Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

Geoscientific Model Development Discussions

Comparison of Different Sequential Assimilation Algorithms for Satellite-derived Leaf Area Index Using the Data Assimilation Research Testbed (lanai)

Xiao-Lu Ling [1,2,3], Cong-Bin Fu [1,2,*], Zong-Liang Yang [3,*], and Wei-Dong Guo [1,2]

[1]Institute for Climate and Global Change Research & School of Atmospheric Sciences, Nanjing University, Nanjing 210023, China
[2]Joint International Research Laboratory of Atmospheric and Earth System Sciences of Ministry of Education, Nanjing 210023, China
[3]Department of Geological Sciences, John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, TX 78705, USA

Correspondence to: Cong-Bin Fu (fcb@nju.edu.cn); Zong-Liang Yang (liang@jsg.utexas.edu)

Abstract. The leaf area index (LAI) is a crucial parameter for understanding the exchanges of momentum, carbon, energy, and water between terrestrial ecosystems and the atmosphere. To improve the ability to simulate land surface water and energy balances, the Data Assimilation Research Testbed (DART) has been successfully coupled to the Community Land Model (CLM) by assimilating global remotely sensed LAI data with explicit carbon and nitrogen components (CLM4CN). The purpose of this paper is to determine the best algorithm for LAI assimilation. Within this framework, four sequential assimilation algorithms, i.e., the Kalman Filter (KF), the Ensemble Kalman Filter (EnKF), the Ensemble Adjust Kalman Filter (EAKF), and the Particle Filter (PF), are applied, thoroughly analyzed and compared. The results show that assimilating remotely sensed LAI data into the CLM4CN is an effective method for improving model performance. In detail, the assimilation accuracies of the ensemble filter algorithms (EnKF and EAKF) are better than that of the KF algorithm because the KF is based on the linear model error assumption. The PF algorithm performs worse than the EAKF and EnKF algorithms because of the gradually reduced acceptance of observations with assimilation steps. In other words, the contribution of the observations to the posterior probability during the assimilation process is reduced. The EAKF algorithm is the best method because the matrix is adjusted at each time step during the assimilation procedure.

1 Introduction

Land surface processes play an important role in the earth system because all the physical, biochemical, and ecological processes occurring in the soil, vegetation, and hydrosphere influence the mass and energy exchanges during land-atmosphere interactions (Bonan, 1995; Pitman, 2003; Pitman et al., 2009, 2012). The leaf area index (LAI) is a key biophysical parameter of vegetation in land surface models (LSMs) and influences their simulation performance. Therefore, high-quality, spatially and temporally continuous LAI inputs are extremely important (Bonan et al., 1992; Li et al., 2015).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

Real-time monitoring of LAI on a large scale is a worldwide problem. The lack of spatial representativeness caused by the sparse distribution of conventional observations makes it difficult to achieve a global observational LAI dataset. Remote sensing can provide global data with high spatial and temporal resolutions, but the inversion accuracy is associated with different plant functional types (PFTs) and vegetation fractions. Furthermore, although advanced LSMs (e.g., the Community Land Model version 4, CLM4) can predict LAI variation, the model performance is greatly affected by the model structure or the initial/forcing/boundary conditions of the input (Dai et al., 2003; Luo et al., 2003; Levis et al., 2004). The Land Data Assimilation System (LDAS), through optimally combining both dynamical and physical mechanisms with real-time observations, can effectively reduce the estimation uncertainties caused by spatially and temporally sparse observations and poor observed data accuracy (Kalnay, 2003).

A complete LDAS is mainly composed of forcing, initial and boundary datasets, parameterization sets, dynamical models as physical constraints, assimilation algorithms, observational data and target output. The widely acknowledged LDASs include the North LDAS (NLDAS, Mitchell et al., 2004; NLDAS-2, Luo et al., 2003; Xia et al., 2012), Global LDAS (GLDAS, Rodell et al., 2004), European LDAS (ELDAS, Jacobs et al., 2008), West China LDAS (WCLDAS, Huang and Li, 2004), and Canadian LDAS (CaLDAS, Carrera et al., 2015).

As a link between observations and dynamic model states, mathematical algorithms play an important role in calculating the increments and adjusting the state vector during assimilation (Kalnay et al., 2007). The two basic data assimilation algorithms are the variational method based on optimal control theory (Dimet and Talagrand, 1986) and sequential algorithms based on the Kalman Filter (KF). To date, the most popular variational algorithms widely utilized in LDAS (Evensen, 2003) are three-dimensional variation (3DVAR, Zhang et al., 2011) and four-dimensional variation (4DVAR) algorithms. For 3DVAR algorithms, the observation operator can be nonlinear, but the background variance is isotropic and does not change with time. The 4DVAR algorithms can employ flow-dependent forecast error covariance but cost more to implement and maintain. The state quantity is estimated by using all possible observations and the statistical characteristics of dynamic model simulations and observations to minimize the estimated error. The KF is the theoretical basis of the sequential data assimilation algorithm. Because the KF algorithm is based on the linear model error assumption, many new sequential algorithms have been proposed. For example, the Extended Kalman Filter (EKF) was developed to meet the need for a nonlinear observation operator, but the tangent operator needs to be developed (Kalnay, 2003). Based on the Monte Carlo method and focused on the nonlinear operator, the Ensemble Kalman Filter (EnKF) was developed (Evensen, 1994) and was first used in the study of atmospheric science (Houtekamer and Mitchell, 1998). Since then, the EnKF has been widely applied for the assimilation of ocean, land surface and atmospheric data (Houtekamer et al., 2005; Evensen, 2007).

Many previous studies focusing on the comparison of variational and sequential algorithms have been conducted to determine the optimal assimilation method (Han and Li, 2008). Wu et al. (2011) systematically compared EnKF and 3DVAR/4DVAR algorithms and found that the EnKF algorithm was better than the 3DVAR method and the same as the 4DVAR method. For this reason, the application of the EnKF algorithm has been expanded quickly, and many other forms of the EnKF method have been

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

developed, such as the Dual EnKF (Li et al., 2014), Ensemble Square Root Filter (EnSRF) (Whitaker and Hamill, 2002), and Ensemble Adjust Kalman Filter (EAKF, Anderson, 2001). At the same time, combinations of variational algorithms and sequential algorithms have also been developed. For example, the maximum likelihood ensemble filter (MLEF, Zupanski, 2005) was developed to find the optimal solution by minimizing the target function for the nonlinear observation operator. The combination of 3DVAR and PF algorithms also showed better results than either single algorithm (Leng and Song, 2013). Furthermore, hybrid variational-ensemble data assimilation methods, i.e., the 4DEnKF (Hunt et al., 2004; Fertig et al., 2007; Zhang et al., 2009) and the DrEnKF (Wan et al., 2009), have been developed at NCEP and applied to improve model predictions (Whitaker et al., 2008).

