Real-time simulation of surface water and groundwater with data assimilation

He, Xin; Lucatero, Diana; Ridler, Marc-Etienne; Madsen, Henrik; Kidmose, Jacob; Hole, Øyvind; Petersen, Claus; Zheng, Chunmiao; Refsgaard, Jens Christian

Published in: Advances in Water Resources

DOI: 10.1016/j.advwatres.2019.03.004

Publication date: 2019

Document version: Publisher's PDF, also known as Version of record

Document license: CC BY

Citation for published version (APA): He, X., Lucatero, D., Ridler, M-E., Madsen, H., Kidmose, J., Hole, Ø., Petersen, C., Zheng, C., & Refsgaard, J. C. (2019). Real-time simulation of surface water and groundwater with data assimilation. Advances in Water Resources, 127, 13-25. https://doi.org/10.1016/j.advwatres.2019.03.004
Real-time simulation of surface water and groundwater with data assimilation

Xin He\textsuperscript{a,b,}\textsuperscript{*}, Diana Lucatero\textsuperscript{c}, Marc-Etienne Ridler\textsuperscript{d}, Henrik Madsen\textsuperscript{d}, Jacob Kidmose\textsuperscript{b}, Øyvind Hole\textsuperscript{e}, Claus Petersen\textsuperscript{f}, Chunmiao Zheng\textsuperscript{g}, Jens Christian Refsgaard\textsuperscript{b}

\textsuperscript{a} Department of Water Resources, China Institute of Water Resources and Hydropower Research, A1 Fuxing Road, Haidian District, 100038 Beijing, China
\textsuperscript{b} Department of Hydrology, Geological Survey of Denmark and Greenland, Copenhagen, Denmark
\textsuperscript{c} Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark
\textsuperscript{d} DHI, Hørsholm, Denmark
\textsuperscript{e} Department of Water, Ramboll, Drammen, Norway
\textsuperscript{f} Department of Research and Development, Danish Meteorological Institute, Copenhagen, Denmark
\textsuperscript{g} Guangdong Provincial Key Laboratory of Soil and Groundwater Pollution Control, Southern University of Science and Technology, Shenzhen, China

A R T I C L E  I N F O

Keywords:
Data assimilation
Hydrological model
Real-time simulation
Model evaluation

A B S T R A C T

Data assimilation (DA) has proven to be a useful technique in real-time hydrological modeling and forecasting. Jointly assimilating both surface water and groundwater data has promising application value for hydrological simulations in areas where surface water and groundwater are closely linked; however, such studies have not been intensively reported. In addition, the role of the quality of precipitation forecast has not been fully addressed in real-time forecasting using a coupled surface water - groundwater model, where the model evaluation includes both deterministic and probabilistic forecasts. In the present study, we use the MIKE SHE hydrological model code in conjunction with the Ensemble Transform Kalman Filter DA technique. The study area is a small urbanized catchment in Denmark. The model is run in simulated real-time using historical numerical weather prediction forecasts. The results show that DA can significantly reduce model bias and thereby improve model performance for both surface water and groundwater simulations. Comparing the impact of DA and rainfall forecast quality, it is found that, for streamflow forecasts, the most important factor is the quality of the rainfall data; whereas for groundwater head forecasts, the initial state at time of forecast is more important. We also find that inclusion of rainfall forecast uncertainty may be important for simulating a single event, however, it is not vital if long-term average model performance is of interest.

1. Introduction

It is commonly acknowledged that climate change is already ongoing and will most likely intensify in the foreseeable future (Hisdal et al., 2001; Zolina et al., 2010). Therefore, climate adaptation strategies in urban planning have now become more frequent in European cities in order to reduce the risk of flooding (Hallegratte et al., 2011; Uittenbroek et al., 2013). One approach to cope with the increased precipitation under the future climate conditions is to use real-time water resources management systems, including real-time hydrological models (Addor et al., 2011; He et al., 2016; Lu et al., 2008). Operation of a real-time system requires higher accuracy from the hydrological models, since the results of the real-time forecasts are used almost immediately for decision making once they are issued.

Data Assimilation (DA) has proven to be a useful technique in real-time hydrological modeling and forecasting (Abaza et al., 2014; Vrugt et al., 2006). DA is the process where observations are incorporated into the state updating of a numerical model, therefore in hydrological modeling, it can help reduce the bias of the model and lead to improved predictive capabilities. So far, most of the modeling studies focused only on a part of the hydrological cycle, e.g. groundwater flow (Franssen et al., 2011), land surface processes (Albergel et al., 2008), and rainfall-runoff generation (Moradkhani et al., 2005a). Studies that assimilate more than one hydrological variable, i.e. multivariate assimilation is already available but not yet a common practice (Kurtz et al., 2014; Lee et al., 2011; Zhang et al., 2016). Jointly considering surface water and groundwater flow in a real case study is even rarer, however important since in many parts of the world, surface water and groundwater are closely linked.

* Corresponding author at: Department of Water Resources, China Institute of Water Resources and Hydropower Research, A1 Fuxing Road, Haidian District, 100038 Beijing, China.

E-mail addresses: hex@sustc.edu.cn, hex@iwhr.com (X. He).

https://doi.org/10.1016/j.advwatres.2019.03.004
Received 18 September 2018; Received in revised form 1 February 2019; Accepted 7 March 2019
Available online 8 March 2019

0309-1708/© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY licence. (http://creativecommons.org/licenses/by/4.0/)
where groundwater is responsible for a large part of the stream flow (Botto et al., 2018; Camporese et al., 2010; Pasetto et al., 2012).

A number of algorithms have been applied for DA in hydrology, among which the most commonly used approach is the ensemble Kalman filter (EnKF) and its variations (Botto et al., 2018; Fertig et al., 2009; Hunt et al., 2007). EnKF is used instead of the ordinary Kalman filter mainly due to the fact that it can deal with nonlinearities encountered in hydrological models, and its ability to drive system state towards the observations even for small ensemble size (Pasetto et al., 2012). In recent hydrological DA studies, the Ensemble Transform Kalman filter (ETKF) has been applied (Rasmussen et al., 2015; Ridler et al., 2018), which is a deterministic formulation of the EnKF that does not require measurement perturbations.

