Geological and Atmospheric Sciences Publications   Geological and Atmospheric Sciences

3-14-2012

An integrated uncertainty and ensemble-based data assimilation approach for improved operational streamflow predictions

M. He
University of California, Los Angeles

T. S. Hogue
University of California, Los Angeles

S. A. Margulis
University of California, Los Angeles

Kristie J. Franz
Iowa State University, kfranz@iastate.edu

An integrated uncertainty and ensemble-based data assimilation approach for improved operational streamflow predictions

Abstract

The current study proposes an integrated uncertainty and ensemble-based data assimilation framework (ICEA) and evaluates its viability in providing operational streamflow predictions via assimilating snow water equivalent (SWE) data. This step-wise framework applies a parameter uncertainty analysis algorithm (ISURF) to identify the uncertainty structure of sensitive model parameters, which is subsequently formulated into an Ensemble Kalman Filter (EnKF) to generate updated snow states for streamflow prediction. The framework is coupled to the US National Weather Service (NWS) snow and rainfall-runoff models. Its applicability is demonstrated for an operational basin of a western River Forecast Center (RFC) of the NWS. Performance of the framework is evaluated against existing operational baseline (RFC predictions), the stand-alone ISURF and the stand-alone EnKF. Results indicate that the ensemble-mean prediction of ICEA considerably outperforms predictions from the other three scenarios investigated, particularly in the context of predicting high flows (top 5th percentile). The ICEA streamflow ensemble predictions capture the variability of the observed streamflow well, however the ensemble is not wide enough to consistently contain the range of streamflow observations in the study basin. Our findings indicate that the ICEA has the potential to supplement the current operational (deterministic) forecasting method in terms of providing improved single-valued (e.g., ensemble mean) streamflow predictions as well as meaningful ensemble predictions.

Hydrology and Earth System Sciences

An integrated uncertainty and ensemble-based data assimilation approach for improved operational streamflow predictions

M. He[1,*,**], T. S. Hogue[1], S. A. Margulis[1], and K. J. Franz[2]

[1]Department of Civil and Environmental Engineering, University of California, Los Angeles, California, USA
[2]Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa, USA
[*]currently at: Office of Hydrologic Development, NOAA National Weather Service, Silver Spring, Maryland, USA
[**]currently at: Riverside Technology, Inc., Fort Collins, Colorado, USA

Correspondence to: T. S. Hogue (thogue@seas.ucla.edu)

Abstract. The current study proposes an integrated uncertainty and ensemble-based data assimilation framework (ICEA) and evaluates its viability in providing operational streamflow predictions via assimilating snow water equivalent (SWE) data. This step-wise framework applies a parameter uncertainty analysis algorithm (ISURF) to identify the uncertainty structure of sensitive model parameters, which is subsequently formulated into an Ensemble Kalman Filter (EnKF) to generate updated snow states for streamflow prediction. The framework is coupled to the US National Weather Service (NWS) snow and rainfall-runoff models. Its applicability is demonstrated for an operational basin of a western River Forecast Center (RFC) of the NWS. Performance of the framework is evaluated against existing operational baseline (RFC predictions), the stand-alone ISURF and the stand-alone EnKF. Results indicate that the ensemble-mean prediction of ICEA considerably outperforms predictions from the other three scenarios investigated, particularly in the context of predicting high flows (top 5th percentile). The ICEA streamflow ensemble predictions capture the variability of the observed streamflow well, however the ensemble is not wide enough to consistently contain the range of streamflow observations in the study basin. Our findings indicate that the ICEA has the potential to supplement the current operational (deterministic) forecasting method in terms of providing improved single-valued (e.g., ensemble mean) streamflow predictions as well as meaningful ensemble predictions.

1 Introduction

Hydrologic forecasting is of primary importance in the context of flood and drought mitigation as well as for optimal water resources planning and management. The National Weather Service (NWS), the US agency responsible for short- and long-term hydrologic forecasting across the nation, primarily applies a joint snow accumulation and ablation model (i.e., SNOW17; Anderson, 1973) and rainfall-runoff model (i.e., SAC-SMA model; Burnash et al., 1973) for operational streamflow prediction in snow-dominated watersheds. For these watersheds, the accuracy of streamflow prediction largely relies on the accuracy of initial snowpack states (Clark and Hay, 2004; Franz et al., 2008). In the current NWS forecasting system, model initialization is conducted by running the calibrated joint model to the beginning of the forecast period. Subsequently, simulated states are manually updated based on the observed states. These state adjustments are typically not archived and are fairly subjective, varying significantly among forecasters who have different understanding and knowledge of the forecast basins, models, and prediction errors. Forecast accuracy thus varies across forecasters with different levels of forecasting skills (Seo et al., 2003). In addition, potential uncertainties in model forcing and parameters are typically ignored in the adjustment process. The non-systematic and manual procedures of the current model initialization method, as well as the need for more systematic uncertainty assessment motivate the development of a coupled automated uncertainty

analysis and data assimilation approach. The developed approach described herein includes the following capabilities: (1) treating errors in model forcing and parameters in a systematic and meaningful way, and (2) automatically merging state measurements into the operational model(s).

Over the past several decades, a variety of uncertainty analysis methods and data assimilation techniques have been developed and reported in the hydrologic literature. Some of the uncertainty analysis methods, among many others, include the generalized likelihood uncertainty estimation (GLUE) (Beven and Binley, 1992), the Bayesian total error analysis (BATEA) (Kavetski et al., 2002), the Integrated Bayesian Uncertainty Estimator (IBUNE) (Ajami et al., 2007), the Framework for Understanding Structural Errors (FUSE) (Clark et al., 2008b; Clark and Kavetski, 2010), and Markov Chain Monte Carlo (MCMC) methods including the Random Walk Metropolis algorithm (Kuczera and Parent, 1998), the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm (Vrugt et al., 2003), and the Differential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt et al., 2008). Most of above methods have not been commonly applied in practice, particularly for the NWS operational models SNOW17 and SAC-SMA. This is partly due to their high computational requirements, since efficiency is extremely important in hydrologic forecasting operations (Weerts and El Serafy, 2006). Most recently, He (2010) proposed an Integrated Sensitivity and UnceRtainty analysis Framework (ISURF). The ISURF first applies the Generalized Sensitivity Analysis (GSA) (Hornberger and Spear, 1981) method as a screening tool to identify sensitive model parameters. Subsequently, ISURF applies the DREAM algorithm (Vrugt et al., 2008) to explicitly quantify the uncertainty structure of these identified parameters. This step-wise method significantly reduces the computational load and has been demonstrated to efficiently and adequately identify the uncertainty structure of the SNOW17 and SAC-SMA parameters at a set of study sites with contrasting hydroclimatic conditions (He, 2010; He et al., 2011b).

Data assimilation techniques have been extensively applied in meteorological and ocean sciences. Currently existing data assimilation techniques were primarily developed for numerical weather prediction (Daley, 1993) and more recently applied in hydrologic fields including operational hydrologic forecasting. The earliest data assimilation approach applied to operational hydrologic models (i.e., SAC-SMA) is the extended Kalman filter (EKF) (Kitanidis and Bras, 1980a,b). Another data assimilation technique applied to an operational model (i.e., SAC-SMA) is the variational method (VAR) (Seo et al., 2003, 2009). A third data assimilation technique applied in hydrologic modeling is the Ensemble Kalman Filter (EnKF) (Evensen, 1994). The EnKF has been applied to both the SNOW17 model (Slater and Clark, 2006) and SAC-SMA model (Vrugt et al., 2006), as well as other hydrologic models including the TOPNET model (Clark et al., 2008a) and the HBV model (Weerts and El Serafy, 2006).

These data assimilation techniques have their own advantages and disadvantages (Liu and Gupta, 2007). The EKF is inherently associated with instabilities and divergence for models with strong non-linearities (Evensen, 1994). The VAR requires the development of the adjoint model which is complicated and labor-intensive (Margulis and Entekhabi, 2001). In comparison to the EKF and VAR, the EnKF does not require reformulation or modification of the original model, which is important in an operational setting. In addition, the EnKF provides more flexibility in explicitly handling various sources of uncertainty. This distinctive characteristic of the EnKF makes it a potentially robust technique for assimilation of a range of hydrologic data into operational forecasting. There are also other ensemble data assimilation algorithms including the Ensemble Kalman Smoother (e.g., Dunne and Entekhabi, 2006), and particle filter (e.g., Moradkhani et al., 2005a) which have similar strengths as the EnKF. However, previous studies have shown the EnKF is easy to implement and is more computationally efficient (Weerts and El Serafy, 2006).

In spite of the above-mentioned attempts to integrate data assimilation approaches into the NWS operational SNOW17 model (Slater and Clark, 2006) and SAC-SMA model (Kitanidis and Bras, 1980a,b; Seo et al., 2003, 2009), most of these studies assumed uncertainty-free model parameters and thus did not capitalize on the full capabilities of the EnKF to propagate parameter uncertainty and subsequently reduce forecasting uncertainty. Recent studies (e.g., Moradkhani et al., 2005a,b; Vrugt et al., 2006; Su et al., 2011) explored the capability of data assimilation techniques (i.e., EnKF and/or particle filter) in sequentially updating model parameters and states at measurement times. However, the traditional premise in operational hydrologic forecasting is that model parameters are time-invariant. Parameter uncertainty can be addressed by defining an ensemble set of parameters which approximately covers the uncertain parameter ranges. This can be achieved by perturbing parameters in the same way as the model forcing (Margulis et al., 2002).

The objectives of the current study are two-fold: (1) to explicitly quantify uncertainty in model parameters; and (2) to develop an ensemble-based data assimilation technique configured with derived parameter uncertainty and evaluate the applicability of this technique in improving NWS operational streamflow predictions. To achieve our objectives, we propose an Integrated unCertainty and Ensemble-based data Assimilation (ICEA) approach for the coupled SNOW17/SAC-SMA model of the NWS. The ICEA has two components, the ISURF framework (He, 2010) and an EnKF framework. The ISURF is applied to identify sensitive model parameters and the uncertainty structure of these parameters. Based on the parameter uncertainty structure identified along with preset uncertainty structure for model forcing, an EnKF framework is developed for the SNOW17 model to generate rain plus snowmelt (defined here as rain falling on bare ground plus melt out of the snowpack) via assimilation of snow water

Table 1. SNOW17 model parameters and state variables that are updated in the data assimilation applications. Parameter ranges are estimated from Anderson (1973) and He et al. (2011a,b).

