Bias-aware data assimilation in integrated hydrological modelling

Marc-Etienne Ridler, Donghua Zhang, Henrik Madsen, Jacob Kidmose, Jens C. Refsgaard and Karsten H. Jensen

ABSTRACT

One of the major challenges in hydrological data assimilation applications is the presence of bias in both models and observations. The present study uses the ensemble transform Kalman filtering (ETKF) method and an observational bias estimation technique to estimate groundwater hydraulic heads. The study was carried out in a relatively complex, groundwater dominated, catchment in Denmark using the MIKE SHE model code. The method is implemented and evaluated using synthetic data and subsequently tested against real observations. The results from the synthetic experiments show that the bias-aware filter outperforms the standard filter, with improved state estimate and correct bias estimate. The assimilation using real observations further demonstrates the robustness of bias-aware ETKF, and the potential improvements using integrated hydrological modelling. Furthermore, the experiments with assimilating over different depths show that the state estimates depend on correlation across layers.

Key words | bias-aware filter, data assimilation, ensemble transform Kalman filter, integrated hydrological modelling, MIKE SHE

Marc-Etienne Ridler (corresponding author)
Henrik Madsen
Karsten H. Jensen
DHI,
Agern Allé 5, Hørsholm 2970,
Denmark
E-mail: mer@dhigroup.com

Donghua Zhang
Department of Geosciences and Natural Resource
 Management,
University of Copenhagen,
Copenhagen,
Denmark

Jacob Kidmose
Jens C. Refsgaard
Geological Survey of Denmark and Greenland
 (GEUS),
Copenhagen,
Denmark

INTRODUCTION

Hydrological models are used extensively to monitor and manage water resources, but are inherently uncertain due to poorly described model physics and parameterization, and imperfect meteorological forcing data and initial conditions. Data assimilation (DA) offers a means to incorporate information from measurements to both correct model forecasts and, importantly, provide quantitative uncertainty estimates useful for decision-makers. One challenge is to assimilate observations that contain bias.

During the past decade, DA methodologies have been developed and applied in groundwater models, almost exclusively in synthetic studies (where the assimilated observation values are taken from a model). Notably, Chen & Zhang (2006) and Hendricks Franssen & Kinzelbach (2008) presented applications of the ensemble Kalman filter (EnKF) to a groundwater flow model updating both state and parameters. Camporese et al. (2009a) applied the EnKF to assimilate pressure head and streamflow in a three-dimensional tilted v-catchment test case, and Camporese et al. (2009b) assimilated synthetic measurements in a small catchment. Sun et al. (2009) compared different types of deterministic EnKFs in a simple conceptual groundwater model. Pasetto et al. (2012) compared EnKF with the particle filter in an integrated groundwater and surface water model. Zhou et al. (2011) and Xu et al. (2013) tested a normal-score EnKF in a synthetic groundwater model. Xue & Zhang (2014) developed a multi-model EnKF using a Bayesian model averaging framework and tested it on a 2D conceptual model. Panzeri et al. (2014) compared the moment equations-based EnKF with

the traditional Monte Carlo-based EnKF in a synthetic 2D flow model. Gharamti *et al.* (2014) developed a hybrid formulation of the EnKF and optimal interpolation for contaminant transport. These synthetic studies have produced novel DA techniques for groundwater models, but have been mainly tested in idealized conditions without the presence of model or observation bias.

A few notable exceptions exist that use real measurements. Zhang *et al.* (2015), in a catchment scale model using an integrated hydrological model, assimilated soil moisture and hydraulic head. The bias in the head observations was pre-processed out such that, on average, it matched model estimates. Hendricks Franssen *et al.* (2011) implemented an operational DA system to monitor groundwater flow and solute transport in the Limmat Valley aquifer in Switzerland. In all test setups, best hydraulic head forecasts were obtained in the synthetic experiments compared to real data. Zovi *et al.* (2017) used the normal-score EnKF to assimilate groundwater levels to update model parameters in a small aquifer in Italy. Again, in the real test case, the inversion approach could not reconstruct the characteristic paleochannel shapes in the model. Kurtz *et al.* (2014) jointly assimilated piezometric head and groundwater temperatures for improving the river–aquifer interactions. They concluded that the 'synthetic setup indicated a potential improvement in parameter estimation, whereas for the real-world setup there is no direct indication that the additional assimilation of groundwater temperatures improves the estimation of unknown hydraulic properties.' These conclusions highlight the need to test DA techniques on real groundwater data that often contains temporal gaps, distinct dynamics, and bias.

Observation bias and model bias normally arise for different reasons, and thus it is important to distinguish between them and to treat them differently. For *in-situ* measurements, biases primarily stem from instrument and scaling errors. For the case of hydraulic head in a grid-based model, scaling errors usually dominate as point measurements are used to represent average values for model grids. For catchment scale models, grid sizes are typically in the order of 100–2,000 m, with potentially significant variation in elevation within the grid cell. For remote sensing observations, observation bias tends to be a consequence of errors in the

retrieval algorithm. A priori rescaling of the observations (e.g., Reichle & Koster 2004) can be applied to remove the bias prior to assimilation. However, the bias in observations may change over time, and therefore online observation bias estimation techniques are usually preferred to dynamically adapt to the transient change of bias (De Lannoy *et al.* 2007). Besides the observation bias, hydrological models are also subject to biased predictions because of the deficiencies in the characteristics of the model (e.g., model structures, model parameters and boundary conditions). All these potential uncertainty sources can contribute to a model forecast error, which, in general, consists of both a random and a systematic component.

There are several approaches to estimate bias using Kalman filter-based methods. Assuming the model is bias-free and the observations are biased, the observation bias parameters can be included in the state vector (Fertig *et al.* 2009; Ridler *et al.* 2014a), and the observation bias estimated online. Under the condition that observations are unbiased, different approaches for model forecast bias estimation have been reported. State augmentation (Drécourt *et al.* 2006; Kollat *et al.* 2008) can be used to include the model errors which are correlated in time. By estimating the correlation between the model state and model error from an ensemble, the spatial distribution of state as well as model error are estimated by the filter. However, model error augmentation increases the computational burden when the state vector size is large, as the dimension of the error covariance matrix increases. Alternatively, other approaches such as 'separate-bias estimation' or 'two-stage estimation' may be applied (Friedland 1969; Drécourt *et al.* 2006). The idea is to estimate the bias separately from the estimation of the biased state. Several variants of this approach using different feedback mechanisms between state update and bias update are compared by De Lannoy *et al.* (2007) in the context of soil moisture assimilation. Another approach that addresses the model bias indirectly is to associate the model bias by the model parameters, which can also be estimated as part of the DA (Moradkhani *et al.* 2005). Recently, Pauwels *et al.* (2013) included the observation bias into a 'two-stage estimation' approach and applied it to a simple rainfall–runoff model, enabling partitioning of the difference between observations and forecasts into

estimation of state variables, forecast bias and observation bias separately.

This study builds on the concepts presented in Rasmussen *et al.* (2015), who applied online bias estimation using the ensemble transform Kalman filter (ETKF) to estimate hydraulic head bias in an integrated hydrological model. The experiments were done in a simple two-dimensional model using real data to estimate saturated zone parameters, hydraulic head, head bias and stream discharge. Online bias estimation showed potential as an effective means to quantify head bias, but because parameters were also updated, it introduced equifinality issues. The filter could account for the discrepancy between model and observation values in two ways: by estimating the observation bias, or by altering the model parameters values. Consequently, the results were not consistent between experiments in terms of the parameter set values and bias calculated.

This study systematically investigates the performance of the bias-aware ETKF by assimilating both synthetic and real data in a complex three-dimensional integrated hydrological model. To avoid equifinality issues, parameter estimation is not included. Particular focus areas are (1) evaluating corrections across layers in the saturated zone, (2) to what extent the filter updates can be attributed to the state and observation bias, and (3) the impact model corrections have on stream discharge.

METHODOLOGY

Study area and hydrological model

The 1,044 km^2 Ahlergaarde catchment in Western Denmark (Figure 1) is characterized by relatively low elevation gradients with sandy and gravelly surface soil. Approximately 61% of the total area is agricultural land. A detailed description of the catchment and geology is given in Kidmose *et al.* (submitted). The stream is predominantly groundwater fed due to the high permeability of the surface soils.

