

## Snow Data Assimilation via an Ensemble Kalman Filter

ANDREW G. SLATER AND MARTYN P. CLARK

*Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado*

(Manuscript received 16 March 2005, in final form 11 August 2005)

### ABSTRACT

A snow data assimilation study was undertaken in which real data were used to update a conceptual model, SNOW-17. The aim of this study is to improve the model's estimate of snow water equivalent (SWE) by merging the uncertainties associated with meteorological forcing data and SWE observations within the model. This is done with a view to aiding the estimation of snowpack initial conditions for the ultimate objective of streamflow forecasting via a distributed hydrologic model. To provide a test of this methodology, the authors performed experiments at 53 stations in Colorado. In each case the situation of an unobserved location is mimicked, using the data at any given station only for validation; essentially, these are withholding experiments. Both ensembles of model forcing data and assimilated data were derived via interpolation and stochastic modeling of data from surrounding sources. Through a process of cross validation the error for the ensemble of model forcing data and assimilated observations is explicitly estimated. An ensemble square root Kalman filter is applied to perform assimilation on a 5-day cycle. Improvements in the resulting SWE are most evident during the early accumulation season and late melt period. However, the large temporal correlation inherent in a snowpack results in a less than optimal assimilation and the increased skill is marginal. Once this temporal persistence is removed from both model and assimilated observations during the update cycle, a result is produced that is, within the limits of available information, consistently superior to either the model or interpolated observations.

### 1. Introduction

In a typical streamflow forecast system, observationally based data is used to force a hydrologic model up to the start of the forecast period, after which output from numerical weather prediction models or historical climate traces are used to project the model into the future. Inherent to such a forecast system there is uncertainty in the structure, parameterizations, and parameters of the hydrologic model; uncertainty in the model forcing data; and uncertainty in the specification of model initial conditions. In snow-dominated catchments, the accuracy of streamflow forecasts depends, in large part, on the accuracy of snowpack estimates at the start of the forecast period (Clark and Hay 2004). This is due to the fact that almost all streamflow in late spring and summer is derived from meltwater of the winter snowpack. Hence, this study seeks to obtain a

more accurate specification of snowpack initial conditions at the start of the forecast; in turn, this will lead to more accurate streamflow forecasts at multiple lead times (days to months).

Recently, much attention has been devoted to the ability of data assimilation to improve estimation of model simulations of the land surface state (e.g., Houser et al. 1998; Reichle et al. 2002a,b). Previous studies have focused largely on assimilation of soil moisture observations into land surface models, with the overall objective of increasing the skill of weather and climate forecasts. These techniques can be used for snow updating in hydrologic models and, when implemented effectively, may result in considerable increases in the skill of streamflow forecasts.

Several prior studies have looked into the problem of assimilating observations of snow into land surface and hydrologic models; these are reviewed below. The two quantities that have been used as update variables are snow-covered area (SCA) and snow water equivalent (SWE). At the simpler end of the spectrum, direct insertion has been employed for updating SCA (the model estimate is simply replaced with the observa-

---

*Corresponding author address:* Andrew G. Slater, National Snow and Ice Data Center, CIRES, University of Colorado, Boulder, CO 80309-0449.  
E-mail: aslater@cires.colorado.edu

tion). In some cases this method was found to improve SWE simulations (Liston et al. 1999; Rodell et al. 2004) but in others, while overall SWE might have improved, streamflow simulations were degraded by using this technique (A. P. Barrett 2005, personal communication). Rodell and Houser (2004) used the satellite-derived Moderate Resolution Imaging Spectroradiometer (MODIS) SCA product in a rule-based assimilation scheme to alter the SWE in a land surface model. They found that results compared more favorably to observations in both coverage and time series, but that the assimilation could unduly influence the snowpack mass balance as SCA contains no information about SWE. A further limitation of rule-based or direct insertion assimilation is that it generally assumes the observations ingested by the model are perfect and error free. Statistical or optimal interpolation (OI) has been used as a snow assimilation tool by Brasnett (1999) to produce an analysis of snow depth within a simple accumulation and melt model. The result was seen to be more skillful than climatology in all tested cases. In OI, the error structure of the observations and model are assumed a priori, and the errors remain constant when new observations are assimilated.

Methods of assimilation that are more adaptive, such as the Kalman filter, do not make fixed assumption about errors—rather they update them during the simulation. The extended Kalman filter (EKF) relies on an adjoint and tangent linear model to propagate the error covariance of model state variables forward in time. Development of this linearized approximation of the original model forms the crux of the method. For snow updating Day (1990) and Sun et al. (2004) have employed EKFs. Although Day found that updating SWE improved ensemble streamflow forecasts, this method is not commonly used operationally both because of the time required for implementation and the limited number of basins for which the method has been tested. Sun et al. (2004) implemented a one-state EKF into a land surface model for snow updating. In synthetic experiments, Sun et al. were able to show improvements in their simulations as the filter chose the optimal balance between the observational errors and model errors.

An attractive alternative is the ensemble Kalman filter (EnKF). The EnKF (Evensen 1994, 2003) obtains error estimates from the variance of an ensemble of model simulations. Thus no adjoint or linearized model is needed for error estimation, and the method provides great versatility as any number of variables can be included in the update procedure. However, obtaining a suitable ensemble that covers all areas of uncertainty in the model system is not a trivial exercise.

Many of the assimilation studies to date have concentrated on technique development and, as such, have used synthetic or “twin” experiments (Reichle et al. 2002a,b; Reichle and Koster 2003; Sun et al. 2004). Under such an experiment, the model is forced with a deterministic set of data that is to be considered “truth.” This forcing data is then degraded and model simulations undertaken; these are sometimes called “open loop” simulations. The degraded simulations then undergo assimilation and the results can be compared to the truth. This study builds on the previous work, but looks toward an operational setting by using real data for both the model forcing and observations.

The overall goal of this study is to develop and apply snow data assimilation methods that improve streamflow forecasts within a distributed model framework. Since the parameters in most operational streamflow forecasting models are calibrated against observed streamflow data, the optimal model SWE may be quite different from the actual (but unknown) SWE in a given model subbasin or grid box. Hence, our objective is not necessarily to improve estimates of actual SWE; rather, we seek to use SWE observations to update the model SWE initial conditions in a way that is consistent with the model’s calibrated parameters. In this paper a method is developed to attain this objective. The method is evaluated using withholding experiments where actual SWE measurements have been transformed to model space.

In the following sections we describe the application of data assimilation to the problem of snow mass estimation using the SNOW-17 model in the Upper Colorado River basin. Section 2 of the paper describes the assimilation methods, section 3 describes the data used in the study, and section 4 gives an overview of the SNOW-17 model and experimental setup. Results and discussion of various simulations are given in section 5, with conclusions given in sections 6.

## 2. The ensemble Kalman filter

The main assimilation tool that we apply here is the ensemble Kalman filter (EnKF), which was first proposed by Evensen (1994). We include only a brief description of the method here—further details are available in Evensen (2003) and Hamill (2006). The Kalman filter is a data assimilation method that produces optimal weighting between a modeled and observed state given estimates of the errors in the model and observations. The Kalman filter is adaptive in that knowledge of errors is updated as the model is integrated forward in time. The advantage of the EnKF over the traditional and/or extended Kalman filters is that it is not limited by the need for a linearized approximation

of the model for the purpose of error propagation, which may be difficult to construct in cases of strongly nonlinear dynamics. The workings of the EnKF are given below.

Given the vector  $\mathbf{x}$  of model state variables (e.g., SWE, snow temperature, etc.), we perform three basic steps for snow updating using the EnKF:

- 1) Run the full nonlinear model forward in time with an ensemble of forcings to obtain a forecast (or “background”) estimate of our model states ( $\mathbf{x}_{f^*}$ ) at time  $t$  for all ensemble members. The ensemble of model forcings is constructed such that the variance of the ensemble is proportional to the estimated errors in the model forcings (see Clark and Slater 2006).
- 2) Next, compute the Kalman gain ( $\mathbf{K}$ ), which represents the optimal weighting between the error covariance of the model states ( $\mathbf{P}_t$ ) and the errors of the observed state ( $\mathbf{R}$ ). In the case of the EnKF,  $\mathbf{P}_t$  is derived from the variance of the ensemble members, whereas in an extended Kalman filter, a linearized model would provide this matrix. The  $\mathbf{H}$  matrix operator usually equates the model states to the observed variables. In our study, equating model and observed states was done as a preprocessing step via a measurement model, with  $\mathbf{H}$  simply indicating which states were observed in the update:

$$\mathbf{K} = \mathbf{P}_t \mathbf{H}^T (\mathbf{H} \mathbf{P}_t \mathbf{H}^T + \mathbf{R})^{-1}. \quad (1)$$

- 3) Finally, obtain the best estimate or “analysis” of the model state ( $\mathbf{x}_t$ ). This is achieved by updating the forecast model state ( $\mathbf{x}_{f^*}$ ). The update is the gain-weighted difference between the observed states ( $\mathbf{z}_t$ ) and the transformed model forecast state ( $\mathbf{H} \mathbf{x}_{f^*}$ ):

$$\mathbf{x}_t = \mathbf{x}_{f^*} + \mathbf{K}(\mathbf{z}_t - \mathbf{H} \mathbf{x}_{f^*}). \quad (2)$$

We then integrate the model forward in time from the analysis until the next update cycle (i.e., start at step 1 again). The filter assumes the model is unbiased.

