Available online at www.sciencedirect.com
ScienceDirect
Advances in Water Resources 31 (2008) 1419–1432
Advances in Water Resources
www.elsevier.com/locate/advwatres

A land surface data assimilation framework using the land information system: Description and applications

Sujay V. Kumar [a,b,*], Rolf H. Reichle [a,c], Christa D. Peters-Lidard [b], Randal D. Koster [c], Xiwu Zhan [d], Wade T. Crow [e], John B. Eylander [f], Paul R. Houser [g]

[a] University of Maryland at Baltimore County, Goddard Earth Sciences and Technology Center, Baltimore, MD 21250, United States
[b] Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771, United States
[c] NASA Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD 20771, United States
[d] NOAA-NESDIS Center for Satellite Applications and Research, Camp Springs, MD 20746, United States
[e] USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705, United States
[f] Air and Space Models Integration Branch, HQ Air Force Weather Agency, AFB, NE 68113, United States
[g] Center for Research on Environment and Water, George Mason University, Calverton, MD 20705, United States

Received 17 May 2007; received in revised form 19 December 2007; accepted 20 January 2008
Available online 6 February 2008

Abstract

The Land Information System (LIS) is an established land surface modeling framework that integrates various community land surface models, ground measurements, satellite-based observations, high performance computing and data management tools. The use of advanced software engineering principles in LIS allows interoperability of individual system components and thus enables assessment and prediction of hydrologic conditions at various spatial and temporal scales. In this work, we describe a sequential data assimilation extension of LIS that incorporates multiple observational sources, land surface models and assimilation algorithms. These capabilities are demonstrated here in a suite of experiments that use the ensemble Kalman filter (EnKF) and assimilation through direct insertion. In a soil moisture experiment, we discuss the impact of differences in modeling approaches on assimilation performance. Provided careful choice of model error parameters, we find that two entirely different hydrological modeling approaches offer comparable assimilation results. In a snow assimilation experiment, we investigate the relative merits of assimilating different types of observations (snow cover area and snow water equivalent). The experiments show that data assimilation enhancements in LIS are uniquely suited to compare the assimilation of various data types into different land surface models within a single framework. The high performance infrastructure provides adequate support for efficient data assimilation integrations of high computational granularity.

Keywords: Land surface modeling; Data assimilation; Remote sensing; Hydrology; Soil moisture; Snow

1. Introduction

The land surface has a profound influence on regional and global weather and climate through the exchanges of moisture and energy between the soil, vegetation and snowpack with the overlying atmosphere. Therefore, a realistic characterization of the land surface moisture and energy stores that control these exchanges is important for improving our understanding and prediction of land–atmospheric interactions. An accurate characterization of the land surface can lead to improvements not only in weather and climate prediction, but also in other applications such as hazard mitigation (floods and droughts), agricultural production and water resources management. The accuracy of predictions from a Land Surface Model (LSM) depends on the model's representation of physical processes, the quality of the model inputs and forcings and the accuracy of the model parameters. Constraining the model predictions with observations via data assimilation

* Corresponding author. Tel.: +1 301 286 8663; fax: +1 301 286 8624.
E-mail address: Sujay.V.Kumar@nasa.gov (S.V. Kumar).

0309-1708/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2008.01.013

methods is an effective way to attenuate model errors and improve the model's predictive skills.

Data assimilation has been used in many scientific applications [25] as a way to improve deterministic model accuracy. Historically, the use of data assimilation techniques in the Earth sciences has been restricted primarily to meteorological and oceanographic applications. In recent years, many researchers have developed data assimilation techniques to exploit the increased availability of remotely sensed land surface variables [26,29,43]. A rapidly growing number of studies evaluate the assimilation of soil moisture, snow and surface skin temperature observations (e.g., [1,3,5,11,24,33,35,38,39,41,42]). These studies not only demonstrate the potential of data assimilation to improve land surface predictions, but also describe the difficulties in managing the complexities associated with data handling, computational burdens and the associated tradeoffs of various algorithms. Such difficulties are especially apparent when large amounts of remotely sensed observations of land surface states are assimilated.

As noted by [43], a comprehensive software framework is required that integrates physical models, observations, assimilation algorithms, and the necessary computing infrastructure to address the complexities associated with land surface data assimilation. The present article describes the development of such a framework as an extension of the existing Land Information System (LIS). As described by [22], LIS is a land modeling system that operates several community land surface models with the required initial and boundary conditions. LIS is designed with advanced software engineering principles and it provides many user extensible interfaces to incorporate diverse data sets from different sources as inputs to the LSMs. LIS also includes generic support for high performance computing, enabling the use of LSMs at global scales with spatial resolutions as high as 1 km. LIS has also been coupled to regional numerical weather prediction models to further enable the investigation of land–atmosphere interactions [21].

The LIS data assimilation extension described here is designed to be a generic, sequential component. Sequential assimilation algorithms step recursively through time, alternating between a model propagation step and a data assimilation update step. Examples of such algorithms are Direct Insertion (DI), the Extended Kalman filter (EKF), particle filters and the Ensemble Kalman filter (EnKF), each of which has been applied successfully to hydrologic data assimilation [36,38,44,45]. A sequential structure is also convenient for processing measurements in real time, such as in an operational setting. To date, the research nature of land surface data assimilation has limited most studies to the assimilation of a single measurement type into a single model using a single methodology. The LIS assimilation extension is designed to facilitate the much needed cross-comparison studies that evaluate the interoperable and integrated use of several kinds of assimilation algorithms, observations data sets and land surface models.

The paper is organized as follows. After a brief review of the Kalman filter (Section 2), we describe the design of the LIS interoperable data assimilation component (Section 3). The capabilities enabled by the design are illustrated with two sets of synthetic experiments that provide new insights in their own right. The first set of experiments investigates the use of different land surface models for soil moisture estimation (Section 4.1). In the second set, we compare the assimilation of two different types of snow observations (Section 4.2). Diagnostics of the assimilation performance and computational aspects are discussed in Section 5. Conclusions and a summary of future planned assimilation extensions to LIS are given in Section 6.

2. Sequential data assimilation

Sequential assimilation algorithms step recursively through time, alternating between a model propagation step and a data assimilation update step. The latter occurs whenever observations are available. In such a progression, the most recent updates reflect the accumulated information from all the observations up to this time. Here, we represent the nonlinear land surface model in the generic form

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{w}_k), \tag{1}$$

where $\mathbf{x}_k$ represents the state vector at time k, $\mathbf{f}(\cdot)$ is the nonlinear model forward operator, and $\mathbf{w}_k$ represents the uncertainties due to errors in the model formulation and uncertain model forcings. The observations at time $k+1$ denoted by $\mathbf{y}_{k+1}$ are connected to the system states via the observation operator $\mathbf{h}_{k+1}$:

$$\mathbf{y}_{k+1} = \mathbf{h}_{k+1}(\mathbf{x}_{k+1}, \mathbf{v}_{k+1}), \tag{2}$$

where measurement errors are represented by $\mathbf{v}_k$. The model and observation errors $\mathbf{w}_k$ and $\mathbf{v}_k$ are typically assumed to be independent Gaussian random vectors with mean zero and covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively, that are also uncorrelated in time.

The propagation step of the sequential assimilation algorithm consists of integrating equation (1) (from an initial state estimate $\hat{\mathbf{x}}_k^+$ at time k and with $\mathbf{w}_k = 0$) to give a model forecast $\hat{\mathbf{x}}_{k+1}^-$ at time $k+1$. In the update step, a new state estimate $\hat{\mathbf{x}}_{k+1}^+$ (also known as the analysis),

$$\hat{\mathbf{x}}_{k+1}^+ = \hat{\mathbf{x}}_{k+1}^- + \mathbf{K}(\mathbf{y}_{k+1} - \mathbf{h}_{k+1}(\hat{\mathbf{x}}_{k+1}^-)), \tag{3}$$

is computed based on the gain matrix $\mathbf{K}$ and on the innovations vector $(\mathbf{y}_{k+1} - \mathbf{h}_{k+1}(\hat{\mathbf{x}}_{k+1}^-)$; that is, the difference between the measurements and the model's prediction of the observations).

The gain matrix is generally a function of the model and observation error covariances. Its computation is the primary difference between the different sequential assimilation algorithms.

In the Kalman filter family of algorithms the error characteristics of the model are dynamically evolved in time, and for linear systems with error characteristics as specified

above, the gain is optimal (in the sense of minimum estimation error variance).

Though the LIS data assimilation framework accommodates different flavors of Kalman filters, we focus primarily on the use of the EnKF in this article. The salient feature of the EnKF is the approximation of model and forcing error covariances through the propagation of an ensemble of model trajectories. Each ensemble member is subject to a different realization of model and forcing errors. Each ensemble member could also employ different sets of model parameters or even entirely different land surface models, although such an approach is not used in this paper. After the assimilation update, the model is again evolved forward from the analysis state to the next observation time and the process is repeated [36].