Recent studies focusing on assimilation in terrestrial systems have tended to add multiple phenological observations to constrain and predict biome variables and further improve model performance (Knyazikhin et al., 1998; Xiao et al., 2009; Viskari et al., 2015). Assimilating satellite-derived LAI and soil moisture products using the Simplified Extended Kalman Filter (SEKF) or EAKF has a strong impact on the LAI data. Furthermore, the abilities to simulate river discharge, land evapotranspiration, and gross primary production have been improved in Europe (Barbu et al., 2011; Albergel et al., 2017). To date, such studies have been conducted using a single sequential algorithm at a single site or on regional scales (Montzka et al., 2012).

The Data Assimilation Research Testbed (DART) is an open source community facility and includes several different types of KF algorithms (Anderson et al., 2009). It has been coupled to many high-order models and observations for ocean, atmosphere, land surface, and chemical constituents. For example, DART has been coupled with CLM4 (DART/CLM4) to improve snow, LAI and soil moisture predictions (Zhang et al., 2014; Kwon et al., 2016; Zhao et al., 2016).

Utilizing coupled DART/CLM4, the Global Land Surface Satellite LAI (GLASS LAI) data are assimilated into the Community Land Model with carbon and nitrogen components (CLM4CN) in the present study to explore the optimal assimilation algorithm for model performance. The experimental design and different assimilation algorithms are described in Sect. 2. Section 3 describes the optimal algorithm for LAI assimilation, and the proportion of observations is discussed in Sect. 4. Conclusions and discussions are given in Sect. 5.

2 Data and Methodology

A complete LDAS is mainly composed of forcing/initial/boundary datasets, parameterization sets, dynamical LSMs, assimilation algorithms, observational data and target output. LSMs play an important role in the LDAS because they can add physical constraints to the control variables during assimilation. In addition, the simulation ability of LSMs can directly affect the output because they provide the associated uncertainty for assimilation.

2.1 CLM4CN

Developed by the National Center for Atmospheric Research (NCAR), the Community Land Model (CLM) can simulate energy, momentum and water exchanges between the land surface and the overlying

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

atmosphere at each computational grid. The CLM is designed mainly for coupling with the atmospheric numerical model and providing the surface albedo (direct and scattered light within the visible and infrared bands), upward longwave radiation, sensible heat flux, latent heat flux, water vapor flux, and east-to-west and south-to-north surface stress needed by the atmospheric model. These parameters are controlled by many ecological and hydrological processes. The model can also simulate leaf phenology and physiological processes, as well as water circulation through plant pores. Ecological differences between vegetation types and thermal and hydrological differences between different soil types are also considered. Each grid cell can be covered by several different land use types. Each cell contains several land units, each land unit contains a different number of soil and snow cylindrical blocks, and each cylindrical block may contain several types of vegetation functions. The CLM employs 10 soil layers to resolve soil moisture and temperature dynamics and uses PFTs to represent subgrid vegetation heterogeneity (Oleson et al., 2010).

There are two ways to update LAI in CLM4. The LAI is treated as a diagnostic variable that is linearly interpolated from a 30-year averaged satellite dataset, and there is no annual LAI variation for CLM4 with Satellite Phenology (CLM4SP) (Lawrence and Chase, 2007). For CLM4CN, the prognostic LAI is calculated by the leaf carbon pool and an assumed vertical gradient of specific leaf area (SLA) (Thornton and Zimmermann, 2007). Carbon and nitrogen are obtained by plant storage pools in one growing season and then retained and distributed in the subsequent year. All carbon and nitrogen state variables in vegetation, litter, and soil organic matter (SOM) are prognostic based on the prescribed vegetation phenology. The CLM4CN offline mode with prescribed meteorological forcing is used in this study.

2.2 DART

DART is developed and maintained by the Data Assimilation Research Section (DAReS) at NCAR. The purpose of DART is to provide a flexible tool for data assimilation (DA), and it has been coupled with many 'high-order' models. As a software environment, DART makes it easy to explore a variety of data assimilation methods and observations with different numerical models. The DART system includes several different types of sequential algorithms, which are selected at runtime by a namelist setting. The detailed settings for DART can be found at https://www.image.ucar.edu/DAReS/DART/.

Currently, the coupled DART/CLM4 model has produced many reanalysis data for snow and soil moisture. It has been found that snow DA can improve temperature predictions, especially over the Tibetan Plateau, implying great implications for future land DA and seasonal climate prediction studies (Lin et al., 2016). Furthermore, the coupled DART/CLM framework would be employed to assimilate other variables, such as LAI, from various satellite sources and ground observations (i.e., truly multimission, multiplatform, multisensor, multisource, and multiscale). Ultimately, this would allow earth system models to be constrained by all types of observations to improve model performance for seasonal and decadal prediction skills.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

2.3 Sequential Assimilation Algorithms

2.3.1 Kalman Filter (KF)

As a theoretical basis of the sequential DA method, the aim of the KF is to achieve the optimal analysis field based on the variance minimization principle (Kalman, 1960). The main KF procedure is as follows: (1) During the forecast stage, the dynamical model produces the forecast variables and associated uncertainties at the next observation time step, and (2) at the analysis stage, updated analyzed variables and associated uncertainties are determined based on the previous information on the uncertainties for each ensemble member.

Compared with the statistical optimal interpolation algorithm, the predicted error changes with the dynamical model for the KF method. Furthermore, the KF method is more easily realized because the adjoint matrix is not needed. However, the KF method is based on the assumption of a Gaussian relation between the variables in the joint stage space prior distribution.

2.3.2 Ensemble Kalman Filter (EnKF)

The KF algorithm has not been widely used because of computing limitations and the linear model error assumption. The EnKF was proposed based on a Monte Carlo approximation, for which the background error covariance is approximated using an ensemble of forecasts (Evensen, 1994). The EnKF algorithm can be utilized for nonlinear systems and can also reduce the computing requirement of DA (Evensen, 2003; 2007).

The EnKF procedure is divided into two stages: prediction and analysis. (1) In the prediction stage, the ensemble forecast field is generated from the ensemble initial condition, and the error covariance matrix of the ensemble forecast is calculated. (2) In the analysis stage, the simulation of each member of the ensemble is updated using the covariance matrix of observation vector error and state vector error.

2.3.3 Ensemble Adjust Kalman Filter (EAKF)

Although the forms of expression are different, the proposed EnSRF (Whitaker et al., 2002) and EAKF (Anderson, 2001) are the same algorithm.

The difference between the EAKF and the traditional EnKF lies in the adjustment of the gain matrix to avoid filtering the divergence problem by increasing the premise of the analysis error covariance (Anderson, 2003, 2007; Wang et al., 2007). In the EAKF algorithm, ensemble observation members are calculated by the observation operator, and the increment of each observation member is calculated as ΔY_i.

