Journal of Hydrology 513 (2014) 127–141

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Assessing model state and forecasts variation in hydrologic data assimilation

Jos Samuel [a,*], Paulin Coulibaly [a,b,c], Gift Dumedah [b], Hamid Moradkhani [c]

[a] McMaster University, Department of Civil Engineering and School of Geography and Earth Sciences, 1280 Main Street West, Hamilton, Ontario L8S 4L7, Canada
[b] Monash University, Department of Civil Engineering, Wellington Road, Clayton, Victoria 3800, Australia
[c] Portland State University, Department of Civil & Environmental Engineering, 1930 S.W. 4th Avenue, Suite 200, Portland, OR 97201, USA

ARTICLE INFO

Article history:
Received 20 July 2013
Received in revised form 20 January 2014
Accepted 22 March 2014
Available online 1 April 2014
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Emmanouil N. Anagnostou, Associate Editor

Keywords:
Combined streamflow and soil moisture assimilation
Data assimilation
Dual state-parameter estimation
Ensemble Kalman Filter
Soil moisture

SUMMARY

Data assimilation (DA) has been widely used in hydrological models to improve model state and subsequent streamflow estimates. However, for poor or non-existent state observations, the state estimation in hydrological DA can be problematic, leading to inaccurate streamflow updates. This study evaluates the soil moisture and flow variations and forecasts by assimilating streamflow and soil moisture. Three approaches of Ensemble Kalman Filter (EnKF) with dual state-parameter estimation are applied: (1) streamflow assimilation, (2) soil moistue assimilation, and (3) combined assimilation of soil moisture and streamflow. The assimilation approaches are evaluated using the Sacramento Soil Moisture Accounting (SAC-SMA) model in the Spencer Creek catchment in southern Ontario, Canada. The results show that there are significant differences in soil moisture variations and streamflow estimates when the three assimilation approaches were applied. In the streamflow assimilation, soil moisture states were markedly distorted, particularly soil moisture of lower soil layer; whereas, in the soil moisture assimilation, streamflow estimates are inaccurate. The combined assimilation of streamflow and soil moisture provides more accurate forecasts of both soil moisture and streamflow, particularly for shorter lead times. The combined approach has the flexibility to account for model adjustment through the time variation of parameters together with state variables when soil moisture and streamflow observations are integrated into the assimilation procedure. This evaluation is important for the application of DA methods to simultaneously estimate soil moisture states and watershed response and forecasts.

1. Introduction

Data assimilation (DA) remains popular with its innovative approach of accounting for uncertainties in model, observation, and forcing data. In hydrology, the aim includes accurate estimation of model states through optimal merging of inaccurate model simulations with uncertain observation (Clark et al., 2008; Moradkhani et al., 2005a, 2012; Reichle, 2008; Moradkhani and Meskele, 2009; Liu et al., 2012; Weerts and El Serafy, 2006). But state estimation in hydro logical DA is problematic, particularly, in the assimilation of streamflow where observation data for the validation of state variables such as soil moisture are usually non-existent or of poor quality. In the absence of soil moisture observation, the typical procedure has been to update state variables using streamflow observations (Clark et al., 2008; Moradkhani et al., 2005a, 2005b; Salamon and Feyen, 2010; Schoups et al., 2010). An apparent implication of this approach is that model states are usually distorted in favor of the model to better fit the streamflow observations alone. It is now well recognized that inaccurate model states have direct impact on model initialization, and subsequently on model forecasts which are made with these initial conditions (Seo et al., 2003; Lee et al., 2011). These have important implications in DA applications for the estimation of watershed state variables and their response to streamflow and model forecasts.

Soil moisture is a key variable responsible for partitioning of water and energy fluxes at the land surface, and has been shown to improve streamflow estimation (Berg and Mulroy, 2006; Brocca et al., 2010; Legates et al., 2010; Reichle et al., 2008a,b), even in ungauged basins (Dumedah and Coulibaly (2012a). Detailed in situ and remote soil moisture data can be more easily obtained through recent technological advances and/or soil moisture sensor installments (Kornelsen and Coulibaly, 2013). Remotely-sensed

* Corresponding author. Current address: Yukon College, Yukon Research Centre, 500 College Drive, Whitehorse, Yukon Y1A 5K4, Canada. Tel.: +1 867 668 8849; fax: +1 867 456 8672.
E-mail address: josmartinussamuel@yahoo.com (J. Samuel).

http://dx.doi.org/10.1016/j.jhydrol.2014.03.048
0022-1694/

soil moisture retrievals have been used in hydrologic DA to improve runoff prediction (Crow and Ryu, 2009) and soil moisture estimates (Crow and Van Loon, 2006; Reichle et al., 2008a,b). However, few studies have actually used soil moisture data to update soil moisture state in the assimilations of streamflow (Aubert et al., 2003; DeChant and Moradkhani, 2011a; Thirel et al., 2010a, 2010b; Lee et al., 2011). Aubert et al. (2003) showed that the integration of soil moisture data in the assimilation procedure allows a better control of the model state time variation. Lee et al. (2011) demonstrated that assimilating in situ soil moisture data in addition to streamflow data significantly improves prediction of soil moisture and streamflow. More recently, Dumedah and Coulibaly (2012b) have shown the potential of soil moisture assimilation in improving flow estimates. However, a detailed examination of model state distortions due to lack of validation data in assimilation procedure, and the impacts on flow and soil moisture forecasts have not been thoroughly conducted in the DA literature. As a result, this study will conduct three experiments to investigate state variations in hydrological DA.

In the first experiment, streamflow only is used to update the model states, whereas the second experiment uses only soil moisture to drive the assimilation. The third experiment uses both soil moisture and streamflow observations to propagate model states and parameters through assimilation time steps. The three experiments are demonstrated for the Sacramento Soil Moisture Accounting (SAC-SMA) model in the Spencer Creek catchment in southern Ontario, Canada. The study estimates soil moisture state in the watershed for each DA experiment in order to quantify the variations of soil moisture and streamflow when streamflow and/or soil moisture is applied to drive the assimilation. Flow and soil moisture forecasts for different lead times are further evaluated to assess the assimilation procedures. These evaluations are important for the application of DA methods to simultaneously estimate soil moisture states and watershed response.

Data assimilation is popular due to its ability to merge diverse range of observations with a dynamic model, and has increasingly been applied to facilitate flow/flood forecasting operations (Weerts et al., 2010; Kitanidis and Bras, 1980; Seo et al., 2003, 2009; Dumedah and Coulibaly, 2012b; McMillan et al., 2013). Many hydrologic studies have reported the use of DA for estimating hydrologic states, parameters and fluxes including streamflow (e.g., Clark et al., 2008; El Serafy and Mynett, 2008; Moradkhani, 2005a,b; Vrugt et al., 2005; Moradkhani and Sorooshian, 2008; Parrish et al., 2012; Moradkhani et al., 2012), soil moisture (e.g., Crow and van Loon, 2006; Houser et al., 1998; Monsivais-Huertero et al., 2010; Montzka et al., 2011; Schaake et al., 2004; DeChant and Moradkhani, 2011b), and snow cover and snow water equivalent (Slater and Clark, 2006; Durand and Margulis, 2006; Andreadis and Lettenmaier, 2006; DeChant and Moradkhani, 2011a, 2011b; Leisenring and Moradkhani, 2011). The most popular data assimilation technique used in many hydrological applications is based on the Kalman filter (Kalman, 1960). The standard Kalman Filter which is used on linear dynamic systems has been modified to the Extended Kalman Filter (EKF) for state estimation in nonlinear systems (Evensen, 1992). However, the EKF is limited to the use of linear approximation which may result in some instability in the estimation when the degree of nonlinearity in the system increases (Dong et al., 2007; Leisenring and Moradkhani, 2011). One recent development of Kalman filtering is the Ensemble Kalman Filter (EnKF) which uses statistical distributions to represent uncertainties of both observations and model errors and to generate ensembles of model forcing and other variables (Evensen, 2003; Leisenring and Moradkhani, 2011; Liang, 2004).

The hydrologic data assimilation for dual state-parameter estimation was developed for better conformity of model output with observation benefiting from temporal organization of the data (Moradkhani et al., 2005b; Moradkhani, 2008). Simultaneous state-parameter estimation using sequential data assimilation is ideal to handle the tight coupling between model states and relevant parameters and allows the parameters dynamically change over time especially when some functional changes in the system of interest is experienced or anticipated (Moradkhani et al., 2012). As a result of these benefits and the ability of this DA approach to merge diverse range of observations (including soil moisture and streamflow) with a dynamic model, the dual EnKF was used in this study to assess soil moisture and flow variations in hydrological DA. Detailed description of the dual-state parameter estimation method based on EnKF can be found in Moradkhani et al. (2005a).

The organization of the paper is presented as follows. Section 2 describes the study area, data and methods. This section also presents the description of the EnKF with dual state-parameter estimation by assimilating streamflow only, soil moisture only and combined streamflow and soil moisture. The extended SAC-SMA model used in this study is also described in this section. Section 3 shows the evaluation of soil moisture and flow variations and forecasts resulting from assimilation of streamflow only, soil moisture only and combined streamflow and soil moisture. The key findings and implications of the study are presented in the conclusion section in Section 4.

2. Data and methods

2.1. Study area

The study is applied in Spencer Creek watershed located in the Southern Ontario, Canada. Spencer Creek flows through the western end of Lake Ontario by Cootes Paradise and has a basin area of approximately 291 km^2. We focus our analysis on three upper sub-basins of this watershed: (1) Westover, (2) Highway-5 and (3) Dundas sub-basins (Fig. 1). Sub-basin area, drainage area, historical flows (for the 1989–2009 period) and dominant land cover and soil types of each sub-basin are presented in Table 1 and the spatial distribution of soil types is shown in Fig. 1b. The watershed consists of mainly wetlands, tallgrass prairie and forests (HRCA, 1990; James, 1994; Sultana and Coulibaly, 2011) and predominant soils are sandy loam, loam, silt loam, and loam (see Table 1). The runoff ratio, measuring the long-term water balance separation between water released from the catchment as runoff and evapotranspiration, is about the same and relatively low (0.15, Table 1). Basins exhibiting low runoff ratios are typically dominated by subsurface flows. The average annual peak flows are slightly different among the sub-basins. The highest average annual peak flow is found in Highway-5 (9.7 mm, Table 1), and the smallest one is found in Westover (6.6 mm) indicating that quick runoff is possibly more dominant in the Highway-5 basin than in the Westover basin. There are two reservoirs along the main channel, those are Valens Reservoir and Christie Dam (in Fig. 1). The Valens reservoir is used for low flow augmentation and to reduce downstream flooding in the spring season. The Christie Dam was constructed to reduce flooding within the Dundas town during the spring (HRCA, 1990; Sultana and Coulibaly, 2011).