The catchment is modelled in MIKE SHE, a deterministic, physically based, distributed modelling system, which simulates the major hydrological processes including evapotranspiration, overland flow, unsaturated flow, groundwater flow, channel flow and their interactions (Abbott *et al.* 1986; Graham & Butts 2005). Groundwater flow in 3D is simulated by solving the Bousinesq equation, which is the combination of mass conservation and Darcy's law. Drain flow (pipes/ditches) is considered when the groundwater table exceeds drain level; a 1D channel flow model based on kinematic routing (in MIKE 11) is used for simulation of river flow; a 2D diffusive wave approximation of the Saint Venant equations is applied for overland flow routing; a simplified two-layer approach is used to simulate unsaturated flow and evapotranspiration (see Graham &

Figure 1 | Ahlergaarde catchment. 'Obs Head' and 'Obs Q' are groundwater head and discharge observations, respectively.

Butts (2005) for descriptions of the different model components).

The model is horizontally discretized in a 200×200 m^2 grid, providing sufficient resolution to incorporate geological stratigraphy, land use and topography from finer resolution maps. The saturated zone is further discretized in six vertical stratigraphic layers based on geographic surveys, summarized in Table 1.

Spatially distributed initial heads are provided for each layer, and a no-flow boundary is defined along the catchment border. The model is set up with a time step of 1 day and forced with grid-based daily values of precipitation, temperature and reference evaporation. The grid-based precipitation $(10 \times 10$ km$^2)$, temperature $(20 \times 20$ km$^2)$ and reference evaporation $(20 \times 20$ km$^2)$ were obtained from the Danish Metrological Institute for the period 1990–2013.

In-situ groundwater heads are measured bi-hourly in ten wells (Figure 1). The first available head data are from November 2012. However, the starting time and the data coverages differ for the different wells. The ten wells are screened at different depths. Wells 5398, 5637, 5353, 5354 and 8008 are located in layer 5; wells 5373, 5647, 5844, 5393 and 5366 in layer 4 (Figure 1). Daily measurements of river discharge are available at several stations, and station 250082 integrates discharge from the entire catchment. A quality control procedure has been performed to filter out erroneous values in the raw data.

Model calibration and uncertainty

Model calibration was performed using an automated optimizer (PEST version 11.8, (Doherty 2010)) for the period

2006–2009. The data used in the calibration were groundwater head and river discharge observations. During this period, a large amount of discrete groundwater head data (466 in total) was available. However, for most wells only one instantaneous value for the entire period was available. For discharge data, time series of daily values from five stations (250018, 250020, 25002, 250733 and 250082) were available. The objective function included several error measures of both head and discharge.

Based on sensitivity analysis, the parameters listed below were selected for calibration and other relevant parameters were tied to these parameters. The selected parameters included horizontal and vertical hydraulic conductivity for the major sediment types (meltwater sand, clay, quartz sand and mica clay), specific yield for meltwater sand, drainage time constant, river–groundwater leakage coefficient, and root depths for the different vegetation types. These parameters describe a particular property class (e.g., geological units, vegetation types and soil types).

The model calibration resulted in all parameter values being within expected ranges. For groundwater head, the Root mean squared error (RMSE) for the calibration period is 2.73 m and the mean error -0.85 m for most locations (90%), with the initial sensitivity analysis of 55 parameters. These parameters include horizontal and vertical hydraulic conductivities, specific yield and storage, root depths, detention storage or overland flow, drainage constant and groundwater–river conductance term. It should be noted that the calibration period is prior to the assimilation period. Therefore, the discrete head observations used to calibrate the model are not used for assimilation. The head observations to be assimilated are from newly installed divers shown in Figure 1. Hence, the model performance during the DA period may not be as good as the model performance for the calibration period.

For DA, an ensemble of model realizations was generated to realistically represent the uncertainty in the system. Based on recommendations from a previous study examining the efficacy of DA with respect to various uncertainty sources in MIKE SHE, both parameters and model forcing were perturbed (Zhang et al. 2015). For model forcings, the grid-based rainfall and reference evaporation were perturbed by multiplying a mean-unity random Gaussian variable for each grid. Parameters were perturbed using

Table 1 | Saturated zone numerical layers based on geological surveys

Layers	Thickness (m)	Notes
6	0–3	Post-glacial sand, clay or peat. Just below the ground surface
5	0–65, avg. 30	Primarily glacial sand. Also minor horizons of glacial clay
4	0–37, avg. 20	Miocene sand
3	0–26, avg. 15	Miocene clay
2	50–125	Miocene sand
1	50–60	Miocene and Oligocene clays

The layers are listed from top to bottom (6 to 1) in the soil column.

Latin hypercube sampling based on the associated parameter covariance obtained from the PEST optimization.

Ensemble transform Kalman filter

The assimilation method used in this study is the ETKF. As a sequential DA method, ETKF approximates the true probability distribution based on an ensemble of model realizations conditioned on a series of observations. Different from EnKF (Evensen 2003), ETKF is one type of deterministic filter which does not require additional observation perturbation. The ETKF was first introduced by Bishop *et al.* (2001), and later modified by Wang *et al.* (2004).

The ETKF uses a state-space formulation of the model

$$X_{t+1} = M(X_t, U_t, \theta) \approx M_d\left(X_t, \widetilde{U_t}, \tilde{\theta}\right) \tag{1}$$

where M is the stochastic model operator, M_d is the deterministic MIKE SHE model operator based on the numerical solution of the equations in the MIKE SHE code, X_t and U_t are the state vector and model forcing at time step t, θ stands for the model parameters. The stochastic model operator M here is approximated by the deterministic MIKE SHE model by taking both model forcing and model parameter uncertainty into account (Zhang *et al.* 2015) as described in the section 'Model calibration and uncertainty'. Therefore, $\widetilde{U_t}$ and $\tilde{\theta}$ are the perturbed forcing and parameters, respectively. Based on Equation (1), an ensemble of model realizations can be generated to capture the uncertainty in the hydrological model.

At time $t + 1$, the observations can be written as:

$$Y_{t+1} = HX_{t+1} + \varepsilon_{t+1}, \ \ \varepsilon_{t+1}\tilde{N}(0, R_{t+1}) \tag{2}$$

where Y is the observation vector, and H denotes the linear mapping operator specifying the deterministic relationship between model state and observations. Here the observations (groundwater hydraulic heads) are of the same type as the state vector, and the measurement operator is linear and can therefore be written as a matrix H. ε is the observation error. The observation error is assumed to be Gaussian, temporally and spatially uncorrelated, with zero-mean and a prescribed constant standard deviation σ_r.

Thus, the observation error covariance matrix R_{t+1} becomes a diagonal matrix with constant values specified along the diagonal (i.e., $R_{t+1} = diag(\sigma_r^2, \ldots, \sigma_r^2)$).

The forecast state distribution is estimated by a finite number m of model realizations as follows (time index omitted in the following):

$$X^f = \left[x_1^f, x_2^f, \ldots, x_m^f\right] \tag{3}$$

where the superscript 'f' stands for 'forecast'. The ensemble forecast is obtained from Equation (1).

The forecast error covariance is written as:

$$P^f = X'^f(X'^f)^T/(m-1) \tag{4}$$

where X'^f is the forecast state ensemble perturbation,

$$X'^f = \left[x_1^f - \overline{X^f}, x_2^f - \overline{X^f}, \ldots, x_m^f - \overline{X^f}\right] \tag{5}$$

and $\overline{X^f}$ is the ensemble mean of the model forecast. After assimilation, both the analysed state mean and the analysed error covariance are updated,

$$\overline{X^a} = \overline{X^f} + K\left(y - H\overline{X^f}\right) \tag{6}$$

$$P^a = (I - KH)P^f \tag{7}$$

where the superscript 'a' stands for 'analysed', and K is the Kalman gain defined as:

$$K = P^f H^T (HP^f H^T + R)^{-1} \tag{8}$$

However, P^a is never explicitly calculated in practice and only the ensemble mean and ensemble anomalies are calculated. Factorizing Equation (7) on both sides we can derive

$$X'^a = X'^f T \tag{9}$$

where

$$T = \left[I + \left(HX'^f\right)^T R^{-1} HX'^f/(m-1)\right]^{-1/2} U \tag{10}$$

and U is an arbitrary orthonormal matrix such that $UU^T = I$.

