Comparison of Different Sequential Assimilation Algorithms for Satellite-derived Leaf Area Index Using the Data Assimilation Research Testbed (version 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. In this study, 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 Kernel Filter (KF), the Ensemble Kalman Filter (EnKF), the Ensemble Adjust Kalman Filter (EAKF), and the Particle Filter (PF), are applied, thoroughly analysed 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. From the aspect of average and RMSE, 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).

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 land surface models (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). Data Assimilation (DA), 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).

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 DA based on optimal control theory and sequential algorithms based on the Kalman Filter (KF) (Dimet and Talagrand, 1986; Gordon et al., 1993; Bannister et al., 2017; Vetra-Carvalho et al., 2018). 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). During recent years, the Monte-Carlo methods were proposed to allow the assimilation of information from sources that have non-Gaussian errors.

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 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), the combination of 3DVAR and PF algorithms (Leng and Song, 2013), 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).

A complete Land Data Assimilation System (LDAS) is mainly composed of forcing, initial and boundary datasets, parameterization sets, dynamical models as physical constraints, assimilation algorithms, observational data and target output. During recent decades, land data assimilation has become very active, though it occurred later than assimilation of atmospheric observations (Lahoz and De Lannoy, 2014). Land data assimilation can assimilate both in-situ observations and remotely sensed data to constrain the physical parametrizations and initialization of land surface states, for example,

assimilating satellite observation of soil moisture, snow water equivalent (SWE), land surface temperature, and so on (Liu et al., 2008; Reichle et al., 2014; Zhang et al., 2014; Zhao et al., 2016; 2018). 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).

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). Joint assimilation of surface incoming solar radiation, soil moisture and vegetation dynamics (LAI) into land surface models or crop models is of great importance that it can improve the model results for national food policy and security assessments (Sabater et al., 2008; Ines et al., 2013; Sawada et al., 2015; Jin et al., 2018; Mokhtari et al., 2018). 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; Sawada et al., 2018).

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 or CLM4.5 to improve snow, soil moisture estimation, as well as land carbon processes (Zhang et al., 2014; Kwon et al., 2016; Zhao et al., 2016; Fox et al., 2018; Zhao et al., 2018). 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 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 (the lanai version)

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 lanai version of DART was used in this study, which supports for many existing models including CESM climate component, the MPAS (Model for Prediction Across Scales) models, the NOAH land model and so on. Released in December 2013, the lanai version of DART can also support for many new observation types and sources, diagnostic routines, as well as new utilities. 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.

2.3 Sequential Assimilation Algorithms

According to Anderson et al. (2001), Equation (1) was used to express how new sets of observations modify the prior joint state conditional probability distribution available from predictions based on previous observation sets.

$$\mathbf{p}(\mathbf{z}_{t,k}|\mathbf{Y}_{t,k}) = \mathbf{p}(\mathbf{y}^{o}_{t,k}|\,\mathbf{z}_{t,k})\;\mathbf{p}(\mathbf{z}_{t,k}\,|\,\mathbf{Y}_{t,k-1})/\;\mathbf{p}(\mathbf{y}^{o}_{t,k}|\,\mathbf{Y}_{t,k-1}) \quad (1)$$

in which $\mathbf{Y}_{t,k}$ is defined as the superset of all observation subsets, $\mathbf{y}^{o}_{t,k}$ is the kth subset of observations at time t, $\mathbf{z}_{t,k}$ is the joint state-observation vector for a given t and k. In ensemble applications, there is generally no need to compute the denominator of (1). Four algorithms for approximating the product in the numerator of (1) are presented below, and detail information can refer to Anderson et al. (2001).

2.3.1 Ensemble Kernel Filter (EKF)

The kernel filter mechanism, first proposed by Lindgren et al. (1993) and developed in Anderson and Anderson (1999), was incorporated in the DART and can be extended to the joint state space. For detail calculation process can refer to Anderson et al. (2001). The kernel filter is potentially general, because the values and expected values of the mean and covariance and higher-order moments of the resulting ensemble are functions of high-order moments of the prior distribution. However, when applied to large models, the algorithm will come to computational efficiency issues.

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 (Burgers et al., 1998; 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. The traditional EnKF was used in this study (Houtekamer and Mitchell1998), which were an ensemble of Kalman Filters, each using a different sample estimate of the prior mean and observations.

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:

$$\Delta X_{ij} = \frac{\sigma_{j_o}^p}{\sigma_o^p + \sigma_{j_o}^p} \Delta Y_i. \quad (2)$$

where *i* indicates the ensemble member, *j* is the state vector member, $\sigma_{j_o}^p$ is the prior covariance of state vector and observation, and σ_o^p is the prior variance of observation.

