We would like to thank the Editor and the two reviewers for their thoughtful comments and suggestions. Based on the feedback, we have made significant changes to the manuscript. We believe that the article is now much improved and again appreciate the help from the reviewers. See below for our detailed responses to all comments.

Note that the reviewer's original comments are in regular black fonts and our responses are in *red italic fonts*.

Editor

I have the feeling that you have addressed most of the aspects mentioned in the reviewer comments, especially of the excellent review of Dr. Newman. Nonetheless, there is one open topic in view of the suggestions of Reviewer one. He stated that he has problems to see what the different forcings are. Can you address this issue as well? If this is done, we can go a step forward.

We have updated the text to include these details for both the synthetic and "real" DA experiments. In the synthetic experiment, the control run is conducted using NLDAS-2, open loop with AGRMET and the forcing ensemble includes AGRMET, GDAS, ECMWF and MERRA-2. For real DA experiments, we use AGRMET as the open loop and AGRMET, GDAS, MERRA-2 and NLDAS-2 as the forcing ensemble. The description related to the synthetic experiment has been updated as:

"The forcing products used in EXP-FENS include the Global Data Assimilation System (GDAS; Derber et al. (1991)) operational outputs from NOAA/NCEP, the Modern Era Retrospective analysis for Research and Applications, version 2(MERRA-2; Bosilovich et al. (2017)) data, the European Center for Medium Weather Forecasting (ECMWF; Molteni et al. (1996)) and AGRMET datasets."

Reviewer #1

This paper presents a study on uncertainties and errors in terrestrial snow assimilation, and is thus within the scope of HESS. It is a well-developed concise article using clear language, and as such, it is a fine addition to the scientific knowledge. However, the experimental set-up lacks some details and references. For example, no information is given about FORCING1, FORCING2 and the 4 other forcing datasets before section 4.

Lines 57-59: This sentence is too speculative; there are many steps to reach that conclusion. Since it is part of the motivation for this article, please expand on the explanation and add references.

The sentence has been modified as follows with the inclusion of the appropriate reference.

"The accuracy of the model error covariance therefore, greatly depends on the accuracy of the forcing input (Reichle and Koster (2003)). "

Equation (2): Please say what the exponent "T" refers to.

Text has been added to say 'exponent T refers to the transpose of a matrix'

Line 124: Add a reference corresponding to the NOAH LSM v3.3.

The reference for Noahv3.3 (Ek et al., JGR, 2003) has been included.

Lines 150 and 160: Do you mean synthetic observations?

Yes. The qualification has been included in these lines.

Lines 152-153: Does it mean the OL was an ensemble run? If so, please justify/clarify.

As noted in the article, OL is conducted as an ensemble run that includes the perturbations. This approach is used to exclude any changes in skill introduced by the perturbation scheme in the evaluation of DA results. The text has been modified as:

"Note that the OL_FSNGL configuration includes the ensemble perturbations to the forcing and model state fields, to exclude any changes in model skill introduced by the perturbations in the evaluation of the DA results"

Lines 179-180: No prior information is given on the forcing datasets.

The section has been updated to include the information about all the forcing datasets. The control run is conducted using NLDAS-2, open loop with AGRMET and the forcing ensemble includes AGRMET, GDAS, ECMWF and MERRA-2. The description in the manuscript says "The forcing products used in EXP-FENS include the Global Data Assimilation System (GDAS; Derber et al. (1991)) operational outputs from NOAA/NCEP, the Modern Era Retrospective analysis for Research and Applications, version 2(MERRA-2; Bosilovich et al. (2017)) data, the European Center for Medium Weather Forecasting (ECMWF; Molteni et al. (1996)) and AGRMET datasets."

Line 213: Please add reference or website.

The references to the AMSR2 product (Oki et al. 2010, Kachi et al. 2013) are given earlier in the text. We have added the reference to the website (http://suzaku.eorc.jaxa.jp/GCOM_W/data/data_w_index.html) within the text.

Table 1 and line 258: Is it cumulative in time? Please clarify.

Yes, the table values are cumulative in time.

Reviewer #2

Title: Role of forcing uncertainty and model error background characterization in snow data assimilation

Authors: S. V. Kumar, J. Dong, C. D. Peters-Lidard, D. Mocko, and B. Gomez

General comments: This study examines the impact of forcing uncertainty/errors on model simulations and the subsequent model error covariances and analysis increments in ensemble snow water equivalent (SWE) data assimilation in an idealized and real data case. They find that accounting for input forcing uncertainty improves both simulations. This is because without forcing uncertainty, the imposed model state perturbations are not large enough to create a realistic background error covariance matrix, and thus the model states receive too much weight relative to the observations, or result in a Kalman gain matrix of zero, so that the analysis increments are essentially zero at some update times.

Overall, the study is easy to read, follow, and the figures and analysis support the conclusions. I think that acknowledging input forcing uncertainty in land-surface and hydrological model data assimilation (DA) in a more realistic way is a key step to developing useful and robust automatic DA systems. I recommend acceptance after the authors address my comments.

Major comments:

1) It would be nice to see some type of plot of analysis increments for the various experiments in Figures 2 and 3. It is clear that changing the input forcing to Noah increases the magnitude of the background error as shown in Figure 3b, thus increasing the analysis increments so that those experiments better match the observations. Analysis increments are another useful way to diagnose what is happening in the system at each DA time, and would be a useful complement to Figure 3b, especially since analysis increments are not shown, yet discussed in many places.

This could be particularly informative for the spatial runs in Figures 5 and 7, where the model performance has some spatial variability.

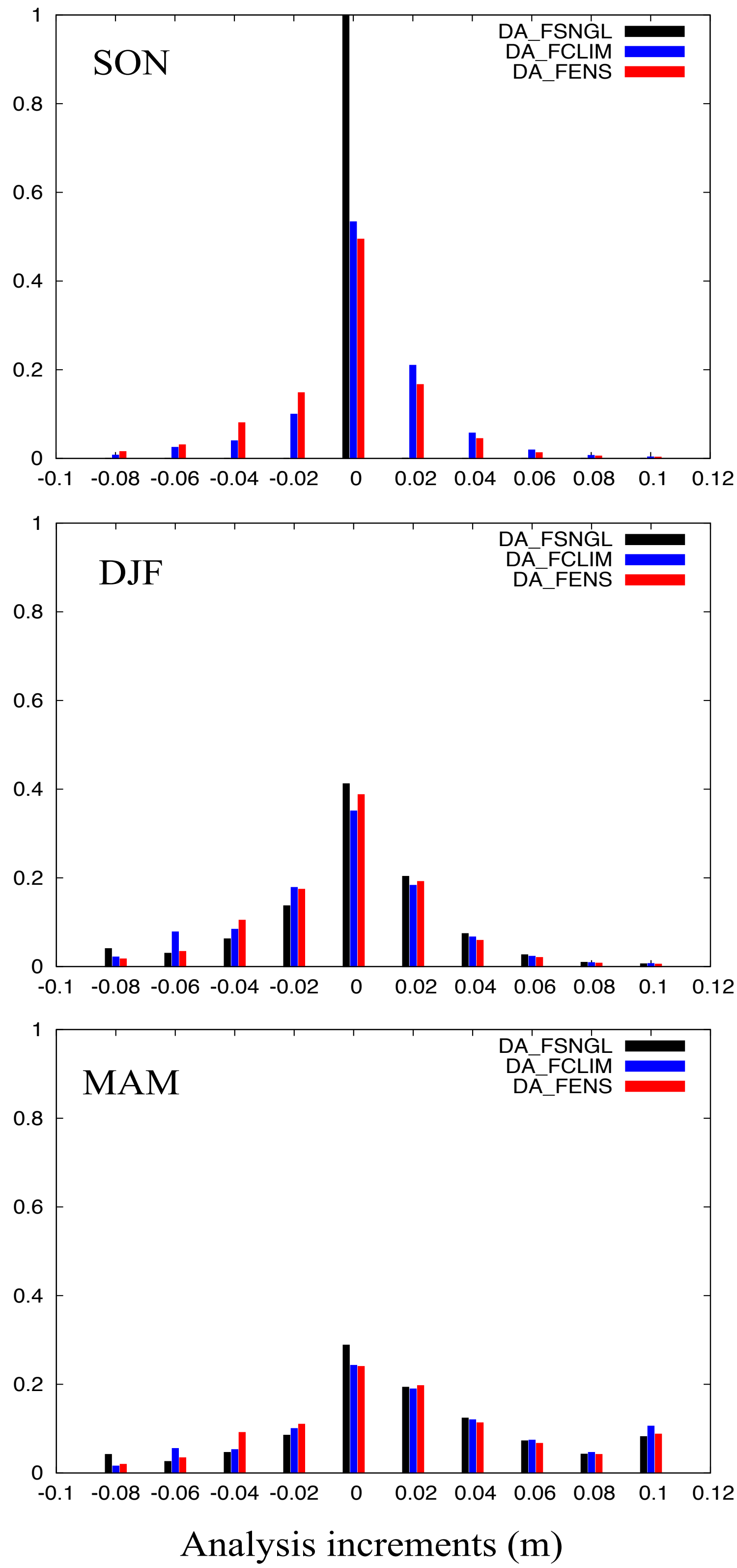
It may also be interesting to examine the spatial changes in the background error at key points during the accumulation and melt season.

Thanks for the excellent suggestion. We have updated Figure 3 to include time series comparisons of the analysis increments from the DA integrations. An additional paragraph describing these plots have been included in the text in Section 3, which reads as follows:

"Comparisons of the analysis increments from DA integrations shown in panels (c) indicate the time periods where the impact of the background model error is more

significant. Generally, the analysis increments from DA_FSNGL and DA_FCLIM are similar, except during the snow accumulation and melt time periods. Comparatively, larger differences in the analysis increments between the DA_FSNGL and DA_FENS integrations are observed, with more prominent differences seen during the accumulation and melt periods. During these times, larger analysis increments are observed in the DA_FCLIM and DA_FENS integrations, reflective of the ability of these configurations to respond to observations due to the improved background model error. It can also be noted that the analysis increments during the peak snow season are generally smaller in DA_FENS and DA_FCLIM integrations compared to that of DA_FSNGL, indicating the contribution of the hybrid forcing inputs for reducing the significant biases in the assimilation system."

We also examined patterns of analysis increments in the DA integrations employing AMSR2 retrievals. Generally, the analysis increments convolve the impact of multiple factors. The analysis increments include the ability/inability of the assimilation system to respond to observations and the contribution of the hybrid forcing ensemble to correcting the biases before observations are assimilated. The Figure below show the distribution of the analysis increments for the accumulation (SON), peak winter (DJF) and melt (MAM) time periods over the Great Lakes region. During the accumulation time period, the FSNGL simulation shows little variability in its distribution (inability to respond to obs), whereas during the other two time periods, the results are more mixed (though DA_FENS generally show greater span over larger analysis increments), likely due to the combined impact of different factors. Therefore, we decided not to include the comparison of analysis increments from the AMSR2 assimilation examples.



