Data Assimilation of Remotely Sensed Snow Observations Using an Ensemble Kalman Filter

Konstantinos M. Andreadis[1], and Dennis P. Lettenmaier[1]

1. Department of Civil and Environmental Engineering, Box 352700, University of Washington, Seattle, WA 98195

CAHMDA-II International Workshop

The Terrestrial Water Cycle: Modelling and Data Assimilation Across Catchment Scales

Oct 25-27, Princeton, NJ

ABSTRACT

Snow is a major component of the hydrologic cycle and can play an important role in water resources management, especially in mountainous areas like the western United States. Current model-based approaches to hydrologic forecasting are limited by model biases and input data uncertainties, while ground based measurements have limited coverage and are unable to capture the spatial and temporal variability of snow properties. Remote sensing offers an opportunity for observation of snow properties, like areal extent and water equivalent, over large areas. The Moderate Resolution Imaging Spectroradiometer (MODIS) has been operational since early 2000, and provides snow cover information at 500 m spatial resolution which is appropriate for regional applications. However, visible wavelength sensors like MODIS are inhibited by cloud cover which causes temporal discontinuities. Furthermore, MODIS provides no information about snow water content. Data assimilation offers a framework for optimally merging information from remotely sensed observations and hydrologic model predictions, and ideally overcoming limitations of both. This work describes the assimilation of MODIS snow areal extent data into a macroscale hydrologic model over the Snake River basin, using an ensemble Kalman filter (enKF). The approach is built around the Variable Infiltration Capacity (VIC) macroscale hydrology model, which balances water and energy over each model grid cell at each timestep. The state variables included snow water equivalent at each model elevation band. Results showed that the enKF is an effective and operationally feasible solution for the assimilation of remotely sensed observations. The filter successfully updated snow cover predictions by the model. Ground observation comparisons using SNOTEL and NCDC Cooperative Observer snow water equivalent and snow depth data, respectively, indicate that the filter estimates are an improvement over the open-loop VIC simulations. Finally, the effect of the assimilation on streamflow and the potential of bias correction using data assimilation are discussed.

1 Experimental Design

The advent of the EOS era and the operational use of very promising remote sensing instruments, such as the MODIS and AMSR-E, has increased the potential for assimilating data products, from those satellites, into land surface models. One of the widely used data assimilation techniques in hydrology, is the ensemble Kalman filter. This work attempts to assess the performance of a data assimilation system that incorporates snow covered area (SCA) information into the VIC model, over a period of four consecutive winters (2000-2003) at the Snake River basin. The latter is a major tributary of the Columbia basin, where about 75% of the annual streamflow is driven by snowmelt. Our experiment uses real observations; in contrast with synthetic experiments the true state of the system is unknown, and hence the performance of the enKF must be evaluated either by comparison with independent data or qualitatively. Surface observations are the only practical option for independent evaluation. We used data from two station networks, SNOTEL and COOP that provide snow water equivalent (SWE) and depth respectively. In order to account for the scaling issues in comparing areal estimates with point measurements, we expressed simulated and SNOTEL SWE as percentiles of their respective climatology (taken from a 20 year record, 1983-2003). Also, stations that had an elevation difference from the model grid cell mean elevation, greater than 200 m were excluded from the comparisons.

2 Hydrologic Model

The hydrologic model used in this study is the Variable Infiltration Capacity (VIC) model (*Liang et al.* 1994). Essentially the model solves a water and energy balance over a grid mesh. VIC accounts for subgrid variability in topography and land cover by representing each grid cell as a number of subgrid tiles of a certain land cover type and elevation zone. SCA is represented indirectly, by assuming that a tile is fully covered if any snow is present. Thus, SCA is just the area-weighted sum of all snow-covered tiles. Snowpack dynamics are modeled using a two-layer energy and mass balance model (Figure 1). The upper layer solves the energy balance between the snowpack and the atmosphere, while the lower layer acts as storage of excess snow and simulates deeper snowpacks. Other processes accounted for include snow densification and interception (*Cherkauer and Lettenmaier*, 2003).