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

$$\mathrm{p}(\mathrm{X}_{\mathrm{k}}^{\mathrm{a}}|\mathrm{Y}_{1:\mathrm{k}}) \approx \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{w}_{\mathrm{i,k}}\, \delta(\mathrm{X}_{\mathrm{k}}^{\mathrm{a}} - \mathrm{X}_{\mathrm{i,k}}^{\mathrm{a}}). \quad (3)$$

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

in which $\mathrm{p}(\mathrm{X}_{\mathrm{k}}^{\mathrm{a}}|\mathrm{Y}_{1:\mathrm{k}})$ is the posterior probability distribution, $\mathrm{X}_{\mathrm{i,k}}^{\mathrm{a}}$ is the particle element, $\mathrm{w}_{\mathrm{i,k}}$ is the weight of each particle, *N* is the number of particles. 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 Datasets

2.4.1 Ensemble Meteorological Forcing and initial conditions

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.

To achieve a steady state solution for all state variables, the CLM4CN was run for 4000 years by Qian's forcing (Qian et al., 2006) at a resolution of 1.9° latitude by 2.5° longitude (Shi et al., 2013). Then the CLM4CN was forced by the ensemble mean of selected 40 members of DART/CAM datasets for 1000 years. At last step, the ensemble simulation during the time period from 1998 to 2001 was treated as spin-up process, and 40 ensemble initial conditions were achieved. Aiming at global scale, only one-

year assimilation and ensemble simulation were conducted in considering of computational cost. We were trying to firstly find out the best experiment, and then conducting a long-term simulation or assimilation in the future, so only one year of assimilation results were presented.

2.4.2 LAI datasets

Global Land Surface Satellite (GLASS) LAI datasets was used in this study as assimilated observation (Zhao et al., 2013). As the ensemble simulation or assimilation was run at a resolution of 0.9° latitude by 1.25° longitude, the original spatial resolution of 0.05° of GLASS LAI is upscaled to the same resolution.

To evaluate the assimilation result, an improved LAI dataset developed from the MODerate Resolution Imaging Spectroradiometer (MODIS) (Yuan et al., 2011) was utilized, which can reduce the spatial and temporal inconsistencies observed at the local spatial or temporal scales by considering the characteristics of the MODIS LAI data and quality control (QC) information (Baret et al., 2013). The resolution is 1 kilometer, and was also upscaled to grid levels to evaluate the analysis LAI and assimilation effect.

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	NO
Algorithms without observation rejection	GLASS LAI	LAI, Leaf C, Leaf N	EAKF	YES

To determine the optimal assimilation algorithm, five experiments corresponding to the KF, EnKF, EAKF and PF methods are designed and showed in Table 1, in which the "Algorithms" experiments would reject some observation under certain conditions using the KF, EnKF, EAKF, and PF algorithms. The expected value of the difference between prior mean and observation is $\sqrt{\sigma_{prior}^2 + \sigma_{obs}^2}$, in which σ_{prior} and σ_{obs} are the standard deviation of prior PDF and observation PDF respectively. DART will reject the observation if the bias of prior mean and observation is larger than 3 times of the expected value. The "Algorithms without observation rejection" experiments would accept all the observed LAI. 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 adjusted DART state vector was resent to the CLM restart files as a new initial condition for the next time step. All the simulation and assimilation were at a spatial resolution of 0.9° latitude by 1.25° longitude. The ensemble assimilation was conducted pointwise, indicating that spatial covariances were not considered.

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 updated 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 region, 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.