2.2. Data

Data used in this study are observed daily precipitation, temperature, in situ soil moisture and streamflow. Precipitation and temperature data are obtained from Environmental Canada weather stations near the basin, namely Hamilton RBG CS, Hamilton Airport stations, and McMaster Britannia weather station (Fig. 1). Streamflow data are obtained from Water Survey Canada (Environmental Canada). If there is missing data on a particular day, data

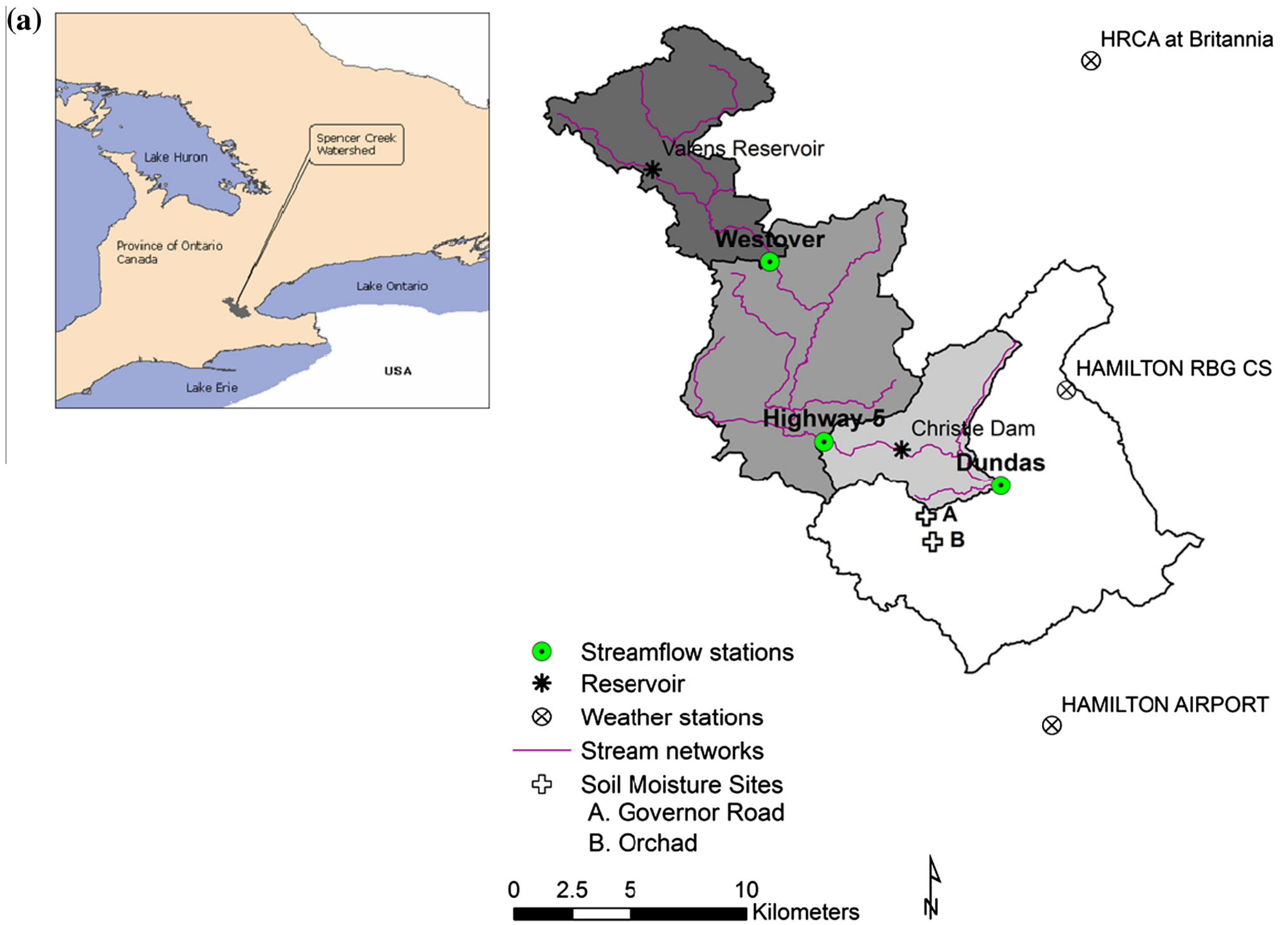

Fig. 1. Location map of Spencer Creek watershed and (1) Westover, (2) Highway 5 and (3) Dundas sub-basins.

are infilled using discharge data obtained from Hamilton Region Conservation Authority (HRCA) at the same sites. Inflow and outflow of water from the Christie Dam were obtained from HRCA. To estimate potential evapotranspiration, we used a simplified version of Thornwaite equation (Samuel et al., 2011; SMHI, 2005; Thornthwaite, 1948). The spatial locations of two in situ (point-based) soil moisture networks, Governor Road and Orchard are shown in Fig. 1. A detailed description of these soil moisture sensor networks can be found in Kornelsen and Coulibaly (2013). The length of data period used for this study is from 23 October 2006 to 31 December 2009 with daily time interval.

2.2.1. Simulated soil moisture with Hydrus

Due to the fact that soil moisture observations were measured only in few points within the study basin, the observed soil moisture does not compare well with the areal averaged soil moisture outputs resulting from the lumped hydrological model (SAC-SMA). Thus, the soil moisture observations should not be directly used to assimilate soil moisture into the SAC-SMA model. In this study, the HYDRUS-1D model was used to estimate the averaged volumetric soil moisture of each sub-basin. This software is widely used for simulating water, heat and solute movement in one-dimensional variably-saturated media (Šimůnek and van Genuchten, 2008; Šimůnek et al., 2012). Hydrus specifically generates a one-dimensional volume of water in the entire flow domain depending on weather conditions and soil types. In a recent study, Montzka et al. (2011) used Hydrus model in a data assimilation framework using particle filtering to explore the potential of using remotely sensed soil moisture observations from different satellite platforms to obtain soil moisture profile and soil hydraulic properties. In this study, the simulated soil moisture data has been validated using the observed (point-based) soil moisture at the two in situ soil moisture sites: Governor Road and Orchard. Fig. 2 illustrates a successful generation of simulated soil moisture data in the Governor Road soil moisture networks. Similar outputs are also found in the Orchard soil moisture networks since both sites have similar weather conditions, soil types and in situ soil moisture variability. To generate the estimates of the average volumetric soil moisture for each sub-basin, (1) each point in the study basins having the same weather conditions and soil types were grouped/merged, (2) each group was simulated in Hydrus to obtain the estimated volumetric soil moisture, and (3) the average of these volumetric soil moistures for each sub-basin was computed. These Hydrus soil moisture data are used as “observation” soil moisture to drive the data assimilation procedure. Note that in this case, the Hydrus volumetric soil moisture are in the same unit as that obtained from the SAC-SMA hydrological model (i.e. in mm).

2.3. Hydrologic modeling using Sacramento Soil Moisture Accounting (SAC-SMA) model

The rainfall-runoff model used in this study is an extended version of the conceptual Sacramento Soil Moisture Accounting (SAC-SMA) model. Detailed description of the original SAC-SMA model can be found in Burnash et al. (1973), Burnash (1995), Koren et al. (2004). We extended the model by adding the snow routine and routing components, and the existence of a reservoir. Fig. 2 presents the schematic of the extended SAC-SMA model, while

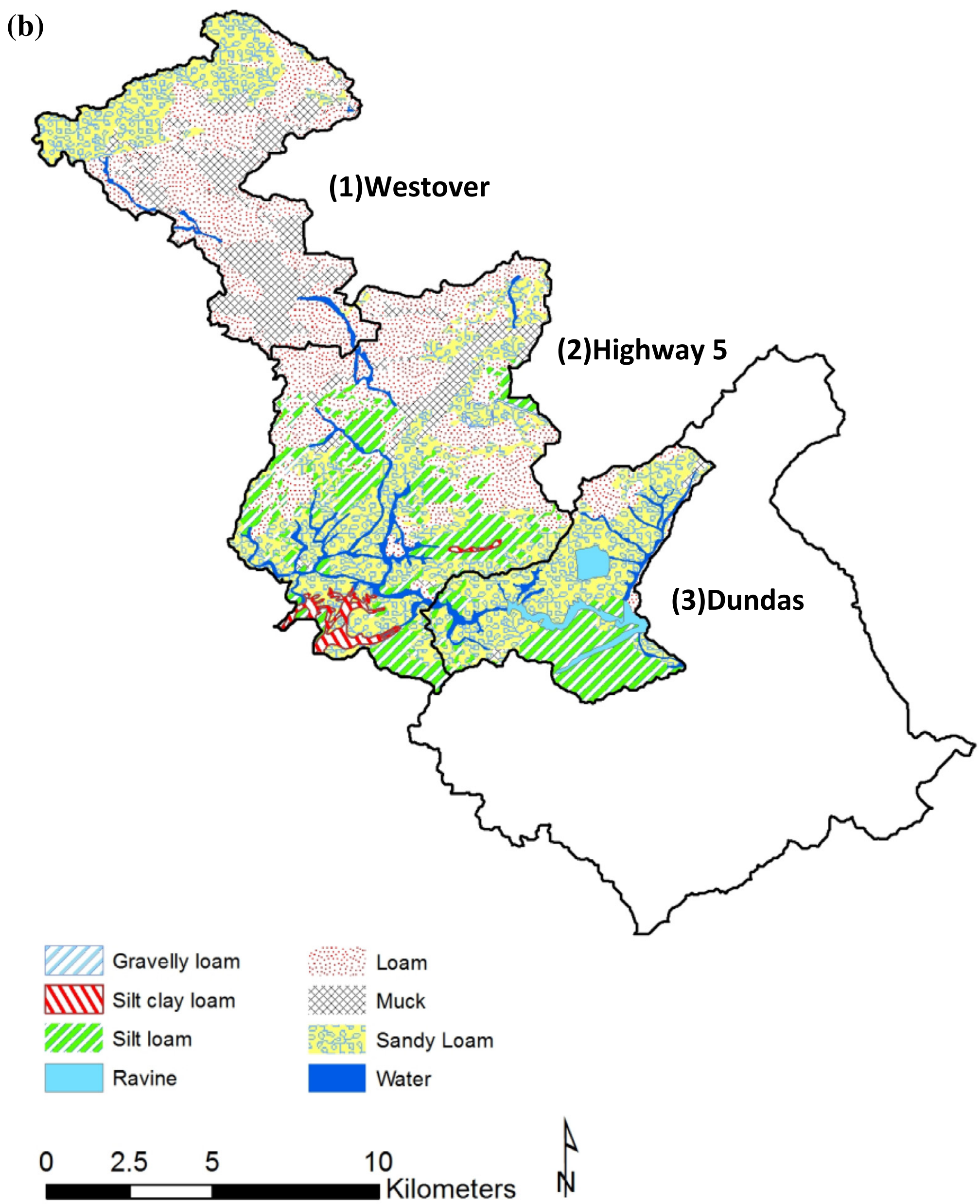

Fig. 1 (*continued*)

Table 1
Basin characteristics.

Gauging station (River)	Sub-basin size (km^2)	Dominant landcover and soil	Historical flow (1989–2009)
Westover (station number: 02HB015)	53.2	Drainage area = 53.2 km^2 *Land cover:* Wetland, deciduous Forest, Tallgrass Woodland *Predominant soil type:* Sandy loam and loam	Average daily flow = 0.62 m^3/s Average annual peak flow = 4.03 m^3/s or 6.6 mm Runoff ratio = 0.15
Highway-5 (station number: 02HB023)	132	Drainage area = 78.8 km^2 *Land cover:* Open Tallgrass Prairie, deciduous forest *Predominant soil type:* Loam and silt loam	Average daily flow = 1.6 m^3/s Average annual peak flow = 14.08 m^3/s or 9.7 mm Runoff ratio = 0.15
Dundas (station number: 02HB007)	170	Drainage area = 38 km^2 *Land cover:* Open Tallgrass Prairie *Soil type:* Sandy loam, silt loam and loam	Average daily flow = 2.0 m^3/s Average annual peak flow = 15.5 m^3/s or 7.8 mm Runoff ratio = 0.15

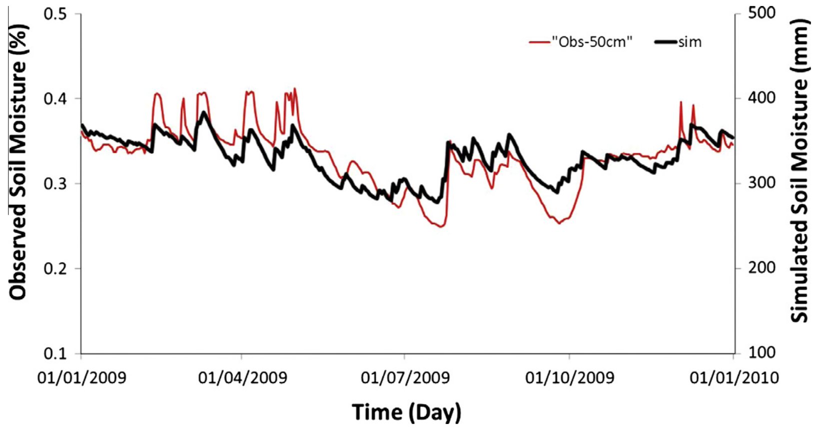

Fig. 2. Validation of soil moisture estimation derived from Hydrus model.

Table 2
Parameters of (a) SAC-SMA, (b) degree-day factor snow routine and (c) Nash-cascade routing routine and their uncertainty ranges.

