Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space

Valentijn R. N. Pauwels[1] and Gabriëlle J. M. De Lannoy[1]

Received 13 November 2008; revised 12 May 2009; accepted 29 May 2009; published 19 August 2009.

[1] The optimization of hydrologic models using the ensemble Kalman filter has received increasing attention during the last decade. The application of this algorithm is straightforward when the relationship between the state variables and the observations is linear, in other words, when the observations can be directly mapped onto the state space. However, when this relationship is nonlinear, a number of methods can be derived in order to perform this transfer. Up till now, it has not been demonstrated which of these methods is recommended for discharge assimilation with the ensemble Kalman filter. The objective of this paper is to analyze these methods for conceptual rainfall-runoff models in a small-scale catchment. The study has been performed in the Bellebeek catchment (86.36 km^2) in Belgium, using two time series models and one conceptual rainfall-runoff model. A first analysis of the algorithms has been performed using the one time step ahead discharge predictions. The results indicate that linearization of the storage-discharge relationship (the observation system) should be avoided if discharge data are assimilated using the ensemble Kalman filter. Further, assimilating discharge data into conceptual rainfall-runoff models for small catchments does not work well when a unit hydrograph is used for runoff routing. This can be explained by the stronger impact of the model error (caused by errors in the forcings, model structure, and parameters), accumulated over the duration of the unit hydrograph, as compared to the impact of erroneous initial conditions. A second analysis using longer lead times has led to the conclusion that, for the type of catchment and model used in this study, the accuracy of the meteorological forcings is more important than an accurate estimation of the model initial conditions through data assimilation.

Citation: Pauwels, V. R. N., and G. J. M. De Lannoy (2009), Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space, *Water Resour. Res.*, *45*, W08428, doi:10.1029/2008WR007590.

1. Introduction

[2] Floods are among the most common natural disasters in the world. For example, in the Northern part of Belgium, eight floods causing severe economic loss occurred during the last 15 years. Besides infrastructure protecting measures, one indispensable tool to manage floods is the use of rainfall-runoff models to predict the arrival of discharge peaks. Typically, such models conceptualize the catchment as a number of reservoirs (for example a soil water reservoir and a deep groundwater reservoir), which receive water either from the atmosphere through infiltration or from other reservoirs, and release water, either to the atmosphere through evapotranspiration, to other reservoirs, or to the catchment river system. Examples of such models are the Hydrologiska Byrns Vattenbalansavdelning (HBV) model [*Linström et al.*, 1997] and the Probability Distributed Model (PDM) [*Moore*, 2007].

[3] In the application of these models, a number of problems arise. A first problem is the determination of the various parameters used in the model equations. Ideally, these parameters should be measured at the study site, and used in the model application. However, because of a number of constraints this approach is not possible. First, a number, if not all, of the model equations require catchment average parameters. These catchment averages can only be obtained through the interpolation of in situ point measurements, which will lead to nonnegligible error in the parameter estimate. Second, a number of the model equations are determined empirically, meaning that they have no real physical basis. It is thus impossible to directly measure the parameters for these models. As a consequence, the only way to determine the model parameters is to compare the modeled discharge values to observations and to adapt the parameter values until an optimal fit has been obtained. A large number of methods have been developed for this purpose, for example the Parameter Estimation (PEST) method [*Doherty*, 2001], Bayesian recursive parameter estimation [*Thiemann et al.*, 2001], the Shuffled Complex Evolution (SCE-UA) algorithm [*Duan et al.*, 1994; *Yapo et al.*, 1998; *Vrugt et al.*, 2003a, 2003b] the Multiple Start

[1]Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium.

0043-1397/09/2008WR007590

Simplex (MSX) and local Simplex methods [*Gan and Biftu*, 1996], simulated annealing [*Thyer et al.*, 1999], genetic algorithms [*Reed et al.*, 2000, 2003], and Particle Swarm Optimization [*Gill et al.*, 2006].

[4] A second problem in the application of rainfall-runoff models, even if they are thoroughly calibrated, is that there will remain a mismatch between the modeled and the observed discharge. This is due to the fact that a model is always a simplification of the physical reality, and to uncertainties in the model initial conditions, parameter values, and meteorologic forcing data. To reduce this mismatch in real time applications, model state variables can be regularly updated using external observations. This is commonly referred to as data assimilation. One type of external observation is the surface soil moisture content. A number of studies have already put the use of remotely sensed soil moisture data for flood forecasting purposes in practice [*Pauwels et al.*, 2001, 2002; *Parajka et al.*, 2006]. However, this type of assimilation requires the inversion of the remotely sensed signal into soil moisture values, either off-line or during the model application, which is a complicated task. An alternative is the use of discharge records to update the model state variables.

[5] In order to assimilate discharge rates into hydrologic models, a number of algorithms can be used. A first possibility is the use of the Extended Kalman Filter (EKF), of which a number of examples can be found in the work by *Wood and Szöllösi-Nagy* [1980]. The major drawbacks of this approach are the high computational demand for the propagation of the background error covariance (especially for large system state vectors), and the neglection of higher-order derivatives for the background error covariance propagation and the mapping of the observational information (the observed discharge) to the model state variables. In order to overcome these drawbacks, the ensemble Kalman filter (EnKF) has been developed [*Evensen*, 1994]. The EnKF propagates an ensemble of model realizations through time, and estimates the background error covariance matrix from the ensemble statistics. Examples of studies that have used to EnKF to assimilate discharge rates into hydrologic models can be found in the work by *Pauwels and De Lannoy* [2006], *Vrugt et al.* [2006], and *Weerts and El Serafy* [2006]. The major drawback of the EnKF is the underlying assumption that the distribution of the model states is Gaussian [*Weerts and El Serafy*, 2006]. The Particle Filter (PF) does not require a specific form for the state distribution, but has as major drawback that the distribution of the particle weights quickly becomes strongly skewed, and a resampling algorithm needs to be applied [*Weerts and El Serafy*, 2006]. Examples of the use of the PF to update the states of hydrologic models using discharge data are given by *Moradkhani et al.* [2005] and *Weerts and El Serafy* [2006]. An intercomparison between the EnKF and the PF was performed by *Weerts and El Serafy* [2006], and led to the conclusion that EnKF is more robust, less sensitive to the choice of the model and measurement errors, and outperformed the PF in real data experiments. For this reason, this paper focusses on the application of the EnKF for hydrologic model state updating using discharge observations.

[6] The most complicated task in the application of the Kalman filter for discharge assimilation is the correct mapping of the discharge observations to the state variables. In the EKF this mapping is performed through the calculation of a Jacobian matrix, which contains the first derivatives of the observation simulations with respect to the state variables. This is referred to as linearization of the observation system. In hydrologic models, the relationship between the storages (or soil moisture values) and the discharge is nonlinear and not explicitly known, and this Jacobian matrix has to be calculated numerically. The EKF is still widely used for operational applications, for example in the European Centre for Medium-Range Weather Forecasts (ECMWF) global data assimilation system [*Reichle*, 2008], or in the French National Center for Meteorological Research (CNRM) Interactions between the Soil Biosphere and Atmosphere (ISBA) model [*Sabater et al.*, 2007]. This implies that, if observations with a nonlinear relationship with the state variables are to be assimilated into these models, a linearization of the observation system has to be performed. On the other hand, if the EnKF is applied, the approach suggested by *Evensen* [1994] can be used, in which the linearization of the observation system is bypassed using the ensemble statistics. This approach avoids the cumbersome calculation of all individual elements of the Jacobian matrix. Although this approach is commonly used in hydrologic applications with nonlinear observation systems [*Reichle et al.*, 2002a; *Weerts and El Serafy*, 2006; *Durand and Margulis*, 2007], it has never been shown whether or not it outperforms a linearization of the observing system which relates the soil water state to the discharge.

[7] The objective of this paper is to assess which of these methods is the most appropriate for ensemble-based discharge assimilation. In order to demonstrate the impact of the linearization described above on the performance of the EnKF, discharge data are first assimilated into a very simple time series–based rainfall-runoff model, constructed with a linear and a nonlinear observation system. The methods are then applied to an operational rainfall-runoff model, the HBV model, which uses unit hydrograph–based routing. This implies that the modeled discharge at a certain time step is a function of not only the storages at that time step, but also of the storages up till a number of time steps before the observation. If discharge rates are assimilated, the state vector thus needs to be distributed in time. The use of these different models allows an assessment of whether the recommended assimilation method depends on the model structure or not.

2. Site and Data Description

[8] The study was performed in a subcatchment of the Dender catchment in Belgium, more specifically the Bellebeek catchment. The elevation ranges between 10 and 110 m. Soil texture is predominantly loam (75%), and the land use is predominantly agriculture (63%) and pasture (22%). The surface area of the catchment is 87.36 km^2. Figure 1 shows the location of the catchment together with a Digital Elevation Model (DEM) of the area.

[9] Discharge was measured at the outlet of the catchment. Precipitation as well as the required meteorologic data to calculate potential evapotranspiration according to the methods of Hargreaves as explained by *Shuttleworth* [1992], were measured at the meteorologic station of

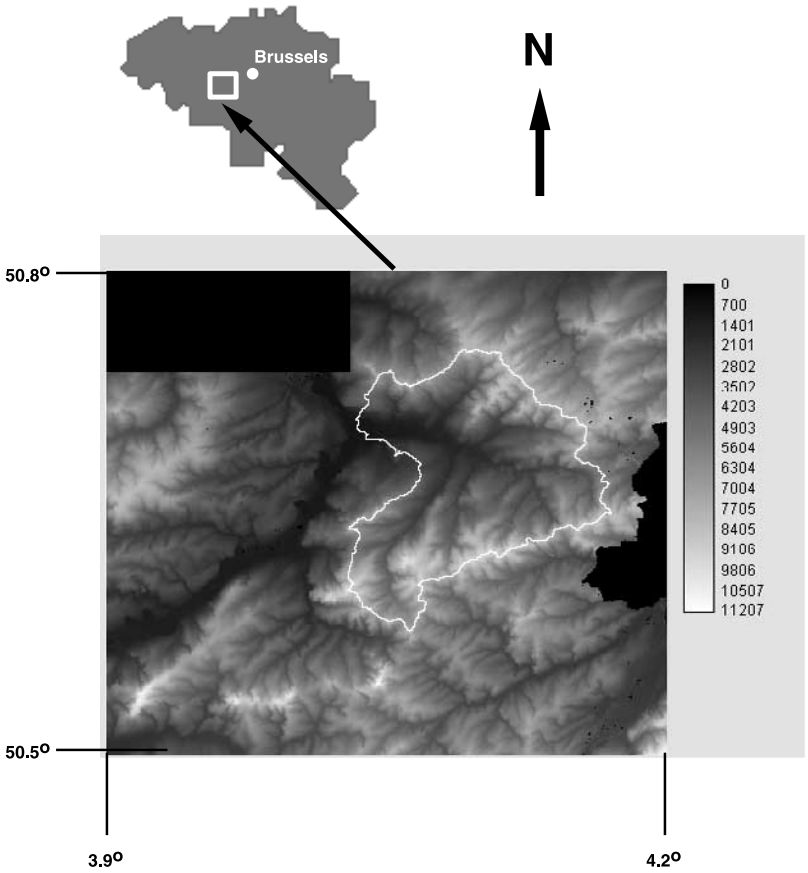

Figure 1. Location of the study site and DEM of the study area. Units are in centimeters above sea level. The white lines in the DEM represent the boundaries of the catchment of the Bellebeek.

Liedekerke, located close to the outlet of the basin. All data sets were available at an hourly time step from August 2006 through July 2007.

3. Rainfall-Runoff Models

[10] Three different rainfall-runoff models are applied in this study, using a 1 h time step. A short description of each model is given in this section, while a listing of all the model equations is provided in Appendix A.

