^*$  has units  $\text{m s}^{-1}$ .

**Table 3.** Default settings for data assimilation with EnKF.

parameter	value
ensemble size	100
observation points	100 (10)
damping factor $\alpha$	0.1
update frequency	10 days
measurement error	0.05 m

For the evaluation of river–aquifer exchange we present either the evolution of the leakage fluxes over time  $Q(t)$  or statistics for the total volume of water that was exchanged between river and aquifer during the whole simulation period  $\Delta V_{\text{tot}}$ :

$$Q(t) = \sum_i^{N_{\text{leak}}} Q_i(t) \quad (11)$$

$$\Delta V_{\text{tot}} = \sum_j^{t_{\text{tot}}} \sum_i^{N_{\text{leak}}} Q_{ij} \Delta t_j, \quad (12)$$

where  $Q(t)$  is the river–aquifer exchange flux over time [ $\text{L}^3 \text{T}^{-1}$ ],  $Q_i(t)$  is the leakage flux for river node  $i$  over time,  $N_{\text{leak}}$  is total number of leakage nodes,  $\Delta V_{\text{tot}}$  is the volume of water that is exchanged between river and aquifer over the whole simulation period [ $\text{L}^3$ ],  $Q_{ij}$  is the volumetric flux between river and aquifer for the  $i$ th leakage node and the  $j$ th time step [ $\text{L}^3 \text{T}^{-1}$ ],  $\Delta t_j$  is time step [T] and  $t_{\text{tot}}$  is total number of time steps.  $Q(t)$  was calculated direction dependent; i.e. fluxes from river to aquifer (positive) and fluxes from aquifer to river (negative) were summed up separately.

#### 4.6 Computational requirements for data assimilation experiments

In general, data assimilation experiments require a high amount of computational resources due to the forward propagation of multiple instances of the same model. For our simulations, one assimilation run consists of 100 forwards runs

of the model (each with 609 time steps), which translates into 16 000 forward runs for all scenarios. In order to reduce the total runtime of the simulations we used a parallel implementation of EnKF in which multiple realizations are processed at once (each model realization using a single CPU). With this implementation, the runtime for one assimilation experiment was about 16 hours using 8 CPUs. Simulations were performed on the supercomputing facilities (JUROPA) of Forschungszentrum Jülich (Germany).

## 5 Results

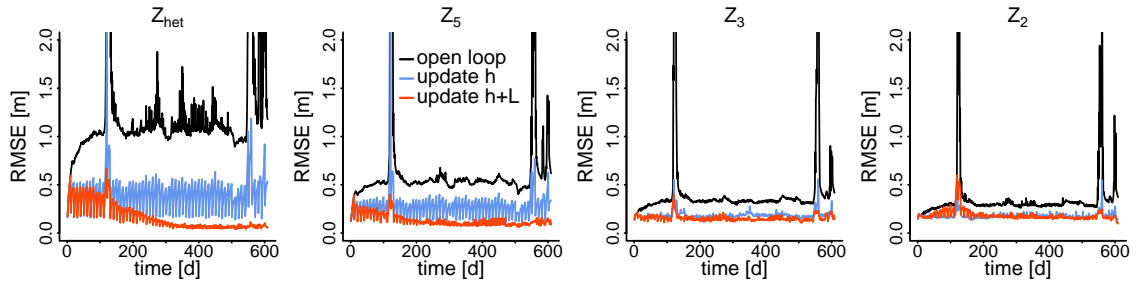
### 5.1 Strongly heterogeneous case (scenario A)

In order to get a first overview on the worth of data assimilation, Fig. 5 compares  $\text{RMSE}_h$  for reference I with open-loop simulations (no update with EnKF) and simulations where only  $h$  was updated with measurements. Generally, head predictions improve through the update with EnKF, and this improvement is stronger when also  $\log_{10}(L^*)$  values are updated. The errors for the open-loop simulations differ between the four ensembles especially between  $Z_{\text{het}}$  and the zonated ensembles. This is related to the fact that the zonated ensembles were derived through spatial averaging of the  $\log_{10}(L^*)$  ensemble of  $Z_{\text{het}}$ . As a consequence, the effect of state and parameter updates is larger for  $Z_{\text{het}}$  than for the other ensembles.

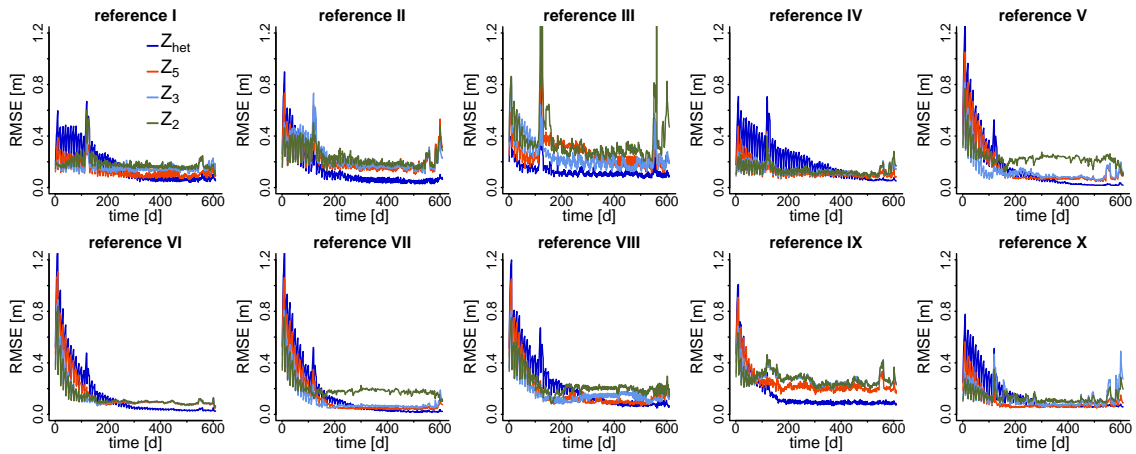
Figure 6 compares  $\text{RMSE}_h$  of the four  $\log_{10}(L^*)$  ensembles for all ten highly heterogeneous reference fields of scenario A for a joint state–parameter update (see also Table 5 for a quantitative overview on  $\text{RMSE}_h$  values). The highest improvement is observed for  $Z_{\text{het}}$  where  $\text{RMSE}_h$  is consistently reduced to about 0.1 m among all references. For the other ensembles the performance in terms of  $\text{RMSE}_h$  is more dependent on the specific reference. For example,  $\text{RMSE}_h$  of  $Z_5$  is similar to the one of  $Z_{\text{het}}$  for some references (e.g. IV, VII and X) but is worse for other references (e.g. II and IX). A similar behaviour can be observed for  $Z_3$  and  $Z_2$ .

The updated net fluxes between river and aquifer are shown in Fig. 7 for reference field I. For this reference field the updating with EnKF led to an improvement of the





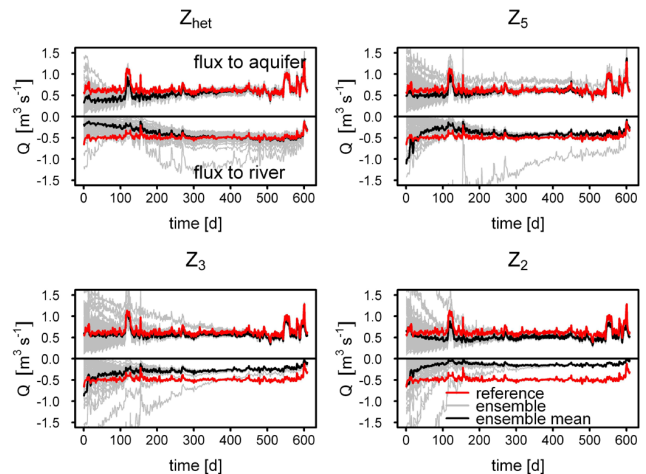
**Fig. 5.**  $RMSE_h$  for open-loop simulations, update of hydraulic heads and joint update of hydraulic heads and leakage coefficients (scenario A). Four different ensembles are compared.