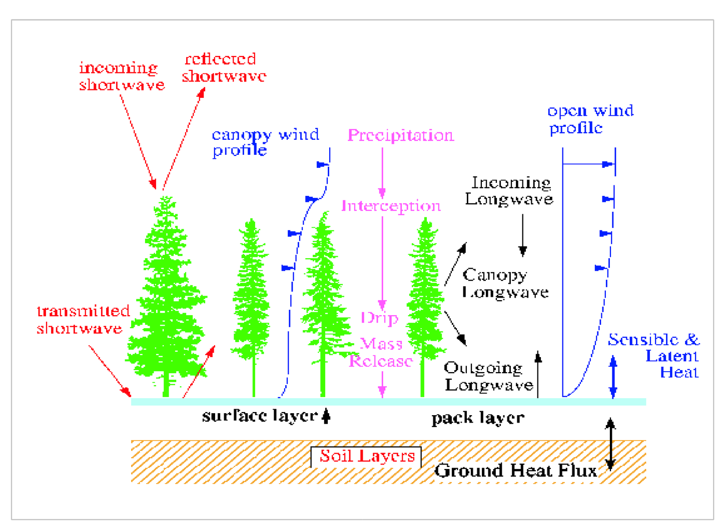

Figure 1. VIC snow model component.

3 Ensemble Kalman Filtering

The Kalman filter solves the optimal estimation problem for linear processes, i.e. the state estimation of a process. The KF accounts for errors in both model and observations, by explicitly propagating the model error covariance information in time *(Gelb*, 1974*)*. This proves to be very expensive computationally for large-scale applications. *Evensen* (1994) developed a Monte Carlo approach to the KF, the ensemble Kalman filter (enKF). This avoids the propagation of the error information, by implicitly calculating the required error covariances from an ensemble of model states. The algorithm starts with the propagation of each model ensemble member to the timestep that an observation becomes available (forecast step). The ensemble is generated at each timestep by treating model parameters as a stochastic variables (e.g. forcing data). At the observation time (analysis step), the calculated error covariance matrix P^f is used to compute the Kalman gain, that weighs the magnitude of the effect of the observations, and the model-predicted state y^f is updated to y^a. H is the observation operator, which relates the state variables to the observations. This allows for assimilation of indirectly observed variables, e.g. assimilation of SCA to update model-predicted SWE. Each ensemble member i is updated separately and the filter estimate is usually taken as the mean of the ensemble values.

At each timestep

Forecast Step :

$$y^f_{t,i} = f\left(y_{t-1,i}, u_i, \alpha, w_i\right) \quad i = 1,\ldots,N$$

$$P^f = \overline{\left(y^f - \overline{y^f}\right)\left(y^f - \overline{y^f}\right)^T}$$

If an observation is available

Analysis Step :

$$z_i = H\left(y_i, v_i\right)$$

$$y^a_{t,i} = y^f_{t,i} + K\left(z_{t,i} - H\left(y^f_{t,i}\right)\right)$$

$$K = P^f H^T \left(HP^f H^T - R\right)^{-1}$$

4 Observation Operator

When updating model predicted SWE by assimilating snow areal extent data, such a non-linear functional is necessary. In this study, a snow depletion curve parameterization scheme developed by *Anderson* (1973) is used, which relates areal average SWE to the snow covered area of the model element. To better account for snow spatial variability, we categorized the VIC subgrid tiles in nine different physiographic classes, based on elevation (between 0, 1500 and 2000 m) and land cover (forest, shrublands, and grasslands). A separate depletion curve was developed for each of these physiographic classes. It is generally difficult to obtain direct observations of both SCA and SWE from which the depletion curve can be estimated. The approach we used makes use of the MODIS dataset and SWE estimates from the prior VIC simulations to infer the parameters and shape of the depletion curve. A full coverage SWE parameter is calculated by examining the snow extent time series and averaging the SWE values that corresponded to full snow coverage immediately before the onset of snowmelt. We chose to fit 2-parameter gamma distributions to the combined MODIS SCA and modeled SWE (Figure 2). Despite the large scatter, which can be partly attributed to discrepancies between VIC simulations and MODIS observations, fitting gamma distributions to the data seemed a reasonable approach.