DA was first applied by the hydrological community for flood forecasting, where only two factors need to be considered: the estimated rainfall and the runoff generation process, early examples include Refsgaard et al. (1983) and Refsgaard (1997). With the development of more integrated hydrological model codes, the quality of precipitation estimation has always been the center of attention when analyzing the quality of hydrological simulations (Borga, 2002; He et al., 2011; McMillan et al., 2012; Entekhabi et al., 1994). For terrestrial water cycle, precipitation is the driving force to other hydrological processes, and therefore the uncertainty of the precipitation product used in integrated hydrological modeling can be propagated and influence the forecast results. In real-time hydrological modeling with DA, uncertainty in observed state variables can be accounted for by using stochastic error models (Clark et al., 2008; Maggioni et al., 2012), and statistical forecasting with an ensemble of hydrological simulations that considers rainfall uncertainty is becoming more common. However, more research is needed on the impact of rainfall forecast uncertainty on updating surface water and groundwater coupled models.

Evaluation of modeled forecast is an important step in the hydrological model development cycle, which ideally requires a large sample of forecast-observation pairs (Hamill et al., 2004; Schaake et al., 2007). Although performance metrics of deterministic simulations can be used for ensemble forecast evaluation by applying them to the ensemble mean, it is often suggested to add metrics for evaluating probabilistic forecast properties (Laio and Tamea, 2007; De Lannoy et al., 2006). This way, additional forecast qualities such as reliability or probabilistic accuracy can be included simultaneously (Murphy, 1993). Probabilistic criteria from the meteorological community have gained increasing popularity in the hydrological community (Bourgin et al., 2014; Jaun and Ahrens, 2009; Roulin and Vannitsem, 2005; Thielen et al., 2008). However, no consensus has yet been reached among hydrologists exactly what kind of statistical values should be chosen and whether the use of complex metrics is useful in the decision-making stage of the forecasting process. Moreover, forecast evaluators are also faced with the limitations of using a small sample size for the evaluation of extreme events such as floods.

The objectives of the present study are: first, to evaluate the operational aspect of real-time hydrological DA framework when assimilating both streamflow and groundwater head measurements at the same time; second to investigate the impact of DA together with rainfall forecasts with respect to performance of streamflow and groundwater head forecasts; and third, to compare difference of model performance using simulations from continuous model runs and flood events based on an evaluation metrics.

2. Methods

2.1. Silkeborg catchment and data

The study area lies in a humid tempered climate area with westerly winds from the North Atlantic facilitating an average precipitation of around 900 mm/y (1960–90), average monthly temperatures between 15 °C (July) and -0.5 °C (February) and a potential evapotranspiration of 550 mm/y (Allerup et al., 1998; Kidmose et al., 2013; Scharling, 2000). The hydrological catchment, seen in Fig. 1, covers an urbanized area of the City of Silkeborg where a motorway was constructed during the period 2013–2016. Both prior to the construction period and until now, hydraulic heads have continuously been measured with automated loggers and manually in boreholes with 1 m screens placed in an upper sandy-gravelly aquifer. Stream discharge has been measured at discharge stations at the two creeks. A significant part of the discharge in the two creeks are generated from urban surface runoff; discharge from more or less impermeable surfaces as roofs, areas around buildings and parking lots. The nature of the routing of water from these impermeable surfaces to the creeks are not well understood, which has a significant impact on the ability to simulate correct flow in the model.

Fig. 1. Location of the study area, the Silkeborg catchment, the rain gauge station that give the rainfall observations, the groundwater wells and the stream discharge stations where data are available and used for data assimilation. Of the two streams, one has 3 discharge stations, the other one has no observation installed.
2.2. Hydrological model

The hydrological model was setup with the MIKE SHE code (Abbott et al., 1986a, 1986b; Graham and Butts, 2005) including simulation of evapotranspiration, overland flow, flow through the unsaturated zone and groundwater flow. Furthermore, the MIKE SHE was coupled with a MIKE 11 model simulating stream and channel flow. Besides water exchange between groundwater and surface waters, the exchange between overland flow and streams is also important because of the urbanized character of the catchment; overland flow describes discharge on the surface of the urbanized, impermeable surfaces as roads, parking and building areas. The model area is 11.7 km², seen in Fig. 1, and 30% is impermeable. In order to have a relatively fast model preprocessing and running time, the model is setup in a 50 x 50 m horizontal grid size. The geology of the area, with Quaternary clay and sand at the surface and Miocene clay, silt and sand at the lower parts, were aggregated into 3 numerical layers with spatial distribution of hydraulic conductivity (K) representing the geology. The model inherits boundary conditions in the lower of the 3 layers from a larger scale model covering the entire Gudenaa river system (Kidmose et al., 2013, 2015). The model was calibrated against 35 hydraulic head time series and two discharge stations using the algorithm AUTOCAL (Madsen, 2000,2003) evaluating model performance from October 2010 to December 2012.

2.3. Precipitation forecast

The precipitation forecasts are issued by a numerical weather prediction (NWP) model named HARMONIE (Bengtsson et al., 2017), which is an operational short-range weather forecast model. HARMONIE is a joint venture by 10 European countries starting in 2011. In Denmark, the model is operated by the Danish Meteorological Institute with the purpose to replace the High Resolution Limited Area Model (HIRLAM), which was initiated in 1985. HARMONIE is able to make predictions of among other variables precipitation, temperature, and solar radiation at various temporal and spatial scales. HARMONIE is a non-hydrostatic model with the potential of better predictions of convective and large-scale precipitation compared to conventional NWP models like HIRLAM. The precipitation product used in this study is a deterministic forecast product projected on 500 m Cartesian grid, with lead time of 48 h and update frequency every 6 h.