3.1. Time Series Model

[11] A first, very simple, rainfall-runoff model applied in this study consists of an Autoregressive Moving Average with Exogenous inputs (ARMAX) expression relating the precipitation to the catchment storage S (mm). This catchment storage is then related to the discharge q ($m^3\ s^{-1}$) using two different observation systems, more specifically a linear and a nonlinear equation. A detailed model description is provided in section A1.

3.2. HBV Model

[12] The Hydrologiska Byrns Vattenbalansavdelning (HBV) model [*Linström et al.*, 1997], of which Figure 2 shows a schematic, divides the catchment into a number of reservoirs. The state variables are the amount of water stored in the soil reservoir ($S(t)$, m^3), the slow reservoir ($S_1(t)$, m^3), and the fast reservoir ($S_2(t)$, m^3). The HBV model allows reinfiltration of the surface runoff, and uses a triangular unit hydrograph for the routing of the overland flow. The time delay between the generation of surface runoff and the arrival at the catchment is thus modeled, albeit in a relatively simple manner. A detailed model description is provided in section A2.

4. Ensemble Kalman Filter

4.1. Description

[13] A description of the ensemble Kalman filter can be found in the work by *Reichle et al.* [2002b]. A short description is given here.

[14] The state variables of a single forecast ensemble member are stored in the vector $\mathbf{x}_k^{if}$. k indicates the time step, i the ensemble member. The superscript ^{f} refers to the forecast variables. The background error covariance $\mathbf{P}_k^f$ is calculated to estimate the forecast uncertainty:

$$\begin{cases} \mathbf{P}_k^f = \dfrac{1}{N-1}\mathbf{D}_k\mathbf{D}_k^{\mathrm{T}} \\ \mathbf{D}_k = \left[\mathbf{x}_k^{1f} - \mathbf{x}_k^f, .., \mathbf{x}_k^{Nf} - \mathbf{x}_k^f\right] \\ \mathbf{x}_k^f = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\mathbf{x}_k^{if} \end{cases} \quad (1)$$

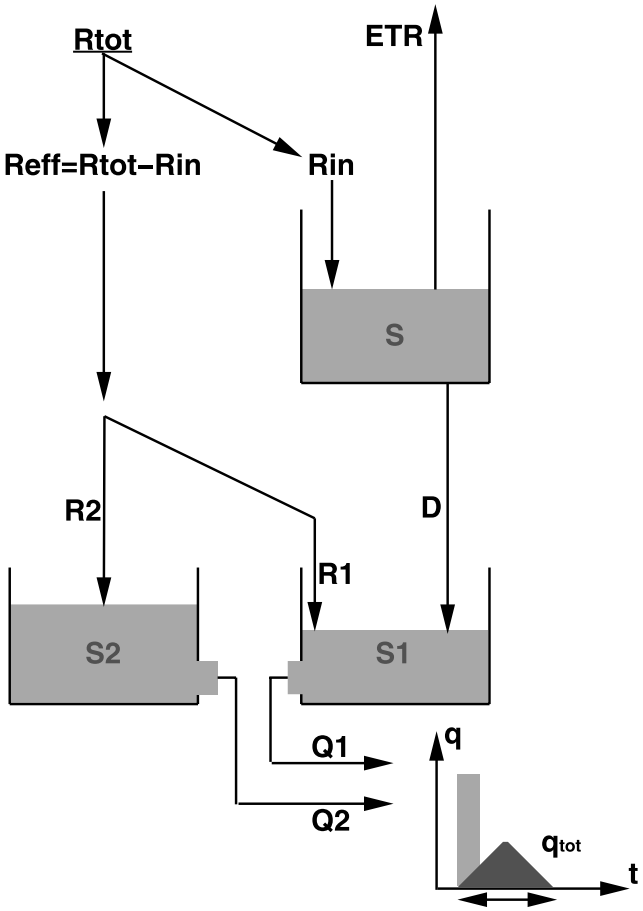

Figure 2. Schematic of the HBV model.

N is the number of ensemble members and the superscript $^{\mathrm{T}}$ indicates the transpose operator. The Kalman gain $\mathbf{K}_k$ is then calculated:

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} \left[\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k \right]^{-1} \tag{2}$$

$\mathbf{R}_k$ is the observation noise covariance.

[15] Equation (2) was originally derived for the optimization of linear systems [*Kalman*, 1960]. In the original formulation, $\mathbf{H}_k$ is a transfer function, describing the linear relationship between the state variables and the observations:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \tag{3}$$

$\mathbf{x}_k$ is the vector with the state variables, $\mathbf{y}_k$ is the vector with the observations, and $\mathbf{v}_k$ is the observation error. In order to allow an application for nonlinear systems, the Extended Kalman Filter was developed [*Gelb*, 1974]. In the application of the Extended Kalman Filter, equation (2) is still valid. Further, $\mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}}$ and $\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}}$ are obtained through explicit matrix multiplications as well, where $\mathbf{H}_k$ is a Jacobian matrix, relating the observations to the state variables. This matrix is calculated as

$$\mathbf{H}_k[i,j] = \left. \frac{\partial \mathbf{h}_k(\mathbf{x}_k^f)[i]}{\partial \mathbf{x}_k[j]} \right|_{\mathbf{x}_k^f[j]} \tag{4}$$

i is the row and j the column number. $\mathbf{H}_k(.)$ is the relationship between the model states and the observations. The linearized observation operator $\mathbf{H}_k$ for the different models in this paper will be derived in section 6. In the application of the Extended Kalman Filter, the model needs to be linearized, in order to propagate the background error covariance matrix $\mathbf{P}_k$. In order to bypass closure problems caused by this linearization, the ensemble Kalman filter has been developed [*Evensen*, 1994]. In the application of the ensemble Kalman filter, the matrix multiplications in equation (2) can be reduced. In order to avoid the linearization of the observation system and to reduce computational requirements, *Evensen* [1994] suggested to calculate $\mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}}$ as the forecast cross covariance between the state and the measurement predictions, and to calculate $\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}}$ as the error covariance matrix of the measurement predictions. It is important to note that it cannot be theoretically proven that this approach is mathematically superior to the explicit calculation of $\mathbf{H}_k$ through equation (4) and the explicit matrix multiplications in equation (2). This covariance-based approach will be further explained in section 5.

[16] Using the Kalman gain, the states of the individual ensemble members are then updated:

$$\mathbf{x}_k^{ia} = \mathbf{x}_k^{if} + \mathbf{K}_k \left[\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k^{if}) + \mathbf{v}_k^i \right] \tag{5}$$

$\mathbf{x}_k^{ia}$ is the analysis vector (after the update), and $\mathbf{v}_k^i$ is a realization of the observation error. Using the analyzed state, the forecast at the next time step can then be calculated:

$$\mathbf{x}_{k+1}^{if} = \mathbf{f}_k(\mathbf{x}_k^{ia}) + \mathbf{w}_k^i \tag{6}$$

$\mathbf{f}_k(.)$ is the model, and $\mathbf{w}_k^i$ a realization of the model error.

4.2. Application in This Study

[17] For the time series models (TSMs), there is at each time step one state variable, so the application of the EnKF is straightforward. More specifically, at each time step $t = k\triangle t$ the state vector can be written as

$$\mathbf{x}_k = [S(k\triangle t)] \tag{7}$$

However, for the HBV model, as stated in section A2, there are three state variables per time step, i.e., the storages in the three reservoirs $S(k\triangle t)$, $S_1(k\triangle t)$, and $S_2(k\triangle t)$. The objective is to update the model at each time step, so one runoff observation per time step is used. However, as stated in section 3, a unit hydrograph is used for runoff routing. This implies that at each time step the modeled discharge is a function of not only the state variables at the time step, but also of the state variables at a number of previous time steps (in this case 14 time steps). As stated in section A1, the amount of time steps to take into account has been determined through calibration of unit hydrograph–based models for the study site [*Ferket*, 2008]. A methodology to take into account this time delay is described by *Pauwels and De Lannoy* [2006]. More specifically, the state vector at

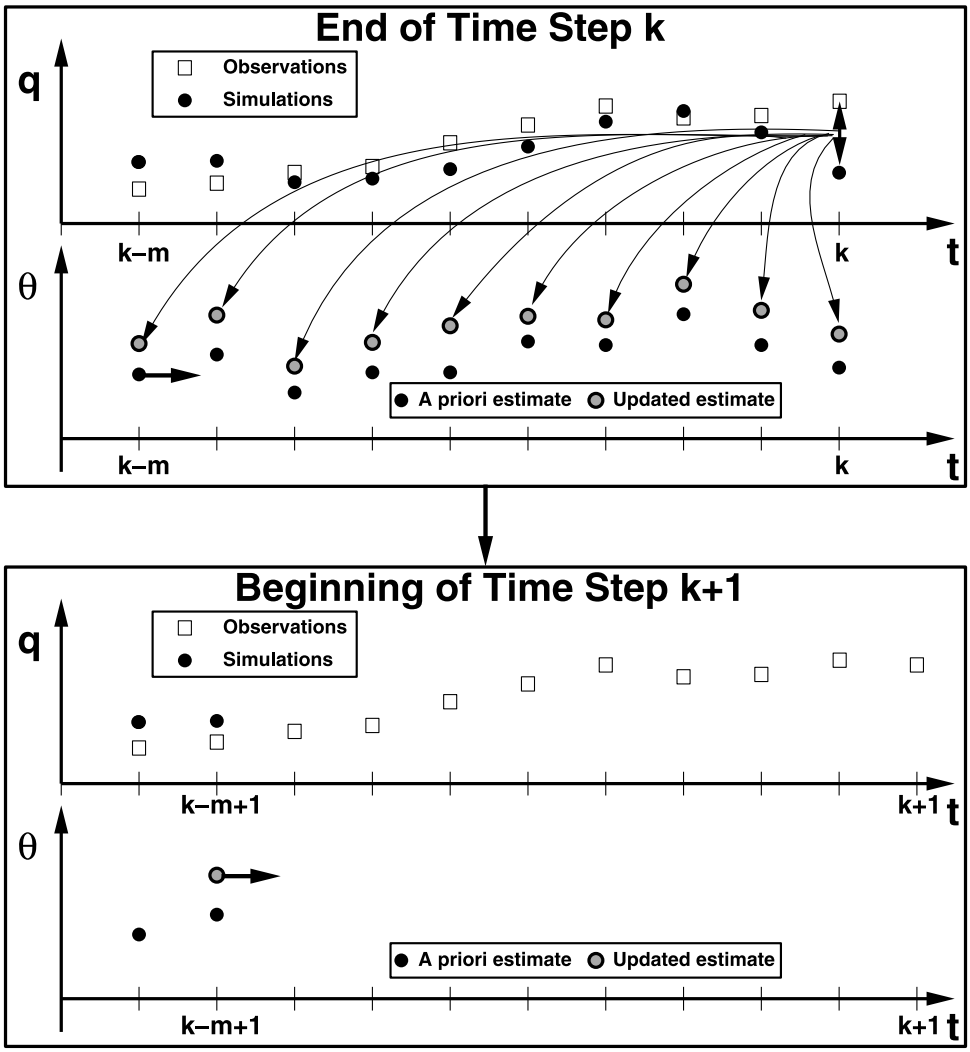

Figure 3. Assimilation of discharge observations into the HBV model, taking into account the time delay between the generation of runoff and the arrival at the catchment outlet. Here q is the routed discharge, and θ is one of the storages (S, S_1, or S_2). Here m is the duration of the unit hydrograph. For simplicity, only one storage is shown.

a certain time $t = k\triangle t$ is expanded by the reservoir storages of time steps $k - 1$ through $t - 13$:

$$\mathbf{x}_k = [S(k\triangle t)S_1(k\triangle t)S_2(k\triangle t) \\ ..S((k-13)\triangle t)S_1((k-13)\triangle t)S_2((k-13)\triangle t)]^{\mathrm{T}} \quad (8)$$

For all models, the observation vector at time step t contains the discharge at time step t:

$$\mathbf{y}_k = [q^o(k\triangle t)] \quad (9)$$

[18] Figure 3 further explains this methodology for the HBV model. At each time step $k - 13$ the model is applied for fourteen extra time steps, and the total discharge at time step k is calculated. The forecasted discharge at time step k is then compared to the observed value, and the ensemble Kalman filter is used to update the storages from time step $k - 13$ through time step k. The analyzed storages at time step $k - 13$ are then the initial conditions for the model at time step $k - 12$. Using this initial condition, the same procedure is then repeated for a new model application, starting at time step $(k - 12)$.