A superior alternative to this traditional EnKF is the use of the square root ensemble Kalman filter (EnSRF) (Whitaker and Hamill 2002; Hamill 2006). Under this method, the ensemble is broken into mean and anomaly portions, and updating is performed separately for the ensemble mean and anomalies:

$$\bar{\mathbf{x}}_t = \bar{\mathbf{x}}_{f^*} + \mathbf{K}(\mathbf{z}_t - \mathbf{H}\bar{\mathbf{x}}_{f^*}), \quad (3)$$

$$\mathbf{x}'_t = \mathbf{x}'_{f^*} - \tilde{\mathbf{K}}\mathbf{H}(\mathbf{x}'_{f^*}). \quad (4)$$

The ensemble mean ( $\bar{\mathbf{x}}$ ) is updated with the traditional gain equation given above [Eq. (2)], while the anomalies ( $\mathbf{x}'$ ) are updated with a reduced gain ( $\tilde{\mathbf{K}}$ ) given by

$$\tilde{\mathbf{K}} = \left( 1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}}} \right)^{-1} \mathbf{K}. \quad (5)$$

The reduced gain, used as the analysis error covariance, is underestimated in the standard EnKF, which in turn can diminish the spread of the ensemble and promote filter divergence (Burgers et al. 1998; Whitaker and Hamill 2002).

While the above equations may appear relatively simple to implement, the crux of the EnKF or EnSRF is ensuring that the ensemble encompasses the full uncertainty of the modeling system and that observational errors can be appropriately determined. These are not trivial tasks in practice, and, as such, errors are commonly prescribed. In this study we explicitly derive error estimates of both forcing and assimilated observations.

### 3. Data

#### a. Data sources and quality control

This study relies on data from the National Weather Service (NWS) Cooperative (COOP) network of weather stations, and the Natural Resources Conservation Service (NRCS) network of snowpack telemetry (SNOTEL) sites. These are the primary observing networks in western Colorado (and throughout the western United States). The COOP stations are predominantly at low elevations, and the SNOTEL stations are predominantly at high elevations (Fig. 1). The most common data elements at COOP stations are precipitation, maximum temperature, minimum temperature, snowfall, and snow depth. The most common data elements at SNOTEL stations are precipitation, maximum temperature, minimum temperature, average temperature, and snow water equivalent (snowfall and snow depth are rarely measured at SNOTEL sites). Data at both SNOTEL and COOP sites are available on a daily basis.

#### b. Generating ensemble forcing data

In our implementation of the ensemble Kalman filter we estimate model error by forcing SNOW-17 with an ensemble of inputs. The ensemble of inputs (precipitation and temperature) produces an ensemble of model states at each time step. The covariance between the ensemble of states provides an estimate of model error (see section 2 for more details). This approach to error estimation clearly ignores model errors associated with the choice of model parameters and inadequacies in model physics; however, other studies have shown that

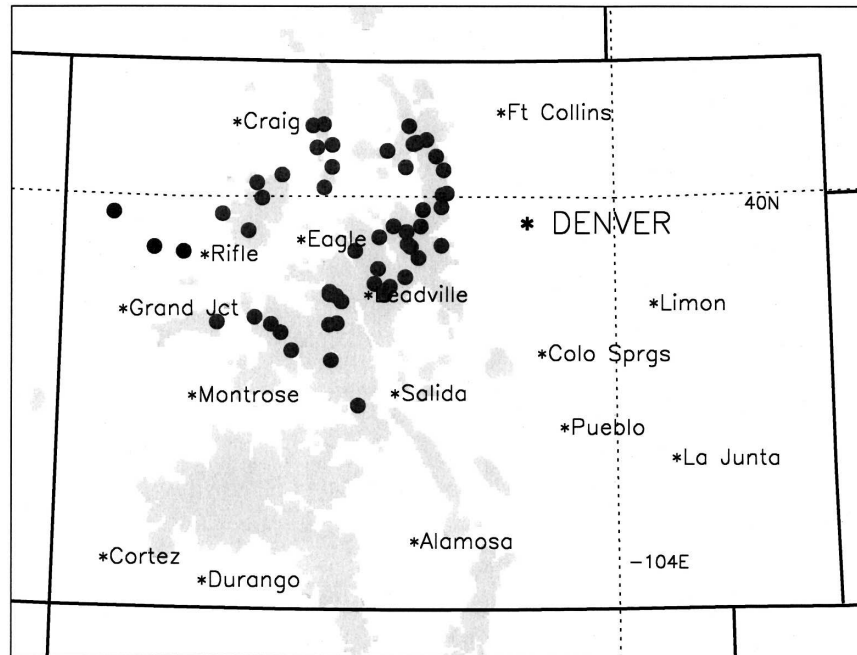


FIG. 1. The state of Colorado: locations of the 53 SNOTEL stations used in this study are marked with dark circles. Areas over 3000-m elevation are shaded light gray.

uncertainties in model forcing data often dominate (e.g., Carpenter and Georgakakos 2004).

In reality observed forcing data is not available for all subbasins or grid boxes; thus a method of estimating the forcing for each computational unit is needed. Below we outline an interpolation method used for obtaining ensemble forcing data, though it is not the focus of this paper. Any method that will provide spatially and temporally coherent forcing fields, *complete with error estimates*, can be used. Previous ensemble simulations using hydrologic and land surface models have specified the error for stochastic ensemble generation (e.g., Reichle et al. 2002a,b; Carpenter and Georgakakos 2004). The geo-statistical method introduced by Clark and Slater (2006) explicitly computes an error field via cross validation; thus we apply it here. 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. Precipitation is generated in a two-step process where logistic regression is used to detect the precipitation occurrence; then ordinary least squares (OLS) regression is used to determine the amount. Spatially explicit temperature estimates are obtained by the OLS approach. Error estimates, which are used to generate probabilistic ensembles, are gained by cross validation of the regression expected values. A complete step-by-

step explanation of the method is available in Clark and Slater (2006).

### c. Generating ensemble observations of snow water equivalent

The ensemble Kalman filter requires observations of snow water equivalent, complete with error estimates [see Eq. (1) in section 2]. These were produced using the precipitation interpolation model—that is, the presence/absence of snow was estimated using locally weighted logistic regression, SWE was estimated using locally weighted OLS, and the error in the SWE interpolation was estimated through cross validation of SWE estimates from surrounding stations. However, we make some minor but important modifications to the interpolation method.

First, instead of performing the interpolation on actual SWE amounts, we interpolate the SWE expressed as a  $Z$  score of each station's long-term record of non-zero SWE values. The nonzero portion is used, as we have already relied on logistic regression to determine SWE occurrence at the interpolate location (e.g., the target model grid cell or subbasin). In turn, the  $Z$  score can be interpreted as a percentile, where the percentile is taken as the cumulative probability of SWE at a given station ( $\times 100$ ). This modification is appropriate because, although SWE can vary significantly from year to year, the spatial distribution of snow remains rela-