The increment ΔX_{ij} for each ensemble sample of each state variable in terms of ΔY_i can then be calculated as follows:

$$\Delta X_{ij} = \frac{\sigma_{jo}^{p}}{\sigma_{o}^{p}} \Delta Y_i. \quad (1)$$

where i indicates the ensemble member, j is the state vector member, σ_{jo}^{p} is the prior covariance of state vector and observation, and σ_{o}^{p} is the prior variance of observation.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

2.3.4 Particle Filter (PF)

The Particle Filter (PF) is also a sequential Monte Carlo method, which is based on the Bayesian sequential importance sampling method (SIS). The PF algorithm finds a set of random samples in the state space to approximate the probability density function and then replaces the integral operation with the sample mean to obtain the process of minimum variance distribution of the state (Moradkhani et al., 2005). The procedure of the PF algorithm can also be divided into two frameworks: forecast and analysis.

If there are enough observations, the posterior density at k can be approximated as

$$p(X_k^a|Y_{1:k}) \approx \sum_{i=1}^{N} w_{i,k}\,\delta(X_k^a - X_{i,k}^a). \quad (2)$$

in which $\delta(*)$ is the Dirac Function and $\sum_{i=1}^{N} w_{i,k} = 1$.

Unlike the EnKF algorithm, the PF method takes into account the weights of different particles and can be better applied to nonlinear systems. However, in association with the DA, there are a limited number of particles with large weights, and too many computing resources are distributed to particles with weights of approximately 0. This situation is called particle degradation (Doucet et al., 2000). Effective methods to solve this issue include resampling or selecting more reasonable importance functions.

2.4 Ensemble Meteorological Forcing

The ensemble initial conditions and background error (Hu et al., 2014) are produced from ensemble analysis products generated by running DART and the Community Atmosphere Model (CAM4) (Raeder et al., 2012). DART/CAM4 produced 80 atmospheric forcing datasets with 6-hour time intervals for the period of 1998-2010. These ensemble meteorological data have been widely employed in DA for ocean, snow, soil moisture, and many other related studies (Danabasoglu et al., 2012). By considering computational cost and filter performance, 40 members among the ensemble forcing datasets are chosen to drive the CLM4CN.

2.5 Experimental Design

Table 1. Experimental design for LAI assimilation using DART/CLM4CN.

Experiment	Assimilated variables	Updated variables	Assimilation algorithm	Accept all observation
Algorithms	GLASS LAI	LAI, Leaf C, Leaf N	EAKF, EnKF, KF, PF	YES
Observation Proportion	GLASS LAI	LAI, Leaf C, Leaf N	EAKF, EnKF, KF, PF	NO

To determine the optimal assimilation algorithm, four experiments corresponding to the KF, EnKF, EAKF and PF methods are designed and shown in Table 1. During assimilation, CLM stops and writes restart and history files at a frequency of 8 days. If there is available observational GLASS LAI data, they are assimilated into the CLM4CN. DART extract state vector, the increments are calculated by filtering at each time step, and the LAI, leaf carbon (Leaf C) and leaf nitrogen (Leaf N) are updated. The

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

adjusted DART state vector is resent to the CLM restart files as a new initial condition for the next time step.

3 The Optimal Algorithm for DART/CLM4CN

The spatial distributions of global LAI in 2002 for (a) observations in July, (b) ensemble mean of simulations in July, (c) observations in November, and (d) ensemble mean of simulations in November are shown in Fig. 1. The observations in Fig. 1 are from the Moderate Resolution Imaging Spectroradiometer (MODIS) LAI dataset with a spatial resolution of 1.0 latitude by 1.0 longitude. There are two latitudinal belts of high LAI values located in the tropics and at 50-65°N. These two regions are mainly dominated by evergreen broadleaf forests and boreal forests, respectively. Because of the presence of deserts, plateaus and bare ground, the LAI is low in western North America, western Australia, southern Africa, and southern South America, where shrubs and/or grass are dominant. Globally, the CLM4CN can simulate the LAI distribution characteristics, except that it systematically overestimates LAI, especially at low latitudes. There are 3 high-LAI regions located in the tropics: the Amazon, central Africa, and some islands in Southeast Asia. The global LAI is lower in November than in July. The LAI values in the high latitudes of the northern hemisphere are higher in July than in November because November is not the growing season for most of the vegetation in the northern hemisphere.

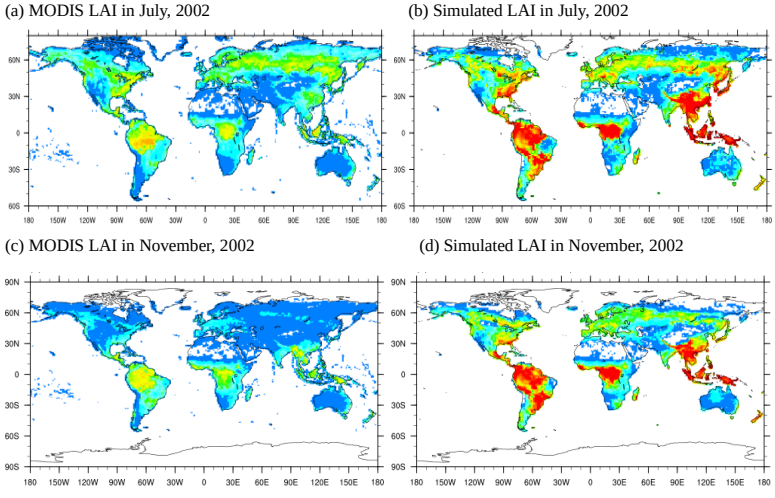

Figure 1: Spatial distributions of global LAI values in 2002 for (a) observations in July, (b) ensemble mean of simulations in July, (c) observations in November, and (d) ensemble mean of simulations in November.

The differences between global LAI from observations and that from assimilation experiments in July 2002 with the methods of (a) EAKF, (b) EnKF, (c) KF and (d) PF are shown in Fig. 2. Globally, assimilation results with the methods of EAKF and EnKF are larger in lower-latitude regions and higher-latitude regions in the Northern Hemisphere. For the EAKF and EnKF algorithms, large negative biases are located in the Amazon region, central Africa, and northeastern China, which are dominated by BET

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

tropical, NET boreal forests and mixed forest types, respectively. The LAI values from the assimilation experiment are always higher in the middle- and high-latitude regions, especially in western North America, the northern Amazon, northwestern China, and western Australia, where open shrublands and grasslands are dominant. The LAI values from the assimilation experiments with the KF and PF algorithms are highly overestimated compared to the observations in the northern and eastern Amazon, central Africa, southern Eurasia, and Southeast Asia. In addition, the LAI values obtained by the EAKF method are more continuous than those obtained by the EnKF method and more consistent with the observations in central South America and central Africa. Notably, the correction of overestimated LAI is significantly better than that of underestimated LAI, which is mainly attributed to the high dispersion of LAI in those regions. In other words, high dispersion is beneficial to assimilation.

