Identification of time-variant river bed properties with the ensemble Kalman filter

Wolfgang Kurtz,[1] Harrie-Jan Hendricks Franssen,[1] and Harry Vereecken[1]

Received 13 December 2011; revised 30 August 2012; accepted 5 September 2012; published 19 October 2012.

[1] An adequate characterization of river bed hydraulic conductivities (L) is crucial for a proper assessment of river-aquifer interactions. However, river bed characteristics may change over time due to dynamic morphological processes like scouring or sedimentation what can lead to erroneous model predictions when static leakage parameters are assumed. Sequential data assimilation with the ensemble Kalman filter (EnKF) allows for an update of model parameters in real-time and may thus be capable of assessing the transient behavior of L. Synthetic experiments with a three-dimensional finite element model of the Limmat aquifer in Zurich were used to assess the performance of data assimilation in capturing time-variant river bed properties. Reference runs were generated where L followed different temporal and/or spatial patterns which should mimic real-world sediment dynamics. Hydraulic head (h) data from these reference runs were then used as input data for EnKF which jointly updated h and L. Results showed that EnKF is able to capture the different spatio-temporal patterns of L in the reference runs well. However, the adaptation time was relatively long which was attributed to the fast decrease of ensemble variance. To improve the performance of EnKF also an adaptive filtering approach with covariance inflation was applied that allowed a faster and more accurate adaptation of model parameters. A sensitivity analysis indicated that even for a low amount of observations a reasonable adaptation of L towards the reference values can be achieved and that EnKF is also able to correct for a biased initial ensemble of L.

Citation: Kurtz, W., H.-J. Hendricks Franssen, and H. Vereecken (2012), Identification of time-variant river bed properties with the ensemble Kalman filter, *Water Resour. Res.*, *48*, W10534, doi:10.1029/2011WR011743.

1. Introduction

[2] Exchange fluxes between surface water and groundwater can have a profound influence on the chemical environment within the hyporheic zone, the riparian ecology, and the local water balance around streams [*Woessner*, 2000; *Sophocleous*, 2002; *Brunke and Gonser*, 1997]. For water management activities close to rivers, like bank filtration, the amount of exchanged water between river and aquifer influences the sustainability of groundwater use and also the quality of pumped groundwater. In order to predict the hydrological situation around rivers it is essential to obtain reliable estimates about the parameters that govern these exchange fluxes between river and groundwater, namely the river bed conductivities.

[3] Information on exchange coefficients can be inferred experimentally from different methods ranging from small scale measurements like permeameter tests, medium-scale information like temperature data, as well as water balance methods for larger scales [*Kalbus et al.*, 2006, and references therein]. On the regional scale, especially when management of groundwater is present, the most common approach is to solve the groundwater flow equation with a numerical model and to calibrate river bed conductivities with observed head (and concentration) data. The calibrated river bed conductivities can then be used for predicting groundwater levels in the postcalibration period.

[4] Different studies have shown that the fluxes between river and groundwater are strongly variable in space and time [e.g., *Conant*, 2004; *Krause et al.*, 2007; *Käser et al.*, 2009; *Rosenberry and Pitlick*, 2009a]. The spatial variability of exchange fluxes is related to the heterogeneity of the river bed and the heterogeneity of the adjacent aquifer [e.g., *Fleckenstein et al.*, 2006; *Kalbus et al.*, 2009; *Frei et al.*, 2009]. Besides their spatial variability river bed characteristics may also change over time. Reasons for changing river bed properties may be flooding events that erode the river bed due to larger shear stress or an enhanced sedimentation during low flow conditions that can lead to a colmatation of the river bed [*Schälchli*, 1992].

[5] In a flume experiment *Rehg et al.* [2005] investigated the effect of clay deposition on exchange fluxes between

[1]Institute for Bio- and Geosciences, IBG-3: Agrosphere, Forschungszentrum Jülich GmbH, Jülich, Germany.

Corresponding author: W. Kurtz, Institute for Bio- and Geosciences, IBG-3: Agrosphere, Forschungszentrum Jülich GmbH, Jülich, Germany. (w.kurtz@fz-juelich.de)

0043-1397/12/2011WR011743

sediment and surface water. They observed the formation of a thin clogging layer that substantially decreased exchange fluxes when no movement of the bed sediment was present. However, when the bed sediment was slightly moved by the stream current and the particle size of the clay was small enough the formation of a clogging layer was not detected. *Rosenberry and Pitlick* [2009b] also used flume experiments to investigate the effect of sedimentation of fine particles on seepage fluxes between sediment and surface water. Their results showed that the vertical hydraulic conductivity of the sediment was decreased during downward flux and remained almost constant during upward flux.

[6] Changes in river bed conductivity have also been observed at the field scale with different measurement and modeling techniques. *Schubert* [2002] investigated the relation between river dynamics and bank filtration activities at the river Rhine and found that the permeability of the clogging layer on top of the river bed varied temporally which was attributed to changes in sediment load of the river, erosion processes in the river bed, and different hydraulic gradients between river and groundwater. *Blaschke et al.* [2003] measured leakage coefficients at an impounded river reach of the Danube and found a decrease of the determined river bed conductivities of about 2 log units within a time frame of 2 years which they attributed to clogging processes. They also found that flooding events led to temporary increases of river bed permeability at their site. *Doppler et al.* [2007] observed a significant change between model predictions and observations after a major flooding event which they attributed to the scouring of an impounded part of the investigated river. *Hatch et al.* [2010] used time series thermal methods to quantify the temporal evolution of river bed conductivities along a river reach. They observed changes in river bed conductivities of up to 1 log unit within a sedimentation period of 150 days and also increasing river bed conductivities due to high flow conditions. *Zhang et al.* [2011] used a 3-D model of a managed site for river bank filtration to investigate the behavior of river bed permeabilities in relation to management activities. They calibrated river bed permeabilities at several times during a 1 year period and observed changes of up to a factor of 3 which they related to high and low flow conditions which were in part induced by management activities. *Genereux et al.* [2008] repeatedly measured streambed hydraulic conductivities over a 1 year period with permeameter tests. At some of their measurement locations they observed nearly no change in river bed conductivities over time, whereas for other observation points river bed conductivities followed different temporal patterns during the measurement campaign which they related to erosion and deposition processes of the sediment. They also observed an increase of river bed permeabilities after the reconstruction of a dam which they attributed to a possible scouring of the river bed. *Mutiti and Levy* [2010] used head and temperature measurements to calibrate river bed conductivities during flooding events. They observed that hydraulic conductivities had to be increased around the discharge peak compared to preflood conditions in order to accomplish a good fit between measured and simulated groundwater heads and temperatures. This was also related to river bed erosion associated with increased shear stress during the flooding event.

[7] These observed changes in river bed properties can have a large impact on the prediction of groundwater or concentration levels near a river. For instance, when a groundwater model is calibrated for a certain time period with a specific, possibly spatially variable, river bed conductivity which is assumed to be constant for further time periods this model will not be able to respond to the changes in model parameters. This will result in systematically erroneous predictions of groundwater levels because the fluxes between river and groundwater are calculated with wrong parameter values for the river bed conductivities. In case of transport calculations, e.g., of contaminants, this wrong parameterization may be even more severe and may result in strongly biased predictions of the extent and breakthrough of contamination plumes.

[8] Until present, changes of river bed hydraulic conductivity are not handled in a systematic way in modeling studies and we are not aware of papers where a systematic procedure is proposed to calibrate time-dependent leakage parameters. In this paper, such an approach is proposed and its performance is tested in detail in a synthetic experiment that mimics a real-world case.

[9] Generally, one possible solution for calibrating time-dependent leakage parameters would be to recalibrate the model whenever deviations between measured and predicted groundwater levels exceed a predefined threshold value. However, these deviations between measurements and model predictions may also arise from measurement errors or predictions errors of the groundwater model (i.e., model structural errors or errors in the forcings terms). One could argue that the measurement errors are constant and known a priori, but the model errors are usually not. Prediction errors depend on many factors and could be temporally variable as well. For example, the uncertainty in timing and magnitude of a precipitation or flooding event could create deviations between measurements and model predictions which would then lead to a recalibration of model parameters although these parameters did not change in reality.

[10] Another approach is to use sequential data assimilation methods like the ensemble Kalman filter (EnKF) [*Evensen*, 1994; *Burgers et al.*, 1998] which has already been used in various studies in groundwater hydrology [e.g., *Chen and Zhang*, 2006; *Hendricks Franssen and Kinzelbach*, 2008; *Nowak*, 2009; *Sun et al.*, 2009; *Huber et al.*, 2011] and is able to update model states as well as model parameters. EnKF offers a flexible framework to jointly handle different types of errors and uncertainty with respect to forcings and model errors is relatively straightforward to incorporate. Since it is a sequential method it might also be more suitable for assessing the transient behavior of river bed conductivities than other calibration methods. Furthermore, EnKF allows an automatization of the adaptation of model states and parameters which is an important issue for real-time modeling and management systems [e.g., *Bauser et al.*, 2010]. Results from *Hendricks Franssen et al.* [2011] indicate that EnKF is able to adapt to seasonal changes in river bed conductivities caused by the temperature dependency of viscosity [*Doppler et al.*, 2007; *Engeler et al.*, 2011] with a time lag of about 3 months which suggests the principal capability of data assimilation to capture changes in river bed conductivities. However, no

systematic investigation of the behavior of EnKF toward time-variant river bed conductivities has been performed so far. Thus, the objectives of this study are to:

[11] 1. Identify whether EnKF is able to capture temporal changes of river bed conductivities under different conditions.

[12] 2. Identify the most important factors that do affect the update of river bed conductivities with EnKF.

[13] 3. Find out under which conditions the characterization of temporally variable leakage coefficients will yield improved flow predictions in practice and will therefore be recommended.

2. Methodology

[14] In this study we use the ensemble Kalman filter to determine time-varying river bed conductivities. EnKF is a Monte Carlo based method in which an ensemble of different model realizations (e.g., with varying model parameters or forcings) is propagated forward in time and updated whenever measurements of the model states (or parameters) become available. In contrast to other calibration techniques EnKF does not adjust parameter values of a model based on the residuals of the whole calibration period. Instead, it steps through time and only updates the model based on the measurements of one time step. EnKF is therefore an interesting method to calibrate time-dependent leakage values.

[15] The basic elements of EnKF are the forecast step, the observation equation, and the analysis step. The forecast step expresses how the states for the current time step are estimated from the past time step, on the basis of the simulation model (which solves numerically the governing equation), initial conditions, boundary conditions, model forcings, and model parameters:

$$h_i^0 = M(h_i^-, p_i, q_i), \tag{1}$$

where i is the stochastic realization ($i = 1, \ldots, N_{\text{real}}$), h_i^0 is the model state vector for the current time step, h_i^- is the model state vector for the previous time step, p_i are model parameters, q_i are model forcings, and M is the forward model which in our case solves the 3-D groundwater flow equation for variably saturated conditions. h_i^0 and h_i^- have a length that corresponds to the size of the problem, i.e., the number of nodes of the model N_{nodes}.

[16] Next the observation equation expresses how the simulation results at locations of observation points for the current time step are related to the whole state vector h_i^0:

$$y_i = H\, h_i^0, \tag{2}$$

where h_i^0 is the simulated state vector for the current time step, y_i is a vector for the simulated states at observation points with a length corresponding to the number of observations N_{obs}, and H is a matrix that extracts or interpolates the simulated results at observation points from the simulated state vector with a dimension of $N_{\text{obs}} \times N_{\text{nodes}}$.

[17] The measured states for the current time step y^0 are perturbed with values from a normal distribution with a mean of 0 and a standard deviation that corresponds to the measurement error:

$$y_i^0 = y^0 + \epsilon_i, \tag{3}$$

where y^0 is the measurement vector for the current time step (with length N_{obs}), ϵ_i is a perturbation vector, and y_i^0 is the perturbed measurement vector for realization i.

[18] Finally, the analysis step expresses how the forecasted states are corrected by the measurements. In the simplest configuration EnKF only updates the model states (in our case hydraulic heads) for the whole domain. However, EnKF has been reformulated so that also model parameters can be updated with an augmented state vector approach [e.g., *Chen and Zhang*, 2006; *Hendricks Franssen and Kinzelbach*, 2008]. The augmented state vector approach implies that the vector with the quantities to be updated in the data assimilation procedure contains not only the model states, but also (part of) the parameters. In our study either hydraulic heads (which correspond to the model state vector h) and river bed hydraulic conductivities (expressed as leakage coefficients L) or hydraulic heads, leakage coefficients and the hydraulic conductivity of the aquifer (K) are updated. Thus the most general form of the state-parameter vector x for our setup is given as

$$x = \begin{pmatrix} h \\ \log_{10}(K) \\ \log_{10}(L) \end{pmatrix}, \tag{4}$$

where h is the model state vector (i.e., hydraulic heads), $\log_{10}(K)$ and $\log_{10}(L)$ are the model parameters that should be updated and x is the augmented state-parameter vector with a total length of $N_{\text{nodes}} + N_{\text{parameters}}$.