2) Are the observations aggregated up to the model resolution for Figures 5 and 7? I believe this is a key point that needs to be clarified. The authors should describe the

aggregation method, or redo the analysis if direct comparisons to the observation points were made.

In these comparisons, observations were aggregated up to the model resolution through simple averaging. We have added the following sentence in Section 4 to clarify this point:

"The available station observations are aggregated up to the model resolution through simple averaging in these comparisons."

Minor comments:

1) Model error background seems to be a non-standard phrasing of the background forecast error covariance matrix (e.g. Hamil et al. 2001, Descombes et al. 2015). I suggest re-phrasing it background model error, or background error.

Thanks the suggestion. We have updated all such references to 'background model error', including the title.

2) It is interesting to me that the article operates with snow depth rather than snow water equivalent (SWE). Could the authors expand on this choice at all? Noah seems to have SWE as a state variable and AMSR2 does have a SWE retrieval as well, so it would be possible to operate using SWE as well, which seems like a more natural state variable to work with.

Thanks for raising this point. There are a couple of reasons for using snow depth as the retrieval variable instead of SWE. In most passive microwave retrieval algorithms (Chang et al. 1987, Kelly et al. 2003, Kelly 2009) compute snow depth first and then derive SWE by using a climatological snow density. The basic retrieval product, in other words, is snow depth. In addition, since most in-situ observations of snow are also available as a depth measurement, the use of snow depth enables a more straightforward evaluation. We have modified the text in Section 3 (first paragraph) as follows:

"We employ snow depth as the measurement variable as most passive microwave retrieval algorithms (Chang et al. (1987); Kelly et al. (2003); Kelly (2009)) compute snow depth first and derive the snow water equivalent (SWE) through a climatological snow density (Brown and Braaten (1998); Krenke (1998, updated 2004)) assumption. In addition, most in-situ observations of snow are also available as depth measurements, allowing a more straightforward evaluation of the results from the model and DA integrations."

3) Line 216: Why is the AMSR2 standard error assumed to be 50 mm when Kachi et al. (2014) cite the standard error as 20 cm (200 mm)? Is the Kachi et al. (2013) citation in the manuscript giving a different standard error than the update?

The reviewer correctly points out that in Kachi et al. (2013), the AMSR2 retrievals satisfy the 20 cm error expectation based on their evaluation against GSOD measurements. The results in that paper also indicate that there is considerable spread in the evaluation of AMSR2 retrievals. In addition, the evaluation was limited to a single water year (2012-2013). In the paper, we use a higher value of standard error, based on the snow DA literature, which generally indicate low skill for passive microwave snow depth retrievals. The higher error standard deviation assumed here is consistent with prior snow DA studies (Liu et al. 2013, Liu et al. 2015, Kumar et al. 2014). We have added the following acknowledgement within the article:

"Note that we use a higher value of observation error standard deviation than that reported by Kachi et al. (2013), based on the previous snow DA studies (Liu et al. (2013, 2015); Kumar et al. (2014, 2015)) that generally assume low skill for passive microwave snow depth retrievals."

4) Line 218: Was the model resolution of 25km selected to match the approximately 30-km footprint of AMSR2? If so, it would be good to state that.

We use a model resolution of 25km, as two key near-real time global modeling environments, the NASA Global Land Data Assimilation System (GLDAS) and the U.S. Air Force 557th Weather Wing operational land data assimilation system) that use LIS are conducted at approximately 25 km resolution.

5) Line 280, change stronger to larger. The authors may want to check the entire paper for instances of this.

Thanks for the suggestion. All such instances have been corrected.

6) Lines 306-308. The two sentences starting with "Though underestimated" and ending with "data assimilation updates" are confusing to me. What are authors trying to describe here?

We agree that these sentences are confusing. The entire paragraph has been rewritten a follows:

"In general, the DA integrations (DA_FSNGL, DA_FCLIM and DA_FENS), have comparable performance at both these locations and they mostly follow the snow evolution patterns in the AMSR2 data. Note that though AMSR2 observations capture the seasonality of snow observations, they show significant underestimation compared to in-situ observations of snow depth. The influence of undersampling the model error background can be observed in the early part of the snow season at location C and during late season at location D, where the DA_FSNGL integrations fail to match the snow events captured by AMSR2. During the peak snow time periods, however, the undersampling of model error background in OL_FSNGL is less of a problem over this domain, as the non-zero model snow states provide an adequate background for

subsequent data assimilation updates. Thus, the evaluation of the snow DA integrations at these two regions …"

7) Figures 5 and 7: The authors may want to consider having a gray color that spans zero as small error differences are likely not significant. The figure is nice as it is with the lighter shades near zero; this is merely a suggestion to look into.

Thanks for the suggestion. We have redone Figures 5 and 7 to have an improved color scheme with a gray color spanning small error magnitudes.

8) The authors may be interested in the article Huang et al. (2016) that is in press in HESS. This article uses an ensemble of forcing data to generate and ensemble of internally consistent (with the forcing traces) initial model states uncertainty for EnKF SWE assimilation. They examine the impact of the relative weighting of the model and observational error covariance matrices. They also find similar results to those stated on lines 365-368 as well, if the open loop simulation has high quality forcing, DA is less beneficial. I am not suggesting the authors need to cite this paper, as I am a co-author on it; it just seems to be very relevant to the study reviewed here and some of their discussion points.

Thanks for the suggestion about this appropriate reference. We have modified the text in Section 5 as follows:

"If the single forcing dataset being used is of high skill, then the added benefit of using the forcing ensemble is likely to be less, consistent with the results of more recent studies to employ an ensemble of forcing data for generating an ensemble of internally consistent model uncertainty representation for applications such as DA (Newman et al. (2015); Huang et al. (2017)). Overall, the results in this article indicate …."

Manuscript prepared for Hydrol. Earth Syst. Sci.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 7 March 2017

Role of forcing uncertainty and background model error characterization in snow data assimilation

Sujay V. Kumar[1], Jiarui Dong[2], Christa D. Peters-Lidard[3], David Mocko[4,1], and Breogán Gómez[5]

[1]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[2]I.M. Systems Group Inc., Environmental Modeling Center, NOAA NCEP, College Park, MD, USA
[3]Hydrosphere, Biosphere and Geophysics, Earth Science Division, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[4]Science Applications International Corporation, McLean, VA, USA
[5]Data assimilation and ensembles, UK Met Office, Exeter, UK

Correspondence to: Sujay V. Kumar
(Sujay.V.Kumar@nasa.gov)

Abstract. Accurate specification of the model error covariances in data assimilation systems is a challenging issue. Ensemble land data assimilation methods rely on stochastic perturbations of input forcing and model prognostic fields for developing representations of input model error covariances. This article examines the limitations of using a single forcing dataset for specifying forcing uncertainty inputs for assimilating snow depth retrievals. Using an idealized data assimilation experiment, the article demonstrates that the use of hybrid forcing input strategies (either through the use of an ensemble of forcing products or through the added use of the forcing climatology) provide a better characterization of the background model error, which leads to improved data assimilation results, especially during the snow accumulation and melt time periods. The use of hybrid forcing ensembles is then employed for assimilating snow depth retrievals from the AMSR2 instrument over two domains in the Continental U.S. with different snow evolution characteristics. Over a region near the Great Lakes where the snow evolution tends to be ephemeral, the use of hybrid forcing ensembles provides significant improvements relative to the use of a single forcing dataset. Over the Colorado Headwaters characterized by large snow accumulation, the impact of using the forcing ensemble is less prominent and is largely limited to the snow transition time periods. The results of the article demonstrate that improving the background model error through the use of a forcing ensemble enables the assimilation system to better incorporate the observational information.

1 Introduction

Land Data Assimilation (DA) methods combine observations of land surface conditions from remote sensing platforms or ground measurements with model forecasts to produce temporally and spatially continuous estimates of land surface fields. The merging of the observations and model forecasts is conducted by weighting them appropriately based on their respective sources of errors. As a result, the skill of the DA systems is critically reliant on the accurate specification of errors in observations and model background.

Despite their importance, the specification of input error covariances is challenging (Dee (1995); Derber and Bouttier (1999); Reichle (2008); Reichle et al. (2008)). The sources of errors in observations include instrument errors, deficiencies of the observation operators (such as radiative transfer models) and representativeness issues from differences in spatial scales (Kumar et al. (2012)). Similarly, uncertainties in model parameters, forcing inputs and deficiencies in model physics contribute to the model background errors. The model error covariance specifications are often made through idealized experiments using analysis of assimilation increments and innovations (Kumar et al. (2008, 2009)). Comparison of model simulations against independent observations is another approach for developing these specifications. However, given the lack of representativeness of the point-scale in situ measurements and the significant heterogeneity of the land surface, developing spatially distributed estimates of these model error covariances are difficult. As noted in Reichle (2008), the specification of input error covariances remains a subjective process in current land data assimilation systems.

Ensemble data assimilation techniques such as the Ensemble Kalman Filter (EnKF) are widely used in land data assimilation applications (Crow and Wood (2003); Reichle et al. (2007); Kumar et al. (2009); Reichle et al. (2010); De Lannoy et al. (2012); Kumar et al. (2014)). The EnKF, a Monte-Carlo variant of the Kalman filter, uses an ensemble of model trajectories to represent the model error structures. The model error covariance is diagnosed as the sample covariance of the ensemble of model forecasts. The ensemble is typically created by adding stochastic noise to the meteorological forcing, propagated to the model fields through the non-linear land surface model (LSM). In addition, stochastic perturbations are also commonly applied to the model prognostic fields.

Perturbations are sampled from randomly generated noise and are directly applied to the forcing and model prognostic fields. The typical approach is to employ either normally distributed additive perturbations or lognormally distributed multiplicative perturbations, depending on the variable. For example, multiplicative perturbations are normally used for fields such as precipitation, since the use of additive noise could generate unphysical values (less than zero) or consistent positive biases during periods where precipitation is absent. In addition, to avoid introducing systematic biases in the perturbed fields, the ensemble-mean of the perturbations are normally constrained to zero and one, for additive and multiplicative perturbations, respectively.

In this article, we examine how the reliance on ensemble perturbations of forcing fields to develop the background model error impacts the performance of data assimilation. Most land data assimilation systems use a single data source as the forcing input and the input forcing uncertainty is characterized by perturbing the meteorological fields from this single data source. The accuracy of the model error covariance therefore, greatly depends on the accuracy of the forcing input (Reichle and Koster (2003)). For example, in a case where precipitation inputs are underestimated, the forcing uncertainty characterized by the resulting ensemble will lead to the underestimation of the model error covariance. In contrast, alternate strategies such as the added use of the forcing climatology or multiple forcing data sources are likely to provide better representations of the forcing uncertainty and a better characterization of the background model error. In this article, we examine the impact of such factors in the context of snow data assimilation case studies.

