Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating

H. J. Hendricks Franssen,[1,2] H. P. Kaiser,[3] U. Kuhlmann,[4] G. Bauser,[1] F. Stauffer,[1] R. Müller,[3] and W. Kinzelbach[1]

Received 27 April 2010; revised 6 December 2010; accepted 30 December 2010; published 23 February 2011.

[1] Urban groundwater is frequently contaminated, and the exact location of the pollution spots is often unknown. Intelligent monitoring of the temporal variations in groundwater flow in such an area assists in selectively extracting groundwater of drinking water quality. Here an example from the city of Zurich (Switzerland) is shown. The monitoring strategy consists of using the ensemble Kalman filter (EnKF) for optimally combining online observations and online models for the real-time characterization of groundwater flow. We conducted numerical simulation experiments for the period January 2004 to December 2007 with a 3-D finite element model for variably saturated groundwater flow. It was found that the daily assimilation of piezometric head data with EnKF results in a better characterization of piezometric heads than does a model which is inversely calibrated with historical data but not updated in real time. The positive impact of model updating with observations can still be observed 10 days after the update. These simulations also suggest that parameters (hydraulic conductivity and leakage) are successfully updated: 1 and 10 day piezometric head predictions are better with than without updating of parameters. Additional experiments with a synthetic model for the same site, in which the only difference is that certain parameter values are selected as the unknown "true" conditions, show that EnKF also successfully updates unknown parameters. However, this is only the case if spatially distributed hydraulic conductivities and leakage coefficients are jointly updated and if a damping parameter is used. The mean absolute error of estimated log leakage coefficients decreased by up to 63%; for log hydraulic conductivity a decrease of up to 27% was observed. From January 2009 the method has been operational at the Water Works Zurich and showed a remarkable performance until present (October 2010).

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

1. Introduction

[2] Groundwater resources supply drinking water for numerous people globally. The quantity and quality of groundwater resources are threatened by overpumping, salinization, and various types of contamination. The groundwater below large cities is often contaminated because of leakages from petrol stations and zones of industrial activities. Therefore, it may seem unsuitable to pump groundwater for drinking water purposes in or near large cities. On the other hand, it is attractive to pump drinking water close to large cities because this limits the transportation costs and large cities are often located close to rivers which recharge an aquifer. If groundwater is pumped in an urban area with multiple pollution sources, the operation requires a sophisticated quality control. In this paper we present a methodology which has been put into practice for the pumping of drinking water in the city of Zurich (Switzerland). The methodology is of interest for a large class of cases involving sites that are relatively close to a larger contamination that is not remediated. It is also of interest for the selective withdrawal of colder winter infiltration water if summer temperatures become too high.

[3] We propose an optimal monitoring strategy for the groundwater flow and solute transport at the site through both online measurements and online models. The groundwater flow and transport models are calibrated with historical information. A monitoring network is developed that measures, in real time, piezometric heads (and also concentrations of solutes or temperature). These data are sent to a central server, where piezometric head data are used to update the model predictions. Data and models can be

[1] Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland.
[2] Now at Agrosphere, IBG-3, Forschungszentrum Jülich GmbH, Jülich, Germany.
[3] Water Supply of Zurich, Zurich, Switzerland.
[4] TK Consult, Zurich, Switzerland.

Copyright 2011 by the American Geophysical Union.
0043-1397/11/2010WR009480

combined in an optimal way by data assimilation methods. In this work, the ensemble Kalman filter (EnKF) [*Evensen*, 1994; *Burgers et al.*, 1998] is used to update the models in real time with observations. EnKF provides an optimal estimate (for linear Gaussian systems) of the current spatial distribution of piezometric heads, concentrations, and temperatures and is therefore very well suited for an optimal characterization of the risk of pumping contaminated water. In this paper we focus on the impact of the assimilation of hydraulic head data on the distribution of piezometric heads.

[4] In this study, the optimal real-time characterization of the groundwater flow situation was used to optimize the management of the water works' well field. The management of the site was adapted if the groundwater flow vector pointed from the city center (with potentially contaminated sites) to the Hardhof area (containing the drinking water well field). The management was adapted for the next few days using fuzzy control techniques and genetic algorithms. The adapted management resulted in a groundwater flow vector that was less likely to introduce potentially contaminated city water into the pumping wells. However, the real-time management is not the subject of this paper. It is described by *Bauser et al.* [2010].

[5] The material of this paper is novel in the sense that EnKF, including parameter optimization, is applied here for a more complex and realistic subsurface flow situation than presented in the literature up to now. This more complex subsurface flow situation includes an unconfined aquifer, the unsaturated zone, and, particularly, river-aquifer interaction. It is shown that information from fluctuating river stages, which strongly influence the groundwater heads, yield important information on river bed and aquifer properties. An additional complicating aspect is that the uncertainty in the values of two different parameters (hydraulic conductivity and leakage coefficients) is roughly equally important in this study. This requires updating (calibration) both uncertain parameter values, which normally is difficult as they are highly correlated and therefore poorly identifiable. An additional important aspect is that we also test the EnKF for a real-world case (Zurich, Switzerland), which presents additional challenges compared to a synthetic study, as new sources of uncertainty might play a role and conceptual model errors cannot be excluded. Moreover, EnKF not only was applied to a real-world case study but was also made operational for the same area. Since January 2009, the subsurface flow situation is calculated in real time, and it is used to adapt the management of the groundwater well field (pumping wells, artificial recharge basins, and artificial recharge wells) in real time. According to our knowledge, this is one of the few cases where stochastic subsurface hydrology is put into practice and the first case where it is made operational as basis for the improved management of a well field. The regulators needed to be convinced of the benefit of this methodology, which includes reducing the risk of pumping contaminated water, while at the same time requiring considerable investment in infrastructure (online sensing equipment, software development, and computing infrastructure). See *Renard* [2007] for an overview of problems associated with putting into practice stochastic groundwater hydrology.

[6] This paper first explains the EnKF for the real-time updating of a groundwater flow model using observations. Then the study area, the groundwater flow model, and the calibration of the model with historical data are introduced. Finally, the results are presented and discussed.

2. Real-Time Modeling With EnKF

[7] The modeling of groundwater flow and solute transport is associated with large uncertainties. These uncertainties may be related to a possible misspecification of the conceptual model (for example, the position of the aquifer boundaries and aquifer bottom and the presence of springs), the unknown external forcing of the aquifer (e.g., recharge and boundary conditions), and, particularly, parameter uncertainties or all of those at once. The most uncertain parameter values of groundwater flow models in humid climates are, in general, the spatially distributed hydraulic conductivities. Hydraulic conductivity is very uncertain due its large spatial variability, the scarcity of measurements, and measurement errors. Model calibration, or inverse modeling, helps to reduce the uncertainty of groundwater flow models by adapting parameters (and possibly boundary conditions, initial conditions, and forcing terms as well) to fit steady state hydraulic heads [e.g., *Kitanidis and Vomvoris*, 1983], historical time series of hydraulic heads [e.g., *Carrera and Neuman*, 1986], or both hydraulic heads and concentration data [e.g., *Medina and Carrera*, 1996]. Inverse methods often focus on obtaining a single best estimate, and the uncertainty of that estimate can be characterized by the posterior covariance matrix obtained from a linearized analysis. However, this uncertainty analysis relies on a Gaussian assumption and underestimates the true variance [*Carrera and Neuman*, 1986]. This and the fact that the calibration process does not have a unique solution were the main motivations to develop methods that generate multiple equally likely solutions to the groundwater inverse problem. This so-called Monte Carlo (MC) type inverse modeling was first formulated for 2-D steady state groundwater flow [*Sahuquillo et al.*, 1992; *RamaRao et al.*, 1995] and later for 2-D transient groundwater flow with the joint calibration of spatially variable transmissivity and storativity fields [*Hendricks Franssen et al.*, 1999], 3-D flow in fractured media [*Gómez-Hernández et al.*, 2001; *LaVenue and de Marsily*, 2001], and coupled groundwater flow and solute transport [*Hendricks Franssen et al.*, 2003; *Wen et al.*, 2003]. All the mentioned MC-type inverse modeling approaches calibrate a large number of spatially distributed fields of hydraulic conductivity (and possibly also other parameters) with derivative-based nonlinear optimization methods, using the adjoint state method to calculate the gradient of the objective function efficiently. The dimensionality of the optimization problem (and therefore also of the gradient vector) is reduced with parameterization techniques that use master blocks or pilot points [*de Marsily*, 1978]. MC-type inverse modeling is not limited to formations with a mild spatial variability of hydraulic conductivity and is very well suited for the characterization of uncertainty. A comparison study showed that MC-type inverse modeling methods outperformed other inverse modeling methods [*Hendricks Franssen et al.*, 2009]. However, some methods like the inverse moment equations method [*Hernandez et al.*, 2003] or the regularized pilot points method in its conditional estimation variant [*Alcolea et al.*,

2006] yielded almost as good results as the MC-type inverse methods but with less CPU time. In principle, MC-type inverse modeling would be suited for the real-time uncertainty characterization, recalibrating the model each time new measurements become available. However, the main limitations of MC- type inverse modeling for real-time modeling are (1) recalibration with all historical data is very CPU intensive, (2) an optimal characterization of the actual conditions is not guaranteed with this approach (the optimum with MC-type inverse modeling being balanced over a historical set of observations), and (3) the application to a physically complex systems (for which an adjoint model has to be formulated) with many different sources of uncertainty is difficult. Alternatives are nonderivative-based MC-type inverse methods like Markov Chain Monte Carlo methods (MCMC) or the EnKF. Currently, MCMC methods are still relatively slow for the inverse modeling of groundwater flow [*Oliver et al.*, 1997; *Fu and Gómez-Hernández*, 2009]. The EnKF [*Evensen*, 1994; *Burgers et al.*, 1998] is a very fast method for the sequential updating of the model states each time new measurements become available. The EnKF was reformulated for subsurface hydrology applications so that with an augmented state vector approach both states and parameters can be updated [*Chen and Zhang*, 2006; *Hendricks Franssen and Kinzelbach*, 2008; *Liu et al.*, 2008; *Nowak*, 2009], an approach that was introduced slightly earlier in petroleum engineering [e.g., *Naevdal et al.*, 2003; *Wen and Chen*, 2006] and surface hydrology [*Moradkhani et al.*, 2005; *Vrugt et al.*, 2005]. For a synthetic study, *Hendricks Franssen and Kinzelbach* [2009] found that EnKF yielded parameter estimates that had approximately the same error as MC-type inverse modeling parameter estimates but with a factor of 80 less CPU time. This was the case for both mildly and strongly heterogeneous transmissivity fields, although EnKF could have been expected to perform worse for more strongly nonlinear problems, as the method relies on Gaussian statistics.

[8] After indicating why we implemented EnKF for the real-time updating of states and parameters, we now present the formulation of EnKF, tailored to our specific problem. The governing equation is the equation for 3-D unsaturated-saturated transient groundwater flow including interaction with rivers [e.g., *Bear*, 1979]:

$$\frac{\partial}{\partial t}[\rho n S(p)] - \nabla\left(\frac{\rho \mathbf{k} k_r(S)}{\mu}(\nabla p + \rho g \nabla z)\right) = q\,, \qquad (1)$$

where S is saturation (dimensionless), ρ is density [$M\,L^{-3}$], n is porosity (dimensionless), p is pressure [$M\,L^{-1}\,T^{-2}$], t is time [T], $\mathbf{k}$ is permeability [L^2], k_r is relative permeability (dimensionless), μ is dynamic viscosity [$M\,L^{-1}\,T^{-1}$], g is the gravitational acceleration [$L\,T^{-2}$], z is the elevation with respect to a reference [L], q represents sinks (abstractions) and sources (recharge) [$M\,L^{-3}\,T^{-1}$], and the Nabla operator [L^{-1}] is three-dimensional, referring to spatial coordinates $\mathbf{x}$. In this paper, the van Genuchten parameterization was applied for modeling the saturation as function of pressure [*van Genuchten*, 1980]. The boundary conditions that are used to solve equation (1) also include the leakage coefficient r [T^{-1}]. The leakage coefficient is given by $Q/(A(h_{\mathrm{river}} - h_{\mathrm{gw}}))$, where Q is the exchange flux between river and groundwater [$L^3\,T^{-1}$], A is the surface for the exchange flux [L^2], h_{river} is the river stage [L], and h_{gw} is the groundwater level [L].