[19] The analysis step itself is calculated with the following equation:

$$x_i^+ = x_i^0 + \alpha G(y_i^0 - y_i), \tag{5}$$

where x_i^0 is the simulated state-parameter vector of the ith ensemble member, x_i^+ is the updated state-parameter vector, G is the Kalman gain matrix, and α is a damping factor that takes values between 0 and 1 and is only used for updates of model parameters. This damping factor is used to reduce filter inbreeding, i.e., the underestimation of ensemble variance in the assimilation process [see *Hendricks Franssen and Kinzelbach*, 2008].

[20] The Kalman gain G in equation (5) is the ratio between the simulated uncertainty and the sum of simulated and measurement uncertainty:

$$G = C(HC + R)^{-1}, \tag{6}$$

$$C = \begin{pmatrix} C_{hy} \\ C_{\log_{10}(K)y} \\ C_{\log_{10}(L)y} \end{pmatrix}, \tag{7}$$

where C is the covariance matrix of the simulated states and parameters and R is the covariance matrix of the state

measurements. The covariance matrix C (equation (7)) is inferred from the states and parameters of the whole ensemble [h, $\log_{10}(K)$ and $\log_{10}(L)$] and the simulated values at observation locations y and has a dimension of $(N_{\text{nodes}} + N_{\text{parameters}}) \times N_{\text{obs}}$. R is inferred from measurement errors at observation points and has a dimension of $N_{\text{obs}} \times N_{\text{obs}}$. For our numerical experiments covariances between the measurement errors at different observation points were set to zero. Therefore, only the variances of observation points were considered in R.

3. Model Description

[21] The updating of time-variant L was tested in a synthetic experiment, but based on real-world data from the Limmat aquifer in Zurich (Switzerland). A schematic representation of the model domain is shown in Figure 1. The model domain covers an area of approximately 6 × 2 km. The rivers Sihl and Limmat are located at the eastern and northern boundaries of the model domain. In the Hardhof area (box in Figure 1) groundwater is extracted for drinking water supply. For that purpose, water is pumped from bank filtration wells located near river Limmat which is redistributed to the aquifer through several recharge wells and three recharge basins south of the river. Drinking water is then extracted through four drinking water wells which are located between the recharge basins and the river Limmat. For our simulations the numerical solution of the groundwater flow equation was calculated with the software SPRING [*Delta h Ingenieurgesellschaft mbH*, 2006] which uses a finite element scheme and is capable of simulating variably saturated flow and river-aquifer exchange fluxes. The model domain was discretized into 92,015 nodes, 173,599 elements, and 25 layers. The average element size was ≈50 m but a higher spatial discretization was present in the Hardhof area where the average element size was ≈20 m and refined up to ≈1 m near wells. Vertical discretization was 1.6 m.

[22] Forcing data for the simulations (river stages, recharge, pumping schedules, lateral inflows) were taken from real-world measurements. A detailed description of how these forcing data are calculated can be found in *Hendricks Franssen et al.* [2011]. As a brief overview, recharge was calculated on the basis of data from the meteorological station in Zurich-Affoltern (MeteoSwiss). For this purpose, potential evapotranspiration (ET) was calculated according to the Penman-Monteith equation. Actual ET was then estimated on the basis of calculated potential ET with the help of a soil water balance model by the FAO56 method [*Allen et al.*, 1998]. With the measured precipitation and calculated actual ET the potential recharge was calculated. Recharge was evenly distributed over the first layer of the model domain but only 15% of the calculated recharge was used because most of the model domain is an urbanized sealed area. Small lateral inflows exist on the south face of the model where water drains from the surrounding hill slopes of the Limmat valley. These lateral inflows were calculated on the basis of the estimated recharge rates. For the management activities in the Hardhof area (river bank filtration, artificial recharge) data for the extracted/redistributed amounts of water were available from Water Works Zurich on a daily basis. River stages were calculated with the help of the hydraulic software FLORIS (FLORIS2000, Software, Scietec Flussmanagement GmbH, Linz, Austria). Used input for these calculations were measured river stages at three locations, daily

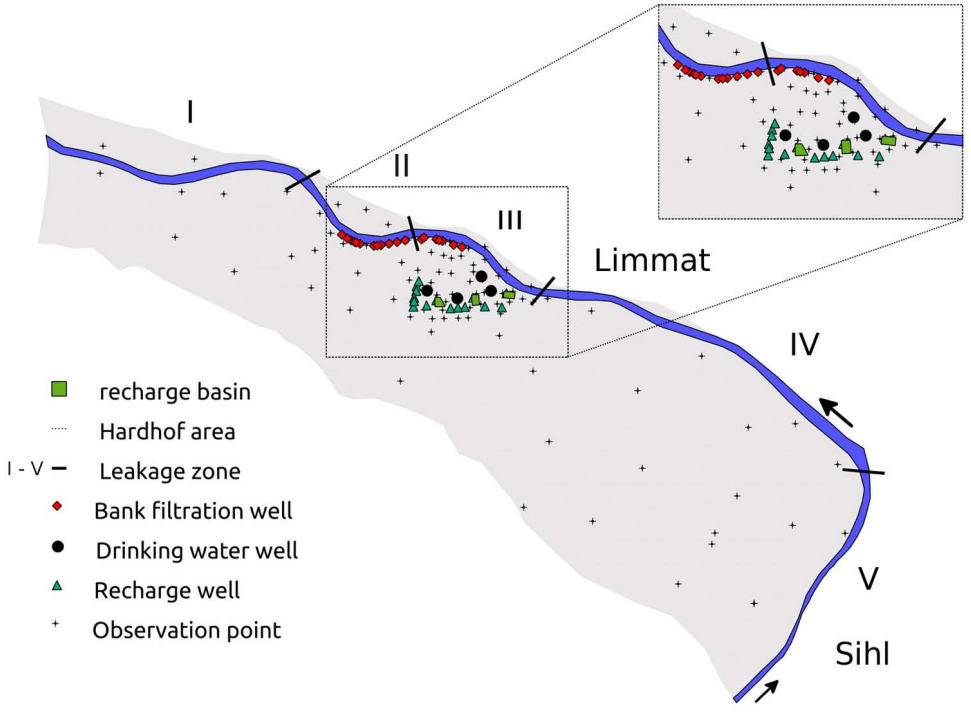

Figure 1. Schematic representation of model domain. River Sihl corresponds to leakage zone V and river Limmat corresponds to leakage zones I–IV.

discharge values for the Limmat and Sihl, and the expected geometry of the rivers Limmat and Sihl (interpolated from a large number of registered profiles along the river courses). Exchange fluxes between river and aquifer are incorporated in SPRING according to the leakage principle

$$Q = LA(h_{\text{river}} - h_{gw}), \tag{8}$$

where Q ($L^3\,T^{-1}$) is the exchange flux between river and aquifer, h_{river} (L) is river stage, h_{gw} (L) is hydraulic head underneath the river node, A (L^2) is surface area through which the flux occurs, and L (T^{-1}) is the leakage coefficient which is a lumped parameter of river bed conductivity and the depth of the river sediment.

[23] The river was implemented into the model with 457 leakage nodes which reside pairwise at the borders of the river. The river nodes were subdivided into five leakage zones with spatially constant L values within each of the zones. These five zones were originally defined for the region in order to capture the possible spatial variability of leakage coefficients. One of the five zones corresponds to the river Sihl, whereas the river Limmat is divided into four zones in correspondence with the position of two weirs and management activities.

[24] Basic model parameters were used from a prior calibration of the model. Porosity was set to a constant value of 0.15. Hydraulic conductivity K and leakage coefficients L were precalibrated with data from 87 piezometers for two calibration periods (June 2004 and July 2005) with the regularized pilot-point method [*de Marsily*, 1978; *Alcolea et al.*, 2006]. The two calibration periods were chosen because they include some main hydrologic features like one flooding event, intensive pumping activities, as well as mean flow conditions.

4. Synthetic Experiments

[25] Multiple reference runs were simulated with the model described above. These reference runs always had a specific evolution of L values which should mimic certain events in the river bed (see Table 1). The starting values of L for the reference runs were always equal to the ones determined in the calibration procedure and simulations were always performed for the period from January 2004 to August 2005 (609 days). From these reference runs daily hydraulic head data from 100 observation points were collected which were then used as conditioning data for the data assimilation with EnKF. The distribution of these observation points is shown in Figure 1. Observation points are mainly concentrated in the Hardhof area where most of the model dynamics takes place which is related to the pumping and artificial recharge activities in this area. For most of the ensemble runs with EnKF only L was assumed to be uncertain. For these scenarios K values and forcing data were equal to the ones in the reference run. Ensembles of L were generated by perturbing the $\log_{10}(L)$ values from the calibration (i.e., the starting values of the reference run) with samples from a normal distribution with a mean value of 0 and a standard deviation of 1 $\log_{10}(\text{s}^{-1})$. A total of 100 ensemble members was generated for the simulations with EnKF. The basic updating scheme for EnKF was to jointly update h and L with data from the 100 observation points every 10 days with a damping factor α of 0.1 and a uniform measurement error of 0.05 m at the observation points.

Table 1. Employed Reference Scenarios for the Temporal Evolution of Leakage Coefficients (L)

Scenario	Description	Temporal Evolution of L
A	Flooding Event	L constant until day 155; increase of L by 1 log unit at day 155; L remains constant until day 609
B	Sedimentation Event	L constant until day 200; linear decrease of $\log_{10}(L)$ by 1 log unit until day 400; L remains constant until day 609
C	Combined Flooding and Sedimentation Event	L constant until day 155; increase of L by 1 log unit at day 155; L remains constant until day 200; linear decrease of $\log_{10}(L)$ (1 log unit) until day 600; L remains constant until day 609
D	Temperature Dependency of L	L is corrected for daily changes in water viscosity with measured temperatures of the river Limmat

[26] In the second part of our study we varied the setup of the experiments in order to investigate the influence of different factors on the update of L with EnKF. This sensitivity study includes the effect of uncertain hydraulic conductivities, effects of spatially varying L values, the influence of a bias in the initial L ensemble and different updating strategies for EnKF. These different simulations always employed the reference scenario A from Table 1. Additionally, we performed simulations with an adaptive filtering approach which has already been used in the atmospheric data assimilation community and could be beneficial in reducing the possible problem of filter inbreeding in our simulations.

[27] The updated ensembles of h and L for the different reference scenarios were mainly evaluated with respect to the temporal evolution of the zonal L ensembles and the temporal evolution of hydraulic head errors at the observation points. For the evaluation of errors the root mean square error (RMSE) was calculated:

$$\text{RMSE}_h = \sqrt{\frac{1}{N_{\text{obs}} N_{\text{real}}} \sum_{i=1}^{N_{\text{obs}}} \sum_{j=1}^{N_{\text{real}}} (h_{ij} - h_i^{\text{ref}})^2}, \tag{9}$$

$$\text{RMSE}_L = \sqrt{\frac{1}{N_{\text{leak}} N_{\text{real}}} \sum_{i=1}^{N_{\text{leak}}} \sum_{j=1}^{N_{\text{real}}} \left[\log_{10}(L)_{ij} - \log_{10}(L)_i^{\text{ref}}\right]^2}, \tag{10}$$

where N_{obs} is equal to the number of observation points, N_{real} is the number of realizations, and N_{leak} is the number of leakage zones.

[28] It has to be noted that the evolution of leakage coefficients was not directly evaluated with the parameter L given in equation (8) but with the slightly modified parameter L^* which is the parameter L multiplied with half of the river width. This was done because the simulation code internally calculates with L^* but that does not directly influence the outcome of our simulations because the river width in the whole model domain is rather constant (about 50 m) and our emphasis was to compare the response of

EnKF to relative changes toward a reference state (i.e., the initial L^* values of the reference runs). Thus the two parameters L and L^* are used interchangeably in the following sections. As there is no explicit information about the thickness of river sediment in our model domain, river bed characteristics are always treated as effective parameters in our model evaluation, i.e., it is not distinguished whether changes in river bed properties are related to changes in bed thickness or changes in river bed conductivities.