5. Full Ensemble-Based Covariance Calculation of the Kalman Gain

5.1. Derivation

[19] Equations (2) and (4) show that, in order to calculate the Kalman gain, a Jacobian matrix is needed. For many applications, as is the case in this study, an explicit analytical expression relating the state to the observation is not available. This first derivative thus needs to be calculated analytically, through a first-order Taylor expansion, which is a computationally demanding task. Therefore, a different methodology to calculate $\mathbf{K}_k$ is proposed, which has been used before to deal with highly nonlinear forward models for satellite observations [*Reichle et al.*, 2002a].

[20] The nominator in the expression for the Kalman gain (equation (2)) can be rewritten as

$$\mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} = \mathrm{E}\left[\left(\mathbf{x}_k - \mathbf{x}_k^f\right)\left(\mathbf{x}_k - \mathbf{x}_k^f\right)^{\mathrm{T}}\right]\left(\left.\frac{\partial \mathbf{h}_k(\mathbf{x}_k^f)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k^f}\right)^{\mathrm{T}} \tag{10}$$

If $\mathbf{h}_k(.)$ is a linear model with respect to the state variables, this can be rewritten as

$$\mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} = \mathrm{E}\left[\left(\mathbf{x}_k - \mathbf{x}_k^f\right)\left(\mathbf{h}_k(\mathbf{x}_k) - \mathbf{h}_k(\mathbf{x}_k^f)\right)^{\mathrm{T}}\right] \tag{11}$$

This is the covariance between the errors in the modeled states and the modeled discharge, $\sigma_{\mathbf{x}_k^f \mathbf{q}_k^f}$. This covariance can be calculated using the ensemble in a similar way as the background error covariance is calculated (equation (1)). Further, the first term in the denominator of equation (2) can be written as

$$\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} = \left(\left.\frac{\partial \mathbf{h}_k(\mathbf{x}_k^f)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k^f}\right)\mathrm{E}\left[\left(\mathbf{x}_k - \mathbf{x}_k^f\right)\left(\mathbf{x}_k - \mathbf{x}_k^f\right)^{\mathrm{T}}\right] \cdot \left(\left.\frac{\partial \mathbf{h}_k(\mathbf{x}_k^f)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k^f}\right)^{\mathrm{T}} \tag{12}$$

Again, for a linear model with respect to the states, this can be rewritten as

$$\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} = \mathrm{E}\left[\left(\mathbf{h}_k(\mathbf{x}_k) - \mathbf{h}_k(\mathbf{x}_k^f)\right)\left(\mathbf{h}_k(\mathbf{x}_k) - \mathbf{h}_k(\mathbf{x}_k^f)\right)^{\mathrm{T}}\right] \tag{13}$$

This is the variance of the modeled discharge ($\sigma^2_{\mathbf{q}_k^f}$). Again, this variance can be calculated in a similar way as equation (1). Using these approximations, the expression for the Kalman gain becomes

$$\mathbf{K}_k = \sigma_{\mathbf{x}_k^f \mathbf{q}_k^f}\left(\sigma^2_{\mathbf{q}_k^f} + \mathbf{R}_k\right)^{-1} \tag{14}$$

5.2. Mathematical Interpretation

[21] Let us assume we have a very simple system, with one state variable (the soil moisture θ) and one model output, the discharge (q). Observations of the discharge are used to update the modeled soil moisture. In this case, the matrix $\mathbf{H}_k$ becomes as scalar variable, and equation (4) can be written as

$$H_k = \left.\frac{\partial q_k^f}{\partial \theta_k}\right|_{\theta_k^f} \tag{15}$$

Using equation (2), the Kalman gain can then be written as

$$K_k = \frac{P_k^f H_k}{H_k^2 P_k^f + R_k} \tag{16}$$

Assume the asymptotical case where the observation is perfect, in other words, R_k is zero. The Kalman gain then becomes

$$\lim_{R_k \to 0} K_k = \frac{P_k^f H_k}{H_k P_k^f H_k} = \frac{1}{H_k} = \left.\frac{\partial \theta_k^f}{\partial q_k}\right|_{q_k^f} \tag{17}$$

The state variable is then updated as (equation (5))

$$\theta_k^{ia} = \theta_k^{if} + \left.\frac{\partial \theta_k^f}{\partial q_k}\right|_{q_k^f}\left(q_k^o - q_k^{if}\right) \tag{18}$$

On the other hand, if equation (14) is used, the Kalman gain in this case becomes

$$K_k = \frac{\sigma_{\theta_k^f q_k^f}}{\sigma^2_{q_k^f} + R_k} \tag{19}$$

Again, assume the asymptotical case where the observation is perfect. The Kalman gain then becomes

$$\lim_{R_k \to 0} K_k = \frac{\sigma_{\theta_k^f q_k^f}}{\sigma^2_{q_k^f}} \tag{20}$$

This equation can be interpreted as follows. If, in a plot with the modeled discharge as abscissa, and with the modeled soil moisture as ordinate, the results of all ensemble members would be plotted, the Kalman gain calculated with equation (20) would simply be the regression through all these model results. In this case the state update equation becomes

$$\theta_k^{ia} = \theta_k^{if} + \frac{\sigma_{\theta_k^f q_k^f}}{\sigma^2_{q_k^f}}\left(q_k^o - q_k^{if}\right) \tag{21}$$

Figure 4 further explains the difference between these two approaches (the use of equation (18) versus equation (21)). Since, in the generation of the ensemble, parameter values as well as forcing variables are disturbed, the relationship between the soil moisture and the discharge will be different for each ensemble member. In the original derivation of the EKF, there is only one ensemble member, and the Kalman gain calculated using equation (17) is nothing else than the slope of the tangent line evaluated at (q_k^{if}, θ_k^{if}). In Figure 4 this slope is denoted α_1. However, in equation (17), ensemble averages of θ_k and q_k are used in order to calculate the first derivative. This slope (denoted α_2 in Figure 4) will thus not be equal to α_1. On the other hand, if the Kalman gain is calculated using the regression through the results of the ensemble members (equation (20)), again a different slope (α_3 in Figure 4) will be obtained. Figure 4 shows that, if either of these latter two slopes (α_2 and α_3) are used, and if the relationship between the discharge and the soil moisture were linear, the assimilation of a perfect observation for each ensemble member would lead to an updated state variable

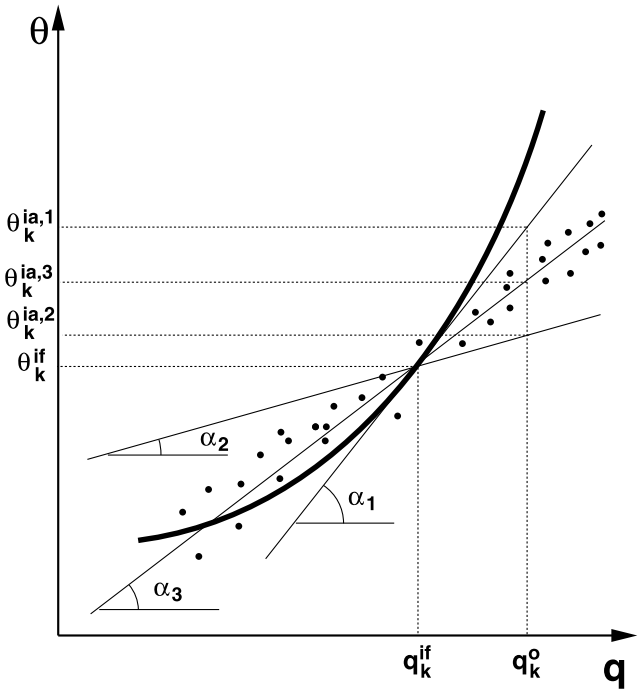

Figure 4. Schematic relationship between the discharge q and the soil moisture content θ for an ensemble member i (thick solid line). The dots represent the results of all individual ensemble members. Here α_1 is the slope of the tangent line using the values from ensemble member i; α_2 is the slope of the tangent line through the point, calculated using the ensemble average values; and α_3 is the slope of the tangent line through the point, calculated using the regression through the ensemble members. Here $\theta_k^{ia,1}$ is the result of the update using α_1, $\theta_k^{ia,2}$ is the result of the update using α_2, and $\theta_k^{ia,3}$ is the result of the update using α_3.

that would not match the observed discharge. A similar reasoning can be made for a system with nonzero observation error, multiple state variables and/or multiple observations. The question thus arises which of the slopes is better to use. For this reason, the results obtained using the Kalman gain calculated using a linearization of the observation system (equation (2)) will be compared to the results obtained using the covariance approximation (equation (14)).

[22] For the reason explained above, we will also assess whether it is better to calculate a separate Kalman gain for each individual ensemble member. In this approach, equation (4) is applied to each individual ensemble member, instead of to the ensemble averaged results. Further, using this Jacobian matrix $\mathbf{H}_k^i$ instead of $\mathbf{H}_k$, a separate Kalman gain is calculated using equation (2). The state update equation then becomes

$$\mathbf{x}_k^{ia} = \mathbf{x}_k^{if} + \mathbf{K}_k^i \left[\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k^{if}) + \mathbf{v}_k^i \right] \tag{22}$$

Figure 4 shows that, if the relationship between the discharge and the soil moisture were linear, the assimilation of a perfect observation for each ensemble member would lead to a state variable that exactly matches the observed discharge. It should be noted that this method will lead to exactly the same results as the ensemble Kalman filter for the TSM with a linear observation system, since in this case p_1 is identical for all ensemble members.

5.3. Implications for the HBV Model

[23] As stated in section 4.2, the state vector for the HBV model at each time step k contains the three reservoir storages from time step k through time step $k - 13$. In the ensemble-based covariance approximation, the covariance between the modeled storages and discharge can be calculated as

$$\sigma_{\mathbf{x}_k^f \mathbf{q}_k^f} = \frac{1}{N-1} \begin{pmatrix} \sum_{i=1}^{N} \left(S_k^{if} - \overline{S_k^f} \right) \left(q_k^{if} - \overline{q_k^f} \right) \\ \sum_{i=1}^{N} \left(S_{1,k}^{if} - \overline{S_{1,k}^f} \right) \left(q_k^{if} - \overline{q_k^f} \right) \\ \sum_{i=1}^{N} \left(S_{2,k}^{if} - \overline{S_{2,k}^f} \right) \left(q_k^{if} - \overline{q_k^f} \right) \\ .. \\ \sum_{i=1}^{N} \left(S_{k-m+1}^{if} - \overline{S_{k-m+1}^f} \right) \left(q_k^{if} - \overline{q_k^f} \right) \\ \sum_{i=1}^{N} \left(S_{1,k-m+1}^{if} - \overline{S_{1,k-m+1}^f} \right) \left(q_k^{if} - \overline{q_k^f} \right) \\ \sum_{i=1}^{N} \left(S_{2,k-m+1}^{if} - \overline{S_{2,k-m+1}^f} \right) \left(q_k^{if} - \overline{q_k^f} \right) \end{pmatrix} \tag{23}$$