Parameter	Definition	Ranges
(a) SAC-SMA model components		
UZTWM	Upper-zone tension water maximum storage (mm)	1–100
UZFWM	Upper-zone free water maximum storage (mm)	1–100
LZTWM	Lower-zone tension water maximum storage (mm)	50–500
LZFPM	Lower-zone free water primary maximum storage (mm)	100–2000
LZFSM	Lower-zone free water supplemental maximum storage (mm)	100–1000
ADIMP	Additional impervious area (–)	0.01–0.1
Recession parameters		
UZK	Upper-zone free water lateral depletion rate (day^{-1})	0.01–0.99
LZPK	Lower-zone primary free water depletion rate (day^{-1})	0.0001–0.02
LZSK	Lower-zone supplemental free water depletion rate (day^{-1})	0.1–0.5
Percolation and other		
ZPERC	Maximum percolation rate (–)	1–50
REXP	Exponent of the percolation equation (–)	1–50
PCTIM	Impervious fraction of the watershed area (–)	0.0–0.1
PFREE	Fraction percolating from upper- to lower-zone free water storage (–)	0–0.5
RIVA	Riparian vegetation area (–)	0
SIDE	Ratio of deep recharge to channel base flow (–)	0
SAVED	Fraction of lower zone free water not transferable to tension water	0
States in soil moisture content		
UZTWC	Upper-zone tension water storage content (mm)	
UZFWC	Upper-zone free water storage content (mm)	
LZTWC	Lower-zone tension water storage content (mm)	
LZFPC	Lower-zone free primary water storage content (mm)	
LZFSC	Lower-zone free secondary water storage content (mm)	
States in additional impervious area content		
ADIMC	Additional impervious area content directly link to the stream network (mm)	
(b) Snow routine components		
DDF	Degree day factor	1–5
SCF	Snowfall correction factor	0.8–1.5
TR	An upper threshold temperature, to distinguish between rainfall, snowfall and a mix of rain and snow	0–2.5
ATHORN	A constant for Thornthwaite's equation	0.1–0.3
RCR	Rainfall correction factor	0.8–1.5
TS	A lower threshold temperature to distinguish between rainfall, snowfall and a mix of rain and snow	0.
TM	Melting threshold temperature	0.
States in the snow routine components		
SWE	Snow water equivalent (mm)	
(c) Nash-cascade routing components		
RQ	Residence time parameters of quick-flow	0.3–0.9
States in the Nash-cascade routing components		
UHG1	Three linear reservoirs to route the upper zone (quick response) channel inflow (mm)	
UHG2		
UHG3		

descriptions of each model parameter and model state are presented in Table 2. In total, the model uses 19 time-variant parameters (*UZTWM*, *UZFWM*, *UZK*, *PCTIM*, *ADIMP*, *ZPERC*, *REXP*, *LZTWM*, *LZFSM*, *LZFPM*, *LZSK*, *LZPK*, *PFREE*, *RQ*, *DDF*, *SCF*, *TR*, *ATHORN*, and *RCR*), 5 fixed parameters (*RIVA*, *SIDE*, *SAVED*, *TM*, and *TS*), and consists of 10 model states (*UZTWC*, *UZFWC*, *LZTWC*, *LZFPC*, *LZFSC*, *ADIMC, SWE, UHG1, UHG2,* and *UHG3*) as described in Table 2. The accumulation of water in soil is represented in 2 layers (upper and lower soil layers). Five storages representing the accumulation of water are used: UZTWC (Upper-Zone Tension Water Content), UZFWC (Upper-Zone Free Water Content), LZTWC (Lower-Zone Tension Water Content), LZFSC (Lower-Zone Free Supplementary Water Content), and LZFPC (Lower-Zone Free Primary Water Content) (see also Fig. 3 for illustration). The first two storages (UZTWC, and UZFWC) constitute the upper soil layer, whereas the last three storages (LZTWC, LZFSC, LZFPC) represent the lower soil layer. The "soil moisture" refers to the sum of water storages content of the upper and lower soil layers. Each soil layer consists of tension and free water storages. Free water reservoir of each layer can generate runoff depending on a depletion coefficient of the storage. The upper zone corresponds to surface soil processes and interception storage. If the upper layer storage is full and precipitation exceeds the interflow and percolation capacities, saturation excess overland flow occurs. The lower soil storages represent deeper soil processes and groundwater storage. The free water storage parameters of the lower soil layers, namely *LZFSM* and *LZFPM* (i.e. Lower-Zone Free Water Supplemental and Primary Maximum Storage, respectively) are modeled to control supplemental (fast) and primary (slow) base flow, respectively. This model uses nonlinear equations to control the partitioning of precipitation into overland flow, infiltration to the upper zone, interflow, percolation to the lower zone, and fast and slow components of groundwater recession baseflow. The fast (*Z2*) and slow (*Z3*) components of groundwater recession baseflow are assumed to flow directly to the watershed outlet (see the illustration in Fig. 3). The runoff from impervious area, direct, surface and interflow runoff are routed using a simple Nash Cascade (NC) of three linear reservoirs to produce quick response channel inflow (*Z1*) (following Vrugt et al., 2006). This channel routing configuration adds one parameter and three state variables to the model.

To represent changes in the snowpack of the basin, a degree-day factor was used in the snow routine. This involves a simple conceptual snow routine since our current study objective is focused on evaluating the soil moisture storages. Three threshold temperatures (namely *TR*, *TS* and *TM*) are used to distinguish between rainfall, snowfall and melting temperature, respectively. Parameter *TR* was assumed variable; whereas 2 other parameters were set constant ($TS = 0$ °C, $TM = 0$ °C). The amount of melt water depends on parameter *DDF* (Degree Day Factor). Changes in snowpacks are computed depending on corrected snowfall and amount of melt water. Details of the applied snow routine can be found in Samuel et al. (2011).

To model water flow from upstream to downstream sub-basins (i.e. from Westover to Highway 5 and Highway 5 to Dundas), we assumed that inflow from a sub-basin flows through a river channel from the outlet of the upstream sub-basin to the stream inlet of the current sub-basin where the local runoff is added. In addition, to model a reservoir, namely Christie Dam in Dundas sub-basin in the hydrological model, we adopted an approach proposed by Payan et al. (2008). This approach requires no additional parameter since it uses the observed values of stored volumes accounting for the behavior of the modeled reservoir. Assuming the water level in the store is V_t at time step t and V_{t+1} at time step $t+1$ are known. If $V_{t+1} > V_t$, i.e., the stored volume increases, a positive quantity of water $|\Delta V| = |V_{t+1} - V_t|$ is taken by the modeled reservoir from the natural system. On the other hand, if $V_{t+1} < V_t$, i.e., the stored volume decreases, amount of water $|\Delta V| = |V_{t+1} - V_t|$ is assumed to leave the modeled reservoir and to feed the natural system. This amount of water is added to the content of the routing routine. If $V_{t+1} = V_t$, there is no water exchange. Note that Valens reservoir is not included in the hydrological model due to limited historical inflow data available for this reservoir.

2.4. Data assimilation for combined state-parameter estimation

The dual EnKF used in this study estimates simultaneous states and parameters using sequential data assimilation (Moradkhani et al., 2005a). Detailed description of this procedure can be found in Moradkhani et al. (2005a). In general, the dual EnKF requires two separate state-space representation for the state and

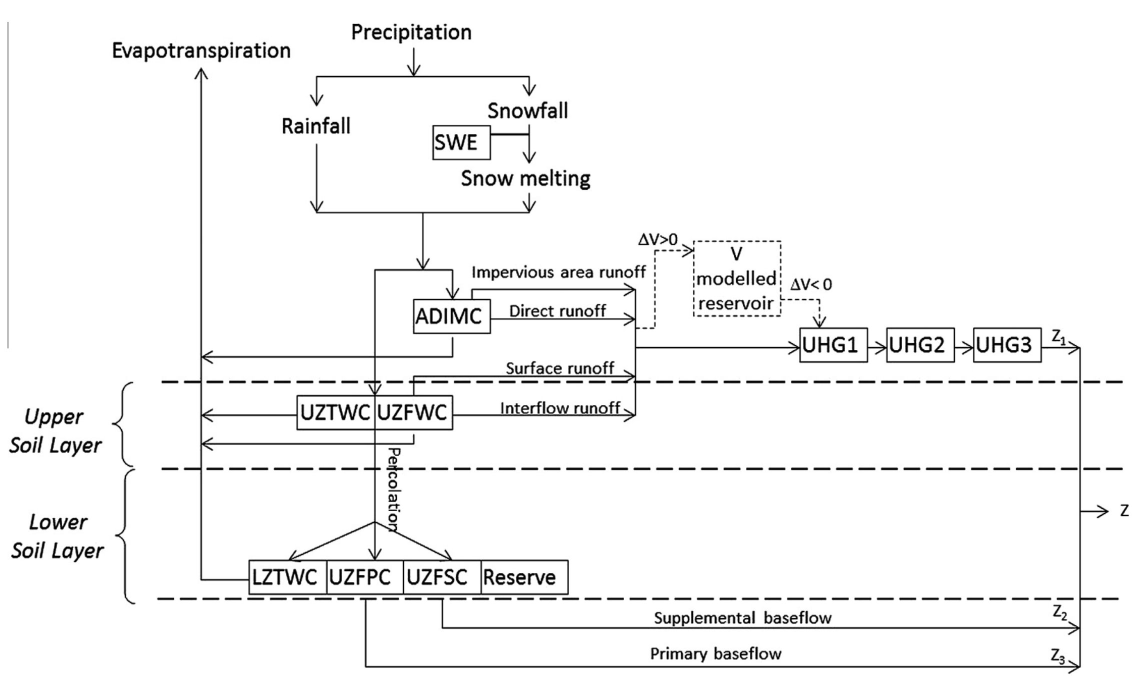

Fig. 3. Schematic representation of extended SAC-SMA model (including snow routine and Nash-cascade routing routine). The rectangles refer to the various model states. *Z*1, *Z*2, and *Z*3 refer to the three main channel components, which sum together to form the streamflow *Z* at the watershed outlet. A modeled reservoir was added for Dundas sub-basin where Christie Dam is located.

parameters through two interactive filters by updating model parameters and states. In this study, model parameters were first updated and then states. Three variations of the dual state-parameter estimation were considered in this study as follows:

(1) Streamflow data assimilation

In this approach, streamflow is the main source of observation used to update both model parameters and states. Given the errors in forcing data and observations, model parameters were first updated followed by state variables for each assimilation time step. Detailed procedure of this approach is presented in Section 2.4.1.

(2) Soil moisture data assimilation

In soil moisture data assimilation approach, "observed" soil moisture is used to update both model parameters and states. Note that "observed" soil moisture is estimated from the Hydrus model (as discussed in Section 2.2), and that observed streamflow data were not used here. Similar to the streamflow assimilation approach, model parameters were first updated and then the state variables for each assimilation time step. This procedure is also repeated until the final period of assimilation is reached (see detailed procedure in Section 2.4.1).

(3) Combined streamflow and soil moisture data assimilation

This approach uses both soil moisture and streamflow observations to update state and model parameters. Model parameters were updated by assimilating streamflow and at the same time states were updated by assimilating soil moisture (combined assimilation). Detailed procedure of this approach is presented in the following section.

2.4.1. Procedure

Here we present the technical approach for dual parameter-state estimation for the above three scenarios. Note that the indices (a), (b) and (c) in Eqs. (5) to (11) are used to indicate that the equations are specifically applied for (a) streamflow only, (b) soil moisture only, and (c) combined assimilation approaches, respectively. In cases where there are no indices, the equations are used for all three approaches.

(1) Specify ensemble size and total period of assimilation. Note that superscript (−) is used to indicate forecasted parameters or states, superscript (+) to indicate updated parameters or states by assimilating measured streamflow and/or soil moisture.
(2) Generate the parameters ensemble by perturbing the updated parameters from previous time step.

$$\theta_{t+1}^{i-} = \theta_t^{i+} + \tau_t^i, \quad \tau_t^i \sim N\left(0, \sum_t^{\theta}\right) \quad i = 1, \ldots, n \tag{1}$$

where θ_{t+1}^{i-} and θ_t^{i+} are the ith ensemble member of forecasted model parameter at time $t+1$ and the ensemble member of updated model parameter at time t, respectively. While, τ_t^i is model parameter error with variance $\sum_t^{\theta}$ which is assumed to be Gaussian and n is ensemble size. The $\sum_t^{\theta}$ of each model parameter was set to be 1% of the variance of the ensemble member of each updated model parameter at time t. It is noteworthy that the initial model parameters were generated randomly using a uniform distribution.