**Fig. 6.**  $RMSE_h$  of conditional simulations with EnKF for ten reference fields of leakage coefficients (scenario A). Four different ensembles are compared. EnKF jointly updated hydraulics heads and leakage coefficients with 100 piezometric head data.

prediction of fluxes from river to aquifer for all four ensembles. Fluxes from aquifer to river were well reproduced by  $Z_{het}$  and  $Z_5$ , whereas for  $Z_3$  and  $Z_2$  a larger deviation between ensemble mean and true values was found. The ensemble variance of leakage fluxes decreases very fast, with most of the decrease happening in the first 100 simulation days. One exception is the flux from aquifer to river for  $Z_{het}$ . The general decrease of variance within the first 100 days is observable for all ten references, whereas the higher variability for  $Z_{het}$  was a special feature of reference I.

An overview of the net exchange between river and aquifer for all ten references is given in Fig. 8. Here the total amount of exchanged water summed over the complete simulation period ( $\Delta V_{tot}$ ) is displayed for each of the reference fields and for each of the zonation approaches. The net exchange for  $Z_{het}$  is very close to the net exchange of the respective reference and thus shows the best performance among the four ensembles.  $Z_3$  and  $Z_2$  show a good fit for some references (e.g. reference IV), but the fluxes are significantly over- or underestimated for other references.  $Z_5$  does not show the partially high deviations from the references that can be observed for  $Z_3$  and  $Z_2$ . However, the performance of  $Z_5$  is not better than that of  $Z_{het}$  in the majority of cases.



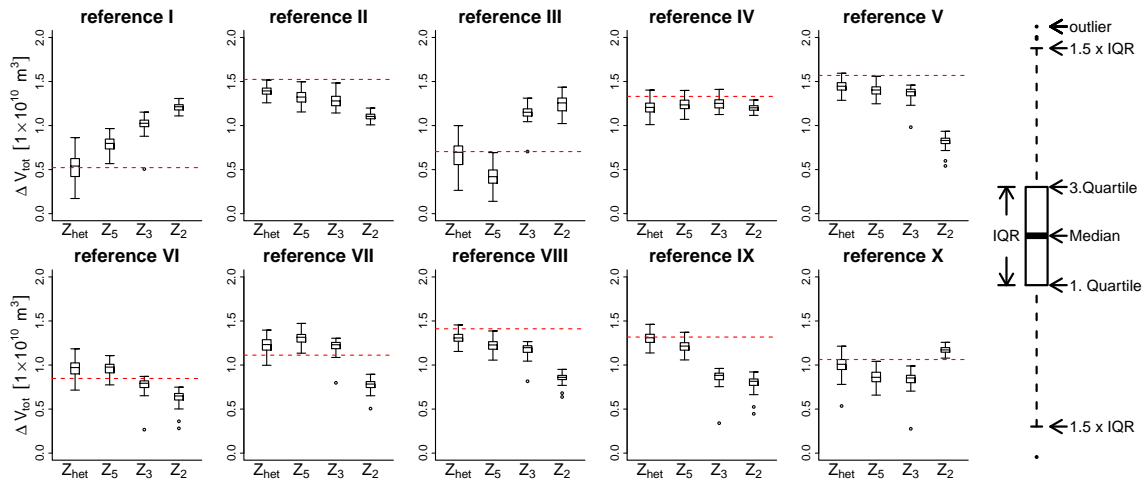
**Fig. 7.** Cumulative directional fluxes between river and aquifer for conditional simulations with EnKF for four representations of spatial heterogeneity (scenario A). Results are shown for reference I. Fluxes from river to aquifer have a positive sign.

**Table 4.** Root mean square error of mean leakage coefficients  $\log_{10}(L^*)$  in [ $\log_{10}(\text{m s}^{-1})$ ] at the end of the simulation period. Results are shown for all scenarios and references.

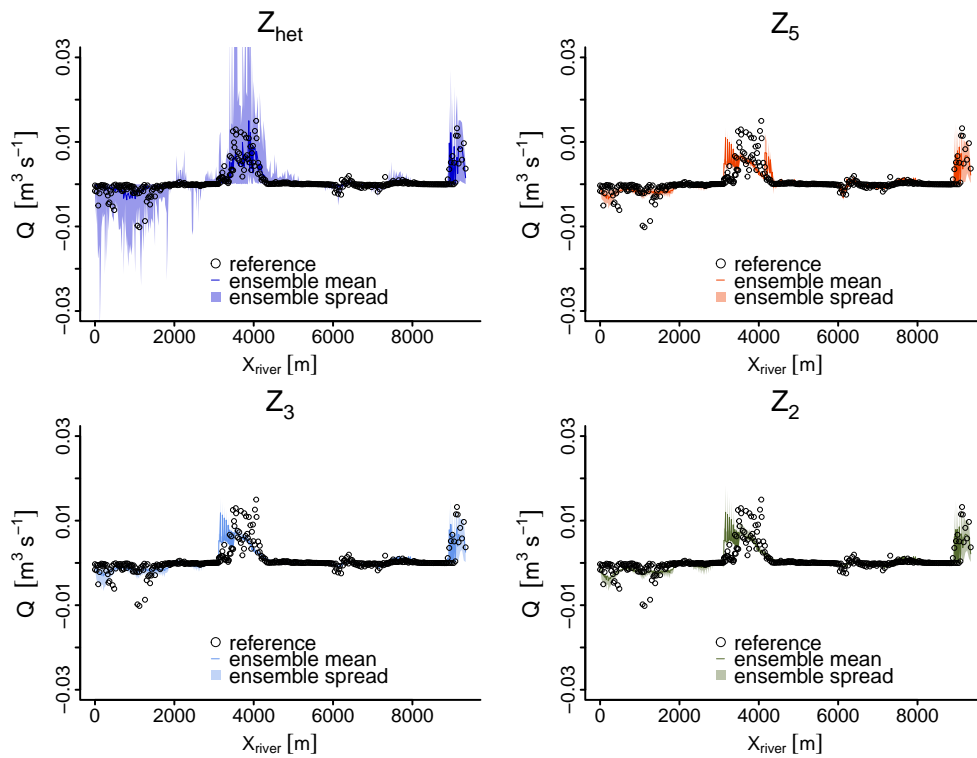
		I	II	III	IV	V	VI	VII	VIII	IX	X
scenario A	$Z_{\text{het}}$	0.964	0.826	0.707	0.859	0.814	0.988	0.750	1.000	0.770	0.689
	$Z_5$	1.710	2.177	1.531	1.070	0.882	1.337	0.732	1.117	0.973	0.852
	$Z_3$	1.093	1.307	0.903	0.809	1.052	1.460	1.079	1.111	1.398	1.133
	$Z_2$	1.087	1.474	1.175	0.834	1.339	1.468	1.296	1.302	1.391	1.138
scenario A 10 observations	$Z_{\text{het}}$	1.121	0.951	0.941	1.138	0.996	1.288	0.916	1.192	1.119	0.803
	$Z_5$	1.136	1.379	0.948	0.786	0.875	1.291	0.754	1.120	1.029	0.902
	$Z_3$	1.114	1.261	0.940	0.801	1.032	1.396	1.052	1.104	1.364	1.127
	$Z_2$	1.099	1.458	1.181	0.829	1.312	1.432	1.264	1.293	1.366	1.138
scenario B	$Z_{\text{het}}$	0.363	0.377	0.338	0.295	0.375	0.336	0.265	0.410	0.452	0.353
	$Z_5$	0.469	0.586	0.319	0.573	0.488	0.412	0.347	0.470	0.630	0.452
	$Z_3$	0.451	0.573	0.404	0.412	0.519	0.579	0.444	0.573	0.733	0.414
	$Z_2$	0.463	0.571	0.338	0.439	0.521	0.483	0.468	0.875	0.629	0.441
scenario C	$Z_{\text{het}}$	1.083	1.065	0.963	1.341	1.243	1.074	1.008	0.991	1.188	1.057
	$Z_5$	1.045	0.937	1.050	1.277	1.148	0.975	1.094	0.922	1.144	0.805
	$Z_3$	1.828	1.467	1.692	2.131	2.240	1.489	1.283	1.466	1.702	1.806
	$Z_2$	1.858	1.507	1.742	2.196	2.600	1.542	1.303	1.493	1.702	1.903
scenario D	$Z_{\text{het}}$	1.046	1.033	1.011	1.111	1.048	1.250	0.909	1.138	1.047	0.948
	$Z_5$	1.407	1.396	0.913	0.811	0.947	1.335	0.754	1.149	1.032	0.931
	$Z_3$	1.092	1.251	0.907	0.787	1.112	1.463	1.097	1.116	1.399	1.128
	$Z_2$	1.083	1.467	1.192	0.836	1.825	1.530	1.564	1.470	1.455	1.136