Parameters	Explanation	Unit	Ranges
SCF	Snow fall correction factor	–	0.7–1.4
MFMAX	Maximum melt factor	mm/6 h/°C	0.5–2.0
MFMIN	Minimum melt factor	mm/6 h/°C	0.05–0.49
UADJ	The average wind function during rain-on-snow periods	mm/mb/°C	0.03–0.19
NMF	Maximum negative melt factor	mm/6 h/°C	0.05–0.50
MBASE	Base temperature for non-rain melt factor	°C	0–1.0
PXTEMP	Temperature that separates rain from snow	°C	−2.0–2.0
PLWHC	Percent of liquid-water capacity	–	0.02–0.3
DAYGM	Daily melt at snow-soil interface	mm/day	0–0.3
TIPM	Antecedent snow temperature index parameter	–	0.1–1.0
SI	Areal SWE above which all the area is covered by snow	mm	0–2000

State variables	Explanation	Unit
TWE	Snow water equivalent	mm
WE	Water equivalent of ice portion of snowpack	mm
LIQW	Amount of liquid water held against gravity drainage	mm
S	Amount of lagged excess liquid water in storage	mm
El	Average hourly lagged excess water for each time step	mm
PACKRO	Snowpack runoff	mm
ROBG	Runoff over bare ground	mm
NEGHS	Heat deficit	mm

equivalent (SWE) observations. The rain plus snowmelt serves as input to the SAC-SMA model to produce streamflow predictions. An operational watershed in California is used as the study basin to demonstrate the applicability of ICEA to the coupled SNOW17/SAC-SMA model in terms of providing streamflow predictions. Results are evaluated against the RFC predictions as well as that from a stand-alone EnKF.

2 Study area, datasets, and models

2.1 NWS operational models

The SNOW17 is a lumped process-based model that simulates snow accumulation and ablation processes (Anderson, 1973). The model requires mean areal precipitation (MAP) and mean air temperature (MAT) as inputs and simulates SWE and a rain plus snowmelt timeseries. Snow is modeled as a single layer in the model. The heat storage of the snowpack, liquid water retention and transmission, and snowmelt are computed using empirical functions, an areal depletion curve (ADC), and 11 parameters (Table 1). The model applies a snow correction factor (SCF (−)) to account for precipitation gage catch deficiencies. The actual snowfall input to the model, P_s (mm), is computed as:

$$P_s = SCF \times P \times f_s \tag{1}$$

where P (mm) represents the observed precipitation (MAP in this study); f_s (−) is fraction of snow in the precipitation and is determined as:

$$f_s = \begin{cases} 1, & \text{when } T_a < \text{PXTEMP} \\ 0, & \text{when } T_a \geq \text{PXTEMP} \end{cases} \tag{2}$$

where T_a (°C) is the observed air temperature (MAT in this study); PXTEMP is the threshold temperature to distinguish snowfall from rainfall in precipitation. Parameters SCF and PXTEMP jointly determine the snowfall input to the model for given precipitation and air temperature forcing.

The SAC-SMA model is a saturation excess rainfall-runoff model that simulates infiltration, percolation, soil moisture storage, drainage, and evapotranspiration processes (Burnash et al., 1973). Inputs to the model typically include MAP (or rain plus snowmelt when used in junction with the SNOW17 model) and potential evapotranspiration (PET). Outputs are estimated evapotranspiration and a basin-averaged runoff depth. A separate unit hydrograph is then applied to convert the runoff depth to streamflow. The model has 16 parameters, 13 of which must be estimated and three are generally set to default values (Hogue et al., 2000, 2006; Ajami et al., 2004).

2.2 Study area and datasets

The current study focuses on the North Fork of the America River Basin (NFARB) located on the western side of the
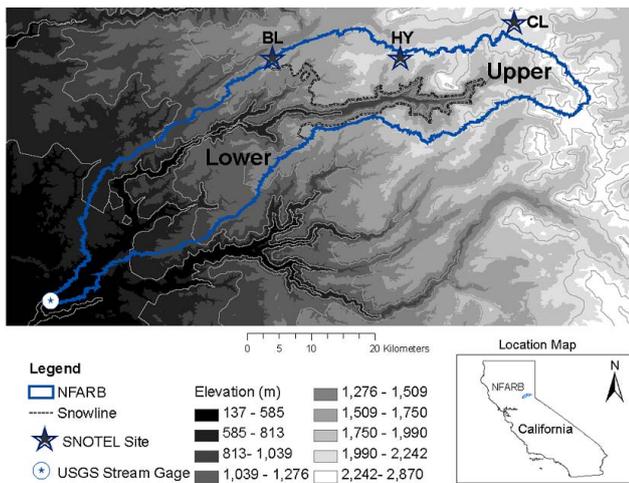
Fig. 1. Location and elevation ranges of the North Fork of the American River Basin (NFARB) as well as locations of streamflow and snow observational stations used in the current study.

Fig. 2. Climatic characteristics of the NFARB sub-basins during the entire study period (WY 1979–2002), training period (WY 1979–1984), and prediction period (WY 1991–1996). Mean annual temperature at various percentiles over the lower **(a)** and upper **(b)** sub-basins are highlighted, as well as the annual areal precipitation at various percentiles over the lower **(c)** and upper **(d)** sub-basins.

Sierra Nevada Mountains in northern California (Fig. 1). The basin is one of the study locations for the Distributed Model Intercomparison Project (DMIP) of the NWS (Koren et al., 2004; Reed et al., 2004; Smith et al., 2004). The NFARB drains into Lake Folsom with an area of 868 km². Elevation in the NFARB ranges from about 200 m at the basin outlet to about 2870 m at the highest boundary. The basin is characterized by deep winter snowpack (November to February) resulting from orographic precipitation processes. The snowpack melts out in the spring (March to June). The long-term annual precipitation and runoff are 1514 mm and 837 mm, respectively. The vegetation is elevation-dependent, with thin alpine tundra forest, dense mixed coniferous forest, and grassland chaparral and woodland species in the upper, medium, and lower elevations of the basin, respectively (Shamir and Georgakakos, 2006).

The snowline at the end of the accumulation period is estimated at around 1500 m. In the NWS modeling system, the NFARB is delineated into the upper sub-basin and lower sub-basin using the snowline as the divide. The upper sub-basin accounts for 37 % of the total basin area. The upper and lower sub-basins receive an annual average precipitation of about 1643 mm and 1397 mm, respectively. The long-term annual average temperatures for the upper and lower sub-basins are approximately 5.7 °C and 10.7 °C, respectively. There are three snow stations which provide daily SWE observations in the upper sub-basin (Fig. 1, Table 1). One of them (CS) is a SNOw TELemtry (SNOTEL) site maintained by the National Water and Climate Center (NRCS). The other two are maintained by the California Data Exchange Center (CDEC). Additional climatic and geographic characteristics of the upper and lower sub-basins are presented in Fig. 2.

Model forcing data, including the historical MAP, MAT, and PET for both upper and lower sub-basins, are avail-
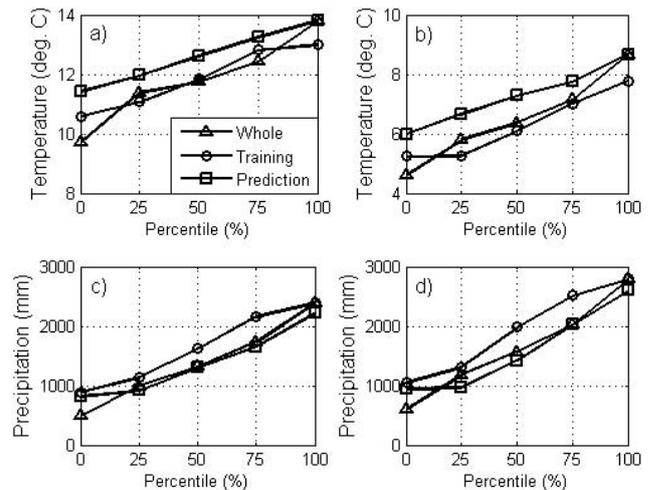
able at a 6-hourly time step from water year (WY) 1979–2002 (hereafter refer to as the "whole period"). The daily streamflow discharge data at the basin outlet (recoded at USGS gage # 11427000) and daily SWE at the snow stations are also available from WY 1979–2002 and WY 1991–2002, respectively. For demonstration purposes, a 6-year training period (WY 1979–1984) is selected to identify parameter uncertainty. An equivalent length of period is selected as the prediction period (WY 1991–1996), when the SWE observations are assimilated to produce streamflow predictions. The MAT and MAP during both training and prediction periods generally resemble their counterparts during the whole period, however, MAT during the prediction period is slightly higher and MAP during the training period is higher at both sub-basins (Fig. 2). Further details on the hydroclimatic characteristics of the study basin (NFARB) during the prediction period are provided in Table 3.

3 Methodology

3.1 Integrated uncertainty and Ensemble-based data Assimilation (ICEA)

The ICEA consists of a parameter uncertainty analysis component, ISURF, and a data assimilation component, the EnKF. These two components are described as follows.

3.1.1 Uncertainty analysis

The ISURF utilizes the GSA (Hornberger and Spear, 1981) as an initial step to screen and identify sensitive parameters.

The GSA first samples the parameter space and then classifies the sampled parameters into behavioral and non-behavioral categories according to the performance of the model configured with these parameters. The behavioral parameter sets are then divided into an amount of equally sized groups. Cumulative distributions of each parameter within each group are calculated and the Kolmogorov-Smirnov (KS) test is applied to quantify the closeness between these distributions. The resulting KS value (from 0 to 1) is the maximum vertical distance observed between these distributions, with higher value indicating higher parameter sensitivity. A KS value greater than 0.25 is utilized as a threshold for parameter sensitivity following He et al. (2011b). More detailed explanation and application of the GSA appears in He et al. (2011b).

