ground water

# The Role of Prior Model Calibration on Predictions with Ensemble Kalman Filter

by E. Huber[1], H.J. Hendricks-Franssen[2], H.P. Kaiser[3], and F. Stauffer[1]

## Abstract

This paper, based on a real world case study (Limmat aquifer, Switzerland), compares inverse groundwater flow models calibrated with specified numbers of monitoring head locations. These models are updated in real time with the ensemble Kalman filter (EnKF) and the prediction improvement is assessed in relation to the amount of monitoring locations used for calibration and updating. The prediction errors of the models calibrated in transient state are smaller if the amount of monitoring locations used for the calibration is larger. For highly dynamic groundwater flow systems a transient calibration is recommended as a model calibrated in steady state can lead to worse results than a noncalibrated model with a well-chosen uniform conductivity. The model predictions can be improved further with the assimilation of new measurement data from on-line sensors with the EnKF. Within all the studied models the reduction of 1-day hydraulic head prediction error (in terms of mean absolute error [MAE]) with EnKF lies between 31% (assimilation of head data from 5 locations) and 72% (assimilation of head data from 85 locations). The largest prediction improvements are expected for models that were calibrated with only a limited amount of historical information. It is worthwhile to update the model even with few monitoring locations as it seems that the error reduction with EnKF decreases exponentially with the amount of monitoring locations used. These results prove the feasibility of data assimilation with EnKF also for a real world case and show that improved predictions of groundwater levels can be obtained.

## Introduction

Predictions of groundwater flow and mass transport in the subsurface are strongly affected by uncertain properties like hydraulic conductivity. In general, only a limited amount of hydraulic conductivity data is available because the drilling of a borehole is an expensive and invasive technique that disturbs the investigated medium. An alternative is noninvasive geophysical techniques (Kowalsky et al. 2004; Rubin and Hubbard 2005; Rings et al. 2010), but these techniques do not yet allow the determination of hydraulic conductivities at large scales. An alternative way to reduce the uncertainty of groundwater flow and mass transport predictions is the incorporation of hydraulic head data. These allow an improved characterization of aquifer properties such as, for example, hydraulic conductivities (de Marsily 1978; Kitanidis and Vomvoris 1983; Carrera and Neuman 1986; La Venue et al. 1995; Gómez-Hernández et al. 1997; Yeh and Liu 2000; Tonkin and Doherty 2005; Alcolea et al. 2006). Recent overviews are given by Carrera et al. (2005) and Hendricks-Franssen et al. (2009).

Inverse modeling techniques are used to calibrate model parameters (e.g., hydraulic conductivity) by incorporating historical data. However, calibrated groundwater flow models deviate from reality if these models are not corrected with new measurement data in real time. This occurs because inverse modeling only improves the

[1]Institute of Environmental Engineering, ETH Zurich, Wolfgang-Pauli-Strasse 15, 8093 Zurich, Switzerland.

[2]Corresponding author: Agrosphere, IBG-3, Forschungszentrum Juelich GmbH, 52425 Juelich, Germany; h.hendricks-franssen@fz-juelich.de

[3]Water Supply of Zurich (WVZ), Hardhof 9, 8023 Zurich, Switzerland.

Received April 2010, accepted December 2010.



doi: 10.1111/j.1745-6584.2010.00784.x

characterization of aquifer properties to a limited extent because of uncertain boundary conditions and forcings and because the aquifer is subject to different flow situations that might not have occurred in the calibration period. Inverse calibrated groundwater models are less suited to deal with on-line measurements and with time-varying parameters in the context of real-time forecasting. Parameters need to be recalibrated over historical time series and the CPU time needed for the reassessment of model parameters is in general elevated, particularly for large coupled surface-subsurface models.

For improved predictions in real time it is important to update models calibrated using inverse methods with new measurement data that become available in real time. Here it is proposed to do this with the help of the ensemble Kalman filter (EnKF) (Evensen 1994; Burgers et al. 1998). The model states are updated with measurements and the prediction uncertainty can be assessed from an ensemble of stochastic realizations. Data assimilation techniques can indeed incorporate different sources of uncertainty in a more flexible manner than inverse models (Hendricks-Franssen and Kinzelbach 2008, 2009). The EnKF seems to be the most adapted approach for a large-size hydrological problem. For small-size linear problems, numerical models can be updated with the classical Kalman filtering in real time (Van Geer et al. 1991). The extended Kalman Filter is able to handle nonlinearities, but is not very suited for large-scale problems (Evensen 1992).

The intent of this paper is to evaluate how the real-time assimilation of hydraulic head data (with EnKF) improves groundwater model predictions. This analysis will be made for models calibrated using inverse methods that are already conditioned to different amounts of (historical) information. Therefore, the capacity of EnKF to improve model predictions will be evaluated as a function of the amount of historical head information contained in the calibration.

An important aspect of this study is that it is based on data from a complex real world case study as almost all published studies relied on synthetic cases (Chen and Zhang 2006; Hendricks-Franssen and Kinzelbach 2008, 2009; Nowak 2009), which in most cases implied that all sources of uncertainty, except the parameter uncertainty, were excluded from the analysis. Liu et al. (2008) applied the EnKF in a real world application, namely on the MADE site. Furthermore, our study includes surface-groundwater interactions, which is new as compared to studies published before, and a verification of the data assimilation is provided based on a large number of monitoring stations. This study uses a three-dimensional (3D) groundwater flow model of a strongly heterogeneous aquifer located in the Limmat Valley in Switzerland. Eighty-seven monitoring hydraulic head locations allow to run several scenarios with specified sets of monitoring locations. The hydraulic head predictions obtained with models calibrated using inverse methods and the prediction improvement brought by the EnKF are evaluated in relation to the amount of monitoring locations used for the calibration and/or the data assimilation. The three studied types of scenarios represent an increase of the a priori historical information used in the model calibration. In the first scenario (M1), the groundwater flow model is run with a uniform hydraulic conductivity estimated from previous knowledge (i.e., averaging the results from different small-scale pumping tests). For the second scenario, a steady-state inverse calibration is performed using all the available knowledge (scenario M2). For the third scenario, several transient inverse calibrations are run using different amounts of monitoring locations (scenario M3). All these scenarios are then updated with the EnKF using alternatively 0, 5, 20, 43, and 85 monitoring locations.

## Case Study and Measurement Data

The study site (about $6.5 \times 1$ km) is located in the Limmat Valley in Zurich (Switzerland, Figure 1). It is delimited by the Sihl River at its south-eastern part and by the Limmat River at its northern part. The modeled domain is limited by the aquifer edges except for the western boundary. The public water supply of Zurich is located close to the river Limmat.

### Hydrogeology

Information on the aquifer properties is available from a large number of boreholes. The aquifer consists of Sihl brash and moraine matter from Würm glaciers and is characterized by a high heterogeneity. The upper layer consists mainly of gravel and has a high permeability. With increasing depth the Limmat Valley brash becomes relatively less permeable and sandier, with less gravel and more silt. The bedrock is made up of molasse. The thickness of the aquifer varies up to 70 m with an average around 20 m.

The aquifer has no hydraulic connection to Lake Zurich. Infiltration from the rivers Limmat and Sihl is the main source of groundwater. The river-aquifer interaction occurs at different scales and is time and space dependent. Doppler et al. (2007) showed that the leakage coefficient is temporally variable related to flood events. They also suggested that the temperature dependence of the leakage coefficient has to be considered for modeling river-aquifer interactions. Engeler et al. (2011) confirmed this by comparing coupled simulations of 3D groundwater flow and heat transport with uncoupled simulations. The temporal variations of the leakage coefficient are neglected in this study. The recharge due to precipitation is small, because a large part of the model domain is urbanized with mainly impervious area, and there is some lateral inflow from the hills located north and south of the study area. The Limmat River is first impounded at the junction point with the Sihl River and also near to the location of the Hardhof Water Works (Hoengg weir). These dams increase locally the groundwater-surface water exchange.