3. Design of the LIS data assimilation component

The LIS software architecture employs object oriented programming paradigms to enable a modular and extensible system [22]. All functional pieces in LIS are implemented as extensible components, including LSMs, meteorological input schemes, sources of land surface parameters, modeling domains and running modes, following the principles of object oriented frameworks [2,28]. This plug-and-play design has enabled the inclusion of several user-defined extensions of each of these functional abstractions. The addition of data assimilation capabilities follows a similar extensible design.

The data assimilation extension in LIS is designed to be a sequential operator, where simulation variables are corrected at every observation time. Three main abstractions encapsulate the overall behavior of the data assimilation process defined in LIS (Fig. 1). These abstractions represent: (1) the data assimilation algorithm, (2) the observations, and (3) the land surface models. The data assimilation algorithm abstraction represents the specific algorithm that is used for the assimilation update, for example direct insertion or the EnKF. Various sources of observational data and corresponding measurement models that provide data to be assimilated are represented by the Observations abstraction. Examples of hydrological remote sensing data include surface soil moisture and snow water equivalent (SWE) retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) as well as snow cover observations from the Moderate Resolution Imaging Spectroradiometer (MODIS). Finally, the Land Surface Models abstraction represents various LSM implementations in LIS (such as the Noah LSM and the NASA Catchment LSM described below), along with the associated model parameters and surface meteorological forcing data.

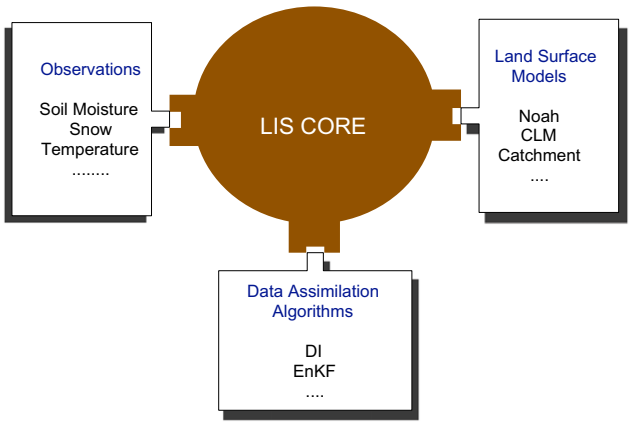

Fig. 1. Data assimilation abstractions in LIS.

The LIS core, which is the primary software that enables the integrated use of various extensible components in LIS, combines the use of these data assimilation abstractions. The abstractions defined by the LIS core need to be extended for each specific data assimilation instance. Data are exchanged between the various LIS components through the constructs of the Earth System Modeling Framework (ESMF, [18]. ESMF provides a standardized, self-describing format for data exchange between the system components shown in Fig. 1 through objects of the *ESMF_State* datatype.

Additional illustration of the sequential assimilation is provided in Fig. 2, which shows the interaction of the data assimilation abstractions for a single cycle of a typical sequential data assimilation operation (using the notation of Section 2). The vertical axis represents the sequence of operations and the horizontal axis represents the interacting modules in LIS as described above (Fig. 1). During a typical cycle (from time k to $k+1$), the forecast step is performed first to project the previous state (and possibly the error covariances) forward in time (from $\hat{\mathbf{x}}_k^+$ to $\hat{\mathbf{x}}_{k+1}^-$) and obtain the a priori estimates for the current timestep. Next, the observations to be assimilated by the `Observation Module` are read and packaged as an *ESMF_State* object, represented by the *OBS_State* y_{k+1}. This is followed by the measurement update (or analysis) step that incorporates the new measurements into the a priori estimate $\hat{\mathbf{x}}_{k+1}^-$ to obtain an improved a posteriori estimate $\hat{\mathbf{x}}_{k+1}^+$, represented by the updated *LSM_State*. This cycle repeats until the end of the simulation. Additional components and interactions are added to simulate the uncertainty and error propagation in specific instances of a sequential data assimilation algorithm such as the EnKF. For example, a `Perturbation Module` is included for the EnKF to control the evolution of model states and error characteristics. For this study, we implemented the EnKF and perturbation software developed by [33] as LIS modules.

The ESMF data objects that enable the exchange of information between data assimilation components contain not only the data but also certain self-describing metadata. For example, the *LSM_State* object contains the prognostic state variables to be assimilated. It also contains attributes describing the data such as the names, units, sign conventions and maximum and minimum acceptable values. Similarly, the *OBS_State* object contains the observation data

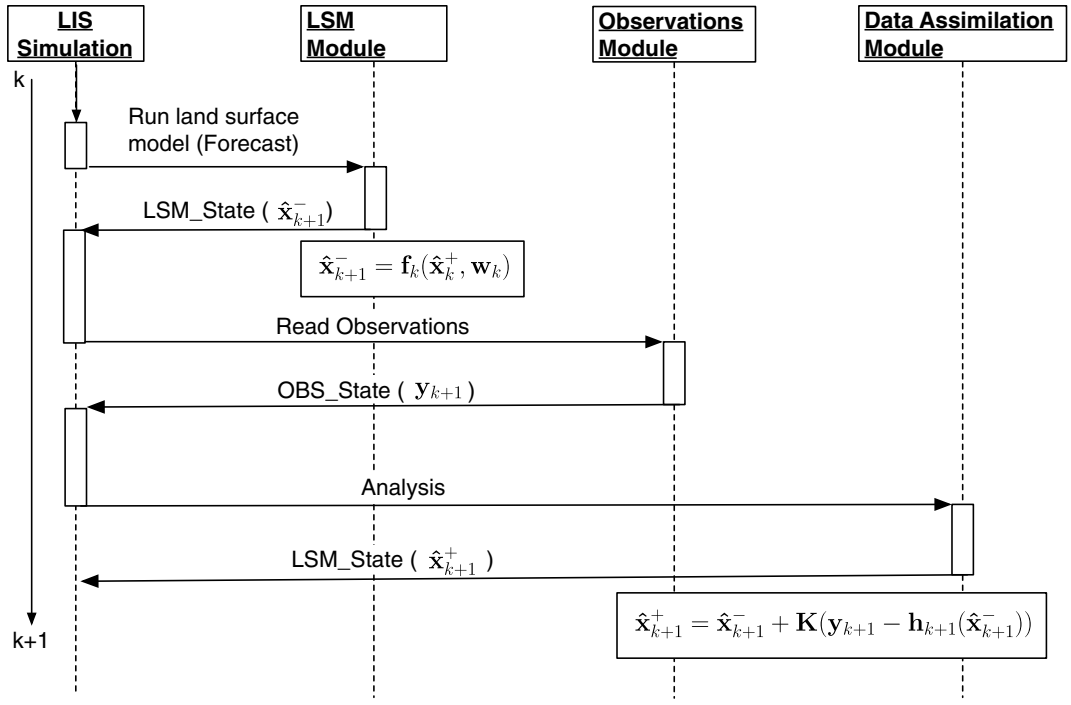

Fig. 2. Sequence of component interactions for a sequential data assimilation cycle.

and several attributes that describe them. These objects are queried to retrieve the data and the metadata by data assimilation specific segments of the framework. In other words, the observations and the relevant LSM prognostic variables must be packaged as self describing *ESMF_State* objects to select a specific data assimilation instance in LIS. The use of *ESMF_State* objects thus provides a flexible representation to describe the data being exchanged.

4. Data assimilation experiments and results

The implementation of data assimilation extensions in LIS facilitates an advanced system that enables the interoperable use of multiple data assimilation algorithms, land surface models and observations in a single framework. In addition, the existing support for computational parallelism enables the use of the system for computationally demanding applications. This section presents two sets of data assimilation experiments that demonstrate these capabilities. The first set of experiments investigates the use of different land surface models for soil moisture estimation (Section 4.1). In the second set, we compare the assimilation of two different types of snow observations (Section 4.2). Each type of snow observation is assimilated using an appropriate assimilation algorithm.

To demonstrate the interoperability of land surface models in a data assimilation mode, two different LSMs are used in the experiments: the Noah LSM [14] and the NASA Catchment LSM [20]. Both models dynamically predict land surface water and energy fluxes, but the model structures and parameterizations differ. For example, the Catchment LSM employs a topographically based hydrologic catchment representation with an explicit treatment of subgrid soil moisture variability and its effect on runoff and evaporation. By contrast, Noah uses a layer-based approach to soil moisture modeling. The Catchment LSM and the Noah LSM include physically based but distinct snow sub-models to represent processes related to snowpack evolution. The Catchment LSM uses three snow layers, whereas Noah uses only one snow layer. The two types of snow observations used in the second set of experiments are snow-cover area (SCA) and SWE.

