

## RESEARCH ARTICLE

10.1002/2013JD021329

## Key Points:

- This work interfaced CLM4 with DART
- MODIS snow cover is assimilated into DART/CLM4
- The RMSE of snow cover and snow depth is reduced

## Correspondence to:

Z.-L. Yang,  
liang@jsg.utexas.edu

## Citation:

Zhang, Y.-F., T. J. Hoar, Z.-L. Yang, J. L. Anderson, A. M. Toure, and M. Rodell (2014), Assimilation of MODIS snow cover through the Data Assimilation Research Testbed and the Community Land Model version 4, *J. Geophys. Res. Atmos.*, 119, 7091–7103, doi:10.1002/2013JD021329.

Received 7 DEC 2013

Accepted 28 MAY 2014

Accepted article online 2 JUN 2014

Published online 18 JUN 2014

## Assimilation of MODIS snow cover through the Data Assimilation Research Testbed and the Community Land Model version 4

Yong-Fei Zhang<sup>1</sup>, Tim J. Hoar<sup>2</sup>, Zong-Liang Yang<sup>1</sup>, Jeffrey L. Anderson<sup>2</sup>, Ally M. Toure<sup>3,4</sup>, and Matthew Rodell<sup>4</sup>

<sup>1</sup>Department of Geological Sciences, John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA, <sup>2</sup>National Center for Atmospheric Research, Boulder, Colorado, USA, <sup>3</sup>Universities Space Research Association, Columbia, Maryland, USA, <sup>4</sup>NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

**Abstract** To improve snowpack estimates in Community Land Model version 4 (CLM4), the Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover fraction (SCF) was assimilated into the Community Land Model version 4 (CLM4) via the Data Assimilation Research Testbed (DART). The interface between CLM4 and DART is a flexible, extensible approach to land surface data assimilation. This data assimilation system has a large ensemble (80-member) atmospheric forcing that facilitates ensemble-based land data assimilation. We use 40 randomly chosen forcing members to drive 40 CLM members as a compromise between computational cost and the data assimilation performance. The localization distance, a parameter in DART, was tuned to optimize the data assimilation performance at the global scale. Snow water equivalent (SWE) and snow depth are adjusted via the ensemble adjustment Kalman filter, particularly in regions with large SCF variability. The root-mean-square error of the forecast SCF against MODIS SCF is largely reduced. In DJF (December–January–February), the discrepancy between MODIS and CLM4 is broadly ameliorated in the lower-middle latitudes (23°–45°N). Only minimal modifications are made in the higher-middle (45°–66°N) and high latitudes, part of which is due to the agreement between model and observation when snow cover is nearly 100%. In some regions it also reveals that CLM4-modeled snow cover lacks heterogeneous features compared to MODIS. In MAM (March–April–May), adjustments to snow move poleward mainly due to the northward movement of the snowline (i.e., where largest SCF uncertainty is and SCF assimilation has the greatest impact). The effectiveness of data assimilation also varies with vegetation types, with mixed performance over forest regions and consistently good performance over grass, which can partly be explained by the linearity of the relationship between SCF and SWE in the model ensembles. The updated snow depth was compared to the Canadian Meteorological Center (CMC) data. Differences between CMC and CLM4 are generally reduced in densely monitored regions.

### 1. Introduction

Snow plays a unique role in the global hydrological cycle, water resources management, and atmospheric predictability. Its special physical properties (high albedo, low thermal conductivity, and ability to change phase) significantly modulate energy and water exchanges between the atmosphere and the land surface [Goodison *et al.*, 1999]. In regions where streamflow is dominated by snowmelt, the performance of hydrological forecasts largely depends on snowpack estimates at the beginning of the forecast period [Clark and Hay, 2004]. Snowpack acts as a key boundary condition for the atmosphere and influences atmospheric predictability. A more realistic simulated snowpack enhances springtime surface air temperature predictability [e.g., Peings *et al.*, 2010]. Furthermore, snowpack impacts atmospheric circulations through teleconnections. Numerous modeling and observational studies have shown an inverse relationship between the winter and springtime Eurasian snow-covered area and the summertime Indian Monsoon rainfall [e.g., Vernek *et al.*, 1995; Bamzai and Shukla, 1999; Turner and Slingo, 2011].

A variety of snowpack products have been generated for hydroclimatic analysis and evaluation of climate models. Ground measurements usually lack spatial representativeness, especially in regions of high heterogeneity [Liston, 2004], and are difficult to obtain in many regions especially in complex terrains; therefore, satellite remote sensing plays an important role in producing global snowpack estimates. Based on the optical properties of snow, observations of visible and near-infrared bands can detect snow extent in

most land surfaces [e.g., Hall *et al.*, 2002], and observations of passive and active microwave can estimate snow mass [e.g., Chang *et al.*, 1982]. However, satellite remote sensing has large errors in certain circumstances. For example, visible and near-infrared observations cannot discriminate snow from clouds [Hall and Riggs, 2007], and microwave-based snow retrievals will greatly underestimate snow mass when water is present in snow [Foster *et al.*, 2005]. Moreover, products from different observing systems may show discrepancies due to their varied retrieval methods and spatial resolutions. With the absence of ground truth, comparing these products is difficult. For example, year-round disagreements among products are reported in the Tibetan Plateau [Savoie *et al.*, 2007].

Land surface models (LSMs) offer another alternative whose advantages include providing spatially and temporally continuous land states that are energy and water balanced. But LSMs contain inevitable errors resulting from biased input forcing data, simplified model structure, and imperfect parameterization schemes [Kato *et al.*, 2007]. We will show that incorporating satellite snow observations into state-of-the-art LSMs is a way to obtain more accurate and spatially temporally continuous snow products.

Data assimilation techniques, which statistically combine model forecasts with observations based on their uncertainties, have been developed to improve our estimates of snowpack [e.g., Andreadis and Lettenmaier, 2006; Slater and Clark, 2006; Su *et al.*, 2008]. A variety of observations have been assimilated into LSMs at regional and continental scales, showing promising results [e.g., Su *et al.*, 2010; De Lannoy *et al.*, 2012]. This study differs from previous studies in the way that we apply a new data assimilation framework with a large ensemble atmospheric forcing and provides future potentials for testing multiple ensemble-based data assimilation algorithms. We also provide analysis of data assimilation performance in various perspectives, which offers ideas in improving snow data assimilation for future work.

This study employs the Data Assimilation Research Testbed (DART) to assimilate the Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover fraction (SCF) data into the Community Land Model version 4 (CLM4) and update snow water equivalent (SWE) and snow depth. Background information including data assimilation techniques, CLM4, DART, and the linkage between DART and CLM4 is provided in section 2. Observational data assimilated into our DART/CLM4 system, meteorological forcings for CLM4, and experimental design are discussed in section 3. Results are analyzed in section 4, followed by conclusions in section 5.

## 2. Background

### 2.1. The Ensemble Adjustment Kalman Filter and DART

The ensemble Kalman filters (EnKFs) uses a Monte Carlo method to estimate model forecast error statistics [Evensen, 1994]. The DART has incorporated a variety of ensemble-based filters, e.g., the traditional EnKF, the deterministic ensemble adjustment Kalman filter (EAKF), and a non-Gaussian rank histogram filter. The EAKF is used in this paper; equations and implementations are documented in Anderson [2001]. As opposed to the traditional EnKF that adds a random perturbation to observation, the EAKF is a deterministic algorithm based on the observation uncertainty and the estimated model forecast error.