[9] Equation (1) is solved using the finite element method. We will refer to the numerical model as M. The hydraulic conductivity $\mathbf{K_c}$ ($\mathbf{K_c} = \mathbf{k}\rho g \mu^{-1}$) and the leakage factor r are stochastic parameters in equation (1), and a large number of stochastic realizations of these parameters are generated. In section 4.2, more details are given on the stochastic generation. The ensemble Kalman filter scheme proceeds in the following steps:

$$\mathbf{x}^0_{i,h} = M\left(\mathbf{x}^-_{i,h}\right), \qquad (2)$$

where i refers to a stochastic realization ($i = 1, \ldots, P$) and $\mathbf{x}_{i,h}$ is part of the vector $\mathbf{x}_i$ and contains states from the previous time step (superscript minus) or the actual time step (superscript 0). An augmented state vector approach is used to update states and parameters jointly. The augmented vector is

$$\mathbf{x_i} = \begin{Bmatrix} \mathbf{x}_{i,h} \\ \mathbf{x}_{i,Y} \\ \mathbf{x}_{i,L} \end{Bmatrix}, \qquad (3)$$

where the subscript Y refers to $\log_{10}$ hydraulic conductivities and L refers to $\log_{10}$ leakage coefficients. The vector $\mathbf{x}_i$ is of dimension $N + E + N_l$, where N is the number of nodes, E is the number of elements, and N_l is the number of leakage zones. The covariance matrix of dimension $((N + E + N_l) \times (N + E + N_l))$ is estimated from the series of stochastic realizations. For the first time step, these stochastic realizations are unconditional or conditioned only on measurements of Y and L. For subsequent time steps the stochastic realizations are also conditional on state information. The covariance matrix is given by

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{hh} & \mathbf{C}_{Yh} & \mathbf{C}_{Lh} \\ \mathbf{C}_{hY} & \mathbf{C}_{YY} & \mathbf{C}_{LY} \\ \mathbf{C}_{hL} & \mathbf{C}_{YL} & \mathbf{C}_{LL} \end{bmatrix} \qquad (4)$$

where the subscript hh refers to covariances between modeled hydraulic heads at two locations (grid nodes), hY refers to cross covariances between a modeled hydraulic head value at one grid node and a log hydraulic conductivity value at an element, hL refers to cross covariances between modeled nodal hydraulic head and the log leakage coefficient for a zone, YL refers to cross covariances between log hydraulic conductivity and log leakage coefficient, YY refers to covariances between log hydraulic conductivities at two locations, and LL refers to covariances between leakage coefficients for two zones. The observations for the current time step are stored in the vector $\mathbf{y}^0$ of dimension n. The data are perturbed (following *Burgers et al.* [1998]) according to

$$\mathbf{y}^0_{\mathbf{i}} = \mathbf{y}^0 + \boldsymbol{\varepsilon}^0_{\mathbf{i}}\,, \qquad (5)$$

where $\boldsymbol{\varepsilon}$ is a vector of random numbers, drawn from a normal distribution with expectation zero and standard deviation equal to the expected measurement error standard deviation.

[10] The model predictions (equation (2)) and the observations (5) are combined to yield an updated ensemble of states (hydraulic heads) and parameters (log hydraulic conductivities and log leakage coefficients), according to

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

where $\mathbf{x}_i^+$ is the updated vector containing updated states (hydraulic heads) and parameters (Y and L) for stochastic realization i, $\boldsymbol{\alpha}$ is a matrix filled with zeros, except for the diagonal elements, which contain damping factors that take a value between 0 and 1 (for the damping factors related to the states the entries are always equal to 1; i.e., no damping is used), and $\mathbf{H}$ is a linear operator ($n \times (N + E + N_l)$) that maps the observations to the state space. The damping factor reduces the correcting influence of the head measurements on updating the log hydraulic conductivity field and the log leakage coefficients. In a previous study by *Hendricks Franssen and Kinzelbach* [2008], damping the perturbation was found to give improved results by reducing filter inbreeding problems. $\mathbf{K}$ is the Kalman gain matrix ($(N + E + N_l) \times n$):

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_h \\ \mathbf{K}_Y \\ \mathbf{K}_L \end{bmatrix}, \tag{7}$$

where $\mathbf{K}_h$ is related to the states (hydraulic heads), $\mathbf{K}_Y$ is related to the log hydraulic conductivities, and $\mathbf{K}_L$ is related to the log leakage coefficients. $\mathbf{K}$ is obtained from

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

where $\mathbf{C}^0$ is the covariance matrix for the actual time step (estimated from the ensemble of stochastic realizations) and $\mathbf{R}^0$ ($n \times n$) is the measurement error covariance matrix for the actual time step, which is estimated a priori.

[11] The ensemble of updated vectors $\mathbf{x}_i$ (hydraulic heads, log hydraulic conductivities, and log leakage coefficients) is the input of the groundwater flow model for the next time step. Equations (2)–(8) are applied each time new observations are available. In several simulation experiments that will be presented in sections 5.1.1 and 5.2.1, only the states are updated and not the parameters, which implies that the equations above are applied excluding log hydraulic conductivities and log leakage coefficients from the expressions.

3. Study Area

[12] The studied aquifer lies below parts of the city of Zurich and is mainly fed by the rivers Sihl and Limmat but also receives water from infiltrating precipitation and lateral inflow from hills. Figure 1 shows an overview of the situation. The river Sihl has, on average, a limited discharge rate (6.8 $m^3\ s^{-1}$ (Swiss Federal Office for the Environment (FOEN)), but it has elevated peaks as a response to intense rainfall. The river Sihl joins the river Limmat in the eastern part of the city. The river Limmat is the outflow of Lake Zurich and has an average discharge of 95.8 $m^3\ s^{-1}$ (FOEN). Two weirs on the river Limmat fall within the study area, with the Hoengg weir being located close to the groundwater well field Hardhof. The rivers Sihl and Limmat infiltrate into the groundwater, except for the downstream, western part, where the aquifer exfiltrates into the river Limmat. The Limmat can show considerable river stage fluctuations, and the aquifer response in most of the study area to these stage fluctuations provides important information about aquifer and river bed hydraulic properties [e.g., *Yeh et al.*, 2009]. No direct measurements of leakage were available, and the leakage coefficient was calibrated for five different river sections via the numerical model. Details will be provided in section 4.

[13] The recharge of the aquifer from rain is limited because of the generally sealed soil surface. Recharge from precipitation is calculated as the difference between precipitation and actual evapotranspiration for the nonsealed areas. Details of the calculations will be given in section 4. The aquifer receives lateral inflow from the hills at its northern boundary (Kaeferberg) and, particularly, at its southern boundary (Uetliberg). The amounts of lateral inflow are, however, not very large. These lateral inflows are calculated as functions of the recharge. The largest inflow occurs from the south (Uetliberg) and in particular close to river Sihl.

[14] The mean hydraulic conductivity of the aquifer is around 2×10^{-3} m s^{-1}. This value is obtained from averaging estimated $\mathbf{K_c}$ from small-scale pumping tests along transects of boreholes. The aquifer consists of coarse material, mainly sandy gravel, which has been deposited by the river Sihl and as glacial moraine [*Kempf et al.*, 1986]. The hydraulic conductivity is, in general, larger for the upper aquifer layers than for the lower aquifer layers. The heterogeneity of the hydraulic conductivity is high, and for small-scale measurements we have $\sigma_Y^2 = 0.51$. There is some evidence that the spatial distribution of $\log_{10}$ conductivity Y shows a complex spatial pattern consisting of small channels and lenses with high and low hydraulic conductivities caused by changing river courses. The aquifer storativity is assumed to be 0.15. The aquifer thickness is on average around 20 m, but in the eastern part of the study area reaches up to 70 m.

[15] Around 20% of the drinking water for the city of Zurich is pumped in the Hardhof area. Figure 2 gives an overview of the Hardhof area. The drinking water is pumped from four horizontal wells. In addition, 19 bank filtration wells along the river Limmat pump water which is used for artificial recharge and distributed over 12 infiltration wells and 3 recharge basins. These artificial recharge facilities are located in the southern part of the Hardhof area and are supposed to create a hydraulic barrier between the city center and the Hardhof area. Below the city center, diffuse pollution is present, which could reach the pumping wells if the abstraction rates are large. Tracer tests and additional analysis on the basis of electrical conductivities revealed that a considerable part (up to 30%) of the water pumped by two of the four horizontal wells originates from the city area.

4. Model Components and Online Data

4.1. Variably Saturated Groundwater Flow Model

[16] A three-dimensional finite element model for variably saturated groundwater flow was developed using the software SPRING [*Delta h, Ingenieurgesellschaft mbH*,

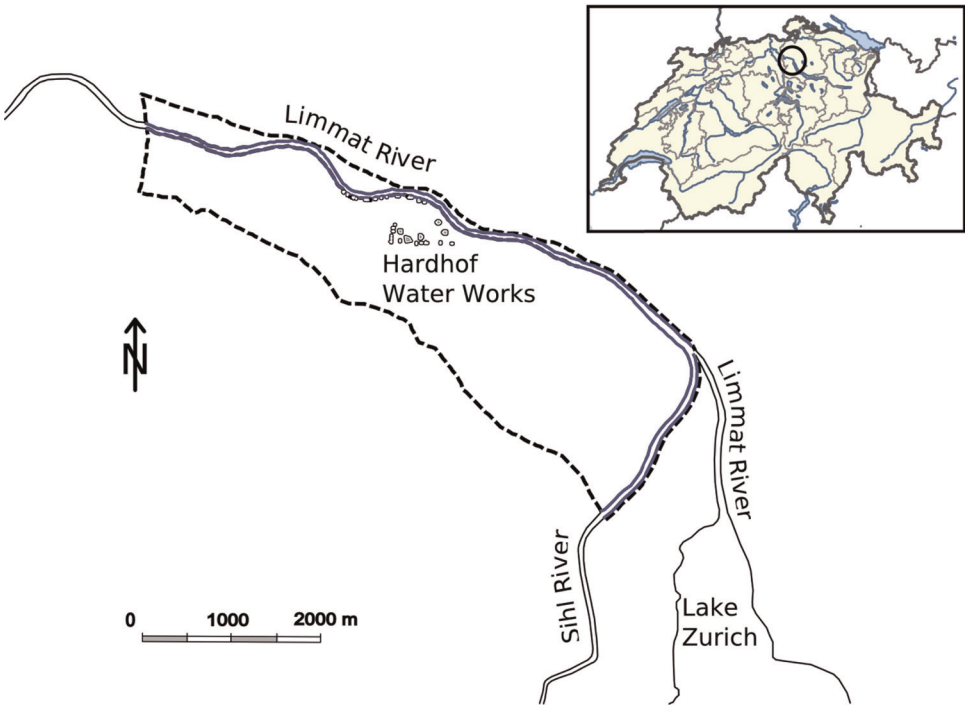

Figure 1. The inset shows Switzerland and the location of the study area (black circle). The study site (black dashed line) corresponds to the boundaries of the simulated domain in the upper Limmat valley in Zurich (Switzerland). Zurich is located at 47.22′ N and 8 32′ E. The Sihl and Limmat rivers are marked by the blue lines. The points in the middle part of the site represent the wells and infiltration basins of the Hardhof Water Works. From *Huber et al.* [2011].

2006]. The size of the triangular elements in the horizontal directions is typically 50 m, but it is 30 m in the Hardhof area and refined (down to 1 m) around wells and basins. The model domain is discretized between the soil surface and the aquifer bottom into layers of 1.6 m depth. The unsaturated zone (often with a depth of 5 m) is therefore discretized in a very coarse manner. However, we did not aim to calculate detailed soil moisture distributions. All we wanted was an approximately correct timing of aquifer recharge. The amount of groundwater recharge was calculated with an external program (see later in this section) and was applied to the top layer of the numerical simulation model. If the aquifer bottom lies below 40 m, this part of the aquifer is neglected in the model. Only a small part of the model domain has an aquifer bottom below 40 m, and this part of the aquifer is relatively far away from the main area of interest (the Hardhof area). The discretization results in 173,599 elements. The temporal discretization for the simulations with data assimilation was 24 h, with two iterations for each time step. One complete model run with 10 iterations showed negligible differences in the solution compared with the solution for two iterations. The model boundaries correspond to the natural boundaries, except for the western boundary. A prescribed head boundary condition is adopted for the western boundary, using the head measurement in a piezometer at this boundary.

[17] The rivers Sihl and Limmat are implemented in the model, and the river-aquifer interaction is modeled with the leakage concept. The rivers Sihl and Limmat are located on (or very close to) the northern and eastern boundaries of the simulation domain (see Figure 1). The leakage is modeled along two lines of nodal points which coincide with the river banks. It was found that the model results show little sensitivity with respect to a varying location of these lines [*Engeler et al.*, 2011]. The software FLORIS [*Scietec Flussmanagement GmbH*, 2000] was used to calculate transient river stages along the two rivers. The program numerically solves the Saint-Venant equations describing unsteady flow in open channels. The one-dimensional hydraulic model used river cross sections (including dam locations) together with measured daily average values of river discharge and river stages as input. These calculations resulted in daily averages of the river water level for each river discretization point over the period January 2004 to August 2005. The results were used to derive quadratic regression equations for each discretization point along the river in order to estimate the daily water levels as a function of the river discharge at the leakage nodes for the complete period September 2005 to December 2007. It is assumed that the river bed shape did not show major modifications over time. Pumping and infiltration rates of the wells and basins in the Hardhof area operated by Zurich Water Works were available on a daily basis.