The two sets of synthetic experiments are designed to demonstrate the sensitivity of the assimilation process to model parameterizations and physical representations. They also provide an opportunity to compare and evaluate the relative merits of assimilating different data types with different assimilation algorithms within a single framework. The basic structure of the experiments is as follows: First, the model is integrated to obtain the assumed "true" state of the land surface, referred to as the control (or truth) run. The observations to be assimilated are then generated from the control run's outputs by introducing realistic retrieval errors. Degraded or "open loop" simulations are conducted using uncertain inputs to degrade the model estimates. The uncertainty in the inputs is introduced by using a meteorological forcing data set for the time period that is different from that of the truth run (e.g. as produced by a different data provider), primarily to ensure that the errors in the open loop simulations are realistic. Finally, the assimilation integrations are conducted by assimilating the synthetic observations into the open loop simulation for each grid cell independently (also

known as "one-dimensional" assimilation, [34]). Since the true fields are known in these experiments, we can easily assess the performance of the assimilation approaches.

4.1. Soil moisture experiment

Soil moisture has a profound influence on the temporal and spatial variability of weather and climate conditions because it plays an important role in controlling the exchange of water and energy between the land surface and the atmosphere through evaporation and plant transpiration. Surface soil moisture observations can be obtained from measurements of microwave radiation emitted by the land surface. Hydrologic data assimilation efforts have focused on improving soil moisture estimates by incorporating such remote sensing data. In this section, we describe a suite of experiments that simulate the assimilation of remotely-sensed surface soil moisture into the Catchment and Noah LSMs.

Moving from the surface downward, the vertical subsurface representation in the Noah LSM includes four soil layers: a 10-cm thick surface layer, a 30-cm thick root zone layer, a 60-cm thick deep root zone layer, and a 1-m thick sub-root zone layer. The model keeps track of eight prognostic soil moisture variables, two for each layer (the total volumetric soil moisture and the liquid fraction of the volumetric soil moisture content). Noah employs the vertically integrated Richards equation [37] to compute the transfer of water within the soil profile.

In contrast, the Catchment LSM uses three bulk moisture variables and topographic characteristics to diagnose subsurface moisture for each catchment (or computational unit). These non-traditional bulk moisture variables represent equilibrium conditions associated with the water table distribution and the non-equilibrium conditions in the root zone and near-surface. The Catchment LSM explicitly computes the sub-catchment spatial distribution of moisture based on topography, subdividing the catchment into three distinct moisture regimes [12]. Soil moisture in a 2-cm surface layer and in a 1-m root zone layer (which includes the surface layer) are diagnosed from the bulk moisture variables.

In both land surface models, soil moisture responds similarly to the meteorological forcing, with soil layers closer to the surface having a faster response and the fluctuations in the deeper layers becoming less variable. The characteristic response time scales and the vertical soil moisture coupling of soil moisture are different for the two models. In the following, surface soil moisture for Catchment refers to its 2-cm surface layer and surface soil moisture for Noah refers to its top 10-cm layer. For the Catchment LSM, root zone soil moisture refers to the model's 1-m root zone layer. For Noah, an equivalent 1-m thick layer is computed as the weighted average of Noah's top three layers and is henceforth referred to as root zone soil moisture.

The modeling domain for the soil moisture experiment covers roughly the Continental United States (CONUS, from 30.5°N,124.5°W to 50.5°N,75.5°W) at 1° spatial resolution. The two land models are spun up from January 1, 2000 using surface meteorological forcings from the Global Data Assimilation System (GDAS; the global operational weather forecast model of the National Center for Environmental Prediction [9]). The eight months from April 1 to December 1, 2003 are used as the experiment period. "True" land surface conditions are simulated by integrating each model with GDAS forcing. Synthetic observations of surface soil moisture for each model are generated from the respective truth integrations by simulating retrieval errors typically associated with soil moisture products from microwave sensors. To account for difficulties in retrieving soil moisture under dense vegetation canopies, the synthetic surface soil moistures are masked out for both experiments when Green Vegetation Fraction values (used in Noah) exceed 0.7. The limitations of soil moisture retrievals in the presence of precipitation or snow are also simulated by introducing data masking for these events. Moreover, random Gaussian noise with an error standard deviation of $0.03\ m^3\ m^{-3}$ (volumetric soil moisture) is added to the synthetic observations to mimic measurement uncertainties. The magnitude of these perturbations is an optimistic estimate of error levels in future surface soil moisture retrievals from space-borne L-band radiometers. Next, open loop simulations are conducted by forcing the land surface models with surface meteorological forcing from the Goddard Earth Observing System (GEOS; [27]). In other words, model errors are represented by the difference between the GDAS and the GEOS forcing data sets.

The synthetic observations from the Catchment LSM are then assimilated into the Catchment LSM open loop model (using GEOS forcing) with the EnKF, once a day at 12Z. Likewise, Noah retrievals are assimilated into the Noah model. Perturbations on meteorological forcing inputs, model prognostic variables and observations are applied to maintain an ensemble of land surface conditions that represent the uncertainty in soil moisture states [33], with details shown in Table 1. Zero-mean, normally distributed additive perturbations were applied to the downward longwave radiation and near surface temperature forcing, and log-normally distributed multiplicative perturbations (with mean 1) are applied to precipitation and downward shortwave radiation forcing. Furthermore, the Catchment LSM prognostic variables catchment deficit and surface excess are perturbed with additive noise. For the Catchment LSM, the surface soil moisture observations are used to adjust all three bulk soil moisture prognostic variables (catchment deficit, root zone excess and surface excess) in the EnKF update step. Similarly, the Noah prognostic variables related to all four soil layers are updated in response to the surface soil moisture observations. Vertical correlations are imposed on the perturbations for the Noah soil moisture prognostic variables for individual layers.

To assess the performance of the assimilation integrations, root mean square errors (RMSEs) with respect to the true simulation are computed for the surface and root

Table 1
Parameters for perturbations to meteorological forcings and model prognostic variables in the EnKF assimilation experiments

Variable	Perturbation type	Standard deviation	Cross-correlations with perturbations in			
Meteorological forcings						
Near surface air temperature	Additive	1 K	n/a			
Downward shortwave	Multiplicative	0.5 (dimensionless)				
Downward longwave	Additive	50 W/m^2				
Precipitation	Multiplicative	0.30 (dimensionless)				
Catchment soil moisture assimilation experiment						
Catchment deficit	Additive	0.02 mm	n/a			
Surface excess	Additive	0.07 mm				
Noah soil moisture assimilation experiment			sm1	sm2	sm3	sm4
Total soil moisture – layer 1 (sm1)	Additive	1.37E − 3 $m^3\ m^{-3}$	1.0	0.6	0.4	0.2
Total soil moisture – layer 2 (sm2)	Additive	0.80E − 4 $m^3\ m^{-3}$	0.6	1.0	0.6	0.4
Total soil moisture – layer 3 (sm3)	Additive	0.60E − 4 $m^3\ m^{-3}$	0.4	0.6	1.0	0.6
Total soil moisture – layer 4 (sm4)	Additive	0.40E − 4 $m^3\ m^{-3}$	0.2	0.4	0.6	1.0
Noah SWE assimilation experiment						
Snow water equivalent	Multiplicative	0.01 m	n/a			
Snowdepth	Multiplicative	0.01 m				

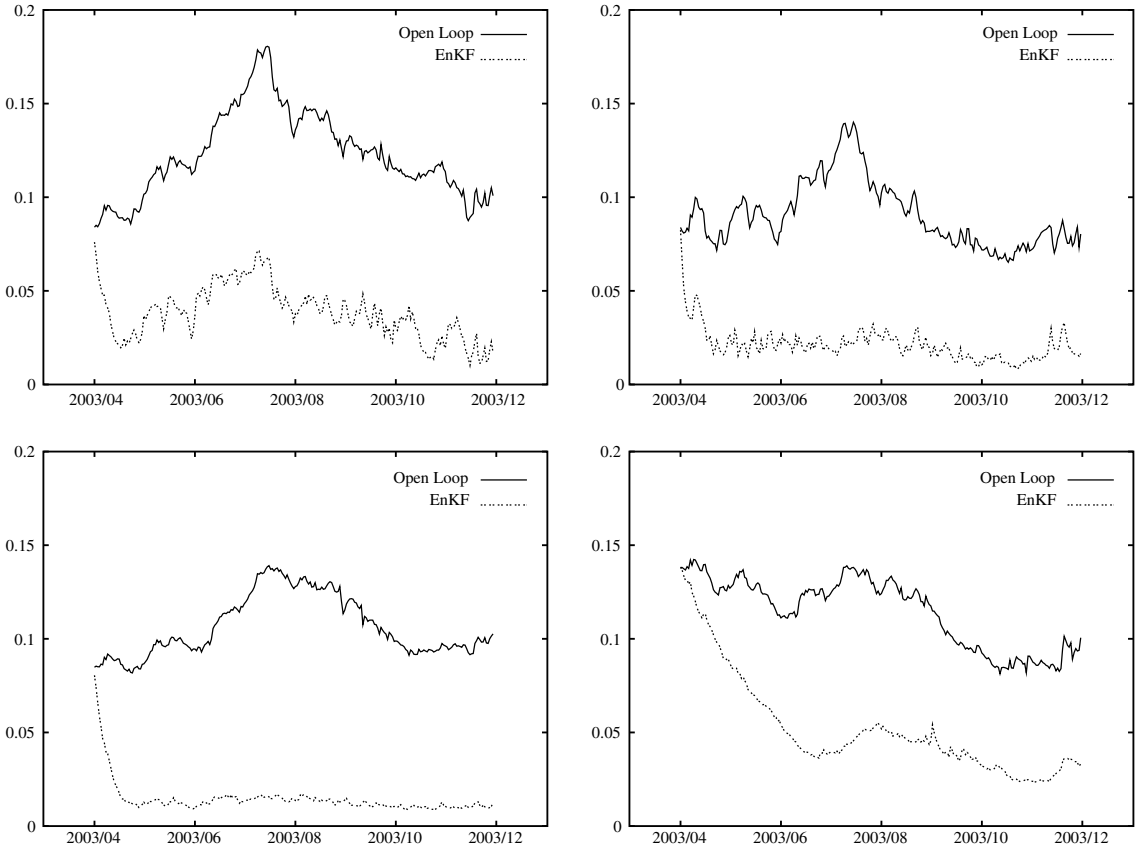

Fig. 3. RMS errors in (top) surface and (bottom) root zone soil moisture from the EnKF and the open loop simulations for the (left) Catchment LSM and the (right) Noah LSM assimilation experiments. Units are volumetric soil moisture ($m^3\ m^{-3}$).