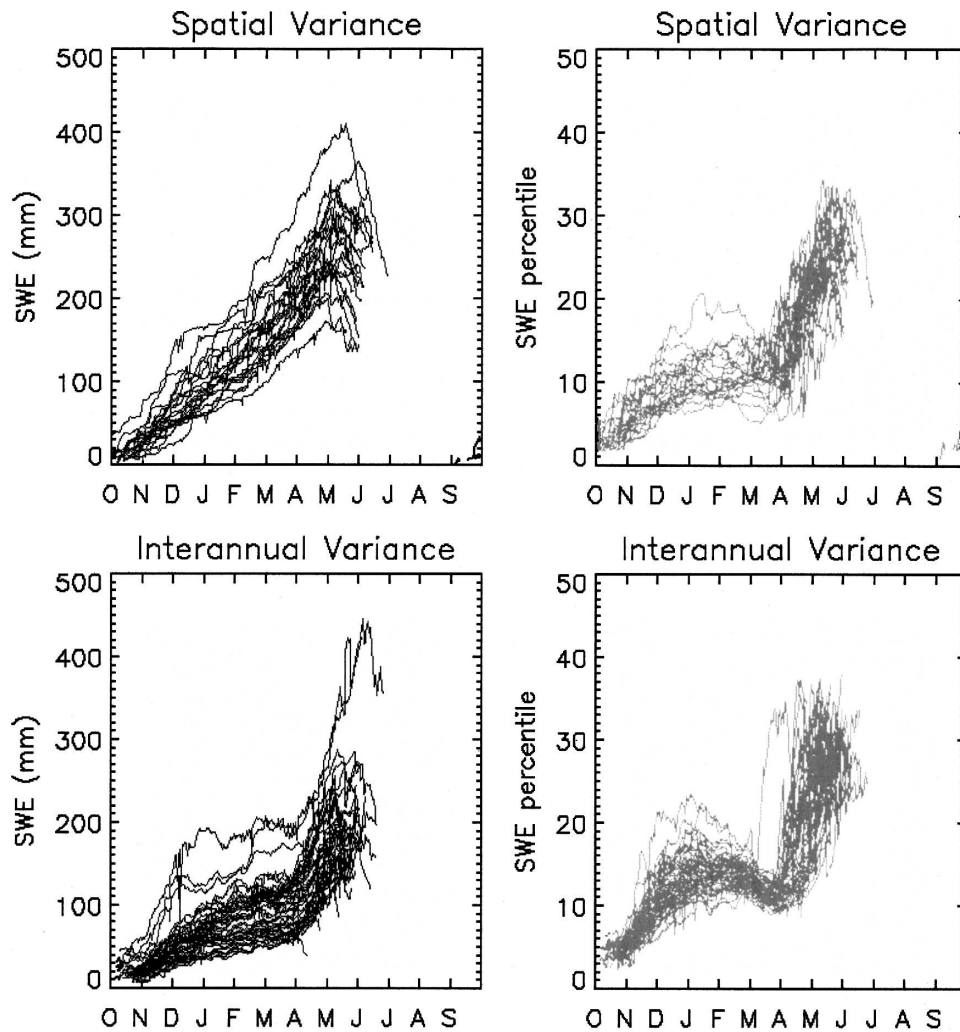


FIG. 2. Spatial and interannual variance of absolute SWE and the SWE percentiles. For the spatial case the standard deviations are computed separately for each day; each line depicts a time series of spatial standard deviations for a given water year. For the interannual case, the standard deviations are computed separately for each station, such that each line depicts the standard deviation through the water year for a given station. Note the relatively lower spatial variance of SWE percentiles over the accumulation season compared to absolute SWE values.

tively constant (see, e.g., Cline et al. 1998; Liston 2004). Figure 2 illustrates the spatial and interannual standard deviation in accumulated SWE and the spatial and interannual standard deviation in SWE percentiles. As expected, the spatial and interannual variability in SWE is highest during spring, when SWE at SNOTEL sites melts out at different rates. However, note that the spatial variability ( $\sigma_{sp}$ ) in SWE is almost twice as large as the interannual variability ( $\sigma_{ia}$ ), but the spatial variability in SWE percentiles is lower than the interannual variability in SWE percentiles (Fig. 2). For example, in March,  $\sigma_{sp}$  for SWE is between 100 and 200 mm for all but three stations, but  $\sigma_{ia}$  is 40–140 mm for all but three stations, with most being below 100 mm. Conversely,

for SWE percentiles over the January to March period,  $\sigma_{sp}$  is 6%–15% while  $\sigma_{ia}$  is 10%–20% for all but a couple of stations. Consequently, our logic is that spatial interpolation is easier for more spatially homogeneous quantities (as is the case with SWE percentiles). Or, put differently, interpolation is more difficult for spatially variable quantities because it is unlikely that the greater heterogeneity is explained by the available topographic information. Additionally, our attempts to interpolate raw SWE values were less accurate (not shown) than using percentiles.

Second, as a consequence of the first modification, we can produce our SWE estimates in “model equivalent” space rather than attempt to determine what real-

world SWE is. Here the interpolated SWE  $Z$  scores at a given model grid box/point are simply back-transformed to model equivalent space using the distribution of modeled SWE for that grid box from a multiyear simulation. Recall that our aim is to use snow assimilation to ultimately improve streamflow, not SWE specifically. The advantages of performing our experiments in the model equivalent space are 1) the calibration removes any bias in the model or forcing data (a necessary condition of the EnKF) and 2) transformation of SWE into model space reduces the impacts of differences in physiographic characteristics between the SNOTEL sites and the model; that is, our estimates of SWE are consistent with the calibrated model parameters, regardless of the physiography in a given model computational unit (e.g., grid box or sub-basin). Note also that the model we use, SNOW-17, is a conceptual snow model that does not have any vegetation; accounting for differences in SWE that reflect different land cover types and so forth will require a more sophisticated physically based model and is well beyond the scope of the current study. Operating in model space makes it possible for us to back-transform the interpolated SWE percentiles at uninstrumented locations. In essence, these modifications can be considered part of the measurement model within the assimilation. Finally, because we operate within the framework of a model calibrated to streamflow, assimilation of actual SWE measurements would be inconsistent with the model and could potentially degrade streamflow forecasts.

The process of generating ensembles of model-equivalent SWE for assimilation is summarized below and illustrated in Fig. 3.

- 1) Collect the current observations of SWE from the observing network. Perform the logistic regression and stochastically generate snow presence or absence of snow for each ensemble member.
- 2) Convert the observed SWE values at each location into their respective  $Z$  score, that is, the current station value less its long-term nonzero mean SWE, all divided by the standard deviation of nonzero SWE at the station.
- 3) Interpolate the  $Z$  scores to a model grid box/point using OLS regression, that is,

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon, \quad (6)$$

where  $Z$  is the interpolated  $Z$  score,  $(\beta_0, \dots, \beta_n)$  are the regression coefficients,  $(x_1, \dots, x_n)$  are the topographic attributes at the model grid box/point, and  $\varepsilon$  are the errors that are assumed to be normally distributed with mean of zero and variance of  $\sigma^2$ . In

Eq. (6), the  $\beta$  coefficients are computed using station data (topographic attributes at station locations are the explanatory variables and  $Z$  scores of SWE at station locations is the dependent variable). The variance in the error term is determined through cross validation. Ensembles of  $Z$  scores are then generated by sampling from a normal distribution defined with mean of  $Z$  and variance  $\sigma^2$ . Sampling is performed using the ‘‘Schaaake shuffle’’ method, which has been shown to preserve the space–time variability in station time series (see Clark et al. 2004 for details).

- 4) Convert each ensemble  $Z$ -score value to a model-equivalent SWE value that will be used in the assimilation. This is done using a ‘‘grand’’ cumulative distribution function (cdf) of SWE from a 19-yr, 100-member ensemble model hindcast simulation at the new location using the unbiased, calibrated model and interpolated stochastic forcing. The percentile corresponding to the ensemble  $Z$  score (obtained from the cumulative probability from a theoretical normal distribution,  $\times 100$ ) is matched with the percentile from the grand cdf from the model hindcast. The model SWE value that corresponds to the  $Z$ -score percentile is taken as the ‘‘model equivalent’’ SWE that is used for assimilation.

The SWE observations at each SNOTEL site were withheld from the interpolation to that site for the purposes of testing our method. As we are operating entirely within the space of the calibrated model, the actual observations were also transformed to model space at that site, so as to provide a measure of truth, and all the simulations are evaluated against this quantity. This transformation was achieved by simply matching the percentiles of nonzero SWE. However as the observations are deterministic in nature, we assume that the mean of the ensemble is the best estimator at any time and use the daily mean of the ensemble to construct the cdf from which we resample the observations. This avoids spurious values, such as the most extreme ensemble member, from being introduced into the observations. This process is shown in Fig. 3 (part 6).

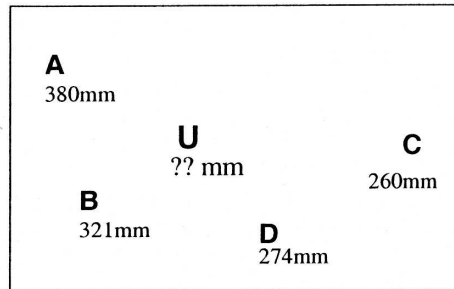
#### 4. Model and experimental setup

##### a. Snow model

The model used in this study is the same as is currently in operational use with the NWS for streamflow forecasting, that is, SNOW-17. A full description is available in Anderson (1973); thus we provide only a brief one here.