The ISURF employs the DREAM algorithm for subsequent uncertainty assessment of sensitive model parameters. The DREAM was recently introduced by Vrugt et al. (2008) to estimate optimal parameter values and their underlying posterior probability density. This is accomplished by minimizing the sum of squared residuals between model-predicted and observed variables. In DREAM, a preset number of Markov Chains (a chain refers to a vector containing model parameters considered) are simultaneously run in parallel. For each chain, a candidate vector is generated by taking a discrete proposal distribution containing a fixed multiple of the difference between randomly chosen chains. The Metropolis ratio is used to decide whether to accept the candidate point or not. The convergence of a DREAM run is monitored with the $\hat{R}$ statistic of Gelman and Rubin (1992). The reader is referred to Vrugt et al. (2008) for detailed descriptions on DREAM. In the current study, we use standard algorithm settings for both the GSA and DREAM applied by He et al. (2011b).

3.1.2 Ensemble Kalman filter (EnKF)

The EnKF, first introduced by Evensen (1994), is a Monte Carlo approach which approximates a Bayesian updating scheme to determine the evolution of model states over time (between measurements) and updating of these states at measurement times. Detailed description of EnKF can be found in Evensen (1994, 2003) and Reichle et al. (2002). Basically, the EnKF involves a recursive propagation and updating process. The propagation step produces an ensemble of model state outputs between measurements based on an ensemble of model inputs as follows:

$$y_j(t) = A\left[\mu_j(t), \theta_j, y_j(t-1)\right]; \quad y_j(t=0) = y_{0j} \quad (3)$$

where A is the model operator investigated (SNOW17 in this case) which is subject to time-varying forcing ($\mu(t)$) and time-invariant parameters (θ); y represents model states (Table 1) and y_0 denotes initial states; j ($\in$ [1, 2, ..., N]), N designates ensemble size) is a single ensemble member.

The probability distributions of model forcing and parameters (from which samples $\mu_j(t)$ and θ_j are generated) need to be specified a priori and are further discussed in Sect. 3.3.3.

The ensemble of model states is systematically updated when the new measurement is available. This update largely relies on the differences between actual and predicted measurements, as illustrated below,

$$y_j^+(t_i) = y_j^-(t_i) + K\left\{z_i + \omega_i - M\left[y_j^-(t_i)\right]\right\} \quad (4)$$

where superscripts "−" and "+" denote the state estimates before and after the update, respectively, at a specific measurement time (t_i); z_i is the i-th measurement (areal SWE in this case); ω_i is a random error term generated by the EnKF and has the same characteristics of the systematic measurement errors associated with z_i; M is a measurement operator (linear in this case) mapping model states to the measurement variable(s); and K is the Kalman gain determined from the ensemble statistics and the measurement error specified. Definition of measurement error is presented in Sect. 3.3. Figure 3 illustrates how EnKF is applied in ICEA.

3.2 Implementation of ICEA and proposed scenarios

Given that most snow events occur above the snowline (i.e., within the upper sub-basin), the current study focus on assimilation of SWE observations from the upper sub-basin. Model parameters (SNOW17 and SAC-SMA) used for the lower sub-basin remain unchanged from those applied in RFC operational forecasting. The two components of ICEA, the ISURF and the EnKF, are applied in the training period and prediction period, respectively. The general application procedure is illustrated in Fig. 4 and described as follows (specific details explained in Sect. 4).

– Step 1:
The ISURF is applied to the joint SNOW17/SAC-SMA model in the upper sub-basin to identify sensitive model parameters, their optimal values, and their uncertainty structures (i.e., distribution type and characteristics) during the training period.

– Step 2:
The determined uncertainty structure of sensitive SNOW17 parameters, along with preset forcing uncertainty and areal SWE data uncertainty (Sects. 3.3 and 3.4), is embedded into the EnKF formulation. The EnKF is subsequently employed to assimilate the daily areal SWE data (during the prediction period) to update SNOW17 model states (Table 1) for the upper sub-basin. The assimilation frequency is set to be one week to mimic the operational environment.

– Step 3:
Rain plus snowmelt prediction is correspondingly produced by the SNOW17 based on the updated states. The rain plus snowmelt timeseries ouput is then utilized
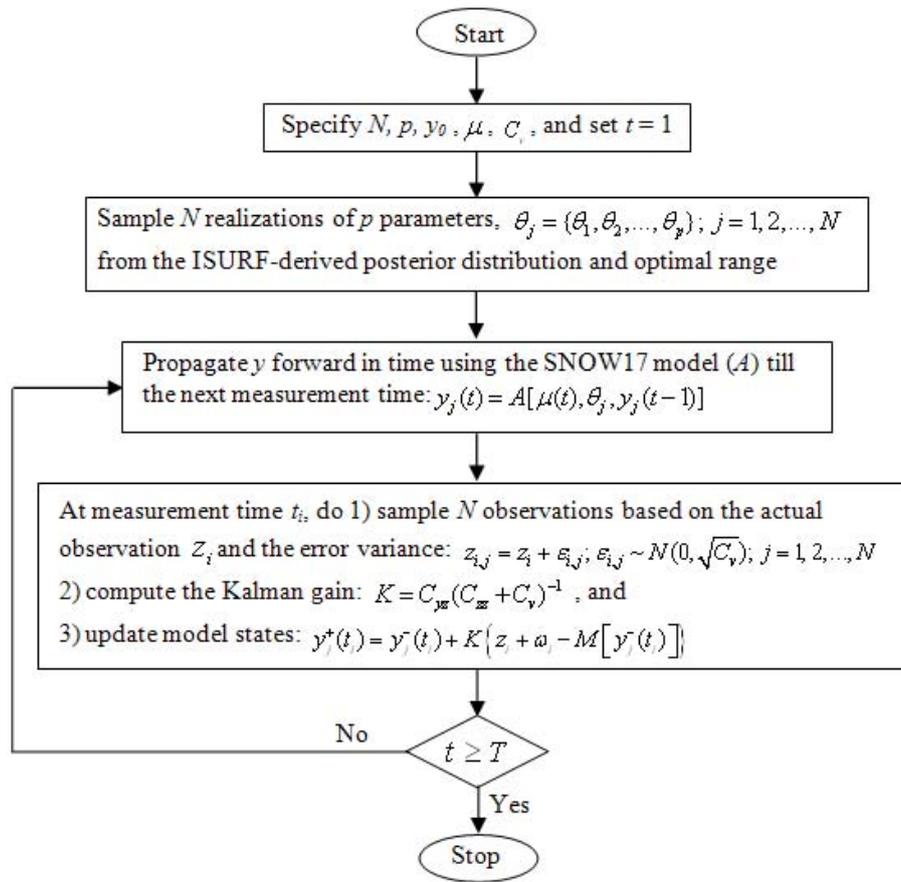
Fig. 3. Flowchart of the EnKF applied in ICEA to recursively update SNOW17 model states with the uncertainty of sensitive model parameters considered. N and p indicate the ensemble size and the number of sensitive parameters, respectively; y_0 and μ represent model initial condition and forcing, respectively; z_i and M designate the observation and measurement operator, respectively; C_{v}, C_{yz}, and C_{zz} denote the variance of observation error, the covariance between model states and observations, and the variance of observations, respectively.

as input to the SAC-SMA model to produce streamflow predictions for the upper sub-basin. The joint SNOW17/SAC-SMA model for the lower sub-basin is run to generate streamflow for the lower sub-basin. The streamflow obtained for both upper and lower sub-basins is combined according to the size (area) of each sub-basin, as is the case in operations. The aggregated flow is routed using the operational unit hydrograph, yielding streamflow predictions at the NFARB outlet.

Four scenarios are considered in our study. The first two scenarios produce single-valued streamflow predictions. The last two involve EnKF applications and thus produce an ensemble of streamflow predictions. The first scenario (entitled "RFC") directly adopts model parameters used in RFC operations. The second scenario (named "ISURF") differs from the first one in that it uses ISURF-derived parameter set with the maximum likelihood (both SNOW17 and SAC-SMA) for the upper sub-basin. The third scenario (entitled "EnKF") applies a stand-alone EnKF technique configured with preset forcing uncertainty and limited information (i.e., the type

of posterior distribution) on parameter uncertainty derived from the ISURF. The fourth scenario (entitled "ICEA") differs from the third one in that it uses both the posterior distribution and optimal posterior parameter ranges derived from the ISURF as inputs to the EnKF. For comparison to the first two deterministic scenarios, the ensemble mean is estimated in the last two scenarios. It is worth noting that, in all scenarios, streamflow (at 6-hourly time step) generated from both the training and prediction periods at the basin outlet are aggregated to the daily time step in order to compare with daily streamflow observations from the USGS gauge.

3.3 Areal SWE data

Snow information, including SWE, is routinely recorded by both in situ observational networks (e.g., the SNOTEL in the western US) and remote sensing platforms (e.g., satellites). However, in situ snow stations are generally sparse and not sufficient to provide accurate areal SWE information. For instance, in the western US there are over 1700 snow stations
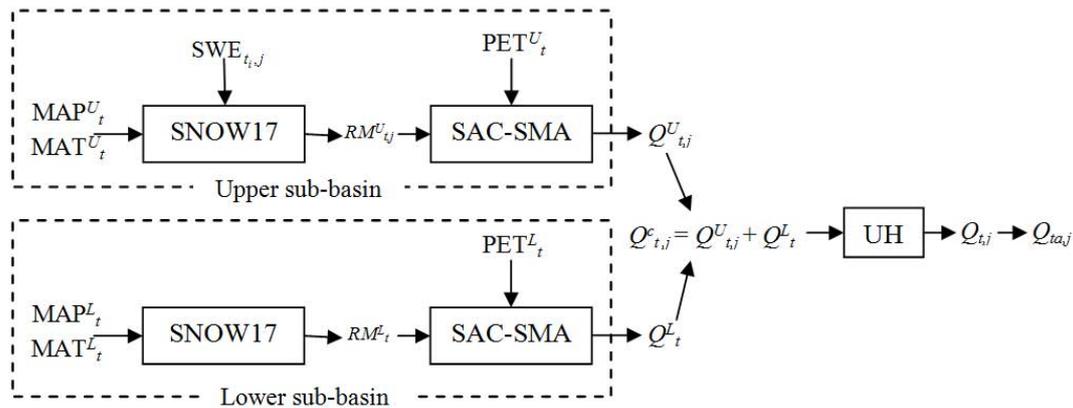
Fig. 4. Flowchart of modeling and assimilation (of areal SWE) procedures in generating ensemble daily streamflow. RM and Q represent rain plus snowmelt and streamflow, respectively; UH stands for unit hydrograph; t indicates time step ($t = 1, 2, ..., T$; 6 hourly); ta is the daily time step aggregated from 6-hourly step; t_i reprsents the measurement time when areal SWE is assimilated (refer to Fig. 2 for detailed assimilation procedure); j is the ensemble number ($j = 1, 2, ..., N$); subscripts U and L denote variables for upper and lower sub-basins, respectively; subscript c designates the combined variable.

which provide SWE observations. Nevertheless, they are still not adequate to resolve the variability of SWE at the basin scale (Bales et al., 2006). There are studies which produce areal SWE estimates from point SWE observations either through interpolation (Fassnacht et al., 2003) or binary tree methods (Molotch and Bales, 2005); however, these studies are limited to specific regions and are less readily applied over broad areas. These conversion methods have also not been applied in operations. There are also remotely sensed SWE data available (e.g., from AMSR-E). However, the data suffers from poor spatial resolution (e.g., AMSR-E at 25 km resolution) and poor accuracy (e.g., Andreadis and Lettenmaier, 2006; De Lannoy et al., 2009). There are also remotely sensed snow cover area (SCA) data available (e.g., from MODIS), which can be transformed to SWE via a snow depletion curve (Andreadis and Lettenmaier, 2006; Durand et al., 2008a,b; Su et al., 2008, 2010; Zaitchik and Rodell, 2009; Thirel et al., 2011). However, the SCA data is error prone (e.g., MODIS SCA data is highly impacted by cloud cover). Furthermore, the relationship between SCA and SWE varies across different hydroclimatic and geographic regimes. Additionally, incorporation of SCA information into hydrologic models may not necessarily lead to improved streamflow simulations (Clark et al., 2006).