Usually the groundwater flows parallel to the valley axis. Along the Limmat River water exchange between groundwater and surface water occurs. This causes

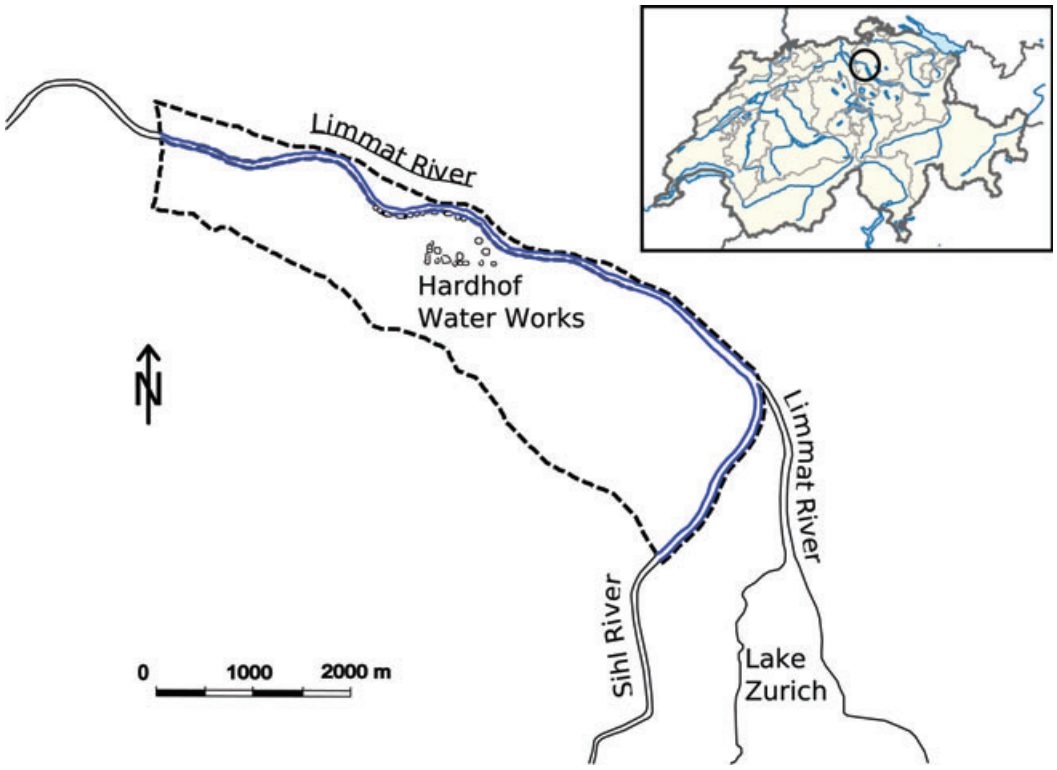


**Figure 1. The small embedded figure shows Switzerland and the location of the study area (black circle). Otherwise, the study site (black lines) corresponds to the boundaries of the simulated area in the upper Limmat Valley in Zurich (Switzerland). The Sihl and Limmat Rivers are delimited by the blue lines. The points in the mid-part of the site represent the wells and infiltration basins of the Hardhof Water Works.**

deviations in the flow direction. The groundwater surface gradient varies between 1‰ and 6‰ (Kempf et al. 1986).

### Hardhof Water Works

About 20% of the drinking water of Zurich is pumped at the Hardhof Water Works. Figure 2 shows a scheme of the Hardhof Water Works, the water supply of Zurich. Water is pumped in 19 bank filtration wells along the Limmat River (10 m from the bank) and re-infiltrated by means of three infiltration basins and 12 wells. The 12 recharge wells have an individual capacity of 12,000 L/min. This artificial recharge improves the groundwater quality through filtration and bioactivity in the soil. The artificially recharged water also helps to create a hydraulic barrier whose goal is to minimize the risk of contamination in case of accidents on the motorway or on the railway (oil and chemical transport) and to divert away contaminated groundwater from the pumping wells. Groundwater below the city center is contaminated due to historical dump sites. The drinking water is pumped by four production wells with an individual maximum capacity of 36,000 L/min.

### Available Data

A dense groundwater monitoring network (87 monitoring locations of hydraulic head, most of them located in the area of the Hardhof Water Works; Figure 3) is available in the study area. Daily mean groundwater hydraulic heads are recorded with an estimated error standard deviation of 5 cm. This error was estimated by practitioners, taking into account sinking of the borehole, errors in the reference height, and small errors in the measurement itself. As convention the locations used for model calibration or real-time data assimilation are called *measurement locations* and the locations only used to evaluate the model predictions are called *control locations*.

Also daily amounts of pumped and artificially recharged water in each of the basins and wells in the Hardhof area, daily meteorological data from Zurich-Affoltern, and daily mean discharge values of the rivers Sihl and Limmat were available. The mean precipitation at Zurich-Affoltern, averaged over the period 1961 to 1990 is 1042 mm/year (MeteoSchweiz 2009). The average discharge of the Sihl River is 6.81 $m^3$/s and for the Limmat River it is 95.8 $m^3$/s. These averages are calculated from a time series of 70 years (Swiss Federal Office for Environment 2008).

## Methodology

### Model Calibration with SPRING (Inverse Modeling)

For the inverse modeling studies, hydraulic conductivities and leakage coefficients were calibrated with the pilot point approach (de Marsily 1978) as implemented in the software SPRING (Delta h, Ingenieurgesellschaft GmbH 2006) using the Levenberg-Marquard method.

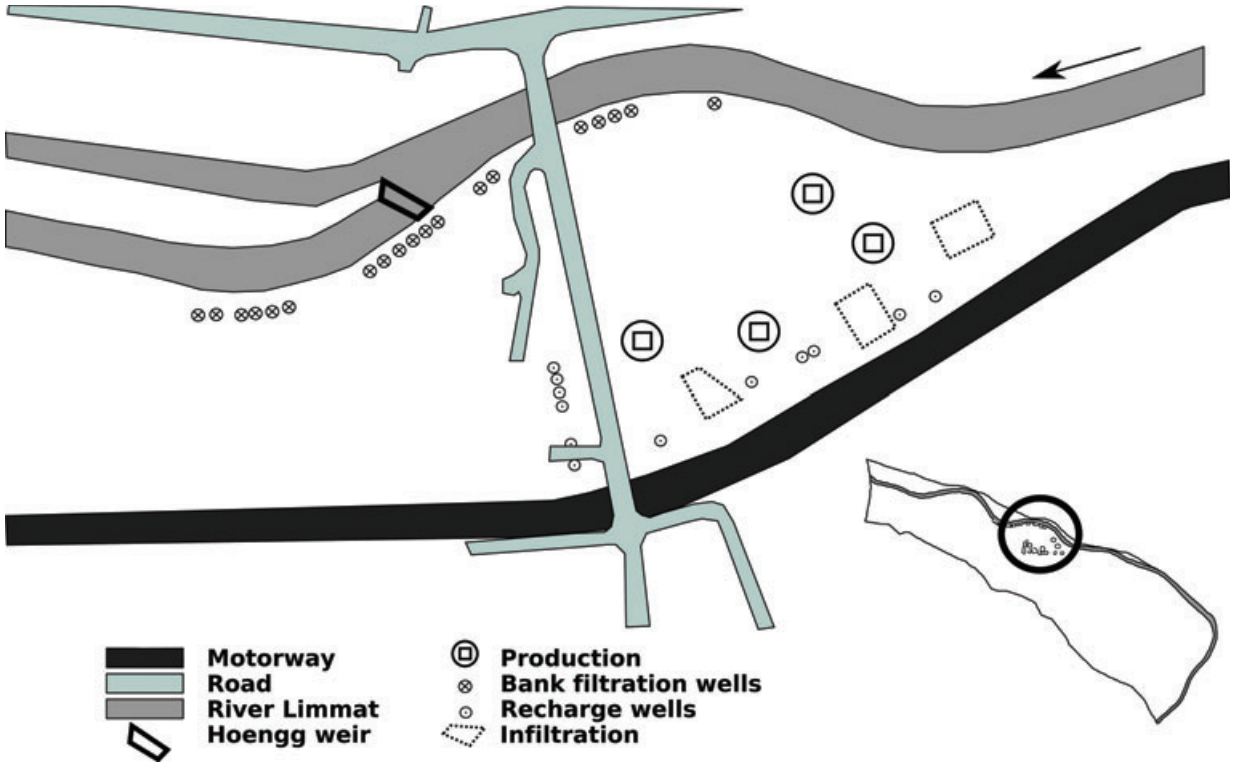


**Figure 2. Scheme of the Hardhof Water Works (from Doppler et al. 2007). The arrow represents the river flow direction.**