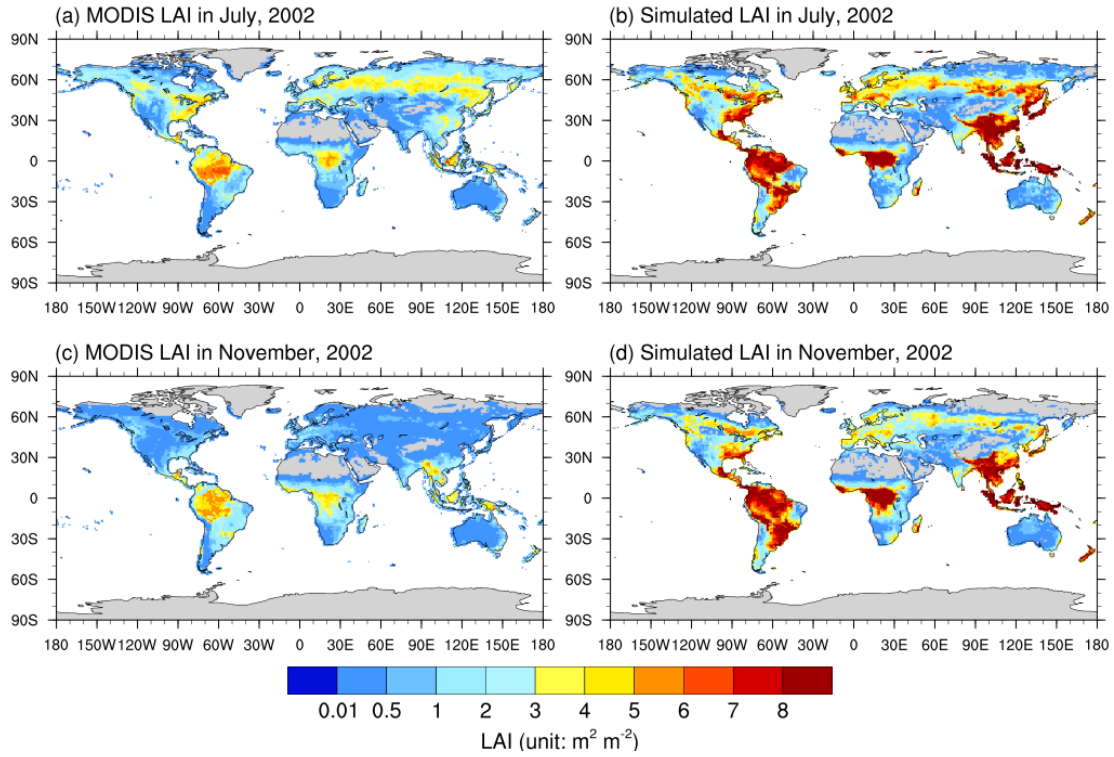

Figure 1: Spatial distributions of global LAI values in 2002 for (a) observations in July, (b) ensemble mean of simulations in July, (c) observations in November, and (d) ensemble mean of simulations in November.

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 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.

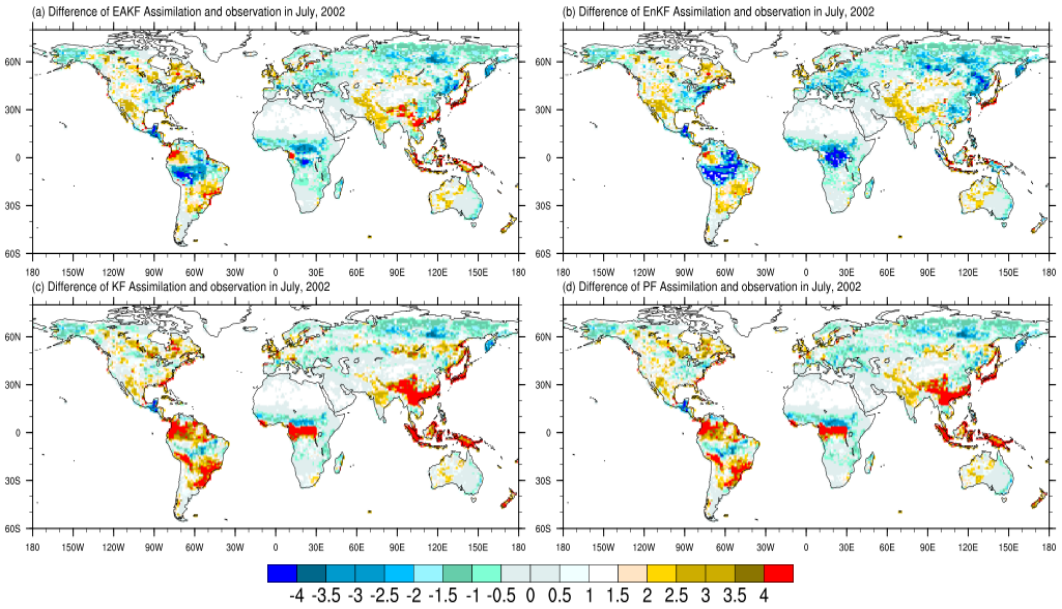

Figure 2: Differences between global LAI from assimilation experiments and that from observations in July 2002 with the methods of (a) EAKF, (b) EnKF, (c) KF and (d) PF.

The results also indicated that the EAKF and EnKF assimilation algorithms are better than the KF and PF algorithms in November (not shown). 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. From the aspect of average and RMSE, the PF algorithm performs worse than the EAKF and EnKF algorithms because of the gradually reduced acceptance of observations with assimilation steps. While average and RMSE only make sense for the Ensemble Kalman Filters, for the PF algorithm, the particle with largest weight (a posteriori maximum for the pdf) should be discussed in the future.