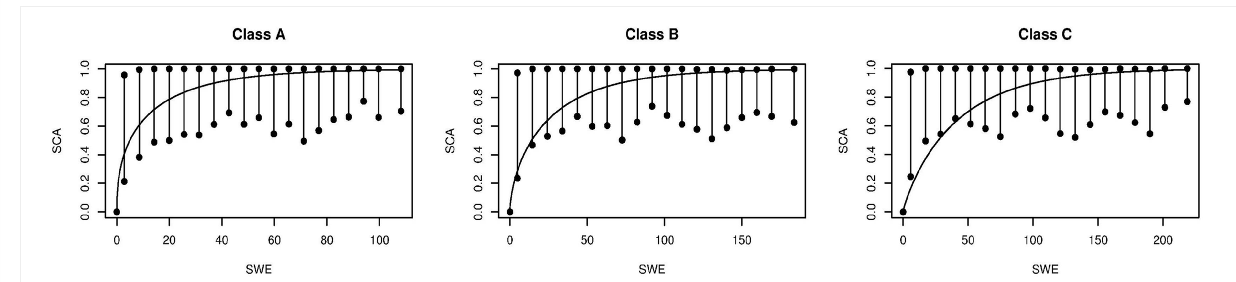

Figure 2. Fitted snow depletion curves for three physiographic classes. Bars show 25th and 75th percentile of the MODIS SCA and VIC SWE (mm).

5 Model Implementation

The model was applied at a spatial resolution of 1/8º and hourly timestep. The model state variables are SWE at each VIC model subgrid tile. The ensemble of model states is generated by treating precipitation and air temperature forcing data as stochastic variables. Log-normally distributed precipitation values were generated and implemented as in (*Nijssen and Lettenmaier*, 2004) $\hat{P} = (\sqrt{1+E^2})^{-1} \exp[\sqrt{\ln(1+E^2)}\epsilon(0,1)]P$, where E is the relative error (taken here as 25%). Minimum and maximum air temperature values were generated by perturbing the daily air temperature mean and range as follows, $\hat{T}_{min,max} = (T_{mean} + \epsilon_1(0,2)) \pm (|T_{range} + \epsilon_2(0,1)|)/2$ ϵ are spatially correlated Gaussian random fields with mean zero, generated with a 2-D turning bands a;gorithm and an arbitrarily selected exponential correlation model. VIC is formulated in a way that it solves eacg grid cell separately; this allows for a small ensemble size (here chosen as 25). The data assimilated were obtained from the MODIS daily snowcover product, which is produced at a 500 m spatial resolution. The data were aggregated to the model resolution, and the end product was a fractional snow cover map of the VIC model elevation bands. In addition, a fractional cloud cover threshold of 20% was used to decide whether to use the observation or not. The observation error was represented as a normally distributed random variable with zero mean and 10% standard deviation.

6 Snow Cover Extent Results

Figure 3 shows spatial maps of the percentage agreement of snow covered days for the prior and filter estimate with MODIS SCA. The days included in the comparison were selected based on whether MODIS imagery was missing or excessively cloud covered. Because these maps only reflect the presence of snow, the cloud cover threshold for the comparison was set to 50%. This figure clearly shows that in general the enKF updates model predicted SCA in a consistent manner. The filter estimates do not match the MODIS observations exactly because the enKF accounts for errors in both model predictions and observations. Next we compare SCA from MODIS and the two model simulations with ground measurements. The same procedure for the selection of days was followed, but the comparisons were limited to the model pixels where a ground station was present. The elevation difference screening reduced the available number of stations to 124 from 257 (elevations from 114 to 2874 m). The table below shows the results of this comparison averaged over all the available stations. Both simulations and MODIS show good agreement with ground observations, however none of them is perfect. On average the enKF simulation shows a small improvement over the prior simulation and MODIS.