5. Results

5.1. Increase of L (Scouring Event)

[29] In order to mimic a scouring of the river bed after a flooding event, L was increased by 1 log unit for all zones of the reference run after a major flooding event on day 155 (scenario A). The evolution of L values of the reference run for this scenario is shown together with the updated L values in Figure 2. For the update head observations were assimilated from the 100 observation points every 10 days and heads and L values were jointly updated at the observation times. It can be seen that the ensemble mean remains fairly constant at the beginning of the simulation. These nearly constant values at the beginning would be the expected behavior of EnKF because the initial ensemble mean for each leakage zone is equal to the leakage values of the reference run. After the jump of L in the reference run at day 155 the ensemble mean of updated L values starts to increase for all five leakage zones. This increase is strongest shortly after the flooding event and then slows down markedly within 150 days following the step change, after which L is nearly constant. The final ensemble means of L at the end of the simulation period (609 days) do not exactly match the postflooding reference values but all L values increase >0.6 log units after the flooding event.

[30] The variability of the ensemble continuously decreases during the course of the simulation (see shaded area in Figure 2). For the period before day 155 where the reference values for L are constant this should be the expected behavior of EnKF because the filter trusts the ensemble mean which is close to the true value. However, after the jump in L this process still continues and the variance in the ensemble of L almost vanished for three of the five zones by the end of the simulation period.

[31] In Figure 3 the temporal evolution of RMSE of forecasted hydraulic heads at observation points for this scenario is compared with the one of an unconditional ensemble simulation (i.e., neither h nor L are updated) and with a simulation run where only h was updated every 10 days. For the unconditional run the error rises by a factor of about 2 during the simulation period compared to the initial error, whereas RMSE(h) for the simulation where h and L are updated simultaneously reaches a value of about 5 cm by the end of the simulation. The error for the sole update of h remains approximately at the same magnitude as the initial error but shows rather high fluctuations between the times when the ensemble is updated. These fluctuations are related to the higher spread of the ensemble of L compared to the simulation where also L was updated. After the flooding event these fluctuations increase in magnitude caused by the fact that systematic errors are also introduced due to the increase of L in the reference run which cannot be captured when only h is updated. This also leads to relatively high errors around day 550 where pumping rates were temporally increased. In contrast, errors for a joint update of h and L did not increase for this event which is related to the adaptation of L and its decreased ensemble variance.

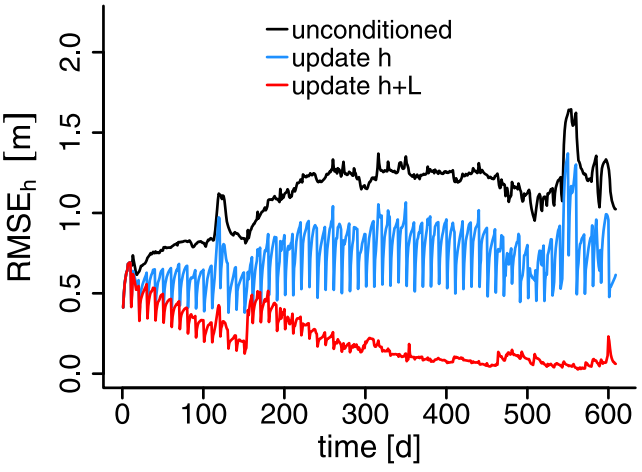

Figure 3. Temporal evolution (period January 2004–August 2005) of root mean square error of hydraulic heads (RMSE$_h$) at observation points for an unconditional simulation (unconditioned, no update of hydraulic heads and leakage coefficients), an update of hydraulic heads every 10 days (update h) and for a joint update of hydraulic heads and leakage coefficients every 10 days (update $h + L$).

[32] Figure 4 gives an overview over the fluxes between river and aquifer. The diagram for the unconditional simulation shows that the variability of positive and negative fluxes is very high for the utilized initial ensemble of L. For

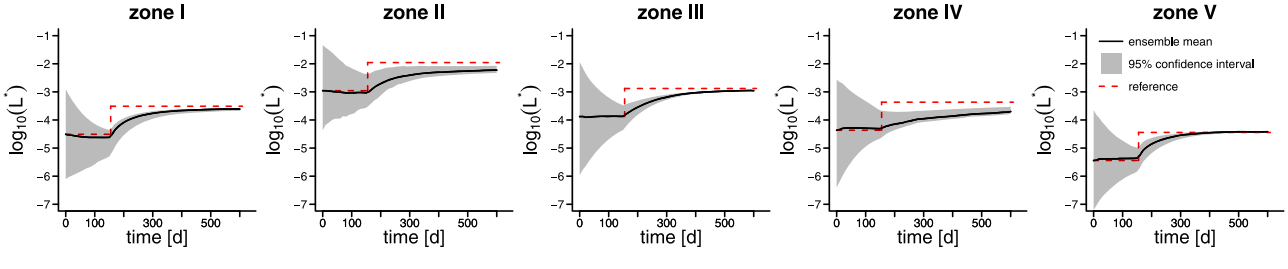

Figure 2. Temporal evolution of zonal leakage coefficients (period January 2004–August 2005) for reference run and update with the ensemble Kalman filter for simulation scenario A (Table 1).

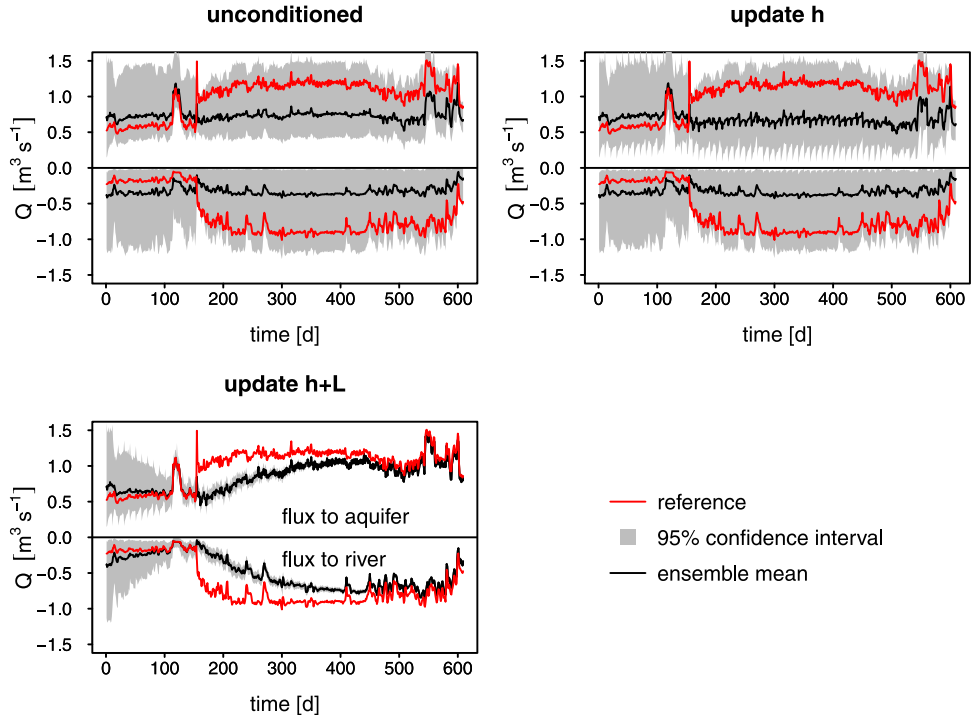

Figure 4. Temporal evolution (period January 2004–August 2005) of fluxes between river and aquifer for an unconditional simulation (unconditioned, no update of hydraulic heads and leakage coefficients), an update of hydraulic heads every 10 days (update h), and for a joint update of hydraulic heads and leakage coefficients every 10 days (update $h + L$). Fluxes in the upper (positive) part of each diagram are from river to aquifer and fluxes in the lower part (negative) are from aquifer to river.

unconditional simulations the mean fluxes for both directions (gaining and loosing conditions of the river) are rather constant throughout the simulation period with the exception of two periods of increased pumping activities around day 120 and day 550. The fluxes for the reference run are markedly increased after the change in L at day 155. Especially when only h is updated the ensemble mean for fluxes from river to aquifer are slightly corrected after each assimilation cycle but it is not possible for EnKF to adjust the fluxes solely by correcting h. For a joint update of h and L the ensemble mean of both flux directions is consecutively adjusted to the reference values which are approximated very closely after day 450.

[33] Figure 5 additionally shows the spatial distribution of averaged exchange fluxes for different time steps along the X axis of the model domain (in the Swiss coordinate system). For time step 100 the distribution of fluxes is very similar for all different updating scenarios and also closely corresponds to the one in the reference run. Shortly after the change of river bed properties (time step 200) the

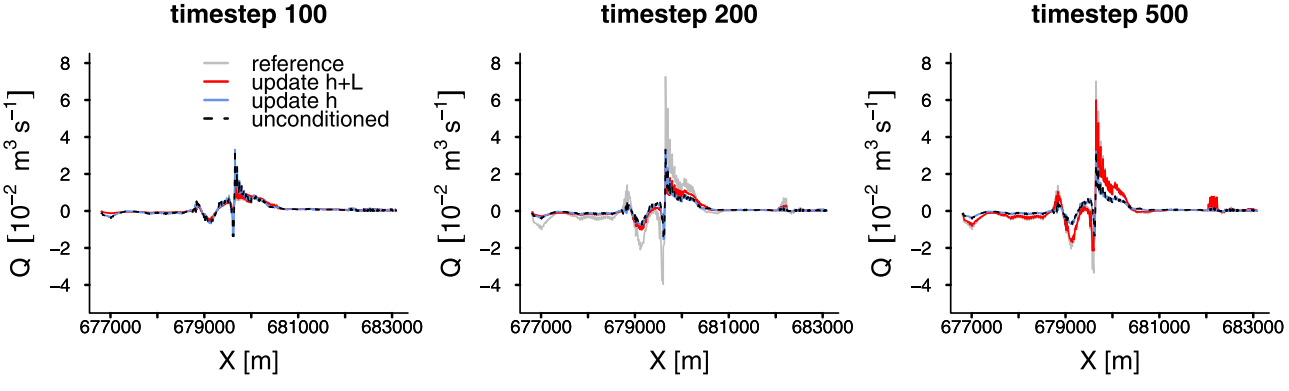

Figure 5. Spatial distribution of river aquifer exchange fluxes for an unconditional simulation (unconditioned, no update of hydraulic heads and leakage coefficients), an update of hydraulic heads every 10 days (update h), and for a joint update of hydraulic heads and leakage coefficients every 10 days (update $h + L$).

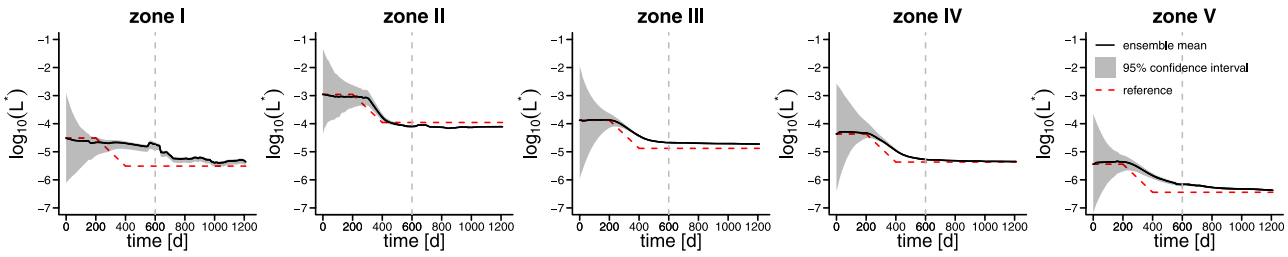

Figure 6. Temporal evolution of zonal leakage coefficients (period January 2004–April 2007) for reference run and update with the ensemble Kalman filter for simulation scenario B (Table 1).

spatial distributions for the different scenarios all deviate from the reference run. An update of states and parameters corrects this deviation as the assimilation proceeds (see time step 500). However, an update of states alone does not correct for the biased exchange fluxes.

5.2. Decrease of *L* (Sedimentation Event)

[34] In a next step a sedimentation event was emulated in order to verify whether EnKF also adapts to slowly decreasing *L* values. A reference run was generated where *L* was decreased by 1 log unit over a period of 200 days (starting from day 200 until day 400) for all five leakage zones (scenario B). The updating strategy was the same as for the scouring test case (scenario A). The evolution of *L* in the reference run together with the updated *L* values are shown in Figure 6. Again the ensemble means for the different zones remain fairly constant before the change in *L*. When *L* starts to decrease after day 200 zonal ensemble means of *L* also decrease during the EnKF update with the exception of the most western zone in the model domain (zone I). In this part of the aquifer the groundwater levels are rather close to the river bottom and the initial *L* values are low. As a consequence, the leakage fluxes between river and aquifer are generally low. The modification of *L* that we simulated in the reference run has a limited impact on the hydraulic heads close to the river and therefore piezometric head measurements will not be so effective (compared to other zones) for adapting the value of *L*. For this experiment, *L* generally does not adapt so well to the reference *L* values because the adaptation at the end of the simulation period is worse for three of the five zones (as compared with the scouring experiment). However, when the simulation time is expanded to 1200 days a similar adaptation as in scenario A was observed for all of the five zones. The variances of the *L* ensembles for the different zones behave similarly to the flooding scenario, i.e., they show a continuous decrease throughout the simulation period. For four of the zones variability within the ensemble has almost vanished at the end of the simulation period.