[18] As explained, the recharge rate was calculated with a separate program and applied to the topsoil layer. Potential evapotranspiration (ET) was calculated according to the Penman-Monteith equation, using measurements from the meteorological station of Zurich-Affoltern. Actual ET was calculated with a soil water balance model, using the Food and Agricultural Organization method [*Allen et al.*,

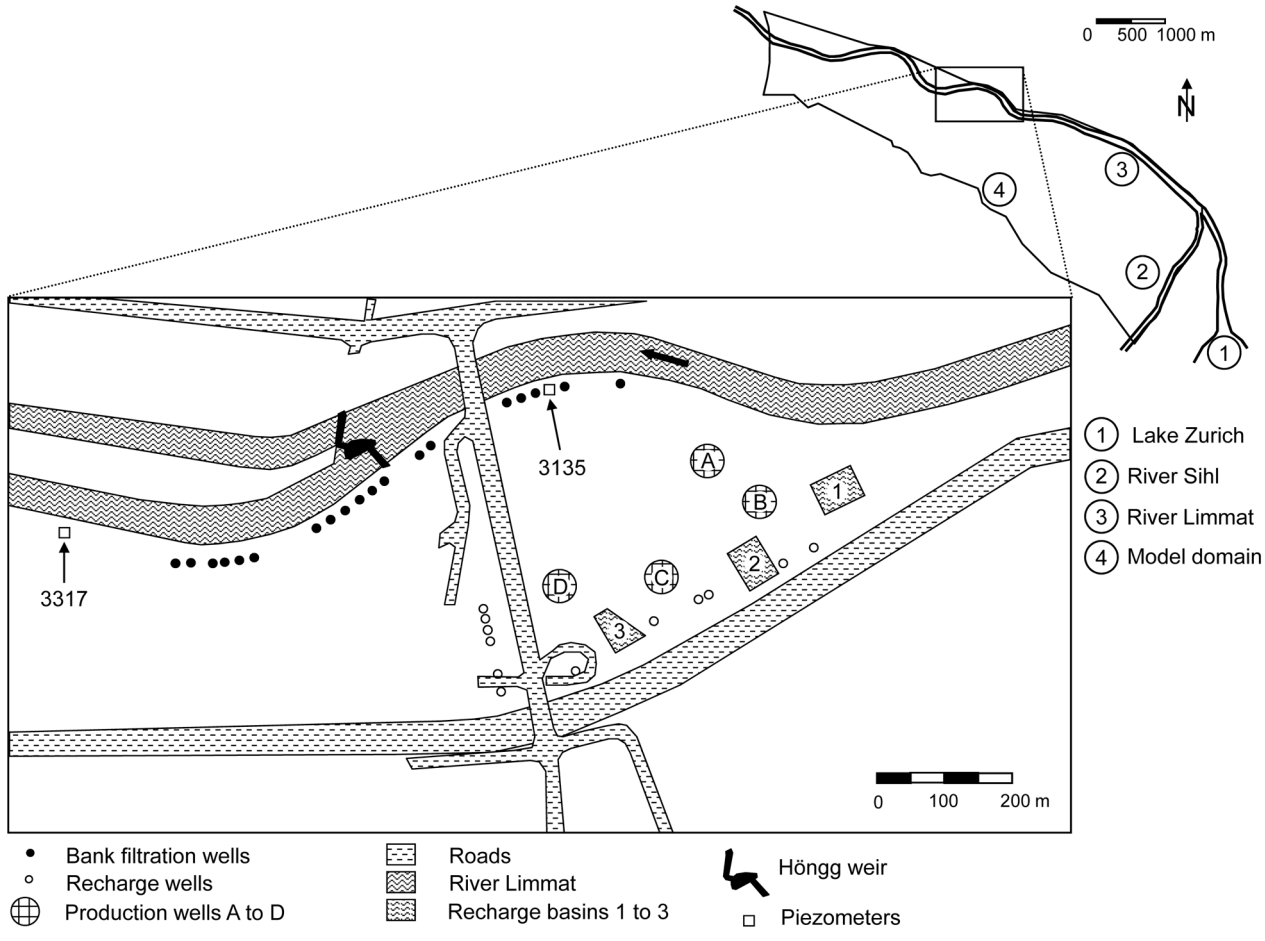

Figure 2. Model area and Hardhof site [from *Engeler et al.*, 2011].

1998]. This method calculates actual ET on the basis of potential ET, applying a reduction depending on the soil moisture content. Input values for soil and vegetation parameters were obtained from *Allen et al.* [1998]. Precipitation measurements (from Zurich-Affoltern) and the calculated actual evapotranspiration resulted in the potential net recharge rate. Only 15% of the calculated value recharged the aquifer in our model, as most of the surface is sealed and drains directly to the Limmat in the model area. The lateral inflow at the model boundaries was calculated as a function of the calculated recharge rates. It implies that the total amount of calculated lateral inflow is distributed over time according to the calculated recharge rate. The lateral inflow is largest on those days where the calculated recharge rate is largest. The lateral inflow originates mainly from precipitation on the hillslopes south and north of the model area. The overall amount of lateral inflow is small compared to the amount of infiltration from the rivers.

[19] The log hydraulic conductivity Y and the log leakage coefficient L were calibrated by transient inverse modeling, taking into account time series of hydraulic head measurements from 87 piezometers. Two calibration periods were chosen: June 2004 and July 2005. These periods were selected because they are characterized by very different hydrological conditions, including a flood, elevated pumping in the Hardhof area, and also typical mean conditions. The fact that the river stage showed large fluctuations over this time period provides important additional information on aquifer and river bed properties [e.g., *Yeh et al.*, 2009] because the groundwater level increases in a large part of the aquifer as a response to floods and subsides later when the river stage has decreased again. The fluctuations of the river stage can be considered as large-scale pumping tests that affect a considerable part of the aquifer. Inverse modeling was carried out with help of a pilot point based approach [*de Marsily*, 1978] including a regularization term [*Alcolea et al.*, 2006]. It would have been preferable to condition multiple equally likely Y fields and L values with transient head data, using the procedure as outlined by *Hendricks Franssen et al.* [1999]. However, a single model calibration took around 16 days of CPU time on a 2800 Hz processor, so the conditioning of a large number of stochastic realizations would have required an excessive amount of CPU time. The leakage coefficient was determined for five zones; a meaningful division of the river into zones was obtained taking into account the position of the weirs. The number of pilot points for the calibration of Y had to be limited, again because of the CPU intensity of the calculations. The reproduction of the heads was, in general, good, not only for the two months of calibration but also for the rest of the 20 month period. Some systematic patterns in the model residuals could be observed; these were analyzed by

Doppler et al. [2007] and *Engeler et al.* [2011] and are not the focus of this work. *Doppler et al.* [2007] and *Engeler et al.* [2011] found strong evidence that the leakage coefficient shows significant temporal fluctuations which are related to (1) floods that transport sediments of the river bed and locally modify its permeability, (2) floods that modify the area over which the river can infiltrate in the aquifer, and (3) river water temperature fluctuations. *Engeler et al.* [2011] demonstrated that the residuals of an isothermally calibrated model could be reduced up to 30% (for measurement points close to the river) if the temperature dependence of the leakage coefficient was taken into account. Those calculations were carried out with a 3-D fully coupled variably saturated groundwater flow and heat transport model.

4.2. Real-Time Model

[20] The numerical model for variably saturated groundwater flow was coupled to a data assimilation routine. This routine (written in C/C++) updates the model states (and, if desired, also Y and/or L) using piezometric head data in the EnKF approach. The model EnKF3d-SPRING runs on a Linux platform, and four processors were used for the calculations. The matrix with damping parameters α, the frequency of assimilation of new observations, and the frequency of updating the parameters are defined by the user.

[21] In most simulation experiments, the calibrated model (with historical data) and 99 stochastic realizations were used together in the EnKF. Therefore, the 99 stochastic realizations characterize the uncertainty and serve to adapt the deterministic model in real time. We will refer to the deterministic model as the "central model run." As the calibration with the pilot points method only yielded conditional mean fields for Y and L, multiple equally likely inverse conditioned stochastic realizations were not available. Stochastic realizations for Y were built by adding a perturbation Y' on top of the calibrated Y field using 857 available small-scale Y measurements. It was assumed that Y' shows a multi-Gaussian distribution, although there is evidence that Y could be non-multi-Gaussian for the site. Not enough data are available for generating meaningful non-multi-Gaussian stochastic realizations of Y'. It is unlikely that the EnKF reveals non-multi-Gaussian structures of Y with the help of head data, as EnKF relies on a Gaussian assumption. Therefore, it is possible that the performance would have been improved by a combination of non-multi-Gaussian stochastic realizations and a data assimilation algorithm that would take into account non-normality. The Y measurements were obtained with small-scale pumping tests for limited sections along the boreholes and are assumed to have a support volume of 1 m $\times$ 1 m $\times$ 10 cm. The difference between the inversely estimated Y field and the Y measurements (which were not used in the calibration) was determined for the measurement locations (Y'). A semivariogram for Y' was estimated on the basis of the Y' data and fitted with a spherical model, a nugget equal to zero, sill equal to 0.584 $\log_{10}(\mathrm{m\ s^{-1}})^2$ and a range of 99 m in the horizontal plane and 3.2 m in the vertical direction. Conditional stochastic realizations were generated on a very fine regular grid of 1 m $\times$ 1 m $\times$ 10 cm on the basis of the estimated variogram [*Gomez-Hernandez and Journel*, 1993] and added to the inversely estimated Y field. The resulting $Y + Y'$ stochastic realizations were upscaled to the finite elements size using simplified renormalization [*Renard et al.*, 2000]. Stochastic realizations for L were built by applying a normally distributed perturbation to the calibrated L for each of the zones. The applied perturbations had an expectation equal to zero and a variance equal to the posterior variance of the inversely estimated L. The stochastic realizations of Y and L formed the starting base for additional conditioning with the help of EnKF. The assimilated observations were randomly perturbed according to *Burgers et al.* [1998]. The standard deviation of the measurement error was set to 5 cm. The largest contribution to the measurement error is the error in the terrain height.

4.3. Synthetic Aquifer Model

[22] In order to gain better insights into the performance of alternative data assimilation strategies and, in particular, the benefit of adapting parameters together with the states, experiments with a synthetic aquifer model for the Zurich site were carried out first. This synthetic model uses exactly the same discretization, boundary conditions, and forcings (pumping, natural recharge, artificial recharge, rivers) as the original model. The difference is that one of the stochastic realizations of Y and L was selected as "truth." This implies that the virtual truth and the 100 stochastic realizations are generated with the same (geo)statistical model. This also eliminates one possible error source that we could have in reality. In this virtual experiment it is expected that the generated ensemble spread is adequate, whereas for the real-world case, that the ensemble spread is too small cannot be excluded. All stochastic realizations are used for a forward model run for the period January 2004 to August 2005 (i.e., a subset of the complete simulation period of 4 years) during which new synthetic observations were generated. The observed values have an expectation value equal to the "true" value (from the forward model run using the true parameters as input) and a standard deviation of 5 cm. These observations were used in subsequent simulation experiments as conditioning information. Note that the simulation experiments for the synthetic aquifer model differ with respect to the original model in the following ways: (1) The model forcings and boundary conditions are without error for the synthetic case, whereas they may contain errors in practice. (2) The geostatistical model for Y and L is used to generate the true Y and L fields in the synthetic case, whereas the reality might exhibit spatial structures of Y and L that are not in correspondence with the adopted geostatistical model. (3) The leakage coefficient L is constant over time, whereas in reality it might show temporal variations, related to floods and temperature variations.

4.4. Simulation Experiments

4.4.1. Experiments With a Synthetic Model (January 2004 to August 2005)

[23] Only part of the simulation experiments with the synthetic model are presented here. We provide results for calculations where a relatively strong damping for the parameters Y and L was used ($\alpha = 0.1$; see equation (6)). It was found that larger α values gave less satisfactory results, which is in correspondence with the findings of *Hendricks*

Franssen and Kinzelbach [2008] and is related to the filter inbreeding problem. For these synthetic experiments, the synthetic model was noncalibrated and used together with stochastic realizations. In most cases, only 100 stochastic realizations were used, although more realizations would have given better results. For 100 stochastic realizations the required CPU time was already 2 weeks on a standard PC.

[24] Table 1 gives an overview over the simulation scenarios in the experiments with the synthetic model. The following variations were chosen.

[25] 1. For conditioning data, a total of four different cases were considered: no conditioning data (i.e., open loop simulations), assimilating all available observations (87) each day, assimilating part of the observations (43) each day, and assimilating all available observations only once every 10 days. The last two experiments are the most suited for verifying the predictions. In the case where 43 observations are used, the other 44 measurements are used only for verifying predictions and are never used for assimilation. Assimilating only every 10 days allows us to verify whether a positive impact of the data assimilation (performed 10 days earlier) can be found for a 10 day prediction.