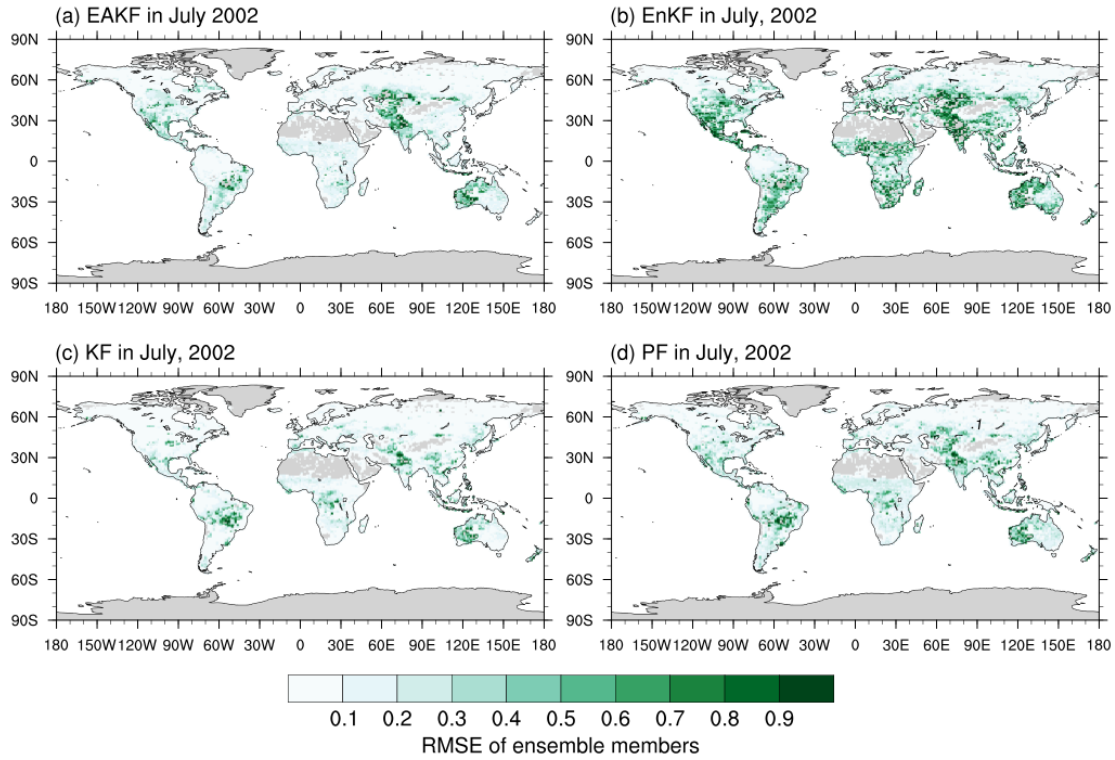

Figure 3: Same as Fig. 2, but for RMSE of ensemble members.

The RMSE of ensemble members were showed in Figure 3 to provide the hints of where assimilation is the most efficient. The RMSE of ensemble members for the EAKF and EnKF algorithm is larger than the KF and PF algorithms, indicating more effective assimilation for these algorithms. In July 2002, the RMSE of ensemble estimates is the largest in lower latitude regions, with particularly high values in central South America, central Africa, and Southeast Asia. The regions with comparatively large ensemble spreads were located in eastern North America and western Europe. The large ensemble spreads dominant area are also transitional regions with different vegetation types, indicating the low capacity of the models to simulate complex vegetation types.

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 differences (RMSDs) 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 RMSDs 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 RMSD relative to the observations is associated with the simulation, followed by the assimilation datasets from the KF and PF algorithms, and the RMSDs are lowest for the EAKF and EnKF methods. During the growing season, the RMSDs 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 RMSD 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. The assimilation is far less efficient in the boreal region than in other places, partly due to the consistently low initial RMSD during non-growing seasons and limited capacity to simulate processes associated with boreal forest types.