Land data assimilation algorithms are commonly embedded within LSMs [e.g., Su *et al.*, 2010], which makes it difficult to maintain the codes if LSMs or data assimilation algorithms are updated frequently. The DART developed at the National Center for Atmospheric Research (NCAR) has overcome this difficulty by developing a well-organized software environment with various ensemble-based assimilation methods that are independent of the details of a dynamical model or biogeophysical model. DART has already been coupled with the Community Atmospheric Model (CAM4) [Raeder *et al.*, 2012] and produced an 80-member ensemble atmospheric reanalysis. The atmospheric ensemble reanalysis output can provide ensembles of atmospheric forcing fields for land or ocean ensemble assimilations. The atmospheric reanalysis samples the uncertainty associated with atmospheric model error and limited atmospheric observations while maintaining appropriate relationships (for instance, approximate geostrophic balance) between the variables in each ensemble member. Many traditional land data assimilation studies add random perturbations to a single set of forcing fields (mainly precipitation and temperature) and may not retain the appropriate balances between different atmospheric variables. Besides providing basic ensemble data assimilation algorithms, DART also includes other algorithms necessary to produce high

quality assimilations in high-dimensional models. More details about DART are documented in an overview paper [Anderson *et al.*, 2009].

## 2.2. CLM4 and Its Snow Model

CLM4 [Lawrence *et al.*, 2011], the land component of the Community Earth System Model (CESM) [Gent *et al.*, 2011], is used in this study for several reasons. First, CLM is a land surface model that benefits from the efforts of the worldwide community. Active improvements and refinements are continuously being added into CLM. Second, CLM4 is a state-of-the-art model that has an advanced hydrology scheme. It deepens soil columns to 42 m using 15 layers, with the first 10 layers (3.8 m) hydrologically active and the bottom 5 layers accounting for thermal interactions with the underlying deep ground. Third, CLM4 has a sophisticated snow parameterization scheme that simulates snowpack with up to five layers depending on the snowpack's thickness. Besides internal physical processes such as water-heat transport, thawing-freezing, liquid water retention, and densification, the CLM4 snow model also accounts for snowpack radiation properties by coupling to the Snow and Ice Aerosol Radiation model [Flanner and Zender, 2005, 2006; Flanner *et al.*, 2007]. Climate models parameterize SCF as a function of grid-mean snow depth, and this functional relationship takes a wide range of forms in literature [Liston, 2004]. Note that these SCF and snow depth relationships are equivalent to the snow depletion curves in watershed-scale snow modeling [Luce and Tarboton, 2004]. Niu and Yang [2007] parameterized SCF as a function of snow density and snow depth based on monthly analyses; this SCF depletion curve was implemented in CLM4 and generates more realistic SCF [Lawrence *et al.*, 2011]. This snow depletion curve is used as the observation operator in the EAKF to simulate the observed state (SCF) based on prognostic variables (snow depth) in CLM4.

CLM4 is run at  $0.9^\circ \times 1.25^\circ$ . The spatial land surface heterogeneity in CLM4 enables that each grid may have multiple columns to capture potential variability in the snow state variables within a grid cell [Oleson *et al.*, 2010]. We area-weight all the columns of SCF in a grid cell to get an observational estimate, which is suboptimal when the observation grid is different from the model grid (e.g., site observations). We directly update SWE via data assimilation and adjust snow depth based on the physical relationship between SWE and snow depth, i.e., the change in snow depth is the change in SWE divided by snow density. Snow density is calculated as prior snow mass divided by prior snow depth. The CLM snow-layer division and combination is handled by comparing the snow depth in each layer with its predetermined minimum and maximum values. From the assimilation perspective, it is desirable to adjust SWE directly and then redistribute the new SWE to the snow depth in each layer according to the above snow-layering scheme.

## 2.3. DART and CLM4

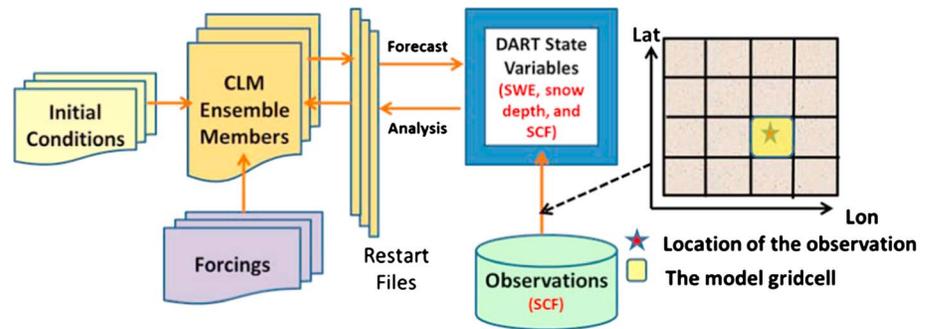
The CESM version 1\_1\_1 fully supports a "multiinstance" capability that enables multiple instances of model components to be run within a single executable. This capability is leveraged by the CLM4/DART assimilation facility. The ensemble of CLM4 instances is stopped every 24 h (at 00 UTC), and restart files are written. DART then reads a subset of the variables from the restart file (the prior or forecast) and a set of observations for that time. An observation operator projects the model states to the observation space, and DART performs assimilation. The updated variables (the posterior or analysis) are then inserted back into the CLM4 restart file to be used for the next forecast cycle (Figure 1).

The snow depletion curve in CLM4 parameterizes SCF, a diagnostic variable, as a function of snow density and snow depth, serving as the observation operator in the EAKF. The model resolution is  $0.9^\circ \times 1.25^\circ$ , and MODIS observations are scaled up to the same resolution. DART only needs to find the grid cells of the observations and performs assimilations there. In rare occasions where CLM4 ensemble members are not consistent in regard to the snow presence, i.e., some members may predict that it is snow-free in a particular grid while others may predict that there is a shallow snow layer in that grid, we do not perform data assimilation.

## 3. Experimental Setup and Data Sets

### 3.1. MODIS Satellite SCF Observations

We assimilated the MODIS SCF observations into CLM4 through DART. MODIS sensors (<http://modis.gsfc.nasa.gov/>) detect snow cover based on the property that snow has higher reflectance in visible bands and lower reflectance in near-infrared bands [Salomonson and Appel, 2004]. This retrieval algorithm is



**Figure 1.** Flowchart showing the coupling of DART and CLM4.

supplemented by several approaches to map snow on cloudy days [Parajka and Blöschl, 2008; Hall *et al.*, 2010].

The MODIS/Terra daily,  $0.05^\circ$  resolution global climate modeling grid snow products (MOD10C2; <http://modis-snow-ice.gsfc.nasa.gov/?c=MOD10C2>) are used in this study. The absolute accuracy of MODIS SCF observations can be about 90% but really depends on snow conditions and land cover types [Hall and Riggs, 2007]. The accuracy is largely limited under cloudy conditions, when trace amounts of snow are present, and over complex terrain and forested regions. To take into account the data quality problem associated with clouds, we followed Rodell and Houser [2004] to upscale the MODIS SCF data from its original resolution to the model resolution ( $0.9^\circ \times 1.25^\circ$ ), discarding the proxy observations with sky visibility less than 20% and average all the other proxy observations within one model grid cell to obtain the MODIS-observed snow cover of that grid cell. A stationary observation error, 0.1, is chosen as suggested by previous studies [e.g., Andreadis and Lettenmaier, 2006].

