



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Advances in Water Resources xxx (2005) xxx–xxx

Advances in
Water Resourceswww.elsevier.com/locate/advwatres

Assimilation of snow covered area information into hydrologic and land-surface models

Martyn P. Clark^{a,*}, Andrew G. Slater^a, Andrew P. Barrett^a, Lauren E. Hay^b,
Gregory J. McCabe^b, Balaji Rajagopalan^a, George H. Leavesley^b

^a *Cooperative Institute for Research in Environmental Sciences, Center for Science and Technology Policy Research, University of Colorado, Boulder, CO 80309-0488, United States*

^b *United States Geological Survey, Water Resources Discipline, Denver, CO, United States*

Received 31 May 2005; received in revised form 29 September 2005; accepted 2 October 2005

Abstract

This paper describes a data assimilation method that uses observations of snow covered area (SCA) to update hydrologic model states in a mountainous catchment in Colorado. The assimilation method uses SCA information as part of an ensemble Kalman filter to alter the sub-basin distribution of snow as well as the basin water balance. This method permits an optimal combination of model simulations and observations, as well as propagation of information across model states. Sensitivity experiments are conducted with a fairly simple snowpack/water-balance model to evaluate effects of the data assimilation scheme on simulations of streamflow. The assimilation of SCA information results in minor improvements in the accuracy of streamflow simulations near the end of the snowmelt season. The small effect from SCA assimilation is initially surprising. It can be explained both because a substantial portion of snowmelts before any bare ground is exposed, and because the transition from 100% to 0% snow coverage occurs fairly quickly. Both of these factors are basin-dependent. Satellite SCA information is expected to be most useful in basins where snow cover is ephemeral. The data assimilation strategy presented in this study improved the accuracy of the streamflow simulation, indicating that SCA is a useful source of independent information that can be used as part of an integrated data assimilation strategy.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Snow data assimilation; Stochastic hydrology; Uncertainty

1. Introduction

Remotely sensed snow covered area (SCA) information has been used in hydrologic simulation models in a large number of applications (e.g., [17,18,21–23,30,9,28,3,4,19]). One approach is to use SCA as a hydrologic model input. For example, in the snowmelt runoff model (SRM) the snowmelt equations are applied to the fraction of the basin that is covered in snow (e.g.,

[18]). In this approach snow water equivalent (SWE) is not computed directly by the model, but can be inferred from snow cover depletion curves [9], facilitating streamflow forecasts.

Another approach is to use SCA information to alter the sub-basin distribution of SWE. For example, Turpin et al. [30] replaced modeled SWE with modeled SWE from a different date that had similar fractional SCA to the SCA in a Landsat image. Barrett [4] and McGuire et al. [19] both used rule-based approaches to modify the sub-basin SWE distribution. Barrett [4] updated snow in two ways. In grid cells where the model estimated snow and the satellite did not, SWE was either removed from

* Corresponding author. Tel.: +1 303 735 3624; fax: +1 303 735 1576.

E-mail address: clark@vorticity.colorado.edu (M.P. Clark).

the basin, or SWE was redistributed to areas where the satellite and model agree (generally higher elevations). Barrett added a thin layer of snow to grid cells where the satellite estimated snow and the model did not. Results showed, somewhat predictably, that streamflow decreased when snow was removed from the basin, and that the peak streamflow was delayed when snow was added to higher elevations. In most cases snow updating decreased the skill of streamflow simulations. Mcguire et al. [19] used the following rules: if the modeled SWE was zero and the satellite estimate of fractional SCA was greater than 50%, then a thin layer of snow was added to the grid cell; if the modeled SWE was non-zero and the fractional SCA was less than 50%, then snow was removed from the grid cell. Mcguire et al. [19] showed that their snow updating procedures improved short-term (14-day) forecasts in a subset of the basins studied.

There are two main problems with these rule-based approaches to snow updating. First, rule-based methods simply replace model simulations with satellite observations [this is known as the “direct-insertion” approach to data assimilation]. The direct-insertion approach assumes that satellite observations are perfect and implies that there is no useful information in the model simulations of snowpack. A second problem with these rule-based snow updating methods is that it is difficult to propagate information to other model state variables. For example, Mcguire et al. [19] assume that in cases when the model shows snow and the satellite does not, that water has “left” the basin. In reality, water may linger in soil and groundwater reservoirs before being “seen” as streamflow at the basin outlet. A superior approach is a Kalman-type data assimilation strategy (e.g., [25,26]) that blends satellite products with model simulations so as to exploit the relative strengths in both the model simulations and satellite observations. Moreover, in the Kalman-type data assimilation methods it is fairly straightforward to update other model states (e.g., soil moisture) based on the modeled covariance across model states.

The purpose of this paper is to demonstrate how satellite observations of SCA can be used as part of a Kalman-type data assimilation strategy to improve model simulations of streamflow. Results are based on sensitivity experiments with a fairly simple snowpack/water-balance model. The remainder of this paper is organized as follows. The hydrologic model is described in the next section. The data assimilation strategy is described in Section 3, and the probabilistic approach to model simulation is described in Section 4. Model covariance is discussed in Section 5, and data assimilation experiments are described in Section 6. The paper concludes with discussion of the potential utility of SCA information for model updates in a range of different snow environments.

2. Model description

The hydrologic model configured for this study uses temperature index methods to model snow accumulation and ablation processes (e.g., [24]). The basin water balance and streamflow are modeled using conceptual storage reservoirs. Fig. 1 presents a conceptual diagram of the model, summarizing the main equations and parameters. The simplicity of this model has some advantages in that the effects of model parameters and state updating are easy to understand. The main advantage is that large ensemble simulations can be performed with little computational effort.

2.1. Snow model

SWE and fractional SCA are modeled using sub-grid parameterizations similar to those described by Luce et al. [16] and Liston [14]. The basic premise is that the variability in snow within a grid cell can be described by a parametric probability distribution function (p.d.f.). Consider the two-parameter lognormal p.d.f. used by Liston [14]:

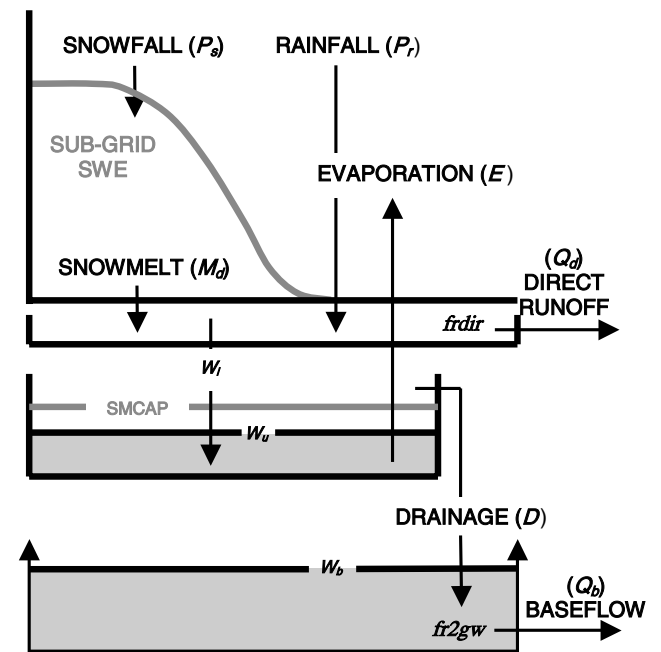


Fig. 1. Conceptual diagram of the snowpack water-balance model configured for this study. A specified fraction of water inputs (rainfall + snowmelt) is extracted as direct runoff (based on parameter fr_{dir}). The soil moisture reservoir is represented as an impervious bucket, and drainage only occurs once it has reached capacity (defined by parameter $smcap$, Eq. (11)). Groundwater storage is infinite, with baseflow parameterized as a function of the groundwater storage (parameter fr_{2gw} ; Eq. (15)). Streamflow is taken as the sum of direct runoff and baseflow (normalized by basin area).

$$f(w) = \frac{1}{w\zeta\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{\ln(w) - \lambda}{\zeta}\right]^2\right\} \quad (1a)$$

with parameters

$$\lambda = \ln(\mu) - \frac{1}{2}\zeta^2 \quad (1b)$$

$$\zeta = \ln(1 + CV^2) \quad (1c)$$

where w is snow water equivalent (mm), μ is the mean of the probability distribution, and CV is the coefficient of variation (parameter scvar).

