JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2020MS002394

Improve the Performance of the Noah-MP-Crop Model by Jointly Assimilating Soil Moisture and Vegetation Phenology Data

Tongren Xu[1], Fei Chen[2], Xinlei He[1], Michael Barlage[2], Zhe Zhang[3], Shaomin Liu[1], and Xiangping He[1]

[1]State Key Laboratory of Earth Surface Processes and Resource Ecology, School of Natural Resource, Faculty of Geographical Science, Beijing Normal University, Beijing, China, [2]National Center for Atmospheric Research, Boulder, CO, USA, [3]School of Environment and Sustainability, University of Saskatchewan, Saskatoon, SK, Canada

Key Points:

- The Multipass Land Data Assimilation Scheme (MLDAS) is proposed based on the Noah-MP-Crop model
- Leaf area index (LAI), soil moisture (SM), and solar-induced chlorophyll fluorescence (SIF) measurements are assimilated into the MLDAS to predict sensible heat flux (*H*), latent heat flux (LE), and gross primary productivity (GPP)
- *H* and LE estimates are more sensitive to SM measurements, while the GPP retrievals are more affected by LAI and SIF observations

Correspondence to:
X. He,
hxlbsd@mail.bnu.edu.cn

Citation:
Xu, T., Chen, F., He, X., Barlage, M., Zhang, Z., Liu, S., & He, X. (2021). Improve the performance of the Noah-MP-Crop model by jointly assimilating soil moisture and vegetation phenology data. *Journal of Advances in Modeling Earth Systems*, *13*, e2020MS002394. https://doi.org/10.1029/2020MS002394

Received 1 NOV 2020
Accepted 27 JUN 2021

Abstract The interactions between crops and the atmosphere significantly impact surface energy and hydrology budgets, climate, crop yield, and agricultural management. In this study, a multipass land data assimilation scheme (MLDAS) is proposed based on the Noah-MP-Crop model. The ensemble Kalman filter (EnKF) method is used to jointly assimilate the leaf area index (LAI), soil moisture (SM), and solar-induced chlorophyll fluorescence (SIF) observations to predict sensible (*H*) and latent (LE) heat fluxes, gross primary productivity (GPP), etc. Such joint assimilation is demonstrated to be effective in constraining the model state variables (i.e., leaf biomass and SM) and optimizing key crop-model parameters (i.e., specific leaf area [SLA], and maximum rate of carboxylation, Vcmax). The performance of the MLDAS is evaluated against observations at two AmeriFlux cropland sites, revealing good an agreement with the observed *H*, LE, and GPP. When using optimized model parameters (SLA and Vcmax) and jointly assimilating LAI, SM, and SIF observations, the MLDAS produces 34.28%, 26.90%, and 51.82% lower root mean square deviations for daily *H*, LE, and GPP estimates compared with the Noah-MP-Crop open loop simulation. Our findings also indicate that the *H* and LE predictions are more sensitive to SM measurements, while the GPP simulations are more affected by LAI and SIF observations. The results indicate that performances of physical models can be greatly improved by assimilating multi-source observations within MLDAS.

Plain Language Summary Accurate estimations of water and carbon fluxes in croplands are required for monitoring crop yield, hydrology, and irrigation scheduling. Thus, many crop models are aimed at modeling crop dynamics to establish more accurate water, carbon, and energy processes. The data assimilation (DA) method can combine land surface and vegetation phenological information with a process-based model to reduce the uncertainty in model variables and optimize model parameters. In this study, multi-source observations of the land surface and vegetation phenology are assimilated into a land-surface-crop model to improve the water, carbon, and energy fluxes over croplands. This research involved the construction of a DA framework with multipass ensemble Kalman filter method. The results show that the assimilation of soil moisture observations significantly improves the surface heat fluxes estimates, while the assimilation of vegetation phenological observations significantly improves the vegetation dynamics and crop yield estimates.

1. Introduction

Accurate predictions of water and carbon fluxes in croplands are essential for determining crop yield, hydrologic components, and irrigation schedules and researching climate change (He, Xu, Xia, et al., 2020; Lokupitiya et al., 2016; Williams, Gornall, et al., 2016; Xu, Guo, et al., 2018). Process-based land surface models (LSMs) have been developed to represent the complex land/hydrological interactions in earth systems (Chen & Dudhia, 2001; Dai et al., 2003; Lawrence et al., 2019; McDermid et al., 2017). In particular, LSMs have been enhanced by incorporating crop-growth and agriculture-management models to capture photosynthesis, plant growth, and crop yield processes in the Simple Biosphere Model (Lokupitiya et al., 2009); the Joint UK Land Environment Simulator (Osborne et al., 2015); the Community Land Model (Drewniak et al., 2012); and the Noah with multiparameterization model, Noah-MP-Crop model (X. Liu

et al., 2016, 2020). The Noah-MP-Crop model was implemented in the High-Resolution Land Data Assimilation System (F. Chen et al., 2007) and Weather Research and Forecasting model (Skamarock et al., 2007). It improves the simulations of vegetation dynamics and surface heat fluxes for corn and soybean at both field and regional scales by using a number of agricultural management data, such as growing degree days (GDDs) and planting/harvesting dates (X. Liu et al., 2016, 2020; Xu, Chen, et al., 2019; Z. Zhang et al., 2020). However, uncertainties in input data, model structures, and model parameters inevitably induce uncertainties in the variables simulated by those process-based models (X. Li et al., 2007; T. Xu et al., 2011; Zhang, Chen, & Gan, 2016; Zhang, Guanter, et al., 2016; Zhang, Xiao, et al., 2016).

Data assimilation (DA) methods originated from estimation theory and cybernetics to merge observational information into process-based models to mitigate uncertainties in model variables and optimize model parameters (X. Li et al., 2020; Liang & Qin, 2008; Xia et al., 2019). This is done within variational-based or ensemble-based DA schemes to improve the model performances (He, Xu, et al., 2019; He, Xu, Bateni, et al., 2020; He et al., 2018; Lu et al., 2016, 2017, 2020; Margulis et al., 2002; Xu, Bateni, et al., 2018; Xu, Chen, et al., 2019; Xu, He, et al., 2019; T. Xu et al., 2011, 2015). Studies have assimilated various observational variables such as land surface temperature (LST), leaf area index (LAI), soil moisture (SM), and solar-induced chlorophyll fluorescence (SIF) into crop models and/or LSMs, which has improved the estimated crop yields (Ines et al., 2013; X. Li et al., 2018; Wang et al., 2014; Xie et al., 2017), vegetation biomass, evapotranspiration (ET), and gross primary production (GPP) within LSMs (Huang et al., 2008; Kumar et al., 2019; Xu, He, et al., 2019; T. Xu et al., 2015). The assimilation of SM observations improved the estimation of SM, runoff, ET (Han et al., 2012; D. Liu & Mishra, 2017; Lu et al., 2016, 2020; Tian et al., 2017), and drought monitoring and forecasting (Kumar et al., 2014; Yan et al., 2017). Remotely sensed SIF, as an indicator of photochemical processes, provides additional constraints for photosynthesis simulation (Camino et al., 2019; Y. Zhang et al., 2014; Zhang, Xiao, et al., 2016). Recently, a linear relationship between SIF and GPP was used to predict GPP under different environmental conditions and vegetation types (Frankenberg & Berry, 2018; He, Chen, et al., 2019; He, Magney, et al., 2020; Sun et al., 2018; Verma et al., 2017; Zhang, Xiao, et al., 2016).

However, these efforts only assimilated a single variable and/or constructed a one or dual-pass DA framework within process-based crop models (Kumar et al., 2019; Norton et al., 2018; Pinnington et al., 2018; Tian et al., 2017; K. Yang et al., 2007). Therefore, this study proposes a multipass land data assimilation scheme (MLDAS), based on the Noah-MP-Crop model and ensemble Kalman filter (EnKF) method so that LAI, SM, and SIF observations are simultaneously assimilated to predict sensible heat flux (H), latent heat flux (LE), gross primary product (GPP), and biomass. The developed MLDAS is tested at two AmeriFlux crop sites, namely, Mead (Nebraska, USA) and Bondville (Illinois, USA). The observed H, LE, and GPP from the eddy covariance (EC) flux towers are used to evaluate the MLDAS performance. The results of EnKF are compared with the ensemble open loop (EnOL) (i.e., no assimilation of LAI, SM, and SIF measurements) approach. Finally, the influence of assimilating LAI, SM, and SIF on the H, LE, and GPP estimates within the MLDAS is assessed.

2. Methodology

2.1. Noah-MP-Crop Model

The Noah-MP LSM is an enhanced Noah LSM (F. Chen & Dudhia, 2001; F. Chen et al., 1996; Ek et al., 2003) with expanded options and improved physics and multiparameterization options (Niu et al., 2011; Z. L. Yang et al., 2011). The improved physics options include a two-source energy balance module, a modified two-stream radiation transfer scheme, a short-term dynamic vegetation growth model, a dynamic groundwater component, and a multilayer snowpack. The multiparameterization options provide users with multiple choices of parameterizations in terms of leaf dynamics, surface-layer turbulence treatment, canopy stomatal resistance, SM limiting factors for stomatal resistance, and runoff and groundwater.

The Noah-MP LSM was recently enhanced with agricultural management models such as crop-growth models (X. Liu et al., 2016, 2020) and a dynamic irrigation model (Xu, Chen, et al., 2019; Z. Zhang et al., 2020). It simulates the process of carbohydrates and distributes them to different crop components, such as leaves, stems, grains, and roots. GDDs are utilized in Noah-MP-Crop to determine the different plant growth stages

from planting to harvest. In addition, numerous agricultural management data (e.g., planting and harvesting dates, fraction of cultivated lands, and fraction of irrigation) are used in Noah-MP-Crop to constrain agriculture models. Noah-MP-Crop simulates the dynamic evolution of various biomasses (e.g., leaf, stem, grain, and root mass) and time-varying LAI for corn and soybean, which affects photosynthesis, surface net radiation, ET, SM, and surface heat fluxes (*H* and LE) (Kumar et al., 2019; X. Liu et al., 2016). This model also estimates grain biomass as the "yield" when the crop reaches physiological maturity. The values of the key parameters in Noah-MP-Crop used in this study are obtained from X. Liu et al. (2016, 2020) and Z. Zhang et al. (2020).

2.2. Multipass Land Data Assimilation Scheme

2.2.1. Data Assimilation Strategy

The MLDAS is based on the Noah-MP-Crop and EnKF method and dynamically estimates model states and parameters independently by assimilating multisource remotely sensed or ground-measured observations in multiple passes. In this study, LAI, SM, and SIF are simultaneously assimilated in the MLDAS to improve the performance of Noah-MP-Crop.

Specific leaf area (SLA), defined as the leaf area to its dry mass (Garnier et al., 2001; Gunn et al., 1999), is used to convert leaf mass into LAI in Noah-MP-Crop and represents leaf functioning and plant productivity, ecological behavior, and growth rate (Cornelissen et al., 2003; Wilson et al., 1999). In this study, SLA is equivalent to the BIO2LAI parameter used in X. Liu et al. (2016). Plants with high SLA values rapidly produce biomass and have high growth rates, while plants with low SLA values efficiently preserve nutrients and decrease photosynthesis and growth rates (Wright & Westoby, 1999). SLA is set to a constant (0.015 for corn) in the Noah-MP-Crop model, but it may vary throughout the growing season and from year to year (Bryant, 2011; Rinaldi, 2003; Zhang, Suyker, et al., 2018; Z. Zhang et al., 2020). Therefore, it is necessary to consider a time-varying SLA within the growing season. This study optimized SLA and updated leaf biomass (LFMASS) simultaneously by assimilating the observed LAI. In fact, three data assimilation strategies are used for LAI assimilation here: Strategy I: only update LFMASS, Strategy II: only optimize SLA, and Strategy III: update both LFMASS and SLA. Moreover, surface SM observations (0–0.1 m) are assimilated into the Noah-MP-Crop model to simultaneously update the four-layer (0–0.1, 0.1–0.4, 0.4–1, and 1–2 m) SM.