5.3. Performance of EnKF for Combined Flooding and Sedimentation Events

[35] The two simulations described above have shown that EnKF is principally capable of adapting toward increases and decreases of *L*. However, natural sediment dynamics often exhibit a sequence of sedimentation and scouring events [e.g., *Blaschke et al.*, 2003]. Therefore, such a scenario was artificially constructed by appending a sedimentation period to scenario A. In this scenario *L* values for all leakage zones were raised by 1 log unit at day 155. *L* is then held constant for 45 days and then again decreased by 1 log unit from day 200 until day 600 in order to simulate a long lasting sedimentation event (scenario C). The evolution of the updated *L* ensembles together with the corresponding reference values is depicted in Figure 7. From all five leakage zones it becomes obvious that EnKF corrects for both changes of *L* within the simulation period. After the flooding event the ensemble means of *L* start to increase. Approximately at day 300 this increase reaches an apex and the ensemble means of *L* begin to adapt to the decreasing values of the reference run. The zonal ensemble means of *L* rise between 0.2 and 0.5 log units in the phase after the flooding event. For all five leakage zones the reference value is approximately intersected at these maximal values of the ensemble mean of *L*. After that peak the ensemble means decrease but only for one zone (which corresponds to the lowest prior increase in *L*) the reference value is reached within the simulation period. For the other four leakage zones the ensemble means adapt rather slowly toward the reference line. This again shows that the adaptation with EnKF toward changes in *L* has a rather long response time which does not optimally capture rapid changes in river bed properties.

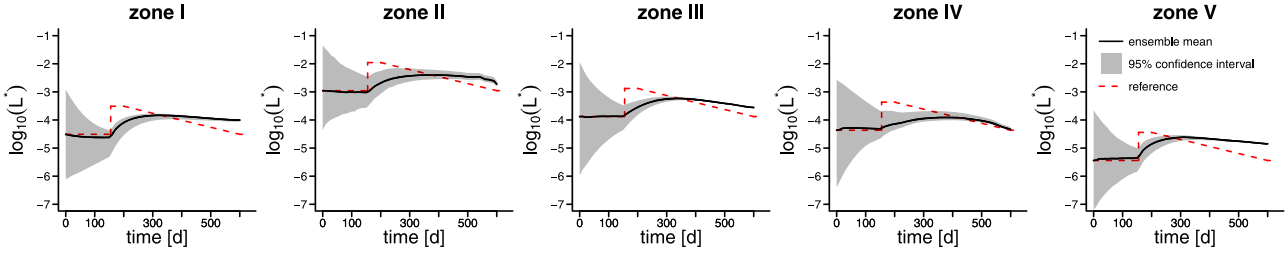

Figure 7. Temporal evolution of zonal leakage coefficients (period January 2004–August 2005) for reference run and update with the ensemble Kalman filter for simulation scenario C (Table 1).

5.4. Temperature Dependency of *L*

[36] Apart from mechanically induced changes of the river bed like scouring and sedimentation the hydraulic conductivity of the river bed can also be influenced on a seasonal scale by temperature variations of the river due to the temperature dependency of viscosity as can be seen from equations (11) and (12) [*Muskat*, 1937]:

$$K(T) = \frac{K_{\text{intr}}\,\rho\,g}{\mu(T)}, \tag{11}$$

$$\mu(T) = 2.414 \times 10^{-5} 10^{\frac{247.8}{T-140}}, \tag{12}$$

where $K(T)$ is temperature dependent hydraulic conductivity ($L\,T^{-1}$), K_{intr} is intrinsic permeability of the porous medium (L^2), ρ is density of water ($M\,L^{-3}$), g is acceleration of gravity ($L\,T^{-2}$), $\mu(T)$ is temperature dependent dynamic viscosity of water ($M\,L^{-1}\,T^{-1}$), and T is temperature (θ).

[37] Within our simulation period the measured temperatures of the river Limmat varied from 4 to 26°C, which translates into a variation of viscosity of up to a factor of 1.7, whereas the change of water density in this temperature range is less than 1%. It has been shown by *Engeler et al.* [2011] that this temperature dependency of L can have a profound influence on the predicted groundwater dynamics close to rivers. In our simulation model L values are not corrected for temperature variations of the river but usually these variations occur in natural settings. Therefore, we created a reference run in which these natural variations of L occur and tested whether EnKF is able to follow these changes. For this purpose, we corrected the L values of all leakage nodes of the reference run with the measured temperature of the river Limmat on a daily basis so that all leakage nodes follow the same variation pattern (scenario D). The temporal evolution of zonal L values of the reference run is shown in Figure 8 together with the updated ensemble values. The L values of the reference run follow a seasonal cycle with higher values during the summer months and lower values during winter. The maximal change of L values for the reference run within the simulation period was 0.23 log units and thus is lower than for the previous scenarios. For zones I and II the update of L with EnKF shows some temporal changes but they do not correspond very well with the temporal dynamics of the reference run. For zones III to V the temporal evolution of updated L values is more close to the reference run especially for zone III. Again a certain time lag is observed between the reference and the updated L values which seems to be a bit higher for zones IV and V than for zone III. The rather high sensitivity of zone III toward the temperature-dependent changes of L may be related to the fact that the highest exchange fluxes between river and aquifer occur in this part of the model domain. As a result, the ensemble of L values for this zone responds well to the low changes of L that are induced by the temperature dependency. For leakage zones I and II the groundwater table is very close to the river bottom and thus the fluxes between river and aquifer are rather low, which might be a cause for the rather low sensitivity of these two zones. For leakage zones IV and V the groundwater table is significantly lower than the river bottom but also the L values are low for these two zones, which results in lower exchange fluxes compared to leakage zone III and may be a cause for the lower sensitivity toward the relatively small changes in L that are induced by river temperature fluctuations.

5.5. Influence of Spatial Patterns

[38] For the previous simulations the change in L was always realized in a spatial homogeneous fashion, i.e., L was changed for all five zones with the same magnitude. In natural systems changes in L due to sedimentation or scouring are expected to happen in a more spatially inhomogeneous fashion. This may happen for example due to meandering of the river which provides spatial sequences of sedimentation and scouring zones which may proceed downstream over time. Also storm events may cause scouring of the river bed preferentially at certain zones where flow velocities and thus shear stress is higher. Therefore, we also investigated how EnKF reacts toward changes in L that occur only at a part of the river reach. For this setup the value of L was only increased for leakage zone III, while for the other zones L remained constant throughout the simulation period. Results for this setup are shown in Figure 9. The ensemble means for the four zones which are not subjected to any changes in L remain almost constant throughout the simulation period. For zone III EnKF adapts the ensemble mean closely to the reference value by the end of the simulation and the course of the adaptation curve is similar to the ones observed when the L values for all zones are increased (see Figure 2). This means that EnKF captures the spatially separated evolution of L for the different zones which is possibly due to the fact that correlations between the hydraulic heads at the observation points and the L ensemble at the four other leakage zones are rather weak. Beforehand it was expected that at least the two neighboring leakage zones would also be affected by the update with EnKF because the filter might give some weight to them in the updating procedure.

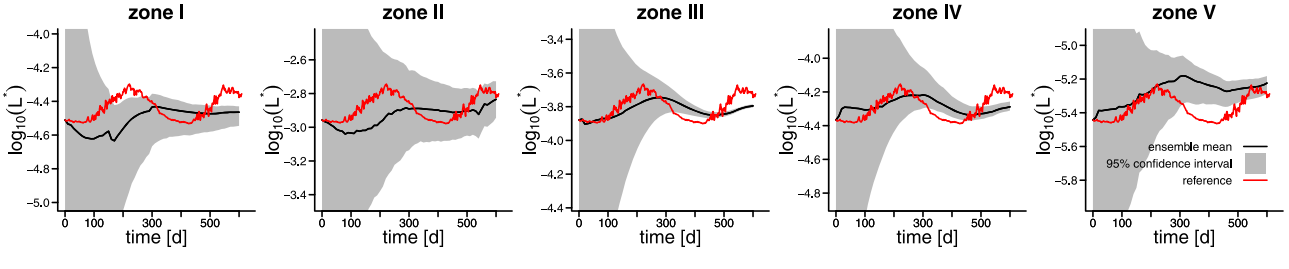

Figure 8. Temporal evolution of zonal leakage coefficients (period January 2004–August 2005) for reference run and update with the ensemble Kalman filter for simulation scenario D (Table 1).

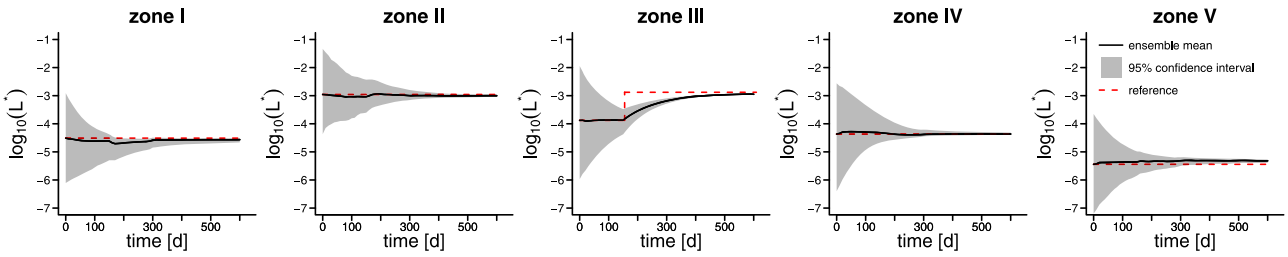

Figure 9. Temporal evolution of zonal leakage coefficients (period January 2004–August 2005) for reference run and update with the ensemble Kalman filter when the zonal leakage coefficient of zone III in the reference run is increased by 1 log unit at day 155.

However, this was not the case in this scenario. One possible reason could be that the leakage zone is close to the managed site (Hardhof area) where the groundwater pumping and artificial recharge takes place. Therefore, the groundwater levels at the observation points in this area which comprise the major part of all observation points are especially affected by changes of the increasing leakage zone.

5.6. Influence of Uncertain Hydraulic Conductivities

[39] In previous simulations L was the only uncertain parameter and hydraulic conductivities of the aquifer were assumed to be known exactly. This assumption was introduced in order to isolate the effect of uncertain L values in the update with EnKF. This simplification is rarely justified for real world examples because the uncertainty regarding K very often governs the uncertainty of model output. Furthermore, it was shown [e.g., *Kalbus et al.*, 2009] that heterogeneity of the underlying aquifer plays a major role for exchange fluxes between river and aquifer and thus should ideally not be neglected for the simulations. For these reasons simulations were performed in which also the ensemble members of K are different among each other. The ensemble of K was generated by conditional sequential Gaussian simulation [*Gómez-Hernández and Journel*, 1993] on a very fine grid which was then upscaled to the simulation grid through simplified renormalization [*Renard et al.*, 2000]. For details of the ensemble generation see *Hendricks Franssen et al.* [2011]. For these realizations the variance in $\ln(K)$ was 2.7.

[40] The setup of this experiment was similar to scenario A with the only difference that the ensemble of K showed the described uncertainty and values of K and L were jointly updated together with hydraulic heads via EnKF. When comparing the temporal evolution of zonal ensemble means of L (Figure 10) only small differences are observable between the runs with and without uncertain K values. The RMSE of hydraulic heads at observation points during the first 100 days was higher for the ensemble with uncertain K values compared to the one with deterministic K values which is due to the additional uncertainty for this parameter. In later steps RMSE for both simulations is very similar which is mainly caused by the adaptation of K values with EnKF which decreases the variance of K in the assimilation process. A comparison of the initial and updated mean fields of $\log_{10}(K)$ is given in Figure 11. It can be seen that the initial mean field of $\log_{10}(K)$ is relatively smooth. During assimilation the structure of the $\log_{10}(K)$ generally gets more patchy because every finite element of the model is allowed to update separately. However, the major distribution of $\log_{10}(K)$ values (increasing values from east to west) is preserved.