Luce et al. [16] and Liston [14] both define the p.d.f. in terms of total accumulation, and assume that melt is spatially uniform over a grid cell. In this implementation it is only necessary to model total snow accumulation, D_a (mm), and total melt depth, D_m (mm). Here, $D_a = \mu$ in Eq. (1), and is defined as the sum of individual snowfall events. Likewise, D_m is defined as the sum of individual melt events. Assuming uniform melt, SCA and SWE are estimated by integrating over the p.d.f. defined by D_a and CV (Eq. (1)), that is

$$SCA = \int_{D_m}^{\infty} f(w) dw \quad (2)$$

$$SWE = \int_{D_m}^{\infty} (w - D_m)f(w) dw \quad (3)$$

Note that D_m is the lower limit of integration. Physically speaking, these equations amount to identifying areas of equal SWE amount across a basin (e.g., 3.5% of the basin has SWE of 500 mm; 2.7% of the basin has SWE of 510 mm; and so forth), and computing the weighted sum of SWE amounts (weighted by area). SCA is simply the sum of areas with SWE on the ground. Liston [14] presents analytical solutions to Eqs. (2) and (3).

P.d.f.s computed using different CV parameters are illustrated in Fig. 2a, and a conceptual example of the resultant distribution of snow is provided in Fig. 2b (the lines in Fig. 2b are essentially the cumulative p.d.f.s flipped on their side). The effects of melt are illustrated for the p.d.f.s in Fig. 2c, and for the example snow distribution in Fig. 2d. Conceptually, the shallow portions of the grid cell will melt first, reducing the snow covered area in the grid cell (geometrically, melt causes

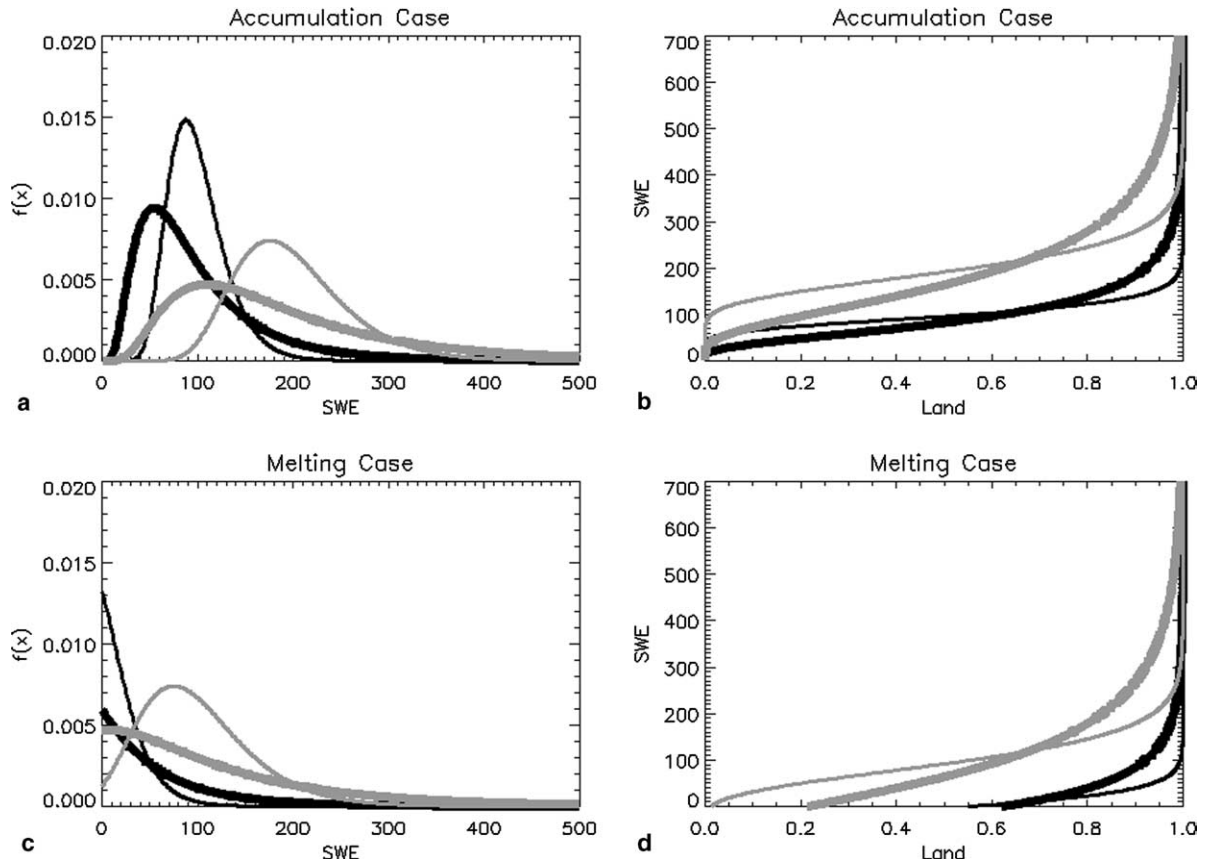


Fig. 2. Various lognormal probability distribution functions (p.d.f.s) describing sub-grid variability in SWE, showing (a) p.d.f.s for mean = 100 mm (dark lines) and mean = 200 mm (light lines), and coefficient of variation parameter = 0.3 (thin lines) and 0.7 (thick lines); (b) conceptual illustration of the distribution of SWE associated with each p.d.f.; (c) effects of melt on the p.d.f.s; and (d) conceptual illustration of the effects of melt on sub-grid parameterization of SWE.

the p.d.f. to shift left (Fig. 2c)). Because the sub-grid variability in SWE is determined based on total accumulation, snow can remain in the grid cell even if total melt is greater than total accumulation.

As just noted, total snow accumulation D_a (mm) is defined as the sum of precipitation from daily snowfall events. Precipitation is modeled as snow if the temperature is below a prescribed threshold; otherwise precipitation is rain:

$$P_s = \begin{cases} P \cdot \text{pbias}, & T_d < \text{tsnow} \\ 0, & T_d \geq \text{tsnow} \end{cases} \quad (4a)$$

$$P_r = \begin{cases} 0, & T_d < \text{tsnow} \\ P \cdot \text{pbias}, & T_d \geq \text{tsnow} \end{cases} \quad (4b)$$

Here P is precipitation (mm d^{-1}), P_s and P_r are precipitation in the form of snow and rain, respectively (mm d^{-1}), pbias is a dimensionless parameter defining the bias in precipitation, T_d is the mean daily temperature ($^{\circ}\text{C}$), and tsnow is a model parameter for the temperature threshold that determines precipitation phase ($^{\circ}\text{C}$). Rain percolates through the snowpack without delay, and is immediately available to the water-balance model.

Snowmelt, M_s (mm d^{-1}), is estimated using temperature index methods described in Rango and Martinec [24]:

$$M_s = \begin{cases} (T_d - \text{tmelt}) \cdot \text{ddpar}, & T_d > \text{tmelt} \\ 0, & T_d \leq \text{tmelt} \end{cases} \quad (5)$$

As before, T_d is the mean daily temperature ($^{\circ}\text{C}$), and ddpar ($\text{mm } ^{\circ}\text{C}^{-1} \text{d}^{-1}$) and tmelt ($^{\circ}\text{C}$) are model parameters. In Rango and Martinec's implementation of this method, ddpar is a parameter that increases through the melt season to account for seasonal changes in solar radiation, snow temperature, and so forth. Here ddpar is estimated based on the mean temperature for the past 30 days (T_{30}):

$$\text{ddpar} = \begin{cases} \kappa(T_{30} - \text{tbase})^{\text{tpowr}}, & T_{30} > \text{tbase} \\ \kappa \text{ tbase}, & T_{30} \leq \text{tbase} \end{cases} \quad (6)$$

where T_{30} tbase are in $^{\circ}\text{C}$ and tpowr is dimensionless. The constant $\kappa = 1$ ($\text{mm } ^{\circ}\text{C}^{-2} \text{d}^{-1}$) is included solely to ensure Eq. (6) is dimensionally correct and is not used as an adjustable parameter. Because the degree-day parameter is frequently set to different values for different times in the melt season [24], Eq. (6) accounts for the seasonality in melt without adding any additional parameters.