The maximum rate of carboxylation (Vcmax) is a critical parameter in the Noah-MP-Crop photosynthesis model because it controls the carbon assimilation process (Farquhar et al., 1980; Y. Zhang et al., 2014). Uncertainty in Vcmax often causes large spreads in GPP estimates across models (Bonan et al., 2011; Hu et al., 2014). In this study, the SIF observations are assimilated to optimize Vcmax and constrain the photochemical process.

Overall, the LAI, SM, and SIF data are simultaneously assimilated into Noah-MP-Crop to constrain the model state variables (i.e., LFMASS and SM) and two critical parameters (i.e., SLA and Vcmax) using MLDAS with four passes and four parallel EnKFs as follows:

Pass I: optimize SLA with the assimilation of LAI observations; Pass II: update LFMASS with the assimilation of LAI observations; Pass III: update four-layer SM with the assimilation of surface SM observations; Pass IV: optimize Vcmax with the assimilation of SIF observations.

The flowchart of MLDAS is shown in Figure 1.

2.2.2. Ensemble Kalman Filter Method

The EnKF method proposed by Anderson (2001) is integrated with Noah-MP-Crop and calculates the model states and parameter error covariance based on the Monte Carlo method, as used in many land data assimilation scheme (LDAS) studies (Koster et al., 2018; Moradkhani et al., 2005; Reichle et al., 2002; T. Xu et al., 2015). In the MLDAS framework, the model state variables (i.e., LFMASS and SM) and two critical parameters (i.e., SLA and Vcmax) are updated separately by using the EnKF method. **T** is considered the state vector of the DA scheme used in the four passes of EnKF, representing LFMASS, SLA, SM, and Vcmax.

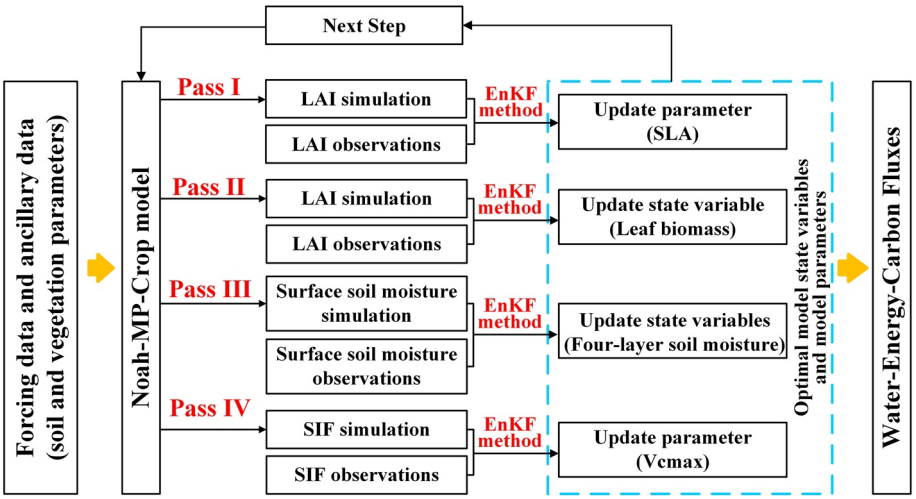

Figure 1. Flowchart of the multipass land data assimilation scheme.

At the beginning of the EnKF method, the initial value ($\mathbf{T}_0$) is used to create the state vector ensembles by adding normally distributed random errors (with zero mean and specified variance):

$$\mathbf{T}_{j,0} = \mathbf{T}_0 + \omega_j \tag{1}$$

where $\mathbf{T}_{j,0}$ represents the jth ensemble member of a state variable at the initial time. ω_j represents the background error, which is assumed to have a normal distribution.

In the forecast step, the state vector is predicted according to:

$$\mathbf{T}^f_{j,k+1} = \mathbf{M}\left(\mathbf{T}^f_{j,k}, F_{k+1}, \alpha_{k+1}\right) + \mu_j \tag{2}$$

where $\mathbf{T}^f_{j,k+1}$ and $\mathbf{T}^f_{j,k}$ are the forecasted state vectors of the jth member at times k+1 and k, respectively. $\mathbf{M}(-)$ represents the model operator (Noah-MP-Crop in this study), and F_{k+1} and α_{k+1} represent the model forcing data and parameters at time k+1, respectively. μ_j is the model error with a zero mean and specified variance.

At times, when measurements of LAI, SM, and SIF is available, the observation operator is described as:

$$Y^f_{j,k} = \mathbf{H}\left(\mathbf{T}^f_{j,k}\right) + \eta_j \tag{3}$$

where $Y^f_{j,k}$ is the model-predicted LAI, SM, and SIF. $\mathbf{H}(-)$ represents the observation operator. For LAI assimilation, the relationship between LAI and LFMASS (LAI = SLA × LFMASS) is described as the observation operator. For SIF assimilation, the linear relationship between SIF and GPP (SIF = a × GPP + b) is described as the observation operator. a and b are calibration parameters. η_j is the observation error with a mean of zero and a covariance of R of the jth member.

The state variable of each member is updated as follows:

$$\mathbf{T}^a_{j,k+1} = \mathbf{T}^f_{j,k+1} + \mathbf{K}\left[Y^o_{k+1} - \mathbf{H}\mathbf{T}^f_{j,k+1}\right] \tag{4}$$

$$\mathbf{K} = P^f\mathbf{H}^T\left(\mathbf{H}P^f\mathbf{H}^T + R\right)^{-1} \tag{5}$$

$$P^f\mathbf{H}^T = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left[\mathbf{T}^f_{j,k+1} - \overline{\mathbf{T}}^f_{k+1}\right]\left[\mathbf{H}\cdot\left(\mathbf{T}^f_{j,k+1} - \overline{\mathbf{T}}^f_{k+1}\right)\right]^T \tag{6}$$

Table 1
Data Assimilation Strategies Discussed in This Study

Data assimilation experiment	Assimilated data	The updated model variables/parameters
EnOL	No data is assimilated	None
EnKF-LAI-I	Leaf area index (LAI)	Optimize specific leaf area (SLA)
EnKF-LAI-II		Update leaf biomass
EnKF-LAI-III		Update both SLA and leaf biomass
EnKF-SM	Surface soil moisture	Update four-layers soil moisture at daily time scale
EnKF-SIF	Solar-induced chlorophyll fluorescence (SIF)	Optimize Vcmax
EnKF-ALL	LAI, surface soil moisture, SIF	Optimize SLA and Vcmax; update leaf biomass and four-layer soil moisture

Abbreviations: EnKF, ensemble Kalman filter; EnOL, ensemble open loop; LAI, leaf area index; SM, soil moisture; Vcmax, maximum rate of carboxylation; SLA, specific leaf area; SIF, Solar-induced chlorophyll fluorescence.

$$\mathbf{H}P^{f}\mathbf{H}^{T} = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left[\mathbf{H}\cdot\left(\mathbf{T}_{j,k+1}^{f} - \overline{\mathbf{T}}_{k+1}^{f}\right)\right]\left[\mathbf{H}\cdot\left(\mathbf{T}_{j,k+1}^{f} - \overline{\mathbf{T}}_{k+1}^{f}\right)\right]^{T} \quad (7)$$

where $\mathbf{T}_{j,k+1}^{a}$ and $\mathbf{T}_{j,k+1}^{f}$ are the analyzed and forecasted state variables of the jth member at time k+1, respectively. $\mathbf{K}$ is the Kalman gain matrix. Y_{k+1}^{o} is the LAI, SM, and SIF observations at time k+1. The LAI, SM, and SIF observations at updated time k+1 are used in the EnKF system to update the model states and parameters. P^{f} is the error covariance matrix of the forecast model states (LFMASS and SM) and parameters (SLA and Vcmax), and $\overline{\mathbf{T}}_{k+1}^{f}$ represents the mean of forecasted state variables at time k+1. $[\cdot]^{T}$ represents the transposed matrix, and N_e is the number of ensembles. In this study, the ensemble size for the MLDAS run is set as 30.

In MLDAS, the model state variables and parameters are updated separately using Equations 4–7. In Passes I, II, III, and IV, the state variable $\mathbf{T}$ in Equation 1 represent SLA, LFMASS, four-layer SM, and Vcmax, respectively.

To apply the EnKF technique, the model ensembles should be perturbed by adding normally distributed random errors to the model states. Random errors within a physically reasonable range reflect the uncertainty in observations, background conditions, and propagated states (Bateni & Entekhabi, 2012). In this study, the model ensembles are obtained by adding normally distributed random errors to the model states (LFMASS and SM) of Passes I and III and the model parameters (SLA and Vcmax) of Passes II and IV in Equations 1 and 2. Additive perturbations with a standard deviation of 10 g m^{-2} are applied to the modeled LFMASS (Kumar et al., 2019). The standard deviations of the four-layer SM (0–0.1, 0.1–0.4, 0.4–1, and 1–2 m) are 0.03, 0.03, 0.025, and 0.02 m^3 m^{-3}, respectively (T. Xu et al., 2015). For SLA and Vcmax, the model parameter standard deviations are set to 10% of the range of the value (T. Xu et al., 2011, 2015). For the observation errors, standard deviations of 0.1, 0.05 m^3 m^{-3}, and 0.15 W m^{-2} nm^{-1} sr^{-1} are used as empirical values for the LAI, SM, and SIF observations, respectively. These standard deviation values are comparable to those values used in prior LAI, SM, and SIF DA studies (Kumar et al., 2019; Norton et al., 2018; Sabater et al., 2008; T. Xu et al., 2011, 2015; K. Yang et al., 2007; Zhang, Joiner, et al., 2018). More details about the MLDAS are listed in Table 1.

3. Experimental Sites and Data Set

The developed MLDAS is tested at the Bondville (US-Bo1) and Mead (US-Ne1) sites. The Mead (US-Ne1, Nebraska) site is covered by corn and irrigated with a central pivot system. Bondville (US-Bo1, Illinois) is characterized by rainfed with annual rotation between corn and soybean. The simulation period for the Mead site is from 2003 to 2005, and for the corn year (2001, 2003, and 2005) at the Bondville site.

Half-hourly (Bondville) and hourly (Mead) meteorological forcing data, such as incoming shortwave and longwave radiation, wind speed, precipitation, air pressure, temperature, and humidity, are used to drive

the Noah-MP-Crop model (http://ameriflux.lbl.gov/). *H*, LE, and net ecosystem exchange (NEE) are recorded by the EC flux towers (http://ameriflux.lbl.gov/). GPP is obtained as the difference between the measured daytime NEE and estimated ecosystem respiration (Hirano et al., 2003; Janssens et al., 2001; Saigusa et al., 2002). The total aboveground biomass, LFMASS, and grain biomass observations are obtained after the plant material is dried to a constant temperature of 105°C (Williams, Lu, et al., 2016). Soil moisture is measured at depths of 0.02, 0.05, and 0.10 m, which is averaged to be consistent with the topsoil layer of Noah-MP-Crop. The LAIs are measured by destructive sampling methods (Ma et al., 2017) and are available for the Mead and Bondville sites from the AmeriFlux databases. The SIF products (with a 5-km spatial resolution and 4-day temporal frequency) are available in the contiguous solar-induced chlorophyll fluorescence archive (Zhang, Joiner, et al., 2018; https://osf.io/8xqy6/?view_only=4d14a64b7be644d39564c82a371e1d20). The measured LFMASS and LAI are used to calculate SLA. The observed *H*, LE, GPP, and aboveground biomass observations are used to evaluate MLDAS simulations. The main soil textures at the Bondville and Mead sites are silt loam and clay loam, respectively (X. Liu et al., 2016).