Bias correction

To estimate the bias through the filter, the observation bias can be included in the state vector using state augmentation (Fertig *et al.* 2009) such that a separate bias is calculated for every measurement. The new state vector consists of both model state and observation bias as follows:

$$Z = \begin{bmatrix} X \\ B \end{bmatrix} \tag{11}$$

where B is the observation bias. Therefore, after assimilation in each time step, the bias is corrected together with the model state. During the model forecast step, we assume the bias is constant:

$$B_t^f = B_{t-1}^a \tag{12}$$

The initial bias for each ensemble member is assumed to be distributed as a zero mean uncorrelated Gaussian random variable.

After augmenting the state vector, the observation operator is modified from H to $\tilde{H}$:

$$\tilde{H}Z = HX + B \tag{13}$$

By doing so, the model equivalence of the biased observation can be calculated by addition of observation bias to the forecasted observation.

The state ensemble can therefore be written as:

$$Z^f = \left[Z_1^f, Z_2^f, \ldots, Z_m^f \right] \tag{14}$$

Experiment design

The implementation of the ETKF and the handling of the model-filter communication are written in a DA library using the Open Model Interface (OpenMI) framework (Moore & Tindall 2005; Ridler *et al.* 2014b). In this study, the MIKE SHE state variable used in the assimilation system is the groundwater hydraulic head. However, as the

groundwater component is fully coupled with the surface water component and other model components, the whole model state may also benefit from updating groundwater head.

From 2010-01-01, the experiment has been divided into two periods: a warm-up period (2010-01-01 to 2012-11-01) and a DA period (2012-11-01 to 2013-12-31). For the warm-up period, each ensemble member started with the same initial condition but was forced with different forcing and parameter values as described in the section 'Ensemble transform Kalman filter'. Given the calibrated model (denoted deterministic model in the following) and the specified model error, around a 3-year warm-up period is chosen to generate the initial forecast ensemble for the DA.

DA is first tested against synthetic biased observations, and subsequently tested against real *in-situ* head observations. In all experiments, groundwater head observations are assimilated every 2 weeks. The initial bias at each observation location for each ensemble member is sampled from a Gaussian distribution with zero mean and a standard deviation of 1 m. The observation error standard deviation is assumed to be 0.15 m. An ensemble size of 60 is used in the experiments. Different ensemble sizes (30, 60, 90, 180) were tested for the synthetic DA experiment and it was found that no significant improvement could be gained by using an ensemble size of more than 60.

Experiments were performed on both a synthetic and real test setup. Figure 2 shows the experimental approach in the synthetic DA experiment, where the biased observations are generated by adding a pre-defined constant bias and observational error to the synthetic true model. The *synthetic* case assumes that the true model is one realization of the forecast ensemble with certain forcing and parameter values. The *real* case assimilates actual *in-situ* measurements, which are assumed to be biased compared to the deterministic model.

For an overview of the experiments performed, see Table 2. For both the real and synthetic setup, experiments were carried out testing both a bias-aware filter (including both state and bias estimation as described in sections 'Ensemble transform Kalman filter' and 'Bias correction') and a bias-blind filter (including only state estimation with the standard ETKF as described in the section 'Ensemble transform Kalman filter').
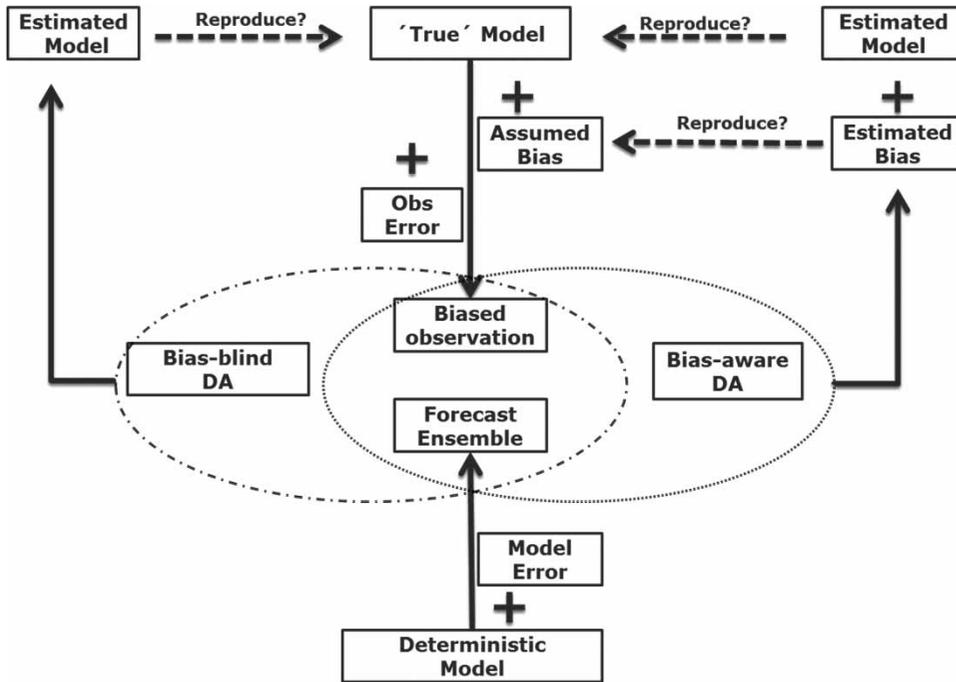
Figure 2 | Schematic diagram of the experimental approach for synthetic DA.

To further examine how the stratigraphic layers in the 3D groundwater model impact assimilation results, an additional experiment was performed (synthetic bias-aware, two-layer) where the state vector is only represented by layer 4 and layer 5. As all observation wells are located in these two layers, updating the entire groundwater zone may not be needed for the bias estimation. In this case, the filter only corrects the state in these two layers, but due to the vertical flow between layers, the hydrological model also updates the remaining layers. Therefore, this experiment is useful to assess how the corrections within two layers are propagated to the entire groundwater zone.

RMSE is used to evaluate assimilation performance. The average RMSE over all available time steps and over the entire domain for all layers is calculated by:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n_t} \sum_{i=1}^{N_{xy}} \left| \bar{X}_{t,i} - X_{t,i}^{true} \right|^2}{n_t N_{xy}}} \qquad (15)$$

where $X_{t,i}^{true}$ represents the true realization of the ith node in the state at time step t, n_t number of time steps, N_{xy} the state size, and $\bar{x}_{t,i}$ the ensemble mean of the ith node in the state at time step t.

RESULTS

Synthetic experiment

All layers in state vector

In the synthetic experiments, one of the aims is to evaluate the performance of the state estimation given the known true model. Figure 3 shows the results for both bias-aware and bias-blind filter at well 5637. The well is located in the northeast of the catchment (see Figure 1) and screened in

Table 2 | Overview of experiments

Observations	Filter	Layers in state vector
Synthetic	Bias-aware	All
	Bias-blind	All
Real	Bias-aware	All
	Bias-blind	All
Synthetic	Bias-aware	Layers 4 and 5
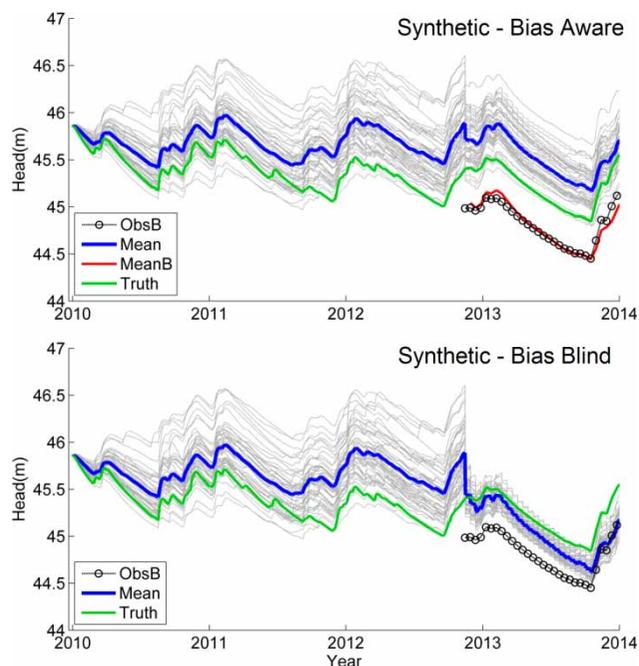
Figure 3 | State/observation estimation at well 5637 for the synthetic experiment. Upper panel: bias-aware filter. Lower panel: bias-blind filter. ObsB: biased observation. Mean: ensemble mean of state. MeanB: ensemble mean of estimated observation. Truth: true model simulation. Grey line: ensemble realization. DA starts from 2012-11-01.

layer 5 at a depth of 16.8 m. There are two aspects worth noticing: the observation bias estimation and what impact that has on the values of the state.