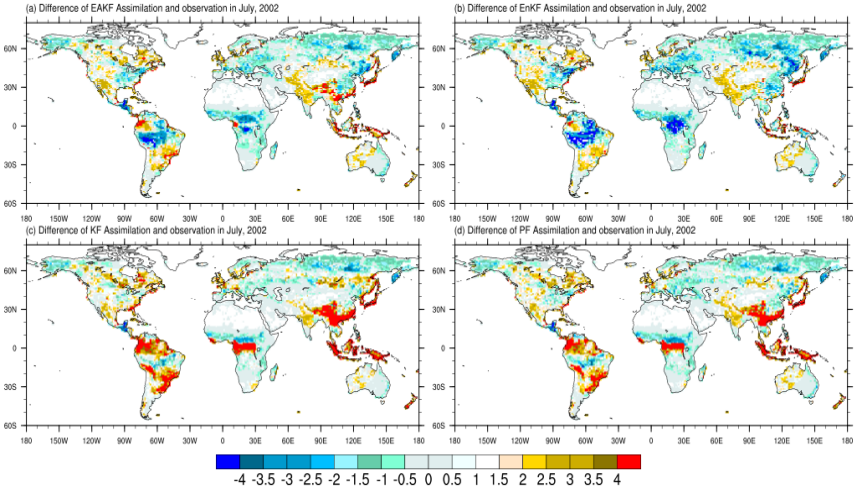

Figure 2: Differences between global LAI from assimilation experiments and that from observations in July 2002 with the methods of (a) EAKF, (b) EnKF, (c) KF and (d) PF.

The differences between global observed LAI values and assimilated LAI values with the methods of (a) EAKF, (b) EnKF, (c) KF and (d) PF in November 2002 are also shown in Fig. 3. Similar to Fig. 2, the results indicate that the EAKF and EnKF assimilation algorithms are better than the KF and PF algorithms. In detail, the EAKF algorithm is better than the EnKF method in November, especially in the Amazon, central Africa, and southern Eurasia. The biases of assimilated LAI relative to the observed LAI are higher in November in the 20-65°N region, which may be because vegetation during this period in the Northern Hemisphere is not lush. In western Australia and central Eurasia, the improvement of the underestimation in November is not as significant as that in July, which indicates that the system has a limited capability to simulate the vegetation process, especially for open shrubland and grassland.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

Geoscientific Model Development Discussions

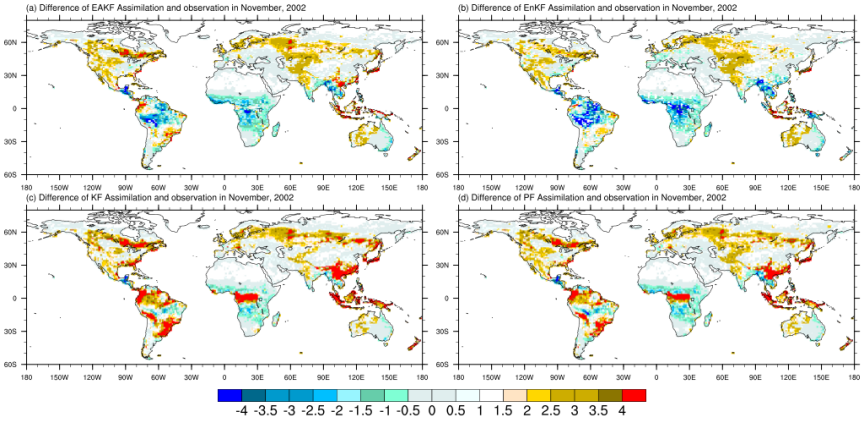

Figure 3: Same as Fig. 2, but for November.

The mean LAI globally and the LAI in five latitudinal bands were chosen for analysis in this study. The five bands are boreal (45-65°N), northern temperate (23-45°N), northern equatorial (0-23°N), southern equatorial (0-23°S), and southern temperate (23-90°S). Figure 4 presents the root mean square errors (RMSEs) of the ensemble means of simulation/assimilation versus observations for (a) global, (b) boreal, (c) northern temperate, (d) northern equatorial, (e) southern equatorial, and (f) southern temperate. Generally, although they all feature similar variation pattern characteristics, the RMSEs of all the assimilation datasets relative to the observations are less than those of the simulation, indicating that all four assimilation algorithms can improve the LAI simulation. The highest RMSE relative to the observations is associated with the simulation, followed by the assimilation datasets from the KF and PF algorithms, and the RMSEs are lowest for the EAKF and EnKF methods. During the growing season, the RMSEs of LAI reach their largest values, especially for the regions in the middle and high latitudes of the Northern Hemisphere and high latitudes of the Southern Hemisphere. In the low-latitude region covered by evergreen or deciduous broadleaf forests, the RMSE does not present an obvious annual change. Because the PF assimilation is heavily dependent on the weights of certain particles and to some degree ignores the importance of observed LAI data, the phenomenon of particle degradation occurs during the assimilation.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

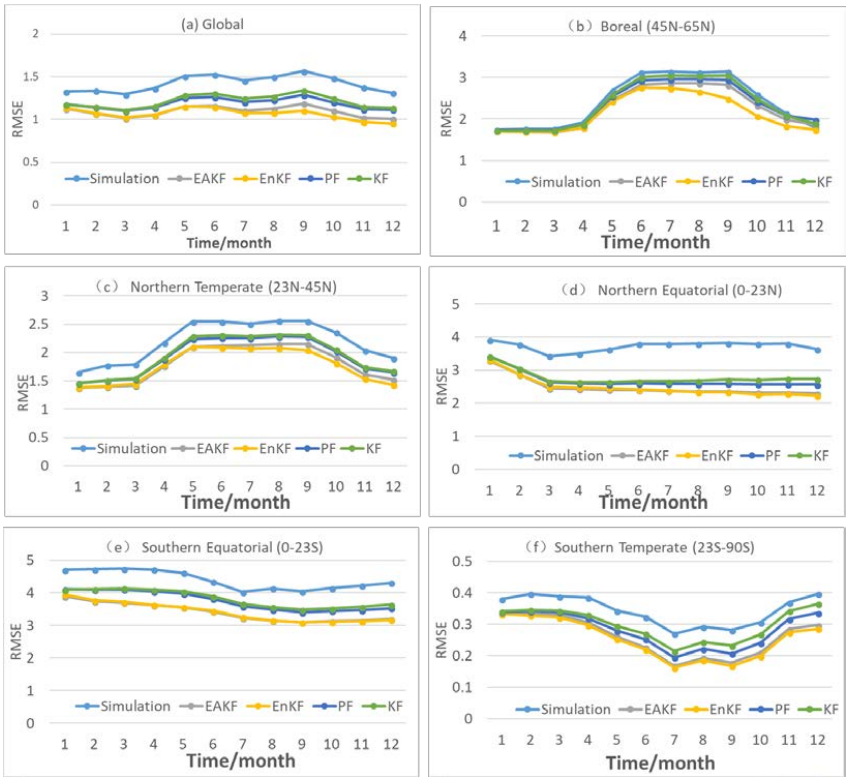

Figure 4: RMSEs of ensemble means of simulation/assimilation versus observations for (a) global, (b) boreal (45-65°N), (c) northern temperate (23-45°N), (d) northern equatorial (0-23°N), (e) southern equatorial (0-23°S), and (f) southern temperate (23-90°S).