2.4. Data assimilation

2.4.1. Local ETKF and bias correction

The MIKE SHE model is coupled with a generic DA library that includes different EnKF based algorithms (Ridler et al., 2018,2014). Here the ETKF is applied. The probability density of the state, x, is estimated by a finite number m of system states (ensemble) such that X = [x1, x2, ..., x_m] where the members are forced with randomly generated system noise and propagated through the model. The forecast error covariance can be obtained from the sample covariance of the ensemble members,

$$P^f = (m-1)^{-1} A^f (A^f)^T$$  (1)

where T denotes the matrix transpose. Here A^f is the matrix of ensemble perturbations, or anomalies defined as:

$$A^f = [x^{f(1)} - x^b, x^{f(2)} - x^b, ..., x^{f(m)} - x^b]$$  (2)

with x^b the ensemble mean:

$$x^b = \frac{1}{m} \sum_{i=1}^{m} x^{f(i)}$$  (3)

With ETKF, a deterministic update of the ensemble anomalies is performed so that the analysis error covariance matches the theoretical value given by the Kalman filter and both the ensemble mean and ensemble anomalies are updated explicitly based on the transform matrix T with

$$A^a = A^f T$$ that must satisfy

$$T = T^TU$$ and $$T^i = \left[ I + \frac{1}{m-1} \left( H A^f \right)^T R^{-1} H A^f \right]^{-1/2}$$  (4)

Where R is the observation error covariance, H the measurement operator, and U an arbitrary orthonormal matrix $$UU^T = I$$, where the solution to $$T^i$$ is symmetric. Variable localization (Zhang et al., 2016) was used so that discharge measurements were only used to correct the river discharge state and the head measurements the saturated zone. A schematic illustration of the ETKF method used in the study is shown in Fig. 2. The variables in the state matrix are: groundwater head, head bias, and discharge. In the MIKE 1D hydrodynamic engine that we used in our study, the model physics ensures the water height in the river crossings corresponds to the discharge. Therefore, the water height gets recalculated to match the updated discharge.

A fundamental assumption in the Kalman filter is that the observations and the model are unbiased. However, in this study, a persistent hydrological head bias is present and must be rectified. Bias correction follows the methodology by Ridler et al. (2018) with observation biases estimated online by augmenting the state space vector such that

$$z = \begin{bmatrix} x \\ \beta \end{bmatrix}$$  (5)

where β is bias. Biases with respect to hydraulic head estimates are assumed to change gradually and are modeled by persistence, with

$$\beta_{t+1}^f = \beta_{t-1}^a.$$  (6)
where \( \delta_{t-1} \) is the updated bias vector from the previous time and \( \delta_t \) is the forecast bias vector at the current analysis time. At the beginning of the data assimilation procedure (\( t = 0 \)), an initial bias \( \delta_0 \) is estimated for each observation location based on a priori knowledge of observational bias. For each ensemble member, the bias uncertainty is modeled by perturbing \( \delta_0 \) with Gaussian noise. The observation operator is modified such that the observation bias is added to the forecasted observations. For a detailed description of the algorithm, see Ridler et al. (2018).

### 2.4.2. Model uncertainty

An ensemble of model realizations was generated to reflect the uncertainty in the catchment model. For optimal DA performance, both meteorological forcing and model parameters were perturbed (Zhang et al., 2015). Parameters were perturbed based on information about parameter uncertainty from a previous parameter optimization performed on this catchment and documented in Kidmose et al. (2015). Details of the parameter perturbations are shown in Table 1.

Daily ET was perturbed using a Gaussian error model at every time step with a relative standard deviation of 0.3. Hourly precipitation was perturbed using a first-order autoregressive AR(1) model according to

\[
\delta_t = \phi \delta_{t-1} + \varepsilon_t, \tag{7}
\]

with \( \delta \) being the perturbation value at time \( t \) and \( t-1 \), \( \phi \) is the AR(1) model parameter with a 24 h decorrelation time, and \( \varepsilon \) is white noise with zero mean and relative standard deviation of 0.2. Precipitation was perturbed by \( \text{Perturbed precipitation}_t = \text{Precipitation value}_t \times (\delta_t + 1) \), with perturbations bounded by 0.5 and 1.5 to avoid extreme values.

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. ETKF is the DA algorithm used by this study. It is a deterministic filter, and used in previous hydrological studies with MIKE SHE (Zhang et al., 2016; Rasmussen et al., 2015; Ridler et al., 2018). Equations are given in the aforementioned literatures which show how observation bias is included in the filter’s state vector and estimated. In our present study, only observation bias is updated during DA model. Parameters models are perturbed in the DA so each ensemble member has a different and fixed parameter set. This will account for some of the structural uncertainty in the system. When the models are run, the rainfall is perturbed for the entire time period. (spin-up, and DA and forecast for the stochastic forecasts) using the AR(1) model.

### 2.5. Performance metrics

#### 2.5.1. Bias

Bias is measured as the mean difference between the ensemble mean and observations for a total of \( N \) forecast-observation pairs. It is computed as follows:

\[
\text{Bias} = \frac{1}{N} \sum_{j=1}^{N} (f_j - y_j) \tag{8}
\]

where \( f_j \) is the ensemble mean for forecast \( j \), and \( y_j \) is the verifying observation for forecast \( j \).

#### 2.5.2. Root mean squared error (RMSE)

The RMSE is the square root of the averaged squared distance between the ensemble mean and the verifying observation. It is computed as follows:

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (f_j - y_j)^2} \tag{9}
\]

where \( N \), \( f_j \), and \( y_j \) are the same as in Eq. (8).

#### 2.5.3. Continuous rank probability score (CRPS)

The Continuous Rank Probability Score is a measure of accuracy for probabilistic forecasts that quantifies the difference between the predicted and the observed cumulative distributions. For a forecast of finite equiprobable members (Hersbach, 2000), the CRPS for forecast-observation pair \( m \) is computed as follows:

\[
\text{CRPS}_m = \sum_{i=1}^{m} \left[ a_i p_i^2 + \beta_i (1 - p_i)^2 \right] \tag{10}
\]

where \( m \) represents the ensemble size, \( p_i \) represents the empirical cumulative distribution value of ensemble member \( i \), \( \beta_i = \frac{1}{m} \text{for } f_{0(i)} < f < f_{(i+1)} \), \( f_{(0)} \) denotes the sorted ensemble member \( i \) (in increasing order). Furthermore, \( a_i \) and \( \beta_i \) take the values in Table 2, depending on the location of the verifying observation \( y_i \) with respect to the ensemble range.