The article presents two sets of experiments: 1) An idealized experiment to demonstrate the impact of model error covariance underestimation and 2) A "real" data assimilation scenario where snow depth retrievals (Oki et al. (2010); Kachi et al. (2013)), from the Advanced Microwave Scanning Radiometer 2 (AMSR2) aboard the Global Change Observation Mission-Water (GCOM-W) satellite are used. The assimilation of AMSR2 data is conducted over two different domains in the continental U.S. with different snow evolution characteristics. The different nature of the snow evolution in these domains is used to investigate the impact of background model error representations in snow data assimilation. All experiments described in this article are conducted using the NASA Land Information System (LIS; Kumar et al. (2006)) which is an observation-driven land surface modeling and data assimilation system. The data assimilation subsystem in LIS (Kumar et al. (2008)) contains algorithms such as the EnKF and supports the assimilation of data from a variety of satellite sensors (Reichle et al. (2010); Liu et al. (2013); Kumar et al. (2014, 2015); Liu et al. (2015); Kumar et al. (2016)).

2 Ensemble Kalman Filter and background error covariance representation

The filtering class of data assimilation algorithms seek the best estimate of the posterior state conditioned on the past observations, using the statistics of the uncertainties in the model and observations. The Kalman Filter (KF) is an optimal estimator for linear dynamical systems driven by Gaussian noise. The EnKF is a reduced-rank variant of the KF, which assumes normality of model and observation errors and typically requires the use of a small number of ensembles to represent these error structures (Reichle (2008)).

EnKF is a sequential data assimilation approach, where the algorithm alternates between a forecast step and an analysis step. In the forecast step, an ensemble of model states is propagated forward in time using the LSM. This is followed by an analysis step where the model forecast is updated based on observations. The analysis step is written in the general form as:

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{K_k}[\mathbf{y}_k - \mathbf{H_k}\mathbf{x}_k^b] \tag{1}$$

where $\mathbf{x}^b$ is the background model state vector, $\mathbf{x}^a$ is the analyzed state vector, $\mathbf{y}$ is the observation vector and $\mathbf{H}_k$ is the observation operator that relates the model states to the observations. The subscript k indicates time and the superscripts b and a refer to the state estimates, before and after the update, respectively. $\mathbf{K_k}$ is the gain matrix, which represents the weighting factor that determines the degree to which the model forecast is adjusted towards the observation. $\mathbf{K_k}$ is expressed as:

$$\mathbf{K}_k = \mathbf{P}_k^b\mathbf{H}_k^T\left[\mathbf{H}_k\mathbf{P}_k^b\mathbf{H}_k^T + \mathbf{R}_k\right]^{-1} \tag{2}$$

where $\mathbf{R}_k$ and $\mathbf{P}_k^b$ are the observation and forecast model error covariances, respectively (exponent T refers to the transpose of a matrix). The model error covariance is computed as the sample covariance of the model ensemble.

EnKF relies on the second order statistics of the noise simulated by ensemble perturbations in the model and observations (drawn from Gaussian distributions), to characterize their probability density functions (PDFs). The accuracy of the sampled model error covariance, in particular, is dependent on the size of the ensemble and the presence of model errors (Li et al. (2009)). Prior studies have used techniques such as covariance inflation (Anderson and Anderson (1999)), to deal with the covariance underestimation. These techniques, however, require significant tuning and rely on the assumption that the observation error covariances are known (Miyoshi and Yamane (2007)). In addition, these inflation techniques are ineffective when the model errors are significant and the resulting model error covariances are close to zero. In the examples below, the impact of underestimating the background model error for snow data assimilation is examined.

Figure 1 shows a schematic of three strategies that are used to examine the issue of model covariance underestimation in this article. The first strategy (A), which is the typical practice in land data assimilation systems, is to use a single forcing dataset to drive the ensemble. The small perturbations applied to the input forcing variables help in simulating the ensemble spread. In the second strategy (B), the ensemble is forced with both the given forcing and a climatology of that forcing. The added use of the forcing climatology helps in incorporating the representation of average conditions within the ensemble and in reducing the covariance underestimation due to the reliance and limitations of a single dataset. In the third approach (C), the model ensemble is driven using an ensemble of forcing products from different sources, providing a more realistic representation of the input forcing uncertainty. Note that small perturbations to the forcing variables are also applied to B and C forcing data to augment the ensemble spread.

3 Assessing the impact of model error covariance underestimation through idealized experiments

In this section, we present an idealized snow depth DA experiment to demonstrate the importance of accurately characterizing the input model error covariances. We employ snow depth as the measurement variable as most passive microwave retrieval algorithms (Chang et al. (1987); Kelly et al. (2003); Kelly (2009)) compute snow depth first and derive the snow water equivalent (SWE) through a climatological snow density (Brown and Braaten (1998); Krenke (1998, updated 2004)) assumption. In addition, most in-situ observations of snow are also available as depth measurements, allowing a more straightforward evaluation of the results from the model and DA integrations. The synthetic experiment is conducted at the Niwot Ridge site in Colorado (40.03°N, 105.5°W), which is part of the NRCS Snow Telemetry (SNOTEL) network. All model simulations are conducted using the Noah land surface model version 3.3 (Ek et al. (2003)). The DA experiment is set up as an identical twin experiment (Kumar et al. (2009)) with the following structure: First, the Noah LSM is run forced with meteorology from the North American Land Data Assimilation System Phase 2 (NLDAS-2; Xia et al. (2012)), and is assumed to represent the "true" state of snow depth evolution at this location. This model integration is termed as the Control or "truth" simulation. Next, a set of synthetic snow depth observations is simulated from this Control run by introducing realistic retrieval errors. Similar to the strategies used in previous studies (Kumar et al. (2008)), to account for the limitations of the passive microwave sensors in retrieving snow depth under dense canopies, the observations are masked out when Green Vegetation Fraction (GVF) values used in the model are greater than 0.6. In addition, observations are degraded by introducing multiplicative random noise with standard deviation of 0.05 to simulate the errors in the snow depth retrievals. An open loop (OL) integration is conducted using the same LSM, but forced with a different meteorology from the Agricultural Meteorology model (AGRMET) of the U.S. Air Force 557th Weather Wing (formerly the Air Force Weather Agency). A data assimilation integration is then conducted by incorporating the simulated observations into the OL configuration using a one-dimensional Ensemble Kalman Filter (EnKF; Reichle et al. (2002)). The modeled estimates from the OL and DA integrations are compared against the true fields from the Control run to evaluate the impact of assimilation.

An ensemble size of 20 is used in the integrations with perturbations applied to both meteorological forcing inputs and model prognostic fields to simulate the background model error. Multiplicative perturbations are applied to the precipitation and downwards shortwave fields with a mean of 1 and standard deviations of 0.3 and 0.5, respectively. Additive perturbations with a standard deviation of 50 W/m^2 are applied to the longwave radiation fields. The Noah LSM model fields of SWE and snow depth are perturbed with multiplicative noise of 0.01 and 0.02, respectively. Time series correlations are imposed via a first-order regressive model (AR(1)) with a time scale of 24 hours for forcing variables and 12 hours for the model fields. The perturbations to the forcing fields are applied hourly, whereas the model prognostic fields are perturbed at three hour intervals, similar to the configurations used in Kumar et al. (2015) and Kumar et al. (2014).

Figure 2 shows a time series of the snow depth fields from model integrations for the 2012-2013 winter season, from the Control, synthetic observations (OBS), open loop simulation forced with a single meteorological dataset (OL_FSNGL), and the data assimilation integration that assimilates observations into the OL_FSNGL configuration (DA_FSNGL). Note that the OL_FSNGL configuration includes the ensemble perturbations to the forcing and model state fields, to exclude any changes in model skill introduced by the perturbations in the evaluation of the DA results. The control and OL_FSNGL runs are significantly different in their simulation of snow depth for this winter season. The OL_FSNGL based snow depth estimates vastly underestimate the snow evolution, likely due to the underestimation of precipitation in the AGRMET data at this location. The assimilation of the observations helps in significantly improving the OL_FSNGL representation, especially during the peak winter months of January through March. The simulation of snow depth during the snow accumulation time periods and the snow melt time periods, however, shows significant differences relative to the Control simulation, though synthetic observations of snow depth exist during these time periods. As shown in Figure 2, the snow accumulation in the OL_FSNGL simulation is significantly delayed relative to the Control. The input model error covariances ($\mathbf{P_k^b}$), therefore, remain close to zero until mid-December 2012, when non-zero snow depth estimates are observed in the OL_FSNGL configuration. These model errors result in the gain matrix ($\mathbf{K_k}$) being zero when the background model error variances are zero. As a result, no non-zero analysis increments are generated from the DA analysis and no changes in the snow depth fields from DA are observed until mid-December, 2012. In contrast, during the peak winter months, the snow depth estimates from DA_FSNGL are closer to the Control simulation, as the availability of a non-zero model error covariance allows DA to compute positive analysis increments. Further, the DA_FSNGL integration also fails to capture the late season snow events (late April and early May 2013), as the deficiencies in the background model error result in the inability of the analysis step to produce meaningful analysis increments.

In the above example, the main source of the model deficiencies is the errors in the forcing inputs, as the same model is used in the Control and open loop integrations. Two variants of this experiment are conducted by: 1) using the forcing climatology in combination with the input forcing to specify the ensemble (EXP-FCLIM) and 2) using an ensemble of forcing datasets to drive the ensemble (EXP-FENS). A climatological forcing dataset is developed by averaging the forcing inputs (used in OL_FSNGL) at each forcing timestep across 4 years (2012 to 2015). In EXP-FCLIM, the forcing climatology is used to drive 10 of the 20 ensemble members with the remaining 10 driven by the OL_FSNGL forcing data. In EXP-FENS, four different forcing datasets (different from the data used in the Control) is used to drive the model ensemble. The forcing products used in EXP-FENS include the Global Data Assimilation System (GDAS; Derber et al. (1991)) operational outputs from NOAA/NCEP, the Modern Era Retrospective analysis for Research and Applications, version 2 (MERRA-2; Bosilovich et al. (2017)) data, the European Center for Medium Weather Forecasting (ECMWF; Molteni et al. (1996)) and AGRMET datasets. Each forcing data is used to drive 5 ensemble members within the 20 member ensemble. As before, perturbations are applied to both forcing and model states. These strategies assume that a better representation of the forcing uncertainty and model error covariance can be developed by augmenting the ensemble through the use of multiple data sources.