The daily depth of snowmelt, M_d (mm d^{-1}) is computed by integrating over the positive portion of the snow depth p.d.f. (see also Eq. (3))

$$M_d = \int_{D_m}^{\infty} (\min(M_s, w - D_m)) f(w) dw \quad (7)$$

which reduces melt over the zero and shallow areas of the grid cell.

The sub-grid snow parameterization must account for situations when new accumulation interrupts melt. In these situations the new snow may fall over the entire grid cell, and then melt rapidly to leave the original distribution of snow before the accumulation event. These situations have historically been handled by keeping track of the new accumulation, and setting snow covered area to 100% until a prescribed percentage of the new accumulation has melted (e.g., 1,13,16,15). In Liston's model, new accumulation is removed from total melt, D_m , until total melt is zero, then new accumulation is added to total accumulation, D_a . This procedure allows the snow covered area to expand and contract over the grid cell during periods when snow is ephemeral. As noted by Liston [14], the procedure seldom permits new accumulation to cover the entire grid cell, which may be unrealistic in some circumstances. However, this procedure preserves the statistical relations between total accumulation and total melt, and SWE and fractional SCA—a feature that is important for the assimilation strategy outlined in Section 3. In this paper we use Liston's method to model transient accumulation events.

2.2. Water-balance model

Rain plus melt from the snow model is used as input to a two-layer water-balance model. A prescribed fraction of rain + melt is extracted for direct runoff, Q_d (mm d^{-1}):

$$Q_d = \text{frdir} \cdot (P_r + M_d) \quad (8)$$

$$W_i = P_r + M_d - Q_d \quad (9)$$

where frdir is a dimensionless model parameter representing the fraction of rain + melt to direct runoff, and W_i is the residual water input to the top layer of the water balance model (mm d^{-1}). As in Section 2.1, P_r is precipitation in the form of rain and M_d is daily snowmelt (both in mm d^{-1}).

The water balance of the top layer is defined by

$$\frac{dW_u}{dt} = W_i - D - E \quad (10)$$

where W_u is the water content of the upper layer (mm), D is the drainage from the upper layer to the lower layer (mm d^{-1}), and E is evaporation (mm d^{-1}). The model uses daily time steps. D is computed as

$$D = \begin{cases} W_u - \text{smcap}, & W_u > \text{smcap} \\ 0, & W_u \leq \text{smcap} \end{cases} \quad (11)$$

where smcap is a model parameter defining the maximum soil moisture capacity (mm). Consequently, drainage only occurs if $W_u > \text{smcap}$. E is computed as

Table 1
Model parameters and parameter bounds

Parameter	Description	Lower bound	Upper bound
pbias	Bias in precipitation	0.5	1.5
tsnow	Temperature threshold for snowfall	−2.0	2.0
tmelt	Temperature threshold for snowmelt	−2.0	2.0
tbase	Base parameter for the melt factor	−20.0	20.0
tpowr	Exponent for the melt factor	0.1	0.9
scvar	Coefficient of variation for sub-grid SWE	0.1	0.9
etpar	Control on the rate of evapotranspiration	0.001	0.1
smcap	Soil moisture capacity	20.0	300.0
frdir	Fraction of rain + melt to direct runoff	0.005	0.1
fr2gw	Fraction of subsurface storage to baseflow	0.005	0.2

$$E = \text{PET} \cdot \frac{W_u}{\text{smcap}} \quad (12)$$

with PET as potential evapotranspiration (mm d^{-1}), which is computed using the Hamon empirical formulation [11]

$$\text{PET} = \text{etpar} \cdot L \cdot \rho_{v(\text{sat})} \quad (13)$$

where L is the length of daylight (in hours) and $\rho_{v(\text{sat})}$ (g m^{-3}) is the saturation absolute humidity, computed using the mean daily temperature (see [8]). Here etpar is an adjustable model parameter.

The water balance of the bottom layer is defined by

$$\frac{dW_b}{dt} = D - Q_b \quad (14)$$

$$Q_b = \text{frtgw} \cdot W_b \quad (15)$$

where W_b is the storage in the lower layer (mm), Q_b is baseflow (mm d^{-1}), and frtgw is a model parameter that defines the fraction of the groundwater storage that drains as baseflow per day (d^{-1}).

Daily runoff, Q (mm d^{-1}) is then

$$Q = Q_d + Q_b \quad (16)$$

Table 1 summarizes model parameters and parameter ranges.

3. Data assimilation strategy

The data assimilation strategy follows the sub-grid SWE parameterization. With the model structure presented in the previous section, differences between remotely sensed and modeled SCA suggests that one or more of three things is inaccurate: (1) snow accumulation; (2) snowmelt; or (3) the variability in sub-grid SWE (see Fig. 2). Thus, SCA information is used to update total accumulation, total melt, and the coefficient of variation parameter. Because snowmelt lingers in soil and groundwater reservoirs, SCA information is also used to update soil and groundwater storages. In this approach SWE is not modified directly, but is altered based on updates to

total accumulation, D_a , total melt D_m and the coefficient of variation parameter, CV (see Eq. (3)).

Our implementation of the ensemble Kalman filter [25] is formalized as follows [10]. Let \mathbf{X}^b be the $m \times n$ matrix of ensemble background model states, that is

$$\mathbf{X}^b = (\mathbf{x}_1^b, \dots, \mathbf{x}_n^b) \quad (17)$$

where $\mathbf{x}_1^b, \dots, \mathbf{x}_n^b$ are vectors of the m model states for each of the n ensemble members. In this study, $n = 100$ ensemble members are generated for $m = 6$ model states, where the model states are (1) total accumulation depth, D_a ; (2) total melt depth, D_m ; (3) the coefficient of variation in sub-grid SWE, CV; (4) the soil moisture content, W_u/smcap ; (5) the water stored in the sub-surface reservoir, W_b ; and (6) the fractional snow covered area, SCA. Section 4 outlines methods for producing ensemble model simulations.

The model error is estimated directly from the ensemble, in which the ensemble mean is used as a index of truth (the notation below follows [10]). The ensemble mean $\bar{\mathbf{x}}^b$ is defined as

$$\bar{\mathbf{x}}^b = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^b \quad (18)$$

The perturbation from the mean for the i th ensemble member is $\mathbf{x}_i^{/b} = \mathbf{x}_i^b - \bar{\mathbf{x}}^b$, and the ensemble of perturbations is defined as

$$\mathbf{X}^{/b} = (\mathbf{x}_1^{/b}, \dots, \mathbf{x}_n^{/b}) \quad (19)$$

An estimate of the $m \times m$ model error covariance, $\hat{\mathbf{P}}^b$, is computed directly from the $\mathbf{X}^{/b}$ ensemble, that is

$$\hat{\mathbf{P}}^b = \frac{1}{n-1} \mathbf{X}^{/b} \mathbf{X}^{/bT} \quad (20)$$

Having $\hat{\mathbf{P}}^b$, it is relatively straightforward to update model states.

The update equation is

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \hat{\mathbf{K}}(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^b) \quad (21)$$

where

$$\hat{\mathbf{K}} = \hat{\mathbf{P}}^b \mathbf{H}^T (\mathbf{H} \hat{\mathbf{P}}^b \mathbf{H}^T + \mathbf{R})^{-1} \quad (22)$$

Here \mathbf{x}_i^b is the m -element background vector of model states for the i th ensemble member, \mathbf{y}_i is the p -dimensional set of observations, \mathbf{H} is the $m \times p$ operator that converts the model state to the observation space, $\hat{\mathbf{K}}$ is the Kalman gain, and \mathbf{x}_i^a is the m -element analysis vector of model states. \mathbf{R} is the $p \times p$ observation error covariance matrix (Eq. (22)). In this case there is a 1:1 correspondence between observations and model states, which avoids the need for interpolation weights in the \mathbf{H} matrix. Because only observations of fractional SCA are available, $\mathbf{H} = [0, 0, 0, 0, 0, 1]$, and $\mathbf{H}\hat{\mathbf{P}}\mathbf{H}^T + \mathbf{R}$ is a scalar representing the sum of model and observation errors. Accordingly, the Kalman gain $\hat{\mathbf{K}}$ is a 6-element column vector in which the covariance between fractional SCA and all other model states, $\hat{\mathbf{P}}\mathbf{H}^T$, is divided by the sum of model and observation errors, $\mathbf{H}\hat{\mathbf{P}}\mathbf{H}^T + \mathbf{R}$ (Eq. (22)).