5.7. Influence of Ensemble Bias

[41] Besides the uncertainty regarding hydraulic conductivities mentioned in section 5.6 the updating of L with EnKF may also be affected by a bias of the initial ensemble of L. In the previous simulations the zonal mean values of the initial ensemble of L were equal to the true values of the reference run. However, in real-world situations information on river bed conductivities within the model domain often is scarce and the generation of the initial ensemble of L may thus only rely on a few or even no measurement data. Therefore, we created an initial ensemble of L where the ensemble means of all five leakage zones were the same and corresponded to the arithmetic mean of $\log_{10}(L)$ values of the reference run (averaged over all five leakage zones). The ensemble variances of the different zones were similar to the ones in the previous simulations. Results for this biased initial ensemble are shown in Figure 12. Because all zonal

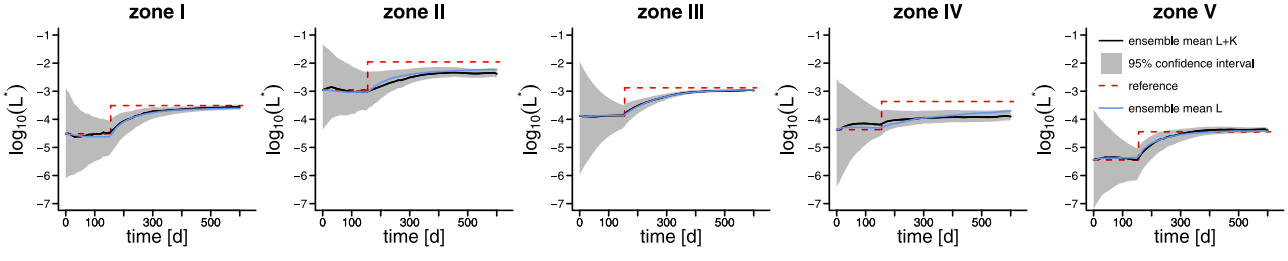

Figure 10. Temporal evolution of zonal leakage coefficients (period January 2004–August 2005) for reference run and an update of either leakage coefficients alone (ensemble mean L) or a joint update of leakage coefficients and hydraulic conductivities of the aquifer (ensemble mean $L + K$) for simulation scenario A (Table 1).

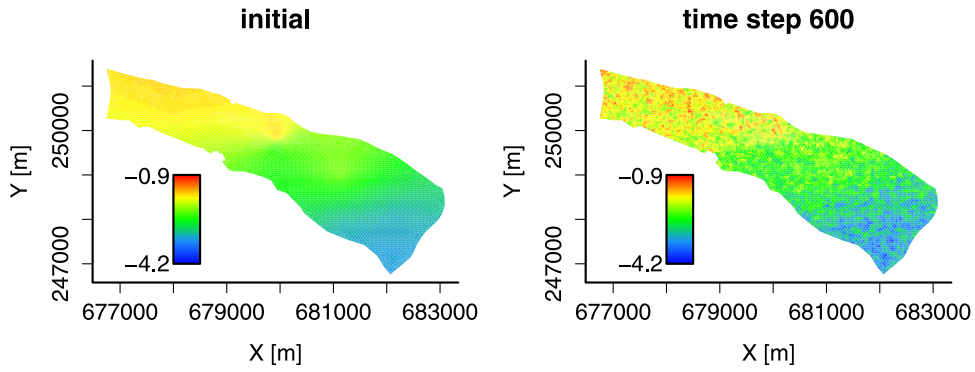

Figure 11. Ensemble mean of $\log_{10}(K)$ for initial ensemble (left) and at time step 600 (right) for model layer 4 when hydraulic conductivities and leakage coefficients are jointly updated every 10 days.

ensemble means of L started with the same value the bias between the initial ensemble mean and the reference value is different among the five leakage zones. During the period before the increase of L in the reference run (up to day 155) the ensemble means of all zones more or less tend to move toward their corresponding reference value. After the jump of L values in the reference run ensemble means of all zones start to increase no matter whether their tendency was to increase or decrease before the flooding event. The evolution of zonal L values looks rather similar to the ones shown before, i.e., a steeper increase at the beginning which flattens after about 100 days. Also the variance of the ensembles behaves similar to the previous examples. Due to the initial bias the absolute performance of EnKF is not as good as for a case where the zonal ensemble means are closer to the true values. However, a distinct reaction of EnKF toward the true values is clearly visible.

5.8. Sensitivity on Updating Strategy

[42] The performance of EnKF in part depends on the amount of available observation data and on filter specific settings like the number of ensemble members, the updating interval, or the damping factor α. Compared to the base scenarios (100 ensemble members, update frequency of h and L: 10 days, $\alpha = 0.1$, 100 head observations) each of these meta parameters was changed in order to see whether they significantly affect the performance of EnKF for our setup. The different parameters for each of these scenarios are summarized in Table 2 where scenario "base" is essentially the base scenario which serves as a reference for the other scenarios.

[43] An increase of the ensemble size to 200 ensemble members (scenario "ens_200") did not significantly improve L compared to the base scenario as the performance for one of the leakage zones slightly decreases.

[44] The effect of updating frequency on the evolution of L can be seen in Figure 13 for four different updating frequencies (1, 2, 5, and 10 days). In general, an increase of the updating frequency for h and L did not lead to an improvement of L updates. For an updating frequency of 5 days the adaptation of L was slightly faster for zone III but the performance for the other zones was equal or slightly worse compared to the base scenario. An updating frequency of 1 or 2 days did not increase performance in any of the five leakage zones. Especially for zones II and IV the performance degrades when h and L are updated very frequently. A reason for this behavior may again lie in the fast decrease of ensemble variance in the updating procedure. In Figure 13 also the ensemble standard deviation is compared for the different updating frequencies. With decreasing updating intervals the ensemble variance also decreases very rapidly for all leakage zones. For updating frequencies of 1 or 2 days the ensemble variance is almost zero before the change in L at day 155, whereas for lower updating frequencies more variability is maintained in the ensemble.

[45] For a sensitivity analysis on the number of observations points, six different configurations were compared where the number of observation points ranged from 5 to 200. The observation points for the scenarios "nobs_5" to "nobs_50" always were a subset of the scenario "nobs_100," which is the configuration used in the previous simulations (equal to the base scenario). For these scenarios the observation points

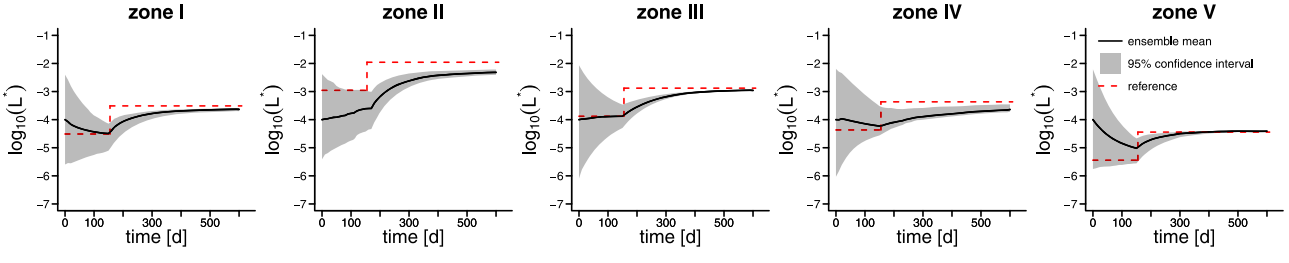

Figure 12. Temporal evolution of zonal leakage coefficients (period January 2004–August 2005) for reference run and update with the ensemble Kalman filter for scenario A when the initial ensemble mean of leakage coefficients was biased compared to the initial values of the reference run.

Table 2. Different Updating Scenarios for Zonal Leakage Coefficients

Name of Scenario	Update Frequency $h + L$ (day)	Number of Ensemble Members	Number of Observation Points	Damping Factor α
Base	10	100	100	0.1
fupd_*xx*	*xx* = 1, 2, 5, 10	100	100	0.1
ens_*xx*	10	*xx* = 100, 200	100	0.1
nobs_*xx*	10	100	*xx* = 5, 10, 20, 50, 100, 200	0.1
damp_*xx*	10	100	100	*xx* = 0.1, 0.2, 0.5, 1.0

were coincident, e.g., the observation points for scenario "nobs_5" were contained in all other scenarios. One exception is the comparability between scenario "nobs_20" and "nobs_50," where both scenarios only shared 10 observation points. For the scenario "nobs_200" 100 additional observation points were added to "nobs_100" and these were spread over the whole model domain. In Figure 14 the temporal evolution of RMSE(L) is compared for the different number of observation points. It is clearly visible that the errors in L are inversely correlated to the number of observation points. For the pre-event period the differences among the scenarios are mostly related to the different decrease in ensemble variance, which means that the ensemble variance in case of few observation points does not decrease as fast as for a larger amount of observation points. In the postevent period also the different updating behavior toward L contributes to the differences among the scenarios. At the end of the simulation period the errors in L follow an exponential decrease with the number of observation points. However, it can be observed that even as little as 10 observation points result in a reduction of RMSE(L) of 70% as compared to the open loop simulations. Nevertheless, RMSE(L) is roughly twice as large at the end of the simulation period for 10 observation points compared with that for 100 observation points.

[46] A closer look at the temporal evolution of the zonal ensemble means of L for the different scenarios (data not shown) reveals that the updating capability of EnKF for the two zones close to the Hardhof area (zones II and III) and for the Sihl (zone V) is rather similar among the scenarios, whereas for the two other zones the updating capability is more strongly dependent on the number of observation points. This might be an indicator that the observation points which were used for the simulations with a lower number of observation points generally had a lower sensitivity toward these two leakage zones. This also means that in order to be able to adapt the L values for all leakage zones a sufficient number and spatial distribution of observation points is needed.

[47] Furthermore, it was tested whether the damping factor α could be raised for the update of L which could possibly lead to a faster adaptation of zonal ensemble means toward reference values. This option was tested for four values of α (0.1, 0.2, 0.5, and 1.0). Results for the different values of the damping factor (data not shown) showed that higher values than 0.1 (the base updating scenario) generally lead to a worse performance of EnKF in updating the zonal ensemble means toward the reference values, which is possibly related to the more rapid decrease in ensemble variance for higher α values which is in accordance with the findings by *Hendricks Franssen and Kinzelbach* [2008].

5.9. Use of Covariance Inflation to Improve Filtering Results of EnKF

[48] In all different scenarios it became obvious that the response time of EnKF to adapt for changing L values is rather long, which is to some part related to the fast decrease in ensemble variance. In order to compensate for this loss of

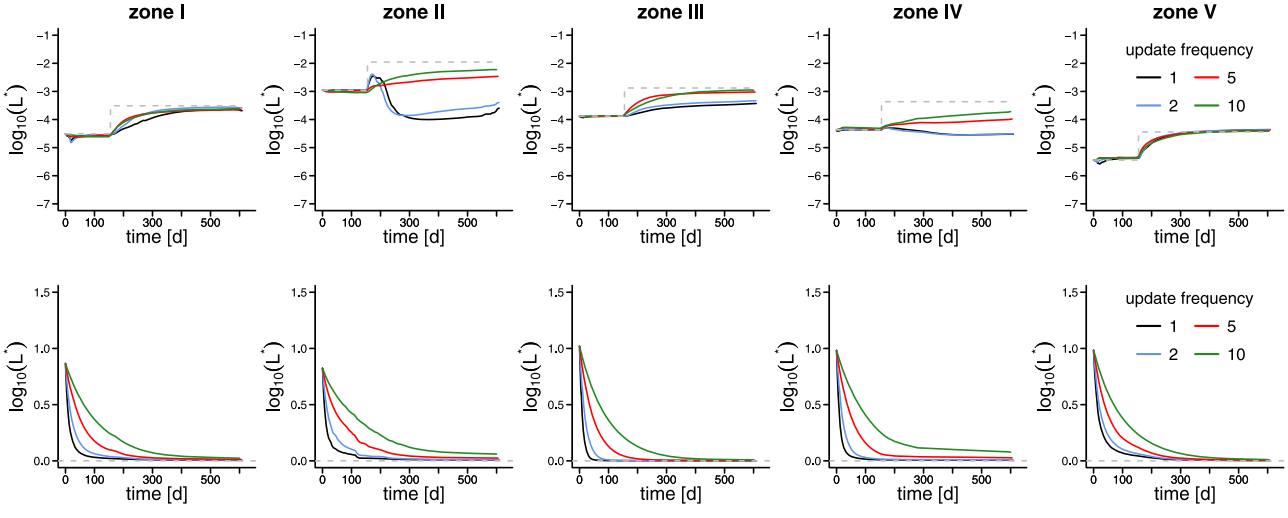

Figure 13. Temporal evolution (period January 2004–August 2005) of ensemble mean (upper row) and ensemble standard deviation (lower row) of zonal leakage coefficients for different updating frequencies of hydraulic heads and leakage coefficients (simulation scenario A).