The simulations are conducted for five months (May 1 to September 30, i.e., DOY 121–273) covering the corn growing season. The LAI observations are assimilated into the Noah-MP-Crop when they are available. The daily averaged (00:00–24:00 local time) SM observations are assimilated every day, and the SIF observations are assimilated every 4 days. All these observations are assimilated into the Noah-MP-Crop model at 00:00 (local time); in addition, the model state variables (LFMASS and four-layer SM) are updated and model parameters (SLA and Vcmax) are constrained. To evaluate the performance of MLDAS, mean absolute error (MAE), root mean square deviation (RMSD), and correlation coefficient (*R*) metrics are used in this study.

4. Results and Discussion

4.1. Assimilation of LAI Observations

Figure 2 compares observations with the LAI estimates from the three assimilation strategies and open loop simulation. EnOL has the largest RMSD, while the three LAI assimilation strategies ensemble Kalman filter-leaf area index (EnKF-LAI-I, EnKF-LAI-II, and EnKF-LAI-III) have much lower RMSDs. Among the three LAI assimilation strategies, EnKF-LAI-III performs best because the DA method from EnKF-LAI-III update both LFMASS and optimizes SLA with the information contained in LAI observations, and retrieves the statistically optimal value for LFMASS and LAI. Moreover, the LAI estimates from EnKF-LAI-III have higher *R* and lower RMSD values than those of EnOL, EnKF-LAI-I, and EnKF-LAI-II at Mead and Bondville. The RMSDs of the EnOL LAI retrievals at the Mead and Bondville sites are 0.73 and 0.62 m^2 m^{-2}, respectively. EnKF-LAI-III reduces the RMSD by 46.58% and 43.55% for the Mead and Bondville sites, respectively.

As shown in Figure 3, EnOL significantly underestimated the LAI during the growing season at the Mead (2004, 2005) and Bondville (2001, 2003) sites but slightly overestimated it during the growing season at Bondville (2005), which are primarily due to inaccurate model parameters (e.g., GDD and planting and harvesting dates). In Noah-MP-Crop, dynamic crop growth parameters, such as the accumulated GDD, is used to determine plant growth stages. The planting and harvesting dates are used to determine the duration of the growing season of corn and soybeans (Levis et al., 2012; X. Liu et al., 2016; Z. Zhang et al., 2020). All these parameters are site-specific and empirical and are not easily obtained and applied to various sites. Furthermore, the specification of SLA requires information on the LFMASS, which is rarely measured for a large area. However, the LAI estimates from EnKF-LAI-III can capture both the seasonal and daily variations in the measurements with parameter optimization and are closer to the measurements than those from the EnOL approach.

Figure 4 shows that the optimized SLA from EnKF-LAI-III is close to the measured values at the two sites. The estimated SLA decreases from 0.03 m^2 g^{-1} at the early stage of the growing season to 0.01 m^2 g^{-1} at the end stage of the growing season, because the corn leaves grow rapidly at the beginning of the growing season (larger LAI) with a small amount of biomass; later, the leaves become thicker (larger biomass), while the LAI increases slightly (Z. Zhang et al., 2020). Overall, the key vegetation parameters of Noah-MP-Crop (i.e., SLA) can be efficiently constrained by assimilating LAI observations. The range of the retrieved

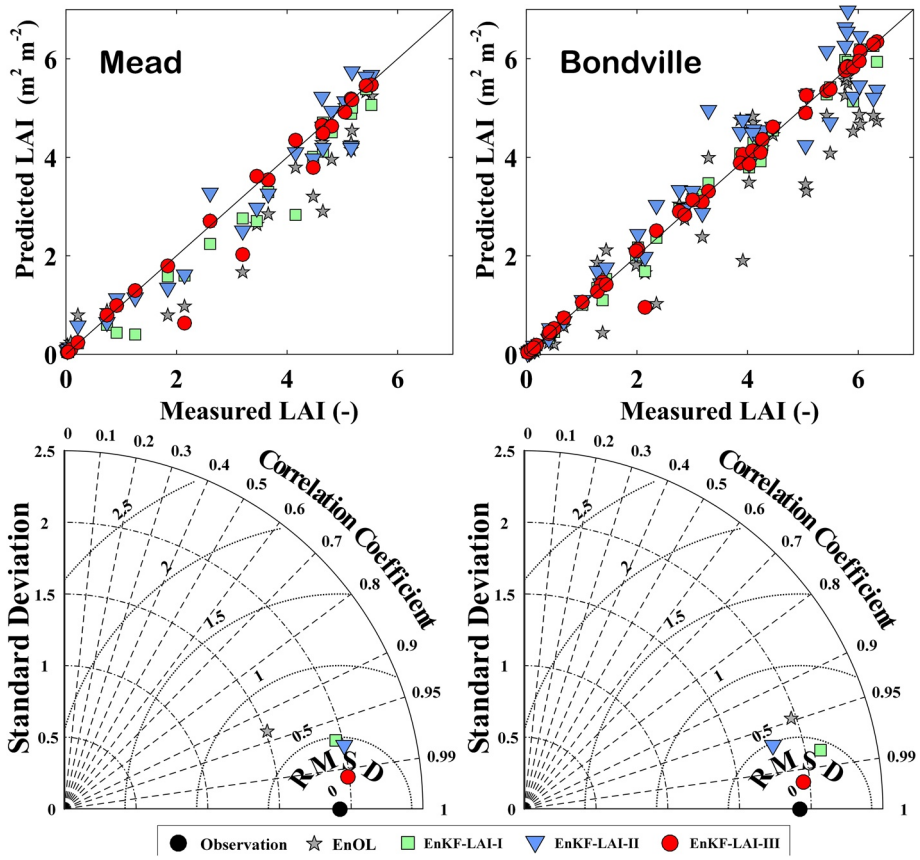

Figure 2. Scatter plots of the estimated daily leaf area index (LAI) values versus the observations (first row) and the statistical metrics of the estimated daily LAI (second row) with the three LAI assimilation strategies.

SLA values in this study is comparable to those reported in the literature (e.g., Bateni et al., 2014; Y. Li et al., 2005; Williams et al., 2017).

4.2. Assimilation of Soil Moisture Observations

As shown in Figure 5, simulated SM from the EnOL method is remarkably lower than observations during the growing season at Bondville (2003) but slightly overestimated during the growing season at Mead (2003, 2005). Both the daily variations and the magnitudes of the SM estimates from the EnKF-SM agree well with observations, which also captures main soil-wetting events (as a response to precipitation) and drying spells at the two sites. EnKF-SM shows slightly worse performance in terms of SM estimates for Bondville in 2003 than in 2001 and 2005. This result occurred because the use of time-invariant SM observation errors and model errors through the modeling period sometimes did not take advantage of the information in the SM observations and model estimates.

Table 2 shows that the two-site average MAE (RMSD) of the SM estimates from EnKF-SM is 0.03 $m^3\ m^{-3}$ (0.04 $m^3\ m^{-3}$), which is 40.00% (63.64%) lower than the MAE (RMSD) of 0.05 $m^3\ m^{-3}$ (0.11 $m^3\ m^{-3}$) from EnOL.

4.3. Assimilation of SIF

In the MLDAS, the linear relationship between SIF and GPP (SIF = $a \times$ GPP + b) was used as the observation operator. The spatial resolutions of GPP measurements from the EC systems and the SIF from satellites are approximately 500–1,000 and 5,000 m (B. Chen et al., 2008, 2011), respectively. Therefore, the linear observation operator includes not only photosynthesis information determined by SIF but also spatial-scale

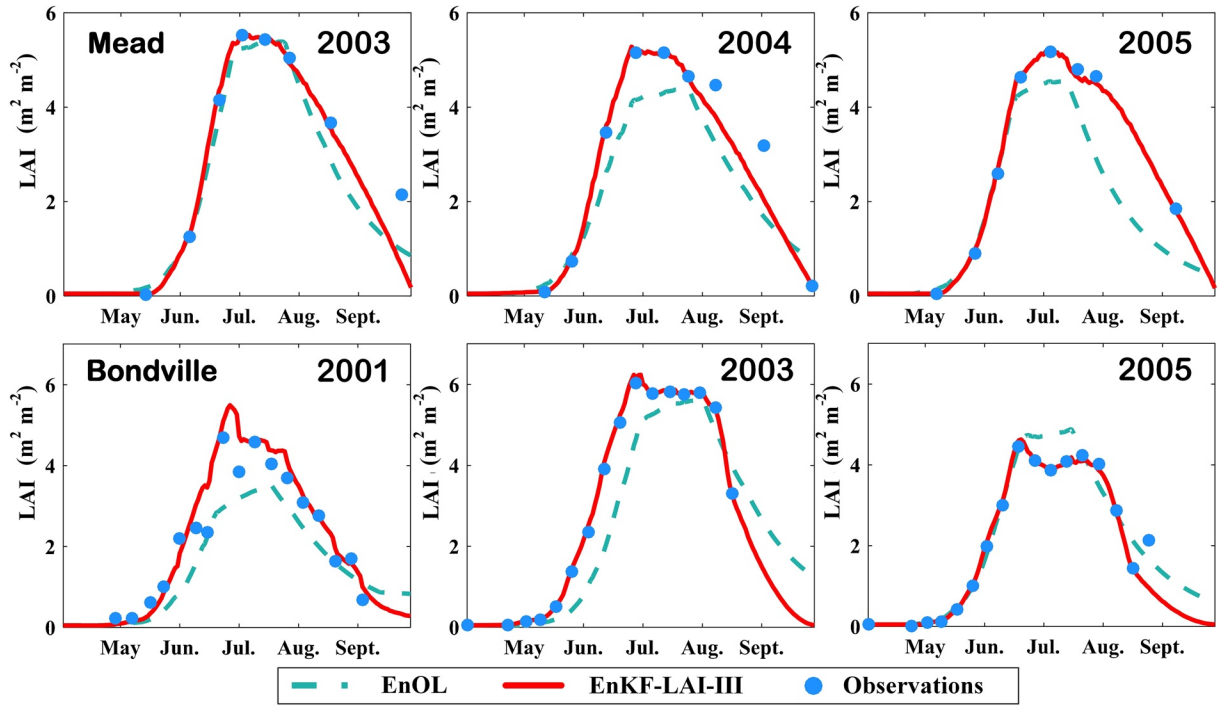

Figure 3. Comparisons of the leaf area index (LAI) values estimated from ensemble Kalman filter-leaf area index-III (EnKF-LAI-III) and ensemble open loop (EnOL) with the observations at the Mead and Bondville sites.

effects. As shown in Figure 6, in general, GPP increases with increasing SIF at the two sites. The 4-day interval SIF is strongly correlated with GPP for Mead ($R = 0.71$) and Bondville ($R = 0.80$). The GPP/SIF ratio at Mead is larger than that at Bondville, most likely due to irrigation, leading to an approximately 30% higher GPP per unit SIF at Mead than at Bondville.