(3) Generate the ensemble of model states (x_{t+1}^{i-}) and streamflow predictions ($\hat{y}_{t+1}^i$) using the forecasted parameters (θ_{t+1}^{i-}) and replicates of forcing data (u_t^i):

$$x_{t+1}^{i-} = f(x_t^{i+}, u_t^i, \theta_{t+1}^{i-}) \tag{2}$$

$$\hat{y}_{t+1}^i = h(x_{t+1}^{i-}, \theta_{t+1}^{i-}) \tag{3}$$

where x_{t+1}^{i-} is the ith ensemble member of forecasted states at time $t+1$ and x_t^{i+} is the ith ensemble member of updated states at time t.

The input uncertainty is characterized by perturbing the forcing data (precipitation and temperature). The normal distribution is used for perturbing temperature and the lognormal distribution is used for perturbing precipitation. The forcing data perturbations are generated at each time step and for all ensemble size as follows:

$$u_t^i = u_t + \varsigma_t^i, \quad \varsigma_t^i \sim N(0, \Sigma_t^u) \tag{4}$$

where ς_t^i and Σ_t^u are the noise and covariance of the forcing data, respectively. The Σ_t^u values for perturbing precipitation and temperature data were set to be 1% of the total daily precipitation and 1% of the mean daily temperature at time t, respectively.

(4) Update the ensemble of parameters according to the standard Kalman equation

$$\theta_{t+1}^{iy+} = \theta_{t+1}^{iy-} + K_{t+1}^{\theta y}(y_{t+1}^{iy} - \hat{y}_{t+1}^i) \tag{5a}$$

$$\theta_{t+1}^{is+} = \theta_{t+1}^{is-} + K_{t+1}^{\theta s}(s_{t+1}^{is} - \hat{s}_{t+1}^i) \tag{5b}$$

$$\theta_{t+1}^{iys+} = \theta_{t+1}^{iys-} + K_{t+1}^{\theta ys}(y_{t+1}^{iys} - \hat{y}_{t+1}^i) \tag{5c}$$

where,

$$K_{t+1}^{\theta y} = \Sigma_{t+1}^{\theta y}[\Sigma_{t+1}^{yy} + \Sigma_{t+1}^{y}]^{-1} \tag{6a}$$

$$K_{t+1}^{\theta s} = \Sigma_{t+1}^{\theta s}[\Sigma_{t+1}^{ss} + \Sigma_{t+1}^{s}]^{-1} \tag{6b}$$

$$K_{t+1}^{\theta ys} = \Sigma_{t+1}^{\theta ys}[\Sigma_{t+1}^{yys} + \Sigma_{t+1}^{ys}]^{-1} \tag{6c}$$

and,

$$y_{t+1}^{iy} = y_{t+1} + \varepsilon_{t+1}^i, \quad \varepsilon_{t+1}^i \sim N(0, \Sigma_{t+1}^y) \tag{7a}$$

$$s_{t+1}^{is} = s_{t+1} + \omega_{t+1}^i, \quad \omega_{t+1}^i \sim N(0, \Sigma_{t+1}^s) \tag{7b}$$

$$y_{t+1}^{iys} = y_{t+1} + \vartheta_{t+1}^i, \quad \vartheta_{t+1}^i \sim N(0, \Sigma_{t+1}^{ys}) \tag{7c}$$

where, $K_{t+1}^{\theta y}, K_{t+1}^{\theta s}, K_{t+1}^{\theta ys}$ are the Kalman gain for correcting the parameter trajectories, and $\Sigma_{t+1}^{\theta y}, \Sigma_{t+1}^{\theta s}, \Sigma_{t+1}^{\theta ys}$ are the cross covariance of parameter ensembles ($\theta_{t+1}^{iy-}, \theta_{t+1}^{is-}, \theta_{t+1}^{iys-}$) for streamflow, soil moisture, and combined assimilation approaches, respectively with measured streamflow (y_{t+1}) and soil moisture (s_{t+1}) and streamflow ($\hat{y}_{t+1}^i$) and soil moisture ($\hat{s}_{t+1}^i$) prediction ensembles; Σ_{t+1}^{yy} is the forecast error covariance matrix of the streamflow prediction ensemble ($\hat{y}_{t+1}^i$), and Σ_{t+1}^{y} is the covariance of the actual streamflow observation. Meanwhile, Σ_{t+1}^{ss} is the forecast error covariance matrix of the soil moisture prediction ensemble ($\hat{s}_{t+1}^i$), and Σ_{t+1}^{s} is the covariance of the actual soil moisture observation. For the combined assimilation, Σ_{t+1}^{yys} and Σ_{t+1}^{ys}, respectively represent the forecast error covariance matrix of the streamflow prediction ensemble and the covariance of the actual streamflow observation. The $\theta_{t+1}^{iy+}, \theta_{t+1}^{is+}, \theta_{t+1}^{iys+}$ are updated parameters of the streamflow, soil moisture, and combined assimilation approaches, respectively.

(5) Compute an ensemble of model states ($x_{t+1}^{iy-}, x_{t+1}^{is-}, x_{t+1}^{iys-}$), streamflow ($\hat{y}_{t+1}^{iy}, \hat{y}_{t+1}^{is}, \hat{y}_{t+1}^{iys}$) and soil moisture ($\hat{s}_{t+1}^{iy}, \hat{s}_{t+1}^{is}, \hat{s}_{t+1}^{iys}$) predictions using updated parameters ($\theta_{t+1}^{iy+}, \theta_{t+1}^{is+}, \theta_{t+1}^{iys+}$)

obtained from step 4 for streamflow, soil moisture and combined assimilation approaches, respectively, and replicates of forcing data (u_t^i) through the forward model:

$$x_{t+1}^{iy-} = f(x_t^{iy+}, u_t^i, \theta_{t+1}^{iy+}) \tag{8a}$$

$$x_{t+1}^{is-} = f(x_t^{is+}, u_t^i, \theta_{t+1}^{is+}) \tag{8b}$$

$$x_{t+1}^{iys-} = f(x_t^{iys+}, u_t^i, \theta_{t+1}^{iys+}) \tag{8c}$$

$$\hat{y}_{t+1}^{iy} = g(x_{t+1}^{iy-}, \theta_{t+1}^{iy+}) \tag{9a}$$

$$\hat{y}_{t+1}^{is} = h(x_{t+1}^{is-}, \theta_{t+1}^{is+}) \tag{9b}$$

$$\hat{y}_{t+1}^{iys} = i(x_{t+1}^{iys-}, \theta_{t+1}^{iys+}) \tag{9c}$$

(6) Update the ensemble of model states according to the standard Kalman equation by assimilating measured streamflow or soil moisture as follows:

$$x_{t+1}^{iy+} = x_{t+1}^{iy-} + K_{t+1}^{xy}(y_{t+1}^{iy} - \hat{y}_{t+1}^{iy}) \tag{10a}$$

$$x_{t+1}^{is+} = x_{t+1}^{is-} + K_{t+1}^{xs}(s_{t+1}^{is} - \hat{s}_{t+1}^{is}) \tag{10b}$$

$$x_{t+1}^{iys+} = x_{t+1}^{iys-} + K_{t+1}^{xys}(s_{t+1}^{iys} - \hat{s}_{t+1}^{iys}) \tag{10c}$$

where,

$$K_{t+1}^{xy} = \Sigma_{t+1}^{xy}[\Sigma_{t+1}^{yy} + \Sigma_{t+1}^{y}]^{-1} \tag{11a}$$

$$K_{t+1}^{xs} = \Sigma_{t+1}^{xs}[\Sigma_{t+1}^{ss} + \Sigma_{t+1}^{s}]^{-1} \tag{11b}$$

$$K_{t+1}^{xys} = \Sigma_{t+1}^{xys}[\Sigma_{t+1}^{yys} + \Sigma_{t+1}^{ys}]^{-1} \tag{11c}$$

$K_{t+1}^{xy}, K_{t+1}^{xs}, K_{t+1}^{xys}$ are the Kalman gain for correcting the model state trajectories, while, $\Sigma_{t+1}^{xy}, \Sigma_{t+1}^{xs}, \Sigma_{t+1}^{xys}$ are the cross covariance of model state ensemble ($x_{t+1}^{iy-}, x_{t+1}^{is-}, x_{t+1}^{iys-}$) for streamflow, soil moisture, and combined assimilation approaches, respectively.

This above procedure is repeated until the final period of the assimilation is reached (that is 1165 number of days). In addition, ensemble size was set to 1000. Ensemble size, range of model parameters and uncertainties in input and output were determined through sensitivity analysis and were the same for each approach. It is noteworthy that the meaningful limits for the variable parameters (as shown in Table 2) were set following the other previous studies such as Samuel et al. (2011), Vrugt et al. (2006) and Dumedah and Coulibaly (2012b). These methods were also evaluated in the open loop simulation (no data assimilation). The optimized model parameters of the open loop simulation were obtained using the Brent's parabolic interpolation method (Samuel et al., 2011) with the ranges of the maximum and minimum model parameters, and the evaluation periods were set same as those which were used to simulate the DA techniques.

2.5. Model validation

The statistics of the assimilation results for the three approaches are evaluated using Root Mean Square Error (RMSE, Eq. (12)), and Volume Error (VE, Eq. (13)). These validation statistics serve well our purpose of comparing relative performances of different DA models.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{sim,i} - x_{obs,i})^2} \tag{12}$$

$$\text{VE} = \sum_{i=1}^{n} x_{obs,i} - \sum_{i=1}^{n} x_{sim,i} \tag{13}$$

where $x_{obs,i}$ and $x_{sim,i}$ are the observed and simulated streamflow or soil moisture, respectively for the ith day, n is the number of days (n = 1097 days, i.e. the 2007–2009 period).

2.6. Forecasting procedures

Variations of streamflow and soil moisture in a forecast mode for different lead times (1, 2, 7, 14, 30, 60 and 90 days) are further investigated. The forecast starts by simulating streamflow from the SAC-SMA rainfall-runoff model using the initial conditions of states and model parameters derived from the DA techniques at the end of the updating date and no additional updating of states and model parameters thereafter. Only precipitation and evaporation are used as input data for the SAC-SMA model to drive the streamflow forecasts. The forecasts are evaluated for each consecutive day in the evaluation period (i.e. from January 1, 2010 to March 31, 2010). For example, for the forecast starting on January 1, 2010, the initial conditions of states and model parameters are obtained from each DA technique at Dec, 31, 2009. This first simulation generates streamflow and soil moisture forecasts for the period of January 1, 2010–March 31, 2010. For the forecast starting on January 2, 2010, the initial conditions of states and model parameters are obtained from each DA technique at January 1, 2010. This second simulation generates streamflow and soil moisture forecasts for the January 2, 2010–April 1, 2010 period. The same simulation is repeated for the other days until March 31, 2010. Therefore, there are 90 values of streamflow and soil moisture forecasts generated for each lead time. This analysis evaluates different forecasting performances of each DA model with different initial conditions of states.