### 3.2. Meteorological Forcings From DART/CAM4

We use an atmospheric ensemble reanalysis to introduce uncertainties to the snow states in CLM4; the ensemble reanalysis products are from DART/CAM4 and preserve the full covariance of each model state while maintaining variability consistent with observational uncertainty. CAM4 is the atmospheric component of the Community Climate System Model, version 4, a general circulation model used widely for both past and current climate studies as well as climate projections [Gent *et al.*, 2011]. DART/CAM4 [Raeder *et al.*, 2012] assimilated observations that are used in the NCEP-NCAR reanalysis plus radio occultation observations from the Constellation Observing System for Meteorology Ionosphere and Climate [Anthes *et al.*, 2008]. The CAM4 ensemble reanalyses have been used to force version 2 of the Parallel Ocean Program (POP2) [Danabasoglu *et al.*, 2012], generating a reasonable ensemble spread and a significantly improved POP2 analysis. We assume that the ensemble spread in CLM4 members is mainly from the meteorological forcing, which is consistent with previous studies [e.g., Carpenter and Georgakakos, 2004; Slater and Clark, 2006] indicating that the uncertainty of hydrological models is dominated by forcing uncertainty. Investigating uncertainties caused by model structures or parameterization schemes is beyond the scope of this study. We use 40 out of the 80-member DART/CAM4 reanalyses randomly to drive the same number of CLM4 ensemble members. We chose 40 ensemble members as a compromise between computational cost and the EAKF performance [e.g., Reichle *et al.*, 2002].

### 3.3. Independent Observation-Based Snow Product for Comparison

The Canadian Meteorological Center (CMC) produces one of the few global snow maps (<http://nsidc.org/data/nsidc-0447>) incorporating site observations, aircraft detections, and snow models [Brown *et al.*, 2003]. CMC snow depth is generated based on a 6 hourly optimal interpolation of extensive in situ snow depth reports from the World Meteorological Organization information system. To obtain a reasonably representative data set, observations taken at elevations 400 m higher or lower than the grid-average elevation are not used, although there is still a low-elevation bias in the data [Brasnett, 1999]. The CMC snow map has been recognized as the best available global data and is used by many studies to validate model results [e.g., Su *et al.*, 2010; Reichle and Koster, 2011]. Daily snow depth and seasonal snow maps were compared directly to our model results.

**Table 1.** List of Eight Experiments With Localization Distance (LD)

LD Experiments	LOC0.01	LOC0.03	LOC0.05	LOC0.07	LOC0.1	LOC0.15	LCC0.2	LOC0.3
Radians	0.01	0.03	0.05	0.07	0.1	0.15	0.2	0.3
Kilometers	60	180	300	420	600	900	1200	1800

### 3.4. Experiments

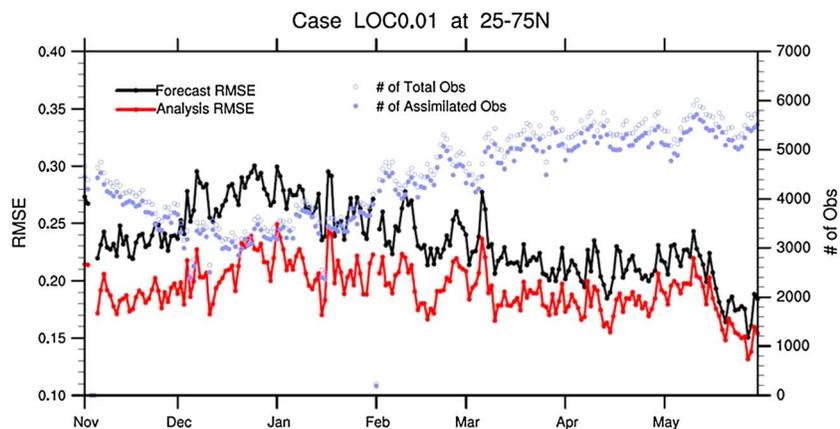
In addition to the open-loop case that uses the standard CLM4 settings, we conducted eight experiments to tune the localization distance [Gaspari and Cohn, 1999] to optimize the data assimilation performance. The experiments are listed in Table 1.

The localization distance limits the influence of observations to nearby grid cells. The influence of observation on nearby grid cells is quantified by their correlation multiplied by a regression weight factor [Gaspari and Cohn, 1999] that decreases with distance. The regression weight decreases to zero at 2 times the localization distance. The significance of spreading information from the observed to the unobserved grids through the background correlations in land data assimilation is identified by Reichle and Koster [2003]; however, spurious correlations in two widely spread variables, due to a limited ensemble size [Anderson, 2007], will degrade the data assimilation performance [Hamill et al., 2001]. The localization distance should be properly set to avoid spurious error correlations and maximize the value of observations. Based on our knowledge of the correlations between horizontal grid cells of CLM4, we tested a set of localization distances (0.01, 0.03, 0.05, 0.07, 0.1, 0.15, 0.2, and 0.3 rad) and chose the distance that optimizes the data assimilation performance as defined in section 4.1.

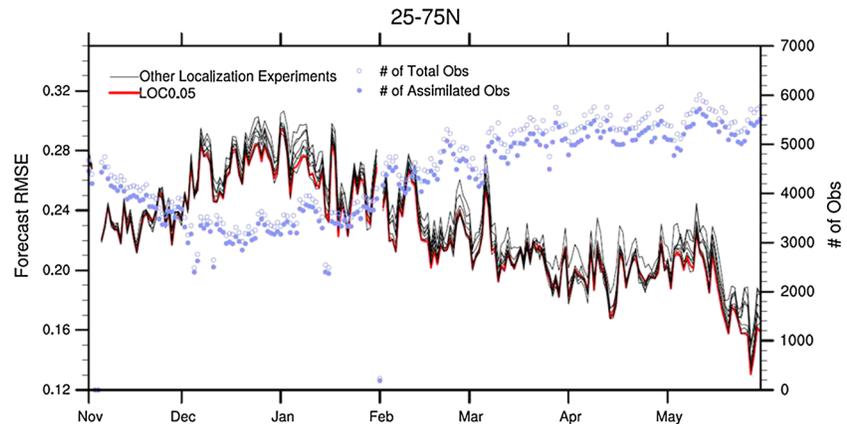
## 4. Results and Discussion

### 4.1. Observation Space Diagnostics

The effectiveness of the data assimilation can be evaluated in “observation space” by comparing the model forecast estimates of the observations to the actual observations. We used root-mean-square error (RMSE), the square root of the average squared difference between the model estimates and observations, to evaluate the data assimilation performance. The evolution of daily RMSE of SCF (Figure 2) shows how the model forecast errors are reduced by the assimilation of observations. The RMSE has little variability as a function of time, indicating that the assimilation is stable, neither drifting away from the observations nor converging to the observations. Figure 2 also shows that the RMSE varies with the number of observations. November and early December have fewer observations which can be associated with increased RMSE in the same time period. Large cloud obscurations are found over the extratropics in the Northern Hemisphere



**Figure 2.** Evolution of daily RMSE of SCF in the latitudinal bands spanning from 25° to 75°N. The black curve is for the forecast (prior) and the red is for the analysis (posterior). Blue circles show the number of observations available, and blue solid dots show the number of observations that are actually assimilated. Time series spans from November 2002 to May 2003. Experiment LOC0.01 is chosen.