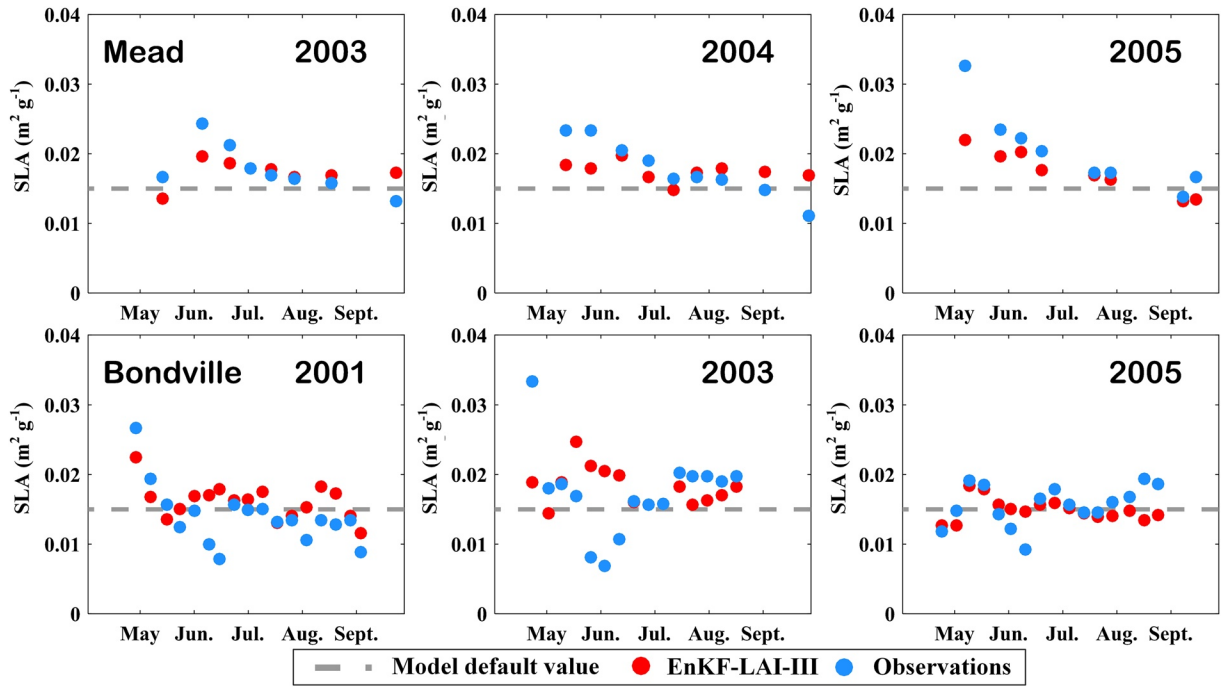

Figure 4. Comparisons of ensemble Kalman filter-leaf area index-III (EnKF-LAI-III) optimized specific leaf area (SLA) with observations at the Mead and Bondville sites.

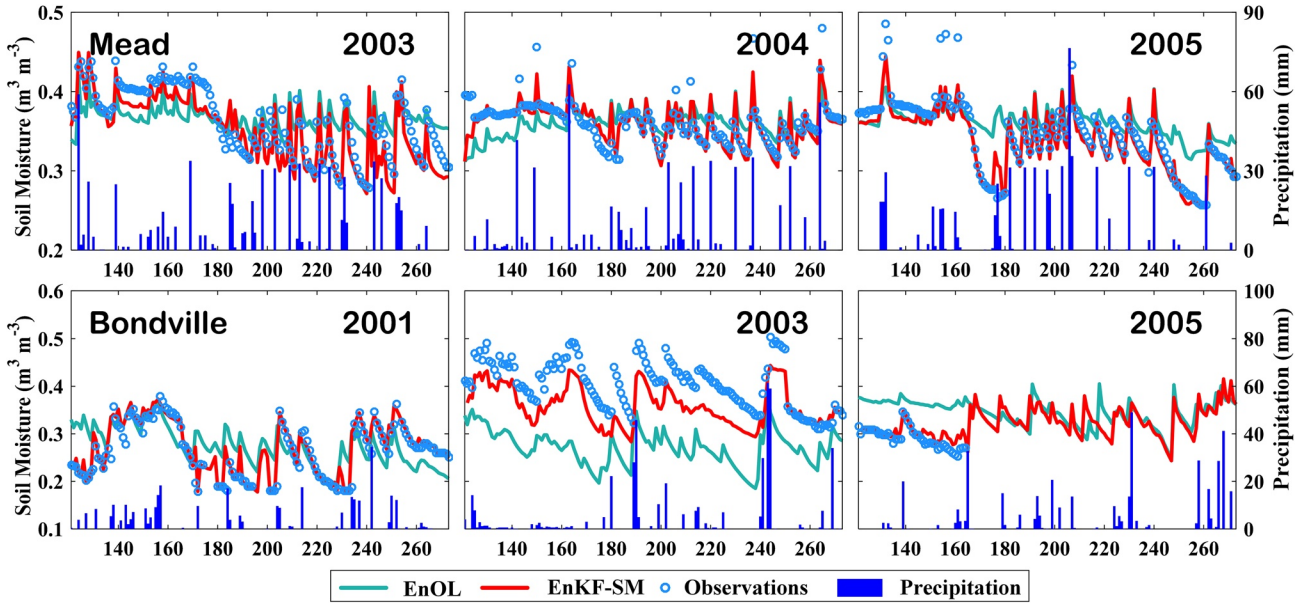

Figure 5. Daily averaged soil moisture (0–10 cm) estimates from ensemble Kalman filter-soil moisture (EnKF-SM) and ensemble open loop (EnOL) versus ground measurements at the Mead and Bondville sites. Blue bars represent the daily total precipitation.

Vcmax, a key parameter in modeling photosynthesis, varies greatly during the growing season and among crop species (Makela et al., 2004; Sellers, 1997) and is set to a wide range of values in different models (from 30 to 150 μmol m^{-2} s^{-1}; Bonan et al., 2011; Z. Zhang et al., 2020). In this study, the dynamic value of Vcmax is updated by assimilating the SIF observations. As shown in Figure 7, Vcmax reveals seasonal variations depending on the vegetation conditions. In comparison to other times, in the middle of the growing season (DOYs 180–220), Vcmax values are higher and produce intense photosynthesis. The dynamic range of Vcmax retrievals is comparable to those of Hu et al. (2014), Y. Liu (2019), and Misson et al. (2006).

4.4. Estimates of Water-Energy-Carbon Variables With MLDAS

This section discusses the analysis of water-energy-carbon variables (i.e., *H*, LE, GPP, aboveground biomass, and grain mass) simulated by MLDAS. As shown in Figure 8, the EnOL approach overestimates (underestimates) *H* (LE) in May and underestimates (overestimates) *H* (LE) in August and September at the Mead site, mainly because the model underestimates (overestimates) SM from May to June (July to September, see Figure 5). In addition, the underestimated LAI from EnOL (July to September; Figure 3) allowed the land surface to absorb more solar radiation, resulting in higher *H* (Hardwick et al., 2015). Compared to the EnOL simulations, the EnKF-ALL simulated diurnal cycles of *H* and LE are closer to the measurements, indicating the benefit of assimilating LAI, SM, and SIF time series in improving the surface fluxes *H* and LE at the Mead site.

Figure 9 shows that the simulated SM from EnOL is much lower than the 2003 observations at the Bondville site, leading to higher *H* and lower LE, and EnKF-ALL performs substantially better. The estimated LE from the EnKF-ALL approach matches the daily variations in precipitation (SM) and LAI. At the Mead and Bondville sites, LE shows a rising trend in the vegetation growing period, reaches its maximum in the middle of July, and decreases as the vegetation senesces. The Mead and Bondville sites have the lowest *H* in July and August as a result of the high LE caused by high LAI and SM availability. In contrast, *H* dominates over LE in May because of sparser vegetation cover.

Table 2
Statistical Indices of the Daily Averaged Soil Moisture Retrievals From the EnOL and EnKF-SM Approaches at the Two Study Sites

	Mead		Bondville	
	EnOL	EnKF-SM	EnOL	EnKF-SM
MAE (m^3 m^{-3})	0.03	0.01	0.07	0.04
RMSD (m^3 m^{-3})	0.09	0.03	0.13	0.05
R (−)	0.72	0.94	0.71	0.92

Abbreviations: EnKF, ensemble Kalman filter; EnOL, ensemble open loop; MAE, mean absolute error; RMSD, root mean square deviation; SM, soil moisture.

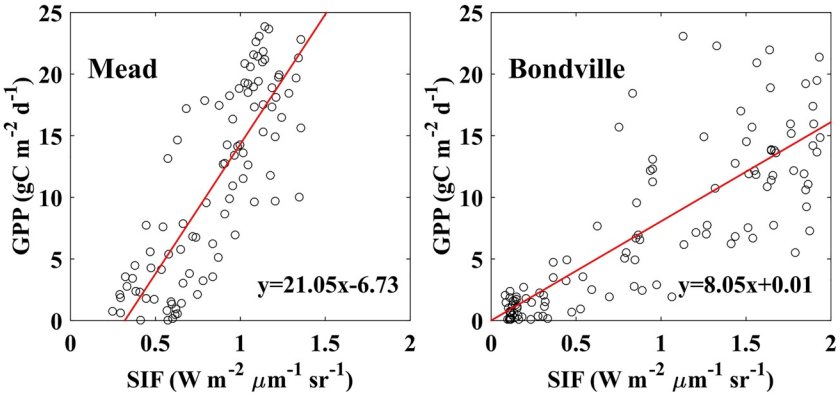

Figure 6. The relationships between gross primary productivity (GPP) and solar-induced chlorophyll fluorescence (SIF). The red line represents the fitted linear regression.

As shown in Table 3, the two-site average MAE (RMSD) of the *H* estimates from EnKF-ALL is 50.56 W m^{-2} (79.42 W m^{-2}), which is 23.00% (27.73%) lower than the MAE (RMSD) of 65.66 W m^{-2} (109.90 W m^{-2}) from EnOL. For LE, the two-site average MAE and RMSD from EnKF-ALL (EnOL) are 54.60 and 89.59 W m^{-2} (68.55 and 116.93 W m^{-2}), respectively.

Figure 10 compares the daily GPP estimates from EnOL and EnKF-ALL with the measurements at the Mead and Bondville sites. The GPP estimates from the EnOL method are remarkably lower than the observations during the growing season at Mead (2004, 2005) and Bondville (2001, 2003). The GPP values are slightly overestimated during the growing season at Bondville (2005), which is consistent with the LAI results from EnOL presented in Figure 3. The discrepancies between the GPP estimates and observations are primarily due to inaccurate model parameters (e.g., GDD, planting and harvesting dates, SLA, and Vcmax). The GPP estimated from the EnKF-ALL approach capture both the seasonal and daily variations in the measurements and is closer to the measurements compared to that from the EnOL approach, implying that the EnKF-ALL approach can exploit implicit information in the LAI and SIF observations to generate more accurate GPP estimates. The estimated GPP values increase continuously from May to June with the growth of the crop, reach their maxima in July (DOYs 182–212) and decrease from July to September because the crop is mature and harvest. The two-site average MAEs (RMSDs) from the EnOL and EnKF-ALL approaches are 3.40 (4.13 gC m^{-2} d^{-1}) and 1.18 gC m^{-2} d^{-1} (1.99 gC m^{-2} d^{-1}), respectively. In comparison

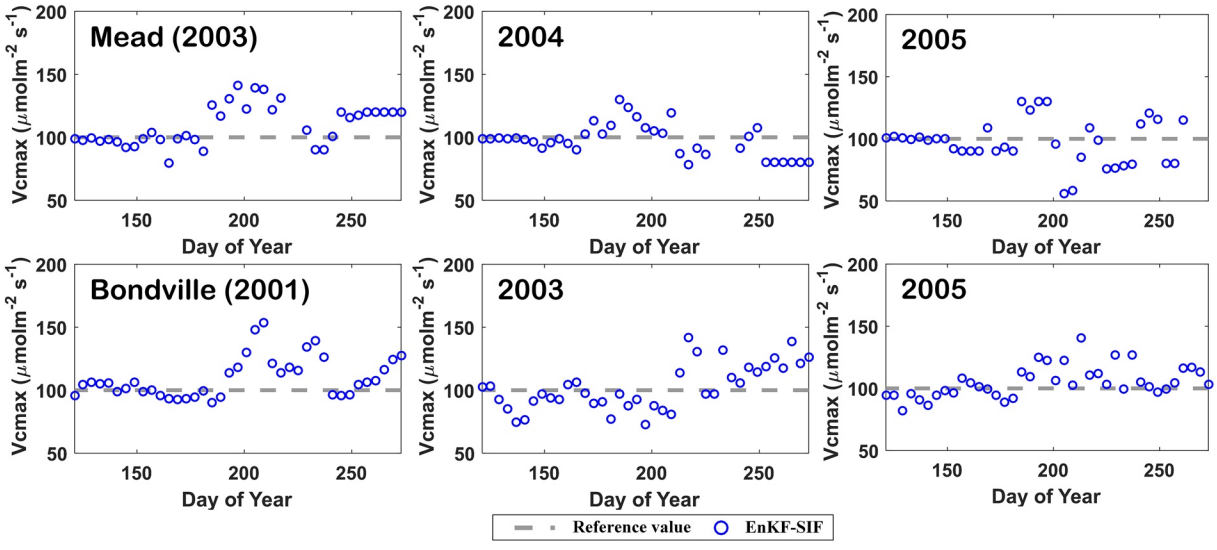

Figure 7. The ensemble Kalman filter-solar-induced chlorophyll fluorescence (EnKF-SIF) optimized maximum rate of carboxylation (Vcmax) values at the Mead and Bondville sites.