---

**Table 1**

Perturbed parameters where SY is Specific Yield, HC is Horizontal Conductivity, VC is Vertical Conductivity.

| Parameter Name               | Unit     | Mean value | Sampling | Comment       |
|------------------------------|----------|------------|----------|---------------|
| Saturated Zone               |          |            |          |               |
| TS (SY)                      |          | 0.3125     |          | Std. dev. = 0.05 | Normal       |
| TS (HC)                      | m/s      | 1.042 × 10⁻³ | 1 × 10⁻⁶ to 0.01 | LogNormal    |
| TS (VC)                      | m/s      |            |          |               |
| "PS1" (HC)                   | m/s      | 4.030 × 10⁻⁵ | 1 × 10⁻⁶ to 0.01 | LogNormal    |
| PRVP (VC)                    | m/s      |            |          |               |
| Drainage Time Constant       | 1/s      | 5.370 × 10⁻⁸ |          | LogNormal     |
| Overland Flow                |          |            |          |               |
| Detention Storage            | mm       | 1.071      |          | Std. dev. = 0.1 | Normal       |
| Manning number               | m¹/²/s   | 1.167      |          | Std. dev. = 0.15 | Normal       |
| Leaf Area Index (LAI)        |          |            |          |               |
| Root Depth                   | m        |            |          | Relative Normal |
| Leakage Coefficient          | 1/s      | 1.000 × 10⁻⁸ | 1 × 10⁻⁶ to 1 × 10⁻⁶ | LogNormal    |

* TS: terrace sand.  
** PS: pre-quaternary sand.  
*** PRVP: terrace sand where test pumping has been performed.
5.5.2. Rank histogram (RH)

Rank Histograms (also called Talagrand diagrams) represent the frequency of the location of the verifying observation within the ensemble range (Hamill, 2001). Rank histograms are used to check for statistical consistency between the ensemble predictive distribution and the observations. In other words, for a forecasting system to be consistent (or reliable), the observed values must be a random sample from the ensemble distribution. When a forecasting system has this quality, the rank histogram is flat. Deviations from flatness tell us about different biases on the mean and dispersion of the ensemble. For example, a U-shaped rank histogram points to problems with biases of the spread (dispersion) of the ensembles, i.e., the system is overconfident. This is a common problem for ensemble forecasts, due to the fact that uncertainty is not fully accounted for Hamill (2001).

2.6. Experimental setup

The ETKF is applied using 30 ensemble members. Uncertainty is added to both input forcing and model parameters as described above. Stream discharge data from 3 gauging stations and hydraulic head data from 13 stations (locations shown in Fig. 1) are the two variables assimilated simultaneously. The assimilation frequency for groundwater head is 1 update/week and for stream discharge is every hour.

The MIKE SHE model was run in simulated real-time forecasting mode starting from 1 November 2015 and finishing on 31 July 2016, a total of 9 months. Before the start time, a 7.5-year warm-up period was simulated with the hydrological model without DA, and then subsequently a 3-month warm-up period with ensemble simulation for propagation of uncertainty. At the beginning of the forecast period, the precipitation product from the HARMONIE numerical weather prediction model is taken as the climate forcing which has a 48 h lead time. This process is repeated every 6 h at 00:00, 06:00, 12:00 and 18:00, according to each time when a weather forecast is issued.

In hindcast simulations, we run the models as they were in operational forecast mode. The real-time climate forcings are obtained every 6 h, therefore, observed stream discharge and groundwater head data being collected during the period of this 6 h are assimilated as they arrive. For the simulated real-time runs, the simulations have only been influenced by DA up to the time of forecast.

### Table 2

| (A) | \( 0 < i < m \) | \( a_i \) | \( \beta_i \) |
|-----|-----------------|----------|----------|
| \( y_i > f_{i+1} \) | \( f_i - f_{i+1} \) | 0 |
| \( y_i > f_0 \) | \( y_i - f_0 \) | \( a + \beta \) |
| \( y_i < f_{i+1} \) | \( 0 \) | \( f_i - f_{i+1} \) |
| Outlier | \( a \) | \( \beta \) |

\[ y_i < f_{i+1} \] \( f_i - f_{i+1} \) \( 0 \)

\[ y_i < f_0 \] \( y_i - f_0 \) \( 0 \)

Note that observed values that lie outside the ensemble range contribute heavier to the CRPS than values within the ensemble range. The score is also sensitive to the ensemble range. Also, the score units are the same as the variable units, in this case m³/s for discharge. The CRPS for a set of \( N \) paired forecasts and verifying observations is:

\[
\text{CRPS} = \frac{1}{N} \sum_{j=1}^{N} \text{CRPS}_j
\]

For a deterministic forecast (i.e., ensemble mean), the CRPS reduces to the mean absolute error (MAE).

### 3. Results

#### 3.1. Hindcast simulations

Discharge station 21,127 on Soholt Beck was chosen since it gave the most trusted measurements based on our past experience. Two events were selected during the 9 months of simulation time: 22–30 December 2015, and 4–10 April 2016. As seen in Fig. 3, the model without DA (deterministic run) is not able to capture the peak flow and the recession flow with high accuracy, and the peaks are usually underestimated and the recession curves overestimated. This is related to the model’s poor ability to simulate the likewise poorly understood surface runoff processes of the two creeks. For both cases, the simulated ensemble with DA is able to successfully encapsulate almost all the observation points, except for few flashy high peaks, and the ensemble mean is a good proxy of the flow if only one result is needed.

Groundwater observation wells (st. 3278 and st. 3398) were chosen since model simulation at st. 3278 gave better performance than the rest of the wells during the calibration period, and st. 3398 was among one of the average performing wells during the calibration period. Simulated groundwater head for both observation wells are presented for the entire 9 months simulation period, since, unlike stream discharge, groundwater head is difficult to separate into events. As seen in Fig. 4, the same tendency is seen in groundwater as in surface water that the model with DA clearly outperforms the model without DA by visual inspection. Similarly, the ensemble mean is able to reproduce well the observed groundwater head in terms of seasonal variations and peaks. The modelled groundwater heads with DA, measured by the ensemble mean, are about 0.5–1 m smaller than those obtained without DA.

Nash-Sutcliffe efficiency (NSE) for stream discharge simulations and root mean squared error (RMSE) for groundwater head simulations are calculated (Table 4). It is seen that DA is able to improve model results, bringing the NSE close to 1 and RMSE near zero. Hence, it is demonstrated that the coupled surface water - groundwater model employing a DA scheme with ETKF and bias adjustment is very effective.