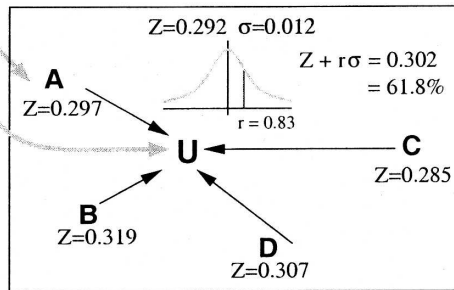
SNOW-17 is a relatively simple one-dimensional con-

1) Obtain spatial set of SWE from SNOTEL stations (e.g. A,B,C,D etc.)



2) Convert SWE to a Z-score from each station's 19 year observational record

3) Interpolate the Z-scores to an uninstrumented point (U) via regression methods; get error ( $\sigma$ ) from cross validation. Sample from the PDF to get ensemble of percentiles



4) Convert each ensemble percentile to an unbiased model equivalent value of SWE by obtaining the corresponding percentile value in the model hindcast

5) Assimilate model equivalent SWE values into the ensemble simulations

6) Point U happens to be a SNOTEL station, which was withheld from the interpolation. Convert the observed value at U to a model equivalent for a measure of "truth" by equating percentiles.

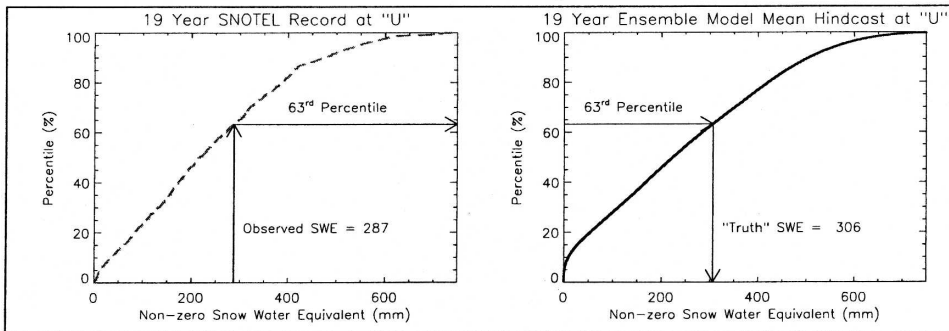
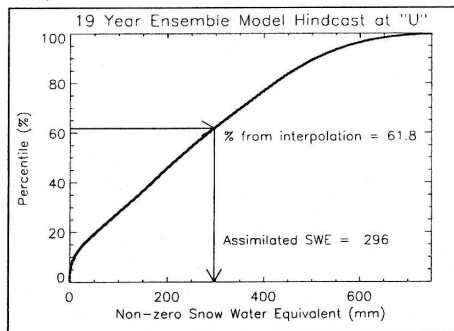


FIG. 3. The method of gaining the interpolated SWE amounts for assimilation and validation. Logistic regression determines the presence/absence of snow for an ensemble member. For cases where snow is present, the above method is used to determine ensemble SWE values. The quantity interpolated is the SWE Z score. Once an interpolated Z score and cross-validated error ( $\sigma$ ) are obtained, we synthesize an ensemble of SWE values by randomly sampling from the SWE pdf. The case shown is for a single ensemble member, with a randomly generated standard-normal  $N[0, 1]$  value of 0.83; thus  $Z + 0.83\sigma = 0.302$ , which gives a Gaussian cumulative probability percentile of 61.8%. Truth values are obtained by converting the actual observed SWE values into model equivalent space.

ceptual model and as such is designed with calibration in mind. Temperature and precipitation are the only required forcing inputs and the model runs at an hourly time step. Snow is modeled as a bulk layer that includes a water holding capacity. Melt is computed differently under rain-on-snow event and nonrain periods. During nonrain periods surface melt is a function of temperature and is computed as

$$M = M_f(T_a - M_{\text{base}}), \quad (8)$$

where  $M$  is the melt rate per unit area and  $M_f$  is the melt factor;  $T_a$  is the air temperature and  $M_{\text{base}}$  a base temperature that is a model parameter. The melt factor is given by

$$M_f = \frac{M_{f\text{MAX}} + M_{f\text{MIN}}}{2} + \sin\left(\frac{n2\pi}{366}\right)\left(\frac{M_{f\text{MAX}} - M_{f\text{MIN}}}{2}\right), \quad (9)$$

where

$M_{f\text{MAX}}$  = maximum melt factor parameter (assumed to occur on 21 June),

$M_{f\text{MIN}}$  = minimum melt factor parameter (assumed to occur on 22 December),

$n$  = day number beginning with 21 March.

During rain-on-snow events, an energy balance approach is applied that requires a calibrated parameter to determine the extent of turbulent transfer.

In total, SNOW-17 has 19 state variables related to mass and energy content. The temperature of the snowpack is modeled as a function of a calibrated antecedent temperature index. The purpose of tracking snow heat storage is for determining the amount of energy required to bring about an isothermal (or “ripe”) snowpack, which means that surface melt water or rainwater will contribute to liquid water storage or outflow rather than refreeze within the snow. Refreezing water releases latent heat and warms the snowpack. The maximum liquid water holding capacity of the snow is a calibrated parameter while the transmission of the liquid is lagged and attenuated according to the ratio of liquid to solid mass.

Twenty-two parameters are required by SNOW-17 (Table 1). Eleven of these define the areal depletion curve (ADC), which determines the extent of snow cover versus bare ground over a particular area. The ADC is important as melt and rain-on-snow are a function of the per unit area of snow. The covered area is computed as the ratio of current SWE relative to the minimum of either the maximum SWE since accumulation or a preset maximum SWE that implies 100%

TABLE 1. Parameters of the SNOW-17 model.

Name	Description
SCF	Factor to correct for precipitation gauge deficiency
$M_{f\text{MAX}}$	Maximum nonrain melt factor (21 June) (mm °C <sup>-1</sup> h <sup>-1</sup> )
$M_{f\text{MIN}}$	Minimum nonrain melt factor (21 December) (mm °C <sup>-1</sup> h <sup>-1</sup> )
$U_{\text{ADJ}}$	Average wind function during rain on snow events (mm mb <sup>-1</sup> )
SI	Average areal SWE above which 100% cover exists
NMF	Maximum of the negative melt factor (21 June) (mm °C <sup>-1</sup> h <sup>-1</sup> )
TIPM	Antecedent temperature index parameter (0.1 < TIPM < 1.0)
$M_{\text{BASE}}$	Base temperature for melt computations during nonrain
PXTEMP	Temperature above which precipitation is assumed rain
PLWHC	Percentage (decimal) liquid water holding capacity
DAYGM	Daily melt at the soil–snow interface (mm)
ADC	Eleven values that define the snow-covered area depletion curve

cover. New snowfall will force 100% coverage until 25% of that event’s snow has melted, after which the model returns to its regular position on the ADC. Two parameters can impact the forcing data, namely, the gauge undercatch parameter (SCF) and the parameter that describes the air temperature at which precipitation is either liquid or solid (PXTEMP). These two parameters can compensate for biases that may result during the estimation of forcing data.

### b. Model simulations and updating

A simple conceptual model, driven by uncertain forcing data, is unlikely to capture all snow processes, and errors in SWE can often be cumulative. Thus, the experiments performed here are based on the premise that calibration of the model can account for low frequency variability, while assimilation is applied so as to improve the high frequency variability. In the eventual application, the calibration would be performed at each location against an objective function involving streamflow, which is the ultimate objective. Such a process should mean that the long-term mean of the model should match that of the observations and no overall bias exists.

We could perform these experiments at any location within the domain of our data; however, we chose 53 SNOTEL station locations in the Upper Colorado River basin (Fig. 1) as these can provide us with “truth” observations that, for each location, are independent of

the model forcing and SWE estimates. We produce a 100-member ensemble for the 19 water years, 1 October 1984 to 29 September 2003.

Many calibration and parameter estimation methods exist [e.g., shuffled complex evolution (Duan et al. 1992); Monte Carlo Markov chains (Vrugt et al. 2003)]; however, in this study we used the SNOW-17 parameters that are implemented operationally in the Colorado Basin River Forecast Center (CBRFC). As an initial control simulation, we applied the model to the stations using parameters that were taken directly from the NWS operational code and applied them to stations on a nearest neighbor basis. In the operational setting SNOW-17 is run for a subbasin or grid box area rather than a point, and calibration is geared toward the final objective of streamflow rather than snow mass. However, we are testing the model at specific station locations. As such, we performed point-based simulations in which the SNOW-17 ADC was replaced with a function that assumes 100% snow cover after 20 mm of SWE is present.