A critical component of the ensemble Kalman filter is the treatment of observations. In its traditional implementation (Eqs. (21) and (22)), the n elements of \mathbf{y} are sampled from a distribution with zero mean and covariance \mathbf{R} , that is $\mathbf{y}_i' \sim N(0, \mathbf{R})$, providing n sets of observations that are used to update each of the n ensemble members (see [5]). Whitaker and Hamill [32] introduced an alternative form of the ensemble Kalman filter that does not require perturbed observations. This is the ensemble square root Kalman filter and is used in this study. Under this method the ensemble is broken into mean and anomaly portions, which are updated separately:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \hat{\mathbf{K}}(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^b) \quad (23a)$$

$$\mathbf{x}'^a = \mathbf{x}'^b + \tilde{\mathbf{K}}\mathbf{H}(\mathbf{x}'^b) \quad (23b)$$

The anomalies (\mathbf{x}') are updated using a reduced gain ($\tilde{\mathbf{K}}$), given by

$$\tilde{\mathbf{K}} = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\hat{\mathbf{P}}\mathbf{H}^T + \mathbf{R}}} \right)^{-1} \hat{\mathbf{K}} \quad (24)$$

The reduced gain is used because the analysis error covariance is underestimated in the standard ensemble Kalman filter, which in turn can diminish the spread of the model ensemble and promote filter divergence [5,32].

For demonstration purposes, \mathbf{R} is specified in this application to be 0.05 (i.e., 5% error). Obtaining reliable spatially explicit error estimates from satellite radiances is a difficult proposition. The biggest challenge involves unscrambling the radiance from snow from other land cover types in a pixel. These problems are compounded when snow cover is obscured by the vegetation canopy. While the analytical framework presented here is applicable for any level of observation error, the Kalman gain clearly depends on the relative magnitude of observation and model error. The 5% error is used only to

demonstrate the applicability of the ensemble Kalman filter for SCA assimilation—we acknowledge that further work on quantifying observational errors is necessary to improve filter performance in real-world applications.

4. Probabilistic model simulations

The $m \times n$ matrix of ensemble background model states, \mathbf{X}^b , described in the previous section, is generated by running the simple water balance model with an ensemble of model inputs and a corresponding ensemble of model parameters.

Generating the necessary ensembles requires real data. Streamflow simulations in this study were produced for Middle Boulder Creek, at Nederland, Colorado (United States Geological Survey gage identification number 06725500). The basin was modeled as a single unit (i.e., lumped rather than distributed simulations). The sub-grid variability in snow is handled using the p.d.f.s defined in Eqs. (1)–(3).

Middle Boulder Creek is a snowmelt-dominated mountain basin with a drainage area of 93.76 km² (Fig. 3). Ensemble precipitation and temperature estimates produced for the Lake Eldora SNOTEL site using the methods described in Section 4.1 were used as forcing data (see also [6,29]). The Lake Eldora SNOTEL is located within the Middle Boulder Creek drainage (Fig. 3). It is assumed that the ensemble precipitation and temperature estimates for Lake Eldora are representative of the entire Middle Boulder Creek drainage [a large number of surrounding stations were used to produce the Lake Eldora ensembles [6]]. Middle Boulder Creek has streamflow measurements extending until 1995. The probabilistic precipitation and temperature estimates for Lake Eldora were produced starting in 1985. This provides 11 years of data (1985–1995) for which streamflow simulations can be compared against observations.

4.1. Ensemble model forcings

Ensemble forcing data were produced using the geostatistical method introduced by Clark and Slater [6]. The method is based on locally weighted regression, in which spatial attributes from station locations (latitude, longitude, elevation) are used as explanatory variables to predict spatial variability in precipitation and temperature. For each time step, regression models are used to estimate the cumulative distribution function (c.d.f.) of precipitation and temperature at a given point (or grid cell).

The precipitation c.d.f. for a given point (and day) is computed using locally weighted logistic regression to estimate the probability of precipitation (POP) and

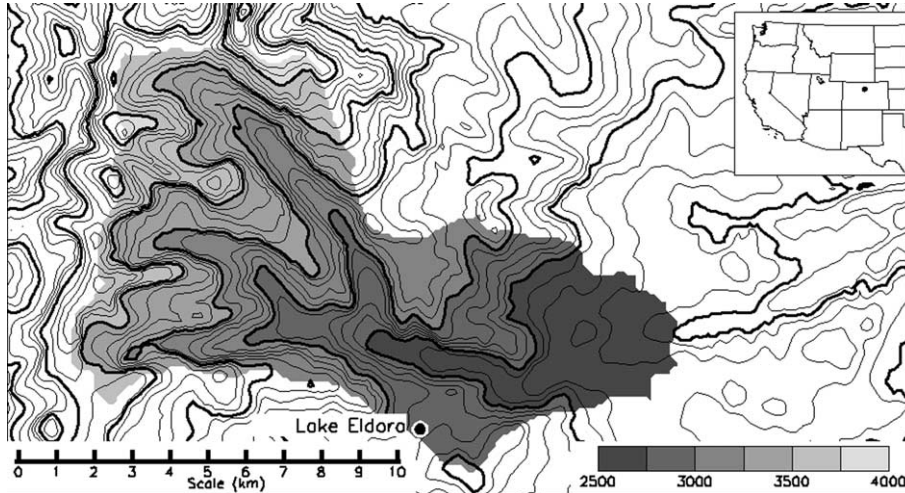


Fig. 3. Contour map of the Middle Boulder Creek drainage. The Basin was modeled as a single units (i.e., lumped rather than distributed simulation).

locally weighted linear regression to estimate precipitation amounts (PCP), with errors (E) in the amounts estimated obtained by locally weighted cross-validated estimates from interpolations at surrounding stations:

$$\text{POP} = \frac{1}{1 + \exp(-Z^T \beta)} \quad (25)$$

$$\text{PCP} = Z^T \beta^a \quad (26)$$

$$E = \left(\frac{\sum_{\text{ista}=1}^{\text{nsta}} W_{\text{ista}} (\text{PCP}_{\text{ista}} - Y_{\text{ista}})^2}{\sum_{\text{ista}=1}^{\text{nsta}} W_{\text{ista}}} \right)^{1/2} \quad (27)$$

where $Z = (1, \text{lat}, \text{lon}, \text{elev})^T$ is the vector of spatial attributes at the target location, β and β^a are the regression coefficients for the logistic and ordinary least squares equations respectively, Y is the precipitation amount at a given station, and W is a vector of weights computed based on the distance from the target station. The regression coefficients (β and β^a) are estimated using local algorithms (see [6] for specific details). The temperature c.d.f. is computed in a similar fashion, although it can be completely defined using Eqs. (26) and (27). Daily precipitation and temperature ensembles are extracted from the estimated c.d.f. at the target location, and the ensembles are re-ordered to preserve the observed space–time correlation structure and correlation among variables (see [7] for more details). Example forcing ensembles are shown for water year 1995 (Fig. 4).

4.2. Ensemble model parameters

The ensemble of model parameters (Table 1) is estimated using Monte Carlo Markov Chains [12,31]. The method proceeds as follows for a given parameter ensemble:

1. Randomly select values for all model parameters from the feasible parameter space (Table 1), use this parameter set to simulate streamflow, and compute the difference between model simulations and observations (summarized using the root mean squared error (RMSE) metric).
2. Randomly select a new parameter set from the feasible parameter space, and compute the RMSE value.
3. Accept the new parameter set if the RMSE value is lower than the RMSE value from the old parameter set.
4. Go back to step 2, and continue until convergence criteria are satisfied or the maximum number of iterations reached (400 iterations in this study).

This process was repeated for 100 ensemble members, producing an ensemble of 100 parameter sets that are consistent with the data. To ensure parameter values were not “over-fit” to the noise in any given forcing ensemble, the RMSE was calculated from streamflow simulations forced with 10 randomly selected forcing ensembles. Fig. 5 shows the RMSE value for each parameter set for the first 100 iterations (later iterations are in gray shades), and the probability distribution function for each parameter after 400 iterations. The range of parameter values (Table 1) was deliberately set to be quite large, partly to test the capability of the Monte Carlo Markov Chain method to select realistic parameter values. Consistent with known precipitation under-catch problems, the method selects pbias values above 1.0 (Fig. 5). Also, note that while some parameter values clearly produce more accurate streamflow simulations, a large number of different parameter sets are consistent with the data.