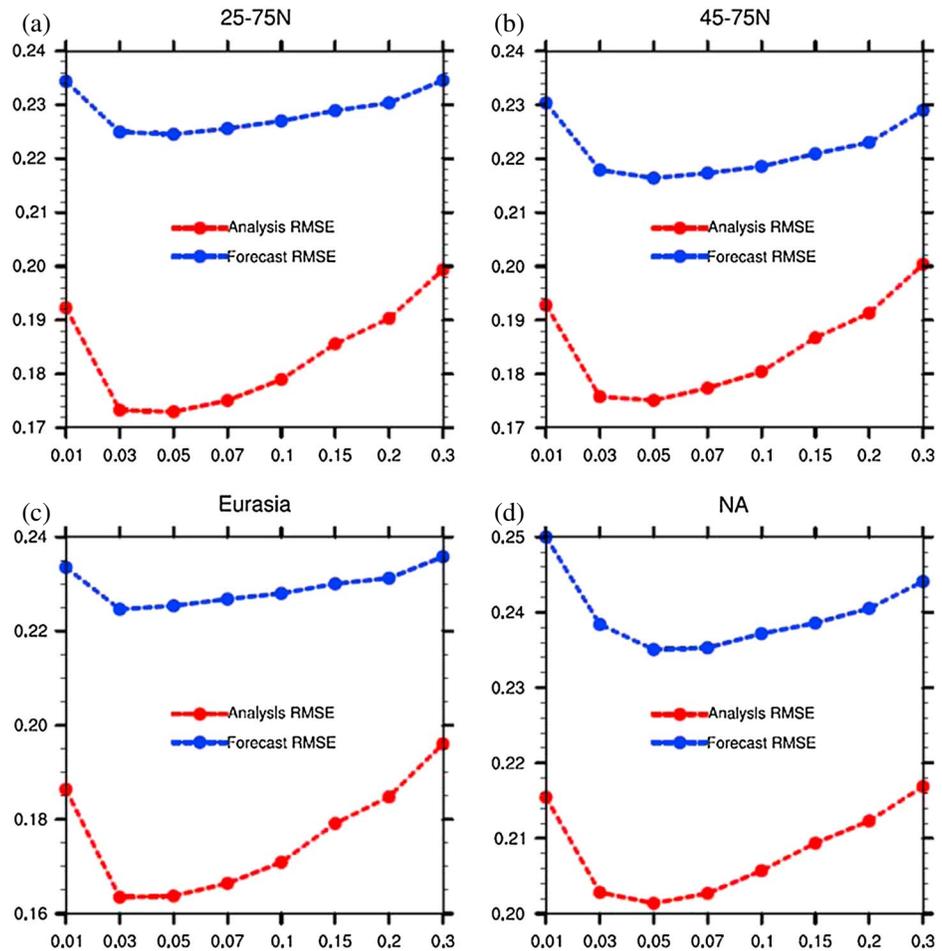
**Figure 3.** Evolutions of daily forecast (prior) RMSE of SCF in the latitudinal band spanning from 25° to 75°N for eight experiments, each has a different localization distance. The black curves are for seven experiments: LOC0.01, LOC0.03, LOC0.07, LOC0.1, LOC0.15, LOC0.2, and LOC0.3. The red line is for LOC0.05. The blue circles show the number of observations available, and the blue solid dots show the number of observations that are actually assimilated.

(figures not shown), impeding snow cover identification in this time period; consequently, the number of reliable SCF observations decreases. As observations become more numerous after late December, there is an obvious decreasing trend in the RMSE.

To determine the localization distance that optimizes the data assimilation performance, we analyzed eight localization experiments (Table 1); the daily forecast RMSE of SCF from these experiments in the latitudinal bands spanning from 25°N to 75°N are shown in Figure 3. The RMSE evolves in time with similar patterns, i.e., increasing as the number of observations decreases and vice versa. The spread of RMSE among these eight experiments becomes large after December. Case LOC0.05 is clearly superior to the other experiments as it produces the smallest RMSE for most days. To better quantify the behaviors of these seven experiments, we calculated area-averaged forecast and analysis RMSE for several regions (Figure 4). It confirms that LOC0.05 outperforms other experiments in the region of 25°N to 75°N, and in the regions of 45°N to 75°N and North America (NA). LOC0.03 produces slightly smaller RMSE than LOC0.05 in Eurasia, which demonstrates that the optimal localization distance may vary with space. Overall, LOC0.05 outperforms other localization distance experiments, which indicates that the horizontal error correlation length of snow states in CLM4 is roughly 0.05 rad (about 300 km). Since the error correlation in the land model is solely dependent on that of the atmospheric forcing, we believe the heterogeneity of localization distance is decided by the atmosphere model. Knowing that, we do not further explore the space-dependent localization distance and choose 0.05 rad (about 300 km) as the optimal localization distance that produces the smallest RMSE globally.

#### 4.2. Model Space Diagnosis

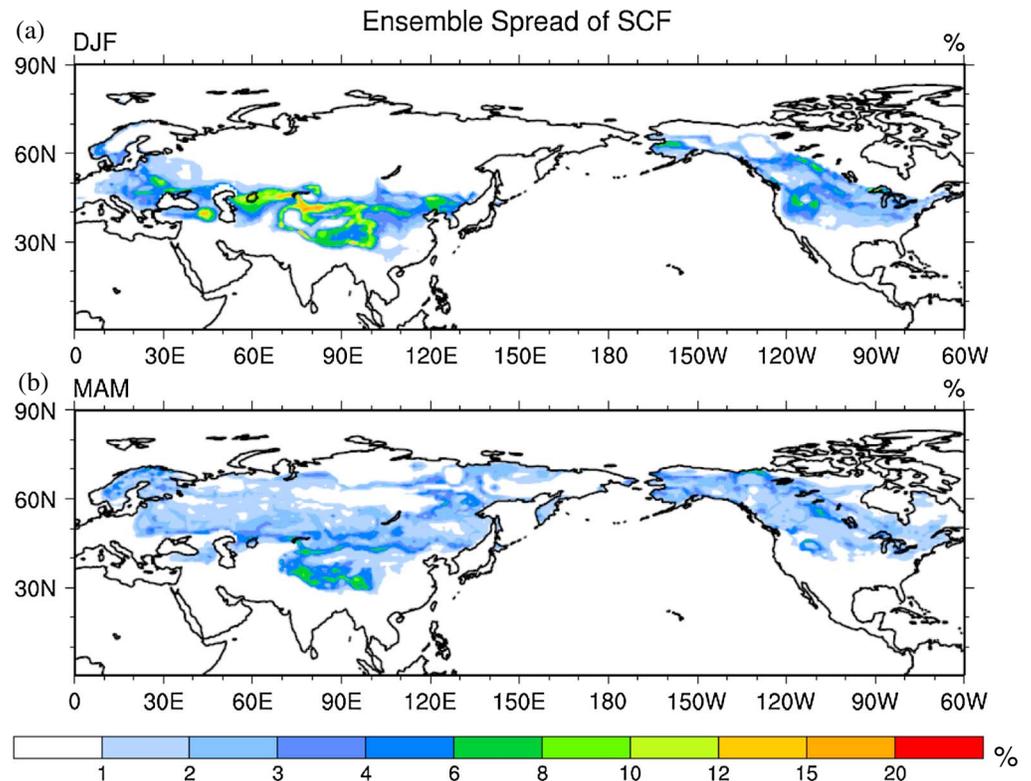
Based on the EAKF theory, data assimilation tends to reduce biases in regions with large model ensemble spread (i.e., uncertainty); hence, the ensemble spread pattern offers hints of where we expect the assimilation to have the greatest impact. In DJF (December–January–February; Figure 5a), the SCF spread is largest in the lower-middle latitudes (23°–45°N), with particularly high values in the Tibetan Plateau, northeastern China, the Caspian Sea and Black Sea region, and the Rocky Mountain areas. The SCF-ensemble spreads in the southern Alaska and Canadian boreal forest regions are also comparatively large. Spread in higher-middle (45°–66°N) and high latitudes is hardly observable; this is simply because the SCF in all the land model members increases to unity quickly in DJF. In MAM (March–April–May; Figure 5b), the ensemble spread increases both in the higher-middle and high latitudes. Conversely, it decreases in the lower-middle latitudes; values are still large in the Tibetan Plateau and part of the Rocky Mountain regions. Figure 5 suggests that the SCF uncertainty is large along snowline (typically over high-elevation regions and dense vegetation-covered regions). The CAM4-induced reanalysis has accordingly large snowfall and surface air temperature spread in high-elevation areas (e.g., the Tibetan Plateau, the Iranian Plateau, the Mongolian



**Figure 4.** Variations of forecast RMSE (blue dots) and analysis RMSE (red dots) for eight experiments with localization distances (radians on the X axis) for four regions: (a) 25°–75°N, (b) 45°–75°N, (c) Eurasia, and (d) NA.