zone soil moisture for both models and compared to the corresponding RMSEs of the open loop simulations. Fig. 3 shows average time series of RMSE for the entire modeling domain. The RMSE values are computed using the ensemble mean values and are based only on times and locations for which retrievals were available. It is clear from the figure that the assimilation simulations for both models rapidly reduce the RMSEs in the surface and root zone estimates, and generate systematic improvements over the open loop simulation throughout the entire simulation

time period of the simulation. The responses to the surface layer assimilation is comparable in Noah and Catchment, even though the representation of surface soil moisture dynamics is different in the two models. The root zone improvements are more rapid in Catchment LSM, presumably due to the stronger vertical soil moisture coupling in the Catchment LSM when compared to Noah. Note that the model error parameters of Table 1 must be carefully specified in order to achieve the documented assimilation improvements. This is particularly true for the vertical error correlations in the Noah soil moisture prognostics. Note also that the realism of the vertical coupling in the two models relative to nature is not being addressed here.

To evaluate the improvements in spatially distributed soil moisture fields, we compute, as an improvement metric, the difference between the RMSEs of the open loop simulation and the EnKF integration at each grid cell over the 8-month experiment period. Fig. 4 shows time average improvement maps of the surface and root zone fields for both LSMs. The improvement metric will be positive if the assimilation provides better estimates than the open loop simulation and will be negative if the assimilation degrades the fields. For both models, the assimilation provides a consistent improvement throughout the domain, although the magnitudes of improvement varies spatially. Generally, the improvements are greater in the West where more surface soil moisture retrievals are available. The generally denser vegetation in the East makes soil moisture retrieval more difficult and limits the potential improvements through data assimilation. Similar to the trends in Fig. 3, the improvements in surface and root zone soil moisture in the two LSMs are comparable. Again, it is important to note that our results do not indicate which land surface model is superior. Rather, our experiment showcases the use of a single system to evaluate data assimilation performance with different model representations and processes. The LIS system is uniquely suited to investigate and compare a variety of data assimilation strategies within a single consistent framework.

4.2. Snow experiment

Snow processes have a large impact on land-atmosphere energy exchanges due to snow's high albedo and insulating properties, which strongly influence surface and ground temperatures. Moreover, the seasonal water storage in the snowpack and associated spring snowmelt dominate the hydrology of many middle- to high-latitude alpine catchments. Presently, snow observations in the form of fractional snow cover and SWE are available from a variety of remote sensing platforms. Fractional snow cover (or snow cover area; SCA) observations are typically obtained from visible or infrared satellite sensors, which provide observations at high spatial resolutions [17]. However, retrievals from these sensors are limited to cloud-free conditions. In contrast, passive microwave sensors can provide quantitative observations of snow water equivalent (SWE; [19]. They can measure snow mass under cloudy and nighttime conditions, but have coarser spatial resolution. Dense vegetation cover and proximity to open water causes large retrieval errors in the passive microwave retrievals [15]. Finally, SWE retrievals are not sensitive to thin snow packs (SWE less than around 10 mm) and saturate for very thick snow packs (SWE above around 200 mm)

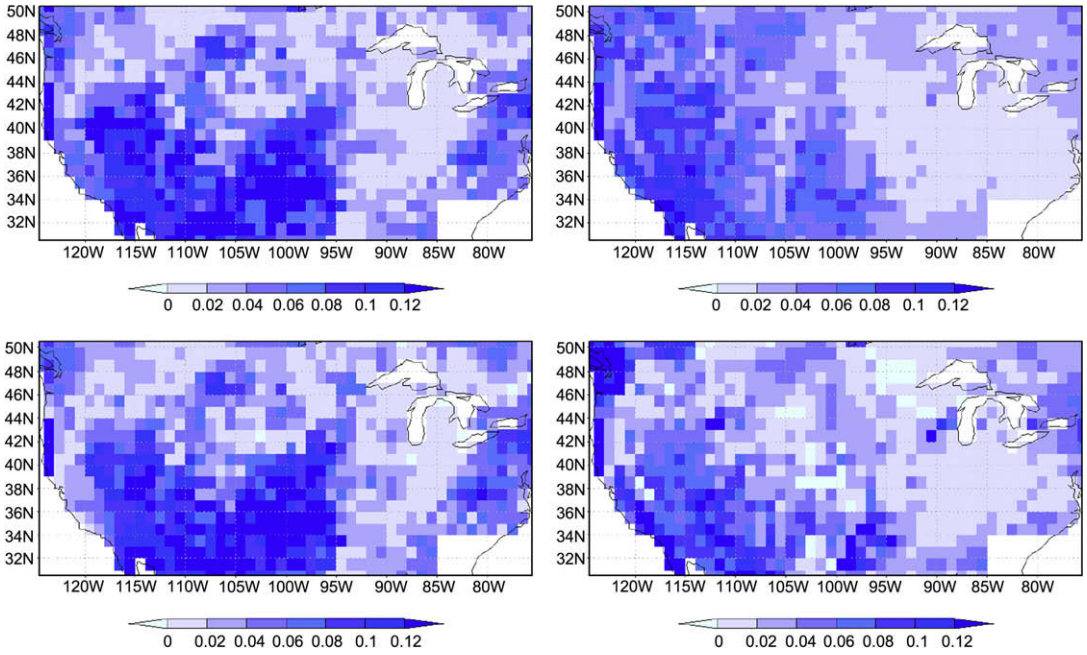

Fig. 4. Time averaged improvement metric (RMSE(open loop) – RMSE (EnKF)) for (top) surface soil moisture and (bottom) root zone soil moisture from the (left) Catchment LSM and the (right) Noah LSM assimilation experiments. Units are volumetric soil moisture ($m^3\ m^{-3}$).

[10]. While SCA and SWE observations have clear limitations, both are viable sources of information to improve model snow estimates [38–40]. Here, the LIS data assimilation extension is used to highlight the relative benefits of both snow observation types, SCA and SWE, within a single modeling and assimilation framework. In this experiment, we use the Catchment LSM with GDAS forcing for the truth simulation, and the Noah LSM with GEOS forcing for the open loop and assimilation integrations.

The snow model component of the Catchment LSM includes three-layer physical representations for snowpack accumulation, evaporation, sublimation, snowmelt, abatement and aging [23,12]. The Noah snow model component uses a single layer and simulates the physical processes of temporally varying snow density, allowing patchy snow cover to evolve as a function of snow depth and vegetation type [13]. The modeling domain covers North America, with a spatial extent from 25.5°N,140.5°W to 70.5°N,60.5°W, at 1° spatial resolution. Both LSMs are spun up from January 1, 2000, with GDAS forcing. The eight months from October 1, 2003 to June 1, 2004 are used as the experiment period. Synthetic snow cover and SWE observations are generated from the Catchment/GDAS truth run as follows: The fractional snow cover data are masked with realistic cloud cover flags from a daily 0.05° resolution MODIS climate-modeling grid-level-3 product (MODIS10C1; [17]). Similarly, the synthetic SWE observations are masked wherever model Leaf Area Index (LAI) values exceed 3.5. The synthetic SWE observations are further degraded by introducing random noise with a 10 mm standard deviation to simulate potentially large errors in the SWE retrievals. Finally, the lack of sensitivity of passive microwave observations to very thin and very thick snow packs is simulated by eliminating SWE observations whenever "true" SWE is below 10 mm or above 200 mm.