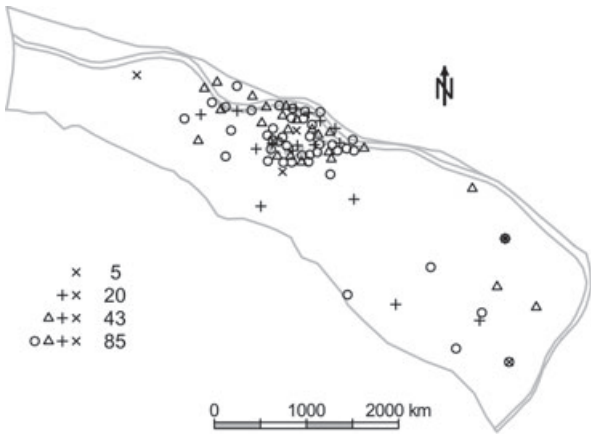


**Figure 3. Location of the sets of measurement locations used for calibration and data assimilation (EnKF). The locations marked by × correspond to the set of 5 measurement locations; the locations marked by + and × to the set of 20 measurement locations; the locations marked by Δ, +, and × to the set of 43 measurement locations; and the locations marked by ○, Δ, +, and × to the set of 85 measurement locations.**

The pilot point method poses the inverse problem as a multiobjective optimization problem where the objective function contains two terms: one for the mismatch of the hydraulic heads and one for differences between prior parameter values and updated parameter values. This objective function is minimized with respect to a limited number of parameters. In our case, these are the hydraulic conductivities at some selected locations, the so-called pilot points, and leakage coefficients for a number of zones. The hydraulic conductivities are not only updated at the pilot points, but also at the other elements of the model on the basis of geostatistical interpolation using a Gaussian variogram. The objective function that has been used includes a plausibility term, whose importance was stressed by Alcolea et al. (2006). It would have been preferable to generate multiple equally likely solutions to the inverse transient groundwater flow problem, for example, with the sequential self-calibration method (SSC) (Gómez-Hernández et al. 1997; Hendricks-Franssen et al. 1999), but this was not possible in this case, as SSC has not been implemented in a software for unsaturated flow conditions and river-aquifer interactions.

### Ensemble Kalman Filter

Data assimilation optimally combines model predictions on one hand and measurement data on the other hand to get updated model predictions that honor recent measurement data. In the EnKF this is done by minimizing the a posteriori error covariance (the error between the predictions and measurements). The EnKF consists of the following steps:

1. The groundwater flow model is used to predict for time step $t = 1$ the groundwater levels. Necessary inputs for these predictions are the initial states (hydraulic heads at $t = 0$), predicted forcings (groundwater recharge, river stages, and lateral inflows), and additional static and dynamic model parameters. The ensemble Kalman method relies on solving the groundwater flow model

many times (100 in our case). In principle, any input parameter or model forcing can be modeled as stochastic. In this study, hydraulic conductivity and leakage coefficient were the uncertain parameters. This was concluded from simulation experiments where the different uncertain parameters (also recharge rate and lateral inflow) were varied in a realistic uncertainty range. The uncertainty of these parameters is quantified with the help of statistical models, in the case of hydraulic conductivity with the help of a multi-Gaussian geostatistical model. With this statistical model different equally likely spatial distributions of hydraulic conductivity and leakage coefficient are generated, which are input to the numerical groundwater flow model. The stochastic realizations of hydraulic conductivity are generated with GCOSIM3D (Gómez-Hernández and Journel 1993) on a very fine grid and upscaled to the size of the finite elements grid using simplified renormalization (Renard et al. 2000). Stochastic realizations of leakage coefficients are generated by sampling randomly and independently for each of five leakage zones from a log-normal distribution of the leakage coefficient. The piezometric head predictions for each of these stochastic realizations will differ. The states covariance matrix **C** can be calculated on the basis of these different head predictions.

2. It is assumed now that at the time step ($t = 1$) measurement data and an estimate of their uncertainties are available. An ensemble of 100 measurement realizations is computed by perturbing the current measurements with a stochastic term drawn from a normal distribution with mean equal to zero and variance equal to the expected measurement uncertainty.
   The measurement data are used to update the model predictions with the following equation:

$$\mathbf{x}^{+} = \mathbf{x}^{0} + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^{0}) \quad (1)$$

   where **x** is a vector containing the piezometric heads at the model nodes. Index "+" stands for an updated vector and index "0" stands for the actual state predictions. Vector **y** contains the measurement data. Matrix **H** relates the modeled heads to the measured heads. Matrix **K** stands for the "Kalman gain" and contains optimized weights of model prediction values and measurement values. The Kalman gain is calculated with

$$\mathbf{K} = \mathbf{C}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \quad (2)$$

   Matrix **R** contains the covariances of the errors of the measurement data. Equation 2 implies the optimal weighting of model-state prediction and measurement values at every model node. All 100 model runs are updated with the help of the Equations 1 and 2.
   The EnKF can also update the parameters such as hydraulic conductivity and leakage coefficient besides the hydraulic head but this was not applied in this study.

3. The results of 100 updated piezometric head distributions are used as initial states for the model runs of the next time step.

The ensemble prediction at each time step may represent the prediction uncertainty. The ensemble size of 100 was chosen to limit the needed CPU time and RAM requirements, although results could have been somewhat better for a larger number of stochastic realizations (Hendricks-Franssen and Kinzelbach 2008).

## Numerical Model

A numerical model for 3D variably saturated groundwater flow including river-aquifer interactions was implemented for the study area. The model was coupled with the software EnKF3d-SPRING (software development by H.-J. Hendricks-Franssen, programmed in C) for assimilation of hydraulic head data and real-time modeling with the EnKF.

The grid has a size of 50 m (in the horizontal direction) and is refined around the wells up to 1 m. The aquifer is discretized into 25 layers of 1.6 m thickness, which resulted in a total of 173,599 finite elements (triangular prisms). The time step is set to 1 day with two iterations per time step.

The model boundaries correspond to the aquifer edge except for the western boundary, which is a time-varying prescribed head boundary condition. The hydraulic head value for this boundary is taken from a monitoring location situated on the boundary. The southern and northern boundaries are prescribed flux boundaries, with fluxes that are variable in space and equal to zero for large parts of the northern boundary. The rivers Sihl and Limmat form the eastern and north-eastern boundaries of the model.

The river is modeled by two lines of leakage nodes. For the period January 2004 to August 2005 (for calibration period, refer section on "Studied Scenarios and Evaluation Measures") the daily mean river stage was computed at each leakage node with the help of the hydraulic software FLORIS (Reichel and Fäh 1995) and used as input for SPRING. The relation between the discharge and river stage (for each of the leakage nodes), as obtained from FLORIS, was fitted with a quadratic regression equation. River stage values for the rest of the simulation period (September 2005 to August 2006; verification period) were obtained from the derived quadratic regression relations and also used as input for SPRING (Doppler et al. 2007). In the simulations presented here, temporal variability of the leakage coefficient was neglected. Five leakage zones were defined, according to hydraulic considerations (e.g., dams, river junction, and bifurcation; Figure 4).

The recharge is computed from the difference between the precipitation and the actual evapotranspiration, the latter being estimated from the potential evapotranspiration and a soil-water balance model. The potential evapotranspiration is calculated with the Penman-Monteith

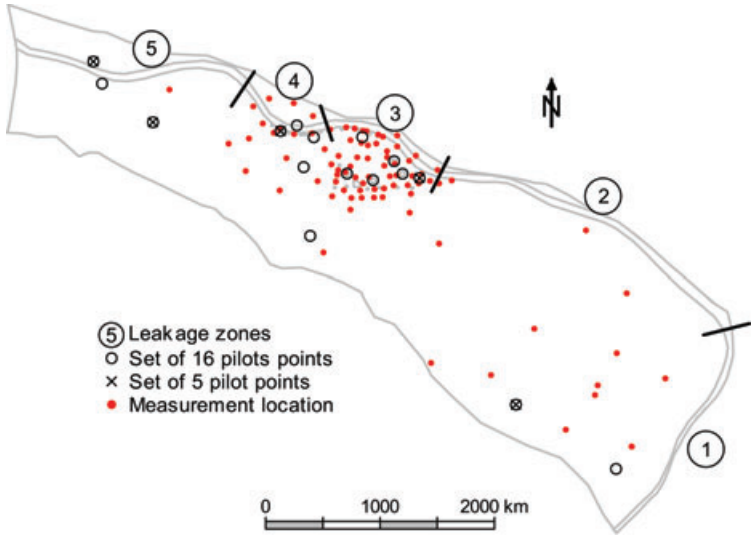


**Figure 4. Position of the five leakage zones, the measurement locations (red dots), the set of 5 pilot points used for the calibration of M3a and M3b (×), and the set of 16 pilot points used for the calibration of M3c (◯) (Table 1).**

method, described in Allen et al. (1998), using meteorological data observed in Zurich-Affoltern as input. The soil-water balance model assumes that the soil is a simple reservoir which is filled by rainfall and from which water can disappear by surface runoff, evapotranspiration, and deep drainage. We estimated the water holding capacity of the soils in the study area from soil-water retention curves (Koorevaar et al. 1983). If the field capacity is exceeded, excess water is drained to the groundwater (recharge). Depending on the soil moisture content, the actual evapotranspiration might be (much) smaller than the potential evapotranspiration, this is especially the case if the soil moisture content approximates the wilting point (Allen et al. 1998). The recharge is also used to determine the lateral inflows from the surrounding hills under the assumption that they are caused by excess precipitation (precipitation minus actual evapotranspiration), as shown by Doppler et al. (2007).