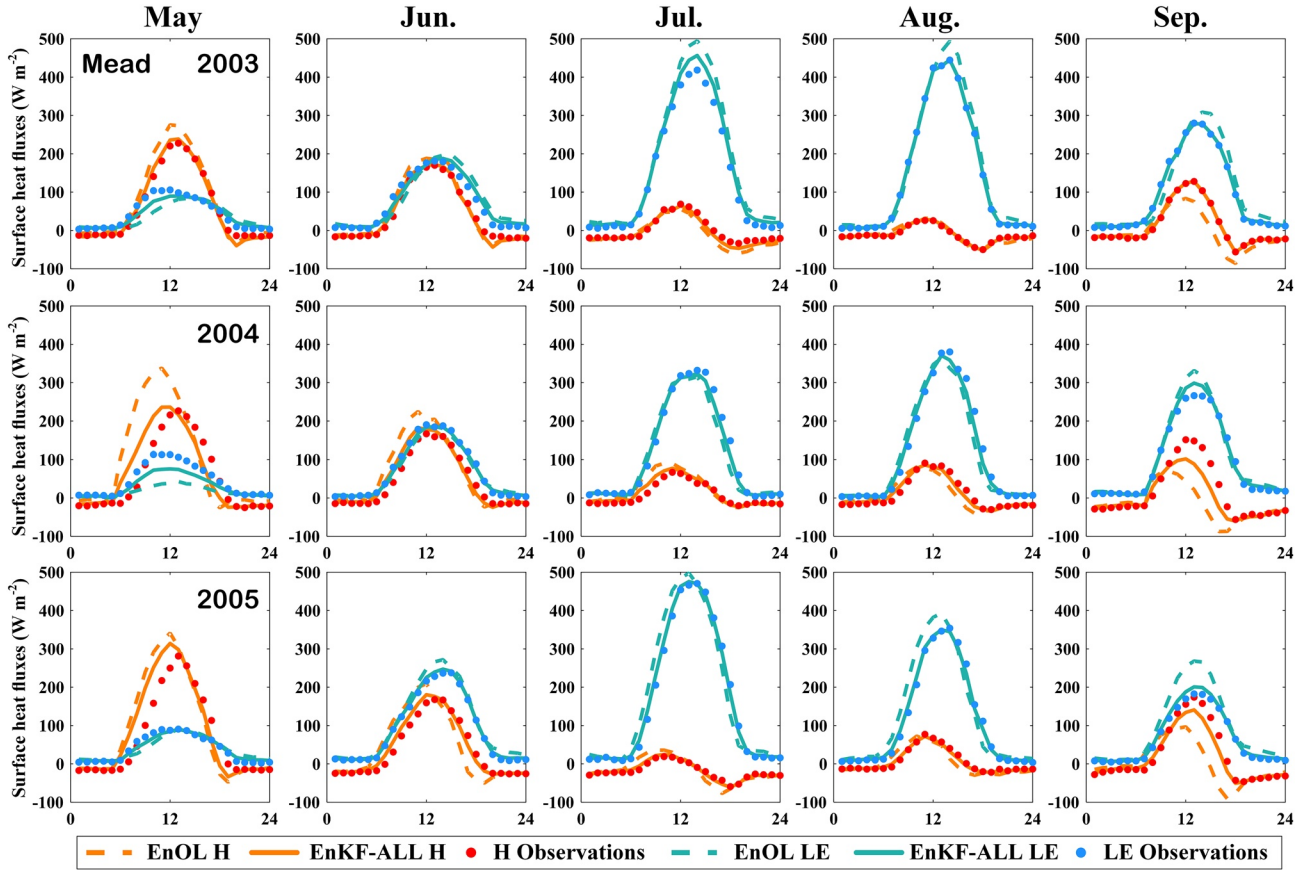

Figure 8. Monthly averaged diurnal cycle of the modeled sensible and latent heat fluxes from ensemble Kalman filter-ALL (EnKF-ALL) (solid lines), ensemble open loop (EnOL) (dashed lines) and corresponding observations (solid dots) at the Mead site.

to the EnOL approach, the EnKF-ALL approach cut down the MAE (RMSD) of the two-site average GPP by 65.29% (51.82%).

Figure 11 shows that the EnOL approach underestimated aboveground biomass at Bondville and overestimated aboveground biomass at Mead, but the performance of the Noah-MP-Crop model is improved with the EnKF-ALL method. The assimilation of LAI, SM, and SIF data effectively captures the rapid crop growth from May to August and the decline in aboveground biomass at the end of August when corn started to ripen and decline at Bondville. The two-site average MAE (RMSD) of aboveground biomass estimates from EnKF-ALL is 100.25 g m^{-2} (153.67 g m^{-2}), which is 39.47% (50.47%) lower than the MAE (RMSD) of 165.63 g m^{-2} (310.25 g m^{-2}) from EnOL. The Noah-MP-Crop model also provides useful crop yield information (grain biomass) for regional-scale agricultural research and management applications (X. Liu et al., 2016; Z. Zhang et al., 2020). The retrieved grain biomass values for Bondville are shown in Figure 11 (bottom line). In Bondville, the EnOL simulation underestimated the grain biomass by approximately 22.30% compared to ground measurements. Compared to the EnOL approach, the EnKF-ALL approach greatly improves corn yield by approximately 23.54%.

The performance of EnKF-ALL is compared with those of EnKF-LAI-III, EnKF-SM, EnKF-SIF, and EnOL, and the results are shown in Figure 12. In comparison to EnOL, EnKF-LAI-III (EnKF-SM) significantly improves the estimation of LAI (SM) at Mead and Bondville. This is because the assimilation of LAI and SM data effectively constrain the parameter (SLA) and state variables (LFMASS and SM) within Noah-MP-Crop.

The RMSD (R) values of the H, LE, and GPP retrievals from the EnKF-ALL approach are lower (higher) than those of EnOL, implying that EnKF-ALL can make full use of information within sequences of LAI, SM, and SIF to predict more accurate H, LE, and GPP. EnKF-SM significantly improve the H and LE re-

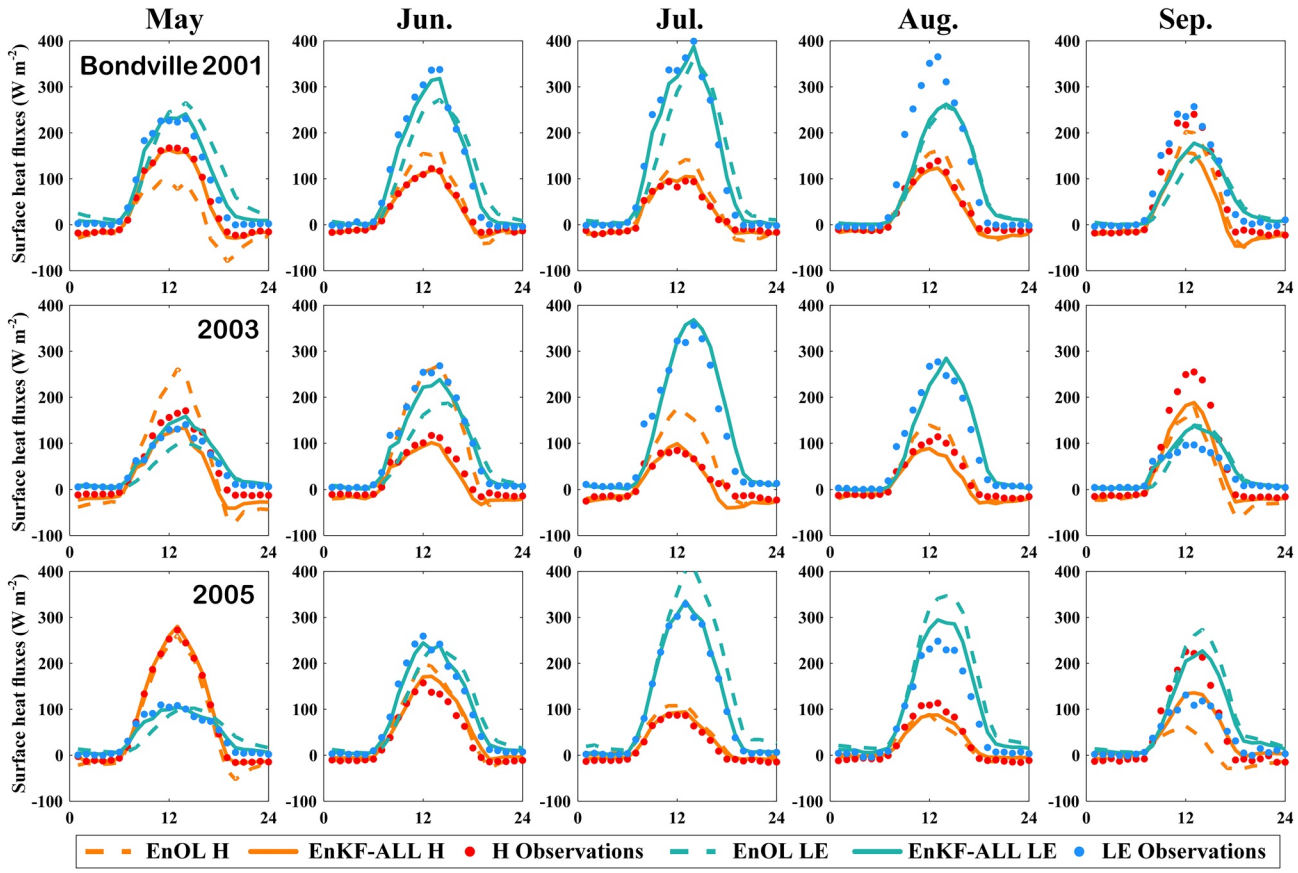

Figure 9. Monthly averaged diurnal cycle of the modeled sensible and latent heat fluxes from ensemble Kalman filter-ALL (EnKF-ALL) (solid lines), ensemble open loop (EnOL) (dashed lines) and corresponding observations (solid dots) at the Bondville site.

trievals, indicating that sequences of SM are of vital importance for *H* and LE retrievals. Similarly, EnKF-LAI-III and EnKF-SIF appropriately improves the *H* and LE retrievals, indicating that changes in vegetation phenology (e.g., LAI dynamics and Vcmax) also affects the partitioning of the available energy between the turbulent heat fluxes (*H* and LE). EnKF-LAI-III, and EnKF-SIF significantly improve the GPP estimates at the two study sites, indicating that changes in vegetation dynamics are vitally important for GPP retrievals. However, the GPP estimates from EnKF-SM are comparable to those from EnOL, which means that GPP is less sensitive to variations in SM in the Noah-MP-Crop model.