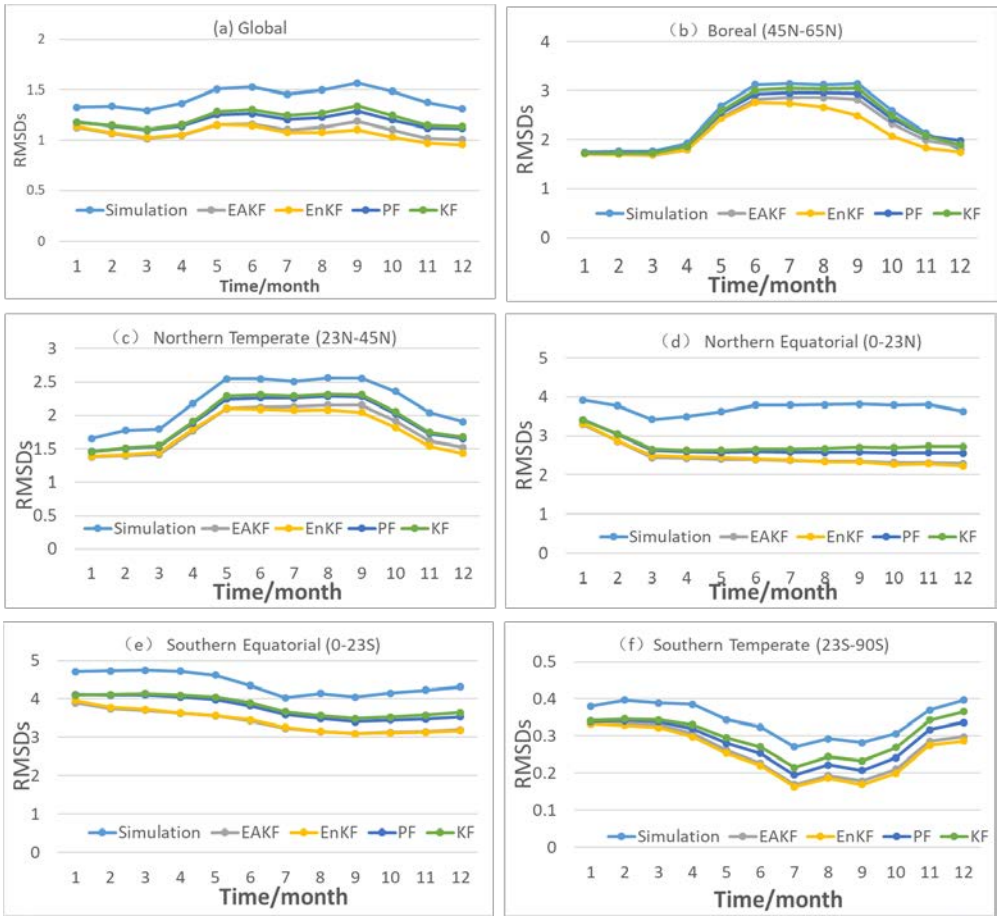

Figure 4: RMSDs of ensemble means of simulation/assimilation versus observations for (a) global, (b) boreal (45-65°N), (c) northern temperate (23-45°N), (d) northern equatorial (0-23°N), (e) southern equatorial (0-23°S), and (f) southern temperate (23-90°S).

Figure 5 shows the globally or regionally averaged RMSDs of simulation/assimilation and observations. The RMSDs 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 RMSDs 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 difference between simulation and all four algorithms in the Northern Equatorial and Southern Equatorial regions is larger than in other regions, indicating assimilation is more efficient.

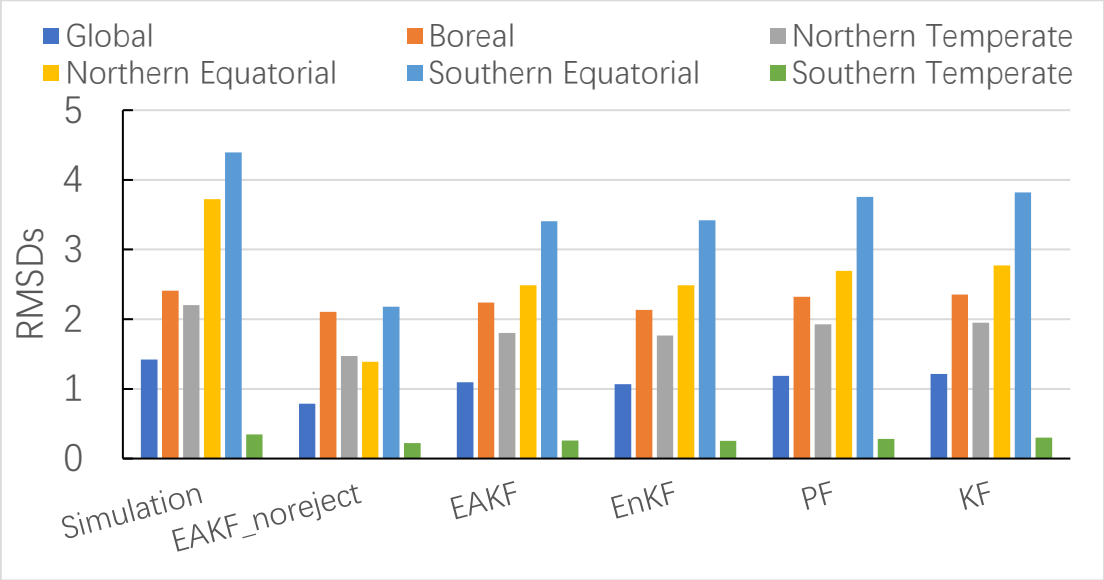

Figure 5: Globally or regionally averaged RMSDs for the simulation/assimilation results.

The background/analysis departures are calculated as (1) innovations, which are the differences between the assimilated LAI 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 similar 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.