Panels (a) in Figure 3 show the time series of snow depth from open loop and DA integrations from the EXP-FCLIM and EXP-FENS experiments, panels (b) show comparisons of the snow depth ensemble spread from DA integrations and panels (c) show comparisons of the analysis increments in snow depth from DA. In the EXP-FCLIM experiment, it can be noted that the added use of forcing climatology with the OL_FSNGL forcing is helpful in increasing the ensemble spread in the DA integrations without a significant change to the mean snow depth estimates. Subsequently, the improved background model error representation leads to improved DA performance, as the DA_FCLIM based estimates are improved relative to the DA_FSNGL estimates. The improvements are more apparent during the snow accumulation (Dec-Jan) and melt (April-May) time periods, though they are significantly underestimated relative to the Control. Quantitatively, the RMSE in the OL_FSNGL and OL_FCLIM integrations for the Oct 2012 to Jun 2013 time period is 85 mm. The DA_FSNGL integration with a single forcing dataset has a RMSE of 55 mm and the added use of the forcing ensemble helps in further reducing the overall RMSE to 48 mm in the DA_FCLIM integration.

Comparatively, the use of an ensemble of forcing products provides significantly improved performance in the assimilation of synthetic observations. First, a significant portion of the bias in the snow depth estimates is reduced by the forcing ensemble based open loop (OL_FENS). The cumulative RMSE of the OL_FENS integration is 56 mm. The use of the forcing ensemble then helps in improving the DA simulations (DA_FENS), as it shows a closer match with the Control relative to all other DA integrations. In particular, DA_FENS shows improvements in the accumulation (Nov-Dec) and snow melt (Mar-Apr) periods, and provides a low RMSE of 29 mm, for the time period of Oct 2012 to June 2013.

Comparisons of the analysis increments from DA integrations shown in panels (c) indicate the time periods where the impact of the background model error is more significant. Generally, the analysis increments from DA_FSNGL and DA_FCLIM are similar, except during the snow accumulation and melt time periods. Comparatively, larger differences in the analysis increments between the DA_FSNGL and DA_FENS integrations are observed, with more prominent differences seen during the accumulation and melt periods. During these times, larger analysis increments are observed in the DA_FCLIM and DA_FENS integrations, reflective of the ability of these configurations to respond to observations due to the improved background model error. It can also be noted that the analysis increments during the peak snow season are generally smaller in DA_FENS and DA_FCLIM integrations compared to that of DA_FSNGL, indicating the contribution of the hybrid forcing inputs for reducing the significant biases in the assimilation system.

4 Impact of forcing ensemble in the assimilation of AMSR2 snow depth retrievals

The idealized experiments presented in the previous section demonstrate that the use of hybrid forcing ensemble strategies is helpful in providing a better characterization of the forcing uncertainty and the background model error. We extend this approach to a "real" data assimilation scenario where passive microwave snow depth observations from the AMSR2 instrument are employed. These retrievals, available from 2012 July onwards, are obtained from the Japan Aerospace Exploration Agency (JAXA; http://suzaku.eorc.jaxa.jp/GCOM_W/data/data_w_index.html). In all the integrations assimilating AMSR2 retrievals, the standard deviation of the observation error is assumed to be 50 mm. Note that we use a higher value of observation error standard deviation than that reported by Kachi et al. (2013), based on the previous snow DA studies (Liu et al. (2013, 2015); Kumar et al. (2014, 2015)) that generally assume low skill for passive microwave snow depth retrievals.

Land surface model simulations using the Noah LSM (version 3.3) are conducted over two regional model domains in the continental U.S. (Figure 4) at 25 km spatial resolution: (1) A region centered around the Great Lakes (GL) and (2) a domain centered around the Colorado Headwaters (CH). The snow evolution in the GL region tends to be ephemeral, wet and shallow whereas the CH region is a high-terrain domain with complex topography and large seasonal snowpacks. The impact of different background model error representations on the assimilation of AMSR2 data is examined over these two domains with contrasting snow development and melt characteristics.

Similar to the synthetic data assimilation experiment presented in Section 3, the model simulations are conducted with a single meteorological forcing dataset, a single meteorological forcing dataset and its climatology, and an ensemble of forcing datasets. The AGRMET data is used as the single meteorological forcing data. In the forcing ensemble based runs, in addition to AGRMET, three other forcing datasets are used, which include the GDAS, NLDAS-2 and MERRA-2 datasets. The LSM simulations are conducted during a time period of October 2012 to Dec 2015 with a time step of 30 min.

We focus first on the GL region by comparing the snow evolution from various model and data assimilation integrations. Figure 5 presents a "RMSE improvement" map (RMSE of DA with the single forcing (DA_FSNGL) minus the RMSE of DA with the hybrid forcing ensemble (DA_FCLIM or DA_FENS)) by comparing to the in-situ snow depth measurements at the Global Historical Climate Network (GHCN; Menne et al. (2012)) sites. The available station observations are aggregated up to the model resolution through simple averaging in these comparisons. The warm colors indicate locations where the DA_FCLIM or DA_FENS has a reduced RMSE compared to DA_FSNGL and the cool colors indicate locations where DA_FSNGL has an increased skill relative to DA_FCLIM or DA_FENS. As the figure indicates, the DA integrations employing hybrid forcing inputs are systematically better than the DA_FSNGL simulation in most parts of the domain. Comparatively, the RMSE improvements are larger in the DA_FENS integration than the DA_FCLIM simulation. Note that the improved skill of DA_FENS in particular, is benefited by both the improved model background and the skill of the precipitation data sources that constitute the forcing ensemble, though it is hard to separate their contributions. This is demonstrated by comparing the time series of model and DA simulations at two locations: Point A, at 45.875 N, 89.375 W and Point B, at 48.875 N, 97.625 W.

As shown in Figure 6 (A), the OL_FSNGL simulations significantly underestimate the snow evolution throughout the winter period of 2012-2013. The added use of the forcing climatology (OL_FCLIM) leads to overestimating the peak season snow (Feb-Mar) and marginally improves the late season snow. Similarly, the use of the forcing ensemble (OL_FENS) marginally improves the OL_FSNGL underestimation (especially during the early snow season), but fails to capture the late season snow events. The AMSR2 retrievals at this location are primarily available in the late snow season and help in improving the snow depth simulation through DA. Overall, the limitations of the OL_FSNGL prevents DA from making a significant impact in the DA_FSNGL simulation. The availability of the improved background in DA_FCLIM and DA_FENS enables them to provide a better match to the relatively large snow events in March and April, compared to other simulations. Table 1 shows a summary of the cumulative RMSE from various simulations at these locations. The cumulative RMSE from the OL_FSNGL is 381 mm, which reduces to 275 mm and 169 mm with the OL_FCLIM and OL_FENS, respectively. The cumulative RMSE in the DA integrations is 266 mm for DA_FSNGL, 262 mm in DA_FCLIM and 244 mm in DA_FENS. Note that the cumulative RMSE does not reflect the obvious improvement during the late season snow periods in DA_FENS (over OL_FENS), as the early season underestimation dominates these statistics.

Figure 6 (B) panel shows a similar time series comparison at point B with larger snow evolution. Similar to point A, OL_FSNGL underestimates the snow evolution throughout the season (RMSE of 252 mm) and is improved by the use of the hybrid forcing ensembles. During the snow accumulation time periods (up to early Feb 2013), the OL_FCLIM (RMSE of 201 mm) and OL_FENS (RMSE of 167 mm) estimates show better agreement with the GHCN measurements. The AMSR2 retrievals show significant underestimation relative to GHCN during the peak snow season, though they are helpful in improving the snow depth simulations in the late snow season (Mar-May). The impact of the improved model background can be noted in the DA_FCLIM and DA_FENS simulations in their ability to provide a better match with the GHCN observations in the late snow season. The single forcing based DA estimate (DA_FSNGL), on the other hand, does a poor job in this time period despite the availability of AMSR2 retrievals that are consistent with GHCN. The cumulative RMSE of the DA_FSNGL integration at this location is 206 mm and it improves to 156 mm and 162 mm in the DA_FCLIM and DA_FENS integrations.

A similar set of evaluations are conducted over the CH domain, an area with deeper seasonal snow accumulation compared to the GL region. Figure 7 presents the RMSE improvement map for the CH domain (similar to Figure 5). Compared to the improvements observed in the GL domain, the patterns of improvements and degradations are more mixed in the CH domain. In addition, larger improvements and degradations are observed in the DA_FCLIM and DA_FENS integrations relative to DA_FSNGL. To examine these patterns, the time series of snow evolution from various integrations is compared at two locations in the CH domain (Point C at 40.375, 106.875 and point D at 45.125, 109.875) and are shown in Figure 8. OL_FSNGL underestimates the snow evolution in both locations (RMSE of 424 mm and 276 mm at C and D, respectively as shown in Table 1). The added use of the climatology (OL_FCLIM) marginally improves the snow simulation at location C (RMSE of 402 mm) and provides more significant improvements at location D (RMSE of 142 mm). The use of the forcing ensemble (OL_FENS) provides a better match to the observations at location C (RMSE of 179 mm), but overestimates the snow accumulation at location D (RMSE of 215 mm). At location C, the assimilation of AMSR2 improves the snow depth estimates in DA_FSNGL (RMSE of 316 mm) and DA_FCLIM (RMSE of 309 mm) integrations relative to their respective OL, whereas DA leads to degradations in the forcing ensemble configuration (RMSE of 285 mm), compared to OL_FENS. At location D, the assimilation of AMSR2 retrievals leads to increased RMSE in the DA integrations (RMSE of 327, 312 and 309 mm for DA_FSNGL, DA_FCLIM and DA_FENS, respectively) These trends are reflective of the fact that the AMSR2 observations underestimate the snow evolution in the peak winter months (Jan-Mar) and overestimates snow estimates in the spring melt time periods (Apr - May), at location C. At location, D, however, the AMSR2 snow observations are generally underestimated. The underestimation of snow at both these locations, is likely due to the fact that passive microwave based retrievals saturate for thick snow packs (Dong et al. (2005)).

In general, the DA integrations (DA_FSNGL, DA_FCLIM and DA_FENS), have comparable performance at both these locations and they mostly follow the snow evolution patterns in the AMSR2 data. Note that though AMSR2 observations capture the seasonality of snow observations, they show significant underestimation compared to in-situ observations of snow depth. The influence of undersampling the background model error can be observed in the early part of the snow season at location C and during late season at location D, where the DA_FSNGL integrations fail to match the snow events captured by AMSR2. During the peak snow time periods, however, the undersampling of background model error in OL_FSNGL is less of a problem over this domain, as the non-zero model snow states provide an adequate background for subsequent data assimilation updates. Thus, the evaluation of the snow DA integrations at these two regions provide valuable insights on the importance of accurately characterizing the background model error. The use of the hybrid forcing ensemble and improved model background is more helpful over the GL domain, where snow evolution is ephemeral. Over regions with large snowpacks such as the CH region, the representation of the model background is more important during the early accumulation and spring melt time periods.