[26] 2. For parameter calibration, a total of four different cases were considered: no parameter calibration (i.e., only updating of states), calibration of Y only, calibration of L only, and calibration of both Y and L.

4.4.2. Off-Line Experiments With Historical Data (January 2004 to December 2007)

[27] For the off-line experiments with historical data a large number of simulation experiments were carried out, part of which are presented here. Again, results are provided only for experiments in which a relatively strong damping for the parameters Y and L was used ($\alpha = 0.1$). In these simulations the calibrated model is used together with 99 stochastic realizations; the stochastic realizations are used to update the calibrated model. Only 99 stochastic realizations were used because the required CPU time for one simulation experiment was already 1 month (for the period January 2004 to December 2007).

[28] Table 2 gives an overview of the simulation scenarios in the experiments with historical data. The same scenarios were chosen as in the experiments with the synthetic model (section 4.4.1), the difference being that in these scenarios a period of 4 years was considered.

Table 1. Different Simulation Scenarios for the Synthetic Case

Simulation Scenario	Parameters Calibrated	Amount of Data Assimilated	Number of Realizations
0.1	-	0	100
1.1	-	87	100
1.2	*Y*	87	100
1.3	*L*	87	100
1.4	*Y, L*	87	100
2.1	-	43	100
2.2	*Y*	43	100
2.3	*L*	43	100
2.4	*Y, L*	43	100
3.1	-	87, every 10 days	100
3.2	*Y*	87, every 10 days	100
3.3	*L*	87, every 10 days	100
3.4	*Y, L*	87, every 10 days	100

Table 2. Different Simulation Scenarios for the Historic Cases[a]

Simulation Scenario	Parameters Calibrated	Amount of Data Assimilated	Number of Realizations
0.1_04-05	-	0	100
1.1_04-05	-	87	100
1.2_04-05	*Y*	87	100
1.3_04-05	*L*	87	100
1.4_04-05	*Y, L*	87	100
2.1_04-05	-	43	100
2.2_04-05	*Y*	43	100
2.3_04-05	*L*	43	100
2.4_04-05	*Y, L*	43	100
3.1_04-05	-	87, every 10 days	100
3.2_04-05	*Y*	87, every 10 days	100
3.3_04-05	*L*	87, every 10 days	100
3.4_04-05	*Y, L*	87, every 10 days	100
0.1_05-07	-	0	100
1.1_05-07	-	87	100
1.2_05-07	*Y*	87	100
1.3_05-07	*L*	87	100
1.4_05-07	*Y, L*	87	100
2.1_05-07	-	43	100
2.2_05-07	*Y*	43	100
2.3_05-07	*L*	43	100
2.4_05-07	*Y, L*	43	100
3.1_05-07	-	87, every 10 days	100
3.2_05-07	*Y*	87, every 10 days	100
3.3_05-07	*L*	87, every 10 days	100
3.4_05-07	*Y, L*	87, every 10 days	100

[a]The scenarios followed by 04-05 correspond to the period 1 January 2004 to 31 August 2005, whereas those with 05-07 correspond to the period from 17 August 2005 to 31 December 2007.

4.4.3. Online Experiments for the Experimental Site (January 2009 to September 2010)

[29] EnKF3d-SPRING was implemented in an online system (see Figure 3). Hydraulic head measurements are communicated mainly by a 4–20 mA signal via cable (all except eight cases) but also by mobile telephone (six cases) or SMS (two cases) and are directly stored in the data warehouse or on an FTP server that is accessed by the Water Supply Zurich. Daily meteorological data are automatically downloaded by FTP from a MeteoSwiss Web site. River discharge measurements are obtained from the electricity company EWZ by a cable connection. During the first part of the simulation period (January 2009 to 14 November 2009), no weather predictions and river discharge predictions were included in the assimilation system. Instead, it was assumed that the next day the conditions would be the same as the previous day. From 15 November 2009 onward, predictions of river discharge for the next few days have already been included in the data assimilation system. The river discharge predictions are obtained from a multimodel ensemble of regional atmospheric models. The predictions are downscaled and provide 1-hourly input for the distributed rainfall-runoff model PREVAH [*Verbunt et al.*, 2007]. At the end of each day (i.e., shortly after midnight), 87 daily averaged piezometric head measurements are assimilated and used to update the states and leakage parameters of the model, using $\alpha = 0.1$ for updating L. The model updating is done only once per day. The updated initial conditions and parameter values are used for predictions of the head distributions for the following day, assuming that the management of the well field is unchanged. The predictions are used to optimize the management of the well field for the

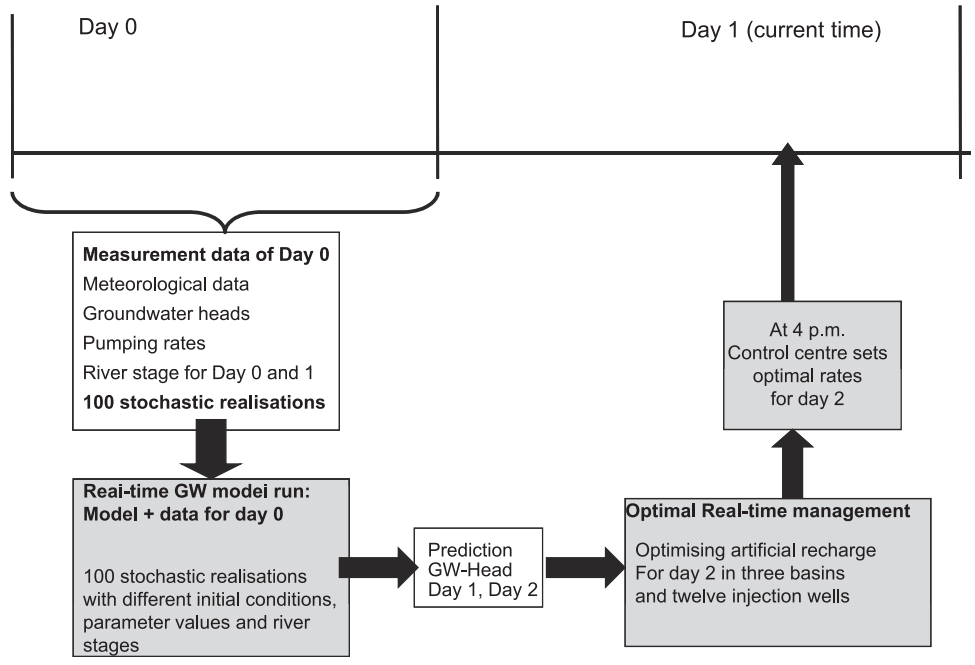

Figure 3. Overview of the different steps implemented for the operational real-time modeling of groundwater flow in the Limmat Valley aquifer and the related real-time control of the groundwater well field Hardhof in the city of Zurich (Switzerland).

next day in an iterative fashion, on the basis of a control engineering approach combining fuzzy-based methods and genetic algorithms [*Bauser et al.*, 2010]. This optimization uses an iterative approach, and the final result (in terms of optimized pumping rates and recharge rates for the different wells and basins) is used for the definitive predictions with the model, as the optimized pumping and artificial recharge rates will be implemented by the water works to operate the well field. The model predictions are compared with the head data that are measured later.

4.5. Evaluation of Model Predictions

[30] The main performance criterion to evaluate the predictions based on the data assimilation algorithm was the mean absolute error (MAE). A distinction is made between predictions by the central model run (for the real world case) and the results from the stochastic realizations (for both the real-world case and the synthetic case):

$$\mathrm{MAE}^{\mathrm{central}}(h) = \frac{1}{n_t n} \sum_{t=1}^{n_t} \sum_{i=1}^{n} \left| h_{i,t}^{\mathrm{sim}} - h_{i,t}^{\mathrm{meas}} \right|, \tag{9}$$

$$\mathrm{MAE}^{\mathrm{stochastic}}(h) = \frac{1}{n_r n_t n} \sum_{r=1}^{n_r} \sum_{t=1}^{n_t} \sum_{i=1}^{n} \left| h_{i,t,r}^{\mathrm{sim}} - h_{i,t}^{\mathrm{meas}} \right|, \tag{10}$$

where sim refers to model predictions (which were made on the basis of the updated initial conditions and possibly also model parameters the day before) and meas refers to measured values (at time t). As indicated, the mean absolute error is averaged over all measurement locations n and measurement times n_t and for the stochastic realizations over the number of realizations n_r. The mean absolute error is also evaluated for individual time steps and locations. In some cases, we will report values for individual time steps and part of the measurement locations, which will be indicated in the text. Moreover, the root-mean-square error was calculated, which was also evaluated separately for the central model run and the stochastic realizations:

$$\mathrm{RMSE}^{\mathrm{central}}(h) = \sqrt{\frac{1}{n_t n} \sum_{t=1}^{n_t} \sum_{i=1}^{n} \left(h_{i,t}^{\mathrm{sim}} - h_{i,t}^{\mathrm{meas}} \right)^2}, \tag{11}$$

$$\mathrm{RMSE}^{\mathrm{stochastic}}(h) = \sqrt{\frac{1}{n_r n_t n} \sum_{r=1}^{n_r} \sum_{t=1}^{n_t} \sum_{i=1}^{n} \left(h_{i,t,r}^{\mathrm{sim}} - h_{i,t}^{\mathrm{meas}} \right)^2}. \tag{12}$$

[31] For the synthetic experiments, the true parameter values are known, and estimates can be compared with the true parameter values. These comparisons are always made for a specific time step and are never averaged over multiple time steps:

$$\mathrm{MAE}(X,t) = \frac{1}{E} \sum_{i=1}^{E} \left| X_{i,t}^{\mathrm{sim}} - X_{i,t}^{\mathrm{true}} \right|, \tag{13}$$

where E is the number of finite elements, true refers to the parameter values from the synthetic reality, and X is either Y or L. Also, the root-mean-square error of the parameter estimates is evaluated but in a slightly different way than for the hydraulic heads:

$$\mathrm{RMSE}(X,t) = \sqrt{\frac{1}{E} \sum_{i=1}^{E} \left(\bar{X}_{i,t} - X_{i,t}^{\mathrm{true}} \right)^2}, \tag{14}$$

where an overbar indicates ensemble average values.

5. Results

5.1. Synthetic Case

5.1.1. Data Assimilation Without Parameter Calibration

[32] First, we will have a look at the prediction of hydraulic heads for the synthetic case. It is important to stress again that in this case, 100 stochastic realizations for Y and L are used, conditioned on information from Y and L, but not conditioned on h, and without a central model run (i.e., without a calibrated model). When no hydraulic head data are used for assimilation, the average absolute error for the predicted hydraulic head, computed from the ensemble, is 46.6 cm. This average is calculated over all 609 time steps, 87 measurement locations, and 100 realizations. Table 3 summarizes all simulation results in terms of hydraulic head characterization for the synthetic case. The results will be discussed in terms of MAE. In most cases, the results in terms of RMSE correspond with the ones for MAE. If this is not the case, a separate comment is added.

[33] Daily assimilation of 87 hydraulic head data reduces the $MAE^{stochastic}(h)$ for 1 day predictions (at the 87 measurement locations) by 59% down to 19.2 cm. Figure 4 shows the temporal evolution of $MAE^{stochastic}(h)$ for this case. Immediately after assimilation (i.e., for the actual situation), $MAE^{stochastic}(h)$ is only 3.3 cm. This illustrates that $MAE^{stochastic}(h)$ increases relatively quickly after assimilating new information, which is related to parameter errors; as in the synthetic model, only Y and L are uncertain. It is possible that assimilation improves the characterization of hydraulic heads only at the measurement locations but not elsewhere. Therefore, the data set was split into 43 hydraulic head data used for assimilation and 44 other locations where the hydraulic head was compared with the predicted values, without using them for assimilation. In that case, $MAE^{stochastic}(h)$ for 1 day predictions at the assimilation locations is 23.9 cm, while it is 28.0 cm at the verification locations. In this case, less information is assimilated (43 instead of 87 data), which likely causes the smaller reduction of $MAE^{stochastic}(h)$ at the measurement locations. On the other hand, at the verification locations the error is only marginally larger than at the assimilation locations, which indicates that the data assimilation is able to improve the estimate of the hydraulic heads in the complete domain. This positive impact of data assimilation at the verification locations can also be observed from the error measure $MAE^{stochastic}(h)$; it decreases from 28.0 cm just before assimilation (1 day prediction) to 19.9 cm right after assimilation.

[34] A concern is that $MAE^{stochastic}(h)$ increases rapidly after assimilation. This is illustrated in Figure 5. This increase could suggest that data assimilation without parameter adaptation gives an improved characterization of the hydraulic heads (compared to the default case without assimilation) only during a relatively short period. However, this is not the case. If hydraulic head data are only assimilated every 10 days, the $MAE^{stochastic}(h)$ for 10 days after the assimilation (i.e., just before a new assimilation) is 36.0 cm, which is still considerably smaller than the value for the case without assimilation (46.6 cm).