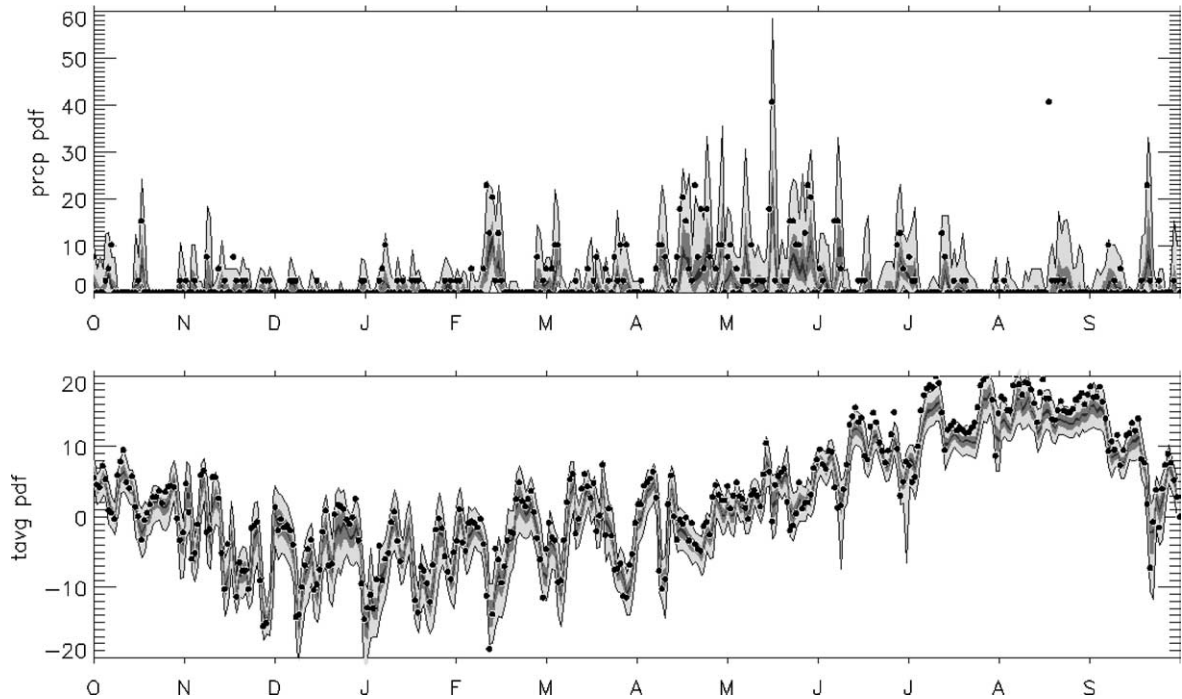


Fig. 4. Example forcing ensembles for the coordinates of Lake Eldora SNOTEL site for water year 1995. The dots are station observations (not used to construct the ensembles), and the shading depicts percentiles from the ensemble (0.05, 0.25, 0.45, 0.55, 0.75, 0.95). While these ensembles were constructed for a specific point, we assume for the purposes of our demonstration that these forcing ensembles are valid for the entire Middle Boulder Creek drainage.

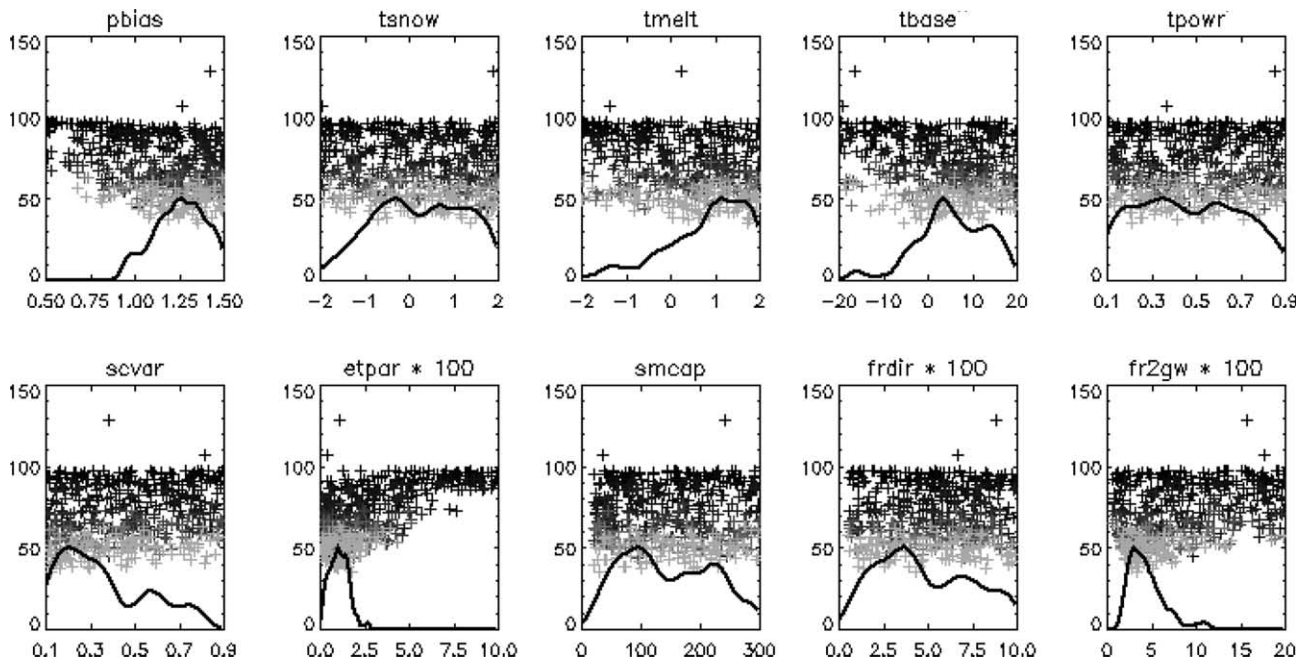


Fig. 5. Root mean squared error for different model parameter sets, as obtained from the Monte Carlo Markov chains. The plus signs depict parameter sets for the first 100 iterations (later iterations are shaded gray), and the lines depict the probability distribution functions for each parameter after 400 iterations. See Table 1 for parameter definitions.

4.3. Implementation

Having estimated an ensemble of model forcings and an ensemble of model parameters, it is now possible to

produce ensemble model simulations that concurrently account for uncertainties in model forcings and uncertainties in the choice of model parameters. To achieve this, we constructed 100 model ensembles in which we