The forecast accuracy is evaluated using RMSE (Eq. (12)) and two additional model performance criteria, namely the Peak Flow Criteria (PFC, Eq. (14)) and Low Flow Criteria (LFC, Eq. (15)), following Coulibaly et al. (2001). Coulibaly et al. (2001) found that the PFC and LFC criteria provide more accurate measures of model performances for the flood and low flow periods, respectively, than the RMSE. A PFC or LFC equal to zero represents a perfect fit.

$$\text{PFC} = \frac{\left(\sum_{i=1}^{Tp}(x_{obs,i} - x_{sim,i})^2 x_{obs,i}^2\right)^{\frac{1}{4}}}{\left(\sum_{i=1}^{Tp} x_{obs,i}\right)^{\frac{1}{2}}} \tag{14}$$

$$LFC = \frac{\left(\sum_{i=1}^{Tl}(x_{obs,i} - x_{sim,i})^2 x_{obs,i}^2\right)^{\frac{1}{4}}}{\left(\sum_{i=1}^{Tl} x_{obs,i}\right)^{\frac{1}{2}}} \tag{15}$$

where T_p is a number of peak flows greater than the peak flow threshold (i.e. 33.3% of the exceedance probability of observed flows for the 23 October 2006 to 31 December 2009 period); while T_l is a number of low flows lower than the low flow threshold (i.e. 66.6% of the exceedance probability of observed flows for the 23 October to 31 December 2009 period); and $x_{obs,i}$ and $x_{sim,i}$ are observed and the forecasted streamflow for different lead time i, respectively. It should be noted that if the forecasted streamflow is higher than the peak flow threshold or lower than the low flow threshold criteria, the PFC and LFC produce no values. This particularly occurs for shorter lead time forecasts.

3. Results and discussion

This section presents three analyses: the evaluation of (1) soil moisture and streamflow estimates, (2) variations of soil moisture for upper and lower soil layers using the three approaches of EnKF described in Section 2.4 and (3) streamflow and soil moisture in a forecast mode. The first analysis focuses on evaluating soil moisture and flow variations and the accuracy of soil moisture and streamflow estimates; in the second analysis, we further investigate how soil moisture vary from upper to lower soil layers when the three DA approaches are applied; and the third analysis focuses on the evaluation of streamfow and soil moisture forecast in different lead times when the DA techniques are used.

3.1. Evaluation of soil moisture and streamflow variations

Soil moisture and flow variations are examined for the three methods: (a) streamflow assimilation, (b) soil moisture assimilation, and (c) combined assimilation of streamflow and soil moisture. The methods were also evaluated against the open loop simulation (no data assimilation). This evaluation examines the impact of soil moisture on state variations, and on the estimation of streamflow and model soil moisture. The assimilation results based on plots of the observed and updated estimates of streamflow and soil moisture are presented in Fig. 4 (for streamflow) and 5 (for soil moisture), and the evaluation statistics are shown in Table 3. The time series of streamflow and soil moisture estimates obtained from the open loop simulation are not shown in Figs. 3 and 4 since they show similar results to those obtained from the streamflow assimilation method. The statistics of model evaluation of these two methods are also relatively similar (Table 3). The streamflow assimilation produces slightly lower RMSEs of flow estimates, particularly in Westover and Highway (Table 3). In Fig. 4, the results show that the estimation of streamflow using the streamflow assimilation, and the combined assimilation methods are superior to estimates from the soil moisture assimilation. The streamflow assimilation and combined assimilation of streamflow and soil moisture provided good results for streamflow estimation with the streamflow residuals averagely varying between −10 and 10 m^3/s in all basins (Fig. 4). On the other hand, the soil moisture assimilation approach has larger errors of streamflow than those of the other two methods. The residuals, for example in Dundas, can reach as high as 25 m^3/s. In this method, the simulated streamflow overestimated the observation in all three study basins (Fig. 4). The evaluation statistics (for the 2007–2009 period) show that the RMSEs of streamflow generated using the soil moisture assimilation are about 2–5 times larger, and the VEs of streamflow are about 2–10 times larger than those obtained from the streamflow assimilation and combined assimilation of streamflow and soil moisture in various basins (see Table 3, first and second rows, respectively). These results show that updating model parameters and soil moisture state in the model by assimilating soil moisture only provides inaccurate estimation of streamflow. In the soil moisture assimilation, soil moisture observations are the only source of observational information applied to update model parameters and states. This deteriorates the propagation of model parameters and states which control the estimated streamflow.

Similarly, the results in Fig. 5 shows that when the streamflow assimilation approach was used, soil moisture estimates deteriorate significantly. The residuals of soil moisture were averagely varying between −50 and +80 mm in Westover (Fig. 5a) and −50 and +250 mm in Highway and +60 and +280 mm in Dundas (Fig. 5b and c, respectively). The RMSEs of soil moisture generated using the streamflow assimilation are 4–6 times larger than those obtained from the soil moisture assimilation and combined assimilations in various basins (Table 3, third rows). This result indicates that states were distorted in favor of assimilating streamflow. On the other hand, the soil moisture assimilation and combined assimilation methods produce more acceptable results. The

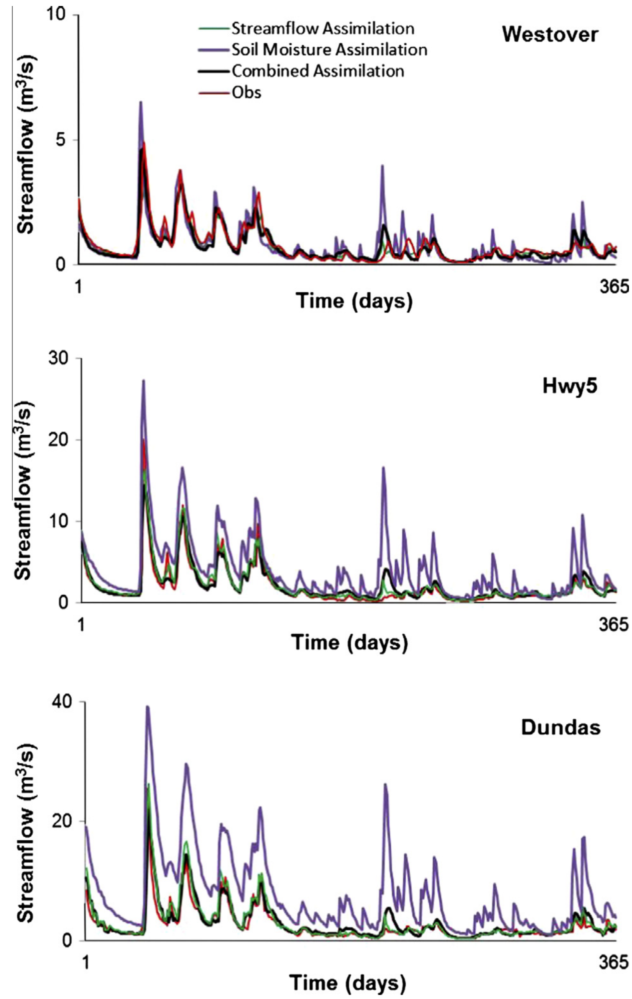

Fig. 4. Observed and mean ensemble of one-day ahead streamflow forecasting of the SAC-SMA in Westover (top figure), Highway-5 (middle) and Dundas (bottom) (water year 2009) generated using assimilation of streamflow, soil moisture and combined assimilation of streamflow and soil moisture.

average residuals are in the ranges between −10 and +7 mm. In addition, the 95% uncertainty bounds of soil moisture estimations are consistent with the observations (later discussed in Fig. 6).

In summary, the results in Figs. 4 and 5 show that in the streamflow assimilation, the soil moisture estimates were markedly distorted; whereas, in the soil moisture assimilation, streamflow estimates were distorted. The use of soil moisture or streamflow observational data alone in DA have direct impact on model parameterization and states, and subsequently on lowering the accuracy of streamflow or soil moisture estimates, respectively. However, this problem can be managed by using the combined assimilation of streamflow and soil moisture. The approach has the flexibility to account for model adjustment through the time variation of parameters together with state variables when both soil moisture and streamflow information were integrated into the system. This combined approach can produce good results for both soil moisture and streamflow estimates. As shown in Table 3 (the last three columns), when the combined assimilation of streamflow and soil moisture approach is used, both streamflow and soil moisture estimates have small errors (about averagely 7 m^3/s for RMSE of streamflow and 30 mm/day for RMSE of soil moisture in the various basins).

The examples of the ensemble means of daily soil moisture storage and their 95% uncertainty bounds for each basin using the combined assimilation of streamflow and soil moisture

Table 3
Model performance statistics for streamflow and soil moisture (water year 2007–2009).

Model performances	EnKF-streamflow only			EnKF-soil moisture only			EnKF-combined soil moisture and streamflow			Open loop		
	Westover	Hwy-5	Dundas	Westover	Hwy-5	Dundas	Westover	Hwy-5	Dundas	Westover	Hwy-5	Dundas
RMSE of streamflow	0.37	0.69	0.82	0.96	1.69	3.17	0.56	0.75	0.87	0.46	0.74	0.69
VE of streamflow (m^3/s)	−68.24	188.74	596.10	−16.54	1791.21	4792.54	−57.28	149.88	625.59	−0.54	172.57	368.69
RMSE of soil moisture storage	60.99	94.93	125.91	13.77	15.06	20.33	14.98	37.73	36.05	63.01	117.21	78.59

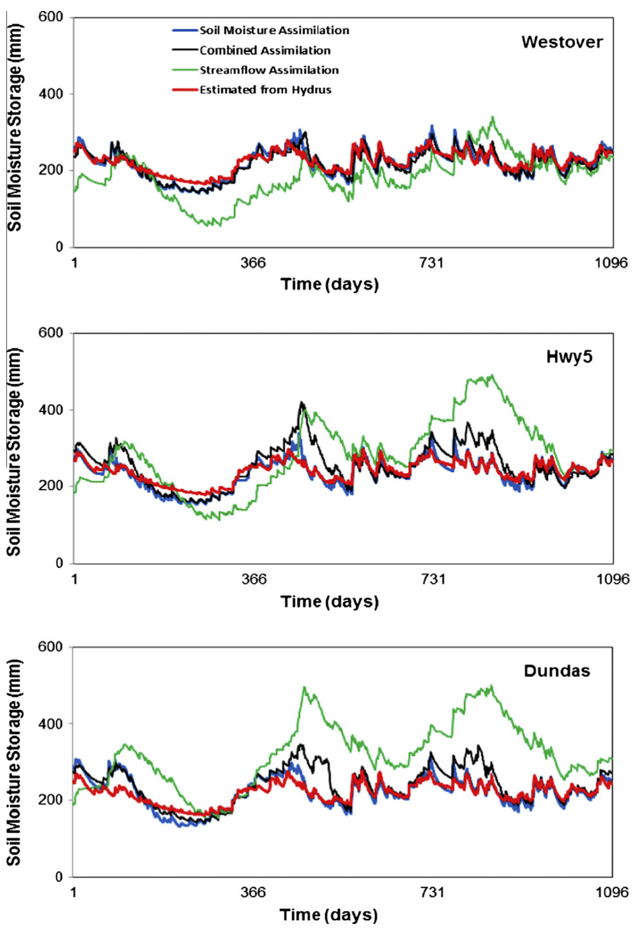

Fig. 5. Comparison of simulated soil moisture obtained from Hydrus and mean ensemble of one-day ahead soil moisture forecasting of the SAC-SMA in Westover (top figure), Highway-5 (middle) and Dundas (bottom) (water year 2009) generated using assimilation of streamflow, soil moisture and combined assimilation of streamflow and soil moisture.