5.1.2. Data Assimilation With Parameter Calibration

[35] If only one of the two unknown parameters (Y or L) is updated (together with the states), the hydraulic head characterization is not better than for the case where only the states are updated. The differences (in terms of $MAE^{stochastic}(h)$ and $RMSE^{stochastic}(h)$) are very small for most of the scenarios (see Table 3). This is in spite of the fact that the parameter values are better or worse than compared to the prior values. If only L is updated, MAE(L, 609) (i.e., the mean absolute error at the end of the simulation period) is between 19% and 24% lower than the prior value. In all three cases (with 87 data, 43 data and assimilation of 87 data every 10 days), MAE(L, 609) reached lower values before the end of the assimilation, with a value that was more than 30% lower than the prior value. Table 4 summarizes all simulation results in terms of parameter characterization for the synthetic case. If only Y is updated, MAE(Y, 609) is larger than the prior value. In the beginning of the assimilation period, MAE(Y) is lower than the prior MAE(Y); after 100 days, the error reduction is 13%. Afterward, the MAE(Y) starts to increase. See Figure 6. This result seems to be related to filter inbreeding. *Hendricks Franssen and Kinzelbach* [2008] found in simulation experiments that for small ensemble sizes and frequent assimilation, the MAE(Y) initially decreases but increases later during the simulation (see their Figure 8 and Table 4). They found that for a large Y variance (like the one in this study) and for a relatively small ensemble of 100 realizations, in the absence of a damping parameter, MAE(Y) was larger at the end of a simulation period of 200 days than at the start. This filter divergence is provoked by a strong underestimation of the Y variance (filter inbreeding), which in turn, is at least in part related to the poor quality of the estimated numerical covariances. *Hendricks Franssen and Kinzelbach* [2008] found that even with a damping parameter (and the ensemble of 100 stochastic realizations), the MAE(Y) increased at later simulation times. However, the MAE(Y) after 200 days was still smaller than the prior MAE(Y). They found that 200 stochastic realizations and a damping parameter avoided filter divergence, and in that case, only a minor underestimation of the variance (filter

Table 3. Performance Measures for the Synthetic Simulation Experiments[a]

	Assimilation Locations		Verification Locations	
Simulation Scenario	$MAE^{stochastic}$ (h) (cm)	$RMSE^{stochastic}$ (h) (cm)	$MAE^{stochastic}$ (h) (cm)	$RMSE^{stochastic}$ (h) (cm)
0.1	46.6	63.4		
1.1	19.2	30.7		
1.2	19.2	30.7		
1.3	21.2	29.3		
1.4	7.2	19.4		
2.1	23.9	35.8	28.0	55.6
2.2	24.0	35.9	27.2	54.8
2.3	23.8	35.6	27.5	55.5
2.4	16.6	26.1	21.2	41.0
3.1	36.0	50.5		
3.2	36.0	50.5		
3.3	36.0	50.5		
3.4	7.4	17.2		

[a]The 1 and 10 day predictions of hydraulic head are given at the same locations where the assimilation was done or at independent verification locations. Note that in all cases, error statistics from predictions are provided, although this corresponds in many cases to the same location where observations were assimilated.

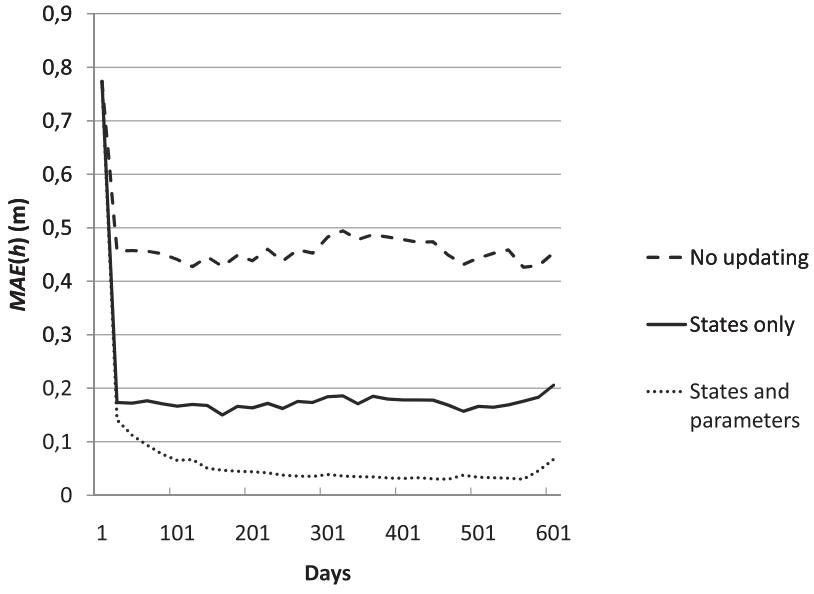

Figure 4. The mean absolute error for 1 day predictions of hydraulic head for the synthetic case as function of time. The mean absolute error is evaluated over all 100 stochastic realizations. Shown are the predictions for the cases where (1) no data assimilation was performed, (2) only states were updated with the help of 87 daily hydraulic head data, and (3) states, *Y,* and *L* were updated with 87 daily hydraulic head data.

inbreeding) was observed. The same is the case for the experiments presented here; the experiments with the damping parameter resulted in a MAE(*Y*) that after 200 days was smaller than the prior one. Experiments without including a damping parameter or daily updating of *Y* showed a worse performance, with only an initial small decrease of MAE(*Y*) and MAE(*Y*) values larger than the prior ones after only 50 and 75 days, respectively.

[36] An alternative explanation for the observed increase of MAE(*Y*) at later simulation times could be that the initial ensemble had too small a spread. As explained, this is not expected because the virtual truth and the 100 stochastic realizations were generated using the same (geo)statistical model. The initial ensemble spread was analyzed for the open-loop simulations where no data were assimilated. This was done using a procedure similar to the one used by *Moradkhani et al.* [2005] and *De Lannoy et al.* [2006]. According to this method, the ratio of the time-averaged RMSE of the ensemble mean and the mean RMSE of the ensemble members was calculated. For 100 stochastic realizations it is expected that this ratio will be 0.711, while for this study a ratio of 0.464 was calculated. This implies that our initial ensemble spread was too large. The explanation for this deviation is most likely that the generated virtual

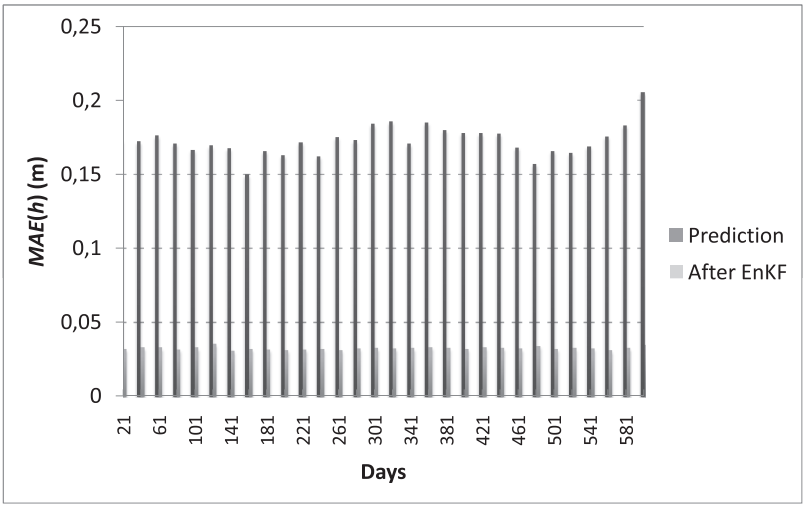

Figure 5. The mean absolute error (as function of time) for 1 day predictions of hydraulic head and after data assimilation with EnKF. The mean absolute error is evaluated over all 100 stochastic realizations. Shown are the results for scenario 1.1 of the synthetic cases.

Table 4. Performance Measures for the Synthetic Simulation Experiments[a]

Simulation Scenario	MAE (Y) ($\log_{10}$(m s^{-1}))	RMSE (Y) ($\log_{10}$(m s^{-1}))	MAE (L) ($\log_{10}$(m s^{-1}))	RMSE (L) ($\log_{10}$(m s^{-1}))
0.1	0.396	0.494	0.296	0.356
1.1	NA	NA	NA	NA
1.2	0.457	0.572	NA	NA
1.3	NA	NA	0.232	0.321
1.4	0.381	0.475	0.142	0.151
2.1	NA	NA	NA	NA
2.2	0.398	0.498	NA	NA
2.3	NA	NA	0.224	0.278
2.4	0.332	0.414	0.168	0.199
3.1	NA	NA	NA	NA
3.2	0.313	0.391	NA	NA
3.3	NA	NA	0.240	0.269
3.4	0.290	0.363	0.112	0.130

[a]A comparison is made between the true parameter values and the estimated parameter values at the end of the simulation period (after 609 days). NA indicates not applicable (i.e., for that scenario the parameter was not calibrated).

truth was by chance a very "average" one, which is therefore well covered by the 100 stochastic realizations. Other measures for analyzing the ensemble spread indicate too large an initial ensemble spread as well. This excessive initial ensemble spread does not explain the later too small ensemble spread (filter inbreeding) and increased MAE(Y).

[37] The observed increase in MAE(Y) could also be linked to nonoptimal conditions under which the ensemble Kalman filter is applied, like bias and nonnormal distributions of parameters and states. For this synthetic case under controlled conditions, only the nonnormal distribution of states might have had a negative impact on the results.

[38] If Y and L were simultaneously updated along with the states, results were much better. For those scenarios, $\mathrm{MAE}^{\mathrm{stochastic}}(h)$ was only 7.2 cm for 1 day predictions in cases where 87 data were assimilated (compared to 19.2 cm without parameter updating). Figure 4 illustrates this case. For the same case, Figure 7 displays explicitly the relative improvement of performance if states, Y, and L are updated simultaneously compared with updating states only. It illustrates that the relative improvement continues to grow during the first 200 days of simulation. For the 10 day prediction the differences were even larger: $\mathrm{MAE}^{\mathrm{stochastic}}(h)$ was only 7.4 cm compared to 36.0 cm without calibration (see Figure 8). It is remarkable that $\mathrm{RMSE}^{\mathrm{stochastic}}(h)$ for 10 day predictions is even lower than for 1 day predictions. This seems to be related to better parameter estimates if hydraulic heads are assimilated only every 10 days. For the experiment with assimilation of 43 head data only and 44 verification locations, the error reductions were smaller: at the verification locations, $\mathrm{MAE}^{\mathrm{stochastic}}(h)$ was 21.2 cm instead of 28.0 cm (without parameter adaptation). The parameter updates also improved for these simulation scenarios. MAE(L, 609) (i.e., the mean absolute error at the end of the simulation period) was between 43% and 62% lower than the prior value, which is a much better result than the one obtained for the cases when only L was calibrated. For all three scenarios (with 87 data, 43 data, and assimilation of 87 data every 10 days) the estimate of L improved throughout the entire simulation period. Also, the estimates of Y improved for all the three simulation experiments were Y was calibrated together with L. However, results still seem to be influenced by filter inbreeding. The best results were found if head data were only assimilated every 10 days: in that case, MAE(Y, 609) was 27% lower than the prior MAE(Y), and an improvement of MAE(Y) was observed throughout the complete simulation period. The worst results were obtained if 87 head data were assimilated each day. In that case, MAE(Y, 609) was only 4% lower than the prior MAE(Y) and increased during the second part

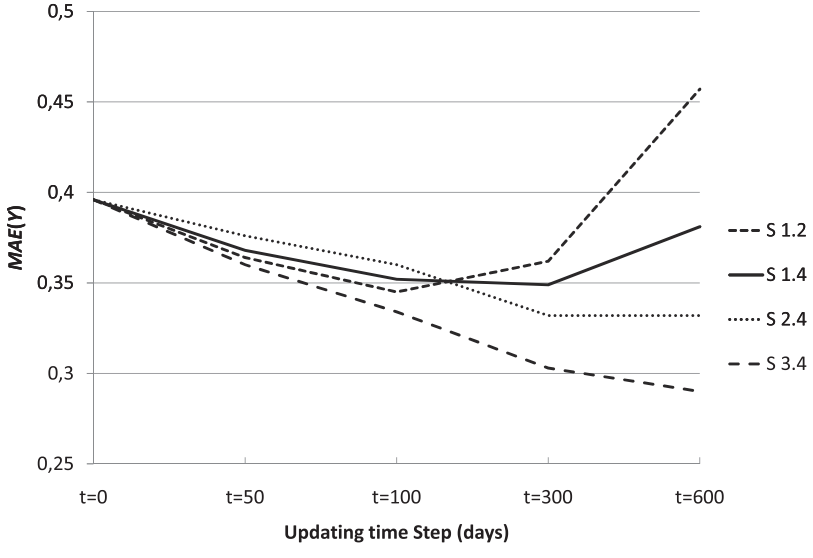

Figure 6. Temporal evolution of mean absolute errors of log hydraulic conductivities (synthetic case), evaluated over all finite elements. Given are results for the case where only Y is updated (although both Y and L are uncertain) and three different cases where both Y and L are updated: with assimilation of 87 daily hydraulic head data (S 1.4), with assimilation of 43 daily hydraulic head data (S 2.4), and with assimilation of 87 hydraulic head data, every 10 days (S 3.4).