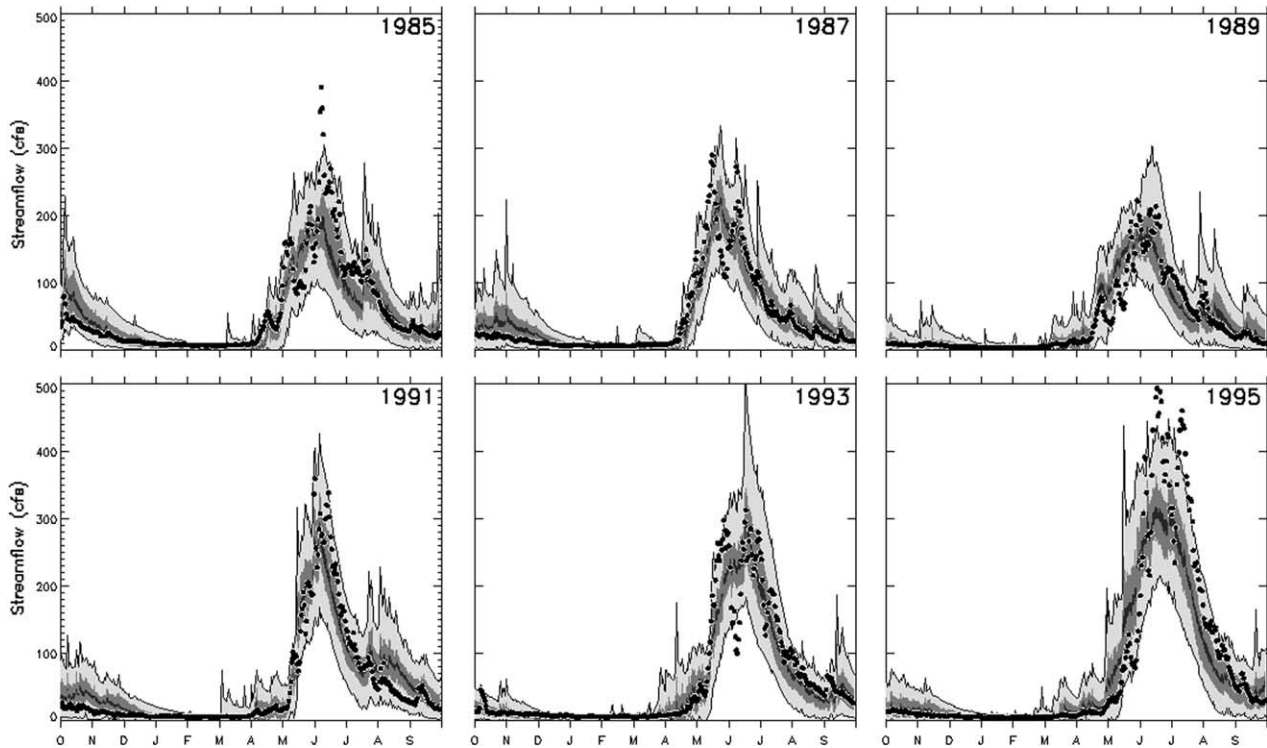


Fig. 6. Ensemble simulations of streamflow for Middle Boulder Creek for the period 1984–1995 (every second year is shown). The dots are streamflow observations, and the shading depicts percentiles from the ensemble (0.05, 0.25, 0.45, 0.55, 0.75, 0.95).

randomly selected (without replacement) one forcing ensemble and one parameter ensemble. This is done only to avoid the extra computational expense of 10,000 ensembles (i.e., all possible combinations of forcing and parameter ensembles). Fig. 6 illustrates probabilistic streamflow simulations for every second year in the period 1985–1995. The dots are streamflow observations from the Middle Boulder Creek gage at Nederland, and the shading depicts percentiles from the simulated ensemble (i.e., 0.05, 0.25, 0.45, 0.55, 0.75, 0.95). The probabilistic estimates encompass the observations fairly well.

5. Model covariance

For assimilation of SCA to have any effect on other model state variables, there must be significant covariance between SCA and the other model states (e.g., see Eq. (22)). To assess covariance across model states, we present in Fig. 7 the correlation (the normalized covariance) between SCA and the other model state variables for each of the 11 years in our study period. The coefficient of variation parameter has generally negative correlations, which implies that ground is exposed earlier (less snow coverage) when snow cover is more variable. Positive correlations for total accumulation imply that more snow accumulation is associated with

more snow coverage, and the negative correlations with total melt imply that more melt is associated with less snow coverage. The correlations between SCA and SWE are positive (note that SWE is not used as a state variable but is derived from total accumulation and total melt).

Turning to the state variables in the water-balance model: correlations are much stronger with the groundwater storage than with soil moisture. This occurs because the soil moisture reservoir is an impervious bucket, in which all water spills to groundwater storage once the bucket has reached capacity (defined by the parameter *smcap*; Eq. (11)). The soil moisture reservoir is filled early in the melt season, which means fractional soil moisture is almost constant across model ensembles. Groundwater storage is represented as a bucket of infinite size, with baseflow dependent on the amount of water in the bucket (see Eqs. (14) and (15)). The negative correlations at the start of the melt season mean that increased snow coverage is associated with decreased groundwater storage—that is, groundwater storage is lower when melt is delayed. The positive correlations at the end of the melt season mean that increased snow coverage is associated with higher groundwater storage—in this case higher snow coverage implies a lengthening of the snow season and maintenance of the high groundwater levels typical of snowmelt periods. Updating state variables in the water-balance model is essential

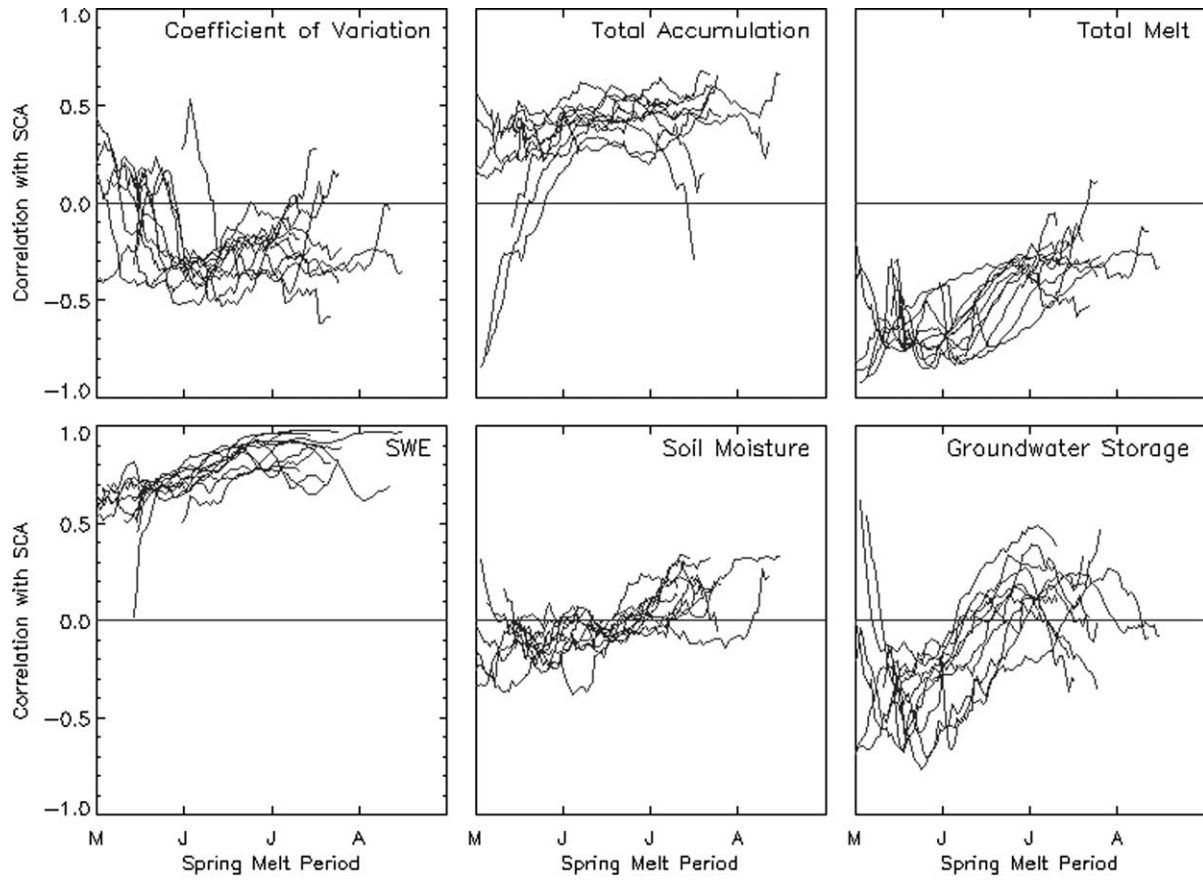


Fig. 7. Correlations among model states for each of the 11 years in the study period. Each year is represented by a separate line.

to keep track of the water in the basin and improve simulations of streamflow.

6. Data assimilation experiments

Synthetic “identical twin” experiments (e.g., [25–27]) were used to evaluate the potential of the proposed data assimilation strategy. In twin experiments the model is first run for the period of record using the best possible forcing data and parameter sets. This serves as the “true” solution and is meant to represent nature [27]. The model is then run with degraded forcing and parameter sets. These simulations, termed “control” or “open loop” model runs, are meant to represent the current situation in which models cannot replicate nature. Finally, model states from the “true” solution are assimilated into the “control” runs. These simulations are termed the “assimilation” model runs. Potential benefits of data assimilation can be evaluated by comparing output from the “control” and “assimilation” runs against model output from the “true” model run. The advantage of twin experiments is that there are synthetic data for all model states, for which observations are often unavailable.

For this study, the twin experiment proceeds slightly differently. It is assumed that each parameter/forcing combination is an equally likely description of nature. The first parameter set and forcing ensemble are selected and the model is run for the 11-year period (1985–1995). This model run is assumed to be truth. Next, the model is run for the other 99-parameter/forcing ensembles. These ensemble simulations are considered as the control run. Finally, the fractional SCA is assimilated from the first ensemble into the control ensemble. This process is repeated for all 100-ensemble members (i.e., each ensemble member is eventually used as “truth”). This experiment provides 1100 synthetic years to evaluate the effect of the assimilation strategy (11 years * 100 ensemble members). Assimilation started on March 1st, to focus on the spring period when snow cover is variable.