The next section describes how hydraulic conductivity and leakage coefficient were estimated for each scenario using inverse modeling or limited prior knowledge.

## Studied Scenarios and Evaluation Measures

### Scenarios and Calibration

Five different scenarios, with different amounts of historical information, are considered. Table 1 summarizes the characteristics of these calibration scenarios.

In the first scenario M1, we assumed that enough information (geology, pumping, and slug tests) was available to estimate an average $K$ value reliably and that also roughly correct estimates of leakage coefficients were possible, but that no model calibration could be made because the number of hydraulic head data is too limited. The uniform hydraulic conductivity value $K$ for M1 was $2.4 \times 10^{-3}$ m/s and the estimated leakage coefficients are given in Table 2.

For the second scenario M2, the model was calibrated at steady state with hydraulic heads from 80 measurement locations averaged over March 2001, 110 pilot points, and 5 leakage zones. The pilot points were laid out in correspondence with the available information on hydraulic heads, with a denser network in the Hardhof area and fewer points in the rest of the study area. The model

**Table 1**
**Overview of the Different Simulation Scenarios, Number of Measurement Locations (87 Are Available), Total Number of Data (Number of Measurement Locations × the Number of Data Available Over Time), Pilot Points, and Leakage Zones Used for the Calibration**

| Scenarios | Calibration | Measurement Locations | Total Number of Data | Pilot Points | Leakage Zones |
|---|---|---|---|---|---|
| M1 | No calibration | — | — | — | — |
| M2 | Steady state | 80 | 80 | 110 | 5 |
| M3a | Transient | 5 | 330 | 5 | 5 |
| M3b | Transient | 20 | 1320 | 5 | 5 |
| M3c | Transient | 87 | 5742 | 16 | 5 |

**Table 2**
**Leakage Coefficients ($10^{-3}$ m/s) of the Five Leakage Zones for the Different Scenarios (After Calibration for M2, M3a, M3b, and M3c)**

| | Leakage Zones | | | | |
|---|---|---|---|---|---|
| **Scenarios** | **1** | **2** | **3** | **4** | **5** |
| M1 | 0.0010 | 0.1000 | 0.1000 | 1.0000 | 0.0100 |
| M2 | 0.0006 | 0.0360 | 0.0429 | 0.5400 | 0.0049 |
| M3a | 2.5180 | 0.2820 | 0.0300 | 0.3430 | 0.0750 |
| M3b | 9.3710 | 0.9670 | 0.1050 | 0.0180 | 0.0180 |
| M3c | 0.0430 | 0.0309 | 0.1322 | 1.1045 | 0.0036 |

forcings for the calibration of M2 (recharge, river stages, lateral inflows, pumping, and artificial recharge) were also computed to be representative of the steady-state conditions. This was achieved by calculating long-term average recharge and lateral inflow on the basis of the same procedure as outlined before. Rivers stages, pumping, and artificial recharge were determined by averaging the calculated (in case of the river stages) or the measured (in case of the pumping and artificial recharge) daily values over March 2001. As explained before, the calibration was carried out using the pilot point method and the reproduction of the head was balanced with the perturbation of the hydraulic conductivity and leakage coefficient (Carrera and Neuman 1986; Alcolea et al. 2006).

The model was also calibrated in transient state (scenario of type M3) with different sets of measurement locations and pilot points (Table 1). The model is calibrated only with hydraulic heads of June 2004 and July 2005, but we will refer to the period January 2004 to August 2005 as calibration period. It was found that the 2-month calibration period also exerts a strong control on the correct reproduction of the hydraulic heads between these calibration months and also on the period between the (deterministic) initial conditions and the first calibration month. These months were chosen as they are representative of the hydraulic and management conditions of the study site. They include a high water table period caused by a flood event in June 2004 as well as a low water table due to high pumping rates in the production wells in July 2005. In the calibration months, also larger periods with average conditions occur. Also in this case the calibration was done with the pilot point method, balancing again the reproduction of the hydraulic heads and the modification of the parameters according to the variances of the states and parameters.

### Simulation and EnKF

EnKF uses as input stochastic realizations of the hydraulic conductivity and leakage coefficient. Each of the scenarios uses a different set of stochastic realizations, consistent with the calibration results. For M1, unconditional stochastic realizations of hydraulic conductivity and leakage coefficient are generated. For the models M2 and M3, stochastic realizations of hydraulic conductivity and leakage coefficient are generated on the basis of the calibrated values (obtained with the pilot point method) plus a perturbation. The perturbations are created on the basis of the posterior statistics from the calibration.

For each of the scenarios the model was first run in transient state from January 1, 2004 until August 31, 2005 (calibration period, 609 days) using the same forcings for all models as described in the section on "Numerical Model." The hydraulic heads at the end of the simulation over the calibration period were used as starting heads for the simulation in the verification period (from September 1, 2005 to August 31, 2006, 365 days). The EnKF assimilates daily new hydraulic head data and predictions are made with the help of the 100 realizations for the next day (the 99 stochastic realizations and the model calibrated with the pilot point method) as described in the section on "Methodology." In these experiments the EnKF only updates the states (i.e., the hydraulic heads) using different amounts of measurement data as a function of the scenarios. For example, M3b_20 is the scenario M3b that assimilates daily 20 measurement data with the EnKF (see Table 3 for an overview of the runs with EnKF).

Four sets of measurement data containing 5, 20, 43, or 85 measurement locations were chosen in a way that they are spread out over the whole area. Like the original data set with 85 data, these measurement locations focus on the Hardhof area, with fewer samples in the rest of the model domain. The Hardhof area is most strongly affected by changes in groundwater level due to the pumping and artificial recharge activities (Figure 4).

### Model Evaluation

The scenarios are evaluated in terms of the characterization of the hydraulic heads with the deterministic model. The following deviance measures were evaluated: the MAE, the mean error (ME), the Nash-Sutcliffe efficiency criterion (NSE) (Nash and Sutcliffe 1970), and the hydrological deviation ($D$) according to Schultz (1967) (Equation 3). These deviance measures can be computed for all the considered monitoring locations at one time

**Table 3**
**Overview of the Sequential Data Assimilation Scenarios for the Verification Period with the Different Number of Hydraulic Head Measurement Locations**

| No. of Measurement Locations | M1 | M2 | M3a | M3b | M3c |
|---|---|---|---|---|---|
| 0 | M1_00 | M2_00 | M3a_00 | M3b_00 | M3c_00 |
| 5 | M1_05 | — | M3a_05 | — | — |
| 20 | M1_20 | — | — | M3b_20 | — |
| 43 | M1_43 | — | — | — | M3c_43 |
| 85 | M1_85 | M2_85 | — | — | M3c_85 |

Note: The entries in the table are the names for the different simulation scenarios.

step $t$ [notation MAE($t$)] or for one particular location $h$ over all the time steps [notation MAE($h$)]. They can be averaged over all the locations and all the time steps (notation: average MAE = AMAE, average ME = AME, average NSE = ANSE, average$D$ = AD).

$$D = 200\frac{\sum_{i=1}^{N} |h_i - \hat{h}_i|\hat{h}_i}{N \cdot \max(\hat{h}_i)} \quad (3)$$

where $h_i$ represents a measurement and $\hat{h}_i$ a model prediction.

The NSE criterion indicates how much of the variability in observations is explained by the simulations. NSE varies between 1 and minus infinity, where NSE = 1 represents a perfect fit. Because of the squared differences, however, NSE is exceedingly sensitive to extreme values. The Schultz criterion varies between 0 (best fit) and infinity. It is best used for event analyses at fine temporal resolution (e.g., daily resolution).

It is of interest to assess how much the assimilation of new measurement data with EnKF can reduce the prediction error within a scenario. For this reason, the relative ($\Delta_r$AMAE, Equation 4) and absolute ($\Delta_a$AMAE, Equation 5) variations of the averaged mean absolute error (AMAE) between a scenario run with data assimilation and an unconditional run (without data assimilation) were computed as follows:

$$\Delta_r\text{AMAE(Mx\_yy)} = \frac{\text{AMAE(Mx\_yy)} - \text{AMAE(Mx\_00)}}{\text{AMAE(Mx\_00)}} \quad (4)$$

$$\Delta_a\text{AMAE(Mx\_yy)} = \text{AMAE(Mx\_yy)} - \text{AMAE(Mx\_00)} \quad (5)$$

with AMAE(Mx_yy) the AMAE for the scenario Mx updated with hydraulic heads from yy measurement locations.