5 Summary

Accurate specification of input model and observations error covariances in data assimilation systems is challenging though these error specifications are critical in the development of a skillful data assimilation system. In offline ensemble land data assimilation systems, the model ensemble and background model error representation are typically generated by applying small perturbations to the model prognostic states and input meteorological forcing fields. Most Land DA studies are reliant on the use of a single forcing dataset to derive their driving meteorology.

In this article, the limitations of using a single forcing dataset as the basis for developing background model error is examined in the context of snow data assimilation. When significant errors are present in the forcing fields (e.g. precipitation), the resulting model and ensemble estimates will have significant errors. In such instances, the use of an ensemble of forcing datasets, either based on climatology or a suite of independent datasets, is likely to provide a better representation of the forcing uncertainty and the background model error. The article demonstrates these issues through both idealized and real data assimilation experiments.

The idealized experiment presents a case where the snow depth estimates are significantly underestimated due to the presence of precipitation biases. The application of stochastic perturbations using this biased precipitation input is inadequate in providing a realistic background model error in the assimilation system. As a result, the snow depth fields in the DA system remain biased, especially during the snow evolution and spring melt periods. In contrast, when an ensemble of forcing datasets is used to drive the model, the representation of the background model error is more realistic. As a result, the assimilation system performs better in incorporating the impact of observations during the snow evolution and ablation periods.

The impact of using a forcing ensemble for developing the background model error is examined for the assimilation of snow depth retrievals from the AMSR2 instrument, over two domains in the Continental U.S. with different snow evolution characteristics. Over the region near the Great Lakes, the snow evolution tends to be shallow, with transitions between snow and no-snow conditions during each snow season. In this region, the added use of the forcing climatology to drive the ensemble leads to improved DA performance, when compared to the in-situ ground observations of snow depth. The DA performance is further enhanced with the use of an ensemble of forcing inputs, partly aided by the enhanced skill of the precipitation inputs. Over the Colorado Headwaters, an area with large seasonal snow packs, the impact of precipitation biases on the simulation of snow states is largely limited to the snow evolution and ablation time periods. As the occurrences of transitions between snow and no-snow states are less common during the peak winter months in this region, the underestimation of the background model error is less problematic in the DA integrations during these time periods. As a result, the positive impact of the use of forcing ensemble is mostly prominent during the accumulation and ablation time periods.

As noted above, the evaluation of snow depth estimates over CH region shows mixed results, with several locations indicating worse performance with the use of the forcing ensemble compared to the use of a single forcing dataset. In regions with large snow accumulation (such as the CH region), passive microwave retrievals such as those from AMSR2 are known to have low skill due to issues such as saturation in deep snowpacks, signal loss in wet snow and overestimation in the presence of large snow grains (Dong et al. (2005); Foster et al. (2005); Durand et al. (2011)). Such limitations contribute to the mixed results seen in these results, especially in the CH domain. In such instances, the poorer performance from the use of the forcing ensemble is a result of the poor skill of the retrievals. To improve the skill of the retrievals themselves, prior studies (Kumar et al. (2014); Liu et al. (2015)) have successfully employed objective analysis techniques such as optimal interpolation to blend in situ measurements with satellite retrievals prior to assimilation. These prior studies and the results of this article suggest that a strategy that combines the use of hybrid forcing inputs (to improve background model error) and in situ data based correction of observations to be assimilated (to enhance the satellite retrievals) is likely to provide a robust configuration for optimal DA performance.

It must be stressed that in the experiments presented in the article, the OL_FSNGL configurations purposely employ an inferior forcing dataset so that the differences between the OL_FSNGL and OL_FCLIM and OL_FENS simulations are more magnified. If the single forcing dataset being used is of high skill, then the added benefit of using the forcing ensemble is likely to be less, consistent with the results of more recent studies to employ an ensemble of forcing data for generating an ensemble of internally consistent model uncertainty representation for applications such as DA (Newman et al. (2015); Huang et al. (2017)). Overall, the results in this article indicate that use of a forcing ensemble is helpful in providing better representations of background model error and more positive and consistent improvements in data assimilation. Note also that the use of an ensemble of forcing products may be practical in operational assimilation environments for centers with ensemble prediction systems. Where not available, the combined use of the forcing climatology along with the single, operational forcing input may be an appropriate strategy to improve the skill of the data assimilation system, as validated by the results in this paper.

Acknowledgements. Funding for this work was provided by the NOAA's Climate Program Office (MAPP program). Computing was supported by the resources at the NASA Center for Climate Simulation. The NLDAS-2 forcing data used in this effort were acquired as part of the activities of NASA's Science Mission Directorate, and are archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC). We are grateful to Dr. Andrew Newman and an anonymous reviewer for their valuable edits and suggestions to the manuscript.

References

Anderson, J. and Anderson, S.: A Monte Carlo implementation of the nonliner filtering problem to produce ensemble assimilations and forecasts, Mon. Wea. Rev., 127, 2741–2758, 1999.

Bosilovich, M. G., Robertson, F. R., Takacs, L., Molod, A., and Mocko, D.: Atmospheric Water Balance and Variability in the MERRA-2 Reanalysis, Journal of Climate, 30, 1177–1196, doi:10.1175/JCLI-D-16-0338.1, http://dx.doi.org/10.1175/JCLI-D-16-0338.1, 2017.

Brown, R. and Braaten, R.: Spatial and temporal variability of Canadian monthly snow depths, Atmosphere-Ocean, 36, 1998.

Chang, A., Foster, J., and Hall, D.: Nimbus-07 SMMR derived global snow cover parameters, Ann. Glaciol., 9, 39–44, 1987.

Cosgrove, B., Lohmann, D., Mitchell, K., Houser, P., Wood, E., Schaake, J., Robock, A., Sheffield, J., Duan, Q., Luo, L., Higgins, R., Pinker, R., and Tarpley, J.: Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), Journal of Geophysical Research, 108, 8845, doi:10.1029/2002JD003316, 2003.

Crow, W. and Wood, E.: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97, Advances in Water Resources, 26, 137–149, 2003.

De Lannoy, G., Reichle, R., Arsenault, K., Houser, P., Kumar, S., Verhoest, N., and Pauwels, V.: Multi-scale assimilation of AMSR-E snow water equivalent and MODIS snow cover fraction observations in Northern Colorado, Water Resources Research, 48, doi:doi:10.1029/2011WR010588, 2012.

Dee, D.: On-line estimation of error covariance parameters for atmospheric data assimilation, Mon. Wea. Rev., 123, 1128–1145, 1995.

Derber, J. and Bouttier, F.: A reformulation of the background error covariance in the ECMWF global data assimilation system, Tellus, 51, 195–221, 1999.

Derber, J. C., Parrish, D., and Lord, S.: The new global operational analysis system at the National Meteorological Center, Weather and Forecasting, 6, 538–547, 1991.

Dong, J., Walker, J., and Houser, P.: Factors affecting remotely sensed snow water equivalent uncertainty, Remote Sensing of Environment, 97, 68–82, 2005.

Durand, M., Kim, E., Margulis, S., and Molotch, N.: A first order characterization of errors from neglecting stratigraphy in forward and inverse passive microwave modeling of snow, Geoscience and Remote Sensing Letters, IEEE, 8, 730–734, doi:10.1109/LGRS.2011.2105243, 2011.

Ek, M., Mitchell, K., Yin, L., Rogers, P., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J.: Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model, Journal of Geophysical Research, 108, doi:10.1029/2002JD003 296, 2003.

Foster, J., Sun, C., Walker, J., Kelly, R., Chang, A., Dong, J., and Powell, H.: Quantifying the uncertainty in passive microwave snow water equivalent observations, Remote Sensing of Environment, 94, 187–203, 2005.

Huang, C., Newman, A. J., Clark, M. P., Wood, A. W., and Zheng, X.: Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States, Hydrology and Earth System Sciences, 21, 635–650, doi:10.5194/hess-21-635-2017, http://www.hydrol-earth-syst-sci.net/21/635/2017/, 2017.

Kachi, M., Naoki, K., Hori, M., and Imaoka, K.: AMSR2 validation results, in: 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS, pp. 831–834, IEEE, 2013.

Kelly, R.: The AMSR-E snow depth algorithm: Description and initial results, Journal of the remote sensing society of Japan, 29, 307–317, 2009.

Kelly, R. E., Chang, A. T., Tsang, L., and Foster, J. L.: A prototype AMSR-E global snow area and snow depth algorithm, IEEE Transactions on Geoscience and Remote Sensing, 41, 230–242, doi:10.1109/TGRS.2003.809118, 2003.

Krenke, A.: Former Soviet Union hydrological snow surveys, 1966-1996. Edited by NSIDC, Digital media, National Snow and Ice Data Center/World data center for glaciology, 1998, updated 2004.

Kumar, S., Peters-Lidard, C., Tian, T., Houser, P., Geiger, J., Olden, S., Lighty, L., Eastman, J., Doty, B., Dirmeyer, P., Adams, J., Mitchell, K., Wood, E., and Sheffield, J.: Land information system: An interoperable framework for high resolution land surface modeling, Environmental Modeling and Software, 21, 1402–1415, 2006.

Kumar, S., Reichle, R., Peters-Lidard, C., Koster, R., Zhan, X., Crow, W., Eylander, J., and Houser, P.: A land surface data assimilation framework using the Land Information System: Description and Applications, Advances in Water Resources, 31, 1419–1432, doi:10.1016/j.advwatres.2008.01.013, 2008.

Kumar, S., Reichle, R., Koster, R., Crow, W., and Peters-Lidard, C.: Role of subsurface physics in the assimilation of surface soil moisture observations, Journal of Hydrometeorology, 10, 1534–1547, doi:10.1175/2010JHM1262.1, 2009.

Kumar, S., Reichle, R., Harrison, K., Peters-Lidard, C., Yatheendradas, S., and Santanello, J.: A comparison of methods for a priori bias correction in soil moisture data assimilation, Water Resources Research, 48, doi:10.1029/2010WR010261, 2012.

Kumar, S., Peters-Lidard, C., Mocko, D., Reichle, R., Liu, Y., Arsenault, K., Xia, Y., Ek, M., Riggs, G., Livneh, B., and Cosh, M.: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation, Journal of Hydrometeorology, 15, 2446–2469, doi:10.1175/JHM-D-13-0132.1, 2014.

Kumar, S., Peters-Lidard, C., Arsenault, K., Getirana, A., Mocko, D., and Liu, Y.: Quantifying the added value of snow cover area observations in passive microwave snow depth data assimilation, J. Hydrometeor., 16, 1736–1741, doi:10.1175/JHM-D-15-0021.1, 2015.

Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B., Jasinski, M., Mocko, D., Getirana, A., Lannoy, G. D., Cosh, M. H., Hain, C. R., Anderson, M., Arsenault, K. R., Xia, Y., and Ek, M.: Assimilation of Gridded GRACE Terrestrial Water Storage Estimates in the North American Land Data Assimilation System, Journal of Hydrometeorology, 17, 1951–1972, doi:10.1175/JHM-D-15-0157.1, 2016.

Li, H., Kalnay, E., and Miyoshi, T.: Simultaneous estimation of covariance inflation and observation errors in with an ensemble Kalman filter, Quarterly Journal of the Royal Meteorological Society, 135, 523–533, doi:10.1002/qj.371, 2009.

Liu, Y., Peters-Lidard, C., Kumar, S., Foster, J., Shaw, M., Tian, Y., and Fall, G.: Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska, Advances in Water Resources, 54, 208–227, doi:10.1016/j.advwatres.2013.02.005, 2013.

Liu, Y., Peters-Lidard, C., Kumar, S., Arsenault, K., and Mocko, D.: Blending satellite-based snow depth products with in-situ observations for streamflow predictions in the Upper Colorado River basin, Water Resources Research, 51, 1182–1202, doi:10.1002/2014WR016606, 2015.

Menne, M.J. and. Durre, I., Vose, R., Gleason, B., and Houston, T.: An overview of the global historical climatology network-daily database, J. Atmos. Oceanic Technol., 29, 897–910, doi:10.115/JTECH-D-11-00103.1, 2012.

Miyoshi, T. and Yamane, S.: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution, Mon. Wea. Rev., 135, 3841–3861, doi:10.1175/2007mwr1873.1, 2007.

Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T.: The new ECMWF ensemble prediction system: methodology and validation, Quarterly Journal of the Royal Meteorological Society, 122, 73–119, 1996.

Newman, A., Clark, M., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrol. Earth Syst. Sci., 19, 209–223, doi:10.5194/hess-19-209-2015, 2015.

Oki, T., Imaoka, K., and Kachi, M.: AMSR instruments on GCOM-W1/2: Concepts and applications, in: Geoscience and remote sensing symposium (IGARSS), 2010 IEEE International, pp. 1363–1366, IEEE, 2010.

Reichle, R.: Data Assimilation methods in Earth Sciences, Advances in Water Resources, pp. 1411–1418, doi:10.1016/j.advwatres.2008.01.001, 2008.

Reichle, R. and Koster, R.: Assessing the Impact of Horizontal Error Correlations in Background Fields on Soil Moisture Estimation, Journal of Hydrometeorology, 4, 1229–1242, doi:10.1175/1525-7541(2003)004<1229:ATIOHE>2.0.CO;2, 2003.

Reichle, R., McLaughlin, D., and Entekhabi, D.: Hydrologic data assimilation with the ensemble Kalman Filter, Monthly Weather Review, 130, 103–114, 2002.

Reichle, R., Koster, R., Liu, P. and. Mahanama, S., Njoku, E., and Owe, M.: Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR), Journal of Geophysical Research-Atmospheres, 112, doi:10.1029/2006JD008033, 2007.

Reichle, R., Crow, W., and Keppenne, C.: An adaptive ensemble Kalman filter for soil moisture data assimilation, Water Resour. Res., 44, doi:10.1029/2007WR006357, 2008.

Reichle, R., Kumar, S., Mahanama, S., Koster, R., and Liu, Q.: Assimilation of satellite-derived skin temperature observations into land surface models, Journal of Hydrometeorology, 11, 1103–1122, doi:10.1175/2010JHM1262.1, 2010.

Rodell, M., Houser, P., Jambor, U., Gottschalk, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J., Walker, J., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381–394, doi:10.1175/BAMS-85-3-381, 2004.

Rodriguez, E., Morris, C., Belz, J., Chapin, E., Martin, J., Daffer, W., and Hensley, S.: An assessment of the SRTM topographic products, Tech. Rep. JPL D-31639, Jet Propulsion Laboratory, Pasadena, CA, 2005.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D.and Koren, V., Duan, Q., K., M., Fan, Y., and Mocko, D.: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System Project Phase 2 (NLDAS-2), Part 1: Comparison Analysis and Application of Model Products, Journal of Geophysical Research: Atmospheres, 117, doi:10.1029/2011JD016048, 2012.

Table 1. Cumulative RMSE (mm) from various model and DA integrations at the four locations in the Great Lakes and Colorado Headwaters domains used in the Figures 6 and 8.

Experiment	GL domain		CH domain	
name	A	B	C	D
OL_FSNGL	381	252	424	276
DA_FSNGL	266	206	316	327
OL_FCLIM	275	201	402	142
DA_FCLIM	262	156	309	312
OL_FENS	169	167	179	215
DA_FENS	244	162	285	309

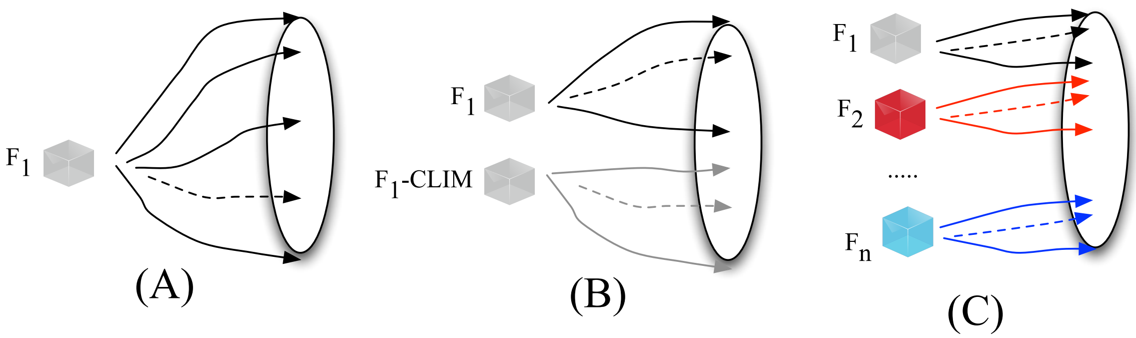


Figure 1. Schematic of the three strategies used to specify forcing uncertainty in the data assimilation integrations: (A) a single forcing dataset, (B) a single forcing dataset and its climatology and (C) an ensemble of forcing products. In all three cases, perturbations are applied to the forcing inputs to generate the ensemble.

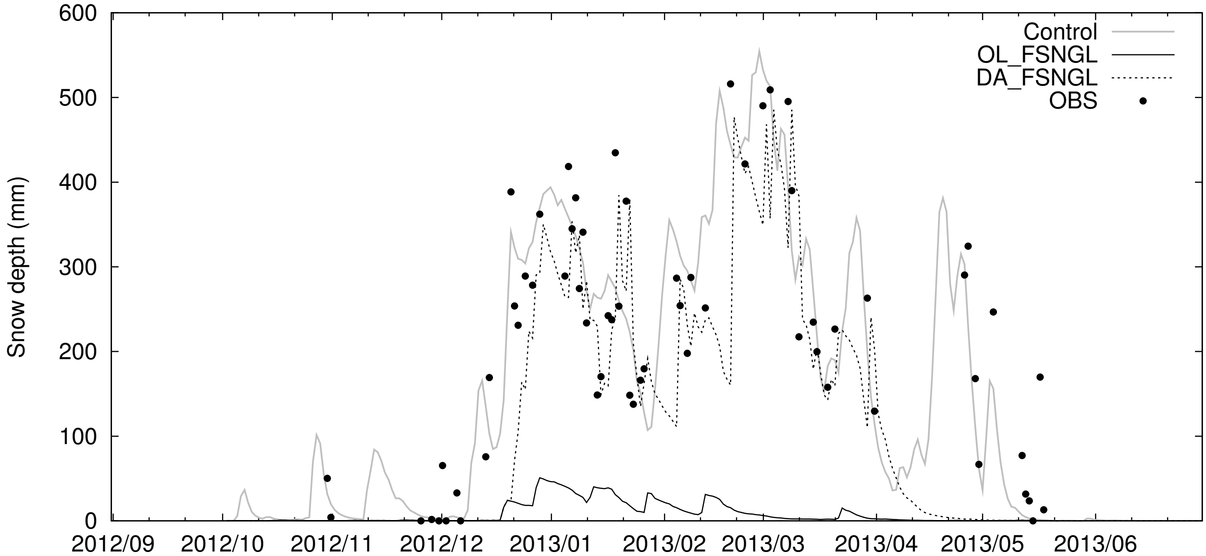


Figure 2. Snow depth time series for the water year of 2012-2013 from the open loop (OL_FSNGL) and data assimilation (DA_FSNGL) integrations using a single forcing dataset, for the synthetic snow data assimilation experiment. The Control simulation and the simulated observations are also shown.