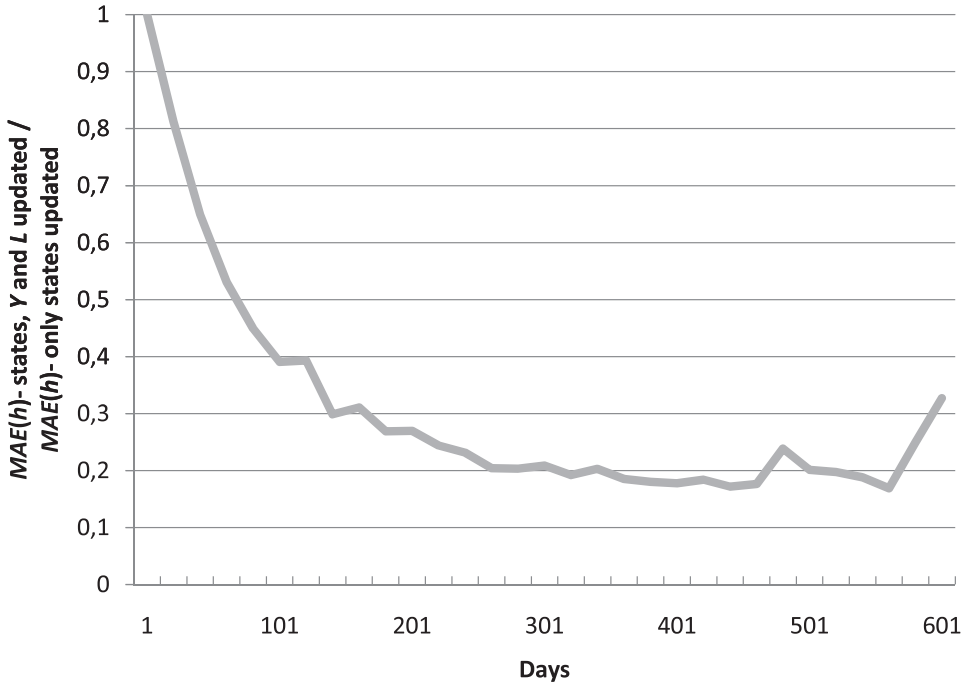

Figure 7. Relative improvement of data assimilation (synthetic case) where both states and parameters are updated, as compared to data assimilation where only the states are updated. The ratio of MAE(h) if states, Y, and L are updated to MAE(h) if only states are updated is given. Shown are the predictions for the case where 87 daily hydraulic head data are assimilated and MAE(h) is calculated over all 100 stochastic realizations.

of the simulation part after reaching a reduction above 12% in the middle of the simulation period. This behavior is also displayed in Figure 6, together with the results for some other scenarios. These findings are typical for filter inbreeding. Nevertheless, all results confirm that jointly updating Y and L yields very good results, which are much better than what is obtained by updating states only or by updating only one of the two parameters.

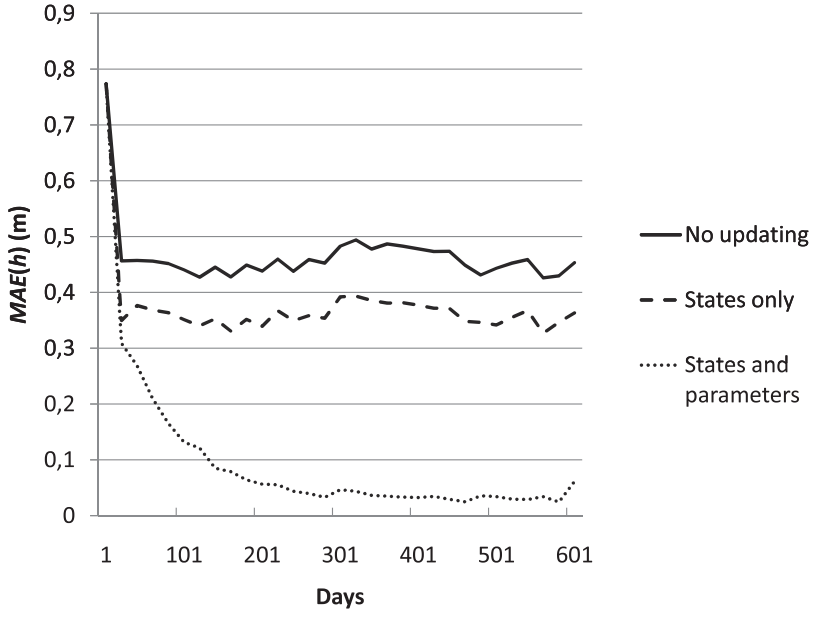

Figure 8. The mean absolute error for 10 day predictions of hydraulic head for the synthetic case as function of time. The mean absolute error is evaluated over all 100 stochastic realizations. Shown are the predictions for the cases where (1) no data assimilation was performed, (2) only states were updated using 87 daily hydraulic head data, and (3) states, Y, and L were updated using 87 daily hydraulic head data.

5.2. Off-Line Experiments With Historical Data (January 2004 to December 2007)

5.2.1. Data Assimilation Without Parameter Calibration

[39] It is stressed that in this case the calibrated model was the central model run, which was updated with observations using an ensemble of 99 stochastic realizations. The results for the updated central model forecast and the forecasts from the stochastic realizations are analyzed separately. Results are also compared separately for the period from January 2004 to August 2005, the period in which 2 months were used for calibration, and the period from September 2005 to December 2007, the verification period. Tables 5 (central model run) and 6 (stochastic realizations) summarize all simulation results in terms of hydraulic head characterization for the off-line experiments.

[40] In the application we find again that data assimilation always reduces MAE(h). This MAE(h) reduction is larger for the stochastic realizations than for the central model run, which could be expected as the central model run was calibrated with historical data. Nevertheless, $MAE^{central}(h)$ still is reduced by 33% (January 2004 to August 2005) or 41% (September 2005 to December 2007) for 1 day predictions (and assimilating 87 hydraulic head data daily) compared to predictions without data assimilation. Data assimilation with EnKF also reduces $MAE^{central}(h)$ if only 43 data are assimilated and for 10 day predictions, both compared to predictions with a calibrated model. For stochastic realizations, $MAE^{stochastic}(h)$ decreases by 53% for 1 day predictions if 87 data are assimilated, both for the period January 2004 to August 2005 and for the period September 2005 to December 2007. The simulation experiment with 44 verification locations showed that the reduction of $MAE^{stochastic}(h)$ (i.e., averaging over the realizations) at the verification locations was stronger (40% for 1 day predictions for the period August 2005 to December 2007) than the $MAE^{central}(h)$ reduction (21% for the same period). Further simulation results are given in Table 5 and illustrate that the hydraulic head characterization is improved with data assimilation both for the real-world case and if an already calibrated model is used.

Table 5. Performance Measures for the Real-World Simulation Experiments for the Periods January 2004 to August 2005 and August 2005 to December 2007[a]

	Assimilation Locations		Verification Locations	
Simulation Scenario	$MAE^{central}$ (h) (m)	$RMSE^{central}$ (h) (m)	$MAE^{central}$ (h) (m)	$RMSE^{central}$ (h) (m)
0.1_04-05	0.195	0.278		
1.1_04-05	0.130	0.168		
1.2_04-05	0.130	0.167		
1.3_04-05	0.128	0.210		
1.4_04-05	0.112	0.145		
2.1_04-05	0.126	0.166	0.206	0.323
2.2_04-05	0.126	0.161	0.204	0.319
2.3_04-05	0.128	0.165	0.191	0.290
2.4_04-05	0.116	0.155	0.185	0.267
3.1_04-05	0.179	0.237		
3.2_04-05	0.179	0.241		
3.3_04-05	0.179	0.236		
3.4_04-05	0.142	0.191		
0.1_05-07	0.257	0.361		
1.1_05-07	0.151	0.214		
1.2_05-07	0.151	0.216		
1.3_05-07	0.195	0.256		
1.4_05-07	0.130	0.181		
2.1_05-07	0.162	0.223	0.202	0.330
2.2_05-07	0.162	0.222	0.200	0.319
2.3_05-07	0.162	0.222	0.203	0.332
2.4_05-07	0.182	0.410	0.218	0.445
3.1_05-07	0.230	0.313		
3.2_05-07	0.230	0.313		
3.3_05-07	0.231	0.313		
3.4_05-07	0.161	0.220		

[a]Here the central model run simulations are evaluated. The 1 and 10 day predictions of hydraulic head are given at the same locations where the assimilation was done or at independent verification locations. Note that in all cases, error statistics from predictions are provided, although this corresponds in many cases to the same location where observations were assimilated.

Table 6. Performance Measures for the Real-World Simulation Experiments for the Periods January 2004 to August 2005 and August 2005 to December 2007[a]

	Assimilation Locations		Verification Locations	
Simulation Scenario	$MAE^{stochastic}$ (h) (m)	$RMSE^{stochastic}$ (h) (m)	$MAE^{stochastic}$ (h) (m)	$RMSE^{stochastic}$ (h) (m)
0.1_04-05	0.506	0.666		
1.1_04-05	0.237	0.318		
1.2_04-05	0.236	0.317		
1.3_04-05	0.250	0.327		
1.4_04-05	0.139	0.181		
2.1_04-05	0.313	0.440	0.394	0.563
2.2_04-05	0.313	0.437	0.392	0.552
2.3_04-05	0.317	0.447	0.384	0.556
2.4_04-05	0.184	0.308	0.262	0.395
3.1_04-05	0.423	0.561		
3.2_04-05	0.419	0.559		
3.3_04-05	0.423	0.563		
3.4_04-05	0.169	0.234		
0.1_05-07	0.525	0.674		
1.1_05-07	0.245	0.324		
1.2_05-07	0.244	0.325		
1.3_05-07	0.274	0.358		
1.4_05-07	0.138	0.197		
2.1_05-07	0.281	0.366	0.314	0.415
2.2_05-07	0.281	0.366	0.313	0.415
2.3_05-07	0.281	0.367	0.314	0.415
2.4_05-07	0.250	0.411	0.315	0.518
3.1_05-07	0.444	0.570		
3.2_05-07	0.444	0.570		
3.3_05-07	0.444	0.570		
3.4_05-07	0.177	0.255		

[a]Here the stochastic simulation runs are evaluated. The 1 and 10 day predictions of hydraulic head are given at the same locations where the assimilation was done or at independent verification locations. Note that in all cases, error statistics from predictions are provided, although this corresponds in many cases to the same location where observations were assimilated.

5.2.2. Data Assimilation With Parameter Calibration

[41] If only Y or L is calibrated, $MAE^{central}(h)$ and $MAE^{stochastic}(h)$ are not reduced compared to simulations where only the states are updated. The 1 day predictions of MAE(h) or RMSE(h) are even slightly worse in two cases when only L is calibrated. These results correspond with the results from the synthetic study. Tables 5 and 6 contain all simulation results for the scenarios where parameters were updated.

[42] Jointly updating Y and L in the real-world case improves the estimates compared to updating the states

only: in terms of $MAE^{central}(h)$ and 1 day predictions with 87 head data the values were 11.2 versus 13.0 cm for January 2004 to August 2005 and 13.0 versus 15.1 cm for September 2005 to December 2007. For 10 day predictions, the corresponding values are 14.2 versus 17.9 cm for January 2004 to August 2005 and 16.1 versus 23.0 cm for September 2005 to December 2007. Updating the parameters for the already calibrated model might also yield better estimates because L is subjected to temporal fluctuations, as mentioned [*Doppler et al.*, 2007; *Engeler et al.*, 2011]. For $MAE^{stochastic}(h)$ the error reduction is larger: for 1 day predictions with 87 data it is 13.9 versus 23.7 cm for January 2004 to August 2005 and 13.8 versus 24.5 cm for September 2005 to December 2007. For 10 day predictions, $MAE^{stochastic}(h)$ is also much lower than for simulations that used the same data without parameter updating: 16.9 versus 42.3 cm for January 2004 to August 2005 and 17.7 versus 44.4 cm for September 2005 to December 2007. It is found that after a certain simulation time, the stochastic realizations, which were not calibrated to h data, yield head predictions that are equally as good as the central model run, which was calibrated to historical head time series. This clearly supports the hypothesis that jointly updating Y and L results in improved parameter estimates. (See Figure 9 for an illustration.)