S_j^{if} is a short notation for $S^{if}(j\triangle t)$. N is the ensemble size, and the overbar indicates the average over the ensemble. q_k^{if} is the forecasted discharge for ensemble member i. $\sigma^2_{\mathbf{q}_k^f}$ is simply the variance of the modeled discharge across the ensemble at the observation time step ($\sigma^2_{q_k^f}$). The state update equation thus becomes

$$\begin{cases} S_k^{ia} &= S_k^{if} + \dfrac{\sigma_{S_k^f q_k^f}}{\sigma^2_{q_k^f} + R_k} \left(q_k^o - q_k^{if} + v_k^i \right) \\ S_{1,k}^{ia} &= S_{1,k}^{if} + \dfrac{\sigma_{S_{1,k}^f q_k^f}}{\sigma^2_{q_k^f} + R_k} \left(q_k^o - q_k^{if} + v_k^i \right) \\ S_{2,k}^{ia} &= S_{2,k}^{if} + \dfrac{\sigma_{S_{2,k}^f q_k^f}}{\sigma^2_{q_k^f} + R_k} \left(q_k^o - q_k^{if} + v_k^i \right) \\ .. & \\ S_{k-m+1}^{ia} &= S_{k-m+1}^{if} + \dfrac{\sigma_{S_{k-m+1}^f q_k^f}}{\sigma^2_{q_k^f} + R_k} \left(q_k^o - q_k^{if} + v_k^i \right) \\ S_{1,k-m+1}^{ia} &= S_{1,k-m+1}^{if} + \dfrac{\sigma_{S_{1,k-m+1}^f q_k^f}}{\sigma^2_{q_k^f} + R_k} \left(q_k^o - q_k^{if} + v_k^i \right) \\ S_{2,k-m+1}^{ia} &= S_{2,k-m+1}^{if} + \dfrac{\sigma_{S_{2,k-m+1}^f q_k^f}}{\sigma^2_{q_k^f} + R_k} \left(q_k^o - q_k^{if} + v_k^i \right) \end{cases} \tag{24}$$

i is the ensemble member number, q_k^o the observed discharge at time step k ($m^3\ s^{-1}$), and v_k^i a Gaussian number with mean 0 and variance R_k. Equation (24) shows that, in the update of each state variable, on top of the observation error and the variance in the modeled discharge, only the covariance of that specific variable and the modeled

Table 1. Parameter Values for the TSM With Linear and Nonlinear Observation Systems[a]

Parameter	Units	Value: Linear Observation System	Value: Nonlinear Observation System
f_1	-	0.923	0.947
m_1	h^{-1}	0.2438	0.5703
m_2	h^{-1}	0.8923	0.8893
m_3	h^{-1}	0.7628	0.8891
m_4	h^{-1}	0.7563	0.4964
m_5	h^{-1}	0.4222	0.0478
m_6	h^{-1}	0.0174	0.0040
m_7	h^{-1}	0.0098	0.0761
m_8	h^{-1}	0.0020	0.0111
m_9	h^{-1}	0.0136	0.0059
m_{10}	h^{-1}	0.0063	0.0020
m_{11}	h^{-1}	0.0194	0.0013
m_{12}	h^{-1}	0.0014	0.0002
m_{13}	h^{-1}	0.0166	0.0003
m_{14}	h^{-1}	0.0007	0.0196
σ_a	h^{-1}	0.06321	0.06120
p_0	$m^3\ s^{-1}$	0.324	0.433
p_1	$m^3\ s^{-1}\ mm^{-1}$	0.083	
p_2	$m^3\ s^{-1}\ mm^{-1.5}$		0.0118

[a]Hyphen indicates a dimensionless parameter.

discharge is used. This is different from the update using the linearization of the observation system, where the gradients of the simulated discharge with respect to the other state variables are used in the calculation of the Kalman gain (equation (2)). This means that, since the HBV model is restarted from time step $k - m + 2$, using the covariance-based approach, only the covariance between the storages at time step $k - m + 1$ and the discharge at time step k need to be calculated, and only the analyzed storages at time step $k - m + 1$ need to be calculated, since these are the initial conditions for the model application at time step $k - m + 2$.

6. Linearized Observation Operators $\mathbf{H}_k$ to Calculate the Kalman Gain

6.1. TSM

[24] For each time step, there is only one state variable, the catchment storage ($S(t)$, mm). Since the relationship between the observations and the state variable is known analytically (equations (A2) and (A3)), the Jacobian matrix (equation (4)) can be calculated analytically. For the TSM with a linear observation system, $\mathbf{H}_k$ becomes

$$\mathbf{H}_k = [p_1] \tag{25}$$

p_1 is the slope of the discharge-storage relationship (equation (A2) in section A1). For the TSM with a nonlinear observation system the Jacobian can be calculated as

$$\mathbf{H}_k = \left[1.5 p_2 \sqrt{S^f(k\triangle t)}\right] \tag{26}$$

p_2 is the parameter multiplied by the storage raised to the power 1.5 in the discharge-storage relationship (equation (A3) in section A1).

6.2. HBV Model

[25] In section 5.3 the distribution of the state vector is explained. Under these circumstances, the Jacobian matrix $\mathbf{H}_k$ can then be calculated as

$$\mathbf{H}_k = \left[\left.\frac{\partial q^f(k\triangle t)}{\partial S(k\triangle t)}\right|_{S^f(k\triangle t)} \left.\frac{\partial q^f(k\triangle t)}{\partial S_1(k\triangle t)}\right|_{S_1^f(k\triangle t)} \left.\frac{\partial q^f(k\triangle t)}{\partial S_2(k\triangle t)}\right|_{S_2^f(k\triangle t)} \cdots \left.\frac{\partial q^f(k\triangle t)}{\partial S((k-13)\triangle t)}\right|_{S^f((k-14)\triangle t)} \left.\frac{\partial q^f(k\triangle t)}{\partial S_1((k-13)\triangle t)}\right|_{S_1^f((k-14)\triangle t)} \cdot \left.\frac{\partial q^f(k\triangle t)}{\partial S_2((k-13)\triangle t)}\right|_{S_2^f((k-14)\triangle t)}\right] \tag{27}$$

An analytical expression for the relationship between the discharge and the storages does not exist. For this reason, the first derivatives are calculated numerically, through a first-order Taylor series expansion.

7. Results

7.1. Calibration of the Rainfall-Runoff Models

[26] All models were applied using an hourly time step from 1 August 2006 through 31 July 2007. The parameters for the models have been estimated using the data from the study period through the Multistart Weight-Adaptive Recursive Parameter Estimation (MWARPE) method, as described by *Pauwels* [2008]. This methodology uses the linear sequential filter equations in an iterative, Monte Carlo framework. Table 1 lists the obtained parameter values for the TSM, with a linear and a nonlinear observation system, and Table 2 for the HBV model. Using these parameters, an ensemble of model applications was generated. For this purpose, the meteorological forcings (precipitation and potential evapotranspiration), the model parameters, and the initial conditions were disturbed by adding a Gaussian random number to their deterministic values. This methodology to generate the ensemble is common practice in the application of the ensemble Kalman filter. Examples can be found, among many others, in the work by *Evensen* [1994], *Reichle et al.* [2002a], *De Lannoy et al.* [2006, 2007], and *Durand and Margulis* [2007]. The standard deviation of this random number was set to a fraction of the parameter value. In order to verify the appropriateness of the ensemble for data assimilation purposes, two different ensemble verification measures were used [*De Lannoy et al.*, 2006]. For this ensemble verification, the ensemble spread ($ensp_k$), the ensemble mean square error (mse_k), and the ensemble skill ($ensk_k$) need to be calculated at each time step k:

$$\begin{cases} ensp_k &= \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\left(q_k^{if} - q_k^f\right)^2 \\ mse_k &= \displaystyle\sum_{i=1}^{N}\left(q_k^{if} - q_k^o\right)^2 \\ ensk_k &= \left(q_k^f - q_k^o\right)^2 \end{cases} \tag{28}$$

In order for the ensemble to have a large enough spread, it is expected that on average the ensemble mean differs from the observation by a value that is equal to the time average

Table 2. Parameter Values and Initial Conditions for the HBV Model Used in the Model Simulations[a]

Parameter	Units	Value
λ	-	1.778
$S_{\max}$	m^3	2.168×10^7
b	-	0.174
α	-	0.414
P	$m^3\ s^{-1}$	13.354
β	-	0.055
γ	-	0.713
$S_{2,\max}$	m^3	4.04×10^6
κ_2	$m^3\ s^{-1}$	411.3
κ_1	s^{-1}	8.065×10^{-6}
$S(t=0)$	m^3	9.143×10^6
$S_1(t=0)$	m^3	9.179×10^4
$S_2(t=0)$	m^3	1×10^{-10}

[a]Hyphens indicate dimensionless parameters.

of the ensemble spread [*De Lannoy et al.*, 2006]. Thus, the following should be true:

$$\frac{\langle ensk \rangle}{\langle ensp \rangle} \approx 1 \tag{29}$$

$\langle . \rangle$ indicates an average over the simulation period. Further, if the truth is statistically indistinguishable from a member of the ensemble, the following should be true [*De Lannoy et al.*, 2006]:

$$\frac{\langle \sqrt{ensk} \rangle}{\langle \sqrt{mse} \rangle} \approx \sqrt{\frac{N+1}{2N}} \tag{30}$$

These verification measures can be interpreted as indications that the ensemble covers the range of the observations. It is therefore important that they are calculated over the entire ensemble, not just the low or high flows. If, for example, they were calculated only over the peak flows, there is a high chance that the ensemble does not cover the low flows at all. This would imply that the model has not been adequately calibrated over the low flows, which would cause bias in the model results, and which could lead to a worse model performance through the data assimilation. The optimal fractions were determined using an interval search. An ensemble size of 32 and 64 members was used, and the ensemble size with the best match for both statistics was used in the data assimilation study.

[27] For the TSM, an ensemble size of 32 members resulted in the best ensemble performance, while for the HBV model the best performance was obtained with 64 members. For the HBV model, a disturbance fraction of 0.11 for the model parameters and 1 for the initial conditions and meteorological forcings was obtained. For the TSM with a linear observation system, the disturbance fraction for the parameters and meteorological forcings was 0.03475. When a nonlinear observation system was used, this fraction was 0.0236. Table 3 shows the ensemble verification statistics obtained for all models. Figures 5 (top) and 6 (top) show the results of the ensemble application for the TSM with linear and nonlinear observation systems, respectively. The Root Mean Square Error (RMSE) between the ensemble mean model results and the observations is calculated as follows:

$$\mathrm{RMSE} = \frac{1}{n_o} \sqrt{\sum_{m=1}^{n_o} (q_m^o - q_m^f)} \tag{31}$$

n_o is the amount of data points, q_m^o the observed discharge, and q_m^f the modeled (forecast) discharge, both in $m^3\ s^{-1}$.