Table 3
The MAE, RMSE, and R of Half-Hourly (Hourly) H and LE Retrievals From the EnOL and EnKF-ALL Approaches at the Two Study Sites

		H (W m^{-2})		LE (W m^{-2})	
Study sites		EnOL	EnKF-ALL	EnOL	EnKF-ALL
Mead	MAE (W m^{-2})	57.59	38.52	55.57	45.29
	RMSD (W m^{-2})	102.43	76.19	108.67	88.30
	R (−)	0.66	0.81	0.72	0.83
Bonville	MAE (W m^{-2})	73.72	62.60	81.52	63.90
	RMSD (W m^{-2})	117.36	82.65	125.19	90.87
	R (−)	0.63	0.77	0.64	0.81
Two-sites average	MAE (W m^{-2})	65.66	50.56	68.55	54.60
	RMSD (W m^{-2})	109.90	79.42	116.93	89.59
	R (−)	0.65	0.79	0.68	0.82

Abbreviations: EnKF, ensemble Kalman filter; EnOL, ensemble open loop; *H*, sensible heat flux; LE, latent heat flux; MAE, mean absolute error; RMSD, root mean square deviation.

To examine the relative contributions of different assimilation strategy to the overall results, the differences in the simulations between EnKF (EnKF-LAI-III, EnKF-SM, EnKF-SIF, and EnKF-ALL) and EnOL are shown in Figure 13. The statistical results are averaged from three-year simulations at the two study sites. The assimilation of LAI and SIF led to increased LAI from July to September. The largest increase in LAI is observed from June to August. As LAI changes, corresponding changes in GPP and LFMASS are also found in the model estimates. The assimilation of LAI leads to slightly increased LE because changes in vegetation dynamics will affect vegetation transpiration (Kumar et al., 2019). The assimilation of SM has a large impact on SM, *H* and LE. Increased (decreased) SM in May (from June to September) leads to decreased (increased) *H* and increased (decreased) LE. The decrease in SM also leads to slightly decreased LAI, LFMASS, and GPP (June and July) because

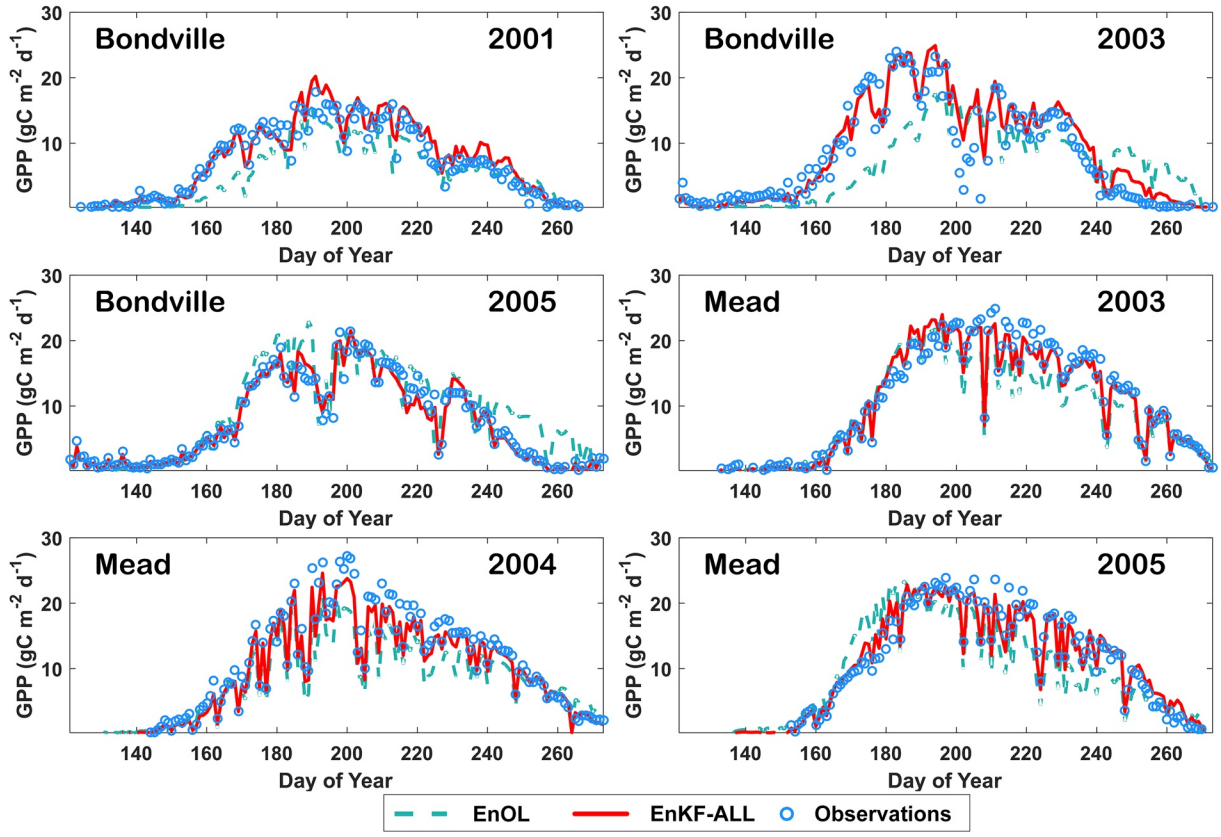

Figure 10. Time series of daily gross primary productivity (GPP) retrievals from the ensemble Kalman filter-ALL (EnKF-ALL) (red solid lines) and ensemble open loop (EnOL) (green dashed lines) approaches at the two study sites. The GPP measurements are indicated by open circles.

of its effects on canopy stomatal resistance and photosynthesis (Zhuo et al., 2019). In contrast, EnKF-ALL showed the opposite trend in LAI, LFMASS, and GPP in July. This is because the assimilation of one variable will be offset by the assimilation of the other variable, and the assimilation of multiple pairs of variables will produce indistinguishable results (Gao et al., 2015; Schneider et al., 2017).

5. Conclusions

In this study, the MLDAS was developed based on the Noah-MP-Crop and EnKF. The LAI, SM, and SIF observations are simultaneously assimilated into the Noah-MP-Crop model within a MLDAS to improve the simulated *H*, LE, and GPP. The performance of the MLDAS is tested at two AmeriFlux cropland sites (Mead and Bondville), and the main results are as follows:

1. The assimilation of LAI helps update LFMASS and optimize SLA and is able to retrieve the statistically optimal value for the LFMASS and LAI predictions. The RMSD of the LAI estimates from EnKF-LAI-III at the Mead (Bondville) site are 0.39 (0.35), representing a reduction of 52.05% (43.55%) in the error compared to the EnOL. In comparison to EnOL, EnKF-LAI-III can capture both the observed seasonal and daily LAI variations with parameter optimization and is closer to the measurements, and the optimized SLA values are close to the ground measurements at the two sites.
2. The SM assimilation results show a more reasonable response to land surface wetting and drying events at the two sites. Compared to EnOL, assimilating surface SM reduces the two-site average MAE (RMSD) of the SM estimates by 40.00% (63.64%).
3. Assimilating SIF helps optimize the seasonal variations in the Vcmax values throughout the whole vegetation growing season.

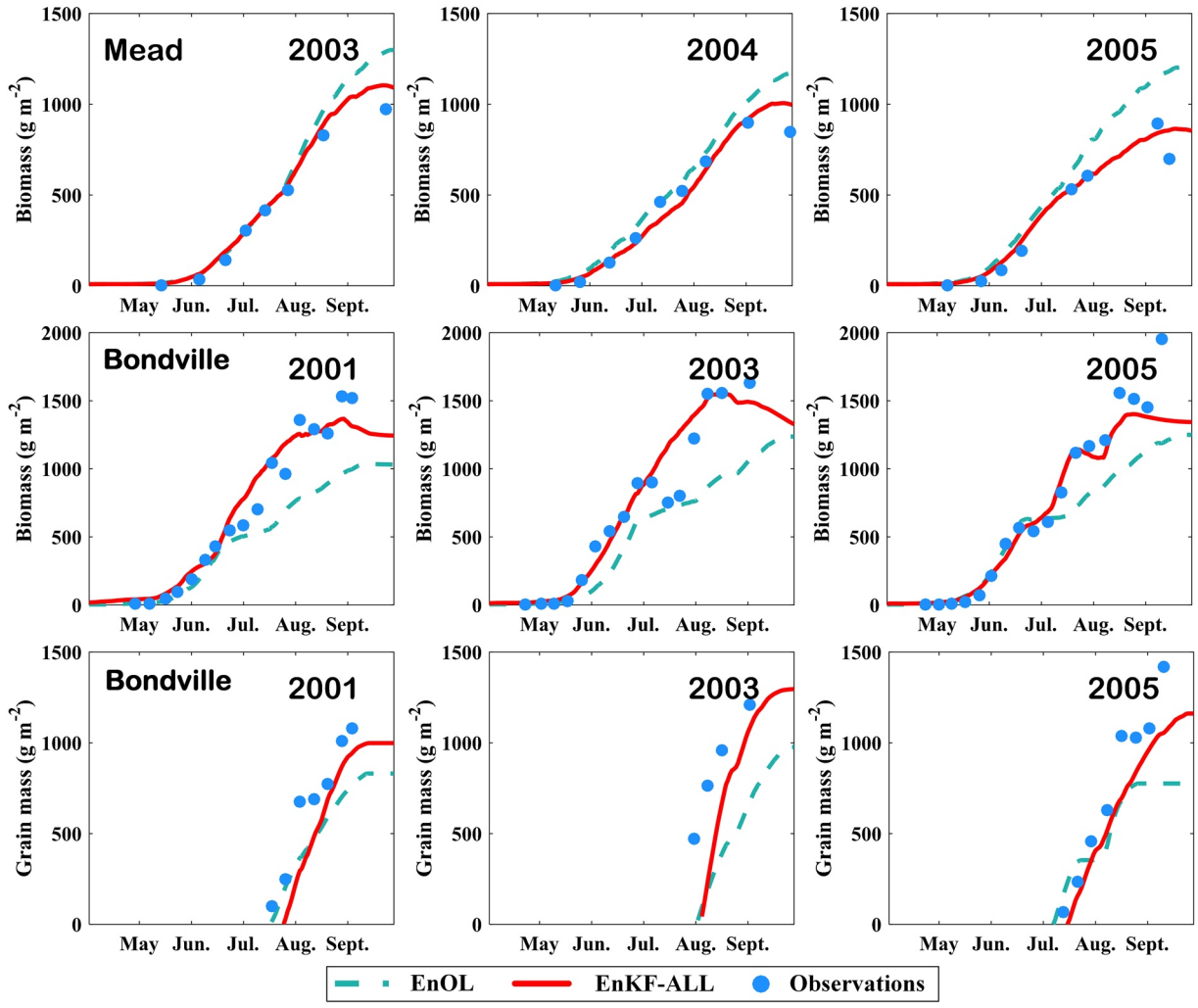

Figure 11. Comparisons of simulated aboveground biomass and grain biomass at the Mead and Bondville sites between the ensemble open loop (EnOL) and ensemble Kalman filter-ALL (EnKF-ALL) experiments.

4. In comparison to those from EnOL, the simulated day-to-day fluctuations in the *H*, LE, and GPP estimates from EnKF-ALL (by jointly assimilating LAI, SM, and SIF) is more consistent with the observations, which reduced the two-site average RMSD of *H* from 109.90 with EnOL to 79.42 W m^{-2} with EnKF-ALL. For LE, the two-site average RMSD from EnKF-ALL (EnOL) is 89.59 W m^{-2} (116.93 W m^{-2}). The two-site average RMSD of daily GPP retrievals from EnKF-ALL is 1.99 gC m^{-2} d^{-1}, which is 51.82% lower than the RMSD of 4.13 gC m^{-2} d^{-1} from EnOL. Moreover, EnKF-ALL also effectively constrained aboveground biomass variations and improved corn yield estimates.