## Results and Discussion

### Results for the Calibration and Verification Period for Each Scenario Without Data Assimilation

We start with presenting results for the calibration period and the verification of the calibrated models for the additional simulation period of 1 year (without real-time updating of the model). The results of the model predictions cannot be interpreted solely as a function of the number of measurement locations because the results of the calibration (different parameters) and calibration characteristics (pilot points) play a role as well.

Figure 5 displays the spatial distributions of the hydraulic conductivities for layer 5 (at 7.2 m depth) of the calibrated models. All models show a spatially smooth distribution of hydraulic conductivity, but M2 shows a wider range of hydraulic conductivities and more important local variations, probably related to instabilities in the calibration. From M3a to M3c, the standard deviation of the calibrated hydraulic conductivities increases. This is related to the fact that the model M3c uses more conditioning data and therefore better captures the spatial variability.

The averaged Nash-Sutcliffe efficiency (ANSE), the averaged Schultz deviation criterion (AD), and the AMAE provide the same model ranking with respect to the ME (Table 4). On average all the scenarios overestimate the hydraulic heads (averaged mean error [AME] negative) except M3b and M3c for the calibration period. For the calibration period, the ranking of the scenarios (from the best to the worst) is M3c > M3b > M3a > M1 > M2. For the verification period the ranking changes only for M3a and M1 (M1 > M3a). The model calibrated in transient state with much information (scenario M3c) shows the best results and has almost a perfect compensation of the residuals. M2 and M3a exhibit much larger errors for the verification period (compared with the calibration period), whereas for the other scenarios the errors in the verification are not so much larger than during the calibration period.

The model M2, calibrated in steady state, has a worse performance than the noncalibrated model M1 (except for AME), which was not calibrated and used a uniform hydraulic conductivity. This is, in particular, the case for the verification period. The choice of the uniform hydraulic conductivity value for the model M1 and the input data and the number/location of the pilot points for the steady-state calibration of M2 influence the results. However, Alcolea et al. (2006) showed that for a properly regularized inverse problem, the number of pilot points can be much larger than the number of measurements, without negative effects on the parameter estimation. However, in this study for M2 some very low calibrated hydraulic conductivities point to the fact that the calibration was maybe not very stable. Therefore, the bad performance of M2 seems to be related to more than just a steady-state flow model. In the literature groundwater flow models are often calibrated in steady state. For dynamic aquifers strongly influenced by interactions with streams, results from a transient calibration seem to be very superior over the ones from a steady-state calibration. The transient changes of the river stage, whose response is observed in the aquifer, provide important information on the hydraulic conductivities of the aquifer (Yeh et al. 2009).

For the calibration and verification periods, the scenarios of type M3 have a smaller error as the number of measurement locations (and the number of pilot points) increases. The transient calibration exhibits better results when more measurement locations are involved, in both the calibration and verification period. The models M3a and M3b present very large calibrated leakage coefficients for the leakage zone 1 (Table 2). In the vicinity of this leakage zone the measurement locations with the largest prediction errors are situated.

Figure 6 illustrates the evolution of MAE with time. It can be seen that from the beginning of the verification period onward MAE is larger than during the calibration period. The MAE also shows stronger

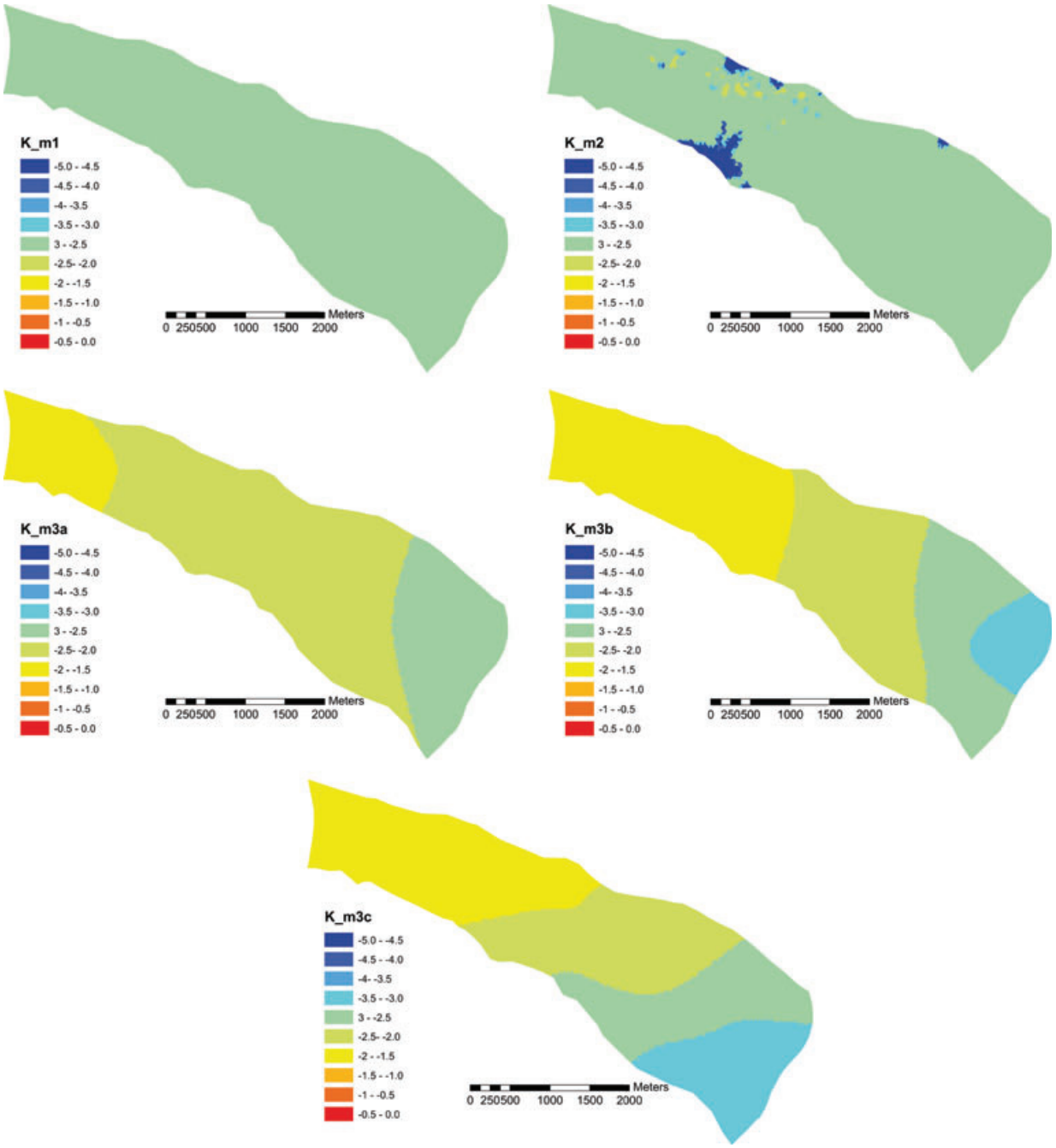


**Figure 5. Spatial distributions of the hydraulic conductivities (at model layer 5 at 7.2 m depth) of the calibrated models.**

variations during the verification period and the difference between the models is larger during the verification period, as compared with the calibration period. This illustrates once more that the calibration quality and prediction quality might be very different (Freyberg 1988).

Only M2 varies already strongly at the end of the calibration period. For the calibration period, M1 and M2 show very similar patterns, whereas for the verification period, M1 shows similar error patterns as M3c (just shifted) while M2, M3a, and M3b also have similar temporal error patterns. The similar temporal error pattern is at least in part related to the assumption of temporally constant leakage coefficients in all models. Other studies (Doppler et al. 2007) showed that the leakage coefficient is temporally variable related to flood events and temperature.