Figure 5 shows the globally averaged RMSEs of simulation/assimilation. The RMSEs of assimilation are lower than those of simulation, implying that assimilating remotely sensed LAI data into the CLM4CN is an effective method for improving the model performance. The RMSEs of assimilation results using the algorithms of EAKF and EnKF are much lower than the KF and PF methods, indicating their better performance in assimilation. The lowest RMSE appears for the assimilation result with the EAKF method, indicating that the EAKF is the best algorithm.

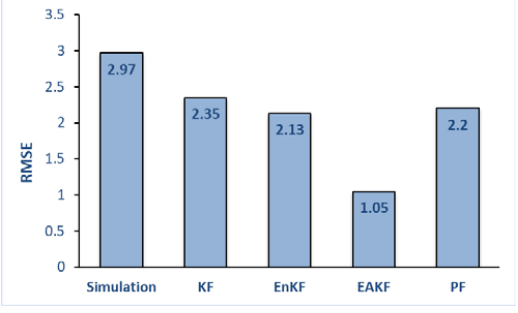

Figure 5: Globally averaged RMSEs for the simulation/assimilation results.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

Geoscientific Model Development Discussions

The background/analysis departures are calculated as (1) innovations, which are the differences between the assimilated observations and model background, and (2) residuals, which are the differences between the assimilated observations and analysis (Barbu et al., 2011). It was concluded that the LDAS system is working well based on the condition that the residuals are reduced compared to the innovations (Albergel et al., 2017). Figure 6 shows the histograms of innovation and residuals of LAI globally and for all subregions during July 2002. Generally, the distribution characteristics of both innovations and residuals are identical for the algorithms of KF and PF, which means that these two algorithms are not very efficient for LAI assimilation. The distribution of residuals is more centered on 0 than that of the innovations for the EAKF and EnKF algorithms, especially for the EAKF algorithm. The innovations dominantly exhibit a large negative bias, indicating that the model always highly overestimates LAI. The residuals can improve this overestimation situation, especially for the EAKF algorithm. The analysis departures show an abnormal high value in the range of -3 to -2 for the boreal and southern equatorial subregions for the EnKF algorithm but not for the EAKF algorithm, implying that the EAKF algorithm is the optimal algorithm for LAI assimilation.

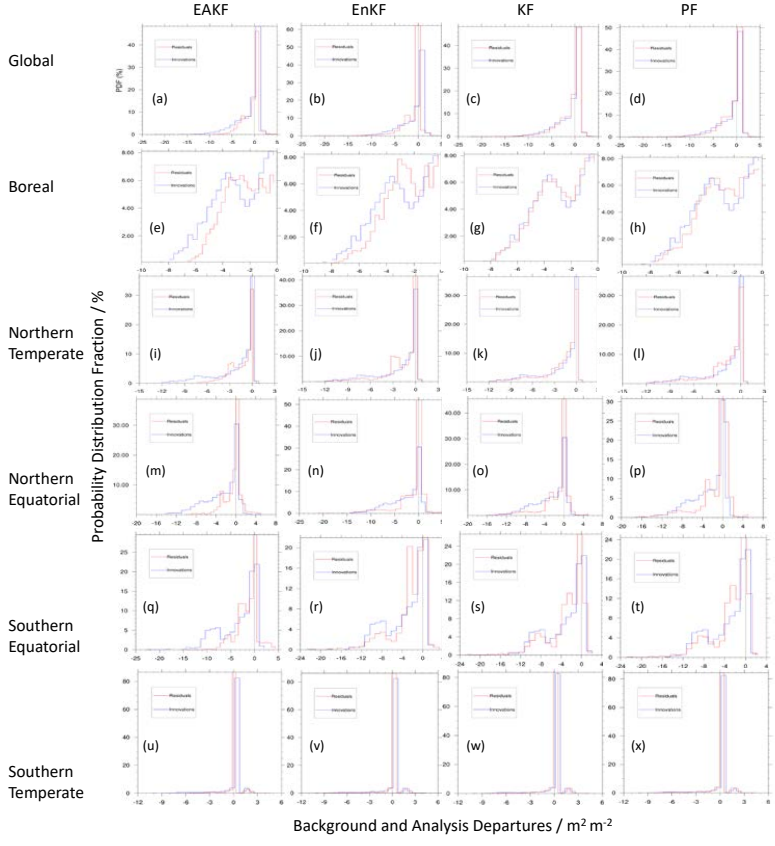

Figure 6: The histograms of innovation and residuals of LAI globally and for all subregions during July 2002. (a-d) Global; (e-h) boreal; (i-l) northern temperate; (m-p) northern equatorial; (q-t) southern equatorial; (u-x) southern temperate

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

Geoscientific Model Development Discussions

4 Effective Observational Proportion

The assimilation results depend not only on the algorithm but also on the observations. This not only requires a sufficiently strong degree of discretization for ensemble simulations but also requires the observational variables to be sufficiently trustworthy. In this section, the proportion of LAI observations that can be accepted for the four algorithms is discussed.

To explain the relationship between assimilation algorithms and observation rejection, Fig. 7 displays the proportion of accepted LAI observations for the four algorithms in the zonal regions. In general, the EnKF and EAKF methods accepted many more observational LAI observations than the PF and KF methods. In the low-latitude regions, the proportion of accepted LAI observations is approximately 75%, which is lower than in the high-latitude regions. This may be because the broadleaf forest in tropical regions can grow unrestrictedly in the model, producing LAI values that are much higher than the observations. At the very beginning of assimilation, DART rejects the largest proportion of LAI observations in the southern equatorial, northern equatorial, and northern temperate zones due to large biases between the simulation and the observations. Over time, the rejection proportion gradually decreases for the northern equatorial, southern equatorial and southern temperate. As ensemble-analyzed LAI values tend to relatively fixed, the rejection proportion increases over regions with small LAI amplitudes, such as the northern temperate and boreal region. From May to September in the boreal region and from April to September in the northern temperate region, the proportion of accepted LAI is much smaller than in the other regions. These two periods are also when the model simulation presents an obvious discrete characteristic. This experiment illustrates the utility of the spin-up process for ensemble initial conditions.