Plateau, and the Rocky Mountains) and are very likely to pass a large uncertainty into CLM. In dense-vegetated regions, the interaction between plants and the land surface are complex, and the parameters are less constrained by observations, resulting in a large uncertainty in CLM4. For example, as small forcing uncertainty (in CAM4) is found in southern Alaska, large snow uncertainty is thus mostly due to the dense vegetation as it has a fairly large portion of needle-leaf-evergreen boreal trees. Based on the SCF spread pattern, we expect that data assimilation will mainly impact the lower-middle latitude regions in DJF and higher-middle and high latitudes in MAM.

Comparisons between modeled and MODIS-observed SCF reveal internal CLM4 biases and the effects of data assimilation. In DJF, a large portion of observations are unavailable due to cloud obscuration and lack of sunlight. We only compare the regions where we have observations throughout more than half of the season. The open-loop run overestimates SCF in the Tibet Plateau, northern and northeastern China, central Great Plains, and underestimates SCF in the southern part of the Rocky Mountains, compared to MODIS SCF (Figure 6a). Discrepancies between model and observations are reduced in the data assimilation run as would be expected (Figures 6c and 6e); however, assimilation updates are limited in the higher-middle and high latitude regions (e.g., central Eurasia, northern Alaska, and western Canada; Figure 7a), due to the small spread of the CLM4 ensembles in these regions and the lack of observations. Almost all ensemble members generate full snowpack and are no longer strongly influenced by the observations. In cases where model members indicate full snowpack while the observations do not, this could possibly be ameliorated by incorporating rule-based algorithms to manually adjust snowpack [e.g., Zaitchik and Rodell, 2009; De Lannoy et al., 2012] or by adding some spread to the ensembles. In

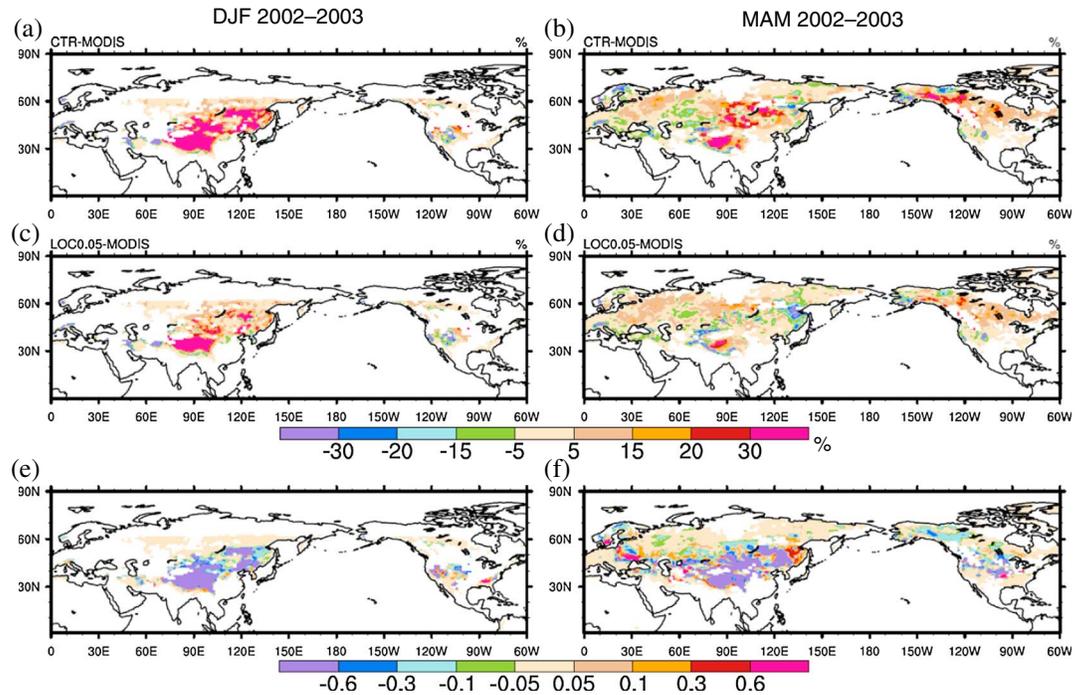


**Figure 5.** The ensemble spread of SCF averaged for (a) DJF and (b) MAM in 2002–2003. The ensemble spread is calculated as the standard deviation in SCF among 40 ensemble members. SCF values are averaged for two seasons before calculating the ensemble spread.

MAM, we have more observations available (Figures 6b and 6d). Data assimilation reduces SCF biases in most observed regions (e.g., the Tibetan Plateau and northern China) but increases SCF biases in some particular regions (e.g., northeastern China; Figure 6f); the adjustments move poleward from DJF to MAM as expected (Figures 7a and 7c). Increased SCF biases in some of the regions suggest that even though we insert a SWE estimate that incorporates MODIS SCF information, a more accurate SCF estimate is not guaranteed.

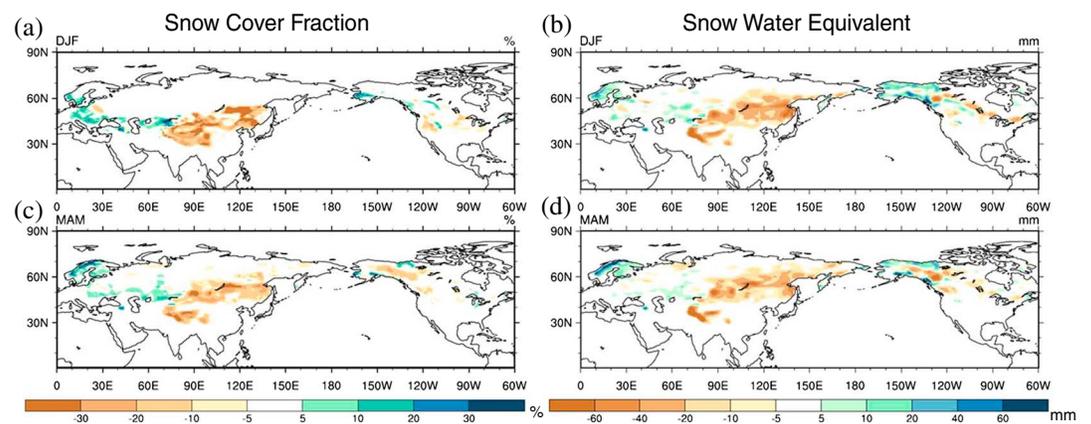
The overall modification of SWE is consistent with that of SCF (Figures 7b and 7d); however, the poleward shift of the pattern from DJF to MAM is not apparent for SWE, which can be attributed to the nonlinear relationship between SWE and SCF in CLM4. As stated in section 2.2, we directly updated the unobserved SWE given the observed SCF through the EAKF, and let the model physics and parameterization schemes propagate the observation information. Hence, the data assimilation performance partially depends on the snow depletion curve. In the lower-middle latitudes, since SCF is generally less than 80%, SCF is very sensitive to SWE and is more sensitive during the snow accumulation season (i. e., DJF; Figure 8a) than during the snow melting season (i.e., MAM). For example, SWE is reduced by almost the same amount in DJF and MAM in the Tibetan Plateau, but SCF is reduced much more in DJF than in MAM. In the higher-middle and high latitudes, SCF is much larger and closer to unity, especially in DJF; as a consequence, SCF is not sensitive to SWE (Figure 8b). Although SWE is changed by a fair amount, SCF remains almost unchanged (Figures 7c and 7d; e.g., over the vast Siberia region).