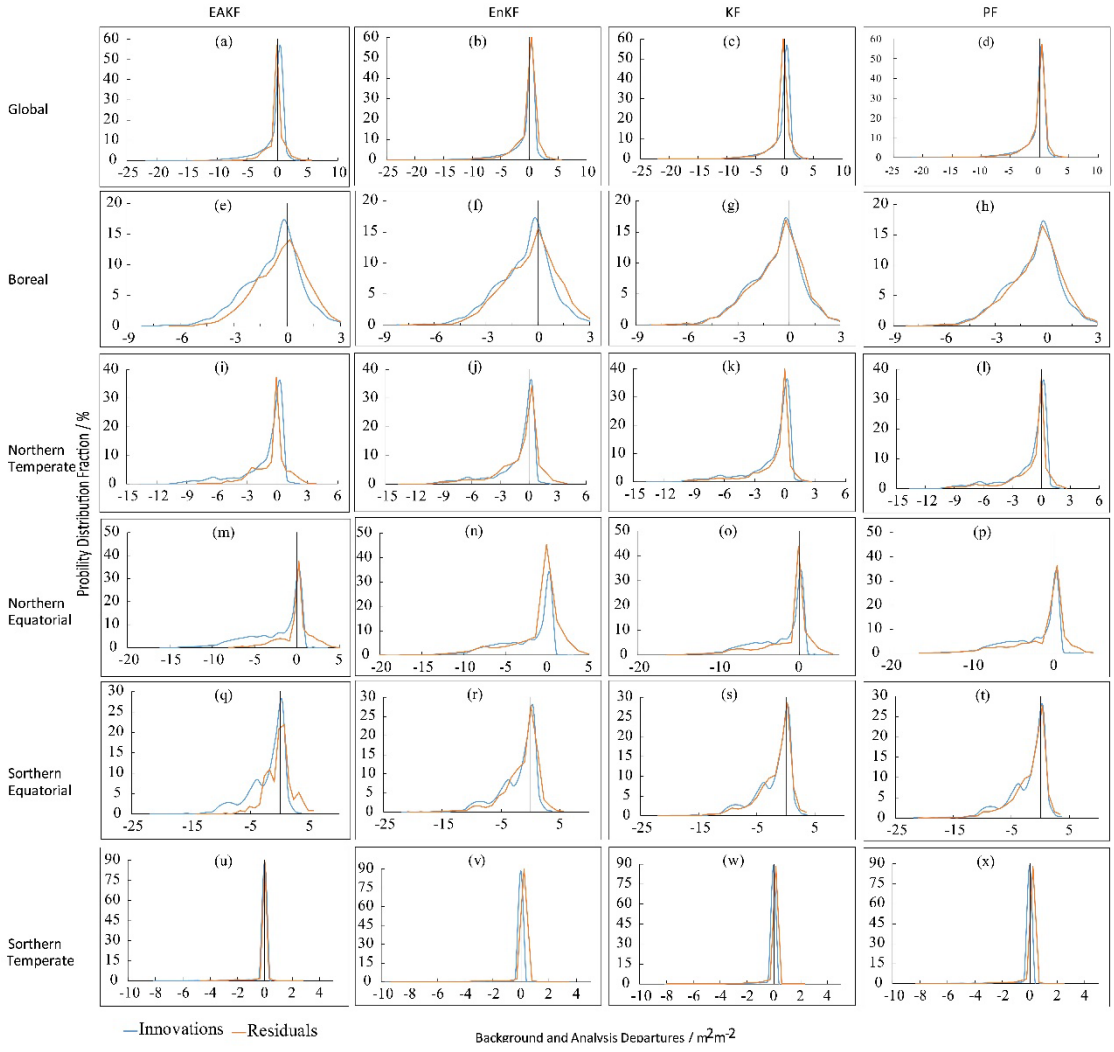

Figure 6: The histograms of innovation and residuals of LAI globally and for all subregions during July 2002. (a-d) Global; (e-h) boreal; (i-l) northern temperate; (m-p) northern equatorial; (q-t) southern equatorial; (u-x) southern temperate

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. During assimilation, DART can calculate the number of non-assimilated observation when the difference of prior mean and observation is larger than 3 times of the expected value. The proportion of of accepted LAI observations is defined as the number of accepted observations divided by the number of total observations.

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.

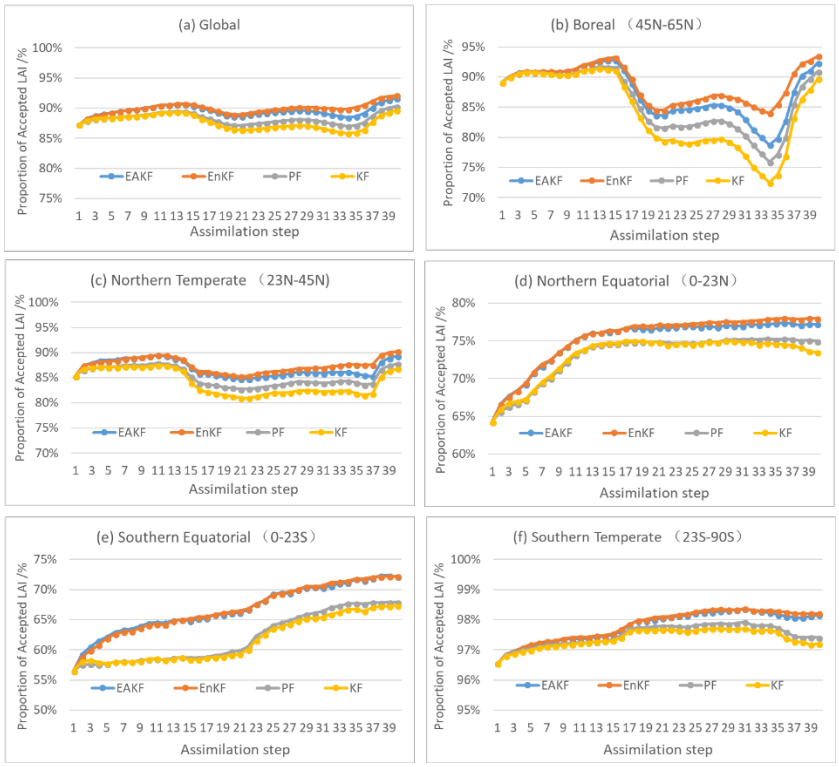

Figure 7: The proportion of accepted LAI observations for the four algorithms in the zonal regions.

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).