### Results for the Verification Period (Unconditional Runs and EnKF)

It is useful to compare the improvement of data assimilation for the different scenarios. For the results presented here, EnKF was always able to yield better predictions than unconditional model runs. The AMAE decreases exponentially as a function of the number of assimilated data. The strongest decrease is observed for

**Table 4**
**Efficiency Criteria Averaged Over Calibration Period and the Verification Period**

| | ANSE (−) | AD (−) | AME (m) | AMAE (m) |
|---|---|---|---|---|
| Calibration | | | | |
| M1 | 0.76 | 135 | −0.60 | 0.68 |
| M2 | 0.69 | 136 | −0.30 | 0.69 |
| M3a | 0.81 | 105 | −0.38 | 0.53 |
| M3b | 0.89 | 77 | 0.33 | 0.39 |
| M3c | 0.95 | 44 | 0.00 | 0.22 |
| Verification | | | | |
| M1_00 | 0.63 | 173 | −0.81 | 0.87 |
| M2_00 | −0.19 | 297 | −1.19 | 1.50 |
| M3a_00 | 0.30 | 247 | −1.23 | 1.25 |
| M3b_00 | 0.75 | 148 | −0.55 | 0.75 |
| M3c_00 | 0.93 | 71 | −0.18 | 0.36 |

Note: During the verification period no data were assimilated.

M1, the scenario with the largest improvement potential (Figure 7).

The EnKF resulted in stronger improvements if the model calibration was based on few data (that means if the calibrated model performed worse). For models that were calibrated with many data, additional information is needed for the reduction of the prediction errors as compared with models that were calibrated with less data. For the same prediction error reduction $\Delta_r$AMAE of 28%, M1_00 needed to be updated with 5 measurement locations (M1_05), whereas M3_00 needed 43 additional data. However, the model ranking is stable: M3c performs better than M2 and M1 if it uses the same amount of measurement locations for assimilation.

### Evaluation of the Performance of EnKF at Control Locations

To assess if the EnKF also improves the characterization of the hydraulic heads at locations that are

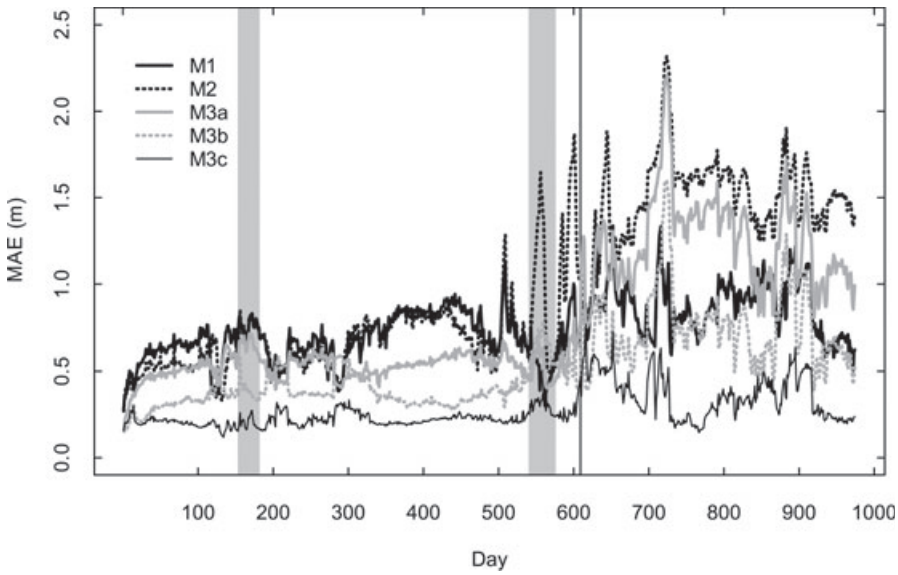


**Figure 6. Variation of MAE($t$) of the calibrated models with time. The calibration period stretches from January 2004 to August 2005 with calibration for the months of June 2004 and July 2005 (represented in the figure with gray bars). The verification period covers the period from September 2005 until August 2006. The beginning of the verification period is represented by a vertical line (Day 609).**

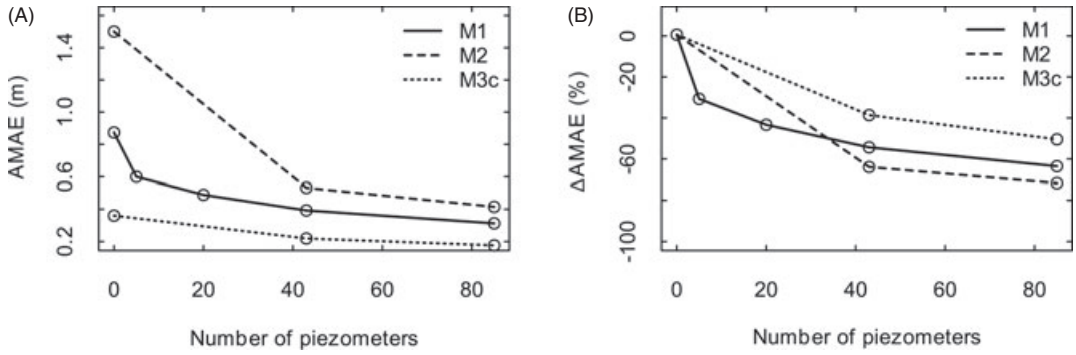


**Figure 7. AMAE (A) and relative variation of AMAE (B) of scenarios M1, M2, and M3c as a function of the number of measurement locations used for data assimilation (verification period). The AMAE is computed using the 85 locations and the AMAE of the unconditional scenarios (M1_00, M2_00, and M3c_00) serve as references for the relative variation of AMAE.**

not used for assimilation (control locations), the relative and absolute variations of MAE($h$) for each location are computed and separately averaged over the measurement locations and the control locations.

The prediction error is reduced by the EnKF not only at the measurement locations but also at the control locations (at least 25% error reduction; Figure 8). On average, the measurement locations (which are used for data assimilation) show a larger improvement in prediction than the control locations. However, this difference in AMAE between the measurement and control locations decreases with the number of measurement locations. For the scenario M1, the difference is 0.10 m for the assimilation of measurement data from 5 locations, 0.08 m for 20 measurement locations, and only 0.03 m for 43 locations. These results show that the EnKF not only updates the hydraulic head successfully at the measurement locations, but also between the measurement locations (at the control locations). If data from more measurement locations are used in the assimilation, the spatial distance between the control locations and measurement locations is on average smaller, and the hydraulic head will be more strongly correlated between verification and measurement locations. As a consequence, in the case of a denser monitoring network the correcting influence of the measurement locations on the control locations is stronger than for a less dense monitoring network. The EnKF does not improve the prediction at all the locations. Few locations (including both measurement and control locations) present a slightly worse prediction with data assimilation. For these locations the EnKF could not find a satisfying solution that handles the measurement value together with the other constraints (from the model and other conditioning data). This might point to an erroneous measurement value or other model errors (e.g., wrong conceptualization of geology and erroneous pumping data). This suspicion is also