In this study we update on a point-by-point basis as the spatial covariance of SWE is largely preserved via the forcing data. Hence our approach can be considered one-dimensional assimilation, where probabilistic spatial fields of SWE are obtained in a preprocessing step. Later studies will look into the computational requirements and impacts of spatial information propagation across a domain (e.g., Reichle and Koster 2003). While total SWE is our only assimilated quantity, we propagate the information of the update across the majority of the model state via the filter. The state variable vector ( $\mathbf{x}$ ) in the assimilation step is of dimension nine and among other items includes total SWE, negative heat storage of the snow, snow depth, and snow temperature. Total SWE in SNOW-17 is the sum of several state variables, that is, the water equivalent of solid phase snow, liquid storage, and a series of seven lagged drainage water stores. Prior to the update, the ratios of these quantities are computed and the updated SWE redistributed accordingly. By equating the model variable with the assimilated quantities, the  $\mathbf{H}$  matrix in the EnKF is actually a row vector with the first element (total SWE) equal to one and all others zero.

Including multiple model state variables in the update step provides two main advantages. First, one can more effectively assimilate observations of different variables during a single update step. Second it represents a deliberate rebalancing the model, rather than letting the model adjust to updated states that may not be in synchronicity with other parts of the model. These advantages are at the expense of increased dimensionality of the assimilation as well as the need to ensure

that the cross covariance of model states do not contravene physical requirements (e.g., by setting bounds on variables). We perform updating every five days in our experiments only when the assimilation SWE quantities are above zero.

## 5. Results and discussion

### *a. Forcing and assimilation data verification*

An important distinction that separates assimilation methods such as the Kalman filter from simpler methods like optimal interpolation is that the model error term is propagated forward in time. Under the EnKF, the effectiveness of this propagation is dependent upon the ensemble spread capturing the uncertainty inherent in the modeling system. In the case of snow-dominated hydrologic forecasting, this uncertainty is largely dependent upon the spread of the forcing data (Carpenter and Georgakakos 2004). Other areas of uncertainty, such as parameter estimation and model structural uncertainty, are not included in the current study. Regardless of whether the experiments have been with real or synthetic data, few studies regarding the EnKF have reported on the methods (above) and/or reliability of the ensemble generation process (below).

Figure 4 shows a rank probability diagram for the interpolated daily forcing and assimilation variables for the seasonal cycle of the water year; the rank refers to where the observation or truth quantity falls within the ensemble. A cross section of the plot can be viewed as the rank histogram (Hamill 2001) or Talagrand diagram (Talagrand et al. 1997) of a given day. The diagram was constructed using a 15-day window centered on a given day of the year for all 19 years of the simulations and across all 53 stations. Twenty bins were used in constructing the probabilities, and ensemble members with equal ranks share the same probability. Thus, if the ensemble has the desired spread such that the observations were to be evenly distributed across it, a probability of 0.05 would be assigned to each of the 20 bins, so as to integrate to 1.0. Instances of under- or overshooting are binned in the end categories.

For mean daily temperature, the ensemble tends to be more centered on the observed value over the winter period, suggesting that our spread is slightly larger than required (Fig. 4a). One reason for this is that the quality of the temperature data from the SNOTEL sites at this time is rather poor. Although we performed careful quality control of the SNOTEL temperature data, we did not catch all spurious values, some of which adversely affect the spread of the temperature ensembles. Diurnal temperature range displays a more uniform distribution (Fig. 4b).

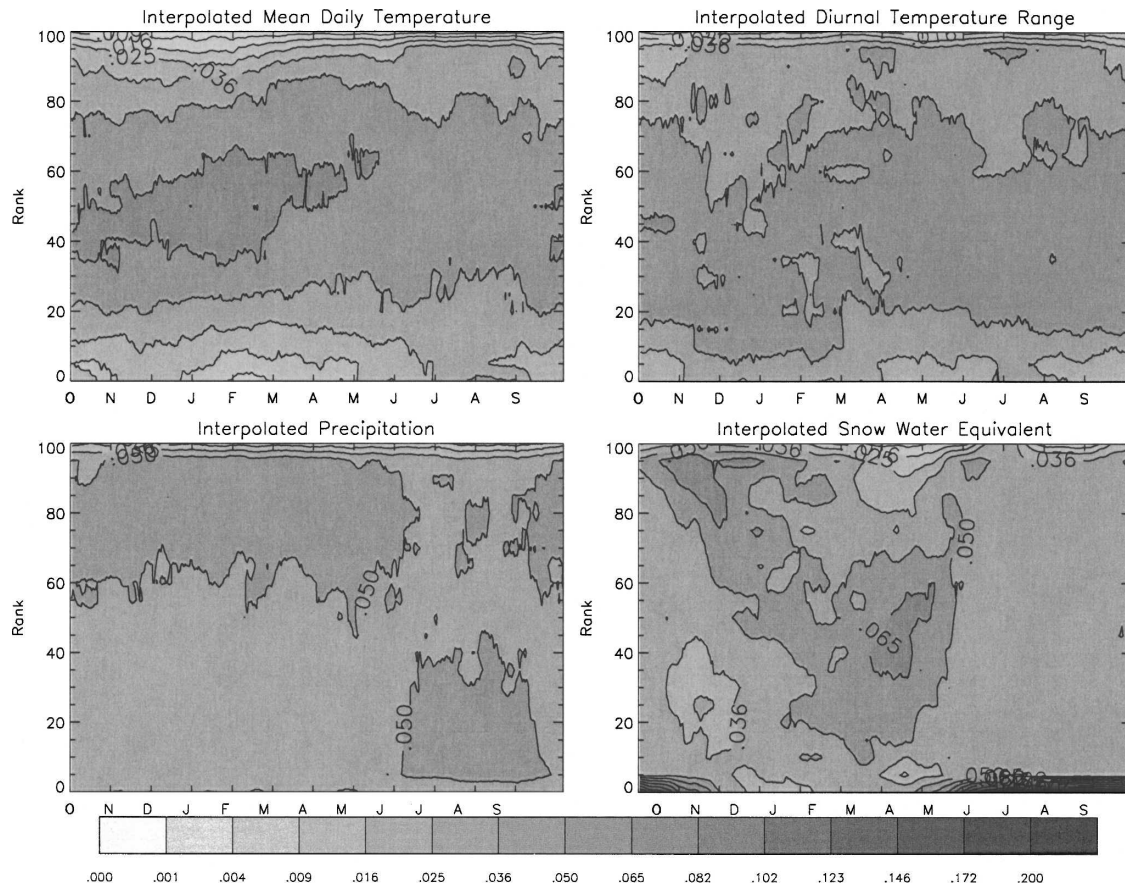


FIG. 4. Seasonal cycles of the ranked histogram of all model inputs. The probability of the observed value having a given rank within the ensemble is the contoured variable. Twenty bins were used in the histogram; thus a probability of 0.05 gives a completely uniform distribution.

Clark and Slater (2006) show that the method of interpolating precipitation used here possesses both reliability, that is, close agreement between the frequency of occurrence of specific precipitation events in different probability categories and the probability that is estimated from the ensemble, and discrimination in the sense that the estimated probabilities differ significantly between cases when specific precipitation events occur and when they do not. Accordingly, the rank histogram plot shows a uniform distribution (Fig. 4c). Additional probabilistic verification for precipitation is shown in Clark and Slater (2006).

The interpolated SWE, which is the data to be assimilated, also displays a reasonably uniform distribution (Fig. 4d). In November the “truth” values tend to be in the upper portion of the ensemble and less in the lower portion, indicating a slight underestimation of SWE. In April, the observations favor the middle of the ensemble, suggesting that the spread is larger than necessary.

Overall, the rank histograms for all model inputs are

dominated by probabilities that suggest an even spread, thus giving us confidence that we have captured the majority of the model forcing uncertainty. For further clarification, an example time series of temperature and precipitation is given for the Kiln station (Fig. 5). The increased spread in temperature over winter is easily seen. The ensemble of precipitation is able to capture specific storm events, for example, February and March 1996, when a step in cumulative precipitation is evident in both the ensemble and observations.