#### 3.2. Simulated real-time hydrological forecast

Simulations of five model scenarios are presented for two cases, forecast on 26 December 2015 at 00:00, and on 6 April 2016 at 18:00, each with lead time of 48 h. These two cases are selected because they represent typical rainfall events in Denmark. Figs. 5 and 6 show simulations of stream discharge (21,127) and groundwater head (3278), respectively.

#### Table 3

Design of model scenarios. Simulations A–E each represents the model runs using observed and forecasted precipitation data, with and without data assimilation (DA) enabled, and with no perturbation of rainfall.

| Precipitation data | DA |
|--------------------|----|
| SIM A Observed     | No |
| SIM B Observed     | Yes|
| SIM C Forecast     | No |
| SIM D Forecast     | Yes|
| SIM E Forecast (with no perturbation of rainfall) | Yes |

Five scenarios (SIM A–E) are developed to investigate the impact of DA and precipitation forcing, respectively (Table 3). Two scenarios are established using observed precipitation data as an indication of the ‘perfect’ rainfall forecast in contrast to the three scenarios using NWP generated data which can be considered as ‘imperfect’ rainfall forecast. The observed rainfall data are obtained from the rain gauges in and around the study area, which can be seen in Fig. 1. Moreover, Scenario SIM E is made with unperturbed rainfall data, in order to demonstrate the impact of perturbation (uncertainty) of rainfall data in DA.
based on the 5 scenarios (SIM A-E) listed in Table 3. It is seen that for stream discharge, SIM A and B with observed rainfall have similar dynamics and are superior to models based on forecasted rainfall (SIM C-E). This is especially the case when simulating peaks, since the models based on forecasted rainfall are not able to reproduce the peak flow to a satisfying extent. In the April 2016 case, the main peak flow is completely missed by the models based on forecasted rainfall. For groundwater simulation, there are only small differences between simulations based on observed and forecasted rainfall. The DA based models (SIM B, D, and E) outperform the model without DA (SIM A, C).

It is suggested from our results that for surface water simulation, the most important factor is the quality of the forcing data, where perfect rainfall (showcased by observed rainfall) with no DA has better performance than models with DA but with imperfect rainfall (showcased by forecast rainfall); whereas for groundwater, the most important factor is the initial groundwater head at time of forecast, since models with DA have better performance than models without DA.

It is noticed that for stream discharge simulations, the models based on forecasted rainfall (SIM C-E) have considerably smaller uncertainty band than the models based on observed rainfall (SIM A, B). This is explained by the spread of the uncertainty band being directly related to the rainfall intensity. As seen in Fig 7, where the hourly observed and simulated rainfall is plotted side by side, the observed rainfall has higher intensity in both cases. Especially, in the April 2016 case, the forecasted rainfall did not perform, which resulted in flow simulations failing to represent the observations. Furthermore, it is seen that the uncertainty band of SIM E is considerably smaller than SIM D, although the same DA algorithm is applied. The only difference between SIM D and E is that in SIM E perturbation of rainfall is turned off in the forecast period. This implies that inclusion of rainfall uncertainty in forecasting stream discharge could be important. However, groundwater level forecasts (Fig. 6) do not seem to be affected by neither the rainfall intensity nor the rainfall uncertainty.

The difference between Figs. 3,4 and Figs. 5,6 can be conceived so that, for the forward model runs, the observed stream discharge and groundwater head are constantly assimilated when new data arrives; whereas for the simulated real-time runs, the simulations have only been influenced by DA up to the time of forecast, whereafter they return to the simulation without DA, which explains why the DA based models in Figs. 5 and 6 performed less well than DA based model in Figs. 3 and 4.
Aggregation (averaged along lead time) of model performance metrics.

| Stream discharge (m³/s) | All data | SIM A  | SIM B   | SIM C   | SIM D   | SIM E   |
|------------------------|----------|--------|---------|---------|---------|---------|
| bias                   | 0.000    | 0.025  | 0.009   | 0.016   | 0.012   | 0.012   |
| RMSE                   | 0.025    | 0.009  | 0.016   | 0.012   | 0.012   | 0.012   |
| CRPS                   | 0.012    | 0.012  | 0.012   | 0.012   | 0.012   | 0.012   |
| Groundwater Head (m)   | All data | SIM A  | SIM B   | SIM C   | SIM D   | SIM E   |
| bias                   | 0.324    | 0.337  | 0.322   | 0.335   | 0.322   | 0.322   |
| RMSE                   | 0.337    | 0.322  | 0.335   | 0.322   | 0.322   | 0.322   |
| CRPS                   | 0.322    | 0.322  | 0.322   | 0.322   | 0.322   | 0.322   |

3.3. Skills of model prediction

Table 5 summarizes the results of the model performance metrics considering bias, RMSE and CRPS for stream discharge st. 21,127 and groundwater well 3278. The computation is done using the complete forecast-observation pairs, which leads to a sample size of 888 for stream discharge and 605 for groundwater head, as well as the event-based forecast-observation pairs, which include 70 for stream discharge and 37 for groundwater head. The complete dataset refers to samples that are attained by the continuous generation of forecasts from 1 November 2015 to 29 July 2016, whereas the event-based dataset represents the forecasts for the five largest events during the simulation period. The results were first calculated at each lead time and then aggregated (averaged) for the 48-hour forecast period.

For stream discharge forecasts, bias is very low for the model evaluation using the complete dataset, which is most likely due to the dominating small values when there is no rain. RMSE and CRPS are notably better for models based on observed rainfall (SIM A, B) than models based on forecasted rainfall (SIM C-E). Such difference is even more obvious for the event-based dataset.

For groundwater head forecasts, DA based models (SIM B, D, and E) show much better performance than deterministic models (SIM A, C). This tendency is found in all of the performance evaluation criteria, and is equally profound for calculations made with both the complete dataset and the event-based dataset.

When it comes to rainfall forecast uncertainty, i.e. SIM D compared to SIM E, it is seen that the difference in performance measures is small for both surface water and groundwater simulations, regardless the type of dataset used. For stream discharge there is a slight increase in CRPS in SIM E without perturbation of the rainfall in the forecast period. These
results indicate that the model forecast is not so sensitive to the rainfall forcing uncertainty spread when averaged at longer time scale.