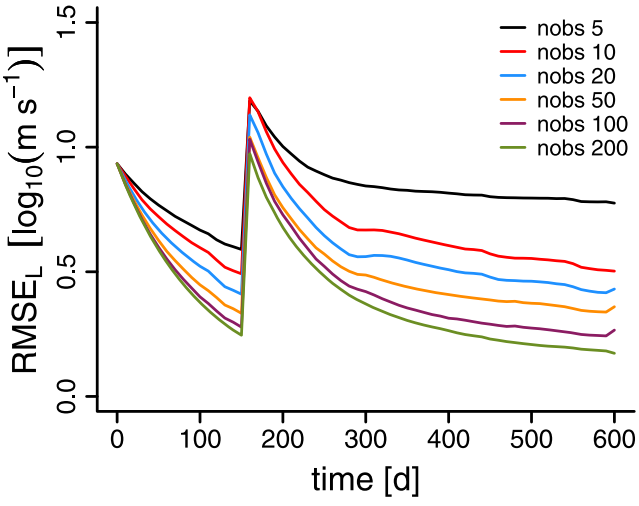

Figure 14. Temporal evolution (period January 2004–August 2005) of root mean square error of leakage coefficients for different numbers of observation points (simulation scenario A).

variance which is related to filter inbreeding, we applied covariance inflation which could possibly improve the response time of EnKF and is a common approach in atmospheric data assimilation [e.g., *Hamill et al.*, 2001; *Anderson*, 2007, 2009]. For covariance inflation an inflation factor λ is used to spread the ensemble around its mean value before every assimilation step:

$$x_{ij} = \lambda(x_{ij} - \overline{x_i}) + \overline{x_i}. \tag{13}$$

[49] Before each assimilation cycle every element i of the state-parameter vector x_{ij} for the jth realization is inflated around the ensemble mean $\overline{x_i}$ with the inflation factor λ. This means that the ensemble mean $\overline{x_i}$ for every element is preserved and only the spread of the ensemble is slightly increased. In atmospheric sciences it was especially in the past common to set the inflation factor λ to a constant value [e.g., *Hamill et al.*, 2001]. Recently, it has also been proposed to make λ temporally and spatially variable [e.g., *Anderson*, 2007, 2009]. We used the inflation method proposed by *Anderson* [2007] in which λ is temporally variable. In this method a Bayesian update is performed on λ every assimilation cycle and the new value of λ is mainly a function of ensemble variance, measurement errors, deviations between observations and simulations, and the variance of λ itself (which is kept constant in our case). This approach allows EnKF to correct for deficiencies in the assimilation process (i.e., filter divergence) by inflating the ensemble based on the residuals at observation points and the ensemble variance. An outline of the used algorithm is given in Appendix A. For a more detailed insight into the methodology we refer to *Anderson* [2007].

[50] Figure 15 compares the evolution of L for scenario A without and with adaptive covariance inflation. For the simulations with adaptive inflation L values for all zones reach the reference values more closely than for simulations without inflation. Furthermore, the adaptation time to reach a certain L value is slightly decreased when covariance inflation is used. The evolution of ensemble variance for covariance inflation also shows the expected behavior. Before the flooding event the decrease in variance is comparable to the simulations without inflation. When L is increased in the reference run λ values increase due to the higher residuals at the observation points which also leads to an increasing ensemble variance which allows EnKF to update L more closely (and faster) to the reference values.

[51] It was already mentioned that an increase of α in the simulations without inflation leads to a worse performance of EnKF. This effect can be seen in the upper row of Figure 16 where α was set to a value of 0.2. In this case the adaptation time for zones III and V is decreased but the absolute adaptation for zones II and IV were worse compared to $\alpha = 0.1$. However, when covariance inflation is used for this example L is adapted faster and more accurate compared to the base case (i.e., no inflation, $\alpha = 0.1$).

[52] We additionally repeated the simulations for the other three base case scenarios (scenario B, C, and D) with covariance inflation for $\alpha = 0.2$ (see Figure 17). For

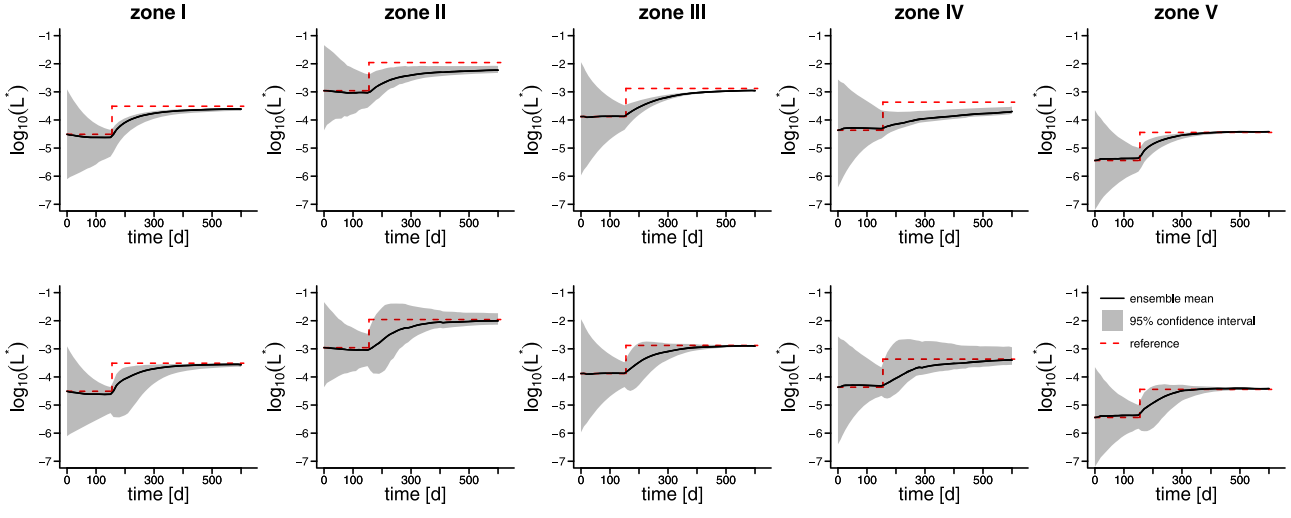

Figure 15. Comparison of temporal evolution of zonal leakage coefficients without (top row) and with (bottom row) adaptive covariance inflation for scenario A (damping factor $\alpha = 0.1$).

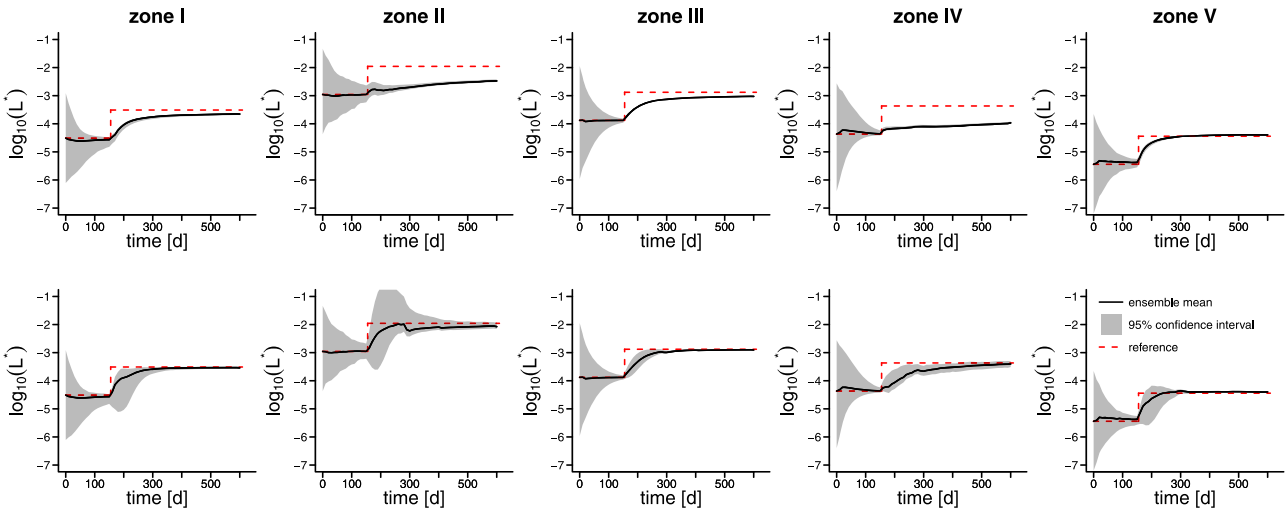

Figure 16. Comparison of temporal evolution of zonal leakage coefficients without (top row) and with (bottom row) adaptive covariance inflation for scenario A (damping factor $\alpha = 0.2$).

scenarios B and C the adaptation time generally decreases and also the accuracy at the end of the assimilation period is improved although for one zone (zone I in scenario B) some slight instability occurs which again is possibly related to the low sensitivity of this zone for that scenario which was also obvious from the corresponding base case scenario. For the variability of L due to temperature changes (scenario D) a positive effect of covariance inflation was not significant which is most likely related to the small changes of L in this scenario.

[53] It has already been mentioned that an increase of updating frequency did not improve the adaptation of L for the base scenario. However, when covariance inflation is used (see Figure 18) the adaptation time can generally be decreased with a higher updating frequency. This can be observed for all of the five leakage zones where updating frequencies of 2 or 5 days consistently performed better than an updating frequency of 10 days. The ensemble standard deviation (lower row in Figure 18) increases after the change in L at day 155 due to covariance inflation.

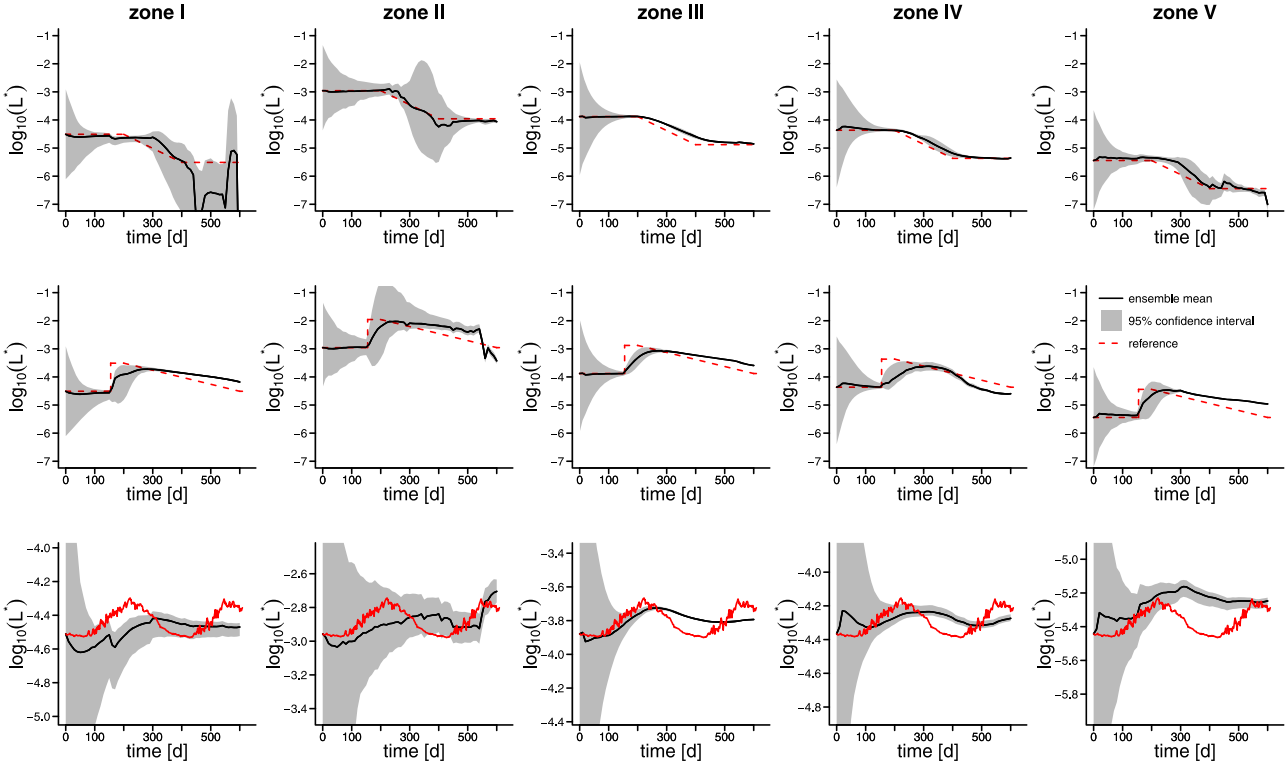

Figure 17. Temporal evolution of zonal leakage coefficients with adaptive covariance inflation for scenario B (top row), C (middle row), and D (bottom row) (damping factor $\alpha = 0.2$).