The EAKF linearly regresses the SCF observational increments onto the increments of the unobserved SWE [Anderson *et al.*, 2009]. As a result, the linearity of the relationship between the ensemble SWE and the ensemble SCF is important in limiting the error of the EAKF. Figure 9 shows the relationship between SWE and SCF over two typical grid cells, one over grass and the other over boreal forest. SCF usually reaches unity quickly in DJF in boreal forest. In the case that the SCF of all the ensemble members equals one, no ensemble spread is present, and hence the MODIS-observed SCF has no effect on the model. However, some ensemble

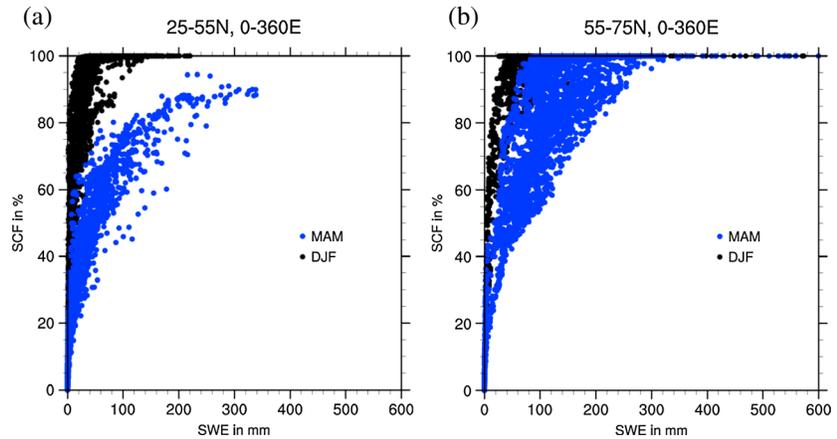


**Figure 6.** SCF biases (i.e., modeled SCF minus MODIS SCF) for the (a and b) open-loop runs and the (c and d) data assimilation runs. (e and f) The differences (i.e., data assimilation minus open loop) in the normalized absolute SCF biases. Figures 6a, 6c, and 6e are for DJF, and Figures 6b, 6d, and 6f are for MAM. Normalized absolute SCF biases are defined as the absolute biases between modeled SCF and MODIS SCF, divided by MODIS SCF. The blank areas in the plots indicate regions where MODIS observations are missing for more than half of the season.

members may diverge from the majority as suggested by Figure 9a. The ensemble spread of SCF is still present but is not Gaussian, nor is the relationship between SCF and SWE linear. The situation becomes better in MAM when snow starts to melt (Figure 9b). Figures 9c and 9d display the relationship between SCF and SWE over one grass-dominant grid cell for DJF and MAM, respectively. The relationship is quite linear, though not perfectly. Consequently, the data assimilation performance may be limited, or even degraded, over boreal forest in DJF, and will show better quality in boreal forest in MAM and over grassland. Quality controls could be introduced to discard the SWE increments when large errors appear (e.g., the linear regression coefficient between SCF and SWE is negative).



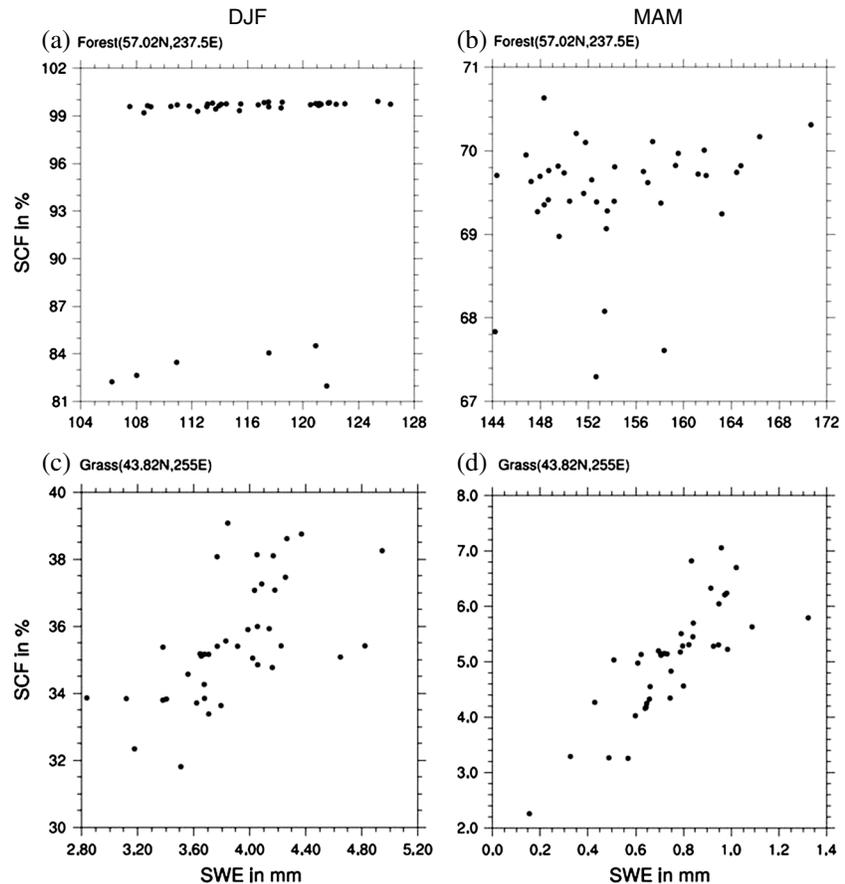
**Figure 7.** Seasonal mean differences (data assimilation minus open loop) for (a and c) SCF and (b and d) SWE. Figures 7a and 7b are for DJF, and Figures 7c and 7d are for MAM.



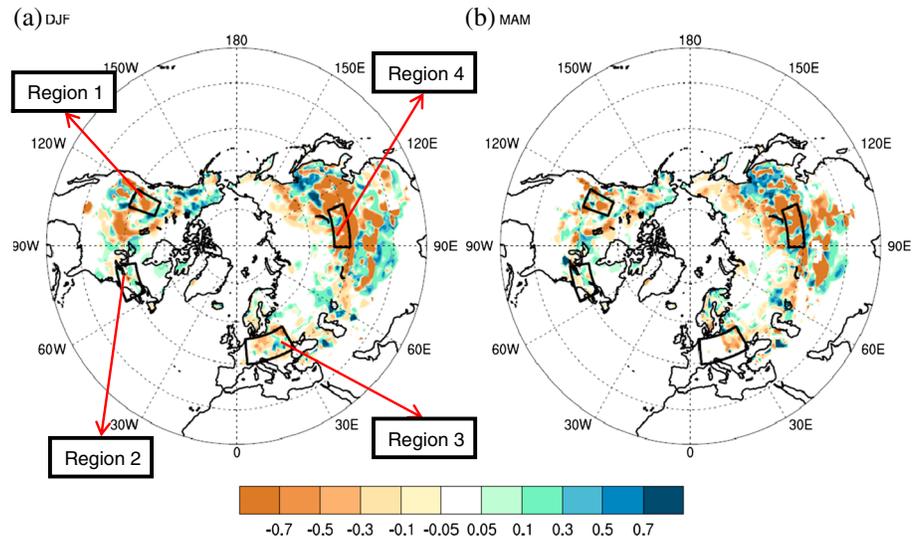
**Figure 8.** Seasonal mean SCF-SWE relationship for two latitudinal bands spanning (a) from 25° to 55°N and (b) from 55° to 75°N. The X axis is SWE in mm, while the Y axis shows the percentage of snow-covered area. Each black dot represents model states over a particular grid for DJF, and each blue dot represents model states over a particular grid for MAM.