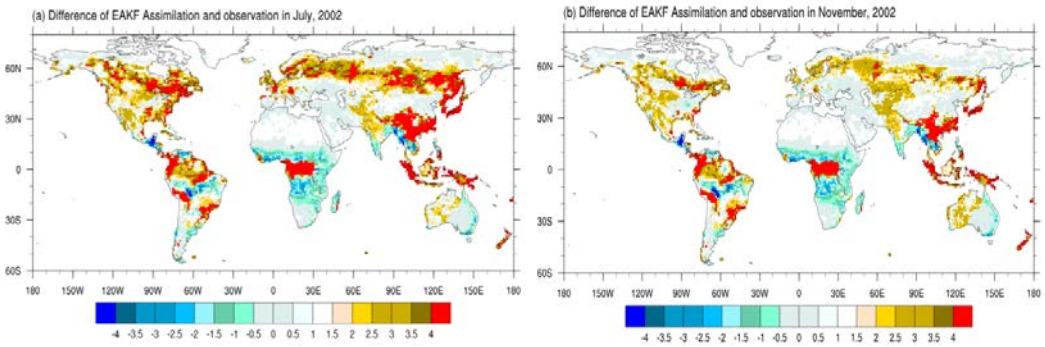

Figure 8: Differences between globally assimilated and observed LAIs for the methods of EAKF in (a) July and (b) November.

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 RMSDs of simulation experiments without/with rejection (EAKF_noreject / EAKF_reject) and MODIS LAI 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, RMSDs for both simulation and assimilation present obvious annual variations, with RMSDs reaching their maximum values in the season with considerable vegetation growth over a large area. The RMSD 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 RMSD relative to the observations was highest for the simulation, followed by the EAKF_reject experiment, and was lowest for the EAKF_noreject experiment. The RMSDs 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 RMSD is smaller than that when rejecting some observations.

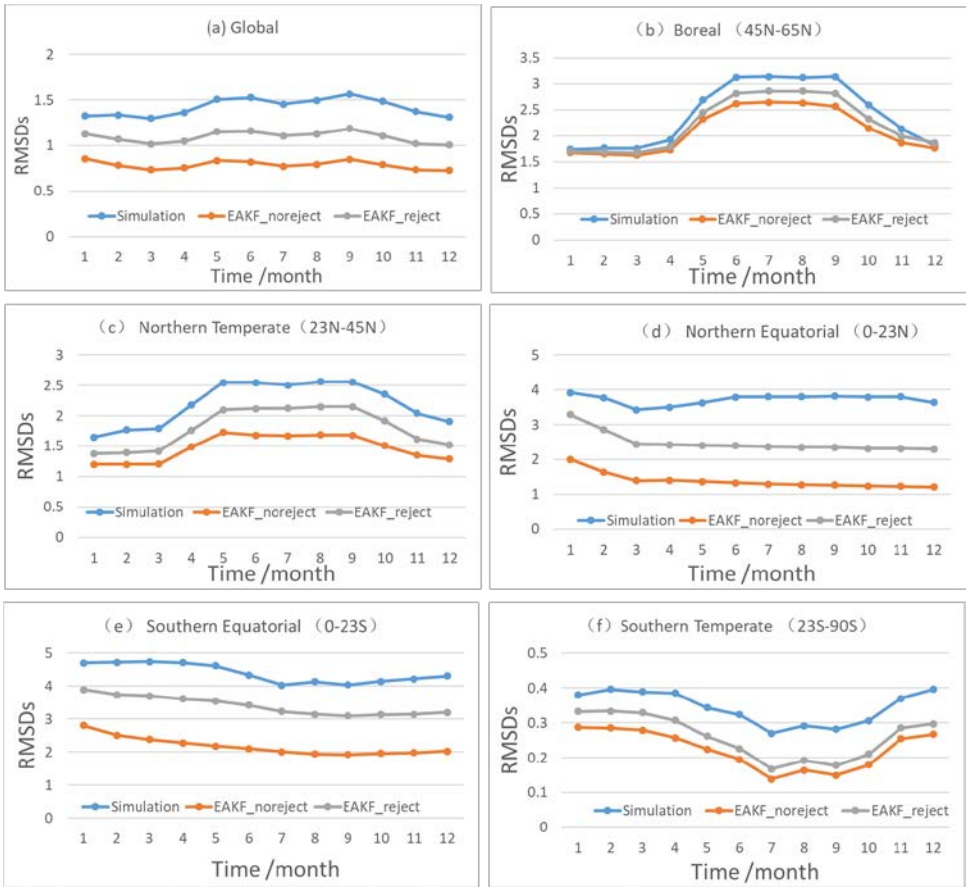

Figure 9: RMSDs of simulation experiments without/with rejection (EAKF_noreject and EAKF_reject) and MODIS LAI for the (a) globe, (b) boreal (45-65°N), (c) northern temperate (23-45°N), (d) northern equatorial (0-23°N), (e) southern equatorial (0-23°S), and (f) southern temperate (23-90°S) regions.

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 (RMSD) is associated with the EAKF algorithm, suggesting that the EAKF algorithm is the best and has a robust performance.

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 RMSD of the results is smaller than that when some observations are rejected.

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.

Bannister, R. N.: A review of operational methods of variational and ensemble variational data assimilation, Q. J. R. Meteorol. Soc., 143, 607-633, 2016.

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.