Parsimonious methods are typically used to derive areal SWE from in situ SWE measurements in operations. For instance, the NWS RFCs in the western US typically use a regression-based technique to reconcile the SNOW17-simulated areal SWE and the SWE observations from available snow stations. In the current study, we use a similar regression method. The method utilizes SWE information from both the SNOW17 model and the snow stations but assumes that the areal SWE is a linear combination of point SWE measurements so that it resembles the SNOW17 simulated SWE as much as possible. Specifically, the areal SWE

is calculated as a linear combination of the SWE observations from three snow stations. The weight associated with each snow station is determined via a non-negative least-squares algorithm as follows:

$$\min \left\| \sum_{t=1}^{T} \left[\sum_{k=1}^{3} \left(C_k \times \mathrm{SWE}_k^t \right) - \mathrm{SWE}_{\mathrm{model}}^t \right]^2 \right\|, \; \forall \, \mathrm{SWE}_k^t \geq 0 \quad (5)$$

where SWE_k^t represents the SWE observations from the k-th snow station ($k = 1, 2, 3$) at time t ($t = 1, 2, ..., T$, where T is the total length of prediction period in days); C_k is the weight corresponding to the k-th snow station; $\mathrm{SWE}_{\mathrm{model}}^t$ denotes the SNOW17-modeled SWE for the upper sub-basin at time t. By applying the above equation, the weights of three snow stations (C_k) are calculated (Table 2). It should be noted that the final weights do not necessarily add up to be 1.

The measurement error associated with the areal SWE is assumed to be normally distributed white noise with zero mean and a standard deviation of 25 mm. Sensitivity tests at six SNOTEL sites illustrates that this error definition leads to satisfactory ensemble SWE estimates via the EnKF in multiple years with contrasting wetness when using an ensemble size of 100 (He, 2010). In the EnKF applications (the last two scenarios) in this study, the ensemble size is also set to be 100.

3.4 Model forcing uncertainty

In previous EnKF applications (e.g., Margulis et al., 2002; Durand and Margulis, 2006; Clark et al., 2008a; Leisenring and Moradkhani, 2011; Wu and Margulis, 2011, among others), model forcing uncertainty is generally considered by treating input variables (i.e., precipitation and air temperature observations) as random variables and perturbing these variables according to predefined error distributions

Table 2. Basic information for the three study snow stations and the corresponding weights used in determining the areal snow water equivalent.

Site	Name	Latitude	Longitude	Elevation (m)	1 April SWE (mm)	Maximum SWE (mm)	Length of melting period (days)	Peak SWE date	Melt out date	Weight
BL	Blue Canyon	39.276	−120.708	1609	142	349	38	February 27	April 5	0.31
HY	Huysink	39.282	−120.527	2011	870	943	61	April 4	June 6	0.54
CL	CSS Lab	39.333	−120.370	2089	873	1016	64	March 21	May 24	0.44

Table 3. Precipitation and observed streamflow characteristics during the prediction period.

Water Year	1991	1992	1993	1994	1995	1996
Precipitation (mm)	957	955	1723	870	2366	1792
Peak Flow ($m^3 s^{-1}$)	256.23	120.05	311.44	31.14	515.29	427.52
Median Flow ($m^3 s^{-1}$)	1.64	2.53	15.09	2.80	26.53	8.96
Flow Volume ($\times 10^8 m^3$)	3.28	2.45	9.22	2.15	15.40	10.10

and characteristics. In operations, however, it may be difficult to determine credible information on error distributions and characteristics that regional forecasters would agree on and implement in their forecasting practices.

This study adopts an alternative, but simpler, method to address the uncertainty of SNOW17 model forcing, which can be easily implemented in an operational environment. This method draws upon how precipitation and air temperature forcing are currently utilized within the SNOW17 model to determine snowfall input. Specifically, the SNOW17 model uses the air temperature and the PXTEMP parameter to determine whether precipitation is snowfall or rainfall (Eq. 2). Rainfall goes directly to runoff while snowfall first accumulates and then melts out. The snowfall input to the model is then adjusted by a snow correction factor (SCF) (Eq. 1). As such, the actual snowfall digested by SNOW17 model is determined by precipitation and air temperature forcing, as well as parameters SCF and PXTMEP. Hence, instead of perturbing precipitation and air temperature timeseries, we determine the uncertainty of parameters SCF and PXTEMP and assume that this implicitly represent the uncertainty in precipitation and air temperature. This is achieved by applying the same method used in defining the uncertainty of these two SNOW17 parameters in one of our previous studies (He et al., 2011a). This method is relatively easier to understand in concept and requires no explicit quantification of the distribution type of precipitation and air temperature.

3.5 Evaluation metrics

In addition to the deterministic metrics utilized in this study, including the correlation coefficient (R), Root Mean Squared Error (RMSE), Nash-Sutcliffe Efficiency (NSE), and percent bias (Bias), we apply two metrics to quantity the spread dispersion of the EnKF-derived ensemble streamflow predictions: the Normalized Root Mean Square Error Ratio (NRR) (Anderson, 2002; Moradkhani et al., 2005b) and the 95th Percentile Uncertainty Ratio (UR95) (Hossain and Anagnostou, 2005; Moradkhani et al., 2006). These two metrics are defined as:

$$\text{NRR} = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\bar{Q}_t - \hat{Q}_t\right)^2}}{\frac{1}{N}\left\{\sum_{i=1}^{N}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Q_t^i - \hat{Q}_t\right)^2}\right\}\sqrt{\frac{N+1}{2N}}} \tag{6}$$

$$\text{UR95} = \frac{\sum_{t=1}^{T}\left(Q_{97.5,t} - Q_{2.5,t}\right)}{\sum_{t=1}^{T}\hat{Q}_t} \times 100\% \tag{7}$$

where N designates the ensemble size ($N = 100$ in this study); T is the length of the prediction period (in days); $\hat{Q}_t$ represents the observed flow at time t; Q_t^i indicates the predicted flow at time t of the i-th ensemble ($1 \le i \le N$);

$\bar{Q}_t$ denotes the ensemble mean flow at time t; $Q_{97.5,t}$ and $Q_{2.5,t}$ signify the 97.5th and 2.5th percentiles of the ensemble flow values at time t, respectively; and subsequently, $(Q_{97.5,t} - Q_{2.5,t})$ is the 95th percentile prediction uncertainty bound of the ensemble at time t.

NRR is a normalized measure of ensemble dispersion relative to the deviation of the ensemble mean (Anderson, 2002). It is always greater than 0, with a perfect score equivalent to 1. A value of NRR greater than 1 signifies too little spread. A value of NRR between 0 and 1 indicates too much spread. UR95 is a measure of the aggregate variability of the 95th percentile prediction range relative to the observations, generally ranging from 0 to 100 % (Moradkhani et al., 2006).

4 Results

4.1 Parameter uncertainty

The GSA method is applied to derive "behavioral" SNOW17 and SAC-SMA model parameters for the upper sub-basin. Subsequent sensitivity analysis focuses on the behavioral SNOW17 parameter set since only SNOW17 parameter uncertainty is addressed in the data assimilation applications. The analysis shows that two forcing-related SNOW17 parameters, SCF and PXTEMP as well as parameter MFMAX (which determines the non-rain melt rate) and parameter SI (which determines the areal extent of the snow cover) are sensitive (Fig. 5). The KS value of other parameters are consistently lower than 0.2. Sensitivity of SCF and PX-TEMP is expected because they control the snowfall input to the model. Sensitivity of MFMAX is also intuitive because non-rain melt is deemed the dominant melting mechanism in western basins (Franz et al., 2008). The sensitivity of SI reflects the importance of the snow cover area parameter when the SNOW17 model is applied at the areal scale. Only these four sensitive parameters are analyzed in the following uncertainty analysis and data assimilation applications. Other SNOW17 parameters and all SAC-SMA parameters for the upper sub-basin are fixed at the values which provide the best streamflow simulations (highest NSE) during the training period.

The DREAM algorithm is subsequently applied to derive posterior information for the four sensitive parameters. The marginal distributions of parameters SCF, MFMAX, and SI generally follow a similar distribution type, which can be roughly identified as a normal distribution (Fig. 6). Parameter PXTEMP shows no apparent distribution in its optimal range. As such, a uniform distribution is assigned for PX-TEMP. SCF is significantly correlated to MFMAX (with a correlation coefficient of 0.77). This is due to the fact that increasing SCF leads to a deeper snowpack. A deep snowpack results in a higher maximum melt rate (MFMAX) in comparison to a shallower snowpack. MFMAX shows medium correlation to SI. The correlation between other parameter sets is
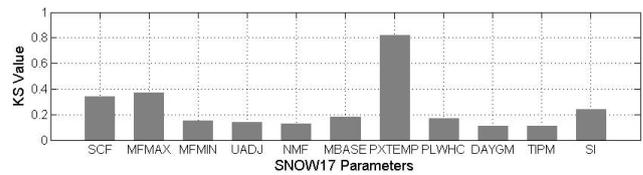

Fig. 5. Kolmogorov-Smirnov (KS) values for SNOW17 parameters for the upper sub-basin.

not significant. Despite the existence of the cross-correlation among sensitive parameters, for simplicity, this correlation is not accounted for when formulating the EnKF in the current study (in the last two scenarios considered). However, implementation of the cross-correlation among parameters into the EnKF is being explored in our ongoing work.