**4.3. Comparisons With CMC Snow Depth**

Our modeled snow depth was evaluated using CMC data. The normalized absolute difference between assimilated output and CMC data is compared to the normalized absolute difference between the open-loop output and CMC data (Figure 10). Negative values in Figure 10 indicate reduced difference between CLM4

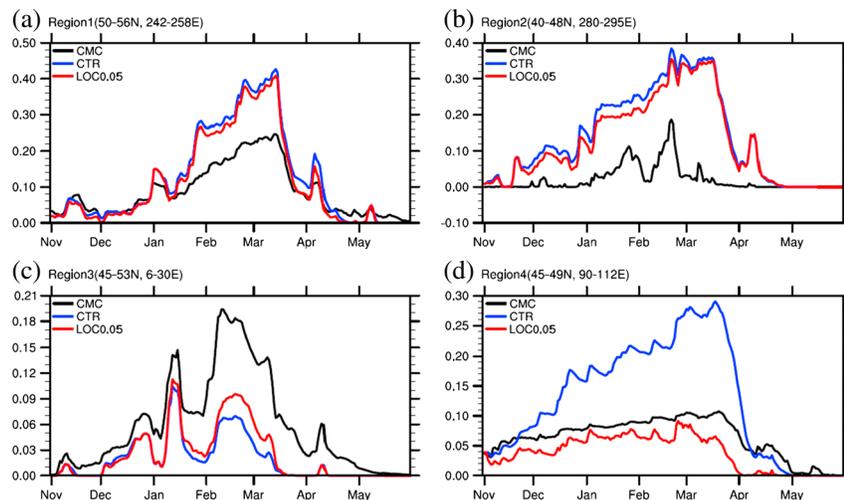


**Figure 9.** Seasonal mean SCF-SWE relationship over one typical forest grid for (a) DJF and (b) MAM, and one typical grass grid for (c) DJF and (d) MAM. Each dot in each plot represents one ensemble member.



**Figure 10.** Differences (data assimilation minus open loop) in the normalized absolute snow depth biases for (a) DJF and (b) MAM. Normalized absolute snow depth biases are defined as the absolute biases between modeled snow depth and CMC snow depth, divided by CMC snow depth.

and the CMC snow depth (i.e., improvement), and positive values indicate degradation. Considering that the CMC network is sparse in much of the Northern Hemisphere, large differences between model and CMC snow depth do not necessarily indicate large model bias. Referring to the site-density map of the CMC data [Reichle and Koster, 2011, Figure 9c], we compared our model results to the CMC data in densely monitored regions, which include eastern Europe, the northern Great Plains in Canada, the northeastern United States, and northern Asia (highlighted in Figure 10 with rectangles). Differences between the model and CMC data are generally reduced after assimilating MODIS in these highlighted regions. The time series comparisons of snow depth from the model runs and from CMC (Figure 11) confirm that the data assimilation results match CMC data better. Over some portions of the observation-sparse regions, typically in dense forests and high-elevation regions, the data assimilation results are further away from the CMC data. A good example is over the Tibetan Plateau, where various snow products are often found to disagree [Fret et al., 2012]. Future work on reducing the uncertainty of simulated snowpack in data-sparse regions is needed.



**Figure 11.** Time series of snow depth (in meters) in four regions (the highlighted rectangles in Figure 10). The black curve represents CMC data, the blue for the open-loop runs, and the red for the data assimilation runs (the localization distance is set to be 0.05 rad).

## 5. Conclusions

This study developed an extensible system, DART/CLM4, for land data assimilation, and tested the system by assimilating MODIS snow cover fraction (SCF) observations. This system is flexible in that land surface models (e.g., CLM) and data assimilation methods (DART) can each be developed independently. The system can be easily updated when new versions of CLM are available or when DART includes new data assimilation algorithms. The large ensemble of reanalysis products used for atmospheric forcing variables overcome the random perturbations introduced in conventional data assimilation approaches. In addition, DART allows tuning of the localization distance to maximize the value of observations and to ameliorate spurious error correlations.

Compared to the open-loop simulations, the RMSE of SCF is largely reduced when MODIS data are assimilated. The ensemble spread of SCF for DJF (December-January-February) is mainly located in the lower-middle latitudes, while for MAM (March-April-May), the ensemble spread extends northward. Consequently, the data assimilation of MODIS SCF is exclusively effective in the lower-middle latitudes in DJF, while notable adjustments appear in the higher-middle and high latitudes in MAM. Because of the small ensemble size, sampling errors exist when using the SCF observational increments to calculate the SWE increments, especially in forests. Compared to the open-loop results, the snow depth simulations from the data assimilation runs are generally closer to CMC data in the densely observed regions but are farther away from the CMC data in some portions of the Northern Hemisphere where observations are sparse (e.g., in dense forests and high-elevation regions). Further work on improving the data assimilation effectiveness in DJF and controlling the errors in the relationship between the ensemble SCF and snow water equivalent (SWE) is needed.

## Acknowledgments

This work is funded by NASA grants NNX09AJ48G and NNX11AJ43G, the NCAR Advanced Study Program, and NSFC grant 91337217. Andrew Fox is thanked for discussing the implementation of DART/CLM4. The DART/CAM4 reanalysis data are prepared by Kevin Raeder (raeder@ucar.edu). Nancy Collins is thanked for part of the coding with DART. MODIS snow data and CMC snow depth data are provided by the National Snow and Ice Data Center. We would like to thank Barton Forman and two anonymous reviewers for their constructive comments.