*b. Control and assimilation simulations*

In this study, the control simulations are defined as those where the areal depletion curve of the model have been modified for a point-based simulation so as to make a fair comparison to SNOTEL station data (see section 4a). Figure 6 shows some example years from various stations for both the control and EnSRF runs. The control ensemble largely envelops the truth, apart from during the early accumulation and late ablation periods. Early accumulation and late ablation pe-

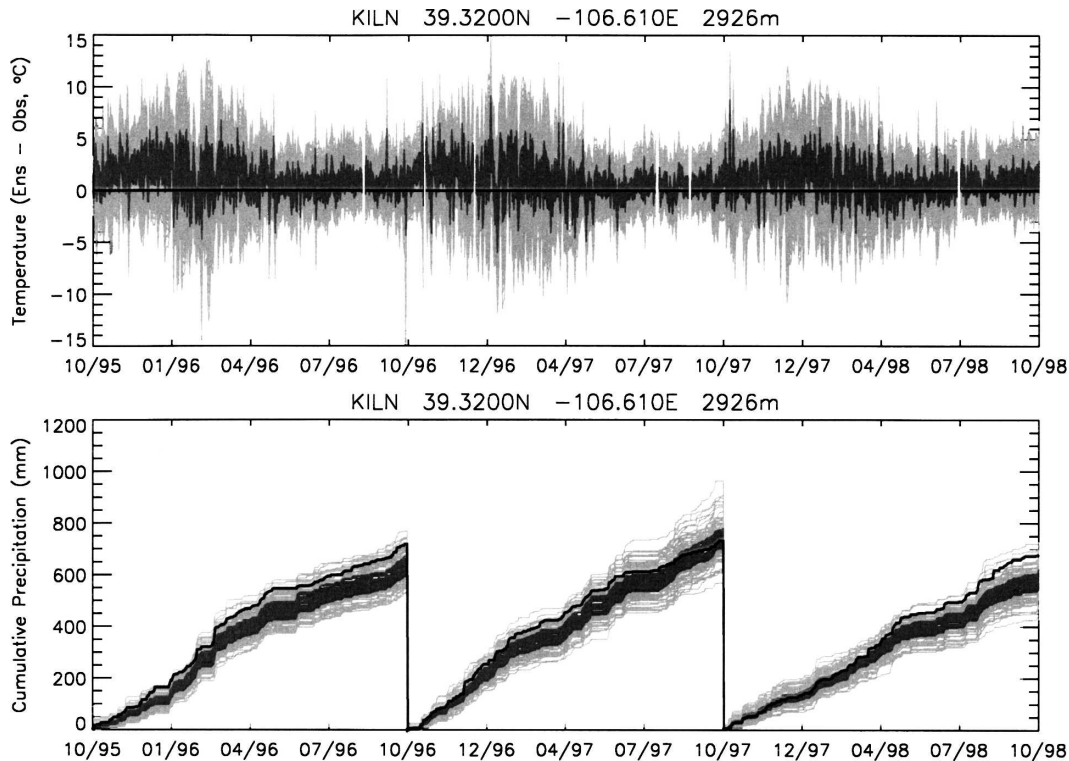


FIG. 5. An example 3-yr time series of model forcing data. Observations are given in the thick bold line. Individual ensemble members are shown in gray, with the darker gray representing the interquartile range of the ensemble. (upper) Mean daily temperature expressed as the difference from observed; (lower) cumulative precipitation over the water.

riods are often difficult to simulate in both conceptual hydrologic and in land surface models (e.g., Slater et al. 2001).

The EnSRF run at Hoosier Pass shows a clear improvement compared to the control, which underestimates SWE through to February (right column of Fig. 6). The assimilated ensemble begins to encompass the truth in December, and by February the model is closer to observations than the interpolated SWE (squares/vertical lines). In contrast, the EnSRF simulation over the midaccumulation season at University Camp demonstrates the rare case of the assimilation resulting in a degradation of skill (middle column of Fig. 6). As expected, the simulation can suffer if the assimilated data is of poor quality or contains a bias or misspecification of error. Typically, the error estimates for the interpolated SWE are much smaller over the accumulation period and largest during the melt, when there is much greater variability between stations. At the stations shown, the control runs either over- or underestimate the melt period.

A feature of the EnSRF simulation is the noticeable decrease in variance of the model ensemble, which may be leading to a less than optimal solution in the assimilation.

Therefore, we need to ensure that we are not over constraining the model due to factors external to the filter.

### c. Temporal correlation

Snow is a cumulative quantity with months of auto-correlation time. Hence, within the EnSRF simulations we are aware that the filter is not functioning in an optimal manner as the innovations possess a red spectra (Daley 1992).

A potentially more serious consequence of the red noise problem is that of filter divergence (Whitaker and Hamill 2002). The Kalman gain applied to the maximum potential increment (i.e., the difference between the model and observation) is essentially a function of the ratio of the model ( $\mathbf{P}$ ) to observational ( $\mathbf{R}$ ) variance. If the analysis variance is underestimated and the model falsely maintains this variance due to temporal persistence, the model variance will keep decreasing with each update. Thus the filter assumes greater confidence in the model and less attention will be paid to the observations, which can eventually lead to a model trajectory that is not related to the observations. For example, if one was to simply apply the EnSRF without

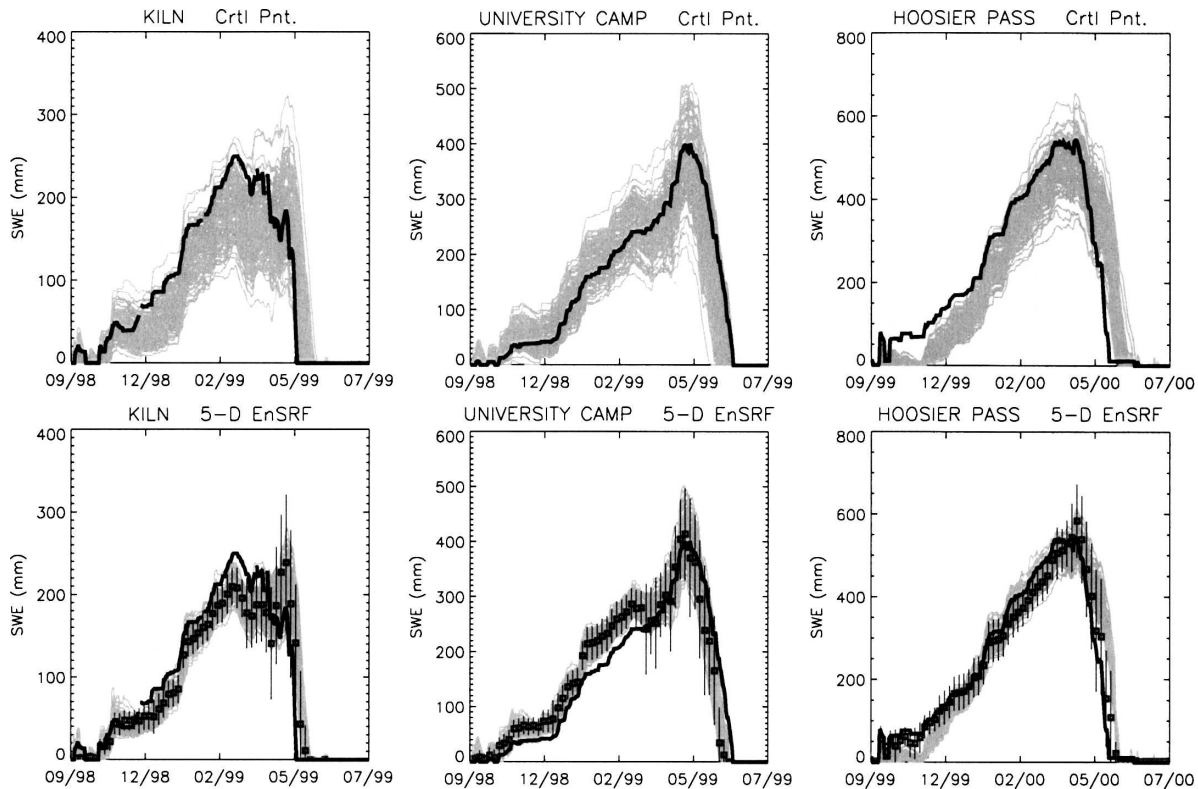


FIG. 6. (upper) Examples of model control simulations and (lower) assimilation runs using the ensemble square root Kalman filter, with updates every 5 days (5-D). The ensemble of model simulations is given in light gray, and truth is given in the thick dark line, and the mean of the assimilated SWE is given by the dark squares complete with one standard deviation shown in the thin dark lines. Periods shown are from 29 September to 31 July of respective water years.

account of temporal correlation and the assimilation step was performed numerous times at the same time step, this positive feedback would cause the spread of the model ensemble to collapse to near zero and the observations would effectively be ignored.