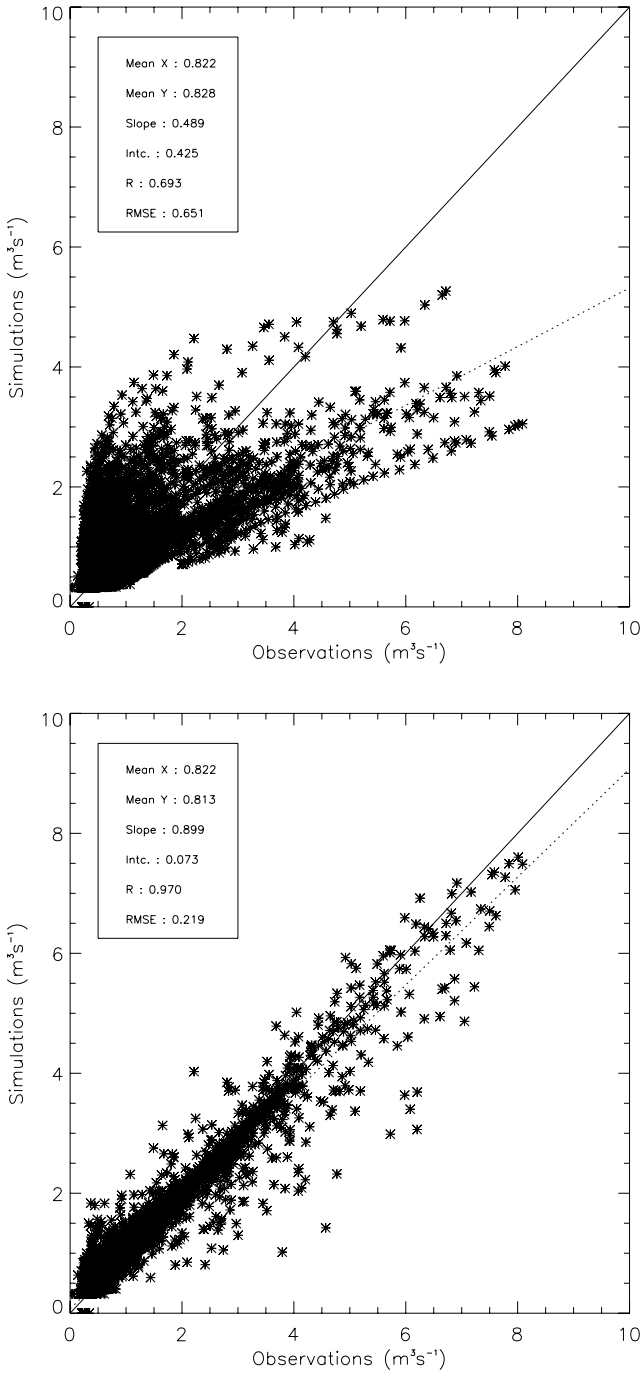

Figure 5. Results of the model applications for the TSMl. (top) Baseline run. (bottom) Assimilation of observations with the ensemble Kalman filter with nonlinear observation system and an observation error of 0.1 $m^3\ s^{-1}$. The solid line is the 1:1 line, and the dotted line is the regression.

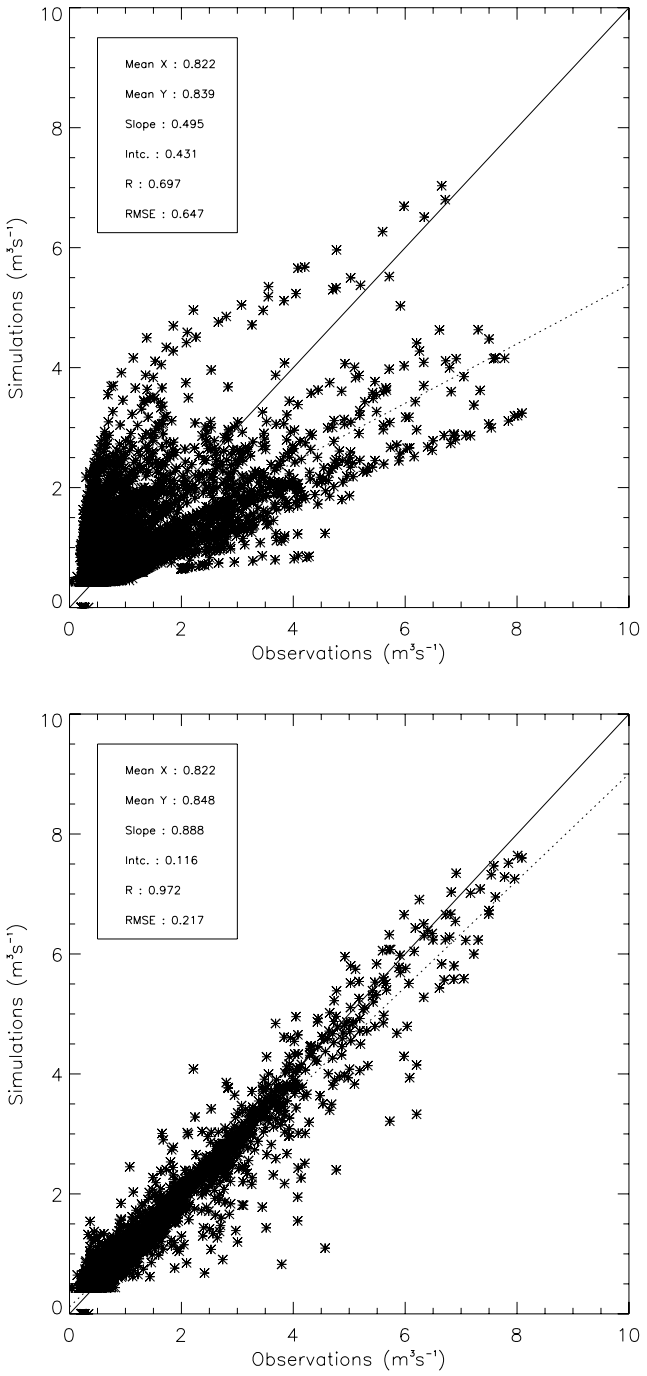

Figure 6. Results of the model applications for the TSMnl. (top) Baseline run. (bottom) Assimilation of observations with the ensemble Kalman filter with nonlinear observation system and an observation error of 0.1 m^3 s^{-1}. The solid line is the 1:1 line, and the dotted line is the regression.

The results of the TSM with a nonlinear observation system are almost identical as when a linear observation system is used. Figure 7 shows the results of the ensemble application for the HBV model. Although some individual peaks are overestimated or underestimated, overall the model simulates the peaks rather well, as is expressed through the regression line which is very close to the 1:1 line. Errors in the model meteorological forcings and the relatively simple model structure are the cause of these mismatches. The motivation to apply data assimilation is to improve the model results for the peaks. Overall, from these results it can be concluded that the HBV model, applied in an ensemble framework, can adequately simulate the rainfall-runoff behavior of the catchment.

[28] In order to demonstrate the impact of the ensemble spread, a second ensemble was generated, in which the

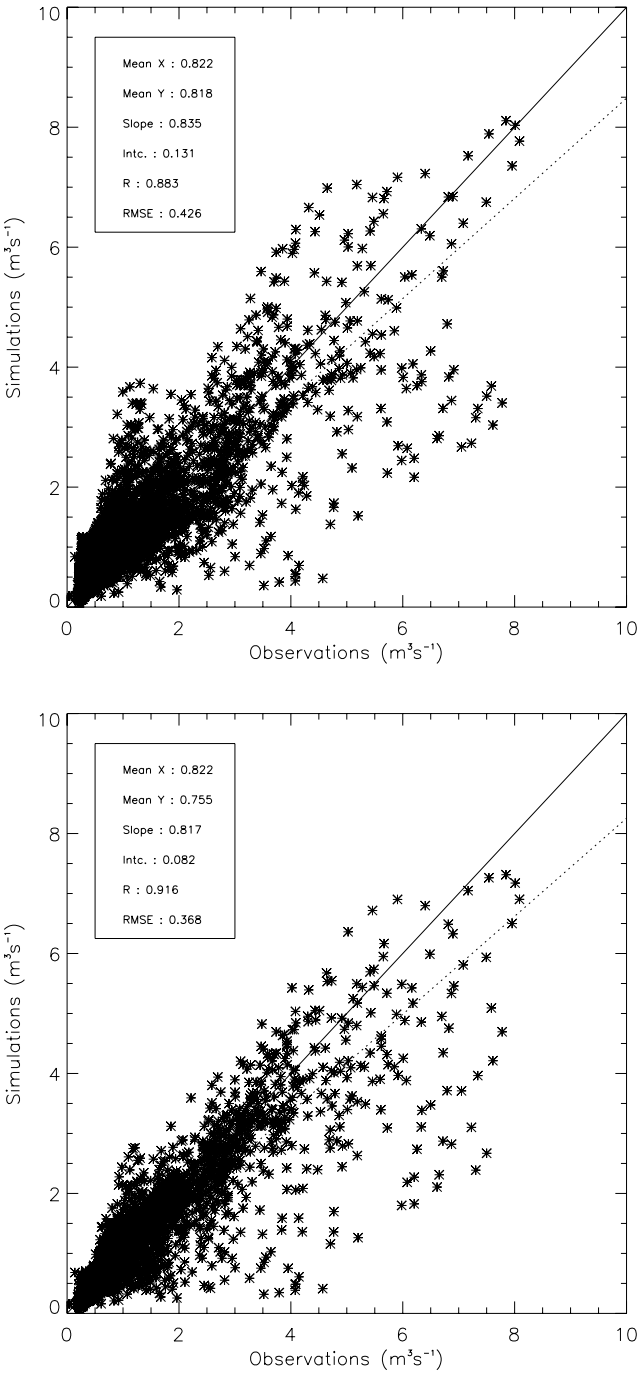

Figure 7. Results of the model applications for the HBV model. (top) Baseline run. (bottom) Assimilation of observations with the ensemble Kalman filter with nonlinear observation system and an observation error of 0.1 m^3 s^{-1}. The solid line is the 1:1 line, and the dotted line is the regression.

Table 3. Ensemble Verification Measures for the Ensembles Used[a]

	HBV		TSMl		TSMnl	
Measure	Ideal Value	Ensemble Value	Ideal Value	Ensemble Value	Ideal Value	Ensemble Value
			Optimal Ensemble			
$\frac{\langle ensk \rangle}{\langle ensp \rangle}$	1	1.15	1	1.00	1	1.05
$\frac{\langle \sqrt{ensk} \rangle}{\langle \sqrt{mse} \rangle}$	0.712	0.639	0.718	0.585	0.718	0.638
			Ensemble With Insufficient Spread			
$\frac{\langle ensk \rangle}{\langle ensp \rangle}$	1	3.32	1	2.73	1	2.41
$\frac{\langle \sqrt{ensk} \rangle}{\langle \sqrt{mse} \rangle}$	0.712	0.795	0.718	0.768	0.718	0.778
			Ensemble With Excessive Spread			
$\frac{\langle ensk \rangle}{\langle ensp \rangle}$	1	0.190	1	0.037	1	0.267
$\frac{\langle \sqrt{ensk} \rangle}{\langle \sqrt{mse} \rangle}$	0.712	0.372	0.718	0.121	0.718	0.382

[a]TSMl stands for the TSM with a linear observation system, and TSMnl stands for the TSM with a nonlinear observation system.

original disturbance factors were decreased by a factor of 1.1 for the TSM, and two for the HBV model. In other words, for both models a second ensemble was generated, with an insufficient ensemble spread. A third ensemble was then also generated, in which the original disturbance factors were multiplied by 1.1 and 2, for the TSM and the other models, respectively. In this case the ensemble spread was thus too large. These different disturbance fractions were chosen so the ensemble statistics for all models were approximately equal. Table 3 shows the statistics of these two extra ensembles for both models. It will be assessed to what extent the quality of the ensemble impacts the performance of the different ensemble Kalman filter approaches for the purpose of discharge assimilation.

7.2. Assimilation at Every Time Step for the TSM

[29] Table 4 shows the results of the discharge data assimilation into the TSM. It should be stressed that these results, as well as all results described in section 7, unless explicitly stated otherwise refer to the one-step-ahead ensemble mean forecasted discharge values, not the analysis values. An error in the observed discharge of 0, 0.1, and 0.5 $m^3\ s^{-1}$ is assumed. Even though zero errors are rarely encountered in reality, this case serves to focus on how the innovation is mapped on the analysis increment. These different error levels are used to assess the impact of the observational error level on the efficiency of the data assimilation algorithms. As can be expected, the results of the ensemble Kalman filter with linearized and nonlinearized observation systems are identical if the observation system is linear. Table 4 shows that these conclusions are independent of the quality of the ensemble used for the data assimilation algorithms.