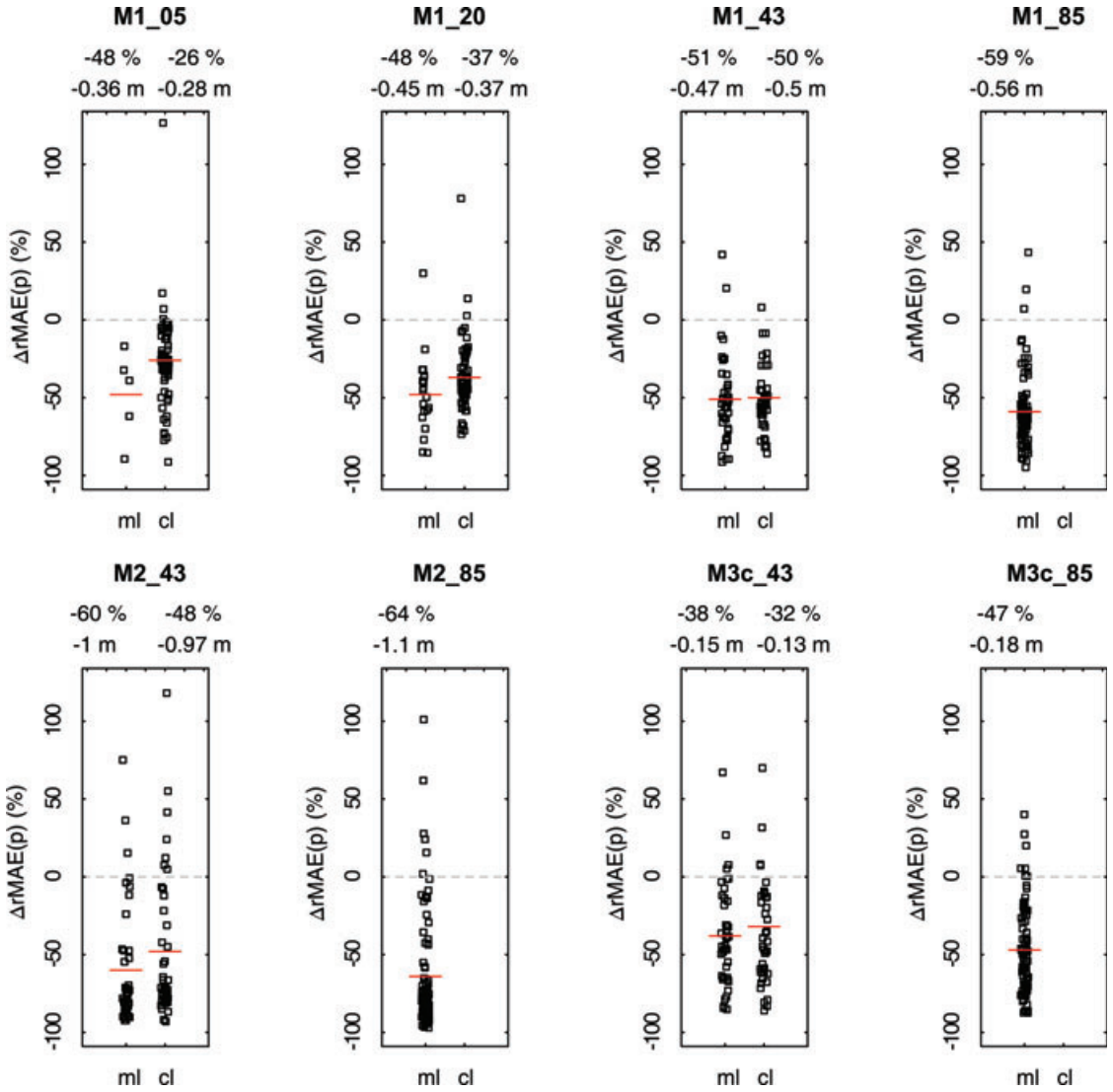


**Figure 8. Relative variation of MAE($h$) [$\Delta_r$MAE($h$)] computed for each location of each scenario over the verification period (EnKF). The mean $\Delta_r$MAE($h$) for the measurement locations (ml) and the control locations (cl) is represented with horizontal red lines. The absolute (m) and relative (%) variation of MAE($h$) averaged over the ml and over the cl are written above the strip chart of ml and cl in each box.**

raised by having a closer look at Figure 8. The calculated MAE($h$) are strongly skewed for almost all scenarios. At most of the measurement points the improvement is considerable, while at a few measurement locations the prediction does not improve or even becomes worse.

The prediction error reductions are smaller than or equal to the ones found in synthetic experiments (Hendricks-Franssen and Kinzelbach 2008, 2009). In those synthetic experiments and for a similar heterogeneity, at control locations an AMAE reduction of around 65% was found; here between 31% (M1_05) and 50% (M1_43). It is a common observation that the results obtained in synthetic experiments are better than the ones found in real world case studies. This is because in real world studies the adopted geostatistical model is not the same as the "true" geostatistical model and maybe the stationarity assumption is not well suited for the study area. Other reasons can be an erroneous model conceptualization, erroneous measurement data, and erroneous data on model forcings.

### Comparison of Inverse Modeling and EnKF

These results show that even if a model is calibrated, it is important to assimilate new measurement data that become available in real time for better predictions. This is illustrated by comparing results for the scenarios M1_05 (no historical calibration, assimilation of data from five locations) and M3a_00 (calibrated with historical data from five measurement locations, no assimilation), as well as M1_20 and M3b_00 (similar, for 20 data). Table 5 summarizes the comparison between calibration and EnKF and presents the AMAE for each of these scenarios. As a reference, M1_00 has an AMAE of 0.87 m. Table 5 illustrates that calibration does not always improve the prediction: five measurement locations (scenario M3a_00) are not enough to catch the variability of the hydraulic heads (43% of error increase, compared with unconditional simulations [M1_00]). The largest prediction errors are found for measurement locations in the south-eastern (very large calibrated leakage coefficients, see earlier discussion) and north-western part of the study site. This reveals the importance of the quality of the calibration: too few measurement locations representative of the hydrological phenomenon result in a poor calibration. The calibration with 20 measurement locations improves the prediction (14% error reduction compared with M1_00). However, a much better relative improvement is achieved with the assimilation of only five measurement locations with EnKF (31% error reduction compared with M1_00). Figure 9 illustrates the temporal fluctuations of MAE($t$) for these different models. The MAE($t$) for the calibration scenarios (M3a_00 and M3b_00) fluctuate around the MAE($t$) of M1_00, whereas the MAE($t$) for the data assimilation (scenarios M1_05 and M1_20) are always smaller than the MAE($t$) of M1_00.

**Table 5**
**Calibration and Updating of M1 with 5 and 20 Measurement Locations**

| | Calibration | Data Assimilation |
|---|---|---|
| 5 measurement locations | M3a_00 AMAE = 1.25 m | M1_05 AMAE = 0.60 m |
| 20 measurement locations | M3b_00 AMAE = 0.75 m | M1_20 AMAE = 0.49 m |

Note: The corresponding scenarios and their AMAE computed over the verification period are given (M1_00 has an AMAE of 0.87 m).

These results illustrate the importance of assimilating data that become available in real time from on-line sensors (Barnhart et al. 2010). Five or 20 measurement locations that are assimilated in real time, for a very simple model with a uniform hydraulic conductivity, are able to yield better predictions of the spatial distribution of hydraulic heads over the aquifer, as compared with models that were calibrated with the same amount of historical measurement data. Note that these inverse calibrated models used a complete time series of hydraulic head data to improve the model and updated parameters such as hydraulic conductivity and leakage coefficients. Here the EnKF was applied in a mode that did not allow the updating of parameters. Hendricks-Franssen et al. (2011) show that updating of hydraulic conductivities and leakage coefficients with the EnKF further improves the model predictions. The update of the leakage coefficient could be a way to integrate its temporal variability due to flood events and temperature dependency. The computation time needed for data assimilation with EnKF (for 100 stochastic realizations) for more than 1 year was 1 week (processor: Intel Core 2 Duo E6600 with 2.40 GHz, operating system: Linux, 3 Gb RAM), whereas the CPU time needed for inverse calibration of a single model for more than 2 months with five pilot points was 2 weeks (processor: Intel Core 2 Duo E6400 with 2.13 GHz, operating system: Windows XP SP3, 2 Gb RAM). A single forward run costs approximately 30 min.

## Conclusions

A variably saturated groundwater flow model, including stream-aquifer interaction, of the Limmat Valley aquifer in Zurich was used to evaluate the hydraulic head prediction errors of a model calibrated using inverse methods and data assimilation with EnKF. These errors were evaluated as a function of the number of measurement locations.

Concerning the calibration, it is highly recommended to perform transient calibration instead of steady-state calibration, for dynamic groundwater systems that show an important interaction with streams. This is well-known from synthetic experiments, but now also clearly shown for a practical case with unknown error sources. A calibration with insufficient measurement locations (e.g., scenario M3a, five measurement locations) can lead to large prediction errors, which appear more accentuated during the verification period.

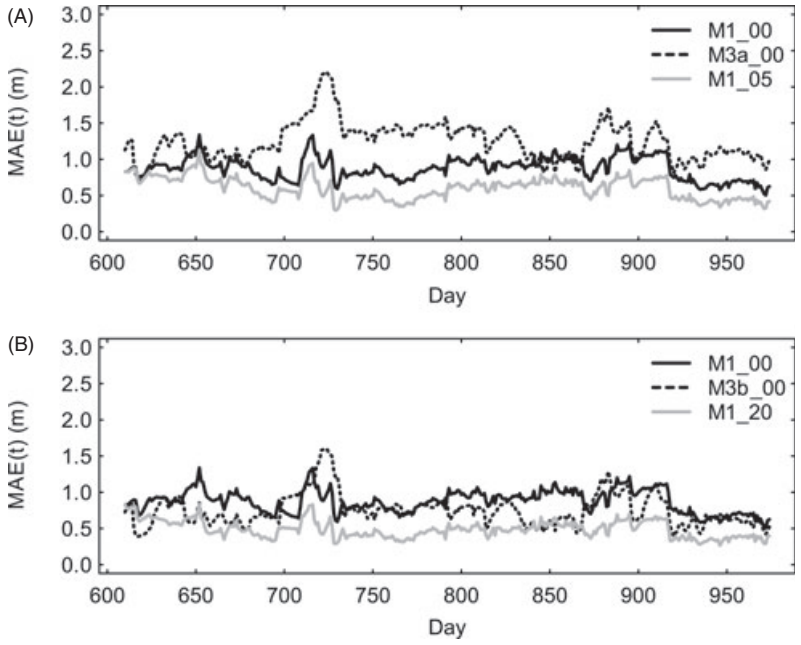


**Figure 9. (A) MAE($t$) over the verification period for M1_00, M3a_00 (=M1 inverse calibrated with five measurement locations), and M1_05 (=M1 updated with EnKF on the basis of five measurement locations). (B) MAE($t$) over the verification period for M1_00, M3b_00 (=M1 inverse calibrated with 20 measurement locations), and M1_20 (=M1 updated EnKF on the basis of 20 measurement locations).**

Real-time hydraulic head data carry important information that is able to improve groundwater flow model predictions, particularly in the context of real-time forecasting. Here the EnKF was used for real-time assimilation of hydraulic head data. This was done for different models that were calibrated with different amounts of historical information. The reduction of 1-day hydraulic head prediction errors with the EnKF (in terms of average MAE) lies within all the scenarios between 31% (scenario M1_05, noncalibrated model updated with 5 measurement locations) and 72% (scenario M2_85, model calibrated in steady state updated with 85 locations).