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., & Smets, B.: GEOVE: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production, Remote Sensing of Environment, 137,299-309. doi: 10.1016/j.rse.2012.12.027, 2013.

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.

Burgers, G., van Leeuwen, P. J. and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Mon. Wea. Rev., 126, 1719-1724, 1998.

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.

De Lannoy, G. J. M., de Rosnay, P. and Reichle, R. H.: Soil moisture data assimilation. In Handbook of Hydrometeorological Ensemble Forecasting, edited by Q. Duan, F. Pappenberger, J. Thielen, A.Wood, H. Cloke and J. C. Schaake. 2016.

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.

Fox, A. M., Hoar, T. J., Anderson, J. L., Arellano, A. F., Smith, W. K., Litvak, M. E., et al.: Evaluation of a data assimilation system for land surface models using CLM4.5. Journal of Advances in Modeling Earth Systems, 10, 2471-2494, 2018.

Gordon, N. J., Salmond, D. J. and Smith, A. F.: Novel approach to nonlinear/non- Gaussian Bayesian state estimation, IEE Proc., 140, 107-113, 1993.

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.

Ines, A. V. M., Das, N. N., Hansen, J. P. and Njoku, E. G.: Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction, Remote Sensing of Environment, 138, 149-164, 2013.

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.

Jin, X., Kumar, L., Li, Z., Xu, X., Yang, G. and Wang, J.: A review of data assimilation of remote sensing and crop models, Eur. J. Agron., 92, 141-152, 2018.

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.

Lahoz, W. A. and De Lannoy, G. J. M.: Closing the Gaps in Our Knowledge of the Hydrological Cycle over Land: Conceptual Problems, Surv. Geophys., 35, 623-660, 2014.

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.

Lindgren, F, Geladi, P, Wold, S.: The kernel algorithm for PLS, J. Chemometrics, 7, 45, 1993.

Liu，Q., L. Gu, R. E. Dickinson, et al.: Assimilation of satellite reflectance data into a dynamical leaf model to infer seasonally varying leaf areas for climate and carbon models, J. Geophy. Res., 113, D19113, doi: 10.1029/2007JD009645, 2008.

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.

Mokhtari, A., Noory, H. and Vazifedoust, M. : Improving crop yield estimation by assimilating LAI and inputting satellitebased surface incoming solar radiation into SWAP model, Agr. Forest Meteorol., 250-251, 159-170, https://doi.org/10.1016/j.agrformet.2017.12.250, 2018.

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.

Qian, T. T., Dai, A. G., Trenberth, K. E., & Oleson, K.W.: Simulation of global land surface conditions from 1948 to 2004: Part I: Forcing data and evaluations, J. Hydrometeor., 7(5), 953-975, doi:10.1175/JHM540.1, 2006.

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.

Reichle, R. H., De Lannoy, G. J. M., Forman, B. A., Draper, C. S. and Liu, Q.: Connecting Satellite Observations with Water Cycle Variables Through Land Data Assimilation: Examples Using the NASA GEOS-5 LDAS, Surv. Geophys., 35, 577-606, 2014.

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.

Sabater, J. M., Rüdiger, C., Calvet, J.-C., Fritz, N., Jarlan, L. and Kerr Y.: Joint assimilation of surface soil moisture and LAI observations into a land surface model, Agr. Forest Meteorol., 148, 1362-1373, 2008.

Sawada, Y., Koike, T. and Walker, J. P.: A land data assimilation system for simultaneous simulation of soil moisture and vegetation dynamics. J. Geophys. Res. Atmos., 120, 5910-5930, 2015.

Sawada, Y.: Quantifying Drought Propagation from Soil Moisture to Vegetation Dynamics Using a Newly Developed Ecohydrological Land Reanalysis, Remote Sens., 10, 1197, 2018.

Shi, M.J., Yang, Z. L., Lawrence, D. M., Dickinson, R. E., & Subin, Z. M.: Spin-up processes in the Community Land Model version 4 with explicit carbon and nitrogen components. Ecological Modelling, 263, 308-325, doi:10.1016/j.ecolmodel.2013.04.008, 2013.

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.

Vetra-Carvalho, S., van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P. Kirchgessner, P. and Beckers, J.-M.: State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems, Tellus A, 70, 1445364, 2018.

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.

Yuan, H., Dai, Y. J., Xiao, Z. Q., Ji, D. Y., & Wei, S. G.: Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling. Remote Sensing of Environment, 115, 1171-1187. doi: 10.1016/j.rse.2011.01.001, 2011.

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.

Zhao, L., and Yang, Z. L.: Multi-sensor land data assimilation: Toward a robust global soil moisture and snow estimation, Remote Sens. Environ., 216, 13-27, https://doi.org/10.1016/j.rse.2018.06.033, 2018.

Zhao, X., Liang, S. L., Liu, S. H., Yuan, W. P., Xiao, Z. Q., Liu, Q., et al.: The Global Land Surface Satellite (GLASS) Remote Sensing Data Processing System and Products. Remote Sensing, 5, 2436-2450. doi:10.3390/rs5052436, 2013.

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