The bias-aware filter (Figure 3) (top) straightaway estimates the head bias, and hence the mean bias (MeanB) matches the observations. In this case, there is not a drastic change in the ensemble mean when observations are first available, in contrast to the bias-blind filter (Figure 3) (bottom) that drives the ensemble states erroneously towards the observations despite there being a bias of −0.87 m.

The quantitative performance of DA is summarized in Table 3. In the model domain, the averaged RMSE is used to calculate the difference between the model simulation and the truth. In the observation domain, the RMSE is calculated by averaging over observation points at observation times.

Table 3 | Averaged RMSE for the entire domain and for the observation locations

	Bias-aware vs truth	Bias-blind vs truth	Open-loop vs truth
Obs wells	0.23 m	0.25 m	0.37 m
Domain	0.22 m	0.35 m	0.30 m

The open-loop simulation refers to the average of the forward simulation of all ensemble members, which is used as a reference. The open-loop RMSE of 0.37 and 0.30 m (compared to the true observation wells and the entire domain in the truth model) represents a realistic model error.

Bias-blind DA reduced the RMSE in the individual wells, but overall in the domain, the RMSE increased from 0.30 to 0.35 m. This deterioration indicates that the filter was not able to correct the model in regions without measurements. The bias-aware filter, however, was able to correct the model at the measurement points as well as the entire domain where the RMSE reduced from 0.30 to 0.22 m. This synthetic experiment provides a strong case for the need for bias-aware filtering in a system with both model error and observation bias present.

To better understand the state estimate improvement across the layers, the RMSE is calculated by averaging over each layer at each time step, for both open-loop and assimilation simulations. The results of the different layers are shown in Figure 4. Note that half of the observations are located in layer 4 and the other half in layer 5. However, the RMSE reductions can be seen for all layers compared to the open-loop results, right after the assimilation starting time (2012-11-01). Among all layers, the largest improvement can be seen for layer 5, where RMSE is reduced by around 0.14 m on average. For the rest of the layers, RMSE is reduced by around 0.1 m on average.

To illustrate bias estimation performance for all observed locations, the average observation bias estimate is calculated for each well and compared to the predefined bias (scatter plot in Figure 5). Although there are some small differences between the estimated bias and the actual bias present at certain wells, the bias-aware filter does estimate the biases accurately on average.

To isolate the filter's impact on state values versus bias estimation, a bar plot is shown in Figure 6 using the open-loop simulation as the reference. During assimilation, the state estimate tends to depart from this open-loop simulation. The overall state correction by the filter at observed locations is calculated as the difference between the average state estimate and open-loop simulation for the ten wells (ETKF-OL), shown in Figure 6. The bias is also estimated (EstBias) at each measurement location as well as the sum of these two (ETKF + EstBias-OL). Finally, (Obs-OL) is the
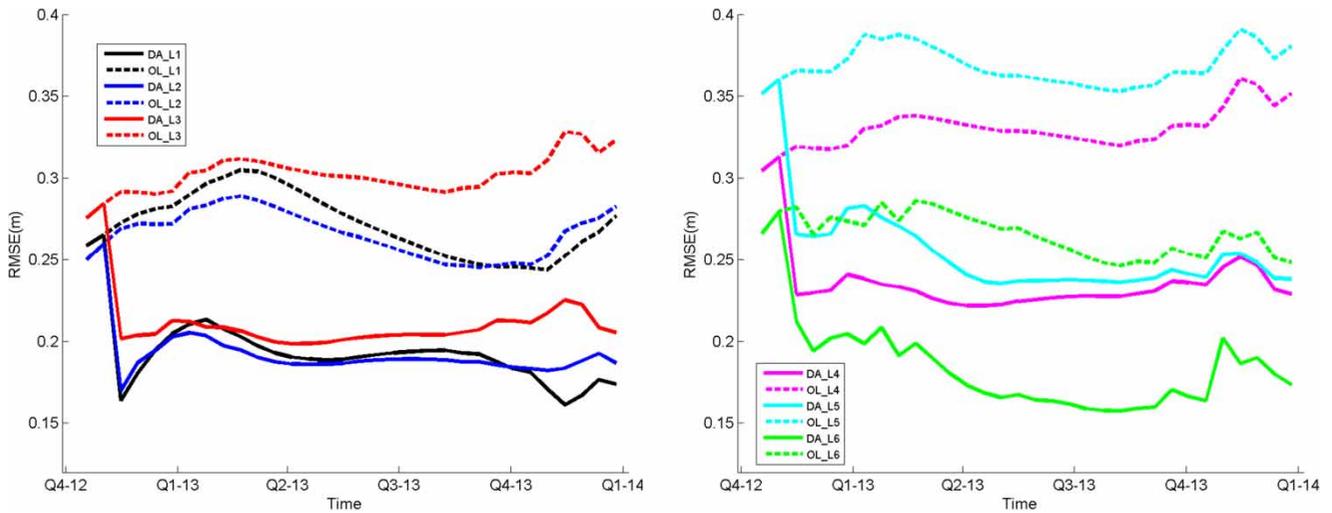
Figure 4 │ Average RMSE in different layers in the synthetic experiment with layers 1–3 on the left and layers 4–6 on the right. The dashed lines are RMSE of open-loop; the solid lines are RMSE of the bias-aware ETKF. Indicated by the legend, for example, 'DA_L1' means the average RMSE using bias-aware ETKF at layer 1 and 'OL_L1' means the average RMSE of open-loop simulation at layer 1. The time axis label 'Q' are the four yearly quarters for years 2012 to 2014.

difference between observation values used for assimilation and open-loop runs. By splitting the estimated observation departure into the two terms (ETKF-OL) and EstBias, we can see to what extent the filter updates can be attributed to the state and observation bias, respectively. For all observations except at M5373, the magnitude of the bias estimated by the filter (EstBias) is less than (Obs-OL). This suggests that simply pre-processing the observations (such that the observation mean matches the open-loop run) would have led to an overestimation of the measurement

bias. In a sense, this figure shows how much of the bias is attributed to observation bias (EstBias) and how much to model bias (ETKF-OL).