[30] In case a nonlinear observation system is used, and the observations are perfect, the use of the ensemble Kalman filter with a linearized observation system leads to an increase in the forecast error. Figure 8 shows the results of the assimilation of perfect observations into the TSM with nonlinear observation system. It should be noted that the constant base flow values around time step 1000 and approximately between time step 6000 and 6750 are caused by the larger value of p_0, as can be seen in Table 1. For these time steps, the storage is approximately zero, which will lead to a discharge close to p_0. Figure 8 shows that this increase in the forecast error can be explained by the strong overesti-

Table 4. Root Mean Square Error Between the Observed and Simulated Discharge for the Baseline Run and the Different Assimilation Runs for the Different Models[a]

		Perfect Observations			Observation Error 0.1 $m^3\ s^{-1}$			Observation Error 0.5 $m^3\ s^{-1}$		
Model	Baseline Run	EnKFl	EnKFnl	MH	EnKFl	EnKFnl	MH	EnKFl	EnKFnl	MH
					Optimal Ensemble					
TSMl	0.651	0.129	0.129	-	0.220	0.220	-	0.354	0.354	-
TSMnl	0.647	0.923	0.537	1.551	0.217	0.217	0.218	0.333	0.332	0.331
HBV	0.426	0.443	0.377	0.437	0.432	0.369	0.434	0.434	0.385	0.425
					Ensemble With Insufficient Spread					
TSMl	0.654	0.129	0.129	-	0.232	0.232	-	0.370	0.370	-
TSMnl	0.651	0.924	0.537	4.144	0.223	0.223	0.223	0.345	0.344	0.342
HBV	0.427	0.447	0.368	0.428	0.445	0.355	0.429	0.453	0.390	0.435
					Ensemble With Excessive Spread					
TSMl	0.625	0.137	0.137	-	0.217	0.217	-	0.409	0.409	-
TSMnl	0.663	0.923	0.537	1.546	0.212	0.212	0.213	0.321	0.320	0.321
HBV	0.466	0.451	0.482	0.443	0.453	0.434	0.447	0.453	0.447	0.437

[a]TSMl stands for the TSM with a linear observing system, and TSMnl stands for the TSM with a nonlinear observing system. EnKFl stands for the EnKF with a linearized observation system, EnKFnl stands for the EnKF with a nonlinearized observation system, and MH stands for multiple Jacobians per ensemble. Units are in $m^3\ s^{-1}$.

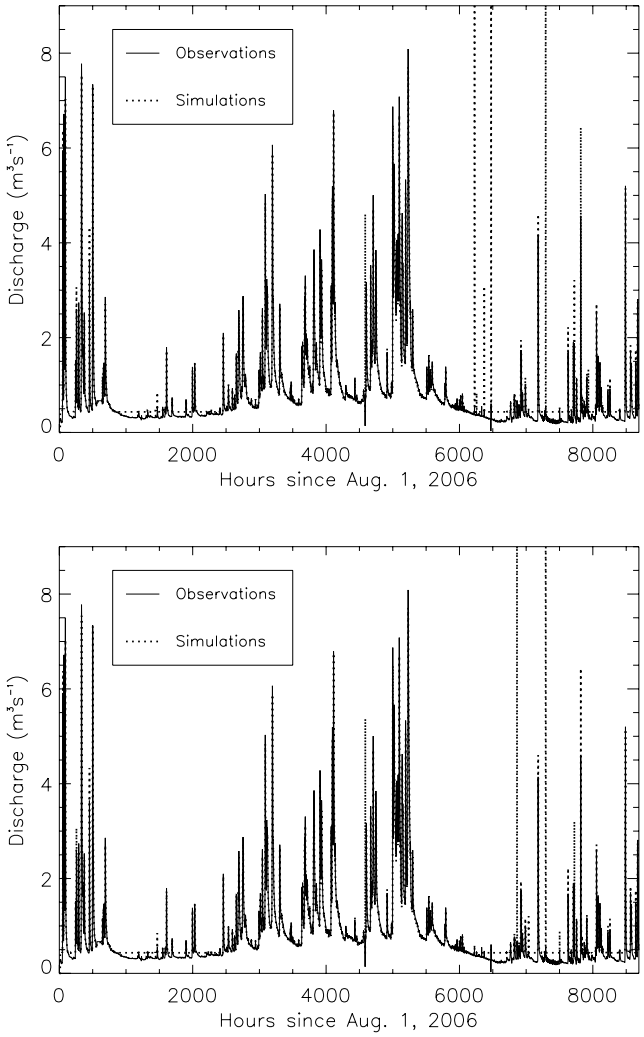

Figure 8. Results of the assimilation of perfect observations into the TSM with nonlinear observation system. (top) Results obtained using the ensemble Kalman filter with linearized observation system. (bottom) Results obtained using multiple Kalman gains.

mation of a number of discharge peaks. This is caused by the effect of the observation system linearization. Table 5 further explains this problem, for the model results around time step 6225 (the first strongly overestimated peak). At time step 6225 the slope H_k is almost zero, while at the same time step the model underestimates the discharge. Since R_k is equal to zero, the Kalman gain becomes very large. The reason for this is that under these conditions H_k^{-1}, which maps the observations to the state, becomes very large. This will lead to a very high value of the analysis storage, and consequently the forecasted discharge. Five time steps are needed in order to correct this overestimation. The use of multiple Kalman gains does not reduce this problem. These results clearly demonstrate the impact of the nonlinearity of the observation system. The baseline results obtained using a linear and a nonlinear observation system are very similar, but the application of the ensemble Kalman filter with linearized observation system leads to strongly different results for both cases. When the observation error is nonzero, both approaches lead to similar results and an improvement of the forecasted discharge.

[31] The results in Table 4 also show that the ensemble Kalman filter with nonlinearized observation system consistently results in the lowest forecast error, regardless of the quality of the ensemble.

[32] Figures 5 (bottom) and 6 (bottom) show that, if discharge rates with a realistic level of uncertainty (0.1 $m^3\ s^{-1}$) are assimilated every time step, and the EnKF with nonlinearized observation system is used, a strong improvement in the model results is obtained, and both the high flows as well as the low flows are simulated accurately.

7.3. Peak Forecasting for the TSM

[33] In order to assess the impact of the data assimilation algorithms on subsequent predictions in a more realistic setting, the data assimilation was halted immediately before the largest peaks in the data set. From this time step on, the model was applied in ensemble forecast mode, without data assimilation. The peaks with maximum values above 3 $m^3\ s^{-1}$ were selected for this purpose. Table 6 lists the start and end times of these peaks, together with the results obtained for all models and data assimilation algorithms.

[34] Table 6 shows that for both the linear and nonlinear observation system similar conclusions can be drawn. For all peaks, and for all assimilation methods, if data are assimilated every time step, the mismatch between the observations and the model simulations is strongly reduced. However, when the data assimilation is halted immediately before the peak, at best a slight reduction in the RMSE and in many cases an increase is obtained. This leads to the conclusion that, for this catchment and these time series models and parameter sets, the accuracy of the precipitation records is more important than the accuracy of the initial storages, if peak discharges need to be predicted using longer lead times.

Table 5. Values of the Different Parts of the State Update Equation for the TSM With Nonlinear Observing System, Optimal Ensemble Spread, Zero Observation Error, and the State Variables for Ensemble Member 1[a]

Time Step k	$q_o(k\triangle t)$ ($m^3\ s^{-1}$)	$q^f(k\triangle t)$ ($m^3\ s^{-1}$)	H_k ($m^3\ s^{-1}\ mm^{-1}$)	K_k (mm s m^{-3})	I_1 ($m^3\ s^{-1}$)	x_k^{1-} (mm)	x_k^{1+} (mm)
6223	0.382	0.433	0.00072	1397.54	−0.0511	0.05230	0
6224	0.428	0.433	0.00005	20028.38	−0.0050	0.00025	0
6225	0.567	0.433	0.00038	2653.55	0.1340	0.01451	355.54
6226	0.395	73.719	0.32534	3.07	−73.087	337.149	112.50
6227	0.374	13.470	0.18297	5.47	−13.038	106.553	35.30
6228	0.374	2.724	0.10248	9.76	−2.3351	33.3839	10.60
6229	0.382	0.810	0.05616	17.80	−0.4283	10.6553	2.45
6230	0.385	0.474	0.02699	37.04	−0.0903	2.34201	0

[a] I_1 stands for the innovation for ensemble member 1.

Table 6. Results of the Model Applications With Data Assimilation Terminated at the Onset of the Peak[a]

				Assimilation Results					
				EnKFl		EnKFnl		MH	
Onset Peak	End Peak	Model	Baseline Run Result	E	S	E	S	E	S
325	350	TSMl	1.812	0.821	1.921	0.821	1.921	-	-
		TSMnl	1.676	0.709	1.878	0.709	1.879	0.707	1.876
		HBV	2.065	2.344	2.347	1.437	1.336	2.328	2.249
493	541	TSMl	1.308	0.635	1.436	0.635	1.436	-	-
		TSMnl	1.279	0.564	1.520	0.565	1.520	0.559	1.520
		HBV	1.330	1.471	1.476	1.251	1.192	1.467	1.424
3116	3150	TSMl	1.356	0.377	1.366	0.377	1.366	-	-
		TSMnl	1.391	0.312	1.255	0.312	1.255	0.312	1.253
		HBV	1.100	0.950	0.952	0.840	1.197	1.236	1.121
4036	4065	TSMl	1.561	0.450	1.225	0.450	1.225	-	-
		TSMnl	1.282	0.362	0.723	0.363	0.723	0.363	0.723
		HBV	1.200	1.433	1.475	1.197	1.271	1.156	1.162
4925	4940	TSMl	2.056	0.991	2.011	0.991	2.011	-	-
		TSMnl	2.209	0.839	1.955	0.840	1.954	0.839	1.955
		HBV	1.800	1.888	1.890	1.853	1.818	1.850	1.770
5020	5042	TSMl	2.642	0.525	2.147	0.525	2.147	-	-
		TSMnl	2.555	0.448	1.833	0.449	1.834	0.450	1.833
		HBV	0.777	0.799	0.897	0.894	1.018	1.073	0.776
5145	5193	TSMl	2.895	0.461	2.919	0.461	2.919	-	-
		TSMnl	2.785	0.390	2.728	0.390	2.728	0.391	2.728
		HBV	0.796	0.829	0.776	0.720	1.352	1.296	0.919

[a]Units are in $m^3\ s^{-1}$. The observation error is 0.1 $m^3\ s^{-1}$. The onsets and ends of the peaks are in time step numbers. E indicates assimilation at every time step, and S indicates assimilation until the onset of the peak.

7.4. Assimilation at Every Time Step for the HBV Model

[35] Table 4 shows the results of the application of the different data assimilation algorithms for the HBV model. Figure 9 shows the results of the assimilation of perfect observations into the HBV model. The impact of the data assimilation is clearly not as strong for the HBV model as for the time series models. This lower impact of the data assimilation for the HBV model can be explained by the impact of the routing algorithm. As explained above, if an observation is assimilated at time step k, the state variables from time step $k - 13$ through k are updated, and the model is restarted at time step $(k - 12)$. One can expect that the correlation between the ensemble mean forecasted state variables at time step $k - 13$ and the ensemble mean forecasted discharge at time step k is not as large as the correlation between the state variables at time step $(k - 1)$ and this discharge. More specifically, for the HBV model baseline run (no data assimilation), the correlation coefficient between $S((k - 13)\triangle t)$ and $q^f(k\triangle t)$ is 0.551. For $S_1((k - 13)\triangle t)$ and $q^f(k\triangle t)$ the correlation coefficient is 0.753, and for $S_2((k - 13)\triangle t)$ and $q^f(k\triangle t)$ it is 0.340. The highest-correlation coefficient is thus 0.753, corresponding to a determination coefficient of 0.56. This implies that changing the state variables 13 time steps before the discharge observation will have a limited impact on the modeled runoff.

[36] In other words, the HBV model is reinitialized 14 time steps prior to the observation. In this case model uncertainty, caused by uncertainty in the parameters, forcings, and model structure, dominates over the uncertainty in the initial conditions.

[37] Table 4 shows that the only assimilation method consistently leading to an improvement in the model forecasts is the ensemble Kalman filter with nonlinearized observation system. The only exception is when the observations are perfect and the ensemble spread is too large. For the other methods, in almost all cases a worsening of the model forecasts has been obtained.