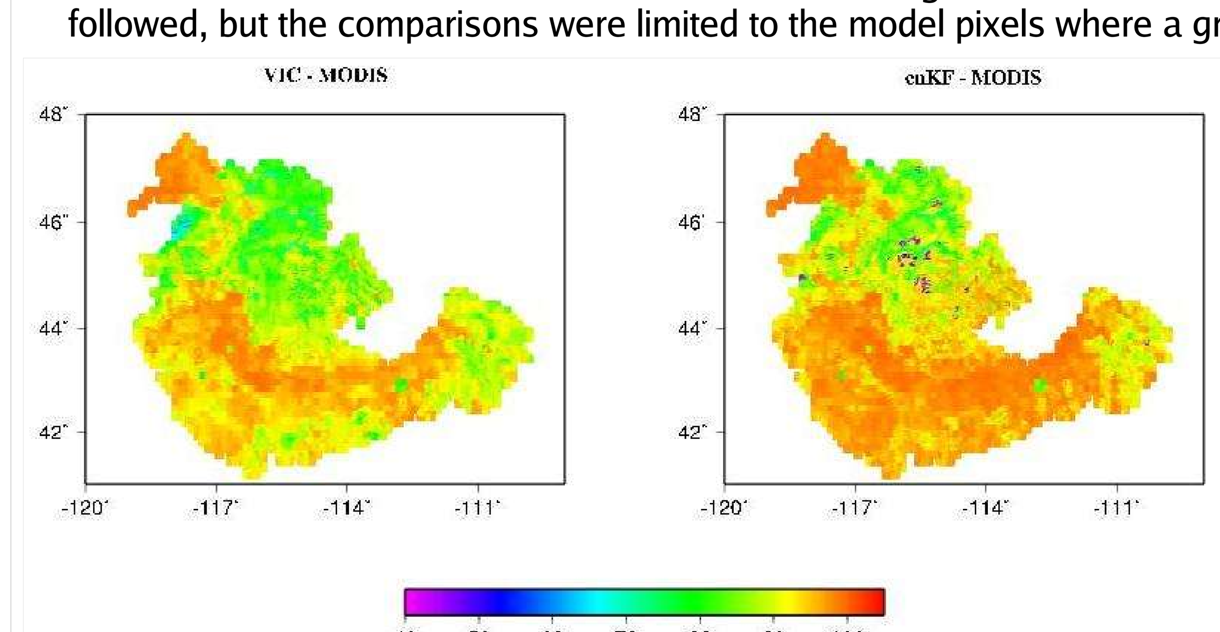

Figure 3. Spatial maps of percentage agreement of snow covered days for VIC (left) and the enKF (right) simulations, with MODIS observations, for the entire simulation period.

Data Source	Min	Mean	Max	Number of stations with minimum misclassification
MODIS	0.542	0.903	1.000	42
VIC	0.778	0.922	0.996	26
VIC-enKF	0.800	0.931	1.000	56

7 Snow Water Equivalent Results

In terms of hydrologic forecasting an interesting variable is the peak seasonal SWE. Figure 4 shows scatterplots of the seasonal maximum SWE for SNOTEL and the two simulations. Even though the results are similar the enKF reduces the scatter somewhat, and the filter-estimated peak SWE is closer to the corresponding SNOTEL value for 58 of the available 66 stations.

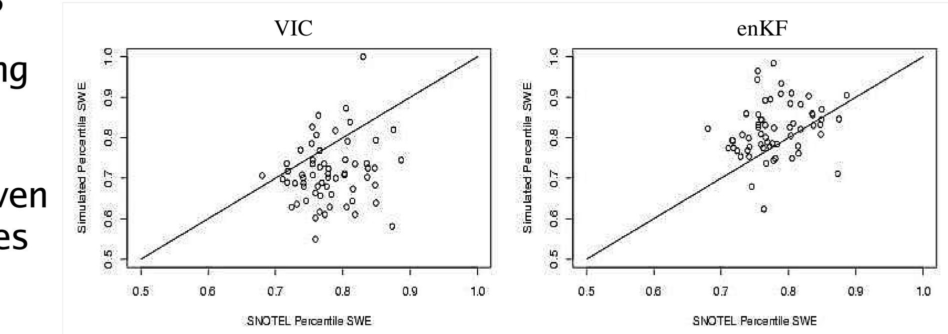

Figure 4. Comparison of seasonal maximum SWE between model simulations and SNOTEL measurements.

Comparing the SWE percentiles for individual SNOTEL stations, we can compute the RMSE. On average the RMSE for both simulations tends to be the same (0.19 versus 0.17, with 40 out of 66 stations having a lower RMSE), however for some stations the enKF estimate has a larger error than the open-loop simulation. Better insight can be obtained by examining a SWE percentile timse series for a specific station (Figure 5, W Yellowstone station, 2010 m). The enKF successfully updated the snow areal extent, which verifies our assumption about the small ensemble size used. However, the filter consistently overestimated SWE when compared to the SNOTEL value. Early in the accumulation season, the model-predicted SCA was less than one and therefore the MODIS observations of full coverage resulted in a relatively large increase of SWE. Subsequently, both the open-loop simulation and enKF (as well as MODIS) were at full coverage, which means that no updates occurred, and SWE was adjusted entirely based on the VIC model physics. The large departure of the enKF SWE at the beginning of the accumulation season may be attributed to the SDC parameters (e.g. low SWE value that corresponds to full snow cover).