Two sets of data assimilation runs are conducted: (1) The synthetic SCA observations are assimilated into Noah/GEOS using the rule-based approach of [38] (hereinafter referred to as "SCA assimilation"), and (2) the synthetic SWE observations are assimilated into Noah/GEOS following an EnKF-based approach (hereinafter referred to as "SWE assimilation"). In each method, the observations are assimilated once a day, at 12Z. The rule-based approach for SCA assimilation directly corrects the SWE fields when there is a mismatch between the observations and the model prediction. If, for example, a SCA observation indicates that snow is absent but the model estimates that snow is present, the model's snow is removed. If, on the other hand, a SCA observation indicates that snow is present but snow is absent in the model, then a nominal amount of 5 mm SWE is added to the model states. Based on the snow density in the model, the snow depth fields are also updated. This empirical approach is used since the SCA observations simply provide information on the presence or absence of snow and do not provide a measure of snow mass. In contrast, SWE observations are assimilated with the EnKF by correcting both the SWE and the snow depth fields based on the sampled ensemble covariance of the observations and the model prognostic variables (Eq. (3)).

Fig. 5 shows the comparison of the time-averaged improvement metric for the SWE and snow depth fields from the SCA and SWE assimilation integrations. The SCA assimilation shows little change when compared to the open loop. By design, SCA assimilation is only effective when there is a transition from snow-free to snow-covered conditions (or vice versa). Improvements through SCA assimilation are therefore limited to regions where snow cover changes frequently throughout the season. The SWE fields from the SWE assimilation show positive improvements throughout the domain, since SWE observations are directly assimilated. The snow depth fields from the SWE assimilation, however, show negative improvements in large areas, mostly at high latitudes. The discrepancies in the updated snow depth fields in the SWE assimilation is due to a mismatch in the relationship between SWE and the snow density in the Catchment LSM and the Noah LSM. Such a mismatch is representative of likely errors in either model when compared to the snow density relationships operating in nature. The discrepancies may also indicate a lack of snow density information in the assimilated SWE observations.

Figs. 6 and 7 provide time series comparisons of both SWE and snow depth at a few selected locations and provide some key insights into the relative merits of both snow observation types. At Plateau Mountain (Fig. 6a and b), the assimilation of SCA provides virtually no improvement in the snow variables since the snow cover observations and the model predictions largely agree on the presence of snow at this location. The evolution of the SWE and snow depth series in the SCA assimilation follows the Noah/GEOS predicted snow evolution, which is very different from the Catchment/GDAS truth. By contrast, the SWE estimates from the SWE assimilation show clear improvements over the open loop simulation and match the truth almost perfectly. However, the snow depth estimates from the SWE assimilation do not capture the true snow depth evolution nearly as closely, which is again related to the differences between the Catchment LSM and the Noah LSM in modeling snow density.

Fig. 6c and d demonstrates that at Chinook, MT, the early season snow buildup is poorly captured by either assimilation scheme. The Noah model driven with the GEOS forcing (open loop) simply does not sustain the snow evolution no matter what observation type is assimilated. As the season progresses, SWE assimilation estimates closely track the open loop, which already provides a reasonable SWE estimate. Later in the season, SWE assimilation again offers clear advantages. By contrast, SWE estimates from the SCA assimilation are worse than the open loop estimates for much of the season and only offer limited improvements late in the season. For snow depth, the assimilation of neither observation type provides much improvement over the open loop.

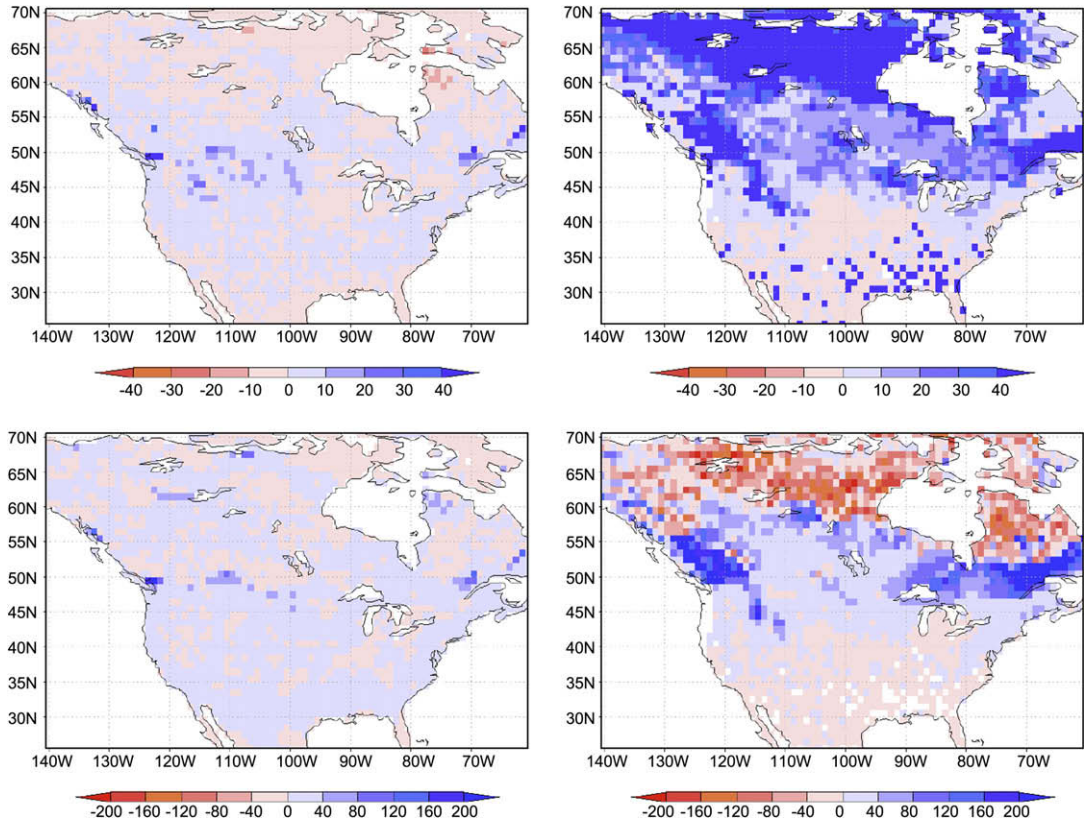

Fig. 5. Time averaged improvement metric [RMSE (open loop) – RMSE (assimilation)] for (top) SWE and (bottom) snow depth from the (left) SCA and (right) SWE assimilation experiments, respectively. Units are mm.

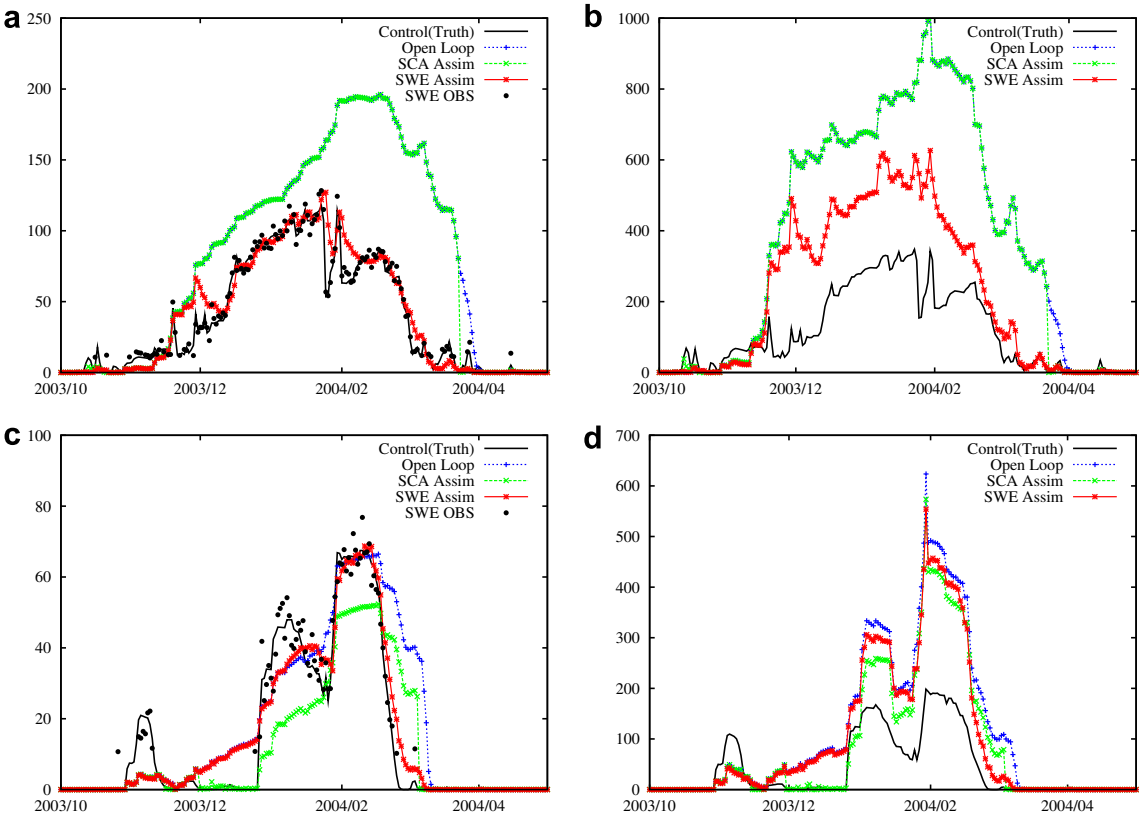

Fig. 6. (a,c) SWE and (b,d) snow depth at (a,b) Plateau Mountain (50.2°N, 116.5°W) and (c,d) Chinook, MT (48.6°N, 109.2°W). Units are mm.