**Table 5.** Root mean square error of mean hydraulic heads (for the whole model domain) in [m] calculated for the second half of the simulation period (day 300 to day 609). Results are shown for all scenarios and references.

		I	II	III	IV	V	VI	VII	VIII	IX	X
scenario A	$Z_{\text{het}}$	0.070	0.055	0.110	0.104	0.034	0.039	0.030	0.090	0.092	0.066
	$Z_5$	0.102	0.184	0.212	0.101	0.078	0.088	0.050	0.113	0.207	0.073
	$Z_3$	0.151	0.182	0.197	0.119	0.092	0.092	0.070	0.145	0.249	0.129
	$Z_2$	0.165	0.203	0.333	0.125	0.222	0.092	0.178	0.200	0.249	0.122
scenario A 10 observations	$Z_{\text{het}}$	0.240	0.101	0.167	0.129	0.082	0.153	0.103	0.118	0.171	0.146
	$Z_5$	0.107	0.180	0.158	0.096	0.081	0.101	0.057	0.104	0.193	0.071
	$Z_3$	0.152	0.167	0.173	0.111	0.095	0.106	0.078	0.104	0.235	0.124
	$Z_2$	0.162	0.191	0.314	0.115	0.225	0.099	0.180	0.167	0.233	0.124
scenario B	$Z_{\text{het}}$	0.100	0.100	0.103	0.080	0.076	0.106	0.095	0.113	0.362	0.071
	$Z_5$	0.149	0.163	0.100	0.110	0.169	0.222	0.151	0.135	0.299	0.154
	$Z_3$	0.178	0.133	0.116	0.115	0.137	0.241	0.115	0.151	1.273	0.168
	$Z_2$	0.175	0.137	0.265	0.118	0.182	0.264	0.186	0.203	0.513	0.142
scenario C	$Z_{\text{het}}$	0.094	0.074	0.326	0.093	0.088	0.088	0.075	0.076	0.091	0.115
	$Z_5$	0.213	0.139	0.265	0.166	0.254	0.214	0.289	0.151	0.219	0.250
	$Z_3$	0.459	0.313	0.550	0.228	0.440	0.406	0.312	0.251	0.346	0.614
	$Z_2$	0.533	0.339	0.552	0.257	0.465	0.412	0.317	0.248	0.435	0.584
scenario D	$Z_{\text{het}}$	0.189	0.138	0.356	0.157	0.124	0.166	0.144	0.168	0.234	0.184
	$Z_5$	0.229	0.344	0.448	0.218	0.153	0.174	0.106	0.210	0.340	0.244
	$Z_3$	0.319	0.345	0.457	0.233	0.166	0.185	0.140	0.191	0.423	0.262
	$Z_2$	0.323	0.299	0.477	0.246	0.278	0.195	0.255	0.286	0.437	0.282



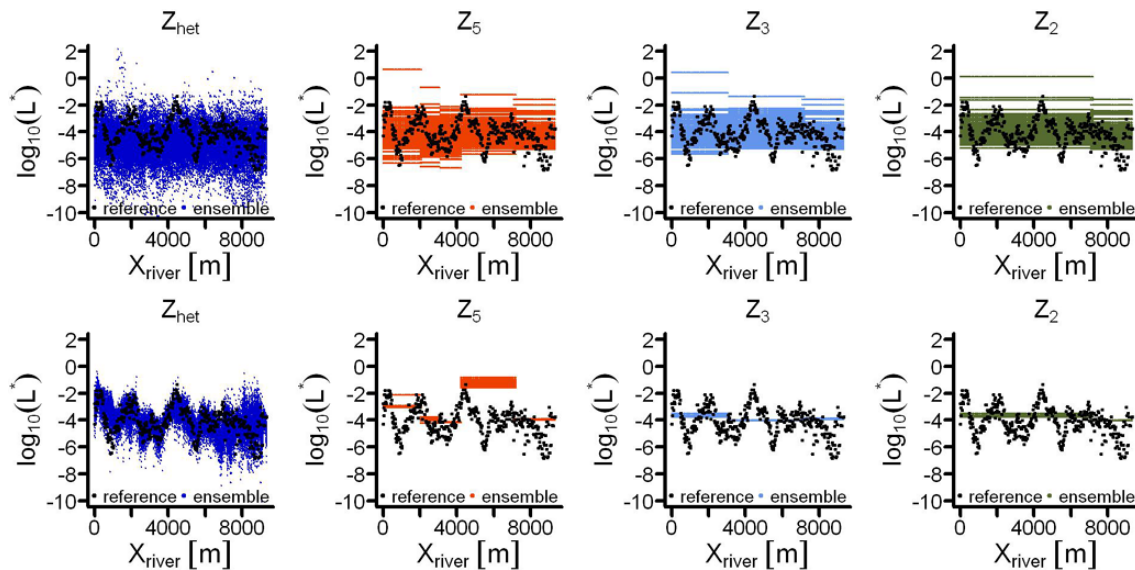
**Fig. 8.** Total amount of water exchanged between river and aquifer over the whole simulation period (609 days) for scenario A. Red lines mark the water exchange for the different reference runs. On the right-hand side the description of the boxplots is illustrated.



**Fig. 9.** Fluxes between river and aquifer along the river reach for conditional simulations with EnKF at simulation day 300 (scenario A, reference IV). Fluxes from river to aquifer have a positive sign. Distance  $X_{river}$  is calculated starting from the two most western river nodes.

Figure 9 gives an example of the spatial distribution of exchange fluxes for time step 300 and reference IV. For  $Z_{het}$  the spatial distribution of exchange fluxes of the reference run is principally captured by the ensemble, and the exchange fluxes of the reference run are within the uncertainty bounds of the ensemble. For  $Z_5$ ,  $Z_3$  and  $Z_2$  some of the principal features of the reference run, i.e. the river parts with the highest

positive exchange fluxes, are also present although the reference fluxes are no longer within the uncertainty bounds of the zoned ensembles. In other parts of the river the exchange fluxes of the reference run are not present in the ensemble calculations of  $Z_5$ ,  $Z_3$  and  $Z_2$ , e.g. the negative fluxes in the western part of the model.



**Fig. 10.** Ensembles of leakage coefficients (coloured) and reference field (black) along the river reach at day 1 (upper row) and at day 600 (lower row) for scenario A. Distance  $X_{\text{river}}$  is calculated starting from the two most western river nodes.  $L^*$  has units  $\text{m s}^{-1}$ .