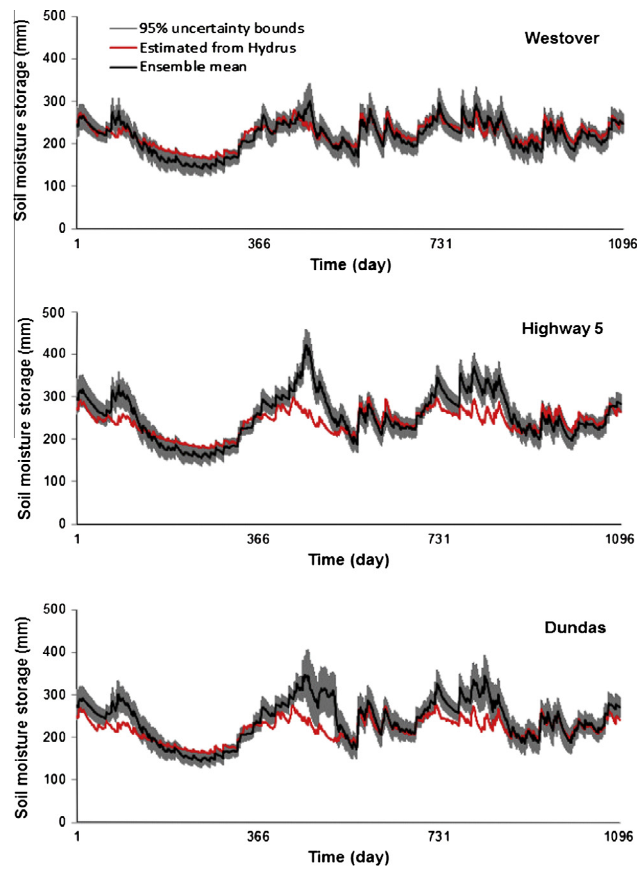

Fig. 6. Comparison of simulated soil moisture generated from Hydrus model, mean value and 95% uncertainty bounds of ensemble of soil moisture storages of the SAC-SMA in Westover (top figure), Highway-5 (middle) and Dundas (bottom) (water year 2007–2009) generated using combined assimilation of streamflow and soil moisture.

approach are plotted in Fig. 6. As shown in Fig. 6, the ensemble mean of daily soil moisture storage and their 95% uncertainty bounds are consistent with the observations for each basin. The observed soil moisture varies between upper and lower uncertainty bounds and captures quite well the observed soil moisture, particularly in Westover (Fig. 6a). This is particularly due to lesser complexity and variations of soil distributions in Westover than in the other sub-basins (see Fig. 2). Overall, the simulated soil moisture storage of the SAC-SMA model follows the variability of soil moisture generated by the Hydrus model in each sub-basin (Fig. 6) relatively well. These results are important for the application of DA methods to simultaneously estimate soil moisture states and watershed response.

The above analyses highlight two important findings of this study. The first is the importance of integrating two observation data sets (i.e. soil moisture and streamflow observation) in the data assimilation procedure to improve flow and soil moisture estimates. By assimilating both soil moisture and streamflow, based on the combined assimilation procedure, the imperfect model outputs are merged with the updated model states to better fit these external observations, and thus reduce the model errors. The second is the importance of the use of adequate data/information in DA to improve the accuracy of model outputs. As shown in the results above, when only streamflow observation was used, the soil moisture estimates degrade. On the other hand, the streamflow estimates degrade when only soil moisture observation was used. Assimilating both soil moisture and streamflow appears a better option in reducing model errors.

3.2. Variations of soil moisture storages of upper and lower soil layers

A recurring objective for most soil moisture assimilation procedures is to extend the forecast skill of surface soil moisture to deeper soil layers. Further analysis was performed to evaluate variations of soil moisture in the different soil layers, i.e. upper and lower soil layers when the three DA approaches were used. As shown earlier in Fig. 5 and Table 3, the combined assimilation and the soil moisture assimilations appear more consistent with "observations" compared to streamflow only assimilation results. The upper soil layer (Fig. 7, top panels) generally exhibits higher variability of soil moisture than the lower soil layers (Fig. 8, bottom panels). This is due to its role to generate fast flow to supply water to the lower zone and to re-supply water to tension water content of upper soil layer. Fig. 7, in general, shows that the variation, magnitude and timing of soil moisture storages of upper soil layer are relatively similar among the three methods and in Westover, Highway 5 and Dundas basins (Fig. 7a–c, top panels). These results indicate that there are no significant differences in soil moisture estimates for the upper soil layer, whatever the data assimilation procedure was used. Slight differences of soil moisture variations of the upper soil layer obtained from the combined assimilation, are found among three study basins. Highway-5 produces slightly higher soil moisture estimates of the upper soil layer than in Dundas and Westover, and consequently slightly higher "quick runoff". This possibly causes slightly higher annual peak flows in

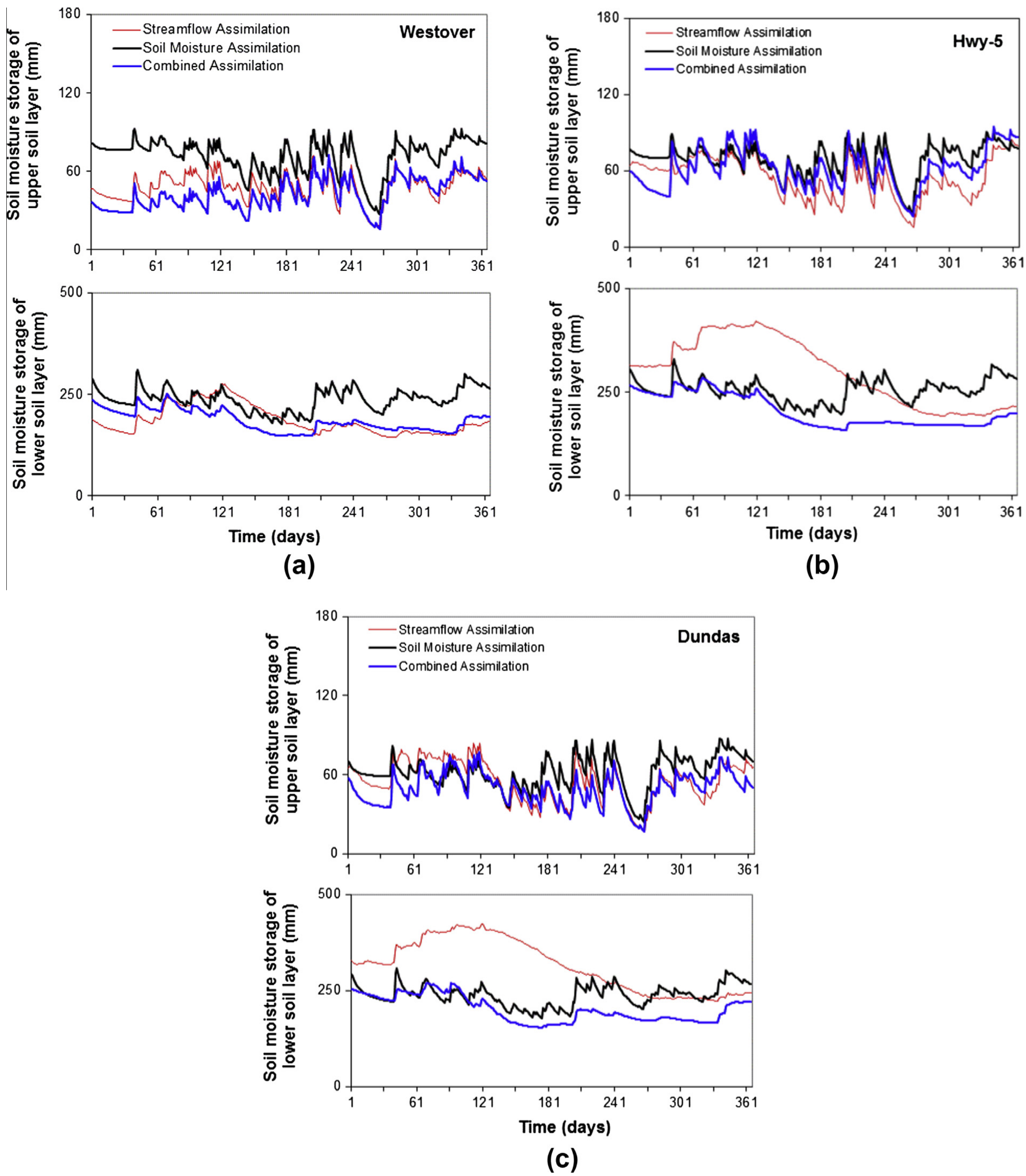

Fig. 7. Soil moisture storages of upper soil layer (top panel) and lower soil layer (bottom panel) for (a) Westover, (b) Highway 5, and (c) Dundas (water year 2009) generated using streamflow assimilation, soil moisture assimilation and combined assimilation of streamflow and soil moisture.

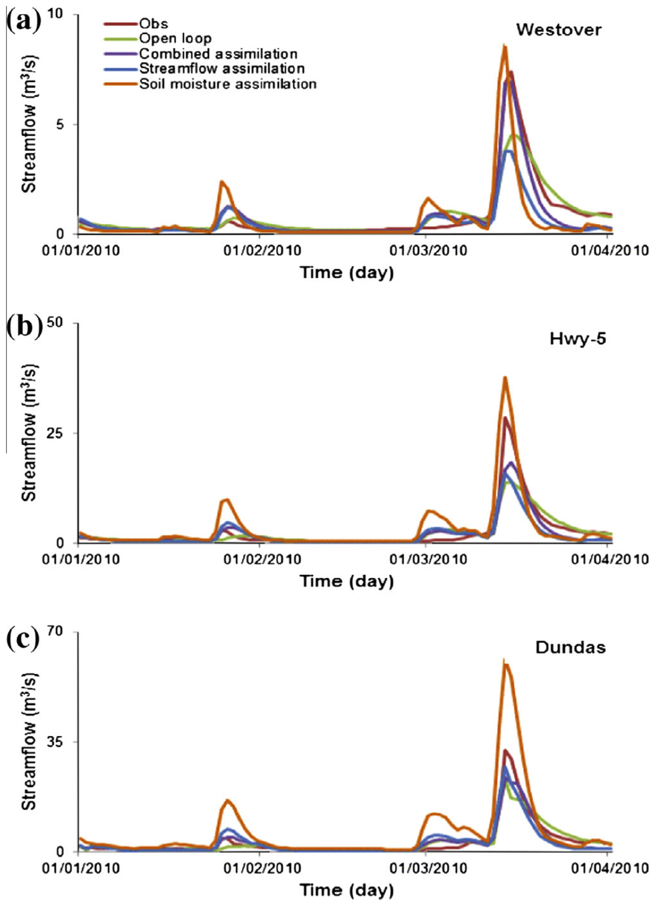

Fig. 8. Streamflow variability comparing those obtained from the open loop simulation, combined streamflow and soil moisture assimilation and streamflow assimilation in the forecast mode for (a) Westover, (b) Highway 5, and (c) Dundas. The initial condition of the forecast is started on Dec 31th, 2009.

Highway-5 than in Dundas and Westover (as described earlier in Section 2.1 and shown in Table 1).

In contrast, there are large differences of soil moisture storages of lower soil layers between the soil moisture assimilation, the combined assimilation, and the streamflow assimilation approach, particularly in Highway-5 and Dundas (Fig. 7b and c, bottom panels). It shows that the accuracy of soil moisture estimates degrades from surface to deeper soil layers when streamflow assimilation is used. This may be, in part, due to the differences in the change of volume of water moving vertically from upper to deeper soil layers (percolation) and the time scale differences between soil moisture and streamflow. Due to slower residence time, the lower zone moisture assimilation would not be that helpful in streamflow forecasting. However, further study is required to investigate the main causes. This may include the uses of additional remotely sensed soil moisture products and soil moisture observations for different soil depths (shallow and deeper soil depths) in the assimilation techniques. Parajka et al. (2009), for example, found that the use of additional ERS (European Remote Sensing) scatterometer data in model calibration reduces the parameter uncertainty of the top soil moisture of the dual soil moisture layers of a lumped hydrological model in the Austrian catchments and provides a more robust water balance simulations. Further study may also include the analysis of the time lag in the assimilation of streamflow observation to update soil moisture state. This analysis may help to better explore the change of volume of water moving vertically from upper to deeper soil layers because the SAC-SMA model was modified to include reservoirs in the system which may delay the streamflow generation. Overall, these results suggest that the use of both soil moisture and streamflow data in data assimilation could produce not only good estimates of both streamflow and soil moisture but also better represent the relationship between soil drainage characteristics and soil moisture conditions.