[43] Results are less favorable for simulations over the period September 2005 to December 2007 for the assimilation of 43 data using 44 data for verification. In that case, $MAE^{central}(h)$ is slightly larger than if only states were updated, and $MAE^{stochastic}(h)$ is only slightly lower. Moreover, $RMSE^{central}(h)$ and $RMSE^{stochastic}(h)$ are much larger than if only states were updated. This holds for both assimilation locations and verification locations. A closer look at the results reveals that until June 2007, $MAE^{central}(h)$, $MAE^{stochastic}(h)$, $RMSE^{central}(h)$, and $RMSE^{stochastic}(h)$ at assimilation and verification locations are clearly lower than the same measurements for simulations without parameter updating. In the second half of 2007, results are, however, much worse, and at some locations MAE(h) and, particularly, RMSE(h) become very large because of the generation of implausible parameter values. This result points to the risk of instabilities that might appear because of the continuous updating of parameter values in real time. For the simulation period January 2004 to August 2005, results are always better if Y and L are jointly updated, at both assimilation and verification locations, and for the 99 stochastic realizations the improvement is significant.

[44] For the real-world case it is not possible to compare estimated and true parameter values. It should also be stressed that for the real-world case there is an increased risk that the prior ensemble has too small a spread, for example, because the adopted geostatistical model may not be adequate. We compare here the estimated parameter values for those simulation scenarios where both Y and L are updated (i.e., different amounts of hydraulic head data and for the two different simulation periods). It is found that the parameter perturbations (updated parameter values minus a priori values) are strongly correlated among the different

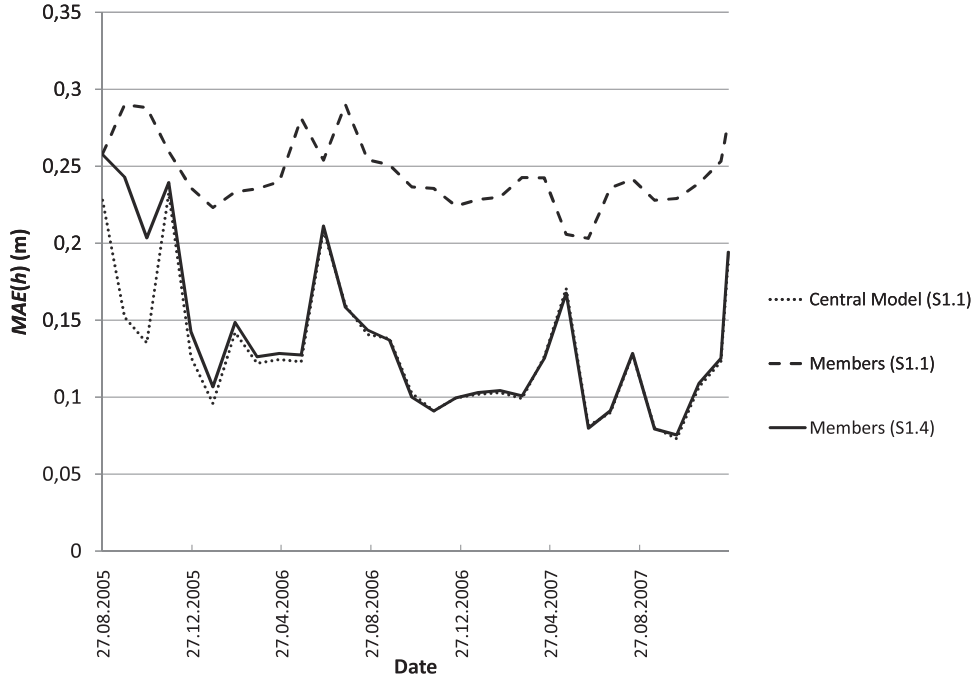

Figure 9. The mean absolute error for 1 day predictions of hydraulic head using 87 hydraulic head data for the real-world case for the period August 2005 to December 2007. Shown are the (1) predictions with the central model (inversely calibrated) whose states are updated daily with EnKF (central model, scenario 1.1), (2) predictions with the 99 stochastic realizations (not calibrated previously) whose states are updated daily with EnKF (members, scenario 1.1), and (3) predictions with the 99 stochastic realizations whose states are updated daily and parameters (Y and L) are updated every 10 days with EnKF (members, scenario 1.4).

scenarios. It is particularly interesting to have a look at L. The different scenarios give consistent results for three of the five zones: in comparison with the a priori values, L shows a strong increase (around 1.0 $\log_{10}$ units on average) for the Limmat west of the Hardhof area, no change for the Limmat in the eastern Hardhof area, and a significant decrease (around 0.7 $\log_{10}$ units on average) for the Limmat east of Hardhof. For the Limmat in the eastern Hardhof a yearly cycle in the leakage coefficient with an average amplitude of 0.2 $\log_{10}$ units is found. The minimal leakage coefficient is found in May in all 4 years, and the maximum value for the leakage coefficient always occurs in November (see Figure 10). This suggests that the updated leakage coefficient lags 3 months behind the true cycle. *Doppler et al.* [2007] and *Engeler et al.* [2011] showed that for this section of the Limmat, the temperature dependence of the leakage coefficient has a significant impact on the hydraulic heads, and minimum L is expected for February, with maximum L expected for August. The time lag might be related to the time needed for the filter to adapt. Further, the hydraulic conductivities somewhat farther away from the river are also influenced by the temperature dependence because the retardation of the temperature wave shows a shifted cycle compared to the leakage factor. Now we have to look at the two sections of the rivers where results are less consistent. For the Limmat in the western Hardhof area the results for the simulation scenarios differ but are consistent for the two simulation periods: no significant change for the simulation period January 2004 to August 2005 and a strong decrease (around 0.9 $\log_{10}$ units on average) for the period August 2005 to December 2007. This decrease seems to be related to modifications of the river bed due to the major flood in August 2005. Visual inspection of model residuals also revealed a sudden jump in the model residuals coinciding with the flood event in August 2005 in this area [*Engeler et al.,* 2011]. It should be stressed that the EnKF needed around 8 months to fully adjust to the new L value. Finally, for the Sihl, three of the scenarios show that the updated L is almost equal to the prior L, whereas for three other scenarios an increase of L was found (on average, 0.3 $\log_{10}$ units). This part of the model domain has fewer observations.

5.3. Online Experiments (June to December 2009)

[45] EnKF3d-SPRING was implemented online using information from online sensors. Since January 2009, the data assimilation has been running online and making daily hydraulic head predictions for the next day. The default implementation included updating leakage coefficients but not updating hydraulic conductivities. As explained in section 4.4.3, initially, in the data assimilation it was assumed that the forcing for the next day would be the same as for the current day. This introduces an error as, for example, the river stage shows large temporal oscillations. This is the main reason why MAE(h) and RMSE(h) are larger than in the off-line experiments for 1 day predictions of hydraulic head. However, from 15 November 2009 onward, predictions of the river stage for the next day are integrated in the data assimilation framework. It was found that these predictions were very reliable during periods with larger variations of the river stage.

[46] Figure 11 shows the temporal evolution of MAE(h) and RMSE(h) for the period of May 2009 to September 2010. Before May 2009, there were strong pumping activities due to dewatering of construction sites that were not accounted for in the data assimilation. Therefore, the period January–April 2009 is not included in the comparison. MAE(h), evaluated over all measurement locations and time steps, is 22.2 cm, whereas RMSE(h) is 30.0 cm. Better results were found for the period where the river stage for

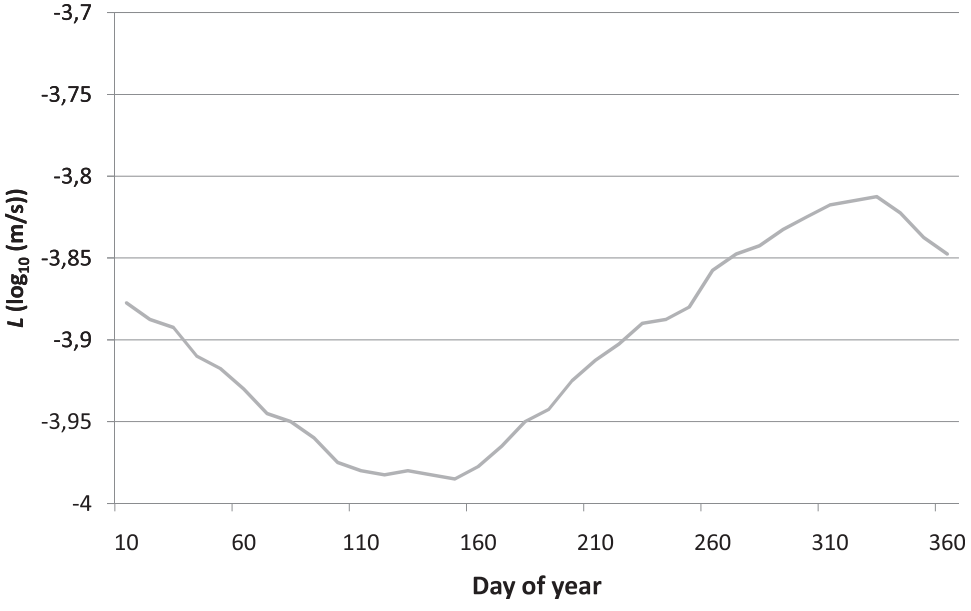

Figure 10. The average yearly cycle of the calibrated leakage coefficient for the real-world case (for the river Limmat in the Hardhof area) based on simulations for the period January 2004 to December 2007. Results are according to simulation scenario 3.4.

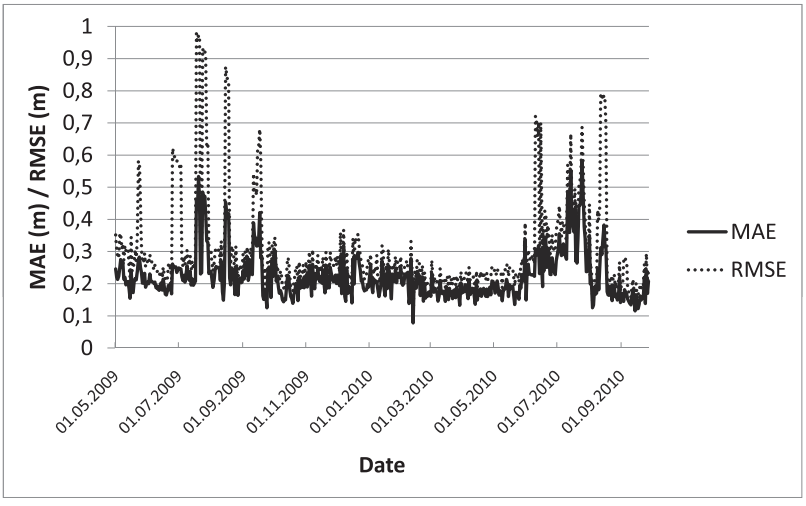

Figure 11. Comparison of online one day hydraulic head predictions with EnKF and measurements. Shown are the absolute mean error and root-mean-square error averaged over all measurement locations and as a function of time.

the next day was predicted. For this period the average MAE(h) and RMSE(h) were 21.5 and 27.7 cm, respectively, as compared to 23.3 and 33.5 cm, respectively, for the period before. The performance of the data assimilation for the period when river stage predictions were included was even better than these numbers suggest. However, the period May–August 2010 showed again strongly increased errors in spite of predicting the daily river stage. In this period all piezometers were replaced with new online sensors, and the information system at the Water Works was also replaced. Many new sensors initially gave erroneous measurement values, and at times more than half of them were not available. This explains the larger errors during this period. Predictions with EnKF for the verification period (August 2005 to December 2007), including calibration of leakage and perfectly known forcings, gave an MAE(h) of 19.5 cm. Predictions without data assimilation but with perfectly known forcings, using a model calibrated with historical data, resulted in an MAE(h) of 25.7cm for the verification period (August 2005 to December 2007).

6. Discussion

[47] These results indicate the positive impact of assimilating hydraulic head data, which yields better predictions of hydraulic head distributions, including heads away from the measurement locations and at a time horizon of prediction of 10 days, than a conventional model calibrated with historical data. Results can be improved further if uncertain parameter values (Y and L) are jointly updated together with the states. Jointly updating Y and L gives better results than updating Y or L alone. We think that this result is related to the fact that both Y and L are affected by errors. If only Y or L is corrected, the updated parameter also accounts for the error in the other parameter. This results in overcorrection and worse results compared to the case where it is acknowledged that both parameters might be affected by errors. However, some questions remain because in the synthetic case, parameter estimates of Y got worse later in the simulation period in one case (out of three). However, also in that case the estimates of Y at the end of the simulation period were still better than the prior values. There is evidence that the increase in the mean absolute error of the hydraulic conductivities in the second half of the simulation period is related to filter inbreeding. In the real-world case the prediction of hydraulic head distributions including parameter updating did not improve in one case (out of six) compared to predictions based on the same amount of data but without parameter updating. It was found that for that specific case, estimated hydraulic head distributions were much better than those obtained without parameter updating for the first 700 days of simulation but much worse for the last 150 simulation days. This result was related to numerical instabilities and points to the risks of continuously updating parameter distributions in real time. We think that the better results for the experiments where the assimilation is carried out every 10 days instead of each day (and parameters are updated) can also be explained by reduced filter inbreeding. Less frequent updating reduces problems with filter inbreeding [*Hendricks Franssen and Kinzelbach*, 2008]. On the other hand, this may not be the only reason. After a longer time period without assimilation, the model residuals are larger and contain clearer signals about too small/too large parameter values.