Figure 10 compares the initial ensembles of leakage coefficients with the updated ones at day 600 for reference I. In general, the updated ensembles at time step 600 have a smaller variance than the initial ensembles. For  $Z_{\text{het}}$  the updated  $\log_{10}(L^*)$  ensemble mostly covers the spatial pattern of the reference field. For some parts of the river  $Z_{\text{het}}$  still has a relatively high variance, while in other parts variance is low and the spatial pattern of the ensemble is close to the pattern of the reference field. For  $Z_5$  the mean values for the different zones remain more or less constant during the updates, while the ensemble variance for the different zones strongly decreases for four of the five zones. For  $Z_3$  and  $Z_2$  also a very strong decrease in ensemble variance is visible. The overview of  $\text{RMSE}_{\log_{10}(L^*)}$  in Table 4 additionally emphasizes that  $Z_{\text{het}}$  usually gives the closest approximation towards the reference fields.

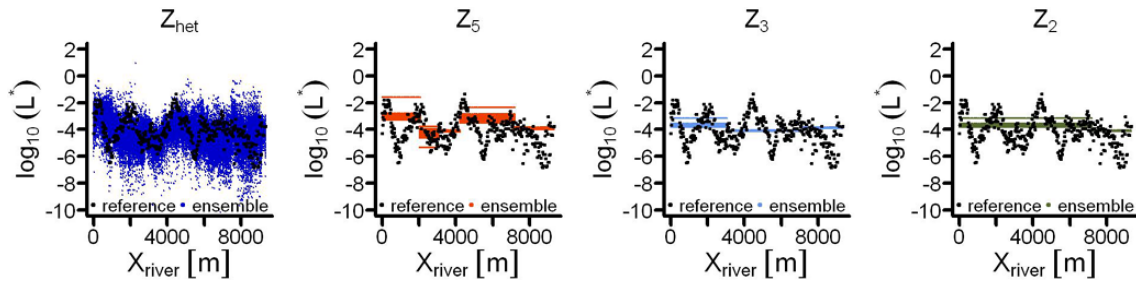
## 5.2 Strongly heterogeneous case (scenario A) with lower observation density

In order to investigate how a lower density of observation points affects the results for the different zonation approaches, the assimilation experiments of scenario A were repeated with hydraulic head time series measured at only 10 points instead of 100. The overall error in terms of  $\text{RMSE}_h$  for  $Z_5$ ,  $Z_3$  and  $Z_2$  was comparable to those in the assimilation experiments with 100 observation points (see Table 5). For  $Z_{\text{het}}$  the overall deviations to the reference were higher when only 10 observation points were used. Especially in the first part of the simulation period the lower observation density led to a slower decrease of  $\text{RMSE}_h$  for  $Z_{\text{het}}$ . At the end of the simulation period,  $\text{RMSE}_h$  for  $Z_{\text{het}}$  was similar to those of the zoned ensembles but not significantly better.

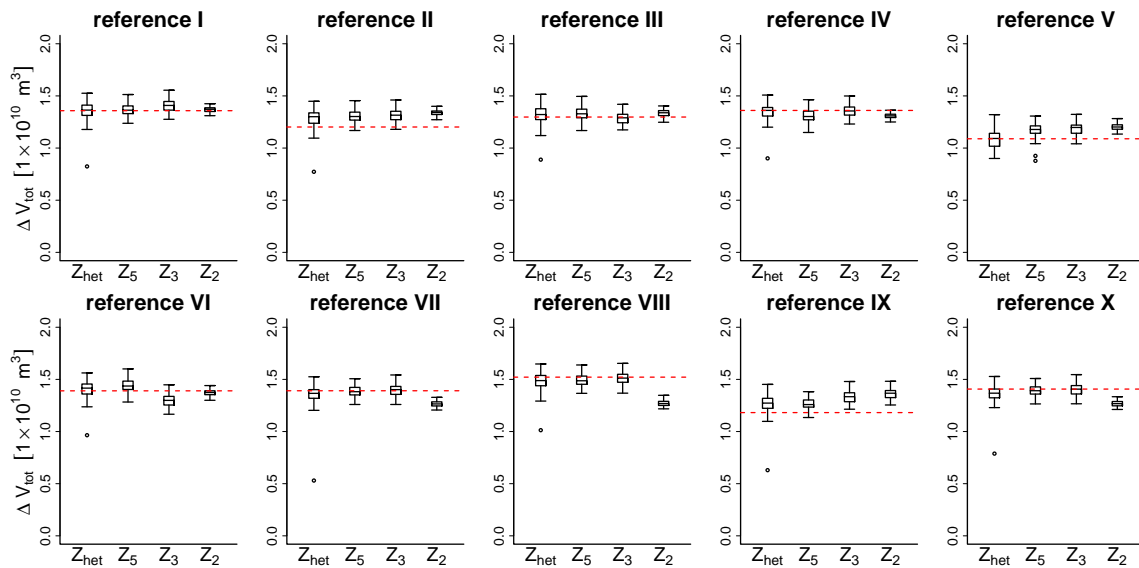
The predicted mean fluxes in both directions (i.e. fluxes to river or to aquifer) for  $Z_{\text{het}}$  were underestimated in most cases, whereas there were no major differences for the other ensembles. Flux predictions for  $Z_{\text{het}}$  were also associated with a higher degree of uncertainty compared to the assimilation of 100 observations. Despite an underestimation of both fluxes from river to aquifer and fluxes from aquifer to river for  $Z_{\text{het}}$ , the predicted net fluxes between river and aquifer were comparable to the simulations with 100 observation points; i.e.  $Z_{\text{het}}$  gave good results for all references. Net fluxes for the other ensembles were also similar to the assimilation of 100 measurements, and especially for  $Z_3$  and  $Z_2$  higher deviations occurred for some references. The update of leakage coefficients for  $Z_{\text{het}}$  was not as good as for the assimilation of 100 observations. The main structural features of the reference fields were captured during the assimilation, but the ensemble variance at the end of the simulation period was significantly higher when only 10 observations were assimilated (see Fig. 11). From Fig. 11 it becomes obvious that there are more extreme values than for the assimilation of 100 observations. This can be seen as a reason for the higher variability of fluxes for  $Z_{\text{het}}$ . The variability of  $\log_{10}(L^*)$  for the other ensembles increased only marginally and also the mean values for  $Z_3$  and  $Z_2$  were similar to the assimilation of 100 measurements.

## 5.3 Mildly heterogeneous case (scenario B)

For scenario B the variability of  $\log_{10}(L^*)$  fields for the references and the initial ensembles was reduced. Results for this case show that  $\text{RMSE}_h$  for  $Z_5$ ,  $Z_3$  and  $Z_2$  correspond more closely to  $\text{RMSE}_h$  of  $Z_{\text{het}}$ . Nevertheless,  $Z_{\text{het}}$  still shows the best performance in terms of  $\text{RMSE}_h$  for all ten references.



**Fig. 11.** Ensembles of leakage coefficients (coloured) and reference field (black) along the river reach at day 600 for scenario A when only 10 observation points are available. Distance  $X_{river}$  is calculated starting from the two most western river nodes.  $L^*$  has units  $m\ s^{-1}$ .



**Fig. 12.** Total amount of water exchanged between river and aquifer over the whole simulation period (609 days) for conditional simulations with EnKF (scenario B). Red lines mark the water exchange for the different reference runs.

The temporal evolution of leakage fluxes for the ten references is captured well by all four ensembles. Compared to scenario A the systematic differences that occurred between reference fluxes and simulated fluxes were reduced (especially for  $Z_3$  and  $Z_2$ ). This is also reflected in the cumulative net exchange over the simulation period (Fig. 12).