Figs. 8 and 9 show rank histograms for the complete and event-based forecast-observation dataset, respectively. The reliability of discharge is affected by the fact that too many observed values lie outside both ends of the ensemble range, as the U-shape of the rank histogram suggests. However, more often, observations are being underestimated, as a larger portion of ranks are situated on the higher end of their distribution. This is also shown in the sign of the bias, which is negative for all ensemble forecast experiments (sim B, D, E, Table 5). The rank histograms for groundwater suggest an overdispersion problem. Too many observations lie in the middle of the forecast range.
4. Discussion

One notable feature in the modeling work presented in this study is that we assimilate both surface water (stream discharge observations) and groundwater (groundwater head observations) data jointly. Although models that include the whole water cycle and simulate multiper hydrological compartments simultaneously are being used more and more, assimilating multiple sources of observed data has not become a common practice in the hydrological community. Rasmussen et al. (2015) applied a joint DA framework for both surface water and groundwater employing the MIKE SHE code, and a synthetic case study was presented to prove the concept. Later, Zhang et al. (2016) applied the DA framework developed by Ridler et al. (2014) where they investigated further details of observation types, ensemble size and localization schemes with applications to both a synthetic and a real case. Our study takes a further step to apply and investigate the DA framework for use in real-time forecasting.

There are other cases that deal with joint data assimilation of two or more variables in hydrological research, e.g., Camporese et al. (2009) tested EnKF on a coupled surface-subsurface model where stream flow and pressure head are jointly assimilated. Lee et al. (2011) used Sacramento SAC in connection with a kinetic wave routing model to assimilate stream flow and soil moisture data based on the VAR technique. Kurtz et al. (2014) employed EnKF for piezometric head and groundwater temperature in an aquifer model in Zurich. It is generally recognized from the studies mentioned above that assimilating two variables jointly brings complementary information to the system and therefore leads to better model predictions of both variables. However, some suggested that assimilating one variable alone does not necessarily improve the other variable, e.g. by assimilating stream discharge alone, it can even worsen the prediction of pressure head (Botto et al., 2018; Camporese et al., 2009). This could be due to the resulting incorrect initial groundwater state for the future time steps after the stream flow data have been assimilated at the current time step. Zhang et al. (2016) also found that assimilation of soil moisture in the unsaturated zone may have a negative impact on groundwater heads and vice versa, and they implemented a regularization procedure to solve this problem. The impact of assimilating both stream flow and groundwater head has not been tested in the present study.

In our ensemble simulations, we chose to use 30 realizations in each ensemble. Generally speaking, increasing the ensemble size may help to increase model prediction accuracy to a certain degree (Zhang et al., 2015; Xie and Zhang, 2010; Chen et al., 2013), however, there is a balance between computational cost and the appropriate ensemble size. The ensemble size is chosen mainly due to computational limitations. We had tested a larger ensemble size, which is 50, and noted only negligible difference. As a result, an ensemble that consists of 30 members is chosen. However, it needs to point out that the determination of reasonable ensemble size is usually a case by case problem (Xie and Zhang, 2010), and it is quite delicate to balance the ensemble size and the efficiency of the forecast analysis (Mitchell et al., 2002).

Data assimilation is known to be useful in hydrological model parameter estimation. In this regard, some refer to inverse problems, either deterministic or stochastic, also as DA, since it is intrinsically a way to incorporate observed data into the model (Liu and Gupta, 2007). In the present study, the hydrological models have been inversely calibrated by using a global search algorithm in AUTOCAL, and thereafter parameters have been fixed to their calibrated values. We use the term DA strictly for the numerical filtering process where hydrological states are being updated each time when new observations are acquired. In principle, hydrological states and model parameters can be updated altogether, for instance, Moradkhani et al. (2005b) introduced a dual state-parameter estimation technique based on EnKF, which has been used in numerous studies e.g. Lu et al. (2013). Other comparable methods, such as state
augmentation approach by Xie and Zhang (2010) and Shi et al. (2014) also showed promising results in reducing representative parameter uncertainty and increasing forecast accuracy. In such sense, the model parameters in the present study are not updated together with the states, which is a subject worth further investigating in our future research.

Based on our previous research, observation bias and model bias normally arise for different reasons, and thus it is important to distinguish between them to treat them differently (Ridler et al., 2018). For in-situ measurements, biases primarily stem from instrument and scaling errors. On the other hand, model structural error is usually caused by insufficient knowledge and/or incorrect conceptualization towards the system being studied. In our present study, there is bound to be some model structural error, which is not accounted for. But since the model has been calibrated, on average the model structural error has been propagated to model parameters, and this part of the uncertainty is to a large extent accounted for by perturbing the parameter values.

Synthetic studies were not performed in the Silkeborg catchment, as previous DA studies in similar hydro-geological areas in Denmark have been reported (Zhang et al., 2016; Rasmussen et al., 2015; Ridler et al., 2018; Rasmussen et al., 2016). The assimilation results from this study were consistent with Ridler et al. (2018) as the bias-aware filter was very responsive, and was able to promptly estimating the bias within one or two update cycles. In contrast, if parameters are included in the update, the bias estimates may take months, even years to resolve, likely due to the system’s increased degree of freedom (Rasmussen et al., 2015). In our opinion, a responsive bias estimate is preferred, especially in cases with drifting or cyclical biases due to instrument failure or seasonal environmental trends. Ridler et al. (2018) also concluded that it is best to include the entire unsaturated zone domain (all layers) in the state vector. We did that in the present study as well, including all 3 unsaturated zone layers. Furthermore, there is a small residual bias in the results, which indicates that bias calculated by the DA filer, and bias calculated by directly comparing simulated and observed values, have small differences.

The impact of DA together with rainfall forecasts with respect to performance of streamflow and groundwater head forecasts is probably one of the aspects that fewer people have looked into previously. In order to fulfill this objective, first, we designed 5 model scenarios, each using observed rainfall (or “perfect” rainfall) with and without DA, as well as forecasted rainfall (or “imperfect” rainfall) with and without DA. Second, we plotted two events for both surface water and groundwater simulations in Figs. 5 and 6, together with the rainfall amount in Fig. 7. This step was mainly for visual inspection. Third, we calculated the performance evaluation using a quite comprehensive metric consisting of bias, RMSE, CRPS, and rank histogram. This step was from a quantitative perspective. However, we realized that this issue can be further investigated e.g. by comparing between different types of rainfall regimes, and different methods for generating rainfall forecast, and it will be a topic for our future research.