[48] To avoid filter inbreeding, we suggested some measures [*Hendricks Franssen and Kinzelbach*, 2008]. The use of a damping parameter is very important in reducing problems with filter inbreeding, and a large ensemble is also important. For large simulation models this requires parallelization of the code and many processors. That setup was not possible in this study, and the number of stochastic realizations (100) was relatively small; better results are expected for a larger number of stochastic realizations. Problems with filter inbreeding can be reduced further by excluding spurious larger positive or negative covariances for large separation distances [*Houtekamer and Mitchell*, 1998]. Results for the synthetic study would have been better if we would have taken additional measures against

filter inbreeding, but nevertheless, all results (in terms of hydraulic heads and parameter updates) were better than simulations with updating of states only.

[49] A bit more complicated is the situation for the real-world case study. Here instabilities developed at the end of a large assimilation experiment over 4 years. In this assimilation experiment, less data were assimilated. As parameter updates improve, especially in the beginning of the assimilation period, and stabilize later (or become worse again in the case of filter inbreeding), our suggestion for calibrating real-world models with EnKF and operational applications is to limit parameter calibration to the beginning of the simulation period with, for example, around 30 parameter updates and then update parameters later in the simulation period only very occasionally. The optimal frequency for parameter updating will depend on the case, and it is important to gain experience from additional simulation experiments to determine the role of the updating frequency and provide measures against possible instabilities that could develop over time.

[50] Nevertheless, EnKF was also successful for updating states and parameters of a subsurface flow system (for both a real-world case and an operational case) which was considerably more complex than the synthetic 2-D saturated groundwater flow problem used by *Hendricks Franssen and Kinzelbach* [2008, 2009]. As mentioned in section 1, real-time modeling is used as a basis for real-time optimization of the water management at the site (fixing the amount to be abstracted and infiltrated but optimizing the spatial distribution of artificial recharge, taking into account constraints like the capacity of artificial recharge basins and wells). Simulations indicate that without optimization (traditional management) the fraction of city water (i.e., potentially contaminated water) is 11% at well C (5% in case of optimization) and 6% at well D (2% in case of optimization). If the amount of water to be infiltrated is allowed to increase, an estimated fraction of city water of 0% at all wells can be achieved. The improved quality of the pumped water in online mode was confirmed by the evolution of its electrical conductivity, indicating a reduced fraction of pumped city water. See *Bauser et al.* [2010] for further details. The next extension of the method to more general problems is under preparation. It concerns the calibration with data assimilation of fully distributed, integral hydrological models that include overland flow as well as evapotranspiration from the unsaturated zone. This is only feasible by assimilating various types of data and by parallelizing the code and running it on a supercomputer (at least for the testing stage) to process sufficiently large ensembles.

7. Conclusions

[51] This paper presents data assimilation with the ensemble Kalman filter (EnKF) for variably saturated subsurface flow including river-aquifer interaction, implemented in a finite element model with 173,599 elements. Simulation experiments are carried out for the Limmat Valley aquifer for the period January 2004 to December 2007 and also for a synthetic case which is similar to the Limmat Valley aquifer, except for the fact that a certain parameter distribution was selected as the virtual truth. Finally, results are presented for the online implementation of this same model for the period May 2009–September 2010. To our knowledge, this is the first online implementation of a data assimilation framework for subsurface flow and the first to adapt online the optimal pumping strategy.

[52] Results indicate that data assimilation with EnKF but without parameter calibration improves 1 day and 10 day hydraulic head predictions, at both assimilation and prediction locations. The reduction of the mean absolute error (MAE) is slightly larger for the synthetic experiments (e.g., 59% reduction of MAE for 1 day predictions and assimilation of 87 hydraulic head data) than for the 99 stochastic realizations from the real-world case (53% MAE reduction for the same case). The difference is larger for 10 day predictions based on 87 assimilated data (23% reduction for the synthetic case versus 15% for the real-world case). Further improvements are obtained if log hydraulic conductivities (Y) and log leakage coefficients (L) are updated along with the states. If Y and L are updated simultaneously, MAE(h) is much lower than for the unconditional case, with an 85% reduction for 1 day predictions and an 84% reduction for 10 day predictions. For the real-world case, MAE(h) is reduced by 73% (1 day predictions) or 66% (10 day predictions) when compared to the same sets of simulation experiments. The improvements are smaller for 1 day predictions at verification locations: 54% MAE(h) reduction for the synthetic case and 44% reduction for the real-world case. The synthetic experiments allowed us to verify that the parameter estimates indeed improved for all experiments where both hydraulic conductivities and leakage coefficients were updated. The improvement was largest if all observations were assimilated but only every 10 days. In that case the MAE reduction for log hydraulic conductivity was 27%, and the MAE reduction for the log leakage coefficient was 63%. It was also observed that for one of the synthetic simulation scenarios the log hydraulic conductivity estimates initially improved strongly but got worse later during the simulation period, although they were still better than the prior estimates at the end of the simulation estimates. It is believed that this behavior is linked to filter inbreeding because the ensemble of 100 realizations was relatively small, resulting in numerical covariances subjected to considerable sampling fluctuations. Also, the fact that states are nonnormally distributed in this synthetic experiment might have contributed to the suboptimal results. In addition, an already inversely calibrated model was updated in the off-line experiments for the real-world case study. This model yielded the best hydraulic head predictions, which were also improved in the case when additional data were assimilated and were further improved if parameters were adapted in real time.

[53] The hydraulic head estimates obtained with the online operational model are better if data assimilation is applied. The absolute errors are larger than for the off-line model because of the up-to-now inferior characterization of the model forcings.

[54] These results corroborate the potential of the ensemble Kalman filter for the operational updating of large-scale groundwater flow models showing a highly dynamic response to surface water bodies and confirm that unknown model parameters can be calibrated and improved in real time. The results also indicate that very frequent updating

of parameters might give less good results than less frequent updating because of filter inbreeding and the risk of numerical instabilities.

[55] **Acknowledgments.** The study was performed within the project "Real-Time Control of a Well-Field Using a Groundwater Model," a cooperation between ETH Zurich, Zurich Water Supply, and TK Consult Zurich. This project is funded by the Swiss Innovation Promotion Agency CTI under contract 7608.2 EPRP-IW.

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*, 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.

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

Bear, J. (1979), *Hydraulics of Groundwater*, McGraw-Hill, New York.

Burgers, G., P. J. van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*, 1719–1724.

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

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

De Lannoy, G. J. M., P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest (2006), Assessment of model uncertainty for soil moisture through ensemble verification, *J. Geophys. Res.*, *111*, D10101, doi:10.1029/2005JD006367.

Delta h, Ingenieurgesellschaft mbH (2006), SPRING 3.2, software, Witten, Germany.

de Marsily, G. (1978), De l'identification des systemes hydrogeologiques, 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*, 177–187, doi:10.1016/j.jhydrol.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*, 295–305, doi:10.1016/j.hydrol.2010.12.007.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods in forecast error studies, *J. Geophys. Res.*, *99*(C5), 10,143–10,162.

Fu, J., and J. Gomez-Hernandez (2009), Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking Markov chain Monte Carlo method, *J. Hydrol.*, *364*, 328–341, doi:10.1016/j.jhydrol.2008.11.014.

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

Gomez-Hernandez, J. J., H. J. Hendricks Franssen, and E. F. Cassiraga (2001), Stochastic analysis of flow response in a three-dimensional fractured rock mass block, *Int. J. Rock Mech. Min. Sci.*, *38*(1), 31–44.

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*, W09408, doi:10.1029/2007WR006505.

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

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

Hendricks Franssen, H. J., J. J. Gómez-Hernández, and A. Sahuquillo (2003), Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data, *J. Hydrol.*, *281*, 281–295, doi:10.1016/S0022-1694(03)00191-4.

Hendricks Franssen, H. J., A. Alcolea, M. Riva, M. Bakr, N. van de Wiel, F. Stauffer, and A. Guadagnini (2009), A comparison of seven methods for the inverse modelling of groundwater flow: Application to the characterisation of well catchments, *Adv. Water Resour.*, *32*, 851–872, doi:10.1016//j.advwatres.2009.02.011.

Hernandez, A. F., S. P. Neuman, A. Guadagnini, and J. Carrera (2003), Conditioning mean steady state flow on hydraulic head and conductivity through geostatistical inversion, *Stochastic Environ. Res. Risk. Assess.*, *17*, 329–338, doi:10.1007/s00477-003-0154-4.

Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, *126*, 796–811.

Huber, E., H. J. Hendricks Franssen, H. P. Kaiser, and F. Stauffer (2011), The role of prior model calibration on predictions with ensemble Kalman filter, *Ground Water*, doi:10.1111/j.17456584.2010.00784.x, in press.

Kempf, T., M. Freimoser, P. Haldimann, V. Longo, E. Müller, C. Schindler, G. Styger, and L. Wyssling (1986), *Die Grundwasservorkommen im Kanton Zürich*, Beiträge Geol. Schweiz, Kümmerly and Frei AG, Bern.

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

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

Liu, G., Y. Chen, and D. Zhang (2008), Investigation of flow and transport processes at the MADE site using ensemble Kalman filter, *Adv. Water Resour.*, *31*, 975–986, doi:10.1016/j.advwatres.2008.03.006.

Medina, A., and J. Carrera (1996), Coupled estimation of flow and solute transport parameters, *Water Resour. Res.*, *32*(10), 3063–3076.

Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Houser (2005), Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, *28*, 135–147, doi:10.1016/j.advwatres.2004.09.002.

Naevdal, G., L. V. Johnsen, S. I. Aanonsen, and E. H. Vefring (2003), Reservoir monitoring and continuous model updating using ensemble Kalman filter, SPE paper 84372 presented at SPE Annual Technical Conference and Exhibition, Soc. of Pet. Eng.

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

Oliver, D. S., L. B. Cunha, and A. C. Reynolds (1997), Markov chain Monte Carlo methods for conditioning a permeability field to pressure data, *Math Geol.*, *29*(1), 61–91.

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

Renard, P. (2007), Stochastic hydrogeology: What professionals really need?, *Ground Water*, *45*, 531–541, doi:10.1111/j.1745-6584.2007.00340.x.

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 porous media, *Water Resour. Res.*, *36*(12), 3567–3580.

Sahuquillo, A., J. E. Capilla, J. J. Gómez-Hernández, and J. Andreu (1992), Conditional simulation of transmissivity fields honoring piezometric data, in *Hydraulic Engineering Software IV, Fluid Flow Modeling*, edited by W. R. Blain and E. Cabrera, pp. 201–214, Kluwer, Dordrecht, Netherlands.

Scietec Flussmanagement GmbH (2000), FLORIS2000, software, Linz, Austria.

van Genuchten, M. T. (1980), A closed form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, *44*, 892–898.

Verbunt, M., A. Walser, J. Gurtz, A. Montani, and C. Schär (2007), Probabilistic flood forecasting with a limited-area ensemble prediction system: Selected case studies, *J. Hydrometeorol.*, *8*(4), 897–909, doi:10.1175/JHM594.1.

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

Wen, X.-H., and W. H. Chen (2006), Real-time reservoir updating using ensemble Kalman filter: The confirming approach, *SPE J.*, *11*(4), 431–442, doi:10.2523/92991-MS.

Wen, X.-H., C. V. Deutsch, and A. S. Cullick (2003), Construction of geostatistical aquifer models integrating dynamic flow and tracer data using inverse technique, *J. Hydrol.*, *255*, 151–168, doi:10.1016/S0022-1694(01)00512-1.

Yeh, T.-C. J., J. Xiang, R. M. Suribhatla, K.-C. Hsu, C.-H. Lee, and J.-C. Wen (2009), River stage tomography: A new approach for characterizing groundwater basins, *Water Resour. Res.*, *45*, W05409, doi:10.1029/2008WR007233.

G. Bauser, W. Kinzelbach, F. Stauffer, Institute of Environmental Engineering, ETH Zurich, Wolfgang-Pauli-strasse 15, CH-8093 Zurich, Switzerland.

H. J. Hendricks Franssen, Agrosphere, IBG-3, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany. (h.hendricks-franssen@fz-juelich.de)

H. P. Kaiser, R. Müller, Water Supply of Zurich, Hardhof 9, CH-8023 Zurich, Switzerland.

U. Kuhlmann, TK Consult, Siewerdtstrasse 7, CH-8050 Zurich, Switzerland.