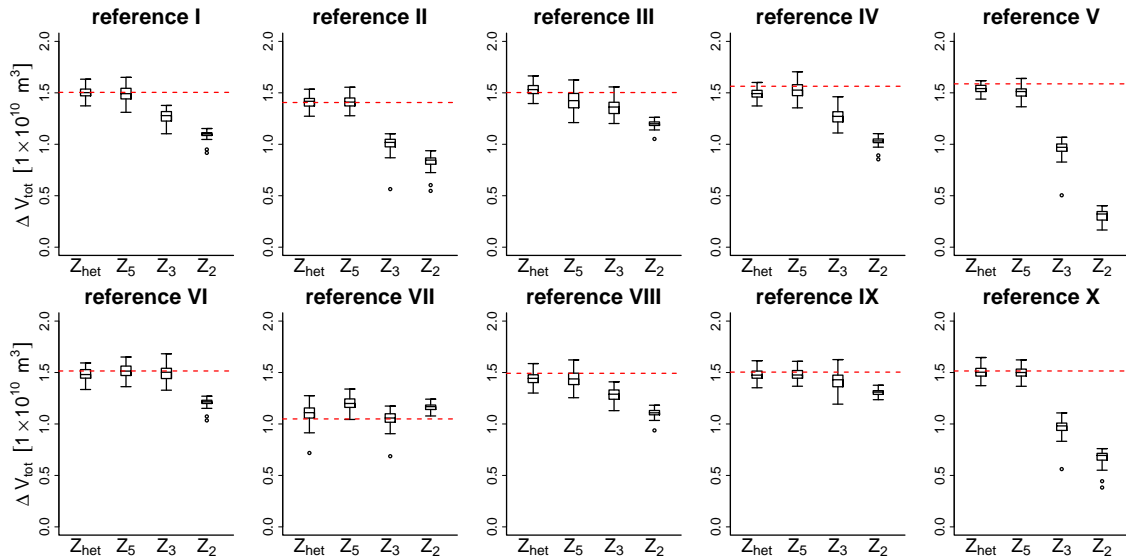
It can be seen that ensembles are generally closer to the reference values, and this is especially pronounced for  $Z_3$  and  $Z_2$ . But still there are some references where the uncertainty bounds of  $Z_2$  do not cover the reference flux. Although the prediction of cumulative net fluxes for scenario B is better for ensembles with a lower number of leakage zones, the spatial representation of fluxes is still worse for  $Z_2$ ,  $Z_3$  and  $Z_5$  than for  $Z_{het}$  where the spatial distribution of leakage fluxes closely corresponds to the reference fluxes.

### 5.4 Strongly heterogeneous case with a predefined zonation (scenario C)

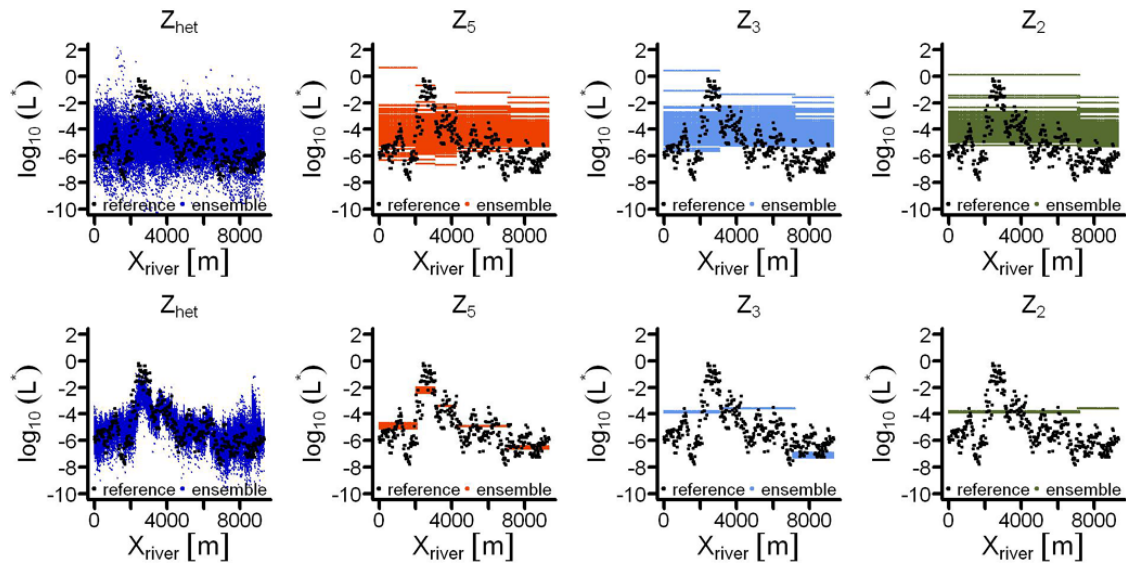
In scenario C the references include a predefined zonation with a relatively high contrast of  $\log_{10}(L^*)$  between the individual zones. A second important feature of this scenario is that for  $Z_5$  the location of leakage zones is similar to those of the references. Thus the initial ensemble of  $Z_5$  includes prior information on the spatial distribution of  $\log_{10}(L^*)$ .

$RMSE_h$  of  $Z_{het}$  and  $Z_5$  are very similar for this scenario with slightly lower errors for  $Z_{het}$ . In contrast,  $Z_3$  and  $Z_2$  perform worse in terms of  $RMSE_h$  compared to scenario A. For the net fluxes between river and aquifer a similar relation is found. Again  $Z_{het}$  and  $Z_5$  show relatively similar values which are very close to the reference values, whereas  $Z_3$  and  $Z_2$  consistently underestimate the net exchange, leading to a higher error compared to scenario A (Fig. 13).

The general worse performance for  $Z_3$  and  $Z_2$  in terms of head and flux predictions is a consequence of the spatial averaging of  $\log_{10}(L^*)$ . For the ensembles used for scenario



**Fig. 13.** Total amount of water exchanged between river and aquifer over the whole simulation period (609 days) for conditional simulations with EnKF (scenario C). Red lines mark the water exchange for the different reference runs.

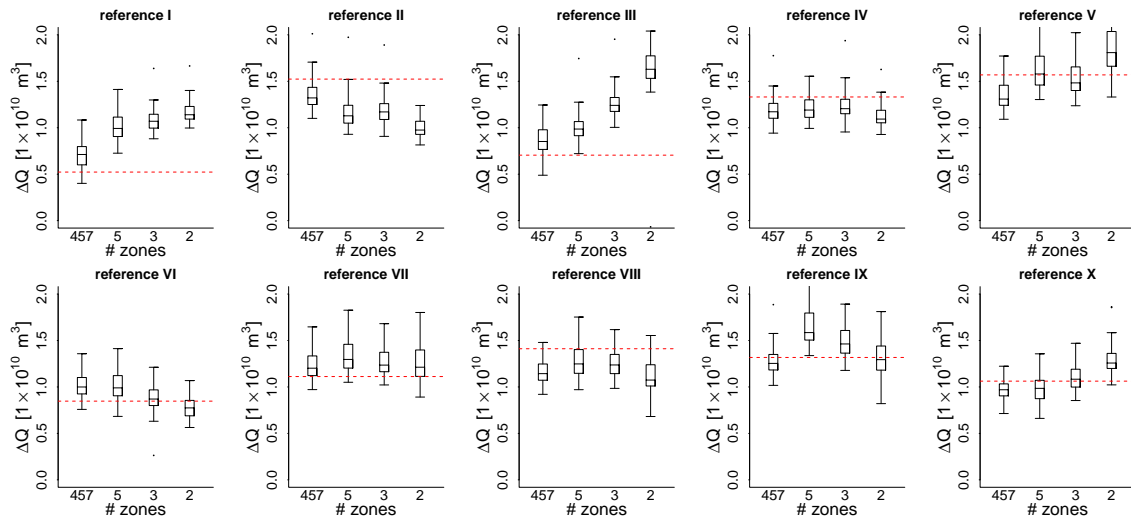


**Fig. 14.** Ensembles of leakage coefficients (coloured) and reference field (black) along the river reach at day 1 (upper row) and at day 600 (lower row) for scenario C. Distance  $X_{\text{river}}$  is calculated starting from the two most western river nodes.  $L^*$  has units  $\text{m s}^{-1}$ .