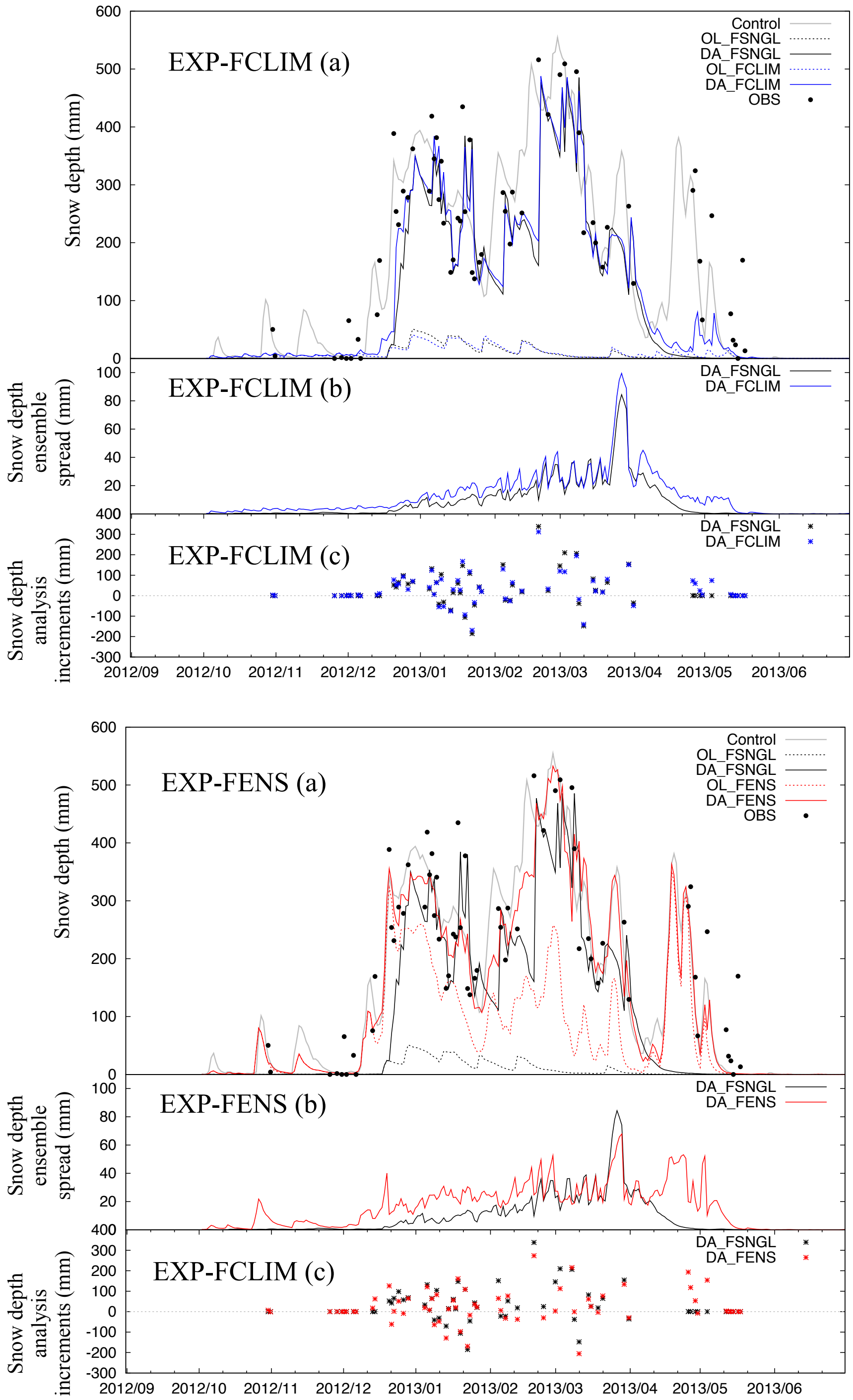


Figure 3. Similar to Figure 2, with the time series of model simulations from EXP-FCLIM and EXP-FENS included. The FCLIM experiments employ the use of a single forcing dataset and its climatology to force the ensemble and the FENS experiments employ the use of an ensemble of forcing datasets. The time series in panel (b) of the top and bottom figures compares the ensemble spread from the DA_FCLIM and DA_FENS integrations to the ensemble spread of DA_FSNGL integration, respectively. Panels (c) show comparison of the analysis increments from DA integrations.

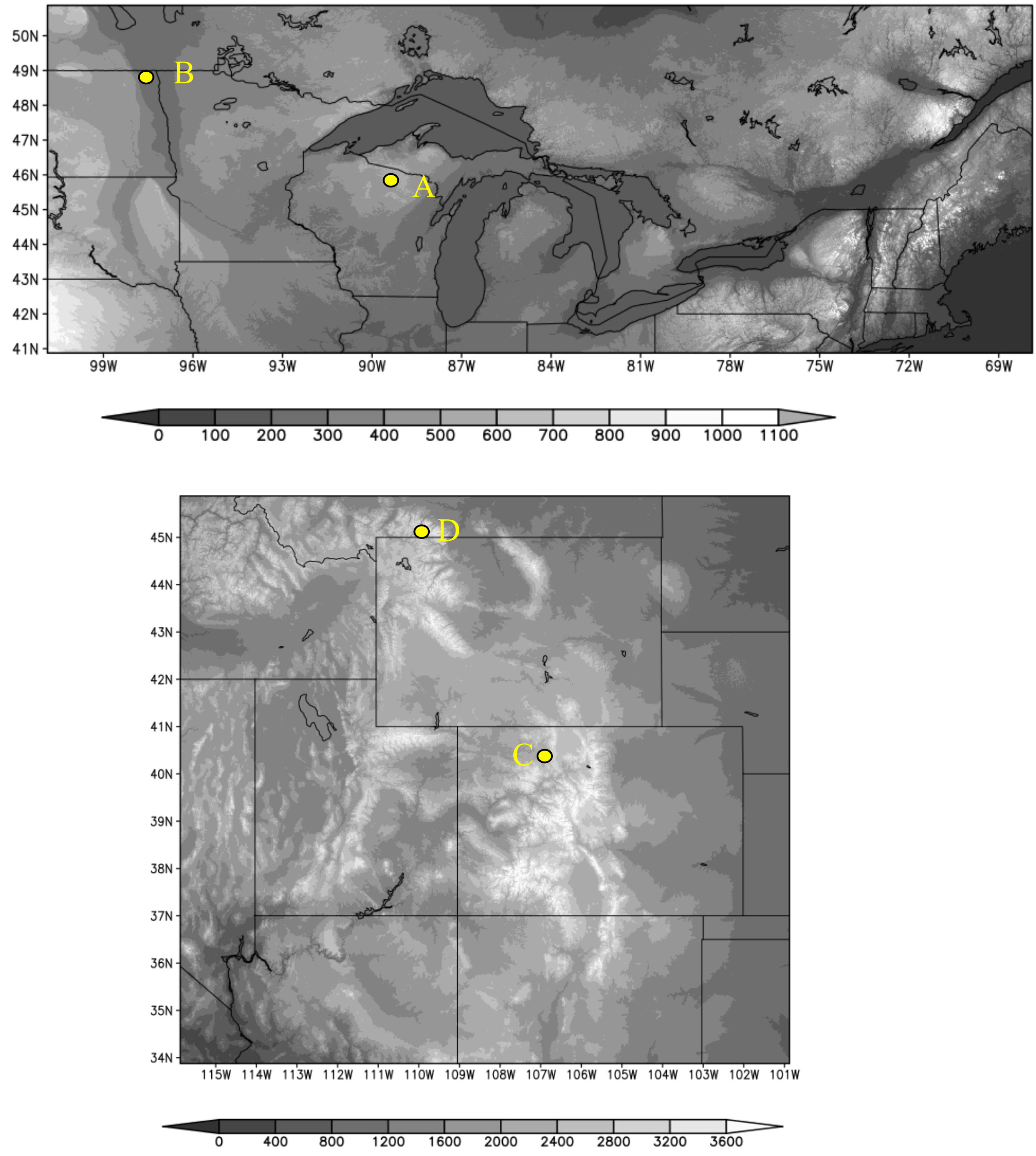


Figure 4. Two study domains with the 1 km terrain elevation (m) as the background: (top) GL domain and (bottom) CH domain. The yellow circles indicate the locations of the grid cells used for time series comparisons.

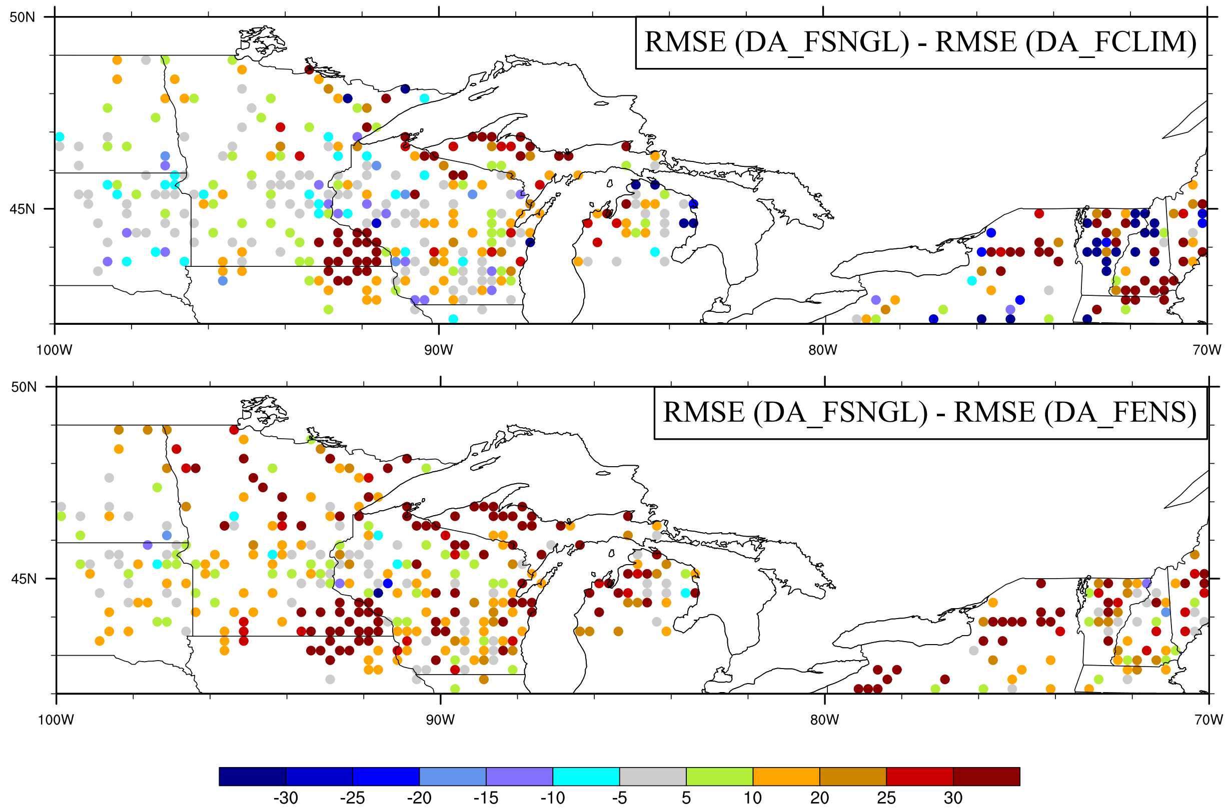


Figure 5. RMSE (mm) differences of snow depth fields from DA integrations using hybrid ensemble forcing strategies (DA_FCLIM and DA_FENS) relative to the DA integration using a single forcing (DA_FSNGL) over the Great Lakes domain, using GHCN data as the reference, for the time period of 2012 to 2015. Warm colors indicate locations where DA_FCLIM or DA_FENS provides a lower RMSE than DA_FSNGL and cool colors indicate locations where DA_FSNGL has a lower RMSE than DA_FCLIM or DA_FENS.