Even if data from few measurement locations are available, it is worthwhile to use them for data assimilation as the prediction errors seem to decrease exponentially as a function of the number of measurement locations. The improvement achieved by EnKF is larger for models that showed larger prediction errors without assimilation or, equivalently, models that were calibrated on the basis of few data. The improvement by the EnKF was observed at both the assimilation locations and control locations.

The fact that it is important to assimilate measurements available in real time from on-line sensors is further illustrated having a closer look at the simulation results. The models that were historically calibrated with 5 or 20 hydraulic head time series have (much) larger errors for a verification period as compared with an uncalibrated model that is supplied with 5 or 20 hydraulic head data (at the same spatial locations as in the historically calibrated models) in real time that are assimilated with EnKF. These results demonstrate the feasibility of data assimilation with EnKF for a real world case and show that improved predictions of groundwater levels can be obtained. Another interesting aspect of EnKF is that the needed computation time is smaller than for inverse calibration, particularly for large-size problems.

The results shown in this paper illustrate that models calibrated with the help of historical information profit from the assimilation of new measurement data available from on-line sensors. This is particularly true for models that were calibrated with limited measurement data; they profit (in relative terms) more from these on-line data.

## Acknowledgments

We would like to thank Steffen Mehl and three anonymous reviewers who reviewed earlier drafts of this manuscript and contributed significantly to its improvement.

## References

Alcolea, A., J. Carrera, and A. Medina. 2006. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Advances in Water Resources* 29, no. 11: 1678–1689.

Allen, R.G., L.S. Pereira, D. Raes, and M. Smith. 1998. Crop evapotranspiration—Guidelines for computing crop water requirements. *FAO Irrigation and Drainage Paper 56*. http://www.fao.org/docrep/x0490e/x0490e00.htm#Contents (accessed August 12, 2010).

Barnhart, K., I. Urteaga, Q. Han, A. Jayasumana, and T. Illangasekare. 2010. On integrating groundwater transport models with wireless sensor networks. *Ground Water* 48, no. 5: 771–780.

Burgers, G., P.J. van Leeuwen, and G. Evensen. 1998. Analysis scheme in the Ensemble Kalman filter. *Monthly Weather Review* 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 Resources Research* 22, no. 2: 199–210.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L.J. Slooten. 2005. Inverse problem in hydrogeology. *Hydrogeology Journal* 13, no. 1: 206–222.

Chen, Y., and D. Zhang. 2006. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources* 29, no. 8: 1107–1122.

Delta h, Ingenieurgesellschaft GmbH. 2006. Spring 3.31.

Doppler, T., H.-J. Hendricks-Franssen, H.-P. Kaiser, U. Kuhlmann, and F. Stauffer. 2007. Field evidence of a dynamic leakage coefficient for modelling river aquifer interactions. *Journal of Hydrology* 347, no. 1–2: 177–187.

Engeler, I., H.-J. Hendricks-Franssen, R. Müller, and F. Stauffer. 2011. The importance of coupled modelling of variable saturated groundwater flow-heat transport for assessing river-aquifer interactions. *Journal of Hydrology*. DOI: 10.1016/j.hydrol.2010.12.007.

Evensen, G. 1992. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research* 97, no. C11: 17905–17924.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99, no. C5: 10143–10162.

Freyberg, D.L. 1988. An exercise in ground-water model calibration and prediction. *Ground Water* 26, no. 3: 350–360.

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

Gómez-Hernández, J.J., A. Sahuquillo, and J.E. Capilla. 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 1. Theory. *Journal of Hydrology* 1–4, no. 203: 162–174.

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 Resources Research* 44: W09408. DOI: 10.1029/2007WR006505.

Hendricks-Franssen, H.-J., and W. Kinzelbach. 2009. Ensemble Kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems. *Journal of Hydrology* 365, no. 3–5: 261–274.

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. *Advances in Water Resources* 23, no. 1: 1–13.

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. *Advances in Water Resources* 32, no. 6: 851–872.

Hendricks-Franssen, H.-J., H.P. Kaiser, U. Kuhlmann, G. Bauser, R. Müller, F. Stauffer, and W. Kinzelbach. 2011. Operational real-time modeling with EnKF of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. *Water Resources Research*. 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 zur Geologie der Schweiz. Geotechnische Serie.* Bern: Kümmerly & Frey.

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 Resources Research* 19, no. 3: 677–690.

Koorevaar, P., G. Menelik, and C. Dirksen. 1983. *Elements of Soil Physics.* Amsterdam, The Netherlands: Elsevier Science Publishers.

Kowalsky, M.B., S. Finsterle, and Y. Rubin. 2004. Estimating flow parameter distributions using ground-penetrating radar and hydrological measurements during transient flow in the vadose zone. *Advances in Water Resources* 27, no. 6: 583–599.

La Venue, A.M., B.S. RamaRao, G. De Marsily, and M.G. Marietta. 1995. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 2. Application. *Water Resources Research* 31, no. 3: 495–516.

Liu, G., Y. Chen, and D. Zhang. 2008. Investigation of flow and transport processes at the MADE site using ensemble Kalman filter. *Advances in Water Resources* 31, no. 7: 975–986.

de Marsily, G. 1978. De l'identification des systèmes hydrogéologiques (tome 1), Université Pierre et Marie Curie—Paris VI, Ph.D. thesis, 58–130.

MeteoSchweiz. 2009. Bundesamt für Meteorologie und Klimatologie MeteoSchweiz. http://www.meteoschweiz.admin.ch (accessed August 12, 2010).

Nash, J.E., and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models, Part I—A discussion of principles. *Journal of Hydrology* 10, no. 3: 282–290.

Nowak, W. 2009. Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator. *Water Resources Research* 45, no. 4. DOI: 10.1029/2008WR007328.

Renard, Ph., 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 Resources Research* 36, no. 12: 3567–3580.

Reichel, G., and R. Fäh. 1995. Programmpaket FLORIS—Grundlagen und praktische Anwendungen, Seminar Mathematische Modelle offener Gerinne, November 21, 1995, Wien, 33–52.

Rings, J., J.A. Huisman, and H. Vereecken. 2010. Coupled hydrogeophysical parameter estimation using a sequential Bayesian approach. *Hydrology and Earth System Sciences* 14, no. 3: 545–556.

Rubin, Y., and S.S. Hubbard, ed. 2005. *Hydrogeophysics.* Dordrecht, The Netherlands: Springer.

Schultz, G.A. 1967. Die Anwendung von Computer-Programmen für das Unit Hydrograph Verfahren am Beispiel eines Donauzubringers. Paper for the Fifth Conference of the Danube River Countries, Bratislava, CSSR.

Swiss Federal Office for Environment. 2008. Hydrologisches Jahrbuch der Schweiz 2007. No. 0824 in Umwelt-Wissen, Bundesamt für Umwelt, Bern, 576.

Tonkin, M.J., and J. Doherty. 2005. A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resources Research* 41, no. 10: W10412. DOI: 10.1029/2005WR003995.

Van Geer, F.C., C.B.M. Te Stroet, and Z. Yangxiao. 1991. Using Kalman filtering to improve and quantify the uncertainty of numerical groundwater simulations: 1. The role of system noise and its calibration. *Water Resources Research* 27, no. 8: 1987–1994.

Yeh, T.-C.J., and S. Liu. 2000. Hydraulic tomography: Development of a new aquifer test method. *Water Resources Research* 36, no. 8: 2095–2105.

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 Resources Research* 45, no. 5: W05409. DOI: 10.1029/2008WR007233.