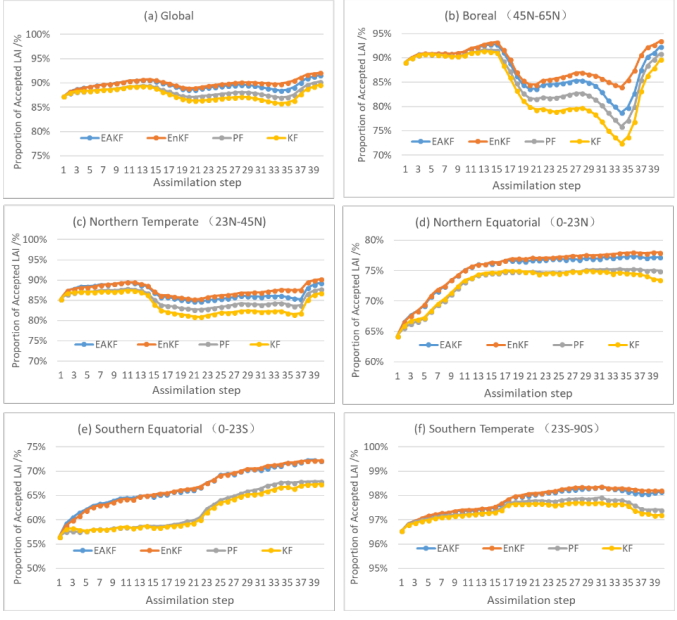

Figure 7: The proportion of accepted LAI observations for the four algorithms in the zonal regions.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

The difference between globally assimilated and observed LAIs with the methods of EAKF (with rejection) in (a) July and (b) November are shown in Fig. 8 to illustrate the role of observation proportion. It can be concluded that when accepting all the observations, the assimilation results seem to be better than when some observations are rejected during assimilation. Large biases occur in the Amazon, central Africa, southern Eurasia, and the boreal region, where the LAI is overestimated in the model. Furthermore, the KF and PF algorithms gradually reduce the acceptance of observations as assimilation progresses, which may partially explain their worse performance than the EnKF and EAKF algorithms (see Fig. 5).

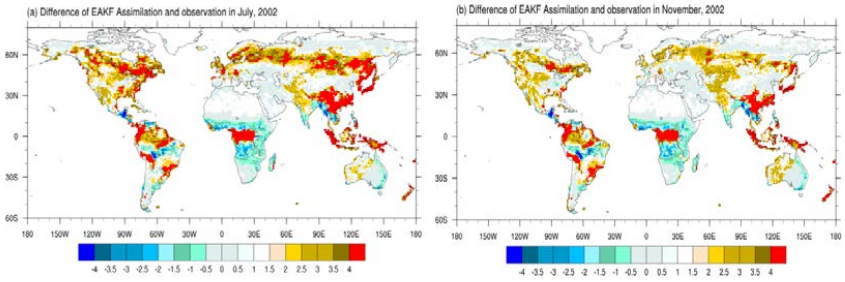

Figure 8: Differences between globally assimilated and observed LAIs for the methods of EAKF in (a) July and (b) November.

During assimilation, the assimilated observations (GLASS LAI) are always treated as "true" values. The question thus becomes how do the true values influence the assimilation results? Figure 9 shows the RMSEs of simulation experiments without/with rejection (EAKF_noreject / EAKF_reject) over the (a) global, (b) boreal, (c) northern temperate, (d) northern equatorial, (e) southern equatorial, and (f) southern temperate regions. In the EAKF_reject experimental design, if the observed LAI is three times larger than the bias between the simulation and the observations, the observation would be rejected by DART, while in the EAKF_noreject experiment, all observed LAIs are assimilated. Generally, RMSEs for both simulation and assimilation present obvious annual variations, with RMSEs reaching their maximum values in the season with considerable vegetation growth over a large area. The RMSE of assimilation is far less than that of the simulation, although their characteristic variation patterns are similar. This demonstrates the effectiveness of assimilation for improving model simulation. The RMSE relative to the observations was highest for the simulation, followed by the EAKF_reject experiment, and was lowest for the EAKF_noreject experiment. The RMSEs are large during the growing season, when LAI values are also high in the boreal and northern temperate regions. During assimilation, when accepting all the observations, the RMSE is smaller than that when rejecting some observations.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

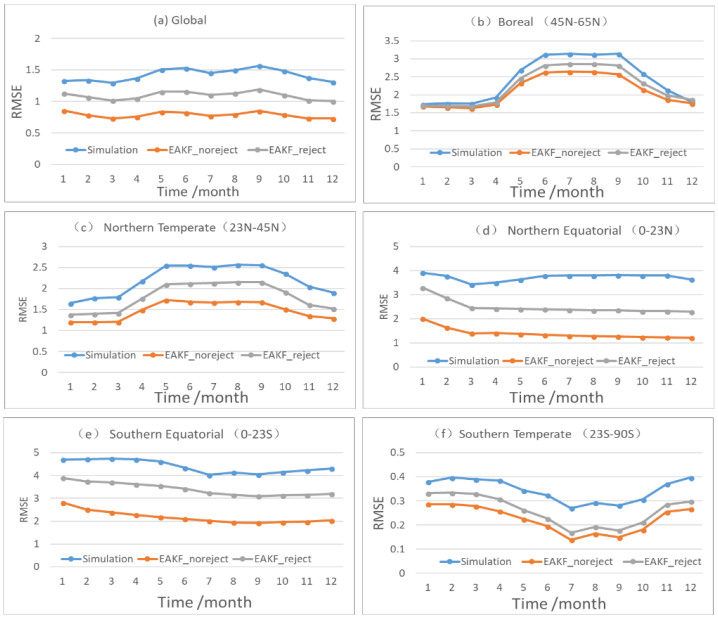

Figure 9: RMSEs of simulation experiments without/with rejection (EAKF_noreject and EAKF_reject) for the (a) global, (b) boreal (45-65°N), (c) northern temperate (23-45°N), (d) northern equatorial (0-23°N), (e) southern equatorial (0-23°S), and (f) southern temperate (23-90°S) regions.

5 Conclusions and Discussion

The Community Land Model version 4 with prognostic carbon-nitrogen components (CLM4CN) is coupled with the Data Assimilation Research Testbed (DART) to determine the optimal assimilation algorithm for leaf area index (LAI). Four different sequential methods, i.e., the Kalman Filter (KF), Ensemble Kalman Filter (EnKF), Ensemble Adjust Kalman Filter (EAKF), and Particle Filter (PF), are discussed in this paper.

The results show that assimilating remotely sensed LAI into the CLM4CN is an effective method for improving model performance. Globally speaking, the EAKF and EnKF assimilation algorithms are better than the KF and PF assimilation algorithms. The LAI obtained by the EAKF method is more continuous than that obtained by the EnKF method and more consistent with observations in central South American and central Africa, whereas the deviation in the EnKF method can be from -4 m^2 m^{-2} to 4 m^2 m^{-2}. Furthermore, the assimilation shows better performance in the vegetation growing season. The lowest root mean square error (RMSE) is associated with the EAKF algorithm, suggesting that the EAKF algorithm is the best and has a robust performance.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-232
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 3 January 2019

The proportion of observations accepted by the model is another topic of this research. The proportion of accepted LAI observations is 10-20% in the low latitudes lower than in the high latitudes because of large biases between the assimilation and the observations. When all the observations are accepted, the RMSE of the results is smaller than that when some observations are rejected.

5

Code availability. The Community Land Model version 4.0 with carbon and nitrogen Components (CLM4CN) is a part of the Community Earth System Model version 1.1.1 (CESM1.1.1) developed by the National Center for Atmospheric Research (NCAR). The CESM code can be downloaded from http://www.cesm.ucar.edu/index.html. Developed and maintained by the Data Assimilation Research Section (DAReS) at NCAR, Data Assimilation Research Testbed (DART version lanai) can be downloaded from https://www.image.ucar.edu/DAReS/DART/.