Two layers in state vector

A set of experiments was performed to investigate how state corrections are propagated across saturated zone layers. Only layers 4 and 5 of the six saturated zone layers (where measurements are present) were included in the state vector.
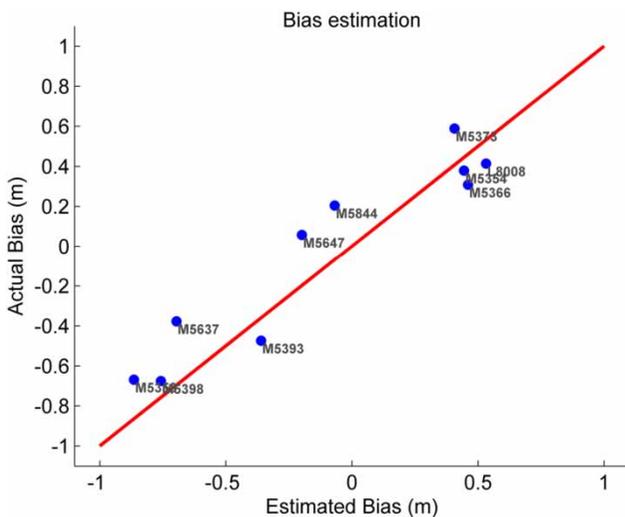
Figure 5 │ Averaged estimated bias vs predefined bias for the ten wells in the synthetic experiment.
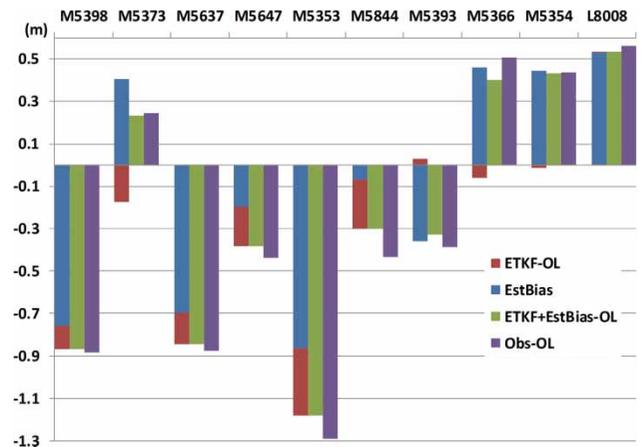
Figure 6 │ Synthetic experiment bias and state correction estimate with respect to the open-loop simulation at the ten wells. ETKF-OL (unbiased state estimate – open-loop simulation) represents the corrections made to the state, EstBias (estimated bias), ETKF + EstBias-OL (unbiased state estimate + estimated bias - open-loop simulation) and Obs-OL (observation – open-loop simulation).

The performance of the filter is comparable to the all-layer experiment described previously. The bias-aware filter estimated the bias in the observations with similar skill, as shown in Figure 7 (left). Furthermore, the proportion of the filter update attributed to bias and state corrections is also similar, as shown in Figure 7 (right). However, the extent of the correction is markedly different from the all-layer experiment. Figure 8 plots the average RMSE in the different layers. In layers 4 and 5, the bias-

aware filter reduced the RMSE to roughly 0.26 and 0.28 m, respectively, but it is not as significant a reduction as 0.22 and 0.24 m for the all-layer experiment. Although the filter corrected only two layers, the model dynamics lead to slightly reduced RMSE in the other layers, especially in layer 3. However, in all the layers, the best performance was obtained by including all layers in the state vector.

The spatial RMSE in the catchment in layer 5 is visualized in Figure 9. The spatial pattern of the open-loop run is
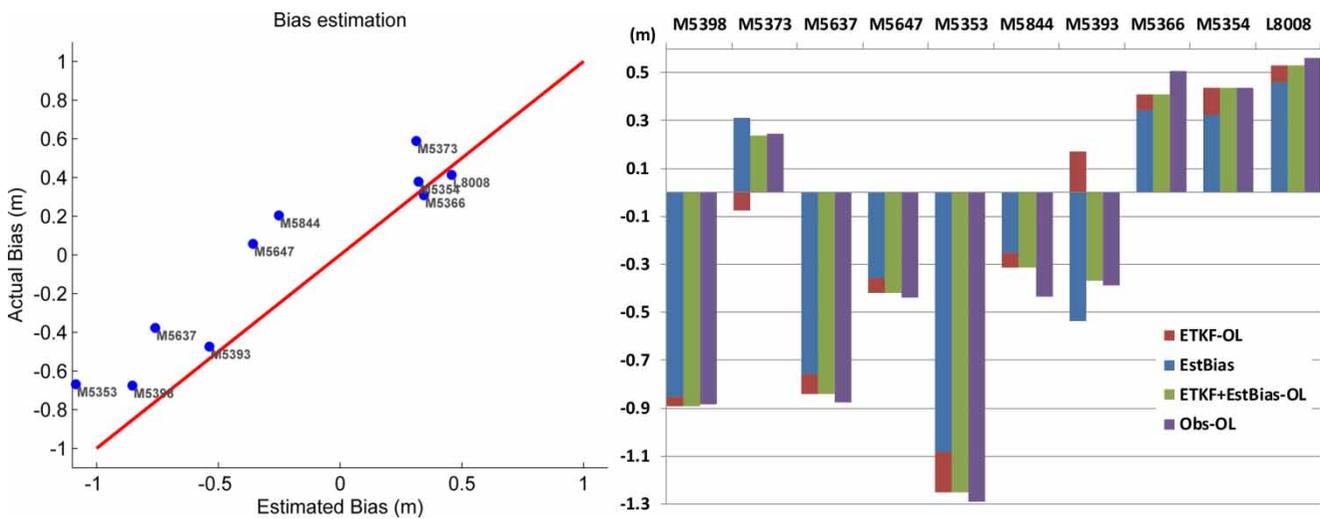
Figure 7 │ Synthetic experiment with two-layer updating. Left: estimated bias vs predefined bias for the ten wells. Right: bias and state correction estimate with respect to the open-loop simulation at the ten wells.
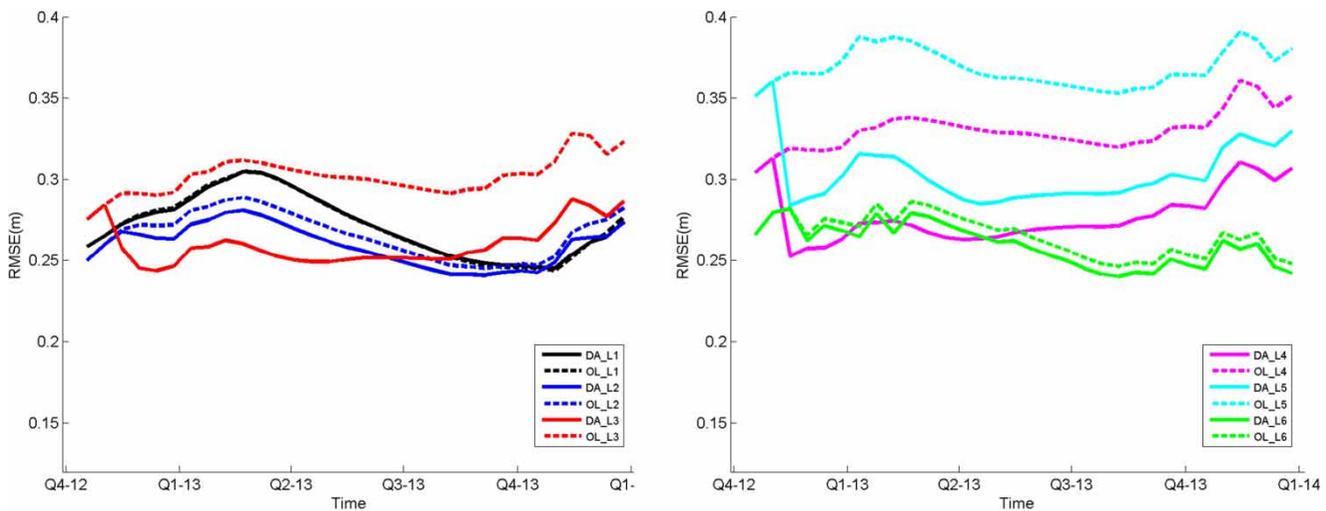
Figure 8 │ Average RMSE in different layers for the synthetic experiment with two-layer updating. The dashed lines are the RMSEs of open-loop while the full lines are the RMSEs of bias-aware ETKF. For example, 'DA_L1' means the average RMSE using bias-aware ETKF at layer 1 and 'OL_L1' means the average RMSE of open-loop simulation at layer 1. The time axis label 'Q' are the four yearly quarters for years 2012 to 2014.
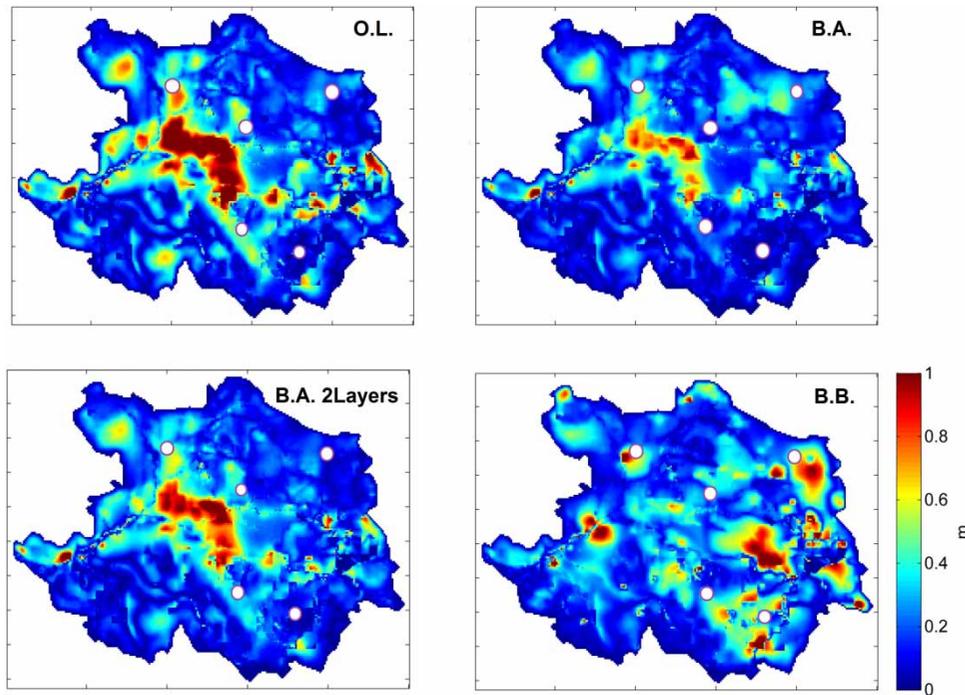
Figure 9 │ Synthetic experiment spatial RMSE in layer 5. Upper left: open-loop simulation; upper right: bias-aware filter with updating of all layers; lower left: bias-aware filter with updating of two layers; lower right: bias-blind filter with updating of all layers. The observation locations in layer 5 are marked with white circles.