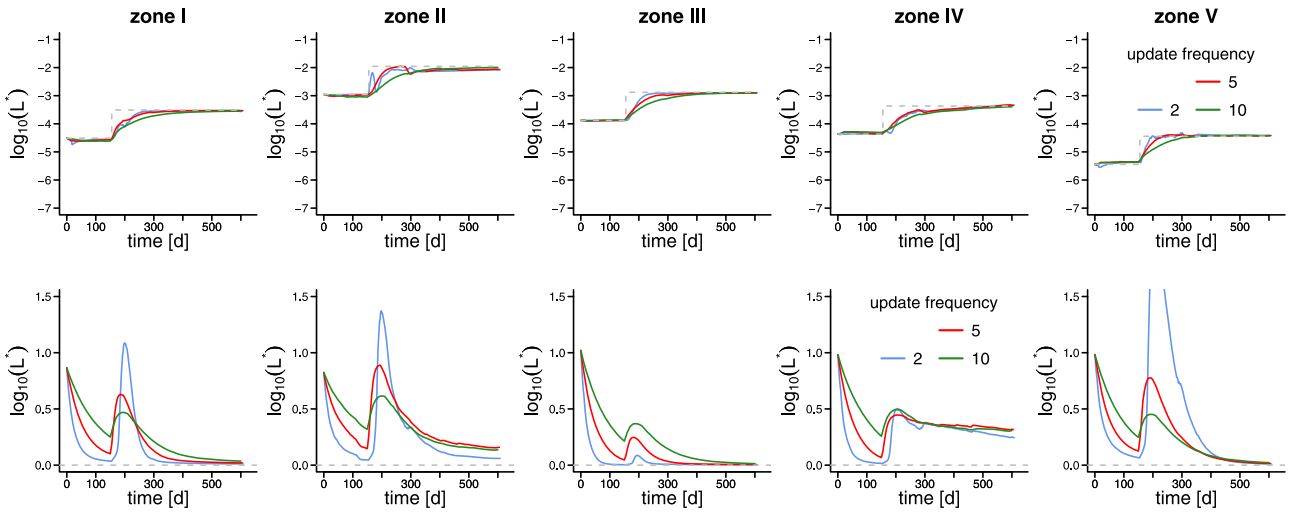

Figure 18. Temporal evolution (period January 2004–August 2005) of ensemble mean (top row) and ensemble standard deviation (bottom row) of zonal leakage coefficients for different updating frequencies of hydraulic heads and leakage coefficients with covariance inflation (simulation scenario A).

Especially for an updating frequency of 2 days this increase in ensemble variance is already rather high for certain leakage zones. For an updating frequency of 1 day (data not shown) this increase in ensemble variance is even higher which leads to numerical problems in the groundwater model due to rather extreme L values. Hence, a simulation with an updating frequency of one day could not be performed successfully. Nevertheless, updating frequencies of 2 or 5 days together with covariance inflation improved the adaptation of L considerably in terms of response time.

6. Discussion

[54] The update of L with EnKF showed some general characteristics among all scenarios. In almost every case all zonal mean values of L were updated by EnKF in order to follow the trend of the reference run. For the scenario with either an increase or decrease of L the updated zonal ensemble means were quite close to the reference values at the end of the simulation period. However, the adaptation time to achieve this improvement was somewhat high within all scenarios. This also became obvious in the scenario where a flooding and a sedimentation event were combined within the simulation period. Here the slow adaptation led to the effect that the changes in zonal ensemble means were within a smaller margin and the extreme values of the reference run were never reached.

[55] A reason for the slow adaptation may lie in the rather rapid decrease of ensemble spread during the simulations. The decrease of variance is already obvious in the time before a change happens in the L values of the reference run and proceeds as the L values are adapted by EnKF. The lowering of the ensemble spread before a change in L occurs in the reference run is a native feature of EnKF. In this case the zonal ensemble means are close to the reference values and because of this the errors in hydraulic heads are rather small. Therefore, the filter "trusts" the ensemble means and adapts the extreme values of the ensemble toward the ensemble mean, which consecutively lowers the ensemble spread. However, this decrease of variance during the period of constant L values then possibly hampers the adaptation by EnKF toward the changed L values. Directly after the sudden change of L, differences between simulated and measured piezometric heads become larger, but the limited ensemble variance implies that the model predictions have a relatively large weight in the EnKF procedure, limiting the influence of the observations and slowing down the adaptation of L. The use of adaptive inflation leads to a faster adaptation of L toward the reference values and at the same time to a more precise determination of L at the end of the simulation period compared to the use of EnKF without inflation. In general, the adaptive inflation method seems to be robust to time-variant model parameters and it also honors changes in the prediction capability of the forward model by increasing the variance for a larger prediction error.

[56] In our experiments we also found that increasing the updating frequency is not necessarily a straightforward solution to achieve a shorter response time of EnKF because increasing the updating frequency leads to a faster decrease of ensemble variance especially in periods where the river bed is stable. This leads to some overconfidence of model parameters which influences the update of parameters negatively when there is a change in river bed conditions. On the contrary, when covariance inflation is used in the EnKF updating scheme a positive effect of updating frequency on the adaptation time of L could be found due to the regulation of ensemble variance by this method. However, even with a higher updating frequency of L together with covariance inflation there is a certain time lag until EnKF responds to instantaneous changes of the river bed. Such instantaneous changes are likely to occur in reality, e.g., as a consequence of flooding events within a typical time period from hours to days. When such changes in the river bed are persistent over a certain time period, EnKF will be able to adapt model parameters gradually within several assimilation cycles (depending on the settings for meta parameters such as updating frequency or damping factor). This means

that predictions of states will also improve step by step until EnKF adapts to the new parameter values. However, when there are very frequent fluctuations in river bed properties that are faster than the assimilation frequency or even the time step of the model, EnKF will only capture the effective changes of the river bed and smooth them temporally. For such changes one would have to increase the temporal resolution of the forward model in order to reduce the effective response time of EnKF but this would probably also require a higher temporal resolution of measurements which is not possible in many cases.

[57] The performance of EnKF with respect to seasonal variation of L caused by temperature changes of the river showed that the ensemble means of L for three of the five leakage zones principally followed the trend of the reference run although the whole magnitude of the change was not reached through updates with EnKF. For the other two zones changes of the ensemble mean of L were observable but did not correspond well to the evolution of L values of the reference run. The absolute changes of L were far lower for this scenario compared to the flooding and sedimentation cases. As a consequence, the errors at observation points as well as the correlations between the hydraulic head data at the observation points and the zonal leakage coefficients were lower, which might have led to a lower adaptation for two of the zones. Nevertheless, the results showed that even for small changes of L a correction with EnKF is principally possible.

[58] The results for the influence of spatial patterns on the update with EnKF showed that EnKF is also able to detect changes in L that only occur at a certain location of the river reach. This might be relevant when the flow regime and thus the sedimentation/scouring regime is not homogeneous within the river, e.g., due to dams or weirs or due to a meandering of the river. However, a prerequisite that allows an optimal spatial update of L with EnKF in a real-world case is that the position of leakage zones corresponds to the sediment dynamics in the river bed. Certainly, alternative parameterization methods like pilot points laid out over stochastic fields of L are an interesting alternative which was not investigated in the context of this study.

[59] EnKF was also able to correct for a bias in the initial ensemble which is important for a real-world scenario because it is often difficult to achieve a good initial guess of L values that provide a good agreement between simulated and measured states due to their high variability in natural settings. However, a correction of ensemble bias will require a sufficient initial ensemble spread of L.

[60] Several different other tests were performed for evaluating the performance of EnKF for updating L. Our ensemble size was found not to be too small, because results for an ensemble twice as large were not better. The damping parameter used was according the suggestions of *Hendricks Franssen and Kinzelbach* [2008] and increasing it gave worse results except for the simulations where adaptive covariance inflation was used.

[61] The sensitivity toward the amount of observation points showed that the ability of EnKF to correct for the time-varying L values generally increased with an increasing number of observation points. However, the results also showed that even with a low number of observation points (e.g., 5 or 10) the time-varying L values of EnKF could be reproduced quite well (150 days after the leakage jump the RMSE for L was reduced 50% for 10 observations and 20% for 5 observations compared to unconditional simulations). This may also be important for real-world applications because usually the amount of available time series of head data for a particular site is rather limited. For our study site the utilization of 100 observation points for the base case scenarios was reasonable because for the real world case 87 piezometer data are available for this site on a daily basis. A doubling of the number of observation points from 100 to 200 did not significantly increase the performance of EnKF. This might be related to the fact that the simulations with 100 observation points already had a relatively high information content, which was sufficient for the observed adaptations. Thus, the additional 100 observations points possibly only contained redundant information.

[62] In general, the calibration of time-dependent L with EnKF using a limited number of piezometers is possible, with the limitations indicated before. We believe that this is the first work where a systematic approach to calibrate time-dependent L was proposed, carefully tested, and shown to be feasible. A point of criticism could be that this approach only adapts L with help of indirect observations, without trying to predict the changes of L directly. However, both direct observations of modifications of L and deterministic prediction of changes of L are difficult and not possible at large scales. If indirect methods can reliably detect changes of L, this will provide new information at larger scales that can be used to better understand the mechanisms behind the changes of L. An additional limitation of the methodology followed in this paper is that L is updated, and not the two parameters which constitute L, the river bed thickness and river bed hydraulic conductivity. Although it would have been desirable to distinguish between the two parameters, this was beyond the scope of this paper:

[63] 1. Measurement data most probably do not allow for differentiating between changes in river bed thickness and river bed hydraulic conductivity.

[64] 2. We would need a very high resolution modeling of the river bed (on the cm scale) with the need to include many more grid cells (now we already have nearly 100,000 nodes and 173,599 elements).

[65] 3. In order to truly represent the dynamics of the river bed, the modeling grid should be adaptive and allowed to change over time. Such an approach is challenging in the context of forward model runs, but in the context of inverse modeling/data assimilation not yet feasible.

[66] It is of course always a question whether the results found in this study are related to specific conditions at this site. It is possible that the groundwater management activities (pumping, artificial recharge) provide additional information which help to constrain the estimation of L. Some of the simulation scenarios were also repeated without these management activities, but results were very similar.

7. Conclusions

[67] In this study we investigated to what extent the ensemble Kalman filter is able to correct states and parameters of a groundwater model for temporal changes in the hydraulic properties of a river bed. For this purpose we

created different synthetic scenarios in which the river bed conductivities followed certain temporal patterns that should imitate natural river bed dynamics. Calculations were based on a 3-D model of the Limmat aquifer (Zurich) and the updating procedure for the model states and parameters with the ensemble Kalman filter was done with hydraulic head data of the corresponding synthetic reference simulations.

[68] Results for the different scenarios indicate the principal capability of EnKF to account for changes in river bed conductivity. This was shown for different types of major changes of the river bed (i.e., erosion of the river bed due to a flooding event, sedimentation, and a combined scenario) in which EnKF correctly adjusted L values of the ensemble toward the reference values with a good overall performance at the end of the simulations. Also seasonal changes with smaller fluctuations of L related to the temperature dependency of L could in part be compensated by EnKF. Furthermore, EnKF was able to handle and correct for different types of uncertainty in the assimilation process (uncertain hydraulic conductivities of the aquifer, biased initial ensemble). One drawback is the relatively long adaptation time that is needed by EnKF to adjust to new L values. In our simulations it took about 150 days until EnKF corrected the ensemble for an instantaneous change in L of 1 log unit. For modifications of L with a longer time duration like a sedimentation event, the delayed response of EnKF also was observed. For real-world applications it will depend on the time scale of river bed dynamics whether EnKF reasonably catches changes in L. Frequent changes in river bed characteristics will only in part be captured by data assimilation if L is not updated very frequently, which means that EnKF just adjusts L for effective changes within a given time period. However, experiments with an adaptive covariance inflation approach suggest that the performance of EnKF to capture time-variant model parameters can be generally improved by this method as it reduced the total adaptation time and increased the overall accuracy of the parameter update.

[69] In summary, from a practical point of view the use of data assimilation with EnKF seems to be a promising way to account for changes of river sediments in real time models because of its capability to account for different changes of the river bed even if they have a low magnitude and because it is also able to handle different sources of uncertainty within the modeling process. Also the sensitivity analysis with respect to the numbers of observation points underpins the usability of EnKF for real time models because it showed that even with a low number of observations, an often encountered situation in practice, it is possible to capture changes in river bed conductivities.