3.3. Evaluation of streamflow and soil moisture in a forecast mode

This section presents the variations of streamflow and soil moisture in a forecast mode for different lead times (1, 2, 7, 14, 30, 60 and 90 days). The examples of the streamflow and soil moisture forecasts with the forecast started at January 1st, 2010 are presented in Figs. 8 and 9, respectively. In general, the streamflow forecasts obtained from the soil moisture assimilation method shows larger errors than those obtained from the other DA methods and open loop in three study sub-basins (Fig. 8). This result is consistent with our previous findings that a large distortion of streamflow variations is found when soil moisture assimilation is used (Section 3.1). In addition, the combined soil moisture and streamflow assimilation generally forecasts better peak and low flows than those of the open loop simulation and streamflow assimilation (Fig. 8). Visually, the rising limbs generated from the streamflow assimilation, open loop and combined assimilation are relatively similar and fit the observation (Fig. 8), but the recession periods of the open loop simulation are relatively slower and longer than the other two data assimilations. On the other hand, the soil moisture assimilation shows better soil moisture forecasts, particularly for shorter lead times (1–14 days) than the other DA

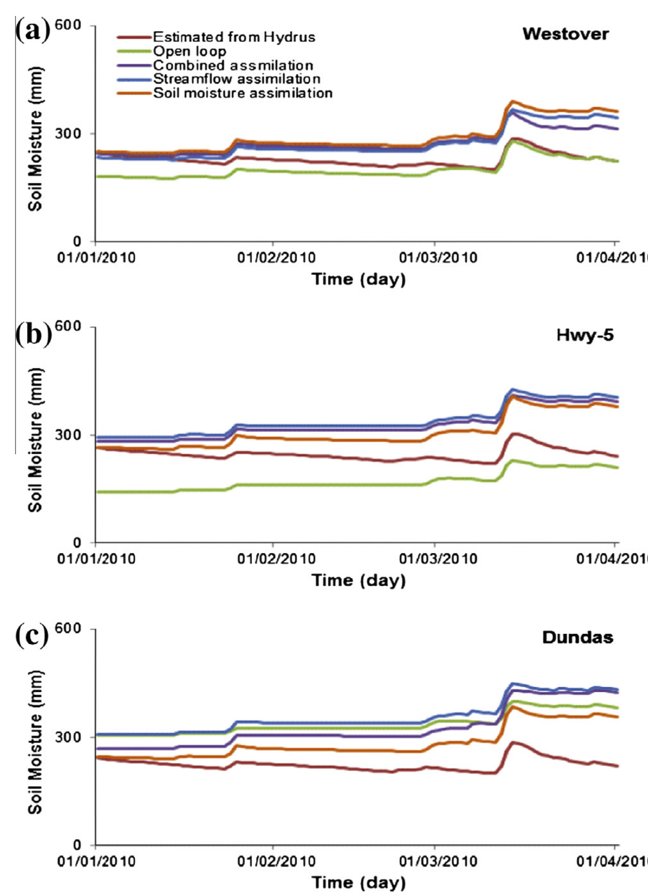

Fig. 9. Soil moisture variability comparing those obtained from the open loop simulation, combined streamflow and soil moisture assimilation and streamflow assimilation in the forecast mode for (a) Westover, (b) Highway 5, and (c) Dundas. The initial condition of the forecast is started on December 31th, 2009.

methods and the open loop simulation in each sub-basin (Fig. 9). The forecasting accuracy of soil moisture estimates decreases for longer lead times for all DA methods. The possible reasons for this decreasing accuracy are discussed in the following paragraphs.

The detailed statistics of model performances for every single date in the evaluation period (i.e. from January 1, 2010 to March 31, 2010) and for each lead time (1, 2, 7, 14, 30, 60 and 90 days) are summarized in Fig. 10. For simplicity, only the median of 90 values of each model criterion during the evaluation period is shown in this figure. The combined soil moisture and streamflow assimilation forecasts better peak and low flows (i.e. the PFCs and LFCs are lower, respectively) than those of the open loop

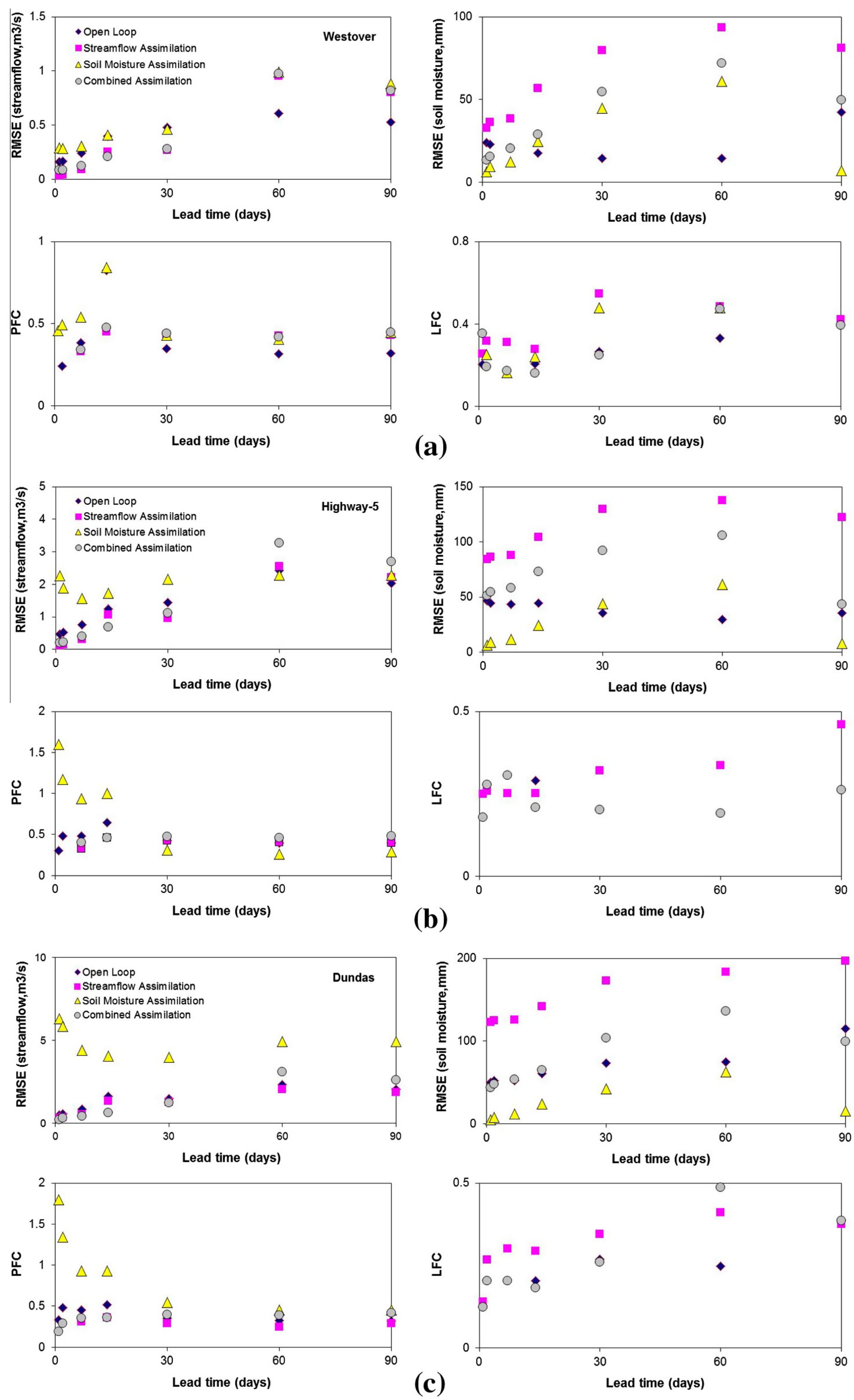

Fig. 10. Median of 90 values of RMSEs of streamflow (top left), RMSEs of soil moisture (top right), Peak Flow Criteria index (bottom left) and Low Flow Criteria index (bottom right) in a forecast mode for different lead times in (a) Westover, (b) Highway 5 and (c) Dundas.

simulation and the other DA methods, particularly for shorter lead times (1–14 days) (Fig. 10). It should be noted that the lower the PFC or LFC value, the closer the model to a perfect fit. As discussed earlier and shown in Fig. 8, the recession periods of the open loop simulation are relatively slower and longer than the streamflow and combined assimilations. This possibly causes worse LFC for the open loop simulation. However, further study is needed to thoroughly explore the different assimilation approaches on these hydrologic processes. Overall, the combined assimilation appears to be capable of accounting for these different extreme phenomena (peak and low flows) better than the other methods.

In addition, the results from the model RMSE statistics indicate that the combined assimilation improves streamflow accuracy from the open loop simulation, and streamflow and soil moisture data assimilations, particularly for shorter lead times (1–14 days, Fig. 10, top left figures). These values slowly diverge to higher values for longer lead times (30–90 days). This is due to the fact that the influence of system and input noise is not filtered out during the forecast. On the other hand, the soil moisture assimilation method shows better model performances for soil moisture forecasts than the other models (Fig. 10, top right figures). It is partly because the soil moisture assimilation model subsequently updates the model parameters and model states to better fit the soil moisture observations, but it leads to degrade streamflow forecasts. Furthermore, the combined assimilation forecasts lower model errors of soil moisture forecasts than the streamflow assimilation and open loop, particularly for shorter lead times. Overall, the combined assimilation produces better both soil moisture and flow forecasts compared to other methods. One possible reason is because soil type information is generally not a direct input to a lumped hydrological model. The use of soil moisture information in the combined assimilation allows the model to reflect the soil type information in the time variation of model parameters and state variables. As a result, the resulting combined assimilation model can produce more accurate soil moisture and flow forecasts, particularly for shorter lead times.

4. Summary and conclusions

Data assimilation has been widely used in hydrological models to improve soil moisture model state and subsequent streamflow estimates. However, due to data unavailability or poor quality of soil moisture data, the state variables are not usually evaluated in hydrological DA particularly, for streamflow assimilation. For example, assimilation of streamflow can produce accurate streamflow estimates; however, the variations of soil moisture could not be accurately estimated. In this study, using observed streamflow and/or soil moisture data in DA, model state variations have been evaluated in assimilations of (1) streamflow, (2) soil moisture and (3) combined assimilation of streamflow and soil moisture. In the first two experiments, only streamflow and only soil moisture observations were used to drive the assimilation, whereas in the third experiment, both soil moisture and streamflow observations were used. As illustrated in this study, there are large differences in soil moisture variations when states and model parameters were updated by assimilations of streamflow, soil moisture, and combined streamflow and soil moisture. In streamflow assimilation, the soil moisture estimates were markedly distorted, especially on the soil moisture storage of lower soil layers. That is, model states are distorted in favor of estimating streamflow. Whereas, in soil moisture assimilation, streamflow estimates were distorted. In this case, streamflow estimates were inaccurate as model parameters were compromised in assimilating soil moisture. Finally, the use of combined streamflow and soil moisture observations to drive the assimilation provide more accurate estimates of both soil moisture and streamflow. This combined approach has the flexibility to account for model adjustment through the time variation of parameters together with state variables when soil moisture and streamflow observations were jointly integrated into the system. The model can balance different sources of model inputs and outputs (i.e. observed and simulated soil moisture and streamflow) of the hydrological model and thus balance the system observations and model outputs. These evaluations are important for the application of DA methods to simultaneously estimate soil moisture states and watershed response. However, this study is limited to using the soil moisture and observed streamflow data with EnKF assimilation technique. Further study should investigate various DA methods, especially in light of recent developments in using the variable variance multipliers (VVM) to avoid the subjectivity in perturbing the model variables (Leisenring and Moradkhani, 2012) and the data assimilation using the particle filter combined with Markov Chain Monte Carlo method which has been shown to significantly improve the effectiveness and accuracy of hydrologic data assimilation (Moradkhani et al., 2012).

Acknowledgements

This research was supported by a grant from the Canadian Foundation for Innovation (CFI) and the Ontario Ministry of Research and Innovation (OMRI). The authors acknowledge Hamilton Region Conservation Authority and Environment Canada for providing the hydrometeorological data for this study.

References

Andreadis, K.M., Lettenmaier, D.P., 2006. Assimilating remotely sensed snow observations into a macroscale hydrology model. Adv. Water Resour. 29, 872–886.

Aubert, D., Loumagne, C., Oudin, L., 2003. Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model. J. Hydrol. 280, 145–161.

Berg, A.A., Mulroy, K., 2006. Streamflow predictability given macro-scale estimates of the initial soil moisture status. HSJ 51 (4), 642–654.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., Hasenauer, S., 2010. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. Hydrol. Earth Syst. Sci. 14 (10), 1881–1893.

Burnash, R.J.C., 1995. The NWS River Forecast System—Catchment Modeling. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Littleton, CO, pp. 311–366.

Burnash, R.J.C., Ferral, R.L., McGuire R.A., 1973. A generalized streamflow simulation system: conceptual models for digital computers. National Weather Service, NOAA, and the State of California Department of Water Resources Tech. Rep., Joint Federal–State River Forecast Center, Sacramento, California, p. 68.