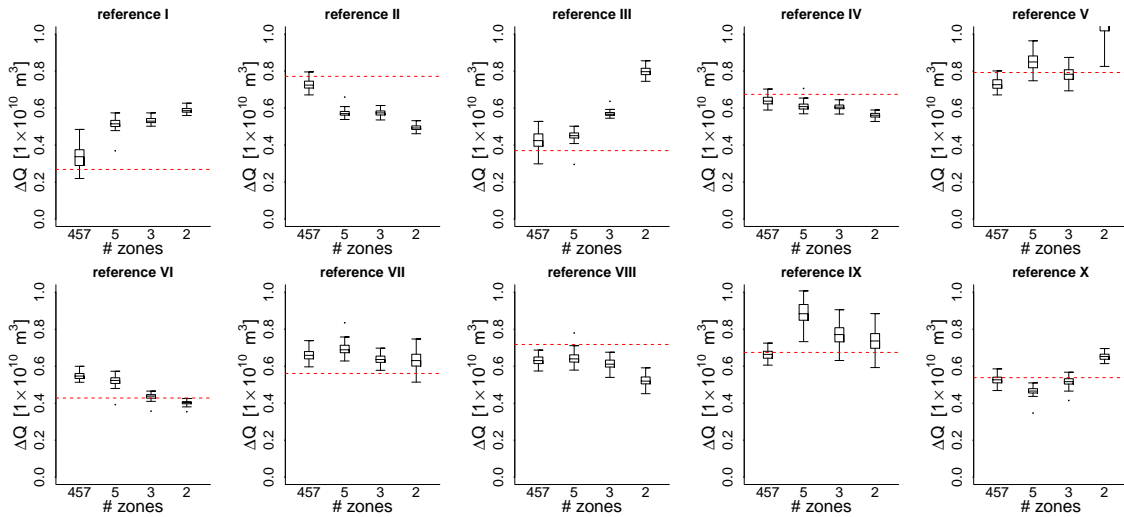
C the spatial contrasts for  $\log_{10}(L^*)$  are higher due to the predefined zonation in the references and initial ensembles. Therefore, the leakage zones of  $Z_3$  and  $Z_2$  cover parts of the reference fields that have very different  $\log_{10}(L^*)$  values. This can be seen in Fig. 14 where the initial  $\log_{10}(L^*)$  ensemble is compared with the updated one at the end of the simulation period. Because  $Z_3$  and  $Z_2$  are not flexible enough to account for the variability of the references due to their limited number of leakage zones the simulated piezometric heads and leakage fluxes deviate stronger from the reference values than  $Z_{\text{het}}$  and  $Z_5$ .

### 5.5 Strongly heterogeneous case with uncertain $K$ values (scenario D)

In order to verify the influence of uncertain hydraulic aquifer properties on the estimation of  $\log_{10}(L^*)$  fields, scenario A was repeated with an uncertain initial ensemble of  $K$  fields (scenario D). In this scenario,  $\log_{10}(K)$  values were jointly updated together with  $h$  and  $\log_{10}(L^*)$ . From Table 5 it is obvious that  $\text{RMSE}_h$  is about two to three times higher for all four  $\log_{10}(L^*)$  ensembles when the additional uncertainty in  $K$  is introduced. In 9 out of 10 cases  $Z_{\text{het}}$  shows the best



**Fig. 15.** Total amount of water exchanged between river and aquifer over the whole simulation period (609 days) for conditional simulations with EnKF (scenario D). Red lines mark the water exchange for the different reference runs.



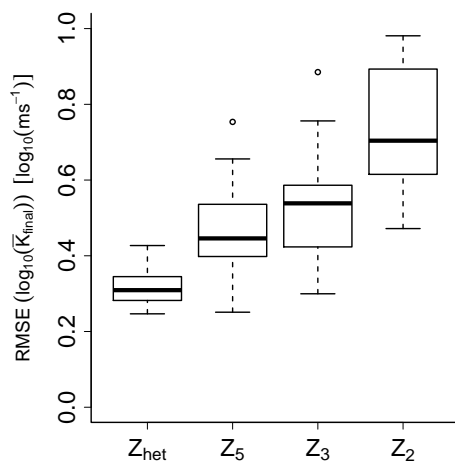
**Fig. 16.** Total amount of water exchanged between river and aquifer over the second half of the simulation period (day 300 to day 609) for conditional simulations with EnKF (scenario D). Red lines mark the water exchange for the different reference runs.

results among the four  $\log_{10}(L^*)$  ensembles, which is similar to the findings for scenario A (with deterministic  $K$  values).

Data for the total amount of exchanged water  $\Delta V_{\text{tot}}$  over the whole simulation period (Fig. 15) generally show a higher variability than for scenario A (Fig. 8). For scenario D  $Z_{\text{het}}$  is significantly closer to the reference values only for references I–III. When  $\Delta V_{\text{tot}}$  is solely calculated for the second half of the simulation period (Fig. 16) the uncertainty in total water exchange is greatly reduced and there are more cases in which the heterogeneous ensemble performs better than the zonated ensembles.

The data for  $\text{RMSE}_{\log_{10}(L^*)}$  from Table 4 show that the average deviation from the reference fields stays about the same for  $Z_3$  and  $Z_2$  but gets slightly worse for  $Z_{\text{het}}$  and

$Z_5$  compared to scenario A. However, for  $Z_{\text{het}}$  the structure of the reference fields is still captured well by the final  $\log_{10}(L^*)$  ensemble with a slightly lower variability than for scenario A. Figure 17 gives some summary information on the update of  $\log_{10}(K)$  fields for the different ensembles. The diagram shows the statistics of  $\text{RMSE}_{\log_{10}(K)}$  for the final  $K$  fields for all ten references. It can be seen that a higher degree of aggregation of leakage parameters also leads to a higher error in the estimated  $K$  fields. The reason is that for the zonated ensembles (especially  $Z_3$  and  $Z_2$ ) the lower flexibility in adapting to the true fields of  $L$  is compensated by a stronger adjustment of  $K$  values (in order to correct the mismatch between observed and simulated heads).



**Fig. 17.** Statistics of root mean square error of mean hydraulic conductivity ( $\text{RMSE}(\log_{10}(\bar{K}_{\text{final}}))$ ) after the last update step calculated from all references of scenario D.

## 6 Discussion

Simulations with EnKF generally lead to an improvement for all four ensembles in terms of  $\text{RMSE}_h$ . It was found that a stochastic field approach (i.e. each discretization point of the model grid has a different leakage value, which results in 457 values in this study) gave the best results. Data assimilation with EnKF made it possible to correct the cumulative fluxes between river and aquifer almost completely. For  $Z_{\text{het}}$  also the spatial distribution of  $\log_{10}(L^*)$  gets quite close to the reference fields during data assimilation, and this is also reflected in the spatial distribution of exchange fluxes which closely coincides with those of the reference runs.

For the three zoned ensembles the overall performance was usually slightly worse than for the ensemble with full heterogeneity. Especially the net fluxes showed significant deviations from the true values for  $Z_3$  and  $Z_2$  for several references. This is reflected in the spatial distribution of fluxes along the river where regions with higher fluxes in the reference runs were not adequately represented by the ensembles  $Z_3$  and  $Z_2$ . EnKF was only partly able to correct for the systematic errors that arose from the wrong spatial distribution of exchange fluxes, and as a result of this the prediction of net fluxes was not as good as for  $Z_{\text{het}}$ . For  $Z_5$  the predicted net fluxes were often similar to the ones of  $Z_{\text{het}}$ , but the spatial distribution of fluxes was not as good as for  $Z_{\text{het}}$ . This was also the case when  $Z_5$  closely matched the spatial distribution of  $\log_{10}(L^*)$  of the reference runs (scenario C). Even with this prior information only the net fluxes were estimated correctly but not their spatial distribution. This also applies for references with a lower degree of heterogeneity (scenario B). Even in this case the predicted spatial distribution of leakage fluxes was better with  $Z_{\text{het}}$  than with the different zonation approaches. A precise estimation of high local leakage fluxes is highly relevant for transport calculations in order to

determine source regions of contaminants or regions of high biogeochemical turnover. For such applications, the usage of effective parameters will therefore only lead to averaged concentration levels that are derived from the net exchange between river and aquifer.