To overcome this red noise problem, during any assimilation step, one should aim to only incorporate information that is additional to that gained at the previous update. A simple method of alleviating the persistence problem is to reduce the values of the Kalman gain proportionally to that of the temporal correlation in the data; that is,

$$\mathbf{K}_{\text{lag}} = \mathbf{K}(1 - \rho^2). \quad (10)$$

The lag 1 correlation,  $\rho$ , is obtained by correlating the current and previous update vector of ensemble members against each other. We apply  $1 - \rho^2$  (sometimes called the coefficient of alienation) as it describes the amount of variance that is unaccounted for from the past update. This correction is applied to both the standard gain for updating the ensemble mean and to the reduced gain for updating the anomalies. It is, in effect,

applied twice: once to account for correlation in the observations and once to account for correlation in the model. The result from the lag(1) EnSRF is substantially more spread in the simulated ensemble compared to the standard EnSRF. This provides the filter with the ability to correctly assess the relative merits of model and observations.

This impact of removing the temporal correlation is demonstrated in Fig. 7 during the melt season at the Freemont Pass station. The control simulation melts considerably later than truth. Applying a standard EnSRF, an improvement in melt timing is evident but the model still melts later than truth, while the interpolated SWE suggest a melt prior to the truth date. The diminished variance of the EnSRF model ensemble is a notable feature, especially in the accumulation season, where no ensemble members fall below the truth. Applying the lag EnSRF, which accounts for temporal correlation, the variance of the ensemble is between that of the control and standard EnSRF. This is evident in both the accumulation period, with some simulations below the truth value, and during ablation, for example, in late May and early June, when the model ensemble

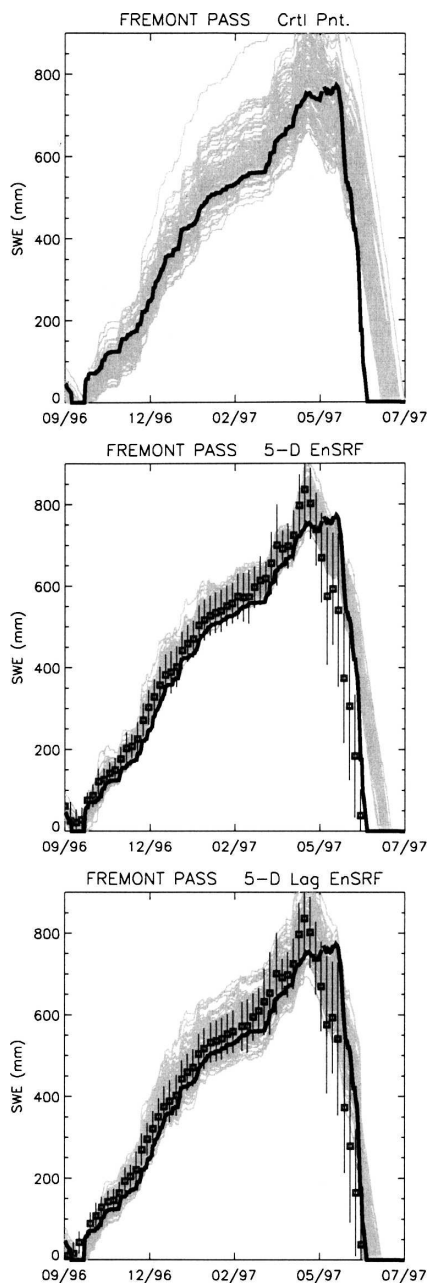


FIG. 7. Plot legends are as for Fig. 5. (upper) A control run, (middle) a standard EnSRF, and (lower) an EnSRF, which accounts for the temporal correlation of model and observation within the assimilation scheme.

now encompasses the mean of the interpolated SWE. Most importantly, the final ablation curve of the ensemble is much closer to truth and more accurately reflects the relative uncertainty of the model and interpolated SWE. Recall that these figures show the entire model ensemble, but only the standard deviation of the interpolated SWE. Alternative methods of dealing with

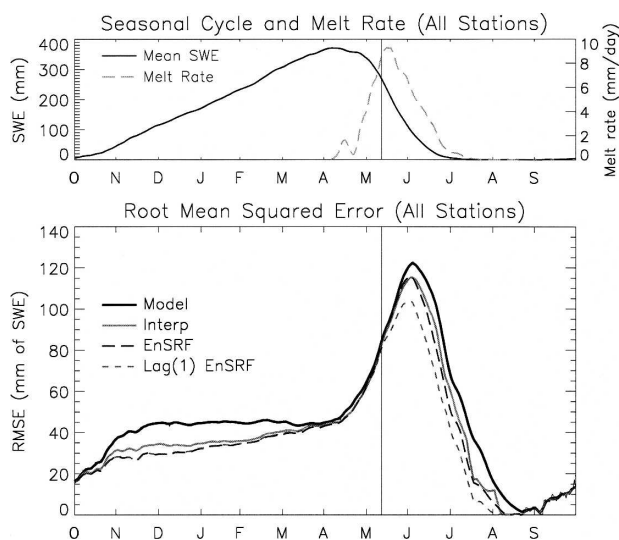


FIG. 8. The rms error of model ensemble means and mean interpolated values. The interpolated SWE (Interp) is always superior to the point-based model control simulation (Point), while the standard EnSRF and the version accounting for temporal correlation [lag(1) EnSRF] show further improvement. The  $x$  axis tick marks indicate the start of the month.

the redness in the increments include state vector augmentation (Reichle et al. 2002a).

#### d. Summary statistics

Adopting a deterministic approach to the estimation problem, one assumes that the mean of the ensemble is the best estimator at any time. Therefore, we can compute the root-mean-square error (rmse) of the various simulations:

$$\text{rmse} = \frac{1}{N} \sqrt{\sum_{i=1}^N (x_i - x_o)^2}, \quad (11)$$

where  $x_i$  is the mean of the ensemble,  $x_o$  is the matching truth observation, and  $N$  is the number of ensemble/observation pairs. For each day of the seasonal cycle, across all stations and all years, the means of the simulation ensembles are compared to the truth observation if it existed. Of the 53 stations, 43 have valid truth data for more than 17 years and the minimum length of valid data is 7.5 years; hence this is a robust summary and is shown in Fig. 8.

In terms of the all station seasonal cycle rmse, the interpolated SWE is superior to the model. The standard EnSRF is either near equal to the interpolated SWE or, more usually, superior. The lag EnSRF is always superior to the interpolated SWE and model. This is a clear demonstration that the filter is working in the expected manner; it combines the respective strengths

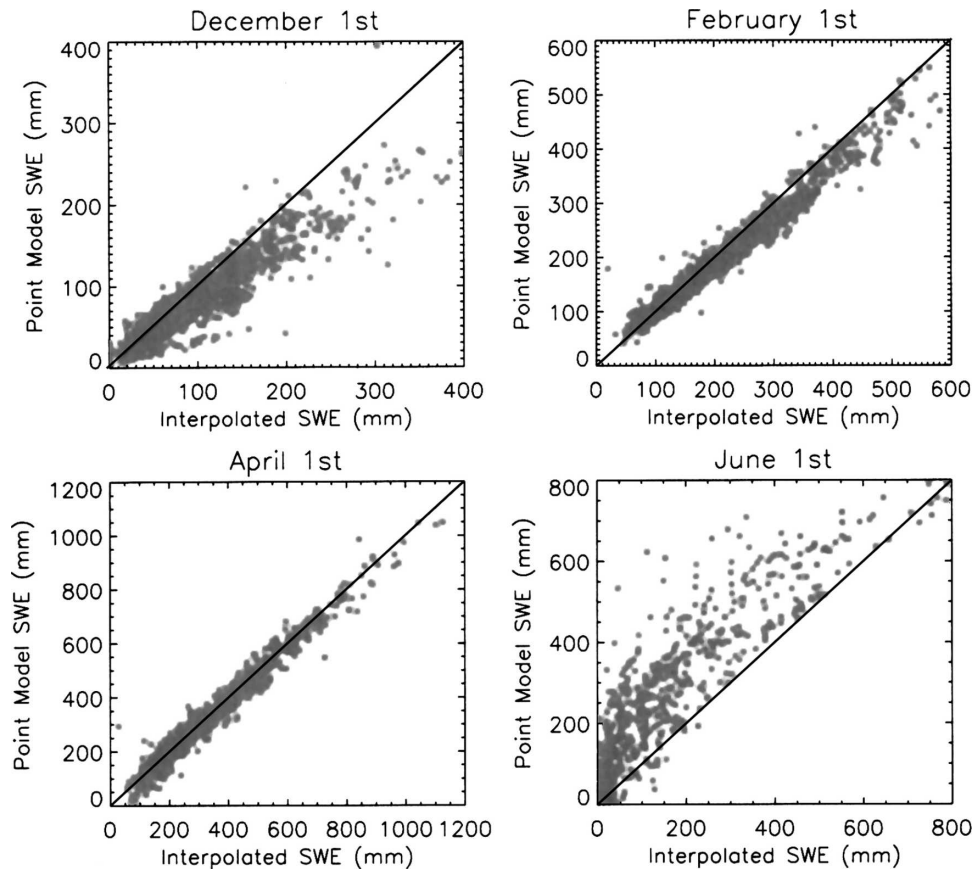


FIG. 9. Scatterplots of the point-based model ensemble mean compared to the mean of the interpolated SWE at four times during the year. Most notable is that around the time of peak accumulation the model and interpolated data generally have the same value.

of model and observations. Recall that all results are cross validated via withholding experiments; that is, the SWE observations for a given station were not used in the SWE interpolations for that station.