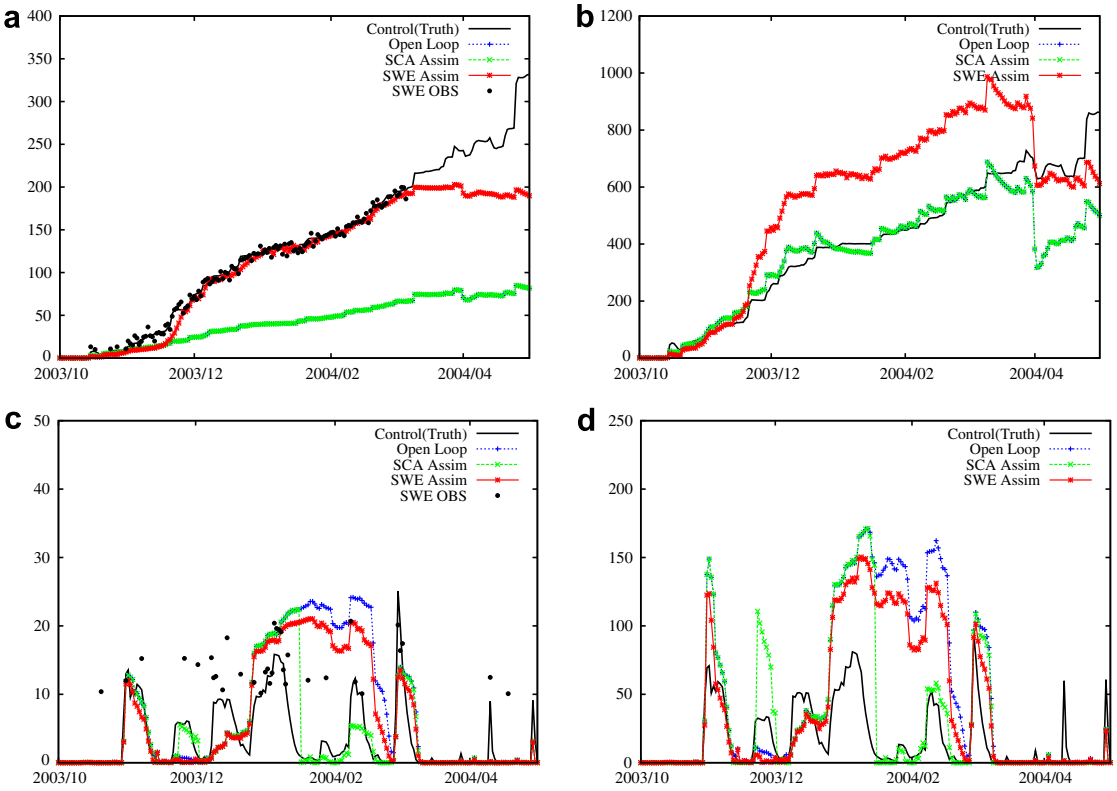

Fig. 7. (a,c) SWE and (b,d) snow depth at (a,b) Churchill RCT (58.8°N, 94.5°W) and (c,d) Riverton, WY (43.0°N, 108.4°W). Units are mm.

At Churchill, RCT, (Fig. 7a and b), a location with large snowfall, SWE assimilation provides an excellent estimate of the SWE time series until SWE reaches 200 mm, which is used as the upper threshold for the SWE retrieval algorithm. As a result, SWE observations are no longer available and the EnKF fails to capture the evolution of late snow season buildup. As for snow depth, SCA assimilation is superior, albeit for the wrong reason. Since at this snowy location the observed and modeled SCA largely agree, the estimates from the SCA assimilation track the open loop estimates very closely. The apparent success in snow depth estimation is because the open loop Noah/GEOS considerably underestimates SWE and even more dramatically underestimates the snow density. The fortuitous combination of these errors happens to provide the correct snow depth by accident, both for the estimates from the SCA assimilation and for the open loop.

Finally, at Riverton, WY, (Fig. 7c and d) estimates from the SCA assimilation capture the evolution of SWE and snow depth better than estimates from the SWE assimilation. The SWE assimilation, just like the open loop Noah/GEOS integration, overpredicts the SWE and SCA during February and March. In this particular climate, SWE oscillates around the minimum threshold of 10 mm. SWE observations over 10 mm are assimilated, but the actual "no snow" conditions are not detected by the SWE retrievals. The SWE assimilation integration is therefore biased high. The SCA assimilation, on the other hand, excels under these conditions. Here, the SCA assimilation accurately detects the mismatch between model and observation and forces the model to adjust towards the true field. As indicated in Fig. 5, however, the superior performance of the SCA assimilation at Riverton is the exception rather than the rule.

The experiments demonstrate the ability of the LIS data assimilation extension system to utilize different sequential assimilation approaches with different kinds of observations in a single framework. Such an integrated system provides the unique ability to compare and contrast the advantages and disadvantages of different observation types and assimilation approaches. It should be noted that the improvement metric in our snow assimilation example is geared towards SWE estimation. In a different context, the SCA assimilation is likely more valuable than our analysis suggests. Improvements in other surface fields (such as albedo) through SCA assimilation should have a positive impact on near surface temperature and heat fluxes in a coupled land–atmosphere system and may improve weather forecasts.

5. Impact of error representations and ensemble size

Sequential data assimilation approaches typically require estimates of the model and observation error covariances to properly merge model predictions with observations. The EnKF, for example, uses an ensemble

approach that dynamically estimates the model and observation error covariances. It is well known that the choice of input error parameters has a critical impact on the accuracy and performance of the assimilation algorithm [6,30]. In this section, we discuss a few strategies to diagnose the filter performance and highlight the computational capabilities of the LIS data assimilation extension.

5.1. Innovation metrics

Filter innovations, defined as the difference between the observations and the corresponding model forecasts (Section 2), are available in most data assimilation algorithms. Also available is the expected covariance of the innovations, which depends on filter input parameters (the model and observation error covariances). The statistics of the innovations can be used to diagnose the performance of the filter [31]. In a linear system with mutually and serially uncorrelated Gaussian model perturbations and observation errors, the innovations, after normalization with their expected covariance, should be serially uncorrelated – and possess a standard normal distribution $N(0,1)$ [16]. While such optimal conditions are seldom seen in real world applications, deviations of the normalized innovations statistics from the theoretical ideals can be used to diagnose suboptimal filter performance and successfully estimate model and noise error variance parameters [4,30].

The statistics of the normalized innovations for the three experiments are presented in Table 2. The distribution of normalized innovations in the soil moisture experiments have a negative bias. The SWE assimilation experiment shows a positive innovations bias.

Such non-zero means are expected because the GEOS forcing data set that is used for the assimilation integration is biased against the "truth" GDAS forcing data set. In the SWE experiment, the use of two entirely different snow models for the truth (Catchment LSM) and the assimilation integration (Noah) is a further source of bias. It is clear from Table 2 that these biases must be addressed as part of the assimilation system in order to improve the performance of the assimilation. Options include: (1) the a priori scaling of biases by matching the climatology of the observations to the climatology of the model as in [32] and (2) on-line bias correction [8,7].

Table 2
Statistics of the distribution of normalized innovations from soil moisture and SWE assimilation experiments

Experiment	Mean	Variance	Lag-one autocorrelation coefficient
Catchment soil moisture assimilation	−0.23	1.07	0.123
Noah soil moisture assimilation	−0.19	1.75	0.335
SWE assimilation	0.31	0.98	0.522

Provided modeling error is approximately unbiased and temporally white, a variance greater than unity and a positive serial correlation of the normalized innovations imply that model errors are underestimated [4]. This is the case for the Noah soil moisture experiment and likely degraded the performance of the filter by placing excessive confidence on the background model forecasts. Generally, better filter performance is diagnosed for the Catchment model (i.e. a variance closer to unity and a serial correlation closer to zero). Consequently, for the experiments to be truly comparable, further calibration or estimation of the respective model and observation errors would have to be conducted, but this is rarely performed for land data assimilation studies [6] and is beyond the scope of this paper. In addition, modeling error should be checked for bias and whiteness before innovation statistics are taken at face value. For the SWE assimilation experiment, the Noah model exhibits acceptable normalized variance, but a clear positive serial correlation. This combination implies that the filter is accurately capturing the total innovation spread, but over-predicting the contribution of observation error to this spread. The fraternal twin aspects of the SWE assimilation experiment, however, complicate such an inference by introducing the strong possibility of serially correlated model error.