For a lower observation density the results for  $Z_5$ ,  $Z_3$  and  $Z_2$  were not very different from the ones for 100 observation points. For  $Z_{\text{het}}$  the uncertainty in the  $\log_{10}(L^*)$  ensemble was larger than for the case with 100 measurement data, and as a consequence also the uncertainty regarding the exchange fluxes was higher. Generally, a lower information content mainly affected the parameter identification for the heterogeneous ensemble. Nevertheless, the errors in head and flux predictions for  $Z_{\text{het}}$  were still comparable to the ones for the zoned ensembles especially at the end of the simulation period. Of course, comparisons with respect to observation density always depend on site-specific conditions like the spatial location of observations and the degree of spatial heterogeneity of the estimated parameter field. In our case the correlation length of the reference fields is relatively high (1000 to 2000 m), and therefore conditioning to a low amount of observation data still gives satisfactory results. However, for cases with a shorter integral scale a higher amount of observations may be necessary.

Overall, the results from these synthetic studies suggest that the high-resolution characterization of river bed properties is feasible because, even with a limited number of measurements, the high-resolution reconstruction resulted in better results than an approach where the spatial variability of the river bed was represented with a few effective parameters only. We expect that in case only very few measurements are available, high-resolution and zonation approaches might give predictions of similar quality, and that in the case of more measurements the high-resolution approach will increasingly outperform the approaches where only a few effective parameters are estimated. It can be important to condition multiple equally likely stochastic high-resolution realizations of river bed properties, because the quality of the estimated net exchange fluxes between aquifer and stream are better with this approach than with a zonation approach. Replacing the heterogeneous streambed with a few effective parameters results in biased predictions of exchange fluxes. Over long streams such a systematic bias might result in an important under- or overestimation of the groundwater infiltration in the river, as well as an under- or overestimation of the loss of river water to the aquifer under flood conditions. Therefore, we recommend the calibration of equally likely stochastic realizations of river bed properties using EnKF together with an augmented state vector approach, for a better characterization of river–aquifer exchange fluxes. This approach is especially needed in the case of very heterogeneous streambeds and in case enough conditioning measurements (hydraulic head data) are available.

Of course, our findings were derived on the basis of a simplified model which only accounts for uncertainties



regarding leakage parameters (for most cases) and where the initial ensemble mean of leakage parameters closely corresponded to the reference values. In real-world applications uncertainties may also arise from the poorly known distribution of aquifer properties, a biased initial parameter ensemble, model structural errors and uncertainties in the determination of forcing terms for the model. Thus, the calibration of  $\log_{10}(L^*)$  distributions with EnKF in real-world cases will probably not be as accurate as in our synthetic simulations due to the higher overall uncertainties. Nevertheless, the incorporation of uncertainty regarding  $K$  in one of the scenarios already showed that the principal differences between simulations with spatially highly resolved ensembles and simulations with zoned  $\log_{10}(L^*)$  ensembles remain. This is related to the higher flexibility of the stochastic field approach, and similar results are also expected if other sources of uncertainty are included.

It is also noteworthy that the differences between the heterogeneous and the zoned ensembles in this study are derived from a pure assimilation experiment where the state variable ( $h$ ) was continuously updated. If the updated  $\log_{10}(L^*)$  ensembles were used in a predictive forward simulation (i.e. with no update of state variables), the differences with respect to exchange fluxes and  $\text{RMSE}_h$  would very likely increase due to the lack of a continuous correction of aquifer states.

## 7 Conclusions

In this study we performed data assimilation experiments with a synthetic river–groundwater interaction problem where piezometric heads and heterogeneous river bed conductivities were jointly updated with EnKF. In this context, different parametrizations of river bed heterogeneity (full heterogeneity versus different amounts of effective parameters) were tested for their effect on the prediction of groundwater levels and river–aquifer exchange fluxes. A high-resolution representation of the river bed hydraulic conductivities (compared to a coarse representation with constant values in a few zones) results in a better characterization of stream–aquifer exchange fluxes, both in terms of the net exchange between the two compartments and the spatio-temporal distribution of this exchange. When the river was divided into a relatively limited number of leakage zones (2, 3 or 5) the net exchange between river and aquifer was not predicted accurately for all references and in general the errors increased with a decreasing representation of heterogeneity. In the case of a reduced amount of head observations (10 instead of 100), the high-resolution calibration of river bed conductivities still outperformed the scenarios with a coarse representation of river bed conductivities. However, differences between the scenarios were smaller. Similar results were found when also uncertain aquifer hydraulic parameters were regarded in the assimilation process. In this

case, the overall mismatch between simulations and reference values generally increased, but the ensemble with a higher spatial resolution consistently performed better than its zoned counterparts.

In summary, it is concluded that a zonation of river bed conductivities should be avoided because small regions with high exchange fluxes might be averaged out by zonation which affects the local water balance. We recommend therefore that for highly heterogeneous river beds a geostatistical simulation approach is used where river bed properties change from grid cell to grid cell. Our simulations showed that parameter updates with EnKF are able to adapt an ensemble of such high-resolution fields of river bed conductivity towards the true reference field. Furthermore, the CPU demand for parameter adaptation with EnKF is not dependent on the number of defined leakage zones as opposed to other calibration techniques. This approach will also avoid a bias in the estimation of the regional water balance.

*Acknowledgements.* We gratefully acknowledge Water Works Zurich for providing data for the study site and Jülich Supercomputing Center for granting computation time on JUROPA. We would also like to thank Matteo Camporese and two anonymous reviewers for their valuable comments and suggestions.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

Edited by: A. Guadagnini

## References

- Alcolea, A., Carrera, J., and Medina, A.: Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Adv. Water Resour.*, 29, 1678–1689, doi:10.1016/j.advwatres.2005.12.009, 2006.
- Boano, F., Camporeale, C., Revelli, R., and Ridolfi, L.: Sinuosity-driven hyporheic exchange in meandering rivers, *Geophys. Res. Lett.*, 33, L18406, doi:10.1029/2006GL027630, 2006.
- Bouwer, H. and Maddock, T.: Making sense of the interactions between groundwater and streamflow: Lessons for water masters and adjudicators, *Rivers*, 6, 19–31, 1997.
- Bruen, M. P. and Osman, Y. Z.: Sensitivity of stream-aquifer seepage to spatial variability of the saturated hydraulic conductivity of the aquifer, *J. Hydrol.*, 293, 289–302, doi:10.1016/j.jhydrol.2004.02.003, 2004.
- Brunke, M. and Gonser, T.: The ecological significance of exchange processes between rivers and groundwater, *Freshwater Biol.*, 37, 1–33, doi:10.1046/j.1365-2427.1997.00143.x, 1997.
- Brunner, P., Cook, P. G., and Simmons, C. T.: Hydrogeologic controls on disconnection between surface water and groundwater, *Water Resour. Res.*, 45, W01422, doi:10.1029/2008WR006953, 2009.
- Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, *Mon.*