4.2 Streamflow simulation

The maximum likelihood values of the four sensitive parameters, along with other SNOW17 parameters and SAC-SMA parameters determined from the GSA for the upper sub-basin, are applied to simulate streamflow at the outlet (scenario "ISURF") following the procedure described in Sect. 3.2. The outflow is compared to that simulated using the RFC parameter set (scenario "RFC"). Observed flow at the outlet is used as the benchmark for comparison. The annual statistics between observed flow and the flow simulated from both scenarios are calculated both on an annual basis as well as for the entire training period (Table 4).

In general, the ISURF parameters outperform the RFC parameters in providing streamflow simulations, not only during the entire training period, but also for each individual year. Particularly, the RFC parameters consistently underestimate streamflow, while the ISURF parameters largely correct these biases. For example, the RFC parameters provide relatively poor streamflow simulations during WY 1979 (e.g., bias of −55.96 % and correlation coefficient of 0.38). In comparison, the ISURF simulation during this year is fairly satisfactory (e.g., with a bias of 1.8 % and a correlation coefficient of 0.77). These results illustrate that there is room to optimize the parameters currently under use by the RFCs and improve model performance. However, it is worth noting that during real-time forecasting, RFC forecasters routinely apply run-time modifications (MODs) in adjusting initial (i.e., initial model states) and boundary (e.g., precipitation forcing) conditions in order to improve the quality of streamflow predictions (Seo et al., 2003). In the current study, no MODs are applied for the RFC simulations.

4.3 Streamflow prediction

For the stand-alone EnKF (scenario "EnKF"), identified sensitive SNOW17 parameters (SCF, PXTEMP, MFMAX, and SI) from the upper sub-basin model are perturbed following

Table 4. Statistics for observed and simulated streamflow using the RFC parameters and ISURF parameters during the training period.

Metrics	Scenarios	Period (Water year)						
		1979–1984	1979	1980	1981	1982	1983	1984
R	ISURF	0.89	0.77	0.96	0.89	0.83	0.87	0.91
	RFC	0.78	0.38	0.86	0.85	0.68	0.76	0.84
Bias	ISURF	−9.02	1.89	−8.48	12.73	−11.11	−13.33	−10.98
(%)	RFC	−31.87	−55.96	−36.17	−16.04	−37.29	−29.33	−14.02
RMSE	ISURF	19.18	10.45	14.32	3.92	35.53	20.01	14.67
($m^3\,s^{-1}$)	RFC	27.62	17.11	28.03	4.55	48.51	27.48	19.22
NSE	ISURF	0.95	0.91	0.98	0.95	0.91	0.94	0.96
	RFC	0.91	0.77	0.97	0.93	0.86	0.91	0.93
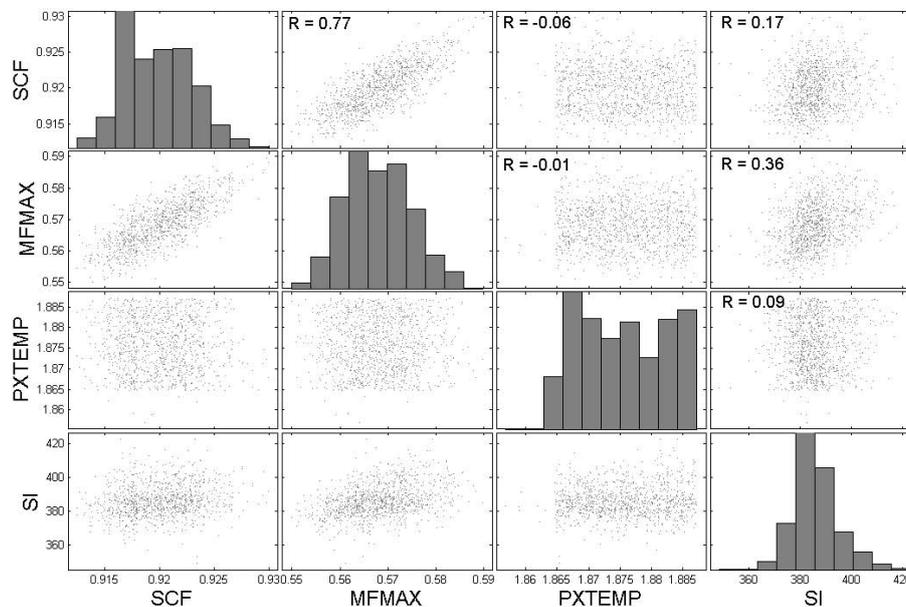
Fig. 6. Marginal distributions (in bars) and scatter plots (in dots) for posterior SNOW17 parameters. R signifies the linear correlation between parameters.

the identified distributions (Fig. 6) within the pre-defined feasible parametric ranges (Table 1) to produce parameter ensembles. Note that it is ideal to sample parameters from their joint distribution. However, joint distribution of DREAM-derived posterior parameters is generally of a complex form and not straightforward to explicitly derive (He et al., 2011b). The last scenario, ICEA, differs from the stand-alone EnKF only in parameter ranges where the perturbations are conducted. The ICEA scenario uses the optimal parameter ranges identified via the ISURF (i.e., the ranges shown in Fig. 6), which is significantly narrower than the feasible parameter ranges (Table 1). In both scenarios, perturbation of SNOW17 forcing (precipitation and air temperature) is assumed to be implicitly included in perturbing parameters SCF and PXTEMP, as discussed in Sect. 3.4. Areal SWE

data for the upper sub-basin is assimilated into the SNOW17 model to produce rain plus snowmelt ensembles, which are consequently applied to generate ensemble streamflow predictions at the basin outlet following the general procedure outlined in Sect. 3.2.

Streamflow prediction is investigated in terms of both deterministic (single-valued) and probabilistic (ensemble) predictions. For the former, streamflow predictions within the prediction period from all four scenarios are compared to the observed streamflow during the same period, using the ensemble mean from the two data assimilation scenarios (i.e., EnKF and ICEA). This comparison focuses on the whole streamflow timeseries as well as the high flows. High flows in each scenario are defined as those greater than the 95th percentile of the observed streamflow timeseries during
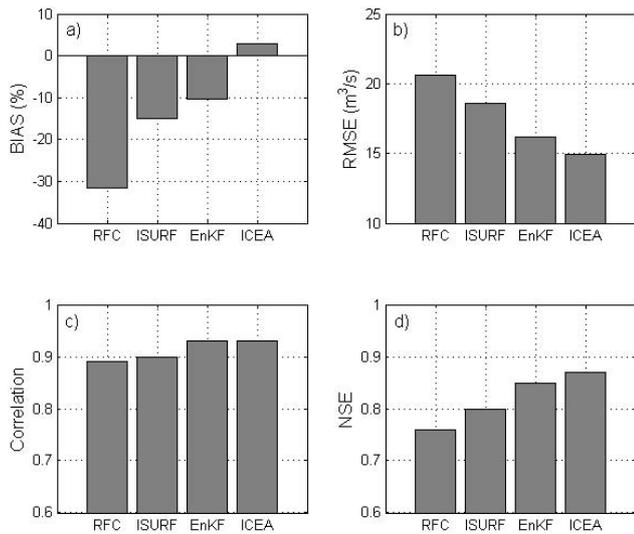
Fig. 7. Statistics between observed and predicted streamflow during the prediction period (WY 1991–1996) under four scenarios considered for Bias **(a)**, RMSE **(b)**, Correlation **(c)** and NSE **(d)**. The ensemble-mean prediction is used in estimating statistics for the EnKF and ICEA scenarios.

Fig. 8. Scatter plot of predicted and observed high flows during the prediction period under scenarios **(a)** RFC, **(b)** ISURF, **(c)** EnKF, and **(d)** ICEA. For the EnKF **(c)** and ICEA **(d)** scenarios, the entire ensemble prediction ranges (error bars) are also shown except for the ensemble mean prediction (dots).

Table 5. Statistics for observed and predicted high flows (above the 95th percentile of observed streamflow) during the prediction period. The EnKF and ICEA statistics are based on the ensemble mean predictions.

	RFC	ISURF	EnKF	ICEA
R	0.80	0.85	0.83	0.87
Bias (%)	−19.58	−13.70	−14.91	−10.02
RMSE ($m^3\,s^{-1}$)	65.37	56.47	58.33	49.59
NSE	0.50	0.62	0.60	0.71

the prediction period. The 95th percentile flow and the number of days (with high flows) are 83 $m^3\,s^{-1}$ and 110, respectively. Streamflow predictions from the two data assimilation scenarios are inter-compared from two perspectives. The first involves the ensemble statistics (i.e., NRR and UR95) as well as hydrographs. The second focuses on the predictability of both scenarios for different lead times. Specifically, since the areal SWE is assimilated once per week (once SNOW17 states are updated, the model is run up to seven days before the next update occurs), lead times are considered varying from one day to seven days with a daily interval. Streamflow prediction with a specific lead time, day L ($L = 1, 2, ..., 7$), is a vector containing predicted flow of the L-th day following each measurement (and update) time during the entire prediction period.