In the soil moisture experiments, the same model that is used to generate the truth and the synthetic retrievals is also used for the assimilation. For these experiments, we find no obvious spatial pattern in the variance of the normalized innovations (not shown). In the SWE assimilation experiment, by contrast, different LSMs are used to generate the truth and the synthetic retrievals on the one hand, and the assimilation estimates on the other hand. Such a fraternal twin experiment is more likely to generate strong biases and complicate the accurate specification of model error parameters, but is more representative of assimilating real observations. Fig. 8 shows a map of normalized innovation variances for the SWE assimilation experiment. Variance values far from unity are observed, and a large-scale spatial pattern emerges that may be useful for further tuning of the input error parameters.

In summary, the innovations statistics can be used as a diagnostic metric for improving the overall performance of the EnKF in a particular assimilation setting. While the interpretation of innovation statistics is complicated by model nonlinearities and by typically biased and correlated errors, innovation-based diagnostics have been successfully applied in simple land data assimilation cases [4], and the development of adaptive strategies for exploiting these diagnostics in more complex land surface models is an ongoing area of study [30,6].

5.2. Ensemble size and computational aspects

The EnKF is a Monte-Carlo approximation of a sequential Bayesian filtering process. As a result, the accuracy of the sampled covariances depends on the ensemble

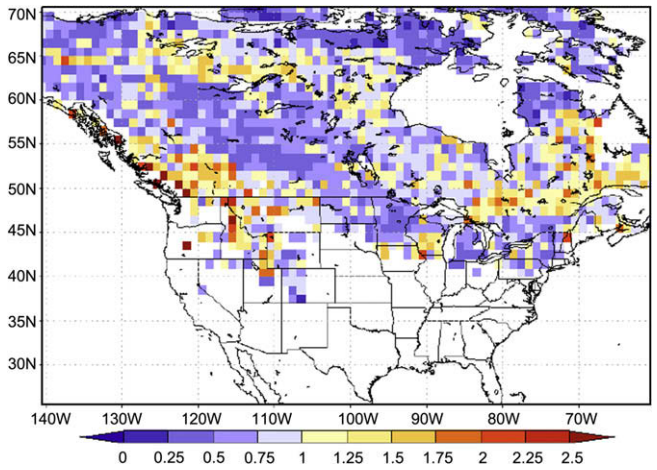

Fig. 8. Variance of normalized innovations (dimensionless) for the SWE assimilation experiment.

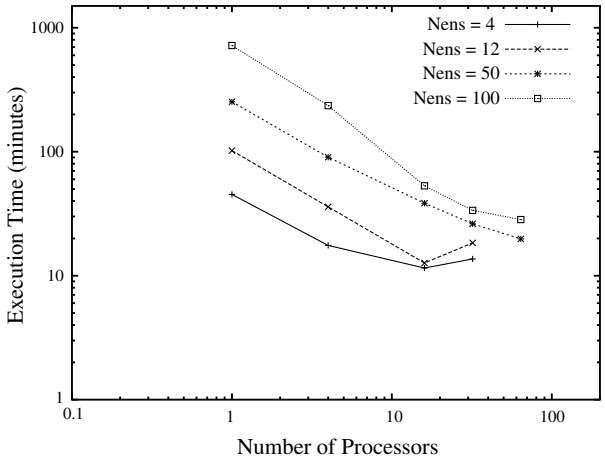

Fig. 10. Computational scaling of the LIS data assimilation framework for the Catchment LSM soil moisture assimilation experiment.

size. Although an increase in the ensemble size generally enables the algorithm to propagate the error information more accurately by reducing sampling noise, this option is usually limited by computational feasibility. In the following, we present a comparison of the computational tradeoff between ensemble size and estimation accuracy and highlight the computational capabilities of the LIS data assimilation extension.

Fig. 9 explores the EnKF's sensitivity to the size of the ensemble for the Catchment soil moisture assimilation experiment. The total RMS error for the entire modeling domain is compared against the ensemble size, which varies from 4 to 100. As expected, the figure shows that the errors decrease with increasing ensemble size. Since we are using a one-dimensional EnKF, the effective size of the EnKF state vector is just three. Therefore, ensemble sizes in excess of 10 provide little additional benefit.

The increase in the ensemble size increases the computational time requirements, which can be mitigated by the use of scalable computing infrastructure in LIS. Fig. 10 shows the scaling performance of the system for different ensemble sizes. When an ensemble size of 100 is employed, a serial simulation on a single processor requires approximately 12 h. This time requirement reduces to about 20–30 minutes with the parallel use of 64 processors. The increase in ensemble size also increases the computational granularity, which is a measure of the incremental computation associated with a given problem. Fig. 10 shows that the system exhibits better scalability with increased granularity. Thus, the high performance computing features of LIS provide adequate computational scaling to support computationally intensive data assimilation simulations.

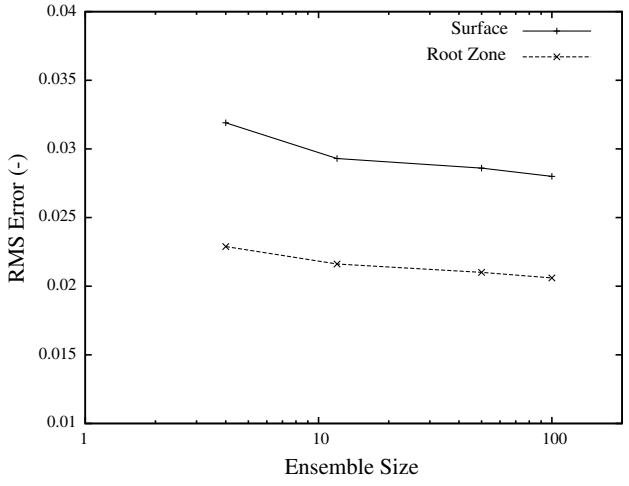

Fig. 9. RMSE of surface and root zone soil moisture (in units of volumetric fraction) for different ensemble sizes for the Catchment LSM soil moisture assimilation experiment.

6. Summary

The article describes a comprehensive land surface data assimilation system enabled by the LIS. The system is designed using advanced software engineering practices to provide a flexible and reusable framework. The data assimilation extensions support the use of multiple sequential data assimilation algorithms, different sources of observations, and their integrated use with different land surface models. The well-defined architecture allows rapid specification of new data assimilation extensions. The high performance infrastructure in LIS enables the use of the assimilation algorithms in a computationally efficient and scalable manner.

The new LIS data assimilation capabilities are demonstrated with a suite of synthetic experiments based on the LIS implementation of the EnKF developed by [33]. The interoperable use of multiple LSMs for data assimilation is demonstrated by a set of soil moisture experiments that demonstrate the assimilation of passive microwave-based soil moisture observations into the Catchment and Noah LSMs. The system also allows for the implementation of multiple data assimilation approaches. This is demonstrated by the implementation of approaches to assimilate

visible/infrared sensor-based fractional snow cover (SCA) and passive microwave sensor-based SWE observations. The (synthetic) SCA observations were assimilated with a rule-based direct insertion approach while the (synthetic) SWE measurements were assimilated with the EnKF.

Comparisons with the synthetic truth are used to illustrate the improved accuracy in the prediction of land surface states through data assimilation. In the soil moisture experiment, both Catchment and Noah's soil moisture estimates are improved by EnKF-based surface soil moisture assimilation. The Catchment LSM has an inherently stronger representation of soil moisture vertical coupling when compared to Noah. With careful specification of error parameters, however, both models provide comparable results. The experiment demonstrates that the performance of data assimilation in improving model estimates is strongly dependent on the model representation as well as the choice of model error parameters. A framework such as LIS is ideal for investigating the sensitivity and influence of these relationships.

The snow experiments provide a comparison of the advantages and disadvantages of assimilating two different observation types, SCA and SWE. SWE observations are typically spatially coarse and error-prone but provide a measure of snow mass. The EnKF can therefore be used as an effective technique for SWE assimilation. SCA only provides qualitative information on the presence or absence of snow. The rule-based assimilation of snow cover data are simple but provides a reasonable assimilation method for this data type. Typically, the improvements in the snow fields are larger when SWE is assimilated than when SCA is assimilated. The experiment also illustrates the impact of model representations of snow processes on the performance of the assimilation algorithms. In particular, the assimilation of SWE or SCA estimates may degrade estimates of snow depth if snow density is poorly modeled.

Innovations statistics from the experiments are used to characterize the performance of the EnKF and reveal deficiencies in model and observation error representations. LIS provides a prototyping environment to calibrate and improve the performance of the assimilation algorithm, using the information from the innovation statistics. As noted by [33], the success of the data assimilation is dependent on the model error parameters which are often calibrated in twin experiments. The rapid prototyping environment offered by LIS is therefore helpful in developing a successful assimilation product.

One of the primary motivations for LIS is the ability to perform global land surface modeling at fine (e.g., 1 km) spatial resolutions, taking advantage of high-resolution remotely sensed observations such as those available from NASA Earth science instruments. With the addition of data assimilation capabilities LIS can be used to study the impact of high resolution observations at the scale of the observations themselves and over large domains. The high performance capabilities in LIS are designed to handle the associated computational requirements. LIS has already been coupled to the (atmospheric) Weather Research and Forecasting model [21]. The data assimilation enhancements in LIS therefore provide the capability to investigate the impact of land surface observations in a coupled land–atmosphere environment.