[38] Figure 7 (bottom) shows that, if discharge rates with a realistic level of uncertainty (0.1 $m^3\ s^{-1}$) are assimilated every time step, and the EnKF with nonlinearized observation system is used, this reduction in the RMSE can be attributed to a slightly better modeling of the low flows as well as the high flows. However, as stated above, the impact of the data assimilation is limited.

7.5. Peak Forecasting for the HBV Model

[39] Table 6 shows that, if the discharge is assimilated at every time step, the EnKFnl outperforms the EnKFl for all peaks, except for the peak starting at time step 5020. However, for this peak, the difference between the results obtained from both methods is relatively small. In general, the HBV baseline run simulates the peaks better than either of the time series models, but when discharge data are assimilated at every time step using both approaches to the EnKF, the opposite conclusion can be drawn. In many cases, the assimilation of discharge rates at every time step using the EnKF increases the mismatch between the observations and the model simulations. However, the impact of the data assimilation is limited, as compared to the time series models. This can be explained by the data assimilation algorithm. In section 5.3, it is explained that for the EnKFnl, the covariance between the state variables at time step $k - 13$ and the discharge at time step k is used, in order to update the state variables at time step $k - 13$. Table 7 lists these covariances, together with the state variables, for the time steps up till the onset of the peak starting at time step 325. As can be expected, and is explained in section 7.4, the

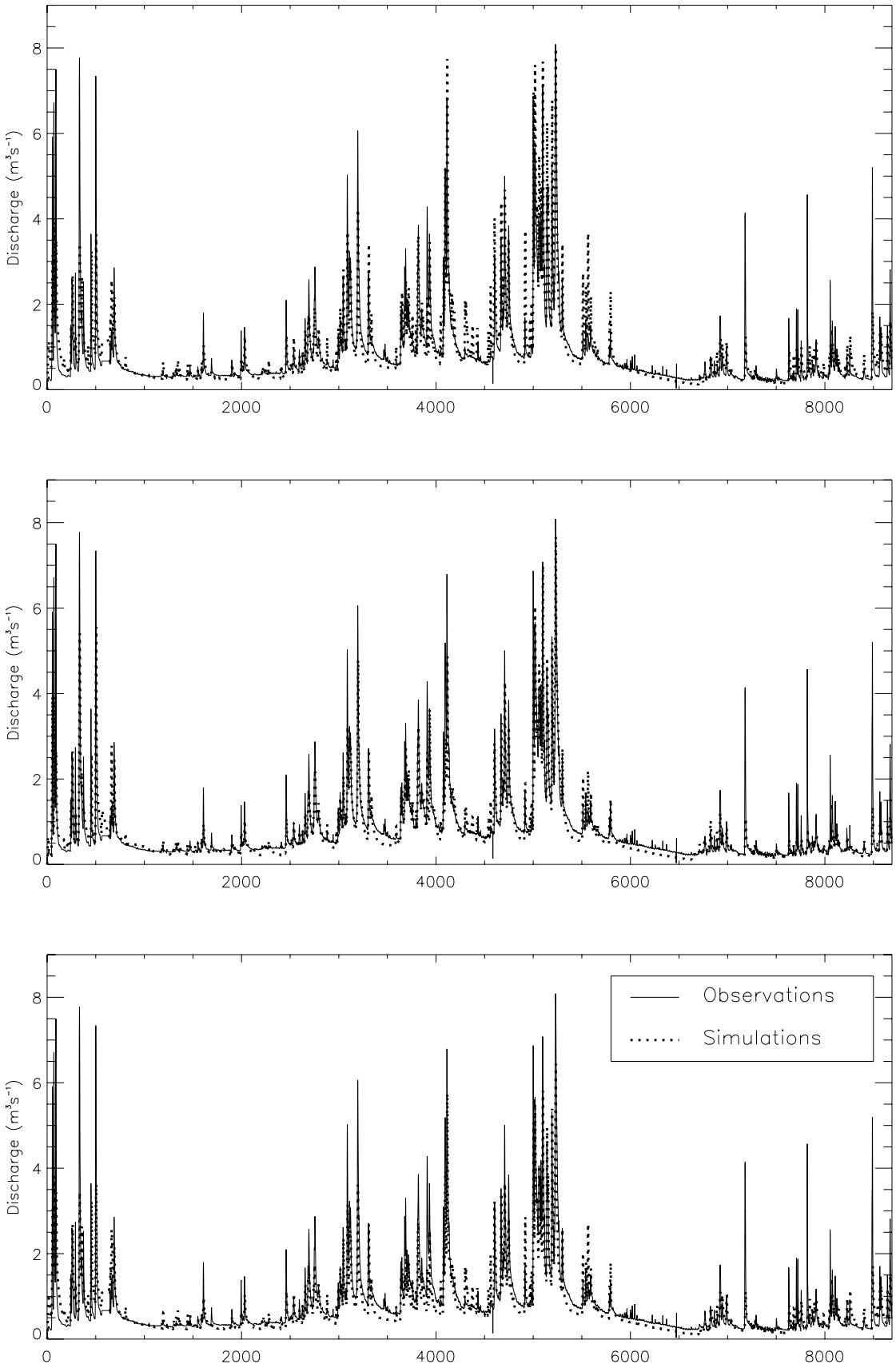

Figure 9. Results of the assimilation of perfect observations into the HBV model. (top) Results obtained using the ensemble Kalman filter with linearized observation system. (middle) Results obtained using the ensemble Kalman with nonlinearized observation system. (bottom) Results obtained using multiple Kalman gains.

Table 7. Analysis of the HBV Model Results for the Peak Starting at Time Step k = 325 Using the Ensemble Kalman Filter With Nonlinearized Observation System[a]

k	P (mm h^{-1})	Q_o (m^3 s^{-1})	$\sigma_{q_{\bar{k}}\ S_{\bar{k}-13}}$ (m^3 s^{-1} m)	$\sigma_{q_{\bar{k}}\ S_{\bar{1},k-13}}$ (m^3 s^{-1} m)	$\sigma_{q_{\bar{k}}\ S_{\bar{2},k-13}}$ (m^3 s^{-1} m)	S^a_k (m)	$S^a_{1,k}$ (m)	$S^a_{2,k}$ (m)
310	0	0.511	-8.1×10^{-5}	2.7×10^{-6}	0	1.9×10^{-1}	7.8×10^{-4}	0
311	0	0.494	-9.1×10^{-5}	1.9×10^{-6}	0	1.9×10^{-1}	7.8×10^{-4}	0
312	0	0.482	-1.0×10^{-4}	1.3×10^{-6}	0	1.9×10^{-1}	7.8×10^{-4}	0
313	0	0.475	-1.2×10^{-4}	6.4×10^{-7}	0	1.9×10^{-1}	7.8×10^{-4}	0
314	0	0.464	-1.3×10^{-4}	5.3×10^{-8}	0	1.9×10^{-1}	7.8×10^{-4}	0
315	0	0.458	-1.3×10^{-4}	1.1×10^{-6}	0	1.9×10^{-1}	7.8×10^{-4}	0
316	0	0.455	-1.2×10^{-4}	2.1×10^{-6}	0	1.9×10^{-1}	7.8×10^{-4}	0
317	0	0.449	-1.1×10^{-4}	3.0×10^{-6}	0	1.9×10^{-1}	7.8×10^{-4}	0
318	0	0.444	-1.0×10^{-4}	3.9×10^{-6}	0	1.9×10^{-1}	7.8×10^{-4}	0
319	0	0.437	-9.0×10^{-5}	4.9×10^{-6}	5.8×10^{-8}	1.9×10^{-1}	7.8×10^{-4}	0
320	0	0.432	-3.3×10^{-5}	4.2×10^{-6}	0	1.9×10^{-1}	7.8×10^{-4}	0
321	0	0.429	-2.6×10^{-5}	5.0×10^{-6}	0	1.9×10^{-1}	7.7×10^{-4}	0
322	0.93	0.425	-2.8×10^{-5}	5.2×10^{-6}	0	1.9×10^{-1}	7.8×10^{-4}	0
323	0	0.455	-2.9×10^{-5}	5.3×10^{-6}	0	1.9×10^{-1}	9.3×10^{-4}	6.9×10^{-5}
324	1.09	0.461	-1.1×10^{-5}	3.1×10^{-6}	0	1.9×10^{-1}	9.3×10^{-4}	2.6×10^{-7}
325	1.03	0.471	1.1×10^{-5}	1.2×10^{-6}	0	1.9×10^{-1}	1.1×10^{-3}	6.8×10^{-5}

[a] P is the precipitation, and q_o is the observed discharge. The observation error is 0.1 m^3 s^{-1}.

magnitude of these covariances is relatively low, which will lead to a negligible update of the state variables.

[40] When multiple Jacobians are used, similar conclusions can be drawn as for the EnKFl. The conclusion of this analysis of the modeled peaks is that, for the EnKF with linearized and nonlinearized observation system, mixed results are obtained, if the data assimilation is stopped at the onset of the peak. This can be explained by the limited updating of the initial conditions 13 time steps before the peak, as is explained above.

8. Computational Efficiency

[41] For all models, when data are assimilated every time step, the EnKF with nonlinearized observation system was computationally the most efficient method. The approaches with a linearized observation system and multiple Kalman gains required more computer time. This can be explained by the higher amount of model applications in order to calculate the Jacobian matrix $\mathbf{H}_k$, and the higher amount of matrix multiplications needed to calculate $\mathbf{K}_k$ in equation (2).

[42] As a summary, the EnKF with nonlinearized observation system is computationally the most efficient method. The approaches in which the observation system is linearized are computationally more demanding, and they suffer from serious drawbacks with respect to model performance.

9. Discussion and Conclusions

[43] The objective of this paper was to examine how discharge data are best assimilated into hydrologic models, if the models are applied in an ensemble framework, and the ensemble Kalman filter is used as assimilation algorithm. The relationship between the observations and the state variables (the observation system) is nonlinear, which has as consequence that a number of approaches can be derived to map the discharge observations to the state variables. The results of the one time step ahead forecasts indicate that the spread and the accuracy of the generated ensemble is in some cases important. The application of two time series models showed the impact of the nonlinearity of the observation system. In all cases, assimilation of discharge into hydrologic models clearly works best if linearization of the observation system is bypassed. If the time delay between the generation of runoff and the arrival at the catchment is modeled using a unit hydrograph, only a marginal improvement in the model results is obtained. This can be explained by the dominating effect of the model error, as compared to the impact of the initial conditions. The model error is caused by errors in the meteorological forcings, model structure, and model parameters, accumulated over the duration of the unit hydrograph. Since the initial conditions need to be updated a number of time steps prior to the observation, the model error can accumulate over a number of time steps, and the improvement in the modeled discharge will be limited. The consequence is that a more realistic model structure will not necessarily lead to an improvement of the results through data assimilation, if discharge is assimilated. This problem can be overcome by using a cascade of linear reservoirs to represent the surface routing processes, as is performed, for example, in the work by *Moradkhani et al.* [2005], *Vrugt et al.* [2006], *Blöschl et al.* [2007], and *Komma et al.* [2008]. Another solution is the approach of *Weerts and El Serafy* [2006], in which the modeled unrouted discharge is compared to the observed discharge a number of time steps later. On the basis of this approach, *Weerts and El Serafy* [2006] concluded that the ensemble Kalman filter outperformed the Particle Filter. Since comparing routed runoff to observations is arguably a more realistic approach, the question arises which of these two filters is more appropriate if this approach is used. Ongoing research is focusing on this question. The final conclusion of this paper is that, if longer lead times are considered, given the model structure, the accuracy of the meteorological forcings is more important than an accurate estimation of the model initial conditions through data assimilation.