Author contributions. All of the authors participated in the development of the paper's findings and recommendations.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgments. This work was jointly supported in part by the National Natural Science Foundation of China (2016YFA0600303 and 2017YFA0604304) and the Jiangsu Collaborative Innovation Center for Climate Change. Kevin Raeder (raeder@ucar.edu) is thanked for providing the DART/CAM4 reanalysis as ensemble meteorological forcing. Tim Hoar, Long Zhao and Yongfei Zhang are thanked for part of the coding and coupling with DART/CLM4CN.

References

Albergel, C., Munier, S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., Gelati, E., Dorigo, W., Faroux, S., Meurey, C., Moigne, P. L., Decharme, B., Mahfouf, J. F., and Calvet, J. C.: Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area, Geosci. Model Develop., 10, 3889-3912, https://doi.org/10.5194/gmd-10-3889-2017, 2017.

Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, Mon. Wea. Rev., 129, 2884-2903, 2001.

Anderson, J. L.: A local least squares framework for ensemble filtering, Mon. Wea. Rev., 131, 634-642, 2003.

Anderson, J. L.: An adaptive covariance inflation error correction algorithm for ensemble filters, Tellus, 59(2), 210-224, http://doi.org/10.1111/j.1600-0870.2006.00216.x, 2007.

Anderson, J. L., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R. and Arellano, A.: The data assimilation research testbed: A community facility, Bull. Am. Meteorol. Soc., 90(9), 1283-1296, https://doi.org/10.1175/2009BAMS2618.1, 2009.

Barbu, A. L., Calvet, J.-C., Mahfouf, J.-F., Albergel, C., and Lafont, S.: Assimilation of Soil Wetness Index and Leaf Area Index into the ISBA-A-gs land surface model : grassland case study, Biogeosciences, 8, 1971-1986, https://doi.org/10.5194/bg-8-1971-2011, 2011.

Bonan, G. B.: Land atmospheric interactions for climate system models: Coupling biophysical, biogeochemical and ecosystem dynamical processes, Remote Sens. Environ., 51, 57-73, 1995.

Bonan, G.B., Pollard, D., Thompson, S.L.: Effects of boreal forest vegetation on global climate, Nature, 359, 716-718, http://doi.org/10.1038/359716a0, 1992.

Carrera, M. L., Belair, S., and Bilodeau, B.: The Canadian Land Data Assimilation System (CaLDAS): Description and Synthetic Evaluation Study, J. Hydrometeo., 16, 1293-1314, http://doi.org/10.1175/JHM-D-14-0089.1, 2015.

Dai, Y. J., Zeng, X. B., and Dickinson, R. E.: The common land model (CLM), Bull. Amer. Meteor. Soc., 84, 1013-1023, http://doi.org/10.1175/BAMS-84-8-1013, 2003.

Danabasoglu, G., Bates, S., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., Peacock, S., and Yeager, S. G.: The CCSM4 Ocean Component, J. Clim., 25, 1361-1389, http://doi.org/10.1175/JCLI-D-11-00091.1, 2012.

Dimet F. X. L., and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, Tellus, 38A, 97-110, https://doi.org/10.3402/tellusa.v38i2.11706, 1986.

Doucet, A., Godsill, S., and Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering, Stat. Comput., 10, 197-208, 2000.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte-carlo methods to forecast error statistics, J. Geophys. Res. Oceans, 99(C5), 10143-10162, 1994.

Evensen, G.: The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation, Ocean Dyn., 53, 343-367, http://doi.org/10.1007/s10236-003-0036-9, 2003.

Evensen, G.: Data Assimilation, the Ensemble Kalman Filter, Springer, pp 279, 2007.

Fertig, E. J., Harlim, J., and Hunt, B. R.: A comparative study of 4D-VAR and a 4D ensemble filter: Perfect model simulations with Lorenz-96, Tellus, 59A, 96-100, http://doi.org/10.1111/j.1600-0870.2006.00205.x, 2007.

Han, X. J., and Li, X.: An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation, Remote Sens. Environ., 112(4), 1434-1449, http://doi.org/10.1016/j.rse.2007.07.008, 2008.

Houtekamer, P. L., and Mitchell, H.: Data assimilation using an ensemble Kalman filter technique, Mon. Wea. Rev., 126(3), 796-811, 1998.

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B.: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, Mon. Wea. Rev., 133, 604-620, 2005.

Huang, C. L., and Li, X.: A Review of Land Data Assimilation System, Remote Sens. Techn. Appl., 19, 424-024.

Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J., and Sauer, T.: Four-dimensional ensemble Kalman filtering, Tellus, 56A, 273-277, 2004.

Hu, S. J., Qiu, C. Y., Zhang, L. Y., Huang, Q. C., Yu, H. P., and Chou, J. F.: An approach to estimating and extrapolating model error based on inverse problem methods: towards accurate numerical weather prediction, Chin. Phys. B, 23, 089201, http://doi.org/10.1088/1674-1056/23/8/089201, 2014.

Jacobs, C. M. J., Moors, E. J., Maat, H. W. Ter., Teuling, A. J., Balsamo, G., Bergaoui, K., Ettema, J., Lange, M., Hurk, B. J. J. M. Van Den, Viterbo, P., and Wergen, W.: Evaluation of European Land Data Assimilation System (ELDAS) products using in situ observations, Tellus, 60A, 1023-1037, http://doi.org/10.1111/j.1600-0870.2008.00351.x, 2008.

Kalman, R. E.: A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng., 82(D), 35-45, 1960.

Kalnay, E.: Amospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, pp 512, 2003.

Kalnay, E., Li, H., Miyoshi, T., Yang, S. -C., and Ballabrera-Poy, J.: 4-D-Var or ensemble Kalman filter?, Tellus, 59A, 758-773, http://doi.org/10.1111/j.1600-0870.2007.00261.x, 2007.

Knyazikhin, Y., Martonchik, J. V., Myneni, R. B., Diner, D. J., and Running, S. W.: Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data, J. Geophys. Res., 103, 32257-32275, 1998.

Kwon, Y., Yang, Z. L., Zhao, L., Hoar, T. J., Toure, A. M., and Rodell, M.: Estimating snow water storage in North America using CLM4, DART, and Snow Radiance Data Assimilation, J. Hydrometeo., 17, 2853-2874, http://doi.org/10.1175/JHM-D-16-0028.1, 2016.

Lawrence, P. J., and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), J. Geophys. Res., 112, G01023, http://doi.org/10.1029/2006JG000168, 2007.

Leng, H. Z., and Song, J. Q.: Hybrid three-dimensional variation and particle filtering for nonlinear systems, Chin. Phys. B, 22, 030505, http://doi.org/10.1088/1674-1056/22/3/030505, 2013.

Levis, S., Bonan, G. B., Vertenstein, M., and Oleson, K. W.: The Community Land Model's Dynamic Global Vegetation Model (CLM-DGVM): Technical Description and User's Guide, Boulder, Colorado: National Center for Atmospheric Research, NCAR/TN-459+IA, 2004.