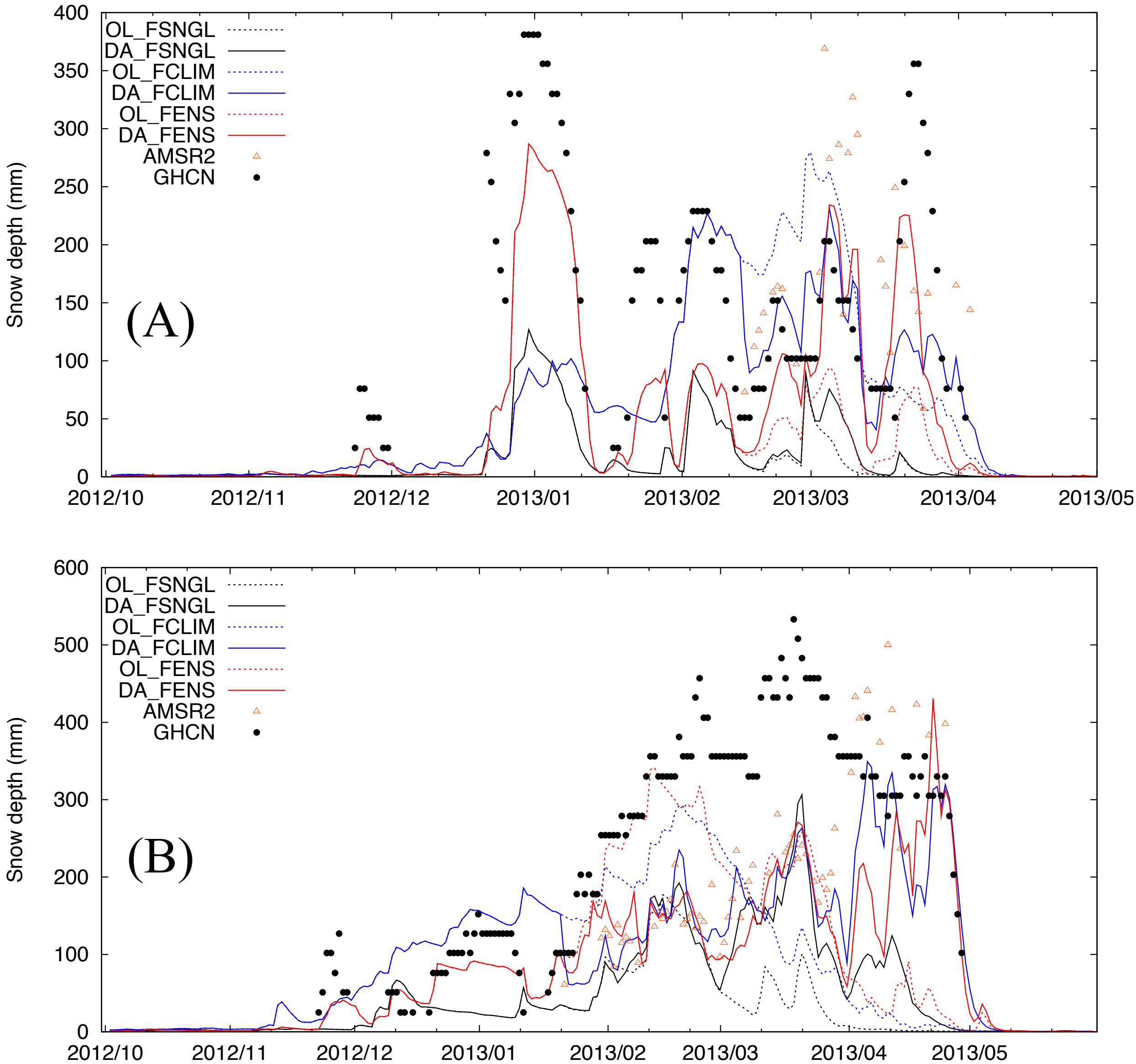


Figure 6. Time series of snow depth fields at location A (top) and B (bottom) from model open loop (OL_FSNGL, OL_FCLIM and OL_FENS), data assimilation (DA_FSNGL, DA_FCLIM and DA_FENS), AMSR2 and in-situ (GHCN).

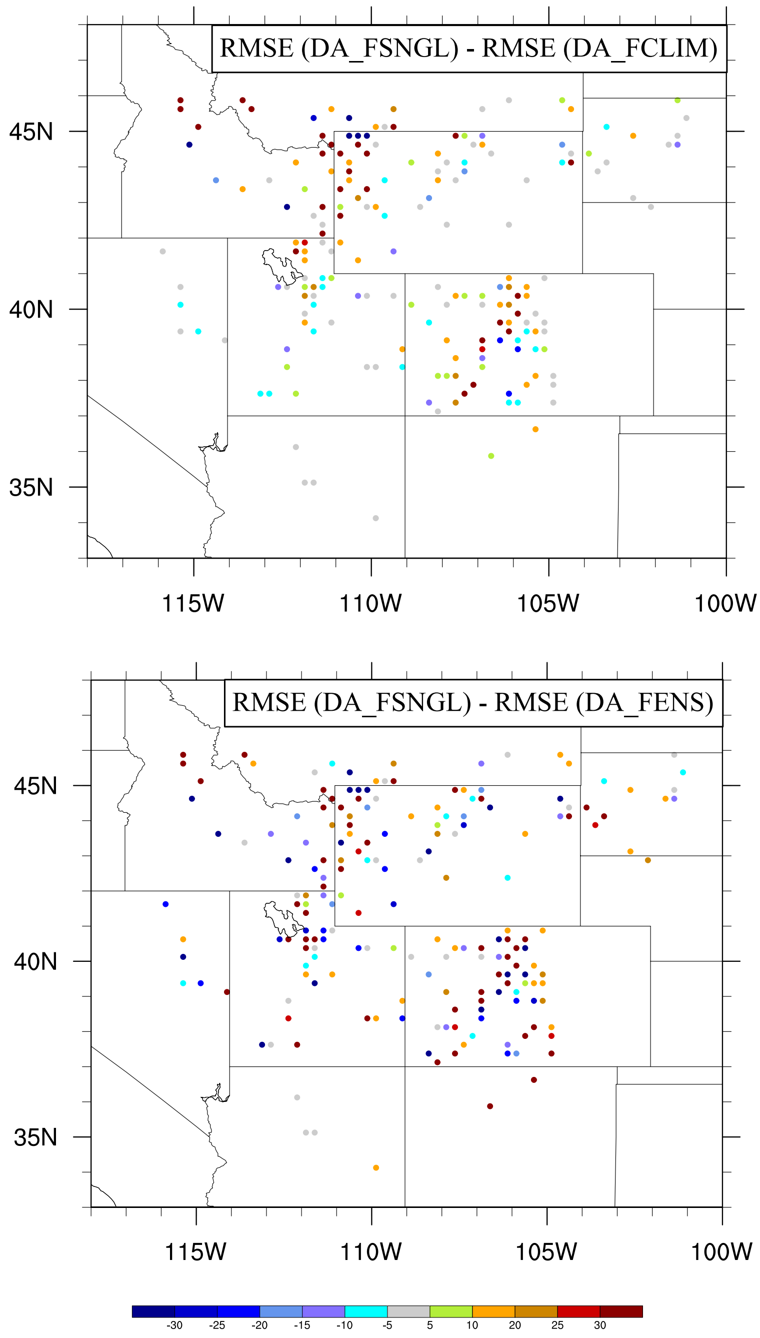


Figure 7. Same as Figure 5, but for the Colorado Headwaters domain.

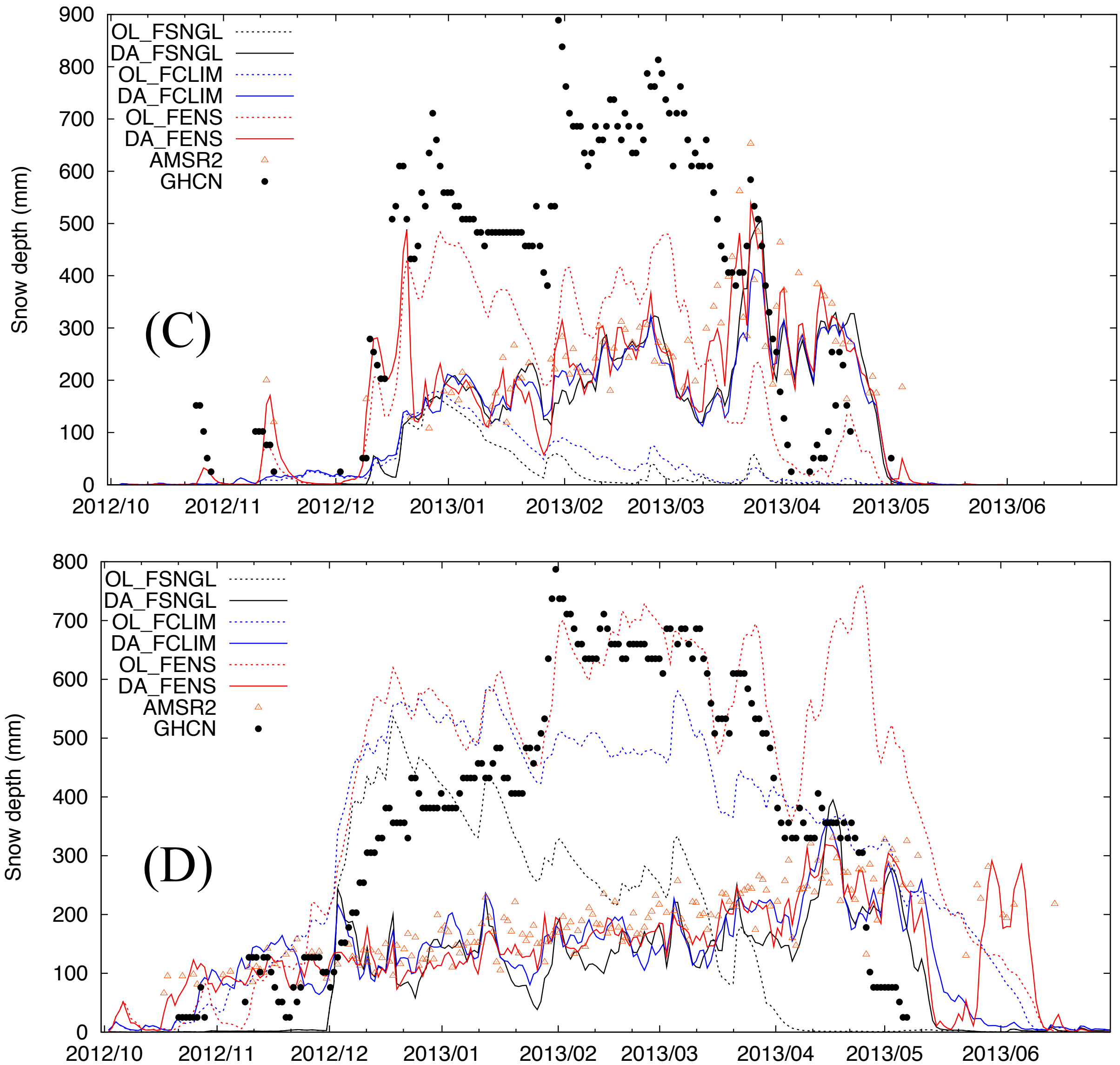


Figure 8. Time series of snow depth fields at location C (top) and D (bottom) from model open loop (OL_FSNGL, OL_FCLIM and OL_FENS), data assimilation (DA_FSNGL, DA_FCLIM and DA_FENS), AMSR2 and in-situ (GHCN).