similar to the two-layer and all-layer experiments, with a noticeable reduction in RMSE in the central region as a result of DA. Again, best results were obtained in the entire catchment using the all-layer state vector. Note that the bias-blind filter (bottom right panel in Figure 9) leads to increased RMSE in several regions of the catchment due to inconsistent corrections from bias in the measurements. This highlights the value of online bias estimation.

Real data experiment

State and bias estimation from the real DA experiment is presented in this section. Large discrepancies in terms of both groundwater levels and dynamics are present when comparing the real observations with the calibrated model. In general, the average absolute difference is over 2.5 m and the observed data show more temporal dynamics than the model. Another challenge with the real DA is that neither the true model nor the bias is known. This makes it difficult to diagnose and evaluate the performance of the assimilation result. Also note that the observation periods of the ten wells may impact the assimilation performance;

larger improvements can be expected in periods where more observations are available and assimilated.

As an example, Figure 10 shows the state and observation estimation for bias-aware and bias-blind runs at wells M5398 and M5637. For the bias-aware case (top panels), the estimated unbiased states are similar to the deterministic model state and the estimated mean bias matches the observations, confirming the efficacy of the bias-aware filter. In contrast, the bias-blind runs (lower panels) produce erratic groundwater levels, evidenced by the zig-zag patterns in hydraulic head levels. In one of the wells (M5637), the head level erroneously swells over 5 m to match the biased measurements.

As the truth is unknown in this case, the calibrated deterministic run is used as a baseline for comparison. Again, the bias-aware filter correctly calculates the observation bias with little residual error (scatter plot in Figure 11). Furthermore, the proportion of the corrections attributed to the bias vs state updating is consistent with the synthetic experiments (bar chart in Figure 11) but with relatively less weight going to state correction (ETKF-OL). Given the considerably large residuals from the actual
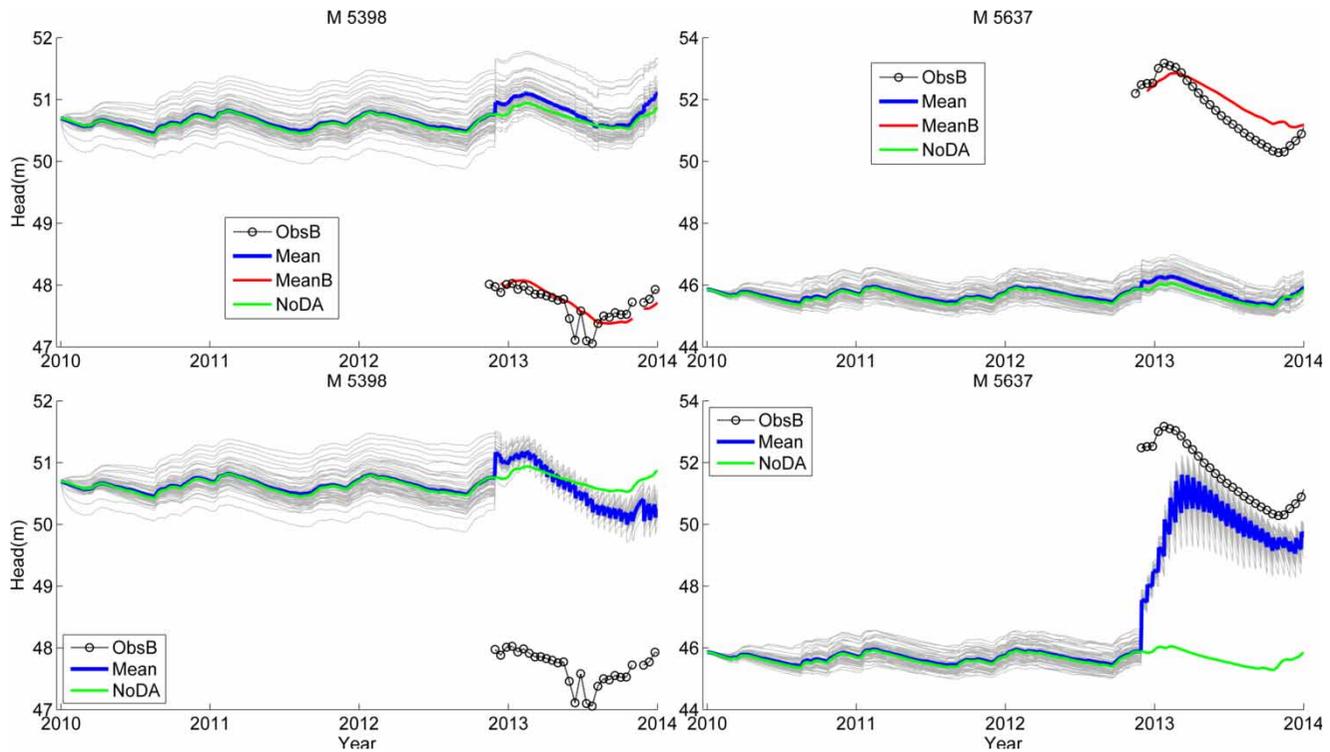
Figure 10 │ State/observation estimation at well 5398 (left) and well 5637 (right) for the real data experiment. Upper: bias-aware. Lower: bias-blind. ObsB: biased observation. Mean: ensemble mean of estimated state. MeanB: ensemble mean of estimated observation. NoDA: deterministic model realization. DA starts from 01/11/2012.

observations, the state corrections at those observed locations are relatively small (<1 m) and most of the corrections are attributed to observation bias. This shows the bias-aware filter is able to estimate large observation bias.