Deterministic metrics (including NSE, RMSE, percent bias, and correlation) from all scenarios during the prediction period are calculated (Fig. 7). In comparison to the RFC
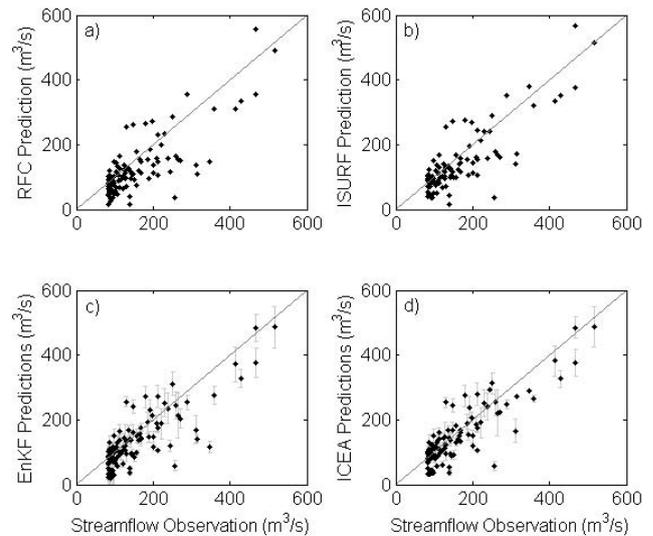
prediction, the ISURF prediction has a lower bias and RMSE and a higher correlation and NSE. However, when compared to the EnKF predictions, the ISURF prediction is poorer. This reflects the potential of data assimilation approaches for providing improved streamflow predictions over traditional (i.e., the RFC parameter calibration method) and advanced (i.e., ISURF) parameter identification methods. The combination of ISURF and EnKF, the ICEA, shows further improvement over the stand-alone EnKF. The NSE of the ICEA prediction is 2.4 % higher than that of the stand-alone EnKF, while the RMSE value of the ICEA prediction is 7.7 % lower. Compared to the RFC prediction, the ICEA NSE is 14.5 % higher and the RMSE is 27.6 % lower. The superior performance of the ICEA is even more significant in terms of percent bias. The ICEA prediction has a bias of 2.8 %, in comparison to −10.4 % for the stand-alone EnKF. As a reference, the bias of the RFC prediction is −31.6 %, suggestive of significant improvements in flow volume using the data assimilation approaches, especially the ICEA.

The performance of the four scenarios in providing high flow predictions is also investigated (Table 5; Fig. 8). Overall, all scenarios underestimate high flows, while the ICEA (mean) prediction outperforms the other three scenarios and the RFC prediction is least satisfactory (Table 5; Fig. 8). The ISURF predictions are comparable to the EnKF mean prediction (Fig. 8), yet the former has slightly better statistics (Table 5). Comparing the two ensemble scenarios specifically, the ICEA considerably outperforms the EnKF in terms of all four metrics (Table 5). Further investigation shows the average replicate-wise RMSE (averaged RMSE of all individual

Table 6. Ensemble performance statistics associated with the EnKF and ICEA scenarios during the prediction period.

Metrics	Scenarios	Period (Water year)						
		1991–1996	1991	1992	1993	1994	1995	1996
UR95	EnKF	25.85	28.15	37.14	25.22	43.19	24.15	21.84
(%)	ICEA	20.58	20.82	8.47	27.12	3.01	29.59	34.43
NRR	EnKF	1.33	1.40	1.35	1.37	1.35	1.30	1.32
	ICEA	1.33	1.40	1.35	1.36	1.39	1.30	1.32

ensemble prediction traces) values associated with the EnKF and ICEA are 62.90 m^3 s^{-1} and 53.18 m^3 s^{-1}, respectively. These observations trace back to the fact that EnKF sample sensitive parameters in wider ranges relative to the ICEA. Part of these derived parameter sets may lead to flow predictions with larger biases. It is thus likely that the (mean) predictive skill of the EnKF is somewhat diluted by those biased predictions. The error bars of the EnKF predictions and ICEA predictions are also presented (Fig. 8c–d), while both the RFC and the ISURF only provide a single-valued prediction. It is evident that the ensemble predictions of both EnKF and ICEAs encompass the highest peak flow.

The dispersion of the ensemble streamflow predictions derived from the two data assimilation applications is also investigated (Table 6). Overall, the stand-alone EnKF ensembles show higher UR95 (25.85 %) when compared to that of the ICEA ensembles (20.58 %), indicating that the 95th uncertainty bound of the EnKF is wider than that of the ICEA (Table 6). At the annual scale, on average, the ensemble spread associated with the EnKF accounts for around 22 % to (WY 1996) to 43 % (WY 1995) of the magnitude of the observations. For the ICEA, the range is from about 3 % (WY 1995) to 34 % (WY 1996). This further indicates that the 95th percentile uncertainty bounds of the ICEA ensembles are relatively narrower. When looking at the entire ensemble, however, the spread of predictions from both scenarios are almost identical; not only during the entire prediction period but also in each individual year. The NRR values are consistently greater than 1, indicating that ensemble predictions of both scenarios have too little spread relative to the deviation of the mean prediction. Additionally, it is evident that the inter-annual variation pattern of NRR is similar for both scenarios. However, the inter-annual variation pattern of UR95 is very different in the two scenarios, suggesting that the stand-alone EnKF and ICEA provide significantly different ensemble predictions though the overall spread of both ensemble sets (NRR) are similar.

To examine the performance of the stand-alone EnKF and ICEA methods at a higher temporal resolution, ensemble streamflow predictions derived from both approaches are investigated at the daily scale against the RFC single-valued predictions as well as observed streamflow.

For demonstration purposes, the results from one year (WY 1995) are presented (Fig. 9). The selection of WY 1995 is based on the facts that (1) it is the wettest year in the prediction period in terms of both annual precipitation and the total annual flow volume; and (2) the maximum peak flow (515.29 m^3 s^{-1}) within the prediction period occurs in this year (Table 3). Results indicate that both the EnKF and ICEA ensembles are fairly narrow, which is consistent with previously discussed results (Table 6). Both ensembles capture the primary peak at day 103 (Fig. 9), while this peak is underestimated by the RFC predictions. Both ensembles also contain the secondary peak at day 162. The high flows at days 106 and 213 are not totally encompassed by both ensembles. Performances of the EnKF and ICEA prior to day 230 are almost identical in terms of the ensemble spread and variation pattern. However, between day 230 and day 289, the performances of both approaches are considerably different. First, from day 230 to day 252, in comparison to the ICEA ensemble, the EnKF ensemble is much wider and well encompasses the observations; second, from day 265 to day 289, the ICEA ensemble reasonably captures the variation pattern of streamflow observations in this period, while the EnKF ensemble follows the variation of RFC predictions which deviate from the observed streamflow with a negative bias.

Recall that both the EnKF and ICEA assimilate areal SWE to update SNOW17 model states and thus produce rain plus snowmelt, which serves as input to SAC-SMA model to generate streamflow. To examine the underlying cause of the deviations in the approaches from days 230 to 289, it is necessary to scrutinize the corresponding SWE and rain plus snowmelt predictions. Given the high variability of the rain plus snowmelt time series, only the ensemble-mean values during the same year are plotted (Fig. 10). It is evident that prior to day 230, the mean SWE predictions from both approaches are almost identical, as are the rain plus snowmelt predictions. After day 230, the EnKF predicted SWE declines faster, which is most likely caused by the inconsistent non-rain melt parameter (MFMAX, which controls the non-rain melt rate of the snowpack) applied in both approaches. The enhanced decline in SWE produces higher rain plus snowmelt onward until day 252, which is the period when the EnKF ensemble streamflow predictions
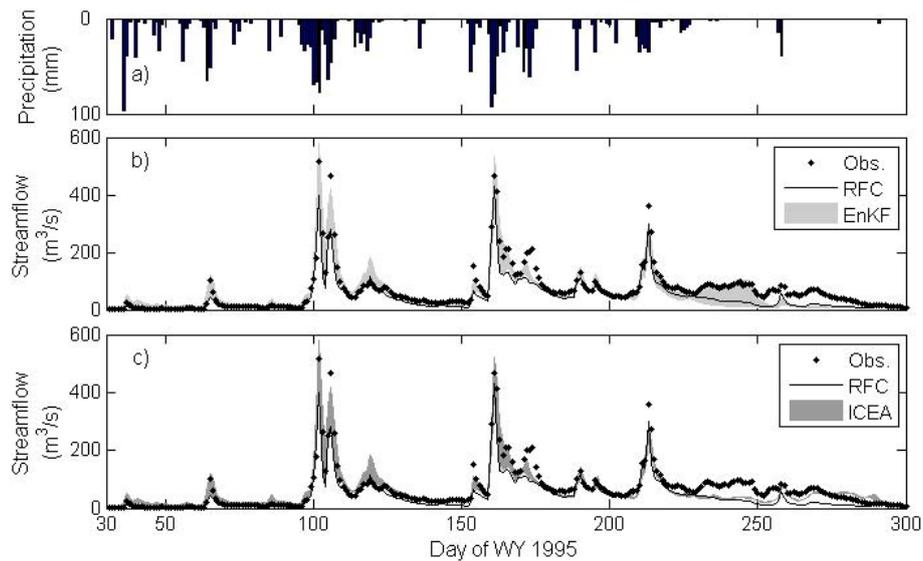
Fig. 9. Basin-average precipitation **(a)**, ensemble streamflow predictions from the stand-alone EnKF **(b)**, and ensemble streamflow predictions from the ICEA **(c)** for water year 1995. Also shown are observed streamflow (circles) and RFC streamflow predictions (black line) during the same period. The grey regions correspond to the entire ensemble range.
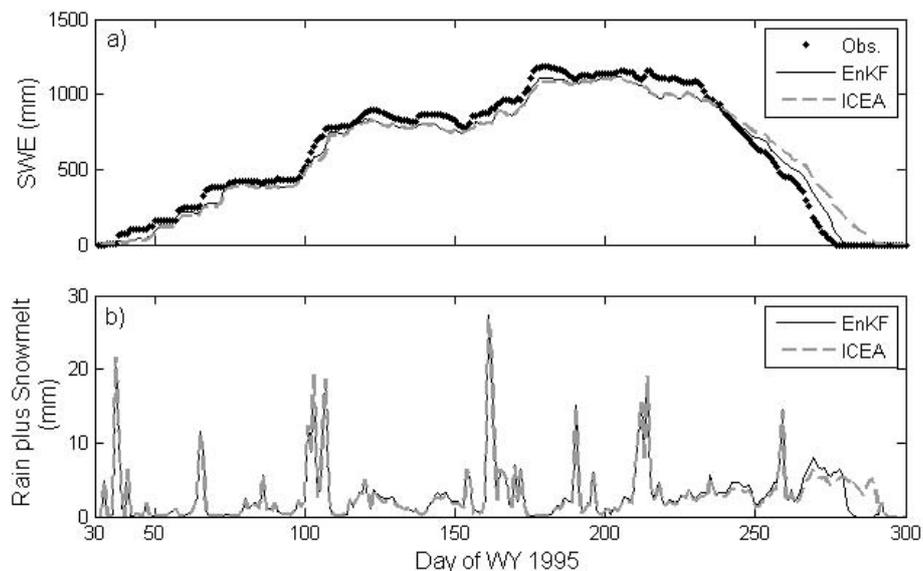

Fig. 10. Ensemble mean SWE **(a)** and Rain plus Snowmelt **(b)** time series for WY 1995 derived from the stand-alone EnKF (black line) and the ICEA (dash line). Also shown in the upper panel is the areal SWE (circles) derived from SWE observations at three snow stations.

significantly spread. From days 253–264, when there is a significant rainfall event (Fig. 9a), rainfall dominates the rain plus snowmelt. As such, the rain plus snowmelt predictions and thus streamflow prediction during this period again become nearly identical for both approaches (Figs. 10b and 9b–c). From day 265 until the snow melts out, there are no major rainfall events (Fig. 9a). Snowmelt thus dominates the rain plus snowmelt. From days 278–289, it is evident that the EnKF predicted SWE reduces to zero earlier than ICEA predicted SWE (Fig. 10a). This is most likely due to the fact

that the EnKF employs parameter values for MFMAX and SI that produce large non-rain melt in this period, particularly given the fact that the EnKF samples these two parameters in significantly larger ranges (Table 1) than the ICEA (Fig. 6). During the same period, the rain plus snowmelt prediction of the ICEA is thus considerably greater than that of the EnKF (Fig. 10b). As a result, the ICEA predicted streamflow ensembles mimics the observed streamflow much better than the EnKF predicted ensembles in this period (Fig. 9b–c).