Several additional enhancements to the system are currently in progress. An online bias correction module will be included to correct the biases that typically exist between satellite retrievals and model estimates of land surface fields. The present article only describes the use of sequential assimilation algorithms. The LIS architecture, however, supports different running modes that are not necessarily sequential in nature. These features can be exploited for the implementation of variational and smoothing approaches for data assimilation. Another functional extension planned for LIS is the support of single- and multi-objective optimization algorithms. The ability to use optimization features in an automated manner allows the adaptive tuning and calibrating of filter parameters to achieve optimal performance. Thus, the LIS data assimilation system is expected to evolve into a framework that can effectively utilize remote sensing data for improving our understanding of land surface water and energy cycling.

Acknowledgements

We gratefully acknowledge support of LIS data assimilation development by the Air Force Weather Agency, internal investment from NASA Goddard Space Flight Center, the NASA Energy and Water Cycle (NEWS) program, the NASA Earth Observing System (EOS) program, and a grant from the Joint Center for Satellite Data Assimilation (JCSDA), among others. We also thank Dr. Sarith Mahanama and James V. Geiger for their help in the implementation of Catchment land surface model in LIS and Dr. Jiarui Dong for helpful discussions.

References

[1] Andreadis K, Lettenmaier D. Assimilating remotely sensed snow observations into a macroscale hydrology model. Adv Water Resour 2005;29:872–86.

[2] Booch G, Maksimchuck RA, Engel MW, Young BJ, Conallen J, Houston KA. Object-oriented analysis and design with applications. Reading (MA): Addison-Wesley; 1993.

[3] Bosilovich M, Radakovich J, da Silva A, Todling R, Verter F. Skin temperature analysis and bias correction in a coupled land–atmosphere data assimilation system. J Meteorol Soc Jpn A 2007;85:205–28.

[4] Crow W, Bolten J. Estimating precipitation errors using spaceborne surface soil moisture retrievals. Geophys Res Lett 2007;34. doi:10.1029/2007GL029450.

[5] Crow W, 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. Adv Water Resour 2003;26:137–49.

[6] Crow WT, Van Loon E. Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. J Hydrometeorol 2006;7:421–32.

[7] De Lannoy GJM, Reichle RH, Houser PR, Pauwels VRN, Verhoest NEC. Correcting for forecast bias in soil moisture assimilation with the ensemble kalman filter. Water Resources Research 2007;43:W09410.
[8] Dee D, da Silva A. Data assimilation in the presence of forecast bias. Quart J Roy Meteorol Soc 1998;124:269–95.
[9] Derber J, Parrish D, Lord S. The new global operational analysis system at the National Meteorological Center. Weather Forecast 1991;6:538–47.
[10] Dong J, Walker J, Houser P. Factors affecting remotely sensed snow water equivalent uncertainty. Remote Sensing Environ 2005;97:68–82.
[11] Drusch M. Initializing numerical weather prediction models with satellite derived surface soil moisture: data assimilation experiments with ECMWF's integrated forecast system. J Geophys Res 2007;112. doi:10.1029/2006JD007478.
[12] Ducharne A, Koster R, Suarez M, Stieglitz M, Kumar P. A catchment-based approach to modeling land surface processes in a general circulation model. 2. Parameter estimation and model demonstration. J Geophys Res 2000;105:24823–38.
[13] Ek M, Mitchell D, Lohmann KE, Wei H. Snowpack advances in the Noah land-surface model. In: Proc of the 18th conference on hydrology. Seattle (WA): American Meteorological Society; 2004.
[14] Ek M, Mitchell K, Yin L, Rogers P, Grunmann P, Koren V, et al. Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model. J Geophys Res 2003;108. doi:10.1029/2002JD003296.
[15] Foster J, Sun C, Walker J, Kelly R, Chang A, Dong J, et al. Quantify the uncertainty in passive microwave snow water equivalent observations. Remote Sensing Environ 2005;94:187–203.
[16] Gelb A. Applied optimal estimation. Cambridge (MA): MIT Press; 1974.
[17] Hall D, Riggs G, Valomonson V, DiGirolamo N, Bayr K. MODIS snow-cover products. Remote Sensing Environ 2002;83:181–94.
[18] Hill C, DeLuca C, Balaji V, Suarez M, da Silva A. The architecture of the earth system modeling framework. Comput Sci Eng 2004:6.
[19] Kelly R, Chang A. Development of a passive microwave global snow depth retrieval algorithm for special sensor microwave imager (SSM/I) and advanced microwave scanning radiometer-EOS (AMSR-E) data. Radio Sci 2003;38. doi:10.1029/2002RS002648.
[20] Koster RD, Suarez MJ, Ducharne A, Stieglitz M, Kumar P. A catchment-based approach to modeling land surface processes in a general circulation model. 1. Model structure. J Geophys Res 2000;105:24809–22.
[21] Kumar S, Peters-Lidard C, Eastman J, Tao W-K. An integrated high resolution hydrometeorological modeling testbed using LIS and WRF. Environ Model Software 2008;23:169–81.
[22] Kumar S, Peters-Lidard C, Tian T, Houser P, Geiger J, Olden S, et al. Land information system: an interoperable framework for high resolution land surface modeling. Environ Model Software 2006;21:1402–15.
[23] Lynch-Stieglitz M. The development and validation of a simple snow model for the GISS GCM. J Climate 1994;7:1842–55.
[24] Margulis SA, McLaughlin D, Entekhabi D, Dunne S. Land data assimilation and estimation of soil moisture using measurements from the southern great plains 1997 field experiment. Water Resour Res 2002;38:1299.
[25] McLaughlin D, O'Neill A, Derber J, Kamachi M. Opportunities for enhanced collaboration within the data assimilation community. Quart J Roy Meteorol Soc 2005;131:3683–93.
[26] McLaughlin M. An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. Adv Water Resour 2002;25:1275–86.
[27] Pfaendtner J, Bloom S, Lamich D, Seablom M, Sienkiewicz M, Stobie J, et al. Documentation of the goddard earth observing system (GEOS) data assimilation system – version 1. Technical report memo 104606, vol 4. NASA Goddard Space Flight Center; 1995.
[28] Pree W. Design patterns for object-oriented software development. Reading (MA): Addison-Wesley; 1995.
[29] Reichle R. Data assimilation methods in the earth sciences. Adv Water Resour 2008 [this issue].
[30] Reichle R, Crow W, Keppenne C. An adaptive ensemble Kalman filter for soil moisture data assimilation. Water Resour Res [in press].
[31] Reichle R, Koster R. Land data assimilation with the ensemble Kalman filter: assessing model error parameters using innovations. In: Hassanizadeh S, Schotting R, Gray W, Pinder G, editors. Developments in water science – computational methods in water resources, vol. 47. New York: Elsevier; 2002. p. 1387–94.
[32] Reichle R, Koster R. Bias reduction in short records of satellite soil moisture. Geophys Res Lett 2004;31:L19501.
[33] Reichle R, Koster R, Liu P, Mahanama S, Njoku E, 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). J Geophys Res – Atmos 2007;112. doi:10.1029/2006JD008033.
[34] Reichle RH, Koster RD. Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. J Hydrometeorol 2003;4:1229–42.
[35] Reichle RH, Koster RD. Global assimilation of satellite surface soil moisture retrievals into the NASA catchment land surface model. Geophys Res Lett 2005;32. doi:10.1029/2004GL021700.
[36] Reichle RH, Walker JP, Koster RD, Houser PR. Extended versus ensemble Kalman filtering for land data assimilation. J Hydrometeorol 2002;3:728–40.
[37] Richards L. Capillary conduction of liquids in porous media. Physics 1931;1:318–33.
[38] Rodell M, Houser PR. Updating a land surface model with MODIS derived snow cover. J Hydrometeorol 2004;5:1064–75.
[39] Slater A, Clark M. Snow data assimilation via an ensemble Kalman filter. J Hydrometeorol 2006;7:478–93.
[40] Sun C, Walker JP, Houser PR. A methodology for snow data assimilation in a land surface model. J Geophys Res 2004:109.
[41] Van Den Hurk BJJM, Jia L, Jacobs C, Menenti M, Li Z-L. Assimilation of land surface temperature data from ATSR in an NWP environment – a case study. Int J Remote Sensing 2002;23:5193–209.
[42] Walker J, Houser P. Hydrologic data assimilation. In: Aswathanarayana A, editor. Advances in water science technologies. The Netherlands: Balkema; 2005. p. 230.
[43] Walker J, Houser P, Reichle R. New technologies require advances in hydrologic data assimilation. Eos 2003;84:545.
[44] Walker JP, Houser PR. A methodology for initializing soil moisture in a global climate model: assimilation of near-surface soil moisture observations. J Geophys Res 2001;106:11761–74.
[45] Zhou Y, McLaughlin D, Entekhabi D. Assessing the performance of the ensemble Kalman filter for land surface data assimilation. Monthly Weather Rev 2006;134:2128–42.