Based on the correlations in Fig. 7, assimilation of SCA should have a dramatic effect on model simulations of snowpack and the basin water balance. Fig. 8 presents example results for one twin experiment for water year 1995, showing the control ensemble (left column), the assimilation ensemble (middle column), and the control and assimilation p.d.f.s for July 1st (right column). Fig. 8 shows that the “true” fractional SCA is lower

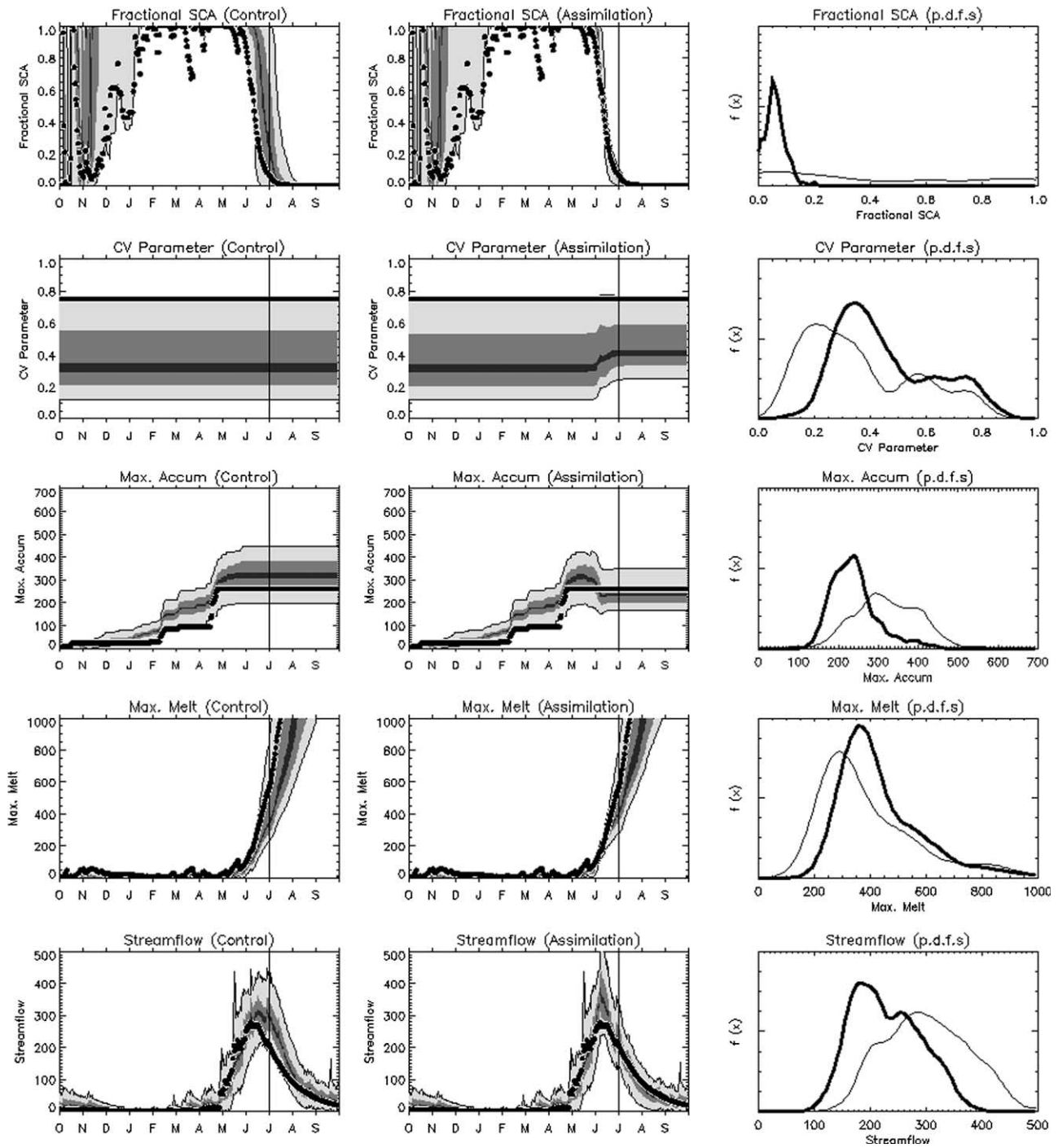


Fig. 8. Example ensemble simulations of model states and streamflow for water year 1995. The left column shows results for the “control” ensemble, the middle column shows results for the “assimilation” ensemble, and the right column shows example p.d.f.s for July 1st (light line is control, dark line is assimilation). Synthetic observations from the “truth” simulation are shown as dots, and the shading depicts percentiles from the ensemble (0.05, 0.25, 0.45, 0.55, 0.75, 0.95). See text for further details.

than most of the control ensembles (top row); that is, bare ground is exposed earlier in the “true” simulation. Early exposure of bare ground may be due to higher variability in sub-grid SWE (i.e., a higher CV parameter), which implies more shallow snow. Consistent with this notion, assimilating fractional SCA from the “true” simulation results in an increase in the CV parameter

(second column of Fig. 8). Increases in the CV parameter are restricted to the month of June when SCA is variable. Bare ground can also be exposed earlier if there is less total snow accumulation in nature than in the control simulations. The assimilation of fractional SCA from the “true” simulation results in a decrease in total snow accumulation (Fig. 8). Finally, bare ground can be

exposed earlier if there is more total melt in nature than in the control simulations. In this example, there are only minor changes in total melt, which may occur as a result of the covariability between melt and other model state variables. The ensemble streamflow simulations (bottom row of Fig. 8) have a better correspondence to the synthetic observations after fractional SCA is assimilated.

Fig. 9 illustrates seasonal cycles in the mean ensemble variance in simulated streamflow (top row), and the RMSE of the ensemble mean streamflow (bottom row), averaged over the 1100 synthetic water years. Both the variance in the ensemble streamflow simulations and the RMSE is slightly lower in the assimilation runs (light line) than in the control simulations (dark line). The reductions in variance and errors are most pronounced during the months of June and July. The relatively small effect of SCA assimilation occurs for two reasons. First, a substantial proportion of spring streamflow occurs before any bare ground is exposed, meaning that assimilating fractional SCA information will always have only a limited effect on ensemble simulations of streamflow. Second, the transition from conditions when 100% of the basin is snow covered to conditions when 100% of

the basin is snow free can occur fairly quickly (e.g., occasionally within a 2-week period), providing a very short temporal window for using fractional SCA information.

7. Summary and discussion

An alternative analytical framework for assimilating fractional snow covered area (SCA) information in hydrologic and land-surface models was presented. The assimilation method uses SCA information in an ensemble Kalman filter to alter the sub-grid probability distribution function (p.d.f.) of snow (defined in terms of total accumulation, total melt, and the variability of SWE). It was demonstrated that assimilating SCA information effectively modifies the sub-grid distribution of SWE, as well as the basin water balance. Assimilation of SCA also modifies the ensemble simulations of streamflow.

The summary statistics presented in this paper demonstrate that for the basin examined in this study assimilating SCA information provides only minor increases in model accuracy. This occurs because a substantial proportion of spring streamflow occurs before any bare ground is exposed, and because the transition from conditions when 100% of the basin is snow covered to conditions when 100% of the basin is snow free often occurs fairly quickly. While these two factors do constrain applications of satellite SCA information in basins worldwide, the relative importance of these two factors will be basin-dependent. We expect satellite SCA information to be most useful in basins where snow cover is ephemeral. The improvements in streamflow simulation demonstrated in this study indicate that fractional SCA is an important source of independent information that is useful as part of an integrated data assimilation strategy.

The results presented in this paper are based on synthetic experiments and are still a few steps away from real-world applications. Andreadis and Lettenmaier [2] recently completed a pilot study assessing the use of the ensemble Kalman filter to assimilate remotely sensed SCA information in the Snake River basin, Idaho, USA. They constructed forcing ensembles by perturbing precipitation inputs with fixed error limits (25%), and specified observation errors at 10%. Comparison with point observations showed the filter significantly reduced errors in model simulations of SWE in the ablation season. Consistent with the results in our study, reduction in errors were minimal in the accumulation season when snow cover is always close to 100%, and the reduction in errors were much lower at high elevations, where snow cover is also more continuous.