## References

- Anderson, J. L. (2001), An ensemble adjustment Kalman filter for data assimilation, *Mon. Weather Rev.*, *129*, 2884–2903.
- Anderson, J. L. (2007), An adaptive covariance inflation error correction algorithm for ensemble filters, *Tellus*, *59*(2), 210–224.
- Anderson, J. L., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano (2009), The data assimilation research testbed: A community facility, *Bull. Am. Meteorol. Soc.*, *90*(9), 1283–1296.
- Andreadis, K. M., and D. P. Lettenmaier (2006), Assimilating remotely sensed snow observations into a macroscale hydrology model, *Adv. Water Resour.*, *29*, 872–886.
- Anthes, R. A., et al. (2008), The COSMIC/FORMOSAT-3 mission: Early results, *Bull. Am. Meteorol. Soc.*, *89*, 313–333.
- Bamzai, A. S., and J. Shukla (1999), Relationship between Eurasian snow cover, snow depth, and the Indian summer monsoon: An observational study, *J. Clim.*, *12*, 3117–3132.
- Brasnett, B. A. (1999), Global analysis of snow depth for numerical weather prediction, *J. Appl. Meteorol.*, *38*, 726–740.
- Brown, R. D., B. Brasnett, and D. Robinson (2003), Gridded North American monthly snow depth and snow water equivalent for GCM evaluation, *Atmos. Ocean*, *41*(1), 1–14.
- 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.
- Chang, A. T. C., J. L. Foster, D. Hall, A. Rango, and B. Hartline (1982), Snow water equivalent determination by microwave radiometry, *Cold Reg. Sci. Technol.*, *5*, 259–267.
- 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.
- Danabasoglu, G., S. Bates, B. P. Briegleb, S. R. Jayne, M. Jochum, W. G. Large, S. Peacock, and S. G. Yeager (2012), The CCSM4 ocean component, *J. Clim.*, *25*, 1361–1389, the CCSM4 Special Collection. [Available at <http://journals.ametsoc.org/page/CCSM4/CESM1>.]
- De Lannoy, G. J. M., R. H. Reichle, K. R. Arsenault, P. R. Houser, S. Kumar, N. E. C. Verhoest, and V. R. N. Pauwels (2012), Multiscale assimilation of Advanced Microwave Scanning Radiometer–EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, *Water Resour. Res.*, *48*, W01522, doi:10.1029/2011WR101588.
- Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162, doi:10.1029/94JC00572.
- Flanner, M. G., and C. S. Zender (2005), Snowpack radiative heating: Influence on Tibetan Plateau climate, *Geophys. Res. Lett.*, *32*, L06501, doi:10.1029/2004GL022076.
- Flanner, M. G., and C. S. Zender (2006), Linking snowpack microphysics and albedo evolution, *J. Geophys. Res.*, *111*, D12208, doi:10.1029/2005JD006834.
- Flanner, M. G., C. S. Zender, J. T. Randerson, and P. J. Rasch (2007), Present-day climate forcing and response from black carbon in snow, *J. Geophys. Res.*, *112*, D11202, doi:10.1029/2006JD008003.
- Foster, J. L., C. Sun, J. P. Walker, R. Kelly, A. Chang, J. Dong, and H. Powell (2005), Quantifying the uncertainty in passive microwave snow water equivalent observations, *Remote Sens. Env.*, *94*, 187–203.
- Fret, A., M. Tedesco, S. Lee, J. Foster, D. K. Hall, R. Kelly, and D. A. Robinson (2012), A review of global satellite-derived snow products, *Adv. Space Res.*, *50*, 1007–1029.
- Gaspari, G., and S. E. Cohn (1999), Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.*, *125*, 723–757.
- Gent, P. R., et al. (2011), The community climate system model version 4, *J. Clim.*, *24*, 4973–4991.
- Goodison, B. E., T. R. Karl, and R. W. Knight (1999), Cryosphere systems, in *EOS Science Plan*, edited by M. D. King, pp. 261–306, NASA Goddard Space Flight Center, Greenbelt, Md.

- Hall, D. K., and G. A. Riggs (2007), Accuracy assessment of the MODIS snow products, *Earth Sci.*, *1547*, 1534–1547.
- Hall, D. K., G. A. Riggs, V. V. Salomonson, N. E. DiGirolamo, and K. J. Bayr (2002), MODIS snow-cover products, *Remote Sens. Environ.*, *83*, 181–194.
- Hall, D. K., G. A. Riggs, J. L. Foster, and S. V. Kumar (2010), Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product, *Remote Sens. Environ.*, *114*(3), 496–503.
- Hamill, M. T., S. J. Whitaker, and C. Snyder (2001), Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Mon. Weather Rev.*, *129*, 2776–2790.
- Kato, H., M. Rodell, F. Beyrich, H. Cleugh, E. van Gorsel, H. Liu, and T. P. Meyers (2007), Sensitivity of land surface simulations to model physics, land characteristics, and forcings, at four CEOP sites, *J. Meteorol. Soc. Jpn.*, *85A*, 187–204.
- Lawrence, D., et al. (2011), Parameterization improvements and functional and structural advances in version 4 of the community land model, *J. Adv. Model. Earth Syst.*, *3*, M03001, doi:10.1029/2011MS000045.
- Liston, G. E. (2004), Representing subgrid snow cover heterogeneities in regional and global models, *J. Clim.*, *17*, 1381–1397.
- Luce, H. C., and D. G. Tarboton (2004), The application of depletion curves for parameterization of subgrid variability of snow, *Hydrol. Process.*, *18*, 1409–1422.
- Niu, G.-Y., and Z.-L. Yang (2007), An observation-based formulation of snow cover fraction and its evaluation over large North American river basins, *J. Geophys. Res.*, *112*, D21101, doi:10.1029/2007JD008674.
- Oleson, K. W., et al. (2010), Technical description of version 4.0 of the community land model (CLM), *NCAR Tech. Note NCAR/TN-478+STR*, doi:10.5065/D6FB50WZ.
- Parajka, J., and G. Blöschl (2008), Spatio-temporal combination of MODIS images – Potential for snow cover mapping, *Water Resour. Res.*, *44*, W03406, doi:10.1029/2007WR006204.
- Peings, Y., H. Douville, R. Alkama, and B. Decharme (2010), Snow contribution to springtime atmospheric predictability over the second half of the twentieth century, *Clim. Dyn.*, *37*(5–6), 985–1004.
- Raeder, K., J. L. Anderson, N. Collins, T. J. Hoar, J. E. Kay, P. H. Lauritzen, and R. Pincus (2012), DART/CAM: An ensemble data assimilation system for CESM atmospheric models, *J. Clim.*, *25*, 6304–6317.
- Reichle, H. R., and D. R. Koster (2003), Assessing the impact of horizontal error correlations in background fields on soil moisture estimation, *J. Hydrometeorol.*, *4*, 1229–1243.
- Reichle, H. R., and D. R. Koster (2011), Assessment and enhancement of MERRA land surface hydrology estimates, *J. Clim.*, *24*, 6322–6338.
- Reichle, H. R., D. B. Mclaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*, 103–114.
- Rodell, M., and P. R. Houser (2004), Updating a land surface model with MODIS-derived snow cover, *J. Hydrometeorol.*, *5*, 1064–1075.
- Salomonson, V. V., and I. Appel (2004), Estimating fractional snow cover from MODIS using the normalized difference snow index, *Remote Sens. Environ.*, *89*(3), 351–360.
- Savoie, M. H., J. Wang, M. J. Brodzik, and R. L. Armstrong (2007), Improved snow cover retrievals from satellite passive microwave data over the Tibet Plateau: The need for atmospheric corrections over high elevations (Poster), *Eos Trans. AGU*, *88*(52), Fall Meet. Suppl., Abstract C23A-0942.
- Slater, G. A., and P. M. Clark (2006), Snow data assimilation via an ensemble Kalman filter, *J. Hydrometeorol.*, *7*, 478–493.
- Su, H., Z.-L. Yang, G.-Y. Niu, and R. E. Dickinson (2008), Enhancing the estimation of continental-scale snow water equivalent by assimilating MODIS snow cover with the ensemble Kalman filter, *J. Geophys. Res.*, *113*, D08120, doi:10.1029/2007JD009232.
- Su, H., Z.-L. Yang, R. E. Dickinson, C. R. Wilson, and G.-Y. Niu (2010), Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information, *J. Geophys. Res.*, *115*, D10104, doi:10.1029/2009JD013035.
- Turner, A. G., and J. M. Slingo (2011), Using idealized snow forcing to test teleconnections with the Indian summer monsoon in the Hadley Centre GCM, *Clim. Dyn.*, *36*, 1717–1735.
- Vernek, A. D., J. Zhou, and J. Shukla (1995), The effect of Eurasian snow cover on the Indian monsoon, *J. Clim.*, *8*, 248–266.
- Zaitchik, B. F., and M. Rodell (2009), Forward-looking assimilation of MODIS-derived snow covered area into a land surface model, *J. Hydrometeorol.*, *10*(1), 130–148.