The stand-alone EnKF and ICEA predictions at different lead times are also examined. Both deterministic (using ensemble mean prediction) and ensemble statistics are calculated for lead times up to seven days (Fig. 11). The ICEA mean prediction on average outperforms the EnKF mean prediction at all lead times investigated (Fig. 11a–d), except at lead time day 2, where the NSE and correlation values of the ICEA mean prediction are slightly lower (Fig. 11c–d). This is similar to what is observed for the ICEA and EnKF mean predictions over the entire prediction period (Fig. 8). Particularly, the ICEA predictions have considerably lower bias (Fig. 11a) and generally smaller RMSE (Fig. 11b) compared to the EnKF predictions, indicating that ICEA outperforms the EnKF in providing both overall and high flow predictions. At all lead times, the 95th percentile ensemble predictions of the EnKF are wider than that of the ICEA (Fig. 11e), which is also the case when looking at the entire prediction period (Table 6). In contrast, the whole EnKF predicted streamflow ensemble is less overconfident than the ICEA ensemble at several lead times (day 2, day 6, and day 7), while the ensemble is more overconfident at other lead times (Fig. 9f). In general, the performance of both approaches remain fairly stable across all lead times (Fig. 11), while the predictive skill of operational models normally decreases with increasing lead times in real-time hydrologic forecasting. This is due to the fact that the precipitation and forcing data applied in the current study are actually observed values. In real-time forecasting, precipitation and air temperature are predicted future values (from numerical weather models) where accuracy decreases with increasing lead times (e.g., Cloke and Pappenberger, 2009; Wu et al., 2011).

5 Discussion and conclusions

The current study proposes an integrated uncertainty and ensemble-based data assimilation framework ICEA. This framework, consisting of a parameter uncertainty analysis algorithm (ISURF) and a data assimilation technique (EnKF), systematically addresses uncertainty in model forcing and parameters for the SNOW17 model. The performance of the framework is evaluated against observed streamflow and compared to the performance of three alternative scenarios: the current RFC operational parameters, the stand-alone ISURF, and the stand-alone EnKF. Datasets from a NWS operational basin are applied in the evaluation and comparison. The key findings of the study include:

1. The automatic RFC prediction is improved by applying a parameter identification tool (i.e., the ISURF) to update model parameters or by employing a data assimilation technique (i.e., the EnKF) which assimilates SWE information to provide updated snow model states, or by using a combination of both (i.e., ICEA). The improvement over the current RFC prediction is most significant for the ICEA (in terms of ensemble-mean prediction),
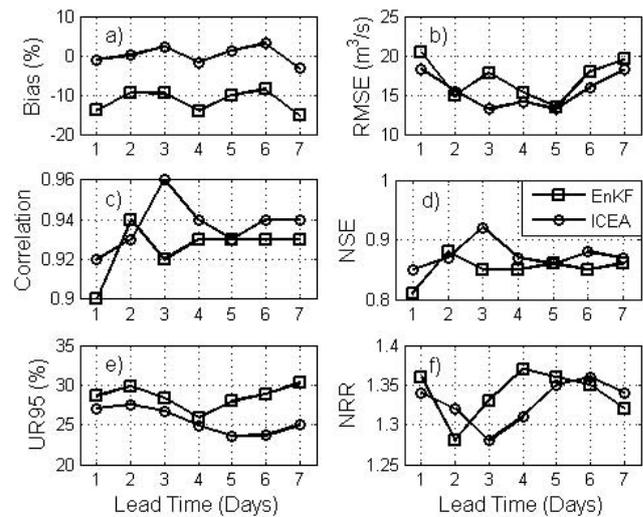

Fig. 11. Statistics for streamflow predictions at various lead times from the stand-alone EnKF and ICEA, including: **(a)** Bias, **(b)** RMSE, **(c)** Correlation, **(d)** NSE, **(e)** UR95, and **(f)** NRR.

followed by the EnKF (ensemble-mean prediction) and then the ISURF. However, when evaluating only high flows (top 5th percentile), ISURF and EnKF (mean) predictions are comparable to each other and the ICEA (mean) prediction shows the most skill.

2. The 95th percentile streamflow prediction range of the stand-alone EnKF is generally wider than that of the ICEA, which is expected since the uncertainty ranges of the four sensitive SNOW17 parameters employed in the EnKF are wider. However, the spread of the whole ensemble (rather than 95th percentile) associated with the EnKF is comparable to that of the ICEA. The spread, based on the ensemble metric calculated (i.e., NRR), is too narrow relative to the (ensemble) mean prediction. This most likely stems from the definitions of uncertainty in forcing as well as areal SWE measurements. Specifically, precipitation and air temperature uncertainty are implicitly accounted for by the uncertainty of parameters SCF and PXTEMP. While this method is straightforward in concept and easy to implement in operations, it limits the spread of forcing ensemble since theoretically there are feasible bounds for those parameter values. Furthermore, a consistent standard deviation is assigned to the SWE measurement error. Sensitivity tests at the point scale illustrate that this error definition provides satisfactory ensemble results, leading to limited variations to large SWE values (He, 2010). Our ongoing work is investigating an alternative definition, namely, assuming this deviation is proportional to the observed SWE value.

3. Both the stand-alone EnKF and ICEA predicted streamflow ensembles contain peak flows during the prediction

period, while the RFC prediction generally underestimate peaks. For the selected (extreme wet) year, both EnKF and ICEA ensembles perform similarly in the snow accumulation period, but dramatically different in the ablation period. This deviation stems from different melt rate values produced by both approaches which ultimately produce different rain plus snowmelt output. The streamflow ensembles predicted by both approaches reasonably capture the variation patterns of the observed streamflow; however, the ensembles are not wide enough to enclose all the observations. This further indicates that the uncertainty defined in the current study can be improved upon.

4. On average, the ICEA mean prediction outperforms the mean prediction of the stand-alone EnKF for all lead times investigated (day 1 to day 7 in with daily interval). However, the EnKF 95th percentile prediction bound is consistently larger than that of the ICEA at all lead times, which is also the case for the entire prediction period as previously discussed. As for the ensemble predictions at different lead times derived from both approaches, their spreads are again too narrow when compared to the (ensemble) mean prediction at each lead time.

Despite its demonstrated advantages, the ICEA can be further improved by addressing several issues in addition to using time-variant error for SWE measurement as previously mentioned. First, it should be highlighted that the SNOW17 model applies air temperature input to calculate melt rate during the snow ablation periods in addition to applying it as a threshold for distinguishing rainfall from snowfall together with model parameter PXTEMP. As such, the uncertainty of parameter PXTEMP only partially mimics uncertainty in air temperature forcing. To investigate the impacts of air temperature measurement error on melt rate, snowmelt, and ultimately streamflow prediction, this error needs to be explicitly defined. This error can be simply assumed to be a systematic error with zero mean and a certain variance, as with previous studies (e.g., Margulis et al., 2002; Durand and Margulis, 2006). Moreover, the uncertainty of parameter SCF derived from ISURF represents only partial uncertainty of precipitation. This is due to the fact that ISURF is applied over the entire training period to produce the uncertainty range and distribution information of SCF. Thus, the uncertainty of SCF is more representative of the average precipitation error in the training period rather than errors associated with individual precipitation events. To more comprehensively account for uncertainty in precipitation, it may be necessary to implicitly consider the error for each precipitation event. This could be achieved by assigned a random multiplier (with a certain distribution) to observed precipitation at each time step following previous studies (e.g., Margulis et al., 2002; Leisenring and Moradhkani, 2011). In addition, it should also be pointed out that model error is not considered in the current study, while model structural error is shown to considerably impact SNOW17 performance and lead to SNOW17-predicted SWE ensemble with significant spread (He et al., 2011a). This error can be addressed by perturbing the predicted SWE using a uniformly distributed and auto-correlated error assumption as explored in previous studies (e.g., Evensen, 1994; Leisenring and Moradkhani, 2011). Addressing this error is promising for resolving the narrow-ensemble issue encountered by the ICEA and stand-alone EnKF. Furthermore, it is worth noting that the ensemble streamflow predictions are evaluated only in terms of dispersion, while other aspects of ensemble predictions (e.g., reliability and discrimination) also contains meaningful information on the predictions (e.g., Franz et al., 2003; Brown et al., 2010; Demargne et al., 2010; Franz and Hogue, 2011). Our ongoing work, which attempts to enhance the ICEA framework to further improve streamflow predictions including real-time predictions, comprehensively considers the uncertainty in forcing, model structure, and measurement, the cross-correlation between parameters, and adopts alternative statistical metrics to verify the ensemble predictions.

Edited by: F. Pappenberger

References

Ajami, N. K., Gupta, H., Wagener, T., and Sorooshian, S.: Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system, J. Hydrol., 298, 112–135, 2004.

Ajami, N. K., Duan, Q., and Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, Water Resour. Res., 43, W01403, doi:10.1029/2005WR004745, 2007.