The performance of EnKF-ALL is compared with those of the EnKF-LAI-III (only LAI assimilation), EnKF-SM (only SM assimilation), EnKF-SIF (only SIF assimilation), and EnOL approaches. The results show that the assimilation of SM significantly improves the *H* and LE retrievals, while the assimilation of LAI and SIF significantly improves the GPP retrievals. Future studies should focus on applying the developed MLDAS approach over large-scale domains using remotely sensed LAI, SM, and SIF data.

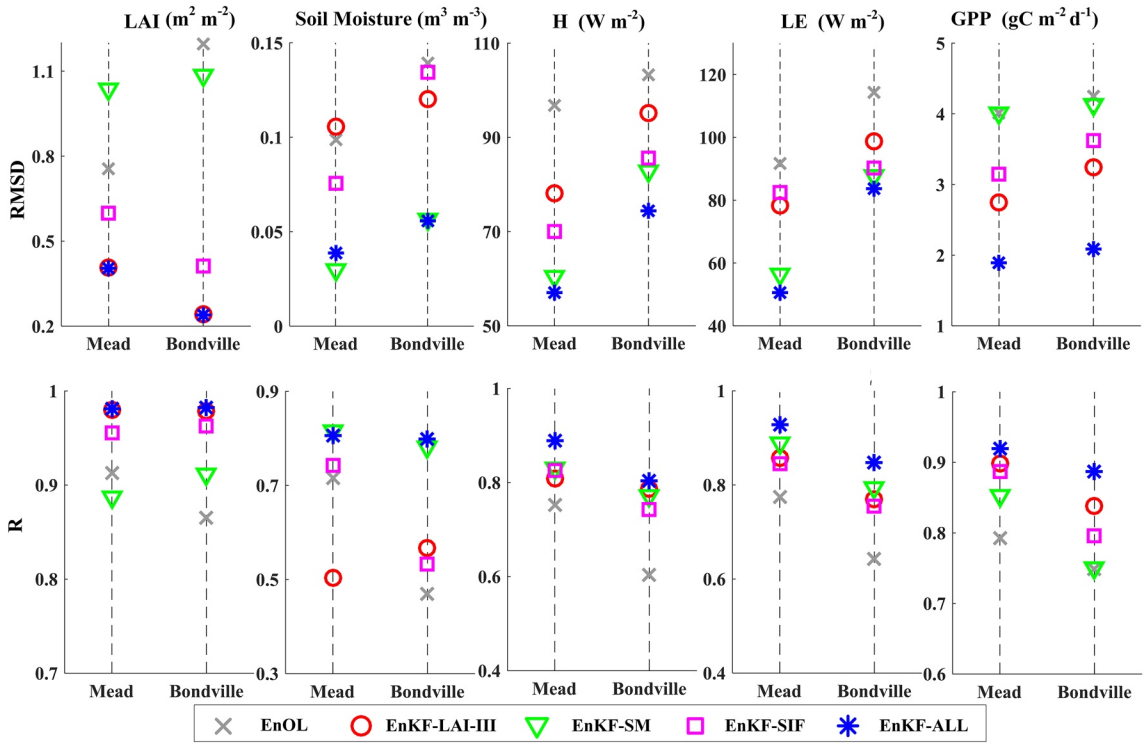

Figure 12. Comparison of the daily leaf area index (LAI), soil moisture (SM), sensible and latent heat flux (*H* and LE), and gross primary productivity (GPP) estimates from ensemble open loop (EnOL), ensemble Kalman filter (EnKF)-LAI-III (only assimilating LAI data), EnKF-SM (only assimilating SM data), EnKF-SIF (only assimilating SIF data), and EnKF-ALL (assimilating LAI, SM, and SIF data) at the Mead and Bondville sites.

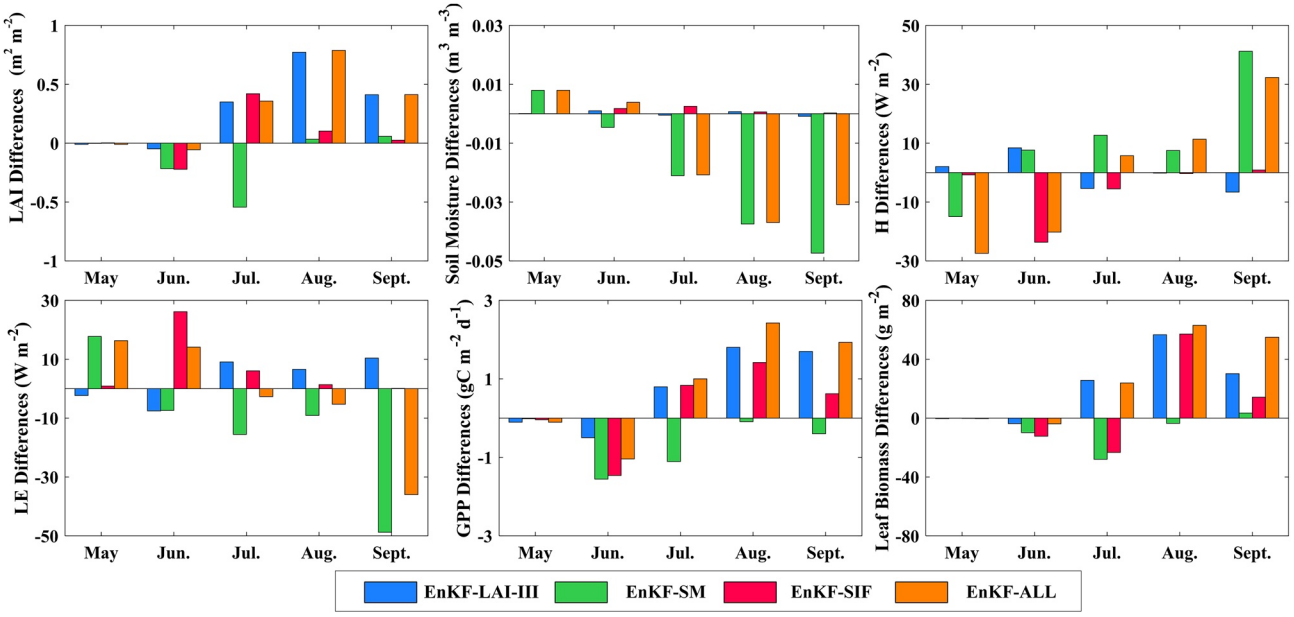

Figure 13. The model variable (leaf area index [LAI], soil moisture [SM], sensible and latent heat flux [*H* and LE], gross primary productivity [GPP], and leaf biomass) differences between the ensemble Kalman filter (EnKF; EnKF-LAI-III, EnKF-SM, EnKF-SIF, and EnKF-ALL) and EnOL methods.

Data Availability Statement

The in situ meteorological variables, turbulent heat fluxes, LAI, and biomass data are downloaded freely from the AmeriFlux Data Center (http://ameriflux.lbl.gov/). The SIF products can be downloaded from the contiguous solar-induced chlorophyll fluorescence (CSIF) archive (https://osf.io/8xqy6/?view_only=4d14a64b7be644d39564c82a371e1d20).

Acknowledgments
This study is supported by the Water System Program at the National Center for Atmospheric Research (NCAR), USDA NIFA Grants 2015-67003-23508 and 2015-67003-23460, NOAA Grants NA18OAR4590381 and NA18OAR4590398, and NSF Grant INFEWS #1739705, and project supported by State Key Laboratory of Earth Surface Processes and Resource Ecology (2021-ZD-04). We would like to thank the high-performance computing support from the Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal University (https://gda.bnu.edu.cn/).

References

Anderson, J. L. (2001). An Ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. https://doi.org/10.1175/1520-0493(2001)129<2884:aeakff>2.0.co;2

Bateni, S. M., & Entekhabi, D. (2012). Surface heat flux estimation with the ensemble Kalman smoother: Joint estimation of state and parameters. *Water Resources Research*, *48*(8). https://doi.org/10.1029/2011wr011542

Bateni, S. M., Entekhabi, D., Margulis, S., Castelli, F., & Kergoat, L. (2014). Coupled estimation of surface heat fluxes and vegetation dynamics from remotely sensed land surface temperature and fraction of photosynthetically active radiation. *Water Resources Research*, *50*(11), 8420–8440. https://doi.org/10.1002/2013wr014573

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., et al. (2011). Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research*, *116*(G2). https://doi.org/10.1029/2010jg001593

Bryant, J. J. (2011). *Biophysical effects on U.S. soybean agroecosystems from increasing carbon dioxide concentration*. Dissertations & theses. Retrieved from University of Minnesota Digital Conservancy. Retrieved from https://hdl.handle.net/11299/104137

Camino, C., Gonzalez-Dugo, V., Hernandez, P., & Zarco-Tejada, P. J. (2019). Radiative transfer Vcmax estimation from hyperspectral imagery and SIF retrievals to assess photosynthetic performance in rainfed and irrigated plant phenotyping trials. *Remote Sensing of Environment*, *231*, 111186. https://doi.org/10.1016/j.rse.2019.05.005

Chen, B., Black, T. A., Coops, N. C., Hilker, T., (Tony) Trofymow, J. A., & Morgenstern, K. (2008). Assessing tower flux footprint climatology and scaling between remotely sensed and eddy covariance measurements. *Boundary-Layer Meteorology*, *130*(2), 137–167. https://doi.org/10.1007/s10546-008-9339-1

Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Barr, A. G., et al. (2011). Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada Research Network based on remote sensing and footprint modeling. *Agricultural and Forest Meteorology*, *151*(1), 87–100. https://doi.org/10.1016/j.agrformet.2010.09.005

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I. Model implementation and sensitivity. *Monthly Weather Review*, *129*, 569–585. https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2

Chen, F., Manning, K. W., LeMone, M. A., Trier, S. B., Alfieri, J. G., Roberts, R., et al. (2007). Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system. *Journal of Applied Meteorology and Climatology*, *46*(6), 694–713. https://doi.org/10.1175/jam2463.1

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., et al. (1996). Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research*, *101*, 7251–7268. https://doi.org/10.1029/95JD02165

Cornelissen, J. H. C., Lavorel, S., Garnier, E., Díaz, S., Buchmann, N., Gurvich, D. E., et al. (2003). A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian Journal of Botany*, *51*(4), 335. https://doi.org/10.1071/bt02124

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., et al. (2003). The Common Land Model. *Bulletin of the American Meteorological Society*, *84*(8), 1013–1024. https://doi.org/10.1175/bams-84-8-1013

Drewniak, B., Song, J., Prell, J., Kotamarthi, V. R., & Jacob, R. (2012). Modeling agriculture in the Community Land Model. *Geoscientific Model Development Discussions*, *5*(4), 4137–4185. https://doi.org/10.5194/gmdd-5-4137-2012

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., et al. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, *108*(D22). https://doi.org/10.1029/2002JD003296

Farquhar, G. D., Caemmerer, von, S., & Berry, J. A. (1980). A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *Planta*, *149*(1), 78–90. https://doi.org/10.1007/bf00386231

Frankenberg, C., & Berry, J. (2018). Solar induced chlorophyll fluorescence: Origins, relation to photosynthesis and retrieval. *Comprehensive Remote Sensing*, 143–162. https://doi.org/10.1016/b978-0-12-409548-9.10632-3

Gao, H., Li, W. G., Cai, L., Berry, C., & Luo, X. Y. (2015). Parameter estimation in a Holzapfel-Ogden law for healthy myocardium. *Journal of Engineering Mathematics*, *95*(1), 231–248. https://doi.org/10.1007/s10665-014-9740-3

Garnier, E., Shipley, B., Roumet, C., & Laurent, G. (2001). A standardized protocol for the determination of specific leaf area and leaf dry matter content. *Functional Ecology*, *15*(5), 688–695. https://doi.org/10.1046/j.0269-8463.2001.00563.x