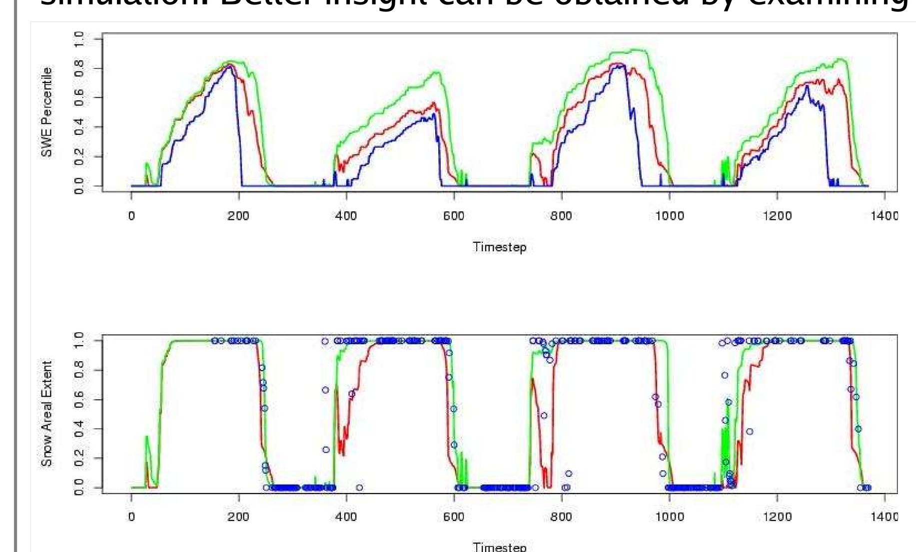

Figure 5. Comparison of SWE percentiles between prior estimate (red line), enKF (green line) and SNOTEL (blue line) for the West Yellowstone SNOTEL station (upper panel). The lower panel displays the model-predicted SCA for VIC (red) and enKF (green), and actual MODIS observations (blue points).

This can also be seen in this table, which shows the SWE RMSE averaged for all SNOTEL stations for different elevation zones and accumulation/ablation periods. We can see that the assimilation had the smallest impact at higher elevations, which can be explained by the fact that snow coverage tends to be 100% most of the time at the highest elevations, and therefore the updates have a smaller effect on SWE. On the contrary, for mid and lower elevations the impact of the assimilation is larger either negative (accumulation) or positive (snowmelt).

	Lower Elevation		Mid Elevation		Higher Elevation	
	Accu-mula-tion	Ablation	Accu-mula-tion	Ablation	Accu-mula-tion	Ablation
VIC	0.144	0.285	0.143	0.310	0.175	0.250
enKF	0.192	0.248	0.188	0.261	0.190	0.243

An important issue arises when assimilating snow observations for streamflow prediction. The data assimilation updates SWE and SCE by compensating for errors in temperature and precipitation forcings. Assuming that the model temperature is biased positively, the model will tend to melt the snowpack earlier. The assimilation of a snow observation will restore the snowpack to its ``true'' state but at the same time it will introduce water balance errors. Similarly, in a cold biased simulation, the model snowpack will persist as long as no observation is available to correct the SWE estimate. The magnitude of the water balance errors will depend on the assimilation frequency. Therefore, it is essential to remove such biases for an operational snow data assimilation application. This can be achieved by constraining the model error (namely precipitation and temperature forcings) as well as the model states, i.e. augmenting the state vector and incorporating temporal correlation to the stochastic forcings.

REFERENCES

- Anderson, E.A. National Weather Service River Forecast System - Snow Accumulation and Ablation Model, *NOAA Technical Memorandum NWS HYDRO-17*, 1973.
- Cherkauer, K.A., and D.P. Lettenmaier, Simulation of spatial variability in snow and frozen soil, *J. Geophys. Res., 108* (D22), 8858, doi:10.1029/2003JD003575, 2003.
- Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res., 99* (C5), 10 143-10162, 1994.
- Gelb, A., Ed., *Applied optimal estimation.* The MIT Press, 1974.
- Liang, X., Lettenmaier, D.P., E.F. Wood, and S.J. Burges,. "A simple hydrologically based model of land surface water and energy fluxes for general circulation models." *J. Geophys. Res.*, 99(D7), 14,415-14,428, 1994.
- Nijssen, B., and D.P. Lettenmaier, Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites, *J. Geophys. Res., 109*, D02103, doi:10.1029/2003JD003497, 2004.