In the present study, we use four verification scores to evaluate the quality of the hydrological forecast, namely Bias, RMSE, CRPS, and RH. We believe this is a comprehensive way to carry out model evaluation since it not only showcases the performance of the ensemble mean, but also the accuracy of the probabilistic forecast as well as the statistical consistency. Although the selection of performance metrics was mainly based on studies of model accuracy from previous research (Addor et al., 2011; Jaun and Ahrens, 2009; Roulin and Vannitsem, 2005; Bartholmes et al., 2009; Thirel et al., 2008; Shrestha et al., 2019), several recent studies suggested that it is worth including scores that consider the value of the forecasts for the end users (Ebert et al., 2013; Lopez et al., 2016). This could be a complementary addition to our subsequent research since increased model accuracy does not necessarily translate to
increased value of the model for the end users (Pappenberger et al., 2011; Roulin, 2007).

5. Conclusions

Increasing the accuracy of hydrological modeling and forecasting has always been a difficult task. One solution towards this problem is to improve the understanding of the physical processes. However, the combined effect of model uncertainties from forcing data, model structures and parameters can never be eliminated entirely, therefore a residual uncertainty will always exist. A complementary approach towards this problem is to use mathematical operations so that the model performance can be optimized. In the present study, we take the second route where the impact of data assimilation (DA) is demonstrated in a coupled surface water - groundwater model. The results show that DA can significantly reduce model bias and therefore improve model performance for both surface water and groundwater simulations.

When a hydrological model is running in forecast mode, usually the meteorological forcing is given, and therefore the impact of rainfall forecast quality is not easily seen. In the present study, we have the possibility of illustrating the impact of rainfall forecast from two aspects in conjunction with DA: First, comparing the impact of numerical weather prediction models (imperfect forecast) on simulated hydrological responses with the simulations based on observed rainfall (perfect forecast); and second comparing the impact of rainfall forecast with and without perturbation. The results in general show that for surface water, the most important factor is the quality of the forcing data, whereas for groundwater benefits brought by DA is more obvious. This is likely due to the difference in dynamics (time scales) in surface water and groundwater. Simulating discharge peaks is much more dependent on the precipitation input, whereas the initial condition is less important.

Groundwater head for the next 48 h, on the other hand, does not respond very much to rainfall in a particular day, but depends almost entirely on initial conditions. In addition, we also find that inclusion of rainfall uncertainty perturbation may be important for hydrological forecast of a single event, but for long-term continues model runs, the impact is insignificant. This could be explained by dividing the rainfall uncertainty into two sources: bias and variance. The perturbation is carried out only to boost the rainfall variance. As a result, for the performance of a single event, both bias and variance can be significant; whereas for the long-term effect, the rainfall bias plays a more dominating role.

We have compared the performance of model predictions using the complete dataset and a number of events, respectively. For both stream discharge and groundwater head simulations, similar conclusions are obtained for the two data sets when comparing performance measures for the different scenarios.

Acknowledgement

This work has been supported by the National Key R&D Program of China, project number 2016YFC0401404, and 2016YFE0102400; the National Natural Science Foundation of China (NSFC), grant number 41701042; and the Innovation Fund Denmark project "HydroCast Hydrological Forecasting and Data Assimilation”, Contract No. 0603-00466B.

References

Abaza, M., Anctil, F., Fortin, V., Turcotte, R., 2014. Sequential streamflow assimilation for short-term hydrological ensemble forecasting. J. Hydrol. 519, 2692–2706.

Abbott, M.B., Bathurst, J.C., Cunge, J.A., Oconnell, P.E., Rasmussen, J., 1986a. An introduction to the European hydrological system - systeme hydrologique Europeen, SHE 1.2. Structure of a physically-based, distributed modeling system. J. Hydrol. 87 (1–2), 61–77.
Kiddmoe, J., Refsgaard, J.C., Trolldborg, L., Seaby, L.P., Escriva, M.M., 2013. Climate change impact on groundwater levels: ensemble modelling of extreme values. Hydro. Earth Syst. Sci. 17 (4), 1619–1634.

Kiddmoe, J., Trolldborg, L., Refsgaard, J.C., Bischoff, N., 2015. Coupling of a distributed hydrological model with an urban storm water model for impact analysis of forced infiltration. J. Hydro. 525, 506–520.

Kurtz, W., Hendrichs Fransson, H.J., Kaiser, H.P., Vereneck, H., 2014. Joint assimilation of temperature and precipitation data for improved modeling of river–interior interactions. Water Resour. Res. 50 (2), 1655–1668.

Laio, F., Tamae, S., 2007. Verification tools for probabilistic forecasts of continuous hydro- logical variables. Hydro. Earth Syst. Sci. 11 (4), 1257–1277.

Lee, H., Seo, D., Koren, V., 2011. Assimilation of streamflow and in situ soil moisture data into operational distributed hydrological models: effects of uncertainties in the data and initial model soil moisture states. Adv. Water Resour. 34 (12), 1597–1615.

Liu, Y.Q., Gupta, H.V., 2007. Uncertainty in hydrological modeling: toward an integrated approach. J. Hydrol. 337 (1–2), 1–18.

Lopez, A., Perez, E.C.d., Bazo, J., Suarez, P., Hurk, B.d., Aalst, M., 2013. Bridging forecast verification and humanitarian decisions: a valuation approach for setting up action-oriented early warnings. Weather Clim. Extremes https://doi.org/10.1016/j.wce.2013.06.006.

Lu, G.H., Wu, Z.Y., Wen, L., Lin, C.A., Zhang, J.Y., Yang, Y., 2008. Real-time flood forecast and flood alert map over the Huaihe river basin in china using a coupled hydro- meteorological modeling system. Sci. China Ser. E Technol. Sci. 51 (7), 1049–1063.