There are a number of ways to improve filter performance. First, it is necessary to account for temporal persistence in the model updates (e.g., [25]). Not doing so

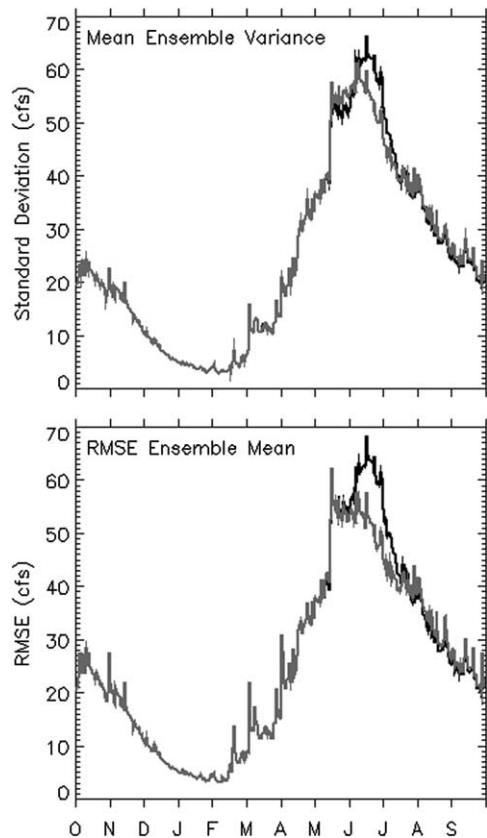


Fig. 9. Mean ensemble variance (top row) and the root mean squared error (bottom row) averaged over 1100 synthetic water years. The control run is depicted by the dark line, and the assimilation run is depicted by the light line.

can be viewed as analogous to repeatedly assimilating the same observation in a single model time step—this will reduce the variance of the ensemble, meaning that the filter will increasingly favor the model predictions. Hence, the filter will operate in a sub-optimal manner. Slater and Clark [29] provide a simple remedy to this issue. Another area that needs more attention is the assumption of normally distributed errors. In many hydrologic systems errors are certainly not normally distributed, which can also lead to sub-optimal model updates. New approaches are emerging that can address more complex error structures (e.g., [33,20]). These issues notwithstanding, the biggest challenge in effectively applying data assimilation techniques lies in quantifying the model and observation error statistics. Too often model and observation errors are either prescribed, or crudely parameterized. It is necessary to attack this error estimation problem with fervor and gusto.

Acknowledgments

This work was primarily supported by the National Weather Service Office of Hydrologic Development (Award NWS4620014). Partial support was also provided by the NOAA GAPP Program (Award NA16GP2806), and the NOAA RISA Program (Award NA17RJ1229). This support is gratefully acknowledged. Three anonymous reviewers provided useful suggestions.

References

- [1] Anderson EA. National Weather Service river Forecast System—snow accumulation and ablation model. NOAA technical memorandum NWS HYDRO-17. Silver Spring, MD: US Department of Commerce; 1973.
- [2] Andreadis KM, Lettenmaier DP. Assimilating remotely-sensed snow observations into a macroscale hydrology model. *Adv Water Resour*, in press. doi:10.1016/j.advwatres.2005.08.004.
- [3] Barrett AP, Leavesley GH, Viger RL, Nolin AW, Clark MP. A comparison of satellite-derived and modeled snow-covered area for a mountain drainage basin. In: Owe M, Brubaker K, Ritchie J, Rango A, editors. *Remote sensing and hydrology 2000* (Proceeding of a symposium held at Santa Fe, New Mexico, USA, April 2000). IAHS Publ. No. 267, 2001. p. 569–73.
- [4] Barrett AP. Integrating remotely sensed snow cover with a distributed hydrologic model. *Hydrol Process*, submitted for publication.
- [5] Burgers G, van Leeuwen PJ, Evensen G. Analysis scheme in the ensemble Kalman filter. *Mon Wea Rev* 1998;126(6):1719–24.
- [6] Clark MP, Slater AG. Probabilistic quantitative precipitation estimation in complex terrain. *J Hydrometeorol*, in press.
- [7] Clark MP, Gangopadhyay S, Hay LE, Rajagopalan B, Wilby RL. The Schaake Shuffle: a method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J Hydrometeorol* 2004;5:243–62.
- [8] Dingman LS. *Physical hydrology*. New Jersey: Prentice-Hall; 1994. 575 p.
- [9] Gomez-Landsea E, Rango A. Operational snowmelt runoff forecasting in the Spanish Pyrenees using the snowmelt runoff model. *Hydrol Process* 2002;16:1583–91.
- [10] Hamill TM. Ensemble-based atmospheric data assimilation. In: Palmer TN, Hagedorn R, editors. *Predictability of weather and climate*. Cambridge Press, 2006, in press.
- [11] Hamon WR. Estimating potential evapotranspiration. *J Hydraul Div, Proc Am Soc Civil Eng* 1961;87:107–20.
- [12] Kuczera G, Parent E. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *J Hydrol* 1998;211:69–85.
- [13] Leavesley GH, Lichty RW, Troutman BM, Saindon LG. *Precipitation-runoff modeling system: users manual*. US Geological Survey Water Resources Investigation Report, 83-4238, 1983.
- [14] Liston GE. Representing subgrid snow cover heterogeneities in regional and global models. *J Climate* 2004;17:1381–97.
- [15] Luce CH, Tarboton DG. The application of depletion curves for parameterization of sub-grid variability of snow. *Hydrol Process* 2004;18:1409–22.
- [16] Luce CH, Tarboton DG, Cooley CR. Sub-grid parameterization of snow distribution for an energy and mass balance snow cover model. *Hydrol Process* 1999;13:1921–33.
- [17] Martinec J. Limitation in hydrological interpretations of the snow coverage. *Nordic Hydrol* 1980;11:209–20.
- [18] Martinec J, Rango A, Roberts R. *Snowmelt runoff model (SRM) users manual*. Geographica Bernensia, Department of Geography, University of Bern; 1994. p. 65.
- [19] Mcguire M, Wood AW, Hamlet AF, Lettenmaier DP. Use of satellite data for streamflow and reservoir storage forecasts in the Snake River basin, ID. *J Water Resour Planning Manage*, 2005, in press.
- [20] Moradkhani H, Hsu K, Gupta H, Sorooshian S. Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter. *Water Resour Res* 2005;41:W05012. doi:10.1029/2004WR003604.
- [21] Rango A. Operational applications of satellite snow cover observations. *Water Resour Bull* 1980;16:1066–73.
- [22] Rango A. Progress in snow hydrology remote sensing research. *IEEE Trans Geosci Remote Sensing* 1986;GE-24:47–53.
- [23] Rango A. Spaceborne remote sensing for snow hydrology applications. *Hydrol Sci J* 1996;41:477–93.
- [24] Rango A, Martinec J. Revisiting the degree-day method for snowmelt computations. *Water Resour Bull* 1995;31:657–69.
- [25] Reichle RH, McLaughlin DB, Entekhabi D. Hydrologic data assimilation with the ensemble Kalman filter. *Mon Wea Rev* 2002;130(1):103–14.
- [26] Reichle RH, Walker JP, Koster RD, Houser PR. Extended versus ensemble Kalman filtering for land data assimilation. *J Hydrometeorol* 2002;3(6):728–40.
- [27] Reichle RH, Koster RD. Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J Hydrometeorol* 2003;4(6):1229–42.
- [28] Rodell M, Houser PR. Updating a land surface model with MODIS-derived snow cover. *J Hydrometeorol* 2004;5:1064–75.
- [29] Slater AG, Clark MP. Snow data assimilation via an ensemble Kalman filter. *J Hydrometeorol*, in press.
- [30] Turpin O, Ferguson R, Johansson B. Use of remote sensing to test and update simulated snow cover in hydrologic models. *Hydrol Process* 1999;13:2067–77.
- [31] Vrugt JA, Gupta HV, Bouten W, Sorooshian S. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour Res* 2003;39(8). doi:10.1029/2002WR001642.
- [32] Whitaker JS, Hamill TM. Ensemble data assimilation without perturbed observations. *Mon Wea Rev* 2002;130(7):1913–24.
- [33] Zupanski M. Maximum likelihood ensemble filter. Theoretical aspects. *Mon Wea Rev* 2005;133:1710–26.