Appendix A: Model Description

A1. TSM

[44] The Time Series Model (TSM) can be written as

$$S(t) = f_1 S(t-1) + \sum_{i=1}^{14} m_i R_{tot}(t-i) + a_k \qquad \text{(A1)}$$

$S(t)$ is the total storage in the catchment at time step t (mm), and R_{tot} is the precipitation (mm h^{-1}). t is equal to $k\triangle t$, with $\triangle t$ the model time step (s). f_1 is the lag one autoregressive parameter (), m_i is the lag i moving average parameter (h), and a_k is a Gaussian random number with zero mean and standard deviation σ_a (mm). The lag of 14 time steps in the moving average model represents the effect of routing and reinfiltration on the catchment discharge. The amount of time steps to take into account has been determined through calibration of unit hydrograph–based models for the study site [*Ferket*, 2008]. The initial storage is assumed to be zero. Two different observation systems were used for this model. A first, linear observation system, relates the storage (the system state) to the catchment discharge as follows:

$$q(t) = p_0 + p_1 S(t) \tag{A2}$$

p_0 and p_1 are parameters, with units $m^3\ s^{-1}$ and $m^3\ s^{-1}\ mm^{-1}$, respectively. A second, nonlinear observation system relates the catchment storage to the catchment discharge as follows:

$$q(t) = p_0 + p_2 S^{1.5}(t) \tag{A3}$$

p_2 is a parameter ($m^3\ s^{-1}\ mm^{-1.5}$). The exponent of 1.5 has been chosen to demonstrate the effect of a relatively weak nonlinear observation system on the performance of the assimilation algorithms.

[45] As a summary, for the TSM with a linear storage-discharge relationship, 18 parameters need to be calibrated: f_1, m_1 through m_{14}, σ_a, p_0, and p_1. If a nonlinear observation system is used, p_2 needs to be determined instead of p_1.

A2. HBV Model

[46] The Hydrologiska Byrns Vattenbalansavdelning (HBV) model, of which Figure 2 shows a schematic, was originally developed by *Linström et al.* [1997]. In this paper, the version of *Matgen et al.* [2006] is applied. The model uses observed precipitation ($R_{tot}(t)$) and potential evapotranspiration ($ETP(t)$) as input, both in $m^3\ s^{-1}$. t is the time in seconds. The catchment is divided into a soil reservoir, a fast reservoir, and a slow reservoir. There are thus three state variables: the amount of water in the soil reservoir ($S(t)$, m^3), the slow reservoir ($S_1(t)$, m^3), and the fast reservoir ($S_2(t)$, m^3).

[47] A number of fluxes are calculated, which depend on the state variables of the system. The actual evapotranspiration $ETR(t)$ ($m^3\ s^{-1}$) is first determined:

$$ETR(t) = \frac{1}{\lambda}\frac{S(t)}{S_{\max}} ETP(t) \tag{A4}$$

λ is a dimensionless parameter, and $S_{\max}$ is the storage capacity of the soil reservoir (m^3). The infiltration $R_{in}(t)$ ($m^3\ s^{-1}$) is calculated as follows:

$$R_{in}(t) = \left(1 - \frac{S(t)}{S_{\max}}\right)^b R_{tot}(t) \tag{A5}$$

b is a dimensionless parameter. After this, the effective precipitation $R_{eff}(t)$ ($m^3\ s^{-1}$) is determined:

$$R_{eff}(t) = R_{tot}(t) - R_{in}(t) \tag{A6}$$

The calculation of the percolation $D(t)$ ($m^3\ s^{-1}$) is then performed:

$$D(t) = Pe\left(1 - e^{-\beta\frac{S(t)}{S_{\max}}}\right) \tag{A7}$$

Pe is a percolation parameter ($m^3\ s^{-1}$), and β a dimensionless parameter. After this, the storage in the soil reservoir at the end of the time step can be calculated as follows:

$$S(t+\triangle t) = S(t) + (R_{in}(t) - ETR(t) - Perc(t))\triangle t \tag{A8}$$

$\triangle t$ is the time step in seconds.

[48] The input in the fast reservoir $R_2(t)$ ($m^3\ s^{-1}$) is then

$$R_2(t) = \alpha\frac{S(t)}{S_{\max}} R_{eff} \tag{A9}$$

α is a dimensionless parameter. The outflow from this reservoir $Q_2(t)$ ($m^3\ s^{-1}$) is then determined:

$$Q_2(t) = \kappa_2\left(\frac{S_2(t)}{S_{2,\max}}\right)^\gamma \tag{A10}$$

$S_{2,\max}$ is the storage capacity of the fast reservoir (m^3), and κ_2 ($m^3\ s^{-1}$) and γ () are model parameters. After this, the storage in the fast reservoir at the end of the time step can be calculated as

$$S_2(t+\triangle t) = S_2(t) + (R_2(t) - Q_2(t))\triangle t \tag{A11}$$

The input in the slow reservoir $R_1(t)$ ($m^3\ s^{-1}$) is then computed:

$$R_1(t) = R_{eff}(t) - R_2(t) \tag{A12}$$

The outflow from this reservoir $Q_1(t)$ ($m^3\ s^{-1}$) can then be calculated as

$$Q_1(t) = \kappa_1 S_1(t) \tag{A13}$$

κ_1 is a model parameter (s^{-1}). Finally, the storage in the slow reservoir at the end of the time step is calculated:

$$S_1(t+\triangle t) = S_1(t) + (R_1(t) - Q_1(t) + Perc(t))\triangle t \tag{A14}$$

The total discharge $q(t)$ is then simply the sum of $Q_1(t)$ and $Q_2(t)$. A triangular unit hydrograph is then used for runoff routing.

[49] As a summary, the model contains ten time-invariant parameters (λ, $S_{\max}$, b, α, P, β, γ, $S_{2,\max}$, κ_2, and κ_1), and three state variables per time step (the storages $S(t)$, $S_1(t)$, and $S_2(t)$).

[50] **Acknowledgments.** The authors would like to express their gratitude to the Department Operational Water Management of the Ministry of the Flemish Community for providing the meteorological and discharge data. Gabriëlle De Lannoy is a postdoctoral research fellow of the Research Foundation Flanders. The research in this paper is partly funded by the Belgian Science Policy Office in the frame of the STEREO II programme (project SR/00/100). We would also like to thank Günter Blöschl and Jürgen Komma for the discussions and comments regarding this paper.

References

Blöschl, G., C. Reszler, and J. Komma (2007), A spatially distributed flash flood forecasting model, *Environ. Modell. Software*, *23*(4), 464–478.

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

De Lannoy, G. J. M., P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest (2007), Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter, *Water Resour. Res.*, *43*, W09410, doi:10.1029/2006WR005449.

Doherty, J. (2001), *PEST-ASP Users' Manual*, Watermark Numer. Comput., Brisbane, Qld, Australia.

Duan, Q. Y., S. Sorooshian, and V. K. Gupta (1994), Optimal use of the SCE-UA global optimization method for calibrating watershed models, *J. Hydrol.*, *158*(3–4), 265–284.

Durand, M., and S. A. Margulis (2007), Correcting first-order errors in snow water equivalent estimates using a multifrequency, multiscale data assimilation scheme, *J. Geophys. Res.*, *112*, D13121, doi:10.1029/2006JD008067.

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

Ferket, B. (2008), The relative importance of accurate catchment averaged evapotranspiration inputs and automatic calibration procedures for operational flood forecasts, M.Sc. dissertation, Fac. of Biosci. Eng., Ghent Univ., Ghent, Belgium.

Gan, T. Y., and G. F. Biftu (1996), Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure, *Water Resour. Res.*, *32*(12), 3513–3524.

Gelb, A. (1974), *Applied Optimal Estimation*, MIT Press, Boston, Mass.

Gill, M. K., Y. H. Kaheil, A. Khalil, M. McKee, and L. Bastidas (2006), Multiobjective particle swarm optimization for parameter estimation in hydrology, *Water Resour. Res.*, *42*, W07417, doi:10.1029/2005WR004528.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *82*, 35–55.

Komma, J., G. Blöschl, and C. Reszler (2008), Soil moisture updating by ensemble Kalman filtering in real-time flood forecasting, *J. Hydrol.*, *357*(3–4), 228–242.

Linström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström (1997), Development and test of the distributed HBV-96 hydrological model, *J. Hydrol.*, *201*(1–4), 272–288.

Matgen, P., J.-B. Henry, L. Hoffmann, and L. Pfister (2006), Assimilation of remotely sensed soil saturation levels in conceptual rainfall-runoff models, in *Prediction in Ungauged Basins: Promise and Progress*, edited by M. Sivapalan et al., *IAHS Publ.*, *303*, 226–234.

Moore, R. J. (2007), The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, *11*(1), 483–499.

Moradkhani, H., K.-L. Hsu, H. Gupta, and S. Sorooshian (2005), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, *41*, W05012, doi:10.1029/2004WR003604.

Parajka, J., R. Merz, and G. Blöschl (2006), Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale, *Hydrol. Earth Syst. Sci.*, *10*(3), 353–368.

Pauwels, V. R. N. (2008), A multistart weight-adaptive recursive parameter estimation method, *Water Resour. Res.*, *44*, W04416, doi:10.1029/2007WR005866.

Pauwels, V. R. N., and G. J. M. De Lannoy (2006), Improvement of modeled soil wetness conditions and turbulent fluxes through the assimilation of observed discharge, *J. Hydrometeorol.*, *7*(3), 458–477.

Pauwels, V. R. N., R. Hoeben, N. E. C. Verhoest, and F. P. De Troch (2001), The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation, *J. Hydrol.*, *251*(1–2), 88–102.

Pauwels, V. R. N., R. Hoeben, N. E. C. Verhoest, F. P. De Troch, and P. A. Troch (2002), Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values, *Hydrol. Processes*, *16*(5), 995–1013.

Reed, P., B. Minsker, and D. Goldberg (2000), Designing a competent simple genetic algorithm for search and optimization, *Water Resour. Res.*, *36*(12), 3757–3761.

Reed, P., B. S. Minsker, and D. E. Goldberg (2003), Simplifying multiobjective optimization: An automated design methodology for the nondominated sorted genetic algorithm-II, *Water Resour. Res.*, *39*(7), 1196, doi:10.1029/2002WR001483.

Reichle, R. H. (2008), Data assimilation methods in the Earth sciences, *Adv. Water Resour.*, *31*(11), 1411–1418.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002a), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*(1), 103–114.

Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser (2002b), Extended versus ensemble Kalman filtering for land data assimilation, *J. Hydrometeorol.*, *3*(6), 728–740.

Sabater, J. M., L. Jarlan, J. C. Calvet, F. Bouyssel, and P. De Rosnay (2007), From near-surface to root-zone soil moisture using different assimilation techniques, *J. Hydrometeorol.*, *8*(2), 194–206.

Shuttleworth, W. J. (1992), Evaporation, in *Handbook of Hydrology*, edited by D. R. Maidment, pp. 4.1–4.53, McGraw-Hill, New York.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521–2536, doi:10.1029/2000WR900405.

Thyer, M., G. Kuczera, and B. C. Bates (1999), Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms, *Water Resour. Res.*, *35*(3), 767–773.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sooroosthian (2003a), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sooroosthian (2003b), A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1201, doi:10.1029/2002WR001642.

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

Weerts, A. H., and G. Y. H. El Serafy (2006), Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water Resour. Res.*, *42*, W09403, doi:10.1029/2005WR004093.

Wood, E.F., and A. Szöllösi-Nagy (1980), *Real-Time Forecasting/Control of Water Resource Systems: Selected Papers From an IIASA Workshop, October 18–21, 1976*, Pergamon, Oxford, U. K.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrologic models, *J. Hydrol.*, *204*(1–4), 83–97.

G. J. M. De Lannoy and V. R. N. Pauwels, Laboratory of Hydrology and Water Management, Ghent University, Coupure Links 653, B-9000 Ghent, Belgium. (valentijn.pauwels@ugent.be)