Over the accumulation period, the assimilation runs give 15 mm lower rmse at 1 December and this is gradually reduced to no appreciable improvement by 1 April. According to Serreze et al. (1999), the mean date of maximum snow accumulation in the Colorado region is 6 April. The lack of any significant improvement between model simulation, assimilation, and interpolated data at the peak of snow accumulation results from the fact that the interpolated SWE gives almost the same mean value as that of the model ensemble. This is perhaps not surprising given that a similar method is used to produce ensemble precipitation forcings and ensemble SWE estimates. In midwinter when most precipitation falls as snow, modeled SWE is predominantly determined by the precipitation ensembles. Hence differences between modeled and interpolated SWE are mostly evident during autumn, before comparative er-

rors between the forcing and interpolated SWE have had a chance to average out, and during spring, when melt processes dominate. Improving SWE simulations at the time of peak accumulation will require an independent SWE data source.

Figure 9 shows comparisons of the model ensemble mean compared to the interpolated SWE at four times during the snow season. Where there is useful additional information to be gained from the interpolated data, for example, early in the season (1 December) or during the melt (1 June), we can see a corresponding increase in skill in the rmse. Where the model and interpolated SWE essentially give a similar result (e.g., 1 April) there is no additional skill to be gained from assimilation.

Greatest uncertainty is associated with the melt process (e.g., Fig. 2); thus it is at the peak of accumulation that the error begins to grow, with rmse increasing noticeably after mid-April. For climate and hydrologic purposes, the melt is of most interest as it represents the period of varied surface conditions and peak stream-

flows. During the melt, an overall increase in skill is most evident after 11 May; a time at which over 75% of the mean snowpack still remains (Fig. 8). This indicates the high utility of our method for hydrologic forecasting. Compared to the control simulation, the lag EN-SRF run gives a rmse that is lower by at least 15% from the time of maximum rmse, 3 June, through to the complete removal of snow.

## 6. Conclusions

A snow data assimilation study has been undertaken in which we update the states of a conceptual snow model by assimilation of SWE. The data used is real, not synthetic, and most of it is available on a near-real time basis. Via the process of cross validation we have explicitly determined the error of both the model forcing and the assimilated data; there are no a priori assumptions on the error structure. The method was tested at points as this is the only source of validation data, but the procedure will directly translate to a model subbasin or grid box.

We have verified that it is possible to gain an overall superior simulation of model space SWE initial conditions when combining the use of model and independent observations. In turn, better specification of initial conditions should lead to a better streamflow forecast. This is especially true during the melt period. The gain in skill with respect to maximum snow volume is rather marginal over the period of late March through April as there is little additional information in the interpolated estimates. The ability of a filter to add skill to a model is dependent upon the quality of the observations that are used and correct estimation of the error structure of both the model and the observations during the assimilation. For short-term forecasting over the melt period, we have a demonstrated improvement in the SWE estimates when we apply a filter that accounts for the temporal correlation in model and observations. When applied to streamflow forecasts, such an improvement can have significant economic potential for reservoir operators are interested in short-term forecasting.

For the assimilation system presented here, the final impact upon streamflow will partly be a function of the density of data sources as this will generally control the levels of uncertainty associated with forcing and/or assimilation data. In this study we have only used station data and are aware that improvement can be made on our current interpolation scheme. In terms of the forcing and SWE data there are additional sources available for those quantities, for example, the Meteorological Assimilation Data Ingest System (MADIS) (<http://www-sdd.fsl.noaa.gov/MADIS/>) real-time network for

temperature and precipitation or the addition of fortnightly COOP snow course data for SWE. In the future we hope to incorporate the use of other data sources, such as visible satellite imagery (e.g., MODIS), to determine snow presence or absence as well as microwave radiances so as to aid in estimation of snow mass [e.g., Advanced Microwave Scanning Radiometer–Earth Observing System (AMSR-E)].

*Acknowledgments.* This work was primarily supported by the National Weather Service Office of Hydrologic Development (Award NWS4620014). Additional support was obtained from NSF Grant OPP-0229769. We thank Dave Brandon for SNOW-17 parameters, and Yun Duan for the SNOW-17 code. We thank Tom Hamill for valuable discussions regarding Kalman filtering. Three anonymous reviewers helped add clarity to this paper.

## REFERENCES

- Anderson, E. A., 1973: National Weather Service River Forecast System—Snow Accumulation and Ablation Model. NOAA Tech. Memo. NWS HYDRO-17, 217 pp.
- Brasnett, B., 1999: A global analysis of snow depth for numerical weather prediction. *J. Appl. Meteor.*, **38**, 726–740.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Carpenter, T. M., and K. P. Georgakakos, 2004: Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model. *J. Hydrol.*, **298**, 202–221.
- Clark, M. P., and L. E. Hay, 2004: Use of medium-range numerical weather prediction model output to produce forecasts of streamflow. *J. Hydrometeorol.*, **5**, 15–32.
- , and A. G. Slater, 2006: Probabilistic quantitative precipitation estimation in complex terrain. *J. Hydrometeorol.*, **7**, 3–22.
- , S. Gangopadhyay, L. E. Hays, B. Rajagopalana, and R. Wilby, 2004: The Schaake shuffle—A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeorol.*, **5**, 243–262.
- Cline, D. W., R. C. Bales, and J. Dozier, 1998: Estimating the spatial distribution of snow in mountain basins using remote sensing and energy balance modeling. *Water Resour. Res.*, **34**, 1275–1285.
- Daley, R., 1992: The effect of serially correlated observation and model error on atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 164–177.
- Day, G. N., 1990: A methodology for updating a conceptual snow model with snow measurements. NOAA Tech. Rep. NWS 43, 133 pp.
- Duan, Q. Y., S. Sorooshian, and V. Gupta, 1992: Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.*, **28**, 1015–1031.
- Evensen, G., 1994: Sequential data assimilations with a nonlinear quasi-geostrophic model using monte-carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10 143–10 162.
- , 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.

- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560.
- , 2006: Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate*, T. N. Palmer and R. Hagedorn, Eds., Cambridge University Press, 712 pp.
- Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich, 1998: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resour. Res.*, **34**, 3405–3420.
- Liston, G. E., 2004: Representing subgrid snow cover heterogeneities in regional and global models. *J. Climate*, **17**, 1381–1397.
- , R. A. Pielke, and E. M. Greene, 1999: Improving first-order snow-related deficiencies in a regional climate model. *J. Geophys. Res.*, **104** (D16), 19 559–19 567.
- Reichle, R. H., and R. D. Koster, 2003: Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J. Hydrometeorol.*, **4**, 1229–1242.
- , D. B. McLaughlin, and D. Entekhabi, 2002a: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103–114.
- , J. P. Walker, R. D. Koster, and P. R. Houser, 2002b: Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeorol.*, **3**, 728–740.
- Rodell, M., and P. R. Houser, 2004: Updating a land surface model with MODIS-derived snow cover. *J. Hydrometeorol.*, **5**, 1064–1075.
- , and Coauthors, 2004: The global land data assimilation system. *Bull. Amer. Meteor. Soc.*, **85**, 381–394.
- Serreze, M. C., M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. L. Pulwarty, 1999: Characteristics of the western U.S. snowpack from snowpack telemetry (SNOTEL) data. *Water Resour. Res.*, **35**, 2145–2160.
- Slater, A. G., and Coauthors, 2001: The representation of snow in land surface schemes: Results from PILPS 2(d). *J. Hydrometeorol.*, **2**, 7–25.
- Sun, C. J., J. P. Walker, and P. R. Houser, 2004: A methodology for snow data assimilation in a land surface model. *J. Geophys. Res.*, **109**, D08108, doi:10.1029/2003JD003765.
- Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *Proc. ECMWF Workshop on Predictability*, Reading, United Kingdom, ECMWF, 1–25. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom.]
- Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian, 2003: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.*, **39**, 1201, doi:10.1029/2002WR001642.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.