From Figure 12 it appears that the patterns of corrections for the two layers are similar. This demonstrates the strong spatially correlated corrections both vertically and horizontally. The largest corrections are made in regions
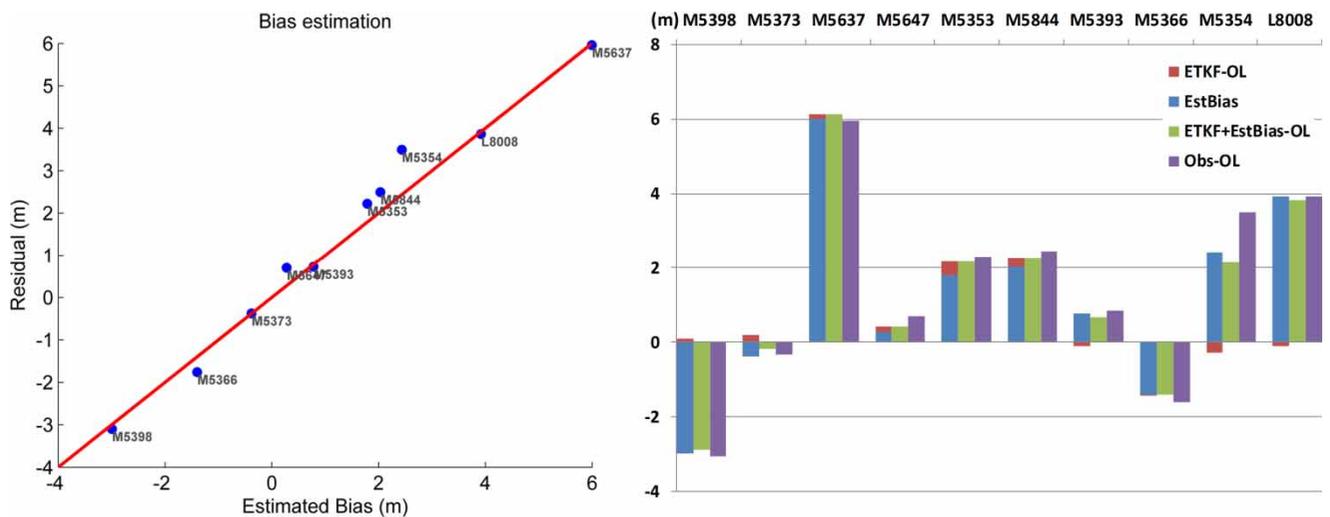
Figure 11 │ Real data experiment. Left: estimated bias vs residual for ten wells. Right: bias and state correction estimate with respect to the open-loop simulation at ten wells. Residual in this case refers to the difference between the assimilation and calibrated deterministic runs.
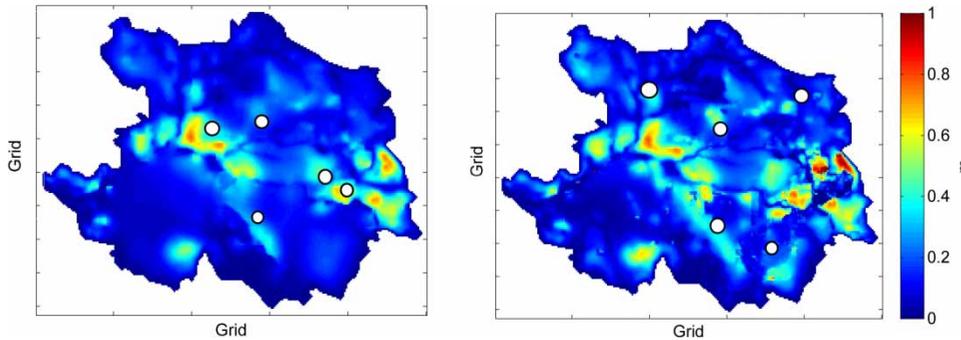
Figure 12 │ Spatial RMSE of groundwater head at layer 4 (left) and layer 5 (right) between the assimilated result and deterministic model output for the real data experiment. The observation locations at each layer are marked with white circles. RMSE is calculated using the calibrated deterministic run as the truth.
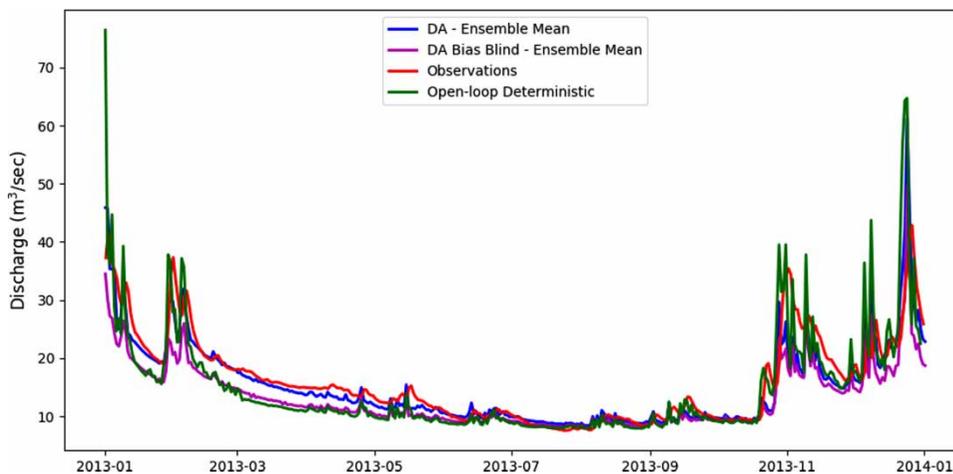
Figure 13 │ Discharge measured at the catchment outlet (250082 in Figure 1). The blue line is the bias-aware assimilated ensemble mean, magenta line is the bias-blind assimilated ensemble mean, red observations, and green the deterministic open-loop run. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2017.117.

close to available observations, implying that in the western and northern part of the domain, where no head observations are available, the corrections are minimal.

Discharge measured in the catchment is not assimilated, and is thus a valuable independent model variable for evaluation. The discharge at the catchment outlet (250082 in Figure 1) is shown in Figure 13. During the DA period with the bias-aware filter, the Nash–Sutcliffe efficiency increased from 0.71 to 0.76, due to improved groundwater estimates in this largely groundwater fed river. A much lesser correction is seen in the bias-blind case. The low-flow periods were especially corrected, improving the otherwise underestimated discharge. DA had a lesser effect on the peak discharges, as the peaks stem from overland flow, which was not a state variable in this experiment.

DISCUSSION AND CONCLUSIONS

Observation bias is a commonly occurring problem in hydrological models that needs to be addressed appropriately for DA. Building on the concepts presented in Drécourt *et al.* (2006) and Rasmussen *et al.* (2015), this is the first study to systematically investigate the performance of a bias-aware filter by assimilating both synthetic and real data in a complex three-dimensional integrated hydrological model. A particular focus was evaluating corrections across layers in the saturated zone.

Unlike most groundwater studies which use filtering as a means of parameter estimation or inverse modelling (Chen & Zhang 2006; Hendricks Franssen & Kinzelbach 2008; Sun *et al.* 2009), this study uses a pre-calibrated model based on geological knowledge of the domain. For this

reason, we assume an unbiased but uncertain model, and use the assimilation algorithm for state correction only. This assumption is, however, over-simplistic, since during calibration we use a set of (independent) measurements that are likely to be individually biased. Even if the calibration takes into account the overall set of measurements, the calibrated model will still contain biases in some areas of the model. Although an assimilation algorithm exists for joint observation-model bias estimation (Pauwels *et al.* 2013), such a method requires knowledge of the ratio of model to observation bias in the system, which is often not knowable in real applications. Therefore, observation bias-aware filtering provides a relatively simple yet consistent approach for uncertain systems.

The bias-aware filter was very responsive, promptly estimating the bias within one or two update cycles (Figures 3 and 10). In contrast, Rasmussen *et al.*'s (2015) bias estimates took months, even years to resolve, likely due to the system's increased degree of freedom from including parameter estimation in the filter. Clearly, a responsive bias estimate is ideal, especially in cases with drifting or cyclical biases due to instrument failure or seasonal environmental trends.

A simple alternative to bias-aware filtering is a priori rescaling of the observations (e.g., Reichle & Koster 2004) to remove the bias prior to assimilation. This pre-processing approach was tested by Zhang *et al.* (2015) to assimilate corrected hydraulic head in an identical setup to this study. The assimilation led to an overall reduction of the hydraulic head RMSE from 0.34 to 0.21 m compared to the *in-situ* (bias corrected) measurements. Encouragingly, the bias-aware filter in this study led to a comparable drop in RMSE, but also provides a continuous estimate of the measurement bias and their correlations to the model state. As shown in the bar graphs in this study, the magnitude of the estimated bias is often different from the observation minus the open-loop simulation, indicating the value of online bias estimation versus pre-processing out observation bias.