Anderson, E. A.: National Weather Service River Forecast System-Snow Accumulation and Ablation Model. National Oceanographic and Atmospheric Administration, Silver Springs, Md., Tech. Memo NWS HYDRO-17, 1973.

Anderson, E. A.: Calibration of Conceptual Models for Use in River Forecasting, Hydrology Lab, NOAA National Weather Service, Silver Spring, MarylandNOAA Technical Report, NWS 45, Hydrology Laboratory, 2002.

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

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., and Dozier, J.: Mountain hydrology of the western United States, Water Resour. Res., 42, W08432, doi:10.1029/2005WR004387, 2006.

Beven, K. and Binley, A.: Future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, 1992.

Brown, J. D., Demargne, J., Seo, D.-J., and Liu, Y.: The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations, Environ. Modell. Softw., 25, 854–872, 2010.

Burnash, R. J. C., Ferral, R. L., and McGuire, R. A.: A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers, US Dept. of Commerce, National Weather Service, 1973.

Clark, M. P. and Hay, L. E.: Use of Medium-Range Numerical Weather Prediction Model Output to Produce Forecasts of Streamflow, J. Hydrometeorol., 5, 15–32, 2004.

Clark, M. P. and Kavetski, D.: Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, Water Resour. Res., 46, W10510, doi:10.1029/2009wr008894, 2010.

Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., and Leavesley, G. H.: Assimilation of snow covered area information into hydrologic and land-surface models, Adv. Water Resour., 29, 1209–1221, 2006.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: 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, 2008a.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00B02, doi:10.1029/2007WR006735, 2008b.

Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, J. Hydrol., 375, 613–626, 2009.

Daley, R.: Atmospheric data analysis, Cambridge University Press, Cambridge, UK, 1993.

De Lannoy, G. J. M., Houser, P. R., Verhoest, N. E. C., and Pauwels, V. R. N.: Adaptive Soil Moisture Profile Filtering for Horizontal Information Propagation in the Independent Column-Based CLM2.0, J. Hydrometeorol., 10, 766–779, 2009.

Demargne, J., Brown, J., Liu, Y., Seo, D.-J., Wu, L., Toth, Z., and Zhu, Y.: Diagnostic verification of hydrometeorological and hydrologic ensembles, Atmos. Sci. Lett., 11, 114–122, 2010.

Dunne, S. and Entekhabi, D.: Land surface state and flux estimation using the ensemble Kalman smoother during the Southern Great Plains 1997 field experiment, Water Resour. Res., 42, W01407, doi:10.1029/2005wr004334, 2006.

Durand, M. and Margulis, S. A.: Feasibility Test of Multifrequency Radiometric Data Assimilation to Estimate Snow Water Equivalent, J. Hydrometeorol., 7, 443–457, 2006.

Durand, M., Molotch, N. P., and Margulis, S. A.: A Bayesian approach to snow water equivalent reconstruction, J. Geophys. Res., 113, D20117, doi:10.1029/2008JD009894, 2008a.

Durand, M., Molotch, N. P., and Margulis, S. A.: Merging complementary remote sensing datasets in the context of snow water equivalent reconstruction, Remote Sens. Environ., 112, 1212–1225, 2008b.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 143–110, 1994.

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

Fassnacht, S. R., Dressler, K. A., and Bales, R. C.: Snow water equivalent interpolation for the Colorado River Basin from snow telemetry (SNOTEL) data, Water Resour. Res., 39, 1208–1218, 2003.

Franz, K. J. and Hogue, T. S.: Evaluating uncertainty estimates in hydrologic models: borrowing measures from the forecast verification community, Hydrol. Earth Syst. Sci., 15, 3367–3382, doi:10.5194/hess-15-3367-2011, 2011.

Franz, K. J., Hartmann, H. C., Sorooshian, S., and Bales, R.: Verification of National Weather Service Ensemble Streamflow Predictions for Water Supply Forecasting in the Colorado River Basin, J. Hydrometeorol., 4, 1105–1118, 2003.

Franz, K. J., Hogue, T. S., and Sorooshian, S.: Snow Model Verification Using Ensemble Prediction and Operational Benchmarks, J. Hydrometeorol., 9, 1402–1415, 2008.

Gelman, A. and Rubin, D. B.: Inference from Iterative Simulation Using Multiple Sequences, Stat. Science, 7, 457–472, 1992.

He, M.: Data assimilation in watershed models for improved hydrologic forecasting, Ph.D., Civil and Environmental Engineering, University of California, Los Angeles, 173 pp., 2010.

He, M., Hogue, T. S., Franz, K. J., Margulis, S. A., and Vrugt, J. A.: Corruption of parameter behavior and regionalization by model and forcing data errors: A Bayesian example using the SNOW17 model, Water Resour. Res., 47, W07546, doi:10.1029/2010WR009753, 2011a.

He, M., Hogue, T. S., Franz, K. J., Margulis, S. A., and Vrugt, J. A.: Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes, Adv. Water Resour., 34, 114–127, 2011b.

Hogue, T. S., Sorooshian, S., Gupta, H., Holz, A., and Braatz, D.: A multistep automatic calibration scheme for river forecasting models, J. Hydrometeorol., 1, 524–542, 2000.

Hogue, T. S., Gupta, H., and Sorooshian, S.: A "User-Friendly" approach to parameter estimation in hydrologic models, J. Hydrol., 320, 202–217, 2006.

Hornberger, G. M. and Spear, R. C.: An approach to the preliminary analysis of environmental systems, J. Environ. Manage., 12, 7–18, 1981.

Hossain, F. and Anagnostou, E. N.: Numerical investigation of the impact of uncertainties in satellite rainfall estimation and land surface model parameters on simulation of soil moisture, Adv. Water Resour., 28, 1336–1350, 2005.

Kavetski, D., Franks, S. W., and Kuczera, G.: Confronting input uncertainty in environmental modelling, in: Calibration of Watershed Models, Water Sci. and Appl., edited by: Duan, Q., Gupta, H., Sorooshian, S., Rousseau, A., and Turcotte, R., Washington, D.C., 49–68, 2002.

Kitanidis, P. K. and Bras, R. L.: Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty, Water Resour. Res., 16, 1025–1033, 1980a.

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

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

Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, J. Hydrol., 211, 69–85, 1998.

Leisenring, M. and Moradkhani, H.: Snow water equivalent prediction using Bayesian data assimilation methods, Stoch. Env. Res. Risk A, 25, 253–270, 2011.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resour. Res., 43, W07401, doi:10.1029/2006WR005756, 2007.

Margulis, S. A. and Entekhabi, D.: A Coupled Land Surface-Boundary Layer Model and Its Adjoint, J. Hydrometeorol., 2, 274–296, 2001.

Margulis, S. A., McLaughlin, D., Entekhabi, D., and Dunne, S.: Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 field experiment, Water Resour. Res., 42, W01407, doi:10.1029/2001WR001114, 2002.

Molotch, N. P. and Bales, R. C.: Scaling snow observations from the point to the grid element: Implications for observation network design, Water Resour. Res., 41, W11421, doi:10.1029/2005WR004229, 2005.

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

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

Moradkhani, H., Hsu, K., Hong, Y., and Sorooshian, S.: Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting, Geophys. Res. Lett., 33, L12401, doi:10.1029/2006gl026855, 2006.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., and Dmip Participants: Overall distributed model intercomparison project results, J. Hydrol., 298, 27–60, doi:10.1016/j.jhydrol.2004.03.031, 2004.

Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic Data Assimilation with the Ensemble Kalman Filter, Mon. Weather Rev., 130, 103–114, 2002.

Seo, D. J., Koren, V., and Cajina, N.: Real-Time Variational Assimilation of Hydrologic and Hydrometeorological Data into Operational Hydrologic Forecasting, J. Hydrometeorol., 4, 627–641, 2003.

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

Shamir, E. and Georgakakos, K. P.: Distributed snow accumulation and ablation modeling in the American River basin, Adv. Water Resour., 29, 558–570, 2006.

Slater, A. G. and Clark, M. P.: Snow Data Assimilation via an Ensemble Kalman Filter, J. Hydrometeorol., 7, 478–493, 2006.

Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S.: The distributed model intercomparison project (DMIP): motivation and experiment design, J. Hydrol., 298, 4–26, 2004.

Su, H., Yang, Z.-L., Niu, G.-Y., and Dickinson, R. E.: Enhancing the estimation of continental-scale snow water equivalent by assimilating MODIS snow cover with the ensemble Kalman filter, J. Geophys. Res., 113, D08120, doi:10.1029/2007jd009232, 2008.

Su, H., Yang, Z.-L., Dickinson, R. E., Wilson, C. R., and Niu, G.-Y.: Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information, J. Geophys. Res., 115, D10104, doi:10.1029/2009jd013035, 2010.

Su, H., Yang, Z.-L., Niu, G.-Y., and Wilson, C. R.: Parameter estimation in ensemble based snow data assimilation: A synthetic study, Adv. Water Resour., 34, 407–416, doi:10.1016/j.advwatres.2010.12.002, 2011.

Thirel, G., Salamon, P., Burek, P., and Kalas, M.: Assimilation of MODIS snow cover area data in a distributed hydrological model, Hydrol. Earth Syst. Sci. Discuss., 8, 1329–1364, doi:10.5194/hessd-8-1329-2011, 2011.

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters, Water Resour. Res., 39, 1201, doi:10.1029/2002WR001642, 2003.

Vrugt, J. A., Gupta, H. V., Nuallain, B., and Bouten, W.: Real-Time Data Assimilation for Operational Ensemble Streamflow Forecasting, J. Hydrometeorol., 7, 548–565, 2006.

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44, W00B09, doi:10.1029/2007wr006720, 2008.

Weerts, A. H. and El Serafy, G. Y. H.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42, W09403, doi:10.1029/2005wr004093, 2006.

Wu, C.-C. and Margulis, S. A.: Feasibility of real-time soil state and flux characterization for wastewater reuse using an embedded sensor network data assimilation approach, J. Hydrol., 399, 313–325, 2011.

Wu, L., Seo, D.-J., Demargne, J., Brown, J. D., Cong, S., and Schaake, J.: Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction, J. Hydrol., 399, 281–298, 2011.

Zaitchik, B. F. and Rodell, M.: Forward-Looking Assimilation of MODIS-Derived Snow-Covered Area into a Land Surface Model, J. Hydrometeorol., 10, 130–148, 2009.