Clark, M.P., Rupp, D.E., Woods, R.A., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J., Uddstrom, M.J., 2008. Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model. Adv. Water Resour. 31, 1309–1324.

Coulibaly, P., Anctil, F., Bobée, B., 2001. Multivariate reservoir inflow forecasting using temporal neural networks. J. Hydrol. Eng. 6 (5), 367–376.

Crow, W.T., Ryu, D., 2009. A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. Hydrol. Earth. Syst. Sci. 13, 1–16.

Crow, W.T., Van Loon, E., 2006. Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. J. Hydrometeor. 7, 421–432.

DeChant, C.M., Moradkhani, H., 2011a. Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. Hydrol. Earth Syst. Sci. 15, 3399–3410. http://dx.doi.org/10.5194/hess-15-3399.

DeChant, C.M., Moradkhani, H., 2011b. Radiance data assimilation for operational snow and streamflow forecasting using the particle filter. Adv. Water Resour. 34 (3), 351–364.

Dong, J., Walker, J.P., Houser, P.R., Sun, C., 2007. Scanning multichannel microwave radiometer snow water equivalent assimilation. J. Geophys. Res. D: Atmos. 112 (D7).

Dumedah, G., Coulibaly, P., 2012a. Integration of evolutionary based assimilation into Kalman-type methods for streamflow simulations in ungauged watersheds. J. Hydrol. 475, 428–440. http://dx.doi.org/10.1016/j.jhydrol.2012.10.033.

Dumedah, G., Coulibaly, P., 2012b. Evolutionary assimilation of streamflow in distributed hydrologic modeling using in-situ soil moisture data. Adv. Water Resour. 53, 231–241. http://dx.doi.org/10.1016/j.advwatres.2012.07.012.

Durand, M., Margulis, S.A., 2006. Feasibility test of multifrequency radiometric data assimilation to estimate snow water equivalent. J. Hydrometeorol. 7, 443–457.

El Serafy, G.Y.H., Mynett, A.E., 2008. Improving the operational forecasting system of the stratified flow in Osaka Bay using an ensemble Kalman filter–based steady state Kalman filter. Water Resour. Res. 44, W06416. http://dx.doi.org/10.1029/2006WR005412.

Evensen, G., 1992. Using the extended Kalman filter with a multi-layer quasi-geostrophic ocean model. J. Geophys. Res. 97, 17905–17924.

Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean. Dynam. 53, 343–367.

Houser, P.R., Shuttleworth, W.J., Famiglietti, J.S., Gupta, H.V., Shed, K.H., Goodrich, D.C., 1998. Integration of soil moisture remotes sending and hydrologic modeling using data assimilation. Water Resour. Res. 34 (12), 3405–3420.

HRCA (Hamilton Region Conservation Authority) – MacLaren Plansearch Lavalin, 1990. Volume 1-Technical Report: Canada/Ontario Flood Damage Reduction Program. Hamilton, Ontario, Canada.

James, W., 1994. Current Practices in Modeling the Management of Stormwater Impacts. CRC Press, Hamilton, Ontario, Canada, p. 91.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82 (1), 35–45.

Kitanidis, P.K., Bras, R.L., 1980. Real-time forecasting with a conceptual hydrologic model. 2. Application and results. Water Resour. Res. 16 (6), 1034–1044.

Koren, V., Reed, S., Smith, M., Zhang, Z., Seo, D.J., 2004. Hydrology laboratory research modeling system (HL-RMS) of the US national weather service. J. Hydrol. 291, 297–318.

Kornelsen, K., Coulibaly, P., 2013. McMaster Mesonet soil moisture dataset: description and spatio-temporal variability analysis. Hydrol. Earth Syst. Sci. 17, 1589–1606. http://dx.doi.org/10.5194/hess-17-1589-2013.

Lee, H., Seo, D.J., Koren, V., 2011. Assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic model: effects of uncertainities in the data and initial model soil moisture states. Adv. Water Resour. 34, 1597–1615.

Legates, D.R., Mahmood, R., Levia, D.F., DeLiberty, T.L., Quiring, S.M., Houser, C., Nelson, F.E., 2010. Soil moisture: a central and unifying theme in physical geography. Prog. Phys. Geogr. 34 (6). http://dx.doi.org/10.1177/0309133310386514.

Leisenring, M., Moradkhani, H., 2011. Snow water equivalent estimation using Bayesian data assimilation methods. Stoch. Env. Res. Risk Assess. 253–270. http://dx.doi.org/10.1007/s00477-010-0445-5.

Leisenring, M., Moradkhani, H., 2012. Analyzing the uncertainty of suspended sediment load prediction using sequential Monte Carlo methods. J. Hydrol. 468–469, 268–282. http://dx.doi.org/10.1016/j.jhydrol.2012.08.049.

Liang, S., 2004. Quantitative Remote Sensing of Land Surfaces. Wiley, Hoboken, NJ.

Liu, Y., Weerts, A.H., Clark, M., Hendricks Franssen, H.J., Kumar, S., Moradkhani, H., Seo, D.J., Schwanenberg, D., Smith, P., van Dijk, A.I.J.M., van Velzen, N., He, M., Lee, H., Noh, S.J., Rakovec, O., Restrepo, P., 2012. Toward advancing data assimilation in operational hydrologic forecasting and water resources management: current status, challenges, and emerging opportunities. Hydrol. Earth Syst. Sci. 16, 3863–3887.

McMillan, H.K., Hreinsson, E.O., Clark, M.P., Singh, S.K., Zammit, C., Uddstrom, M.J., 2013. Operational hydrological data assimilation with the recursive ensemble Kalman filter. Hydrol. Earth Syst. Sci. 17, 21–38.

Monsivais-Huertero, A., Graham, W.D., Judge, J., Agrawal, D., 2010. Effect of simultaneous state–parameter estimation and forcing uncertainties on root-zone soil moisture for dynamic vegetation using EnKF. Adv. Water Resour. 33, 468–484.

Montzka, C., Moradkhani, H., Weihermuller, L., Canty, M., Hendricks Franssen, H.J., Vereecken, H., 2011. Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. J. Hydrol. 399, 410–421.

Moradkhani, H., 2008. Hydrologic remote sensing and land surface data assimilation. Sensors 8, 2986–3004.

Moradkhani, H., Meskele, T., 2009. Probabilistic assessment of the satellite rainfall retrieval error translation to hydrologic response, in satellite applications for surface hydrology, Springer. Water Sci. Technol. Library. http://dx.doi.org/10.1007/978-90-481-2915-7, 229–242.

Moradkhani, H., Sorooshian, S., 2008. General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis, in hydrological modeling and water cycle, coupling of the atmospheric and hydrological models, Springer. Water Sci. Technol. Libr. 63, 1–23.

Moradkhani, H., Hsu, K., Gupta, H.V., Sorooshian, S., 2005a. Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using particle filter. Water Resour. Res. 41, W05012. http://dx.doi.org/10.1029/2004WR003604.

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

Moradkhani, H., DeChant, C.M., Sorooshian, S., 2012. Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. Water Resour. Res. 48, W12520. http://dx.doi.org/10.1029/2012WR012144.

Parajka, J., Naeimi, V., Blöschl, G., Komma, J., 2009. Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria. Hydrol. Earth Syst. Sci. 13, 259–271.

Parrish, M.A., Moradkhani, H., DeChant, C.M., 2012. Toward reduction of model uncertainty: integration of Bayesian model averaging and data assimilation. Water Resour. Res. 48, W03519. http://dx.doi.org/10.1029/2011WR011116.

Payan, J.L., Perrin, C., Andreassian, V., Michel, C., 2008. How can man-made water reservoirs be accounted for in a lumped rainfall-runoff model? Water Resour. Res. 44, W03420. http://dx.doi.org/10.1029/2007WR005971.

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

Reichle, R.H., Crow, W.T., Koster, R.D., Sharif, H.O., Mahanama, S.P.P., 2008a. Contribution of soil moisture retrievals to land data assimilation products. Geophys. Res. Lett. 35 (L01404). http://dx.doi.org/10.1029/2007GL031986.

Reichle, R.H., Crow, W.T., Keppenne, C.L., 2008b. An adaptive ensemble Kalman filter for soil moisture data assimilation. Water Resour. Res. 44, W03423. http://dx.doi.org/10.1029/2007WR006357.

Salamon, P., Feyen, L., 2010. Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation. Water Resour. Res. 46. http://dx.doi.org/10.1029/2009WR009022.

Samuel, J.M., Coulibaly, P., Metcalfe, R., 2011. Estimation of continuous streamflow in Ontario ungauged basins: comparison of regionalization methods. J. Hydrol. Eng. – ASCE 16 (5), 447–459.

Schaake, J.C., Duan, Q., Koren, V., Mitchell, K.E., Houser, P.R., Wood, E.F., Robock, A., Lettenmaier, D.P., Lohmann, D., Cosgrove, B., Sheffield, J., Luo, L., Higgins, R.W., Pinker, R.T., Tarpley, J.D., 2004. An intercomparison of soil moisture fields in the North American Land Data Assimilation System (NLDAS). J. Geophys. Res. 109, D01S90, doi: http://dx.doi.org/10.1029/2002JD003309.

Schoups, G., Vrugt, J.A., Fenicia, F., van de Giesen, N.C., 2010. Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models. Water Resour. Res. 46. http://dx.doi.org/10.1029/2009WR008648.

Seo, D.J., Koren, V., Cajina, N., 2003. Real-time variational assimilation of hydrologic and hydrometeorolical data in to operational hydrologic forecasting. J. Hydromet. 4, 627–641.

Seo, D.J., Cajina, L., Corby, R., Howieson, T., 2009. Automatic state updating for operational streamflow forecasting via variational data assimilation. J. Hydrol. 367, 255–275.

Šimůnek, J., van Genuchten, M.Th., 2008. Modeling nonequilibrium flow and transport with HYDRUS. Vadose Zone J.. http://dx.doi.org/10.2136/VZJ2007.0074, Special Issue "Vadose Zone Modeling", 7(2), 782–797.

Šimůnek, J., Šejna, M., Saito H., Sakai, M., van Genuchten, M. Th., 2012. The HYDRUS-1D software package for simulating the movement of water, heat, and multiple solutes in variably saturated media, Version 4.15. HYDRUS Software Series 3, Department of Environmental Sciences, University of California Riverside, Riverside, California, USA, p. 338.

Slater, A.G., Clark, M.P., 2006. Snow data assimilation via an ensemble Kalman filter. J. Hydrometeorol. 7, 478–493.

SMHI, 2005. Integrated hydrological modelling system (IHMS). Manual version 5.8. Norrkoping, Sweden.

Sultana, Z., Coulibaly, P., 2011. Distributed modeling of future changes in hydrological processes of Spencer Creek watershed. Hydrol. Process. 25 (8), 1254–1270.

Thirel, G., Martin, E., Mahfouf, J.F., Massart, S., Ricci, S., Habets, F., 2010a. A past discharges assimilation system for ensemble streamflow forecasts over France Part 1: description and validation of the assimilation system. Hydrol. Earth Syst. Sci. 14 (8), 1623–1637.

Thirel, G., Martin, E., Mahfouf, J.F., Massart, S., Ricci, S., Regimbeau, F., Habets, F., 2010b. A past discharge assimilation system for ensemble streamflow forecasts over France Part 2: impact on the ensemble stream-flow forecasts. Hydrol. Earth Syst. Sci. 14 (8), 1639–1653.

Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. Geogr. Rev. 38 (1), 55–94.

Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. Water Resour. Res. 41, W01017. http://dx.doi.org/10.1029/2004WR003059.

Vrugt, J.A., Gupta, H.V., Nualláin, B.Ó., Bouten, W., 2006. Real-time data assimilation for operation ensemble streamflow forecasting. J. Hydrometeorol. 7 (3), 548–565. http://dx.doi.org/10.1175/JHM504.1.

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

Weerts, A.H., ElSerafy, G., Hummel, S., Dhondia, J., Gerritsen, H., 2010. Application ofgeneric data assimilation tools (DATools) for flood forecasting purposes. Comput. Geosci. 26, 453–463.