- Weather Rev., 126, 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2, 1998.
- Calver, A.: Riverbed Permeabilities: Information from Pooled Data, *Ground Water*, 39, 546–553, doi:10.1111/j.1745-6584.2001.tb02343.x, 2001.
- Camporese, M., Cassiani, G., Deiana, R., and Salandin, P.: Assessment of local hydraulic properties from electrical resistivity tomography monitoring of a three-dimensional synthetic tracer test experiment, *Water Resour. Res.*, 47, W12508, doi:10.1029/2011WR010528, 2011.
- Cardenas, M. B., Wilson, J. L., and Zlotnik, V. A.: Impact of heterogeneity, bed forms, and stream curvature on subchannel hyporheic exchange, *Water Resour. Res.*, 40, W08307, doi:10.1029/2004WR003008, 2004.
- Chen, Y. and Zhang, D.: Data assimilation for transient flow in geologic formations via ensemble Kalman filter, *Adv. Water Resour.*, 29, 1107–1122, doi:10.1016/j.advwatres.2005.09.007, 2006.
- Conant, B.: Delineating and quantifying ground water discharge zones using streambed temperatures, *Ground Water*, 42, 243–257, doi:10.1111/j.1745-6584.2004.tb02671.x, 2004.
- Delta h Ingenieurgesellschaft mbH: SPRING 3.3, Software, Witten, Germany, 2006.
- Evensen, G.: Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistics, *J. Geophys. Res.-Oceans*, 99, 10143–10162, doi:10.1029/94JC00572, 1994.
- Fleckenstein, J. H., Niswonger, R. G., and Fogg, G. E.: SWGW interactions, geologic heterogeneity, and low-flow management, *Ground Water*, 44, 837–852, doi:10.1111/j.1745-6584.2006.00190.x, 2006.
- Frei, S., Fleckenstein, J. H., Kollet, S. J., and Maxwell, R. M.: Patterns and dynamics of river-aquifer exchange with variably-saturated flow using a fully-coupled model, *J. Hydrol.*, 375, 383–393, doi:10.1016/j.jhydrol.2009.06.038, 2009.
- Genereux, D. P., Leahy, S., Mitasova, H., Kennedy, C. D., and Corbett, D. R.: Spatial and temporal variability of streambed hydraulic conductivity in West Bear Creek, North Carolina, USA, *J. Hydrol.*, 358, 332–353, doi:10.1016/j.jhydrol.2008.06.017, 2008.
- Gómez-Hernández, J. and Journel, A.: Joint sequential simulation of multi-Gaussian fields, in: *Geostatistics Tróia '92*, edited by: Soares, A., Vol. 1, 85–94, Kluwer Acad., New York, 1993.
- Hatch, C. E., Fisher, A. T., Ruehl, C. R., and Stemler, G.: Spatial and temporal variations in streambed hydraulic conductivity quantified with time-series thermal methods, *J. Hydrol.*, 389, 276–288, doi:10.1016/j.jhydrol.2010.05.046, 2010.
- Hendricks Franssen, H. J. and Kinzelbach, W.: 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, W09408, doi:10.1029/2007WR006505, 2008.
- Hendricks Franssen, H. J., Kaiser, H. P., Kuhlmann, U., Bauser, G., Stauffer, F., Mueller, R., and Kinzelbach, W.: Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating, *Water Resour. Res.*, 47, W02532, doi:10.1029/2010WR009480, 2011.
- Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an Ensemble Kalman Filter technique, *Mon. Weather Rev.*, 126, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, 1998.
- Huber, E., Hendricks Franssen, H. J., Kaiser, H., and Stauffer, F.: The role of prior model calibration on predictions with Ensemble Kalman Filter, *Ground Water*, 49, 845–858, doi:10.1111/j.1745-6584.2010.00784.x, 2011.
- Irvine, D. J., Brunner, P., Hendricks Franssen, H. J., and Simmons, C. T.: Heterogeneous or homogeneous? Implications of simplifying heterogeneous streambeds in models of losing streams, *J. Hydrol.*, 424–425, 16–23, doi:10.1016/j.jhydrol.2011.11.051, 2012.
- Jafarpour, B. and Tarrahi, M.: Assessing the performance of the ensemble Kalman filter for subsurface flow data integration under variogram uncertainty, *Water Resour. Res.*, 47, W05537, doi:10.1029/2010WR009090, 2011.
- Kalbus, E., Schmidt, C., Molson, J. W., Reinstorf, F., and Schirmer, M.: Influence of aquifer and streambed heterogeneity on the distribution of groundwater discharge, *Hydrol. Earth Syst. Sci.*, 13, 69–77, doi:10.5194/hess-13-69-2009, 2009.
- Kurtz, W., Hendricks Franssen, H.-J., and Vereecken, H.: Identification of time-variant river bed properties with the ensemble Kalman filter, *Water Resour. Res.*, 48, W10534, doi:10.1029/2011WR011743, 2012.
- Li, L., Zhou, H., Gómez-Hernández, J. J., and Hendricks Franssen, H. J.: Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter, *J. Hydrol.*, 428–429, 152–169, doi:10.1016/j.jhydrol.2012.01.037, 2012.
- Liu, G., Chen, Y., and Zhang, D.: Investigation of flow and transport processes at the MADE site using ensemble Kalman filter, *Adv. Water Resour.*, 31, 975–986, doi:10.1016/j.advwatres.2008.03.006, 2008.
- Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R.: Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, 28, 135–147, doi:10.1016/j.advwatres.2004.09.002, 2005.
- Nowak, W.: Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator, *Water Resour. Res.*, 45, W04431, doi:10.1029/2008WR007328, 2009.
- Renard, P., Le Loc'h, G., Ledoux, E., de Marsily, G., and Mackay, R.: A fast algorithm for the estimation of the equivalent hydraulic conductivity of heterogeneous media, *Water Resour. Res.*, 36, 3567–3580, doi:10.1029/2000WR900203, 2000.
- Rosenberry, D. O. and Pitlick, J.: Local-scale variability of seepage and hydraulic conductivity in a shallow gravel-bed river, *Hydrol. Process.*, 23, 3306–3318, doi:10.1002/hyp.7433, 2009.
- Schoeniger, A., Nowak, W., and Hendricks Franssen, H. J.: Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography, *Water Resour. Res.*, 48, W04502, doi:10.1029/2011WR010462, 2012.
- Schubert, J.: Hydraulic aspects of riverbank filtration-field studies, *J. Hydrol.*, 266, 145–161, doi:10.1016/S0022-1694(02)00159-2, 2002.
- Sophocleous, M.: Interactions between groundwater and surface water: the state of the science, *Hydrogeol. J.*, 10, 52–67, doi:10.1007/s10040-002-0204-x, 2002.
- Springer, A. E., Petrouson, W. D., and Semmens, B. A.: Spatial and temporal variability of hydraulic conductivity in active reattachment bars of the Colorado river, Grand Canyon, *Ground Water*, 37, 338–344, doi:10.1111/j.1745-6584.1999.tb01109.x, 1999.

- Sun, A. Y., Morris, A. P., and Mohanty, S.: Comparison of deterministic ensemble Kalman filters for assimilating hydrogeological data, *Adv. Water Resour.*, 32, 280–292, doi:10.1016/j.advwatres.2008.11.006, 2009.
- Winter, T. C.: Relation of streams, lakes, and wetlands to groundwater flow systems, *Hydrogeol. J.*, 7, 28–45, doi:10.1007/s100400050178, 1999.
- Woessner, W. W.: Stream and fluvial plain ground water interactions: Rescaling hydrogeologic thought, *Ground Water*, 38, 423–429, doi:10.1111/j.1745-6584.2000.tb00228.x, 2000.
- Zhang, Y., Hubbard, S., and Finsterle, S.: Factors Governing Sustainable Groundwater Pumping near a River, *Ground Water*, 49, 432–444, doi:10.1111/j.1745-6584.2010.00743.x, 2011.
- Zhou, H., Gómez-Hernández, J. J., Hendricks Franssen, H. J., and Li, L.: An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering, *Adv. Water Resour.*, 34, 844–864, doi:10.1016/j.advwatres.2011.04.014, 2011.