Lu, H.S., Hou, T., Horton, R., Zhu, Y.H., Chen, X., Jia, Y.W., Wang, F., Xu, X.L., 2013. The streamflow estimation using the xiaojiang rainfall runoff model and dual state-parameter estimation method. J. Hydro. 480, 102–114.

Madsen, H., 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. J. Hydro. 235 (2–4), 276–288.

Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Adv. Water Res. 26 (2), 205–212.

Maggioni, V., Angnapost, E.N., Reichle, R.H., 2012. The impact of model and rainfall forcing errors on characterising soil moisture uncertainty in land surface modelling. Hydro. Earth Syst. Sci. 16 (10), 3949–3951.

McMillan, H., Krueger, T., Freer, J., 2012. Benchmarking observational uncertainties for hydrology: rainfall, discharge and water quality. Hydro. Process. 26 (20), 4078–4111.

Mitchell, H.L., Houtekamer, P.L., Pellerin, G., 2002. Ensemble size, balance, and model–error representation in an ensemble Kalman filter. Mon. Weather Rev. 130 (11), 4171–4190.

Moradkhani, H., Hsu, K.L., Gupta, H., Sorooshian, S., 2005a. Uncertainty assessment of hydrological model states and parameters: sequential data assimilation using the particle filter. Water Resour. Res. 41 (5), 1–17.

Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005b. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Adv. Water Res. 28 (2), 135–147.

Murphy, A.H., 1993. What is a good forecast? – An essay on the nature of goodness in weather forecasting. Weather Forecast. 8 (2), 281–293.

Pappenberger, F., Thielen, J., Del Medico, M., 2011. The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the European River Basins. J. Hydrol. 404 (1–2), 1091–1102.

Pasetto, D., Camporese, M., Putti, M., 2012. Ensemble kalman filter versus particle filter for a physically-based coupled surface–subsurface model. Adv. Water Res. 47, 1–12.

Rasmussen, J., Madsen, H., Jensen, K.H., Refsgaard, J.C., 2015. Data assimilation in integrated hydrological modeling using ensemble kalman filtering: evaluating the effect of ensemble size and localization on filter performance. Hydro. Earth Syst. Sci. 19 (7), 2999–3013.

Rasmussen, J., Madsen, H., Jensen, K.H., Refsgaard, J.C., 2016. Data assimilation in integrated hydrological modelling in the presence of observation bias. Hydro. Earth Syst. Sci. 20 (5), 2103–2118.

Refsgaard, J.C., 1997. Validation and intercomparison of different updating procedures for real-time forecasting. Nord. Hydro. 28 (6), 68–84.

Refsgaard, J.C., Ronberg, D., Markussen, L.M., 1983. Application of Kalman Filter to Real-Time Operation and to Uncertainty Analyses in Hydrological Modelling. IAHS, pp. 275–282. Publication No. 147.

Ridler, M.E., van Velzen, N., Hummel, S., Sandholt, L., Falk, A.K., Heemink, A., Madsen, H., 2014. Data assimilation framework: linking an open data assimilation library (OpenDA) to a widely adopted model interface (OpenPfri). Environ. Model. Softw. 57, 76–89.

Ridler, M.E., Zhang, D., Madsen, H., Kiddmoe, J., Refsgaard, J.C., Jensen, K.H., 2018. Bias-aware data assimilation in integrated hydrological modelling. Hydro. Res. 49 (4), 989–1004.

Roulin, E., 2007. Skill and relative economic value of medium-range hydrological ensem- ble predictions. Hydro. Earth Syst. Sci. 11 (2), 725–737.

Roulin, E., Vannitsem, S., 2005. Skill of medium-range hydrological ensemble forecasts. J. Hydrometeorol. 6 (5), 729–744.

Schaeke, J.G., Hamill, T.M., Buizza, R., Clark, M., 2007. The hydrological ensemble pre- diction experiment. Bull. Am. Meteorol. Soc. 88 (10), 1541–1557.

Scharling, M., 2000. Climate Grid – Denmark, Norms 1961–90, Month and Annual Values. Danish Meteorological Institute, Copenhagen, Denmark Technical report.

Shi, D., Davis, K.J., Zhang, F.Q., Duffy, C.J., Yu, X., 2014. Parameter estimation of a physically based land surface hydrological model using the ensemble Kalman filter: a synthetic experiment. Water Resour. Res. 50 (1), 706–724.
Shrestha, D.L., Robertson, D.E., Wang, Q.J., Pagano, T.C., Hapuarachchi, H.A.P., 2013. Evaluation of numerical weather prediction model precipitation forecasts for short-term streamflow forecasting purpose. Hydrol. Earth Syst. Sci. 17 (5), 1913-1931.

Thielen, J., Schaake, J., Hartman, R., Buizza, R., 2008. Aims, challenges and progress of the hydrological ensemble prediction experiment (HEPEX) following the third HEPEX workshop held in Stresa 27 to 29 June 2007. Atmos. Sci. Lett. 9 (2), 29–35.

Thirel, G., Rousset-Regimbeau, F., Martin, E., Habets, F., 2008. On the impact of short-range meteorological forecasts for ensemble streamflow predictions. J. Hydrometeorol. 9 (6), 1301–1317.

Uittenbroek, C.J., Janssen-Jansen, L.B., Runhaar, H.A.C., 2013. Mainstreaming climate adaptation into urban planning: overcoming barriers, seizing opportunities and evaluating the results in two Dutch case studies. Reg. Environ. Change 13 (2), 399–411.

Vrugt, J.A., Gupta, H.V., Nuallaine, B.O., 2006. Real-time data assimilation for operational ensemble streamflow forecasting. J. Hydrometeorol. 7 (3), 548–565.

Xie, X.H., Zhang, D.X., 2010. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Adv. Water Resour. 33 (6), 678–690.

Zhang, D.H., Madsen, H., Ridler, M.E., Kidmose, J., Jensen, K.H., Refsgaard, J.C., 2016. Multivariate hydrological data assimilation of soil moisture and groundwater head. Hydrol. Earth Syst. Sci. 20 (10), 4341–4357.

Zhang, D.H., 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. Adv. Water Resour. 86, 400–413.

Zolina, O., Simmer, C., Gulev, S.K., Kollet, S., 2010. Changing structure of European precipitation: longer wet periods leading to more abundant rainfalls. Geophys. Res. Lett. 37, 1-5.