Appendix A: Algorithm for Adaptive Covariance Inflation

[70] In accordance with *Anderson* [2007] a Bayesian updating scheme is used to calculate the inflation factor λ for each assimilation step. In this updating scheme the simulated ensemble estimate at observation points y, the measurement at observation points y^0, and their respective variances σ_p^2 and σ_0^2 are used to estimate a new value of λ in order to correct the model ensemble for effects of filter divergence. The basic updating equation for the assimilation time step $t_{a,i}$ can be stated as

$$p(\lambda, t_{a,i}) \sim p(y^0|\lambda)\, p(\lambda, t_{a,i-1}), \tag{A1}$$

where $p(\lambda, t_{a,i})$ is the posterior probability distribution of λ, $p(\lambda, t_{a,i-1})$ is the prior probability distribution, and $p(y^0|\lambda)$ is a likelihood term that describes the probability that y^0 is observed given a certain λ.

[71] The prior distribution in equation (A1) is assumed to follow a normal distribution:

$$p(\lambda, t_{a,i-1}) = N(\overline{\lambda_p}, \sigma^2_{\lambda,p}). \tag{A2}$$

[72] It is also assumed that the prior of λ is identical to the posterior from the last assimilation cycle. The likelihood term in equation (A1) is also expressed as a normal distribution:

$$p(y^0|\lambda) = (\sqrt{2\pi}\theta)^{-1} \exp(-D^2/2\theta^2), \tag{A3}$$

where D is the actual distance between simulated ensemble mean at observation points $\overline{y_p}$ and the measured value y^0:

$$D = |\overline{y_p} - y^0| \tag{A4}$$

and θ is a measure for the variability of this distance for a certain value of λ:

$$\theta = \sqrt{\lambda\sigma_p^2 + \sigma_0^2}. \tag{A5}$$

[73] Inserting the equations of the prior and the likelihood function into equation (A1) gives

$$p(\lambda, t_{a,i}) \sim (\sqrt{2\pi}\theta)^{-1} \exp(-D^2/2\theta^2) N(\overline{\lambda_p}, \sigma^2_{\lambda,p}). \tag{A6}$$

[74] For the determination of the inflation factor $\overline{\lambda}$ for the current assimilation cycle $\overline{\lambda}$ is set equal to the mode of equation (A6) which is found by differentiating the right-hand side of (A6) and setting it to zero. This results in a cubic equation of the form

$$x^3 - \left(\sigma_0^2 + \overline{\lambda}\sigma_p^2\right)x^2 + \frac{1}{2}\sigma_\lambda^2\sigma_p^4 x - \frac{1}{2}\sigma_\lambda^2\sigma_p^4 D^2 = 0, \tag{A7}$$

where $x = \theta^2$. Solving this equation with the cubic formula and substituting the results into equation (A5) gives the value of λ for the respective observation.

[75] The utilized algorithm for covariance inflation in our study can be summarized as follows:

1. Propagate the ensemble forward until the next observations y^0 are available.
2. Determine $\overline{y}$ and σ_p^2 from the ensemble.
3. For each observation of y^0 do the following steps:
   (a) Determine D with equation (A4).

(b) Solve equation (A7) for x and insert the result in equation (A5) to determine λ.

4. From the distribution of λ calculate $\overline{\lambda}$ and update the state-parameter vector x according to equation (13).

[76] In our implementation of the algorithm we set the variance of the inflation factor σ_{λ}^2 to a constant value of 0.25 instead of using a model for the evolution of this parameter as it was done by *Anderson* [2007]. This was also suggested as an algorithmic variant by *Anderson* [2009]. Additionally, we put a constraint on the values of $\overline{\lambda}$ by setting values of less than 1, which could occur when the filter is very confident (i.e., low residuals, to a value of 1).

[77] **Acknowledgments.** We would like to acknowledge Water Works Zurich for providing the management data of the study site, Uli Kuhlmann from TK Consult (Zurich) for technical assistance with the modeling software EnKF3d-SPRING, and Jülich Supercomputing Center for granting the required computation time on JUROPA. We also would like to thank three anonymous reviewers and the Associate Editor for their helpful comments.

References

Alcolea, A., J. Carrera, and A. Medina (2006), Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Adv. Water Resour.*, *29*(11), 1678–1689, doi:10.1016/j.advwatres.2005.12.009.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration: Guidelines for computing crop water requirements, *FAO Irrig. Drainage Pap.* 56, Food and Agric. Organ., Rome.

Anderson, J. L. (2007), An adaptive covariance inflation error correction algorithm for ensemble filters, *Tellus A*, *59*(2), 210–224, doi:10.1111/j.1600-0870.2006.00216.x.

Anderson, J. L. (2009), Spatially and temporally varying adaptive covariance inflation for ensemble filters, *Tellus A*, *61*(1), 72–83, doi:10.1111/j.1600-0870.2008.00361.x.

Bauser, G., H. J. H. Franssen, H. P. Kaiser, U. Kuhlmann, F. Stauffer, and W. Kinzelbach (2010), Real-time management of an urban groundwater well field threatened by pollution, *Environ. Sci. Technol.*, *44*(17), 6802–6807, doi:10.1021/es100648j.

Blaschke, A. P., K. H. Steiner, R. Schmalfuss, D. Gutknecht, and D. Sengschmitt (2003), Clogging processes in hyporheic interstices of an impounded river, the Danube at Vienna, Austria, *Int. Rev. Hydrobiol.*, *88*(3–4), 397–413, doi:10.1002/iroh.200390034.

Brunke, M., and T. Gonser (1997), The ecological significance of exchange processes between rivers and groundwater, *Freshwater Biol.*, *37*(1), 1–33, doi:10.1046/j.1365-2427.1997.00143.x.

Burgers, G., P. J. van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*(6), 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Chen, Y., and D. Zhang (2006), Data assimilation for transient flow in geologic formations via ensemble Kalman filter, *Adv. Water Resour.*, *29*(8), 1107–1122, doi:10.1016/j.advwatres.2005.09.007.

Conant, B. (2004), Delineating and quantifying ground water discharge zones using streambed temperatures, *Ground Water*, *42*(2), 243–257, doi:10.1111/j.1745-6584.2004.tb02671.x.

Delta h Ingenieurgesellschaft mbH (2006), *Spring 3.3, Software, Witten*, Germany.

de Marsily, G. (1978), De l'identificaton des systèmes hydrogéologiques, Ph.D. thesis, Paris VI, Paris.

Doppler, T., H. J. Hendricks Franssen, H. P. Kaiser, U. Kuhlman, and F. Stauffer (2007), Field evidence of a dynamic leakage coefficient for modelling river-aquifer interactions, *J. Hydrol.*, *347*(1–2), 177–187, doi:10.1016/j.jhydrot.2007.09.017.

Engeler, I., H. J. Hendricks Franssen, R. Müller, and F. Stauffer (2011), The importance of coupled modelling of variably saturated groundwater flow-heat transport for assessing river-aquifer interactions, *J. Hydrol.*, *397*(3–4), 295–305, doi:10.1016/j.jhydrol.2010.12.007.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*(C5), 10,143–10,162, doi:10.1029/94JC00572.

Fleckenstein, J. H., R. G. Niswonger, and G. E. Fogg (2006), River-aquifer interactions, geologic heterogeneity, and low-flow management, *Ground Water*, *44*(6), 837–852, doi:10.1111/j.1745-6584.2006.00190.x.

Frei, S., J. H. Fleckenstein, S. J. Kollet, and R. M. Maxwell (2009), Patterns and dynamics of river-aquifer exchange with variably-saturated flow using a fully-coupled model, *J. Hydrol.*, *375*(3–4), 383–393, doi:10.1016/j.jhydrol.2009.06.038.

Genereux, D. P., S. Leahy, H. Mitasova, C. D. Kennedy, and D. R. Corbett (2008), Spatial and temporal variability of streambed hydraulic conductivity in West Bear Creek, North Carolina, USA, *J. Hydrol.*, *358*(3–4), 332–353, doi:10.1016/j.jhydrol.2008.06.017.

Gómez-Hernández, J., and A. Journel (1993), Joint sequential simulation of multi-Gaussian fields, in *Geostatistics Tróia '92*, Vol. 1, edited by A. Soares, pp. 85–94, Kluwer Academic, New York.

Hamill, T. M., J. S. Whitaker, and C. Snyder (2001), Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Mon. Weather Rev.*, *129*(11), 2776–2790.

Hatch, C. E., A. T. Fisher, C. R. Ruehl, and G. Stemler (2010), Spatial and temporal variations in streambed hydraulic conductivity quantified with time-series thermal methods, *J. Hydrol.*, *389*(3–4), 276–288, doi:10.1016/j.jhydrol.2010.05.046.

Hendricks Franssen, H. J., and W. Kinzelbach (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, doi:10.1029/2007WR006505.

Hendricks Franssen, H. J., H. P. Kaiser, U. Kuhlmann, G. Bauser, F. Stauffer, R. Mueller, and W. Kinzelbach (2011), Operational real-time modeling with EnKF of variably saturated subsurface flow including stream-aquifer interaction and parameter updating, *Water Resour. Res.*, *47*, W02532, doi:10.1029/2010WR009480.

Huber, E., H. Hendricks-Franssen, H. Kaiser, and F. Stauffer (2011), 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.

Kalbus, E., F. Reinstorf, and M. Schirmer (2006), Measuring methods for groundwater–surface water interactions: A review, *Hydrol. Earth Syst. Sci.*, *10*(6), 873–887.

Kalbus, E., C. Schmidt, J. W. Molson, F. Reinstorf, and M. Schirmer (2009), Influence of aquifer and streambed heterogeneity on the distribution of groundwater discharge, *Hydrol. Earth Syst. Sci.*, *13*(1), 69–77.

Käser, D. H., A. Binley, A. L. Heathwaite, and S. Krause (2009), Spatio-temporal variations of hyporheic flow in a riffle-step-pool sequence, *Hydrol. Processes*, *23*(15), 2138–2149, doi:10.1002/hyp.7317.

Krause, S., A. Bronstert, and E. Zehe (2007), Groundwater-surface water interactions in a north German lowland floodplain—Implications for the river discharge dynamics and riparian water balance, *J. Hydrol.*, *347*(3–4), 404–417, doi:10.1016/j.jhydrol.2007.09.028.

Muskat, M. (1937), *The Flow of Homogeneous Fluids through Porous Media*, McGraw-Hill, New York.

Mutiti, S., and J. Levy (2010), Using temperature modeling to investigate the temporal variability of riverbed hydraulic conductivity during storm events, *J. Hydrol.*, *388*(3–4), 321–334, doi:10.1016/j.jenvman.2011.03.017.

Nowak, W. (2009), Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator, *Water Resour. Res.*, *45*, W04431, doi:10.1029/2008WR007328.

Rehg, K. J., A. I. Packman, and J. H. Ren (2005), Effects of suspended sediment characteristics and bed sediment transport on streambed clogging, *Hydrol. Processes*, *19*(2), 413–427, doi:10.1002/hyp.5540.

Renard, P., G. Le Loc'h, E. Ledoux, G. de Marsily, and R. Mackay (2000), A fast algorithm for the estimation of the equivalent hydraulic conductivity of heterogeneous media, *Water Resour. Res.*, *36*(12), 3567–3580, doi:10.1029/2000WR900203.

Rosenberry, D. O., and J. Pitlick (2009a), Local-scale variability of seepage and hydraulic conductivity in a shallow gravel-bed river, *Hydrol. Processes*, *23*(23), 3306–3318, doi:10.1002/hyp.7433.

Rosenberry, D. O., and J. Pitlick (2009b), Effects of sediment transport and seepage direction on hydraulic properties at the sediment-water interface of hyporheic settings, *J. Hydrol.*, *373*(3–4), 377–391, doi:10.1016/j.jhydrol.2009.04.030.

Schälchli, U. (1992), The clogging of coarse gravel river beds by fine sediment, *Hydrobiologia*, *235*, 189–197, doi:10.1007/BF00026211.

Schubert, J. (2002), Hydraulic aspects of riverbank filtration-field studies, *J. Hydrol.*, *266*(3–4), 145–161, doi:10.1016/S0022-1694(02)00159-2.

Sophocleous, M. (2002), Interactions between groundwater and surface water: The state of the science, *Hydrogeol. J.*, *10*(1), 52–67, doi:10.1007/s10040-002-0204-x.

Sun, A. Y., A. P. Morris, and S. Mohanty (2009), Comparison of deterministic ensemble Kalman filters for assimilating hydrogeological data, *Adv. Water Resour.*, *32*(2), 280–292, doi:10.1016/j.advwatres.2008.11.006.

Woessner, W. W. (2000), Stream and fluvial plain ground water interactions: Rescaling hydrogeologic thought, *Ground Water*, *38*(3), 423–429, doi:10.1111/j.1745-6584.2000.tb00228.x.

Zhang, Y., S. Hubbard, and S. Finsterle (2011), Factors governing sustainable groundwater pumping near a river, *Ground Water*, *49*(3), 432–444, doi:10.1111/j.1745-6584.2010.00743.x.