Another approach to reconcile the difference between model and measurement values is joint parameter-state assimilation. This method assumes unbiased observations, and calibrates the model parameters such that model forecasts match measurement values. In this study,

however, the model was calibrated using an independent set of data points and during a different time period. Given the spatial discretization of this catchment-sized model (200×200 meters), the biases are largely assumed to stem from the mapping between observation and model space.

The synthetic studies demonstrated that the bias-aware filter clearly outperforms the bias-blind filter. Updating the states toward largely biased observations can lead to unstable model dynamics, which deteriorate the assimilation results. The bias-aware filter, on the other hand, correctly accounted for the bias in the observations. Results show that all layers benefit from assimilation, even though deeper and surface layers are not measured. The two-layer experiment led to improved RMSE statistics in the entire catchment, also in layers not observed (and not included in the state vector), due to vertical propagation of the corrections in the numerical model. However, it is recommended to include the entire saturated zone in the state vector, as this leads to the greatest reduction in RMSE in the entire model domain.

This study assimilated real measurement data containing temporal gaps, distinct dynamics and bias in an integrated hydrological model to correct groundwater head. In summary, our main findings are as follows:

(1) Bias-aware filtering quickly and accurately estimates the hydraulic head observation bias.
(2) Best results are obtained by including the entire saturated zone model domain in the state vector, not only the layers being measured.
(3) In this integrated model, corrections to the groundwater state lead to improved discharge estimates, especially during low-flow periods.

ACKNOWLEDGEMENTS

REFERENCES

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. & Rasmussen, J. 1986 An introduction to the European Hydrological System – Systeme Hydrologique Europeen, 'SHE', 2: structure of a physically-based, distributed modelling system. *Journal of Hydrology* **87**, 61–77.

Bishop, C. H., Etherton, B. J. & Majumdar, S. J. 2001 Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review* **129** (3), 420–436.

Camporese, M., Paniconi, C., Putti, M. & Salandin, P. 2009a Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow. *Water Resources Research* **45** (10), W10421.

Camporese, M., Paniconi, C., Putti, M. & Salandin, P. 2009b Comparison of data assimilation techniques for a coupled model of surface and subsurface flow. *Vadose Zone Journal* **8** (4), 837–845.

Chen, Y. & Zhang, D. 2006 Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources* **29** (8), 1107–1122.

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Pauwels, V. R. N. & Verhoest, N. E. C. 2007 Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. *Water Resources Research* **43** (9), W09410.

Doherty, J. 2010 *PEST, Model-Independent Parameter Estimation, User Manual*, 5th edn. Watermark Numerical Computing, Brisbane, Australia.

Drécourt, J.-P., Madsen, H. & Rosbjerg, D. 2006 Bias aware Kalman filters: comparison and improvements. *Advances in Water Resources* **29** (5), 707–718.

Evensen, G. 2003 The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics* **53** (4), 343–367.

Fertig, E. J., Baek, S.-J., Hunt, B., Ott, E., Szunyogh, I., Aravequia, J., Kalnay, E., Li, H. & Liu, J. 2009 Observation bias correction with an ensemble Kalman filter. *Tellus Series A – Dynamic Meteorology and Oceanography* **61** (2), 210–226.

Friedland, B. 1969 Treatment of bias in recursive filtering. *IEEE Transactions on Automatic Control* **14** (4), 359–367.

Gharamti, M. E., Valstar, J. & Hoteit, I. 2014 An adaptive hybrid EnKF-OI scheme for efficient state-parameter estimation of reactive contaminant transport models. *Advances in Water Resources* **71**, 1–15.

Graham, D. N. & Butts, M. B. 2005 Flexible, integrated watershed modelling with MIKE SHE. In: *Watershed Models* (V. P. Singh & D. K. Frevert, eds). CRC Press, Boca Raton, FL, pp. 245–272.

Hendricks Franssen, H. J. & Kinzelbach, W. 2008 Real-time groundwater flow modeling with the Ensemble Kalman Filter: joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research* **44** (9), W09408.

Hendricks Franssen, H. J., Kaiser, H. P., Kuhlmann, U., Bauser, G., Stauffer, F., Müller, R. & Kinzelbach, W. 2011 Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. *Water Resources Research* **47** (2), W02532.

Kidmose, J., Refsgaard, J. C., Troldborg, L. & Stisen, S. 2017 Seasonal memory of a groundwater-surface water system. *Hydrogeology Journal*.

Kollat, J. B., Reed, P. M. & Rizzo, D. M. 2008 Addressing model bias and uncertainty in three dimensional groundwater transport forecasts for a physical aquifer experiment. *Geophysical Research Letters* **35** (17), L17402.

Kurtz, W., Hendricks Frankssen, H.-J., Kaiser, H.-P. & Vereeken, H. 2014 Joint assimilation of piezometric heads and groundwater temperatures for improved modeling of river-aquifer interactions. *Water Resources Research* **50** (2), 1665–1688.

Moore, R. V. & Tindall, C. I. 2005 An overview of the open modelling interface and environment (the OpenMI). *Environmental Science & Policy* **8** (3), 279–286.

Moradkhani, H., Sorooshian, S., Gupta, H. V. & Houser, P. R. 2005 Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources* **28** (2), 135–147.

Panzeri, M., Riva, M., Guadagnini, A. & Neuman, S. P. 2014 Comparison of Ensemble Kalman Filter groundwater-data assimilation methods based on stochastic moment equations and Monte Carlo simulation. *Advances in Water Resources* **66**, 8–18.

Pasetto, D., Camporese, M. & Putti, M. 2012 Ensemble Kalman filter versus particle filter for a physically-based coupled surface–subsurface model. *Advances in Water Resources* **47**, 1–13.

Pauwels, V. R. N., De Lannoy, G. J. M., Hendricks Franssen, J. H. & Vereecken, H. 2013 Simultaneous estimation of model state variables and observation and forecast biases using a two-stage hybrid Kalman filter. *Hydrology and Earth System Sciences* **17** (9), 3499–3521.

Rasmussen, J., Madsen, H., Jensen, K. H. & Refsgaard, J. C. 2015 Data assimilation in integrated hydrological modelling in the presence of observation bias. *Hydrology and Earth System Sciences* **20** (5), 2103–2118.

Reichle, R. H. & Koster, R. D. 2004 Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters* **31** (19), L19501.

Ridler, M. E., Madsen, H., Stisen, S., Bircher, S. & Fensholt, R. 2014a Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark. *Water Resources Research* **50** (11), 8962–8981.

Ridler, M. E., van Velzen, N., Hummel, S., Sandholt, I., Falk, A. K., Heemink, A. & Madsen, H. 2014b Data assimilation framework: linking an open data assimilation library

(OpenDA) to a widely adopted model interface (OpenMI). *Environmental Modelling & Software* **57**, 76–89.

Sun, A. Y., Morris, A. & Mohanty, S. 2009 Comparison of deterministic ensemble Kalman filters for assimilating hydrogeological data. *Advances in Water Resources* **32** (2), 280–292.

Wang, X., Bishop, C. H. & Julier, S. J. 2004 Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Monthly Weather Review* **132**, 1590–1605.

Xu, T., Jaime Gómez-Hernández, J., Zhou, H. & Li, L. 2013 The power of transient piezometric head data in inverse modeling: an application of the localized normal-score EnKF with covariance inflation in a heterogenous bimodal hydraulic conductivity field. *Advances in Water Resources* **54**, 100–118.

Xue, L. & Zhang, D. 2014 A multimodel data assimilation framework via the ensemble Kalman filter. *Water Resources Research* **50** (5), 4197–4219.

Zhang, D., Madsen, H., Ridler, M. E., Refsgaard, J. C. & Jensen, K. H. 2015 Impact of uncertainty description on assimilating hydraulic head in the MIKE SHE distributed hydrological model. *Advances in Water Resources* **86**, 400–413.

Zhou, H., Jaime Gómez-Hernández, J., Hendricks Franssen, H.-J. & Li, L. 2011 An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in Water Resources* **34** (7), 844–864.

Zovi, F., Camporese, M., Franssen, H. J. H., Huisman, J. A. & Salandin, P. 2017 Identification of high-permeability subsurface structures with multiple point geostatistics and normal score ensemble Kalman filter. *Journal of Hydrology* **548**, 208–224.