Li, X. J., Xiao, Z. Q., Wang, J. D., Qu, Y., and Jin, H. A.: Dual Ensemble Kalman Filter assimilation method for estimating time series LAI, J. Remote. Sens., 18, 27-44, http://doi.org/10.11834/jrs.20133036, 2014.

Li, Y., Zhao, M. S., Motesharrei, S., Mu, Q. Z., Kalnay, E., and Li, S. C.: Local cooling and warming effects of forests based on satellite observations, Nat. Commun., 6, 6603, http://doi.org/10.1038/ncomms7603, 2015.

Lin, P. R., Wei, J. F., Yang, Z. L., Zhang, Y. F., and Zhang, K.: Snow data assimilation-constrained land initialization improves seasonal temperature prediction, Geophy. Res. Lett., 43, 11423, http://doi.org/10.1002/2016GL070966, 2016.

Luo, L. F., Robock, A., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Lohmann, D., Cosgrove, B., Wen, F. H., Sheffield, J., Duan, Q. Y., Higgins, R. W., Pinker, R. T., and Tarpldy, D.: Validation of the North American Land Data Assimilation System (NLDAS) retrospective forcing

over the southern Great Plains, J. Geophys. Res. Atmos., 108, 8843, http://doi.org/10.1029/2002JD003246, 2003.

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q. Y., Luo, L. F., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marchall, C. H., Entin, J. K., Pan, M. Koren, V., Meng, J., Ramsay, B. H., and Bailey, A. A.: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, J. Geophys. Res., 109, D07S90, http://doi.org/10.1029/2003JD003823, 2004.

Montzka, C., Pauwels, V. R. N., Franssen, H.-J. H., Han, X.-J., and Vereecken. H.: Multivariate and multiscale data assimilation in terrestrial systems: a review, Sensors, 12, 16291-16333, https://doi.org/10.3390/s121216291, 2012.

Moradkhani, H., Hsu, K. L., Gupta, H., Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resour. Res., 41(5), W05012, http://doi.org/10.1029/2004WR003604, 2005.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., Levis, S., Swenson, S. C., Thornton, P. E., Dai, A. G., Decker, M., Dickinson, R., Feddema, J., Heald, C. L., Hoffman, F., Lamarque, J. -F., Mahowald, N., Niu, G. Y., Qian, T. T., Randerson, J., Running, S., Sakaguchi, K., Slater, A., Stöckli, R., Wang, A. H., Yang, Z. L., Zeng, X. D., and Zeng, X. B.: Technical Description of Version 4.0 of the Community Land Model, NCAR Tech Note (NCAR/TN-478 + STR) 257pp, 2010.

Pitman, A. J.: The evolution of, and revolution in, land surface schemes designed for climate models, Int. J. Climatol., 23(5), 479-510, http://doi.org/10.1002/joc.893, 2003.

Pitman, A. J., Noblet-Ducoudré, N. de, Cruz, F. T., Davin, E. L., Bonan, G. B., Brovkin, V., Claussen, M., Delire, C., Ganzeveld, L., Gayler, V., van den Hurk, B.J.J.M., Lawrence, P. J., van der Molen, M. K., Müller, C., Reick, C. H., Seneviratne, S. I., Strengers, B. J., and Voldoire, A.: Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study, Geophys. Res. Lett., 36, L14814, http://doi.org/10.1029/2009GL039076, 2009.

Raeder, K., Anderson, J. L., Collins, N., Hoar, T. J., Kay, J. E., Lauritzen, P. H., and Pincus, R., DART/CAM: An Ensemble Data Assimilation System for CESM Atmospheric Models, J. Clim., 25, 6304-6317, http://doi.org/10.1175/JCLI-D-11-00395.1, 2012.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The global land data assimilation system, Bull. Amer. Meteor. Soc., 85, 381-394, http://doi.org/10.1175/BAMS-85-3-381, 2004.

Thornton, P. E., and Zimmermann, N.E.: An improved canopy integration scheme for a land surface model with prognostic canopy structure, J. Clim., 20, 3092-3923, http://doi.org/10.1175/JCLI4222.1, 2007.

Viskari, T., Hardiman, B., Deasi, A. R., and Dietz, M. C.: Model-data assimilation of multiple phenological observations to constrain and predict leaf area index, Ecol. Appl., 25, 546-558, 2015.

Wan, L. Y., Zhu, J., Wang, H., Yan, C. X., and Bertino, L.: A "dressed" ensemble Kalman filter using the hybrid coordinate ocean model in the pacific, Adv. Atmos. Sci., 26(5). 1042-1052, http://doi.org/10.1007/s00376-009-7208-6, 2009.

Wang, X. G., Hamill, T. M., Whitaker, J. S., and Bishop, C. H.: A Comparison of Hybrid Ensemble Transform Kalman Filter-Optimum Interpolation and Ensemble Square Root Filter Analysis Schemes, Mon. Wea. Rev., 135, 1055-1076, http://doi.org/10.1175/MWR3307.1, 2007.

Whitaker, J. S., and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Mon. Wea. Rev., 130, 1913-1924, 2002.

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., and Toth, Z.: Ensemble data assimilation with the NCEP global forecasting system, Mon. Wea. Rev., 136, 463-482, http://doi.org/10.1175/2007MWR2018.1, 2008.

Wu, X. R., Han, G. J., Li, D., and Li, W.: A hybrid ensemble filter and 3D variational analysis scheme, J. Trop. Oceanogr. (in Chinese), 30(6), 24-30, 2011.

Xia, Y. L., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L. F., Alonge, C., Wei, H., Meng, J., Livneh, B., Duan, Q. Y., and Lohmann, D.: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow, J. Geophys. Res., 117, D03109, http://doi.org/10.1029/2011JD016051, 2012.

Xiao, Z. Q., Liang, S. L., Wang, J. D., and Wu, X. Y.: Use of an ensemble Kalman Filter for real-time inversion of Leaf Area Index from MODIS time series data, IEEE Trans. Geosci. Remote Sens., 4: 73-76, 2009.

Zhang, F. Q., Zhang, M., and Hansen, J. A.: Coupling Ensemble Kalman Filter with Four-dimensional Variational Data Assimilation, Adv. Atmos. Sci., 26, 1-8, http://doi.org/10.1007/s00376-009-0001-8, 2009.

Zhang, L., Huang, S. X., Shen, C., and Shi, W. L.: Variational assimilation in combination with the regularization method for sea level pressure retrieval from QuickSCAT scatterometer data I: Theoretical frame construction, Chin. Phys. B, 20(11), 119201, http://doi.org/10.1088/1674-1056/20/11/119201, 2011.

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

Zhao, L., Yang, Z. L., Hoar, T. J.: Global Soil Moisture Estimation by Assimilating AMSR-E Brightness Temperatures in a Coupled CLM4-RTM-DART System, J. Hydrometeo., 17, 2431-2454, http://doi.org/10.1175/JHM-D-15-0218.1, 2016.

Zupanski, M.: Maximum likelihood ensemble filter: Theoretical aspects, Mon. Wea. Rev., 133, 1710-1726, 2005.