Gunn, S., Farrar, J. F., Collis, B. E., & Nason, M. (1999). Specific leaf area in barley: Individual leaves versus whole plants. *New Phytologist*, *143*(1), 45–51. https://doi.org/10.1046/j.1469-8137.1999.00434.x

Han, X., Li, X., Franssen, H. J. H., Vereecken, H., & Montzka, C. (2012). Spatial horizontal correlation characteristics in the land data assimilation of soil moisture. *Hydrology and Earth System Sciences*, *16*(5), 1349–1363. https://doi.org/10.5194/hess-16-1349-2012

Hardwick, S. R., Toumi, R., Pfeifer, M., Turner, E. C., Nilus, R., & Ewers, R. M. (2015). The relationship between leaf area index and microclimate in tropical forest and oil palm plantation: Forest disturbance drives changes in microclimate. *Agricultural and Forest Meteorology*, *201*, 187–195. https://doi.org/10.1016/j.agrformet.2014.11.010

He, L., Chen, J. M., Liu, J., Zheng, T., Wang, R., Joiner, J., et al. (2019). Diverse photosynthetic capacity of global ecosystems mapped by satellite chlorophyll fluorescence measurements. *Remote Sensing of Environment*, *232*, 111344. https://doi.org/10.1016/j.rse.2019.111344

He, L., Magney, T., Dutta, D., Yin, Y., Köhler, P., Grossmann, K., et al. (2020). From the ground to space: Using solar-induced chlorophyll fluorescence to estimate crop productivity. *Geophysical Research Letters*, *47*, e2020GL087474. https://doi.org/10.1029/2020GL087474

He, X., Xu, T., Bateni, S., Neale, C., Auligne, T., Liu, S., et al. (2018). Evaluation of the weak constraint data assimilation approach for estimating turbulent heat fluxes at six sites. *Remote Sensing*, *10*(12), 1994. https://doi.org/10.3390/rs10121994

He, X., Xu, T., Bateni, S. M., Ek, M., Liu, S., & Chen, F. (2020). Mapping regional evapotranspiration in cloudy skies via variational assimilation of All-Weather land surface temperature observations. *Journal of Hydrology*, *585*, 124790. https://doi.org/10.1016/j.jhydrol.2020.124790

He, X., Xu, T., Bateni, S. M., Neale, C. M. U., Liu, S., Auligne, T., et al. (2019). Mapping regional turbulent heat fluxes via assimilation of MODIS land surface temperature data into an Ensemble Kalman smoother framework. *Earth and Space Science*, *6*, 2423–2442. https://doi.org/10.1029/2019ea000705

He, X., Xu, T., Xia, Y., Bateni, S. M., Guo, Z., Liu, S., et al. (2020). A Bayesian Three-Cornered Hat (BTCH) method: Improving the terrestrial evapotranspiration estimation. *Remote Sensing*, *12*(5), 878. https://doi.org/10.3390/rs12050878

Hirano, T., Hirata, R., Fujinuma, Y., Saigusa, N., Yamamoto, S., Harazono, Y., et al. (2003). CO_2 and water vapor exchange of a larch forest in northern Japan. *Tellus B: Chemical and Physical Meteorology*, *55*(2), 244–257. https://doi.org/10.1034/j.1600-0889.2003.00063.x

Hu, S., Mo, X., & Lin, Z. (2014). Optimizing the photosynthetic parameter Vcmax by assimilating MODIS-fPAR and MODIS-NDVI with a process-based ecosystem model. *Agricultural and Forest Meteorology*, *198–199*, 320–334. https://doi.org/10.1016/j.agrformet.2014.09.002

Huang, C. L., Li, X., & Lu, L. (2008). Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. *Remote Sensing of Environment*, *112*(4), 1320–1336. https://doi.org/10.1016/j.rse.2007.03.028

Ines, A. V. M., Das, N. N., Hansen, J. W., & Njoku, E. G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment*, *138*, 149–164. https://doi.org/10.1016/j.rse.2013.07.018

Janssens, I. A., Lankreijer, H., Matteucci, G., Kowalski, A. S., Buchmann, N., Epron, D., et al. (2001). Productivity overshadows temperature in determining soil and ecosystem respiration across European forests. *Global Change Biology*, *7*(3), 269–278. https://doi.org/10.1046/j.1365-2486.2001.00412.x

Koster, R. D., Liu, Q., Mahanama, S. P. P., & Reichle, R. H. (2018). Improved hydrological simulation using SMAP data: Relative impacts of model calibration and data assimilation. *Journal of Hydrometeorology*, *19*(4), 727–741. https://doi.org/10.1175/jhm-d-17-0228.1

Kumar, S. V., Mocko, D. M., Wang, S., Peters-Lidard, C. D., & Borak, J. (2019). Assimilation of remotely sensed Leaf Area Index into the Noah-MP land surface model: Impacts on water and carbon fluxes and states over the Continental U.S. *Journal of Hydrometeorology*, *20*(7), 1359–1377. https://doi.org/10.1175/jhm-d-18-0237.1

Kumar, S. V., Peters-Lidard, C. D., Mocko, D., Reichle, R., Liu, Y., Arsenault, K. R., et al. (2014). Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *Journal of Hydrometeorology*, *15*(6), 2446–2469. https://doi.org/10.1175/jhm-d-13-0132.1

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*, 4245–4287. https://doi.org/10.1029/2018ms001583

Levis, S., Bonan, G. B., Kluzek, E., Thornton, P. E., Jones, A., Sacks, W. J., & Kucharik, C. J. (2012). Interactive crop management in the Community Earth System Model (CESM1): Seasonal influences on land-atmosphere fluxes. *Journal of Climate*, *25*(14), 4839–4859. https://doi.org/10.1175/jcli-d-11-00446.1

Li, X., Du, H., Mao, F., Zhou, G., Chen, L., Xing, L., et al. (2018). Estimating bamboo forest aboveground biomass using EnKF-assimilated MODIS LAI spatiotemporal data and machine learning algorithms. *Agricultural and Forest Meteorology*, *256–257*, 445–457. https://doi.org/10.1016/j.agrformet.2018.04.002

Li, X., Huang, C., Che, T., Jin, R., Wang, S., Wang, J., et al. (2007). Development of a Chinese land data assimilation system: Its progress and prospects. *Progress in Natural Science*, *17*(8), 881–892. https://doi.org/10.1080/10002007088537487

Li, X., Liu, F., & Fang, M. (2020). Harmonizing models and observations: Data assimilation in Earth system science. *Science China Earth Sciences*, *63*, 1059–1068. https://doi.org/10.1007/s11430-019-9620-x

Li, Y., Johnson, D. A., Su, Y., Cui, J., & Zhang, T. (2005). Specific leaf area and leaf dry matter content of plants growing in sand dunes. *Botanical Bulletin of the Academica Sinica*, *46*, 127–134.

Liang, S., & Qin, J. (2008). Data assimilation methods for land surface variable estimation. In S. Liang (Ed.), *Advances in Land Remote Sensing: System, Modeling, Inversion and Applications* (pp. 319–339). Springer. https://doi.org/10.1007/978-1-4020-6450-0_12

Liu, D., & Mishra, A. K. (2017). Performance of AMSR_E soil moisture data assimilation in CLM4.5 model for monitoring hydrologic fluxes at global scale. *Journal of Hydrology*, *547*, 67–79. https://doi.org/10.1016/j.jhydrol.2017.01.036

Liu, X., Chen, F., Barlage, M., & Niyogi, D. (2020). Implementing dynamic rooting depth for improved simulation of soil moisture and land surface feedbacks in Noah-MP-Crop. *Journal of Advances in Modeling Earth Systems*, *12*(2). https://doi.org/10.1029/2019ms001786

Liu, X., Chen, F., Barlage, M., Zhou, G., & Niyogi, D. (2016). Noah-MP-Crop: Introducing dynamic crop growth in the Noah-MP land surface model. *Journal of Geophysical Research: Atmospheres*, *121*(23), 13953–13972. https://doi.org/10.1002/2016jd025597

Liu, Y. (2019). *Global retrieval of the maximum carboxylation rate (Vcmax) of plant leaves by solar-induced chlorophyll fluorescence from TROPOMI*. Retrieved from http://hdl.handle.net/1807/100591

Lokupitiya, E., Denning, A. S., Schaefer, K., Ricciuto, D., Anderson, R., Arain, M. A., et al. (2016). Carbon and energy fluxes in cropland ecosystems: A model-data comparison. *Biogeochemistry*, *129*(1–2), 53–76. https://doi.org/10.1007/s10533-016-0219-3

Lokupitiya, E., Denning, S., Paustian, K., Baker, I., Schaefer, K., Verma, S., et al. (2009). Incorporation of crop phenology in Simple Biosphere Model (SiBcrop) to improve land-atmosphere carbon exchanges from croplands. *Biogeosciences*, *6*, 969–986. https://doi.org/10.5194/bg-6-969-2009

Lu, Y., Dong, J., Steele-Dunne, S. C., & van de Giesen, N. (2016). Estimating surface turbulent heat fluxes from land surface temperature and soil moisture observations using the particle batch smoother. *Water Resources Research*, *52*(11), 9086–9108. https://doi.org/10.1002/2016wr018943

Lu, Y., Steele-Dunne, S. C., & De Lannoy, G. J. M. (2020). Improving soil moisture and surface turbulent heat flux estimates by assimilation of SMAP brightness temperatures or soil moisture retrievals and GOES land surface temperature retrievals. *Journal of Hydrometeorology*, *21*(2), 183–203. https://doi.org/10.1175/jhm-d-19-0130.1

Lu, Y., Steele-Dunne, S. C., Farhadi, L., & van de Giesen, N. (2017). Mapping surface heat fluxes by assimilating SMAP soil moisture and GOES land surface temperature data. *Water Resources Research*, *53*(12), 10858–10877. https://doi.org/10.1002/2017wr021415

Ma, H., Liu, Q., Liang, S., & Xiao, Z. (2017). Simultaneous estimation of leaf area index, fraction of absorbed photosynthetically active radiation, and surface albedo from multiple-satellite data. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(8), 4334–4354. https://doi.org/10.1109/tgrs.2017.2691542

Makela, A., Hari, P., Berninger, F., Hanninen, H., & Nikinmaa, E. (2004). Acclimation of photosynthetic capacity in Scots pine to the annual cycle of temperature. *Tree Physiology*, *24*(4), 369–376. https://doi.org/10.1093/treephys/24.4.369

Margulis, S. A., McLaughlin, D., Entekhabi, D., & Dunne, S. (2002). Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water Resources Research*, *38*(12). https://doi.org/10.1029/2001wr001114

McDermid, S. S., Mearns, L. O., & Ruane, A. C. (2017). Representing agriculture in Earth System Models: Approaches and priorities for development. *Journal of Advances in Modeling Earth Systems*, *9*, 2230–2265. https://doi.org/10.1002/2016MS000749

Misson, L., Tu, K. P., Boniello, R. A., & Goldstein, A. H. (2006). Seasonality of photosynthetic parameters in a multi-specific and vertically complex forest ecosystem in the Sierra Nevada of California. *Tree Physiology*, *26*(6), 729–741. https://doi.org/10.1093/treephys/26.6.729

Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, *28*(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*(D12). https://doi.org/10.1029/2010jd015139

Norton, A. J., Rayner, P. J., Koffi, E. N., & Scholze, M. (2018). Assimilating solar-induced chlorophyll fluorescence into the terrestrial biosphere model BETHY-SCOPE v1.0: Model description and information content. *Geoscientific Model Development*, *11*(4), 1517–1536. https://doi.org/10.5194/gmd-11-1517-2018

Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., & Wheeler, T. (2015). JULES-crop: A parametrization of crops in the Joint UK Land Environment Simulator. *Geoscientific Model Development*, *8*(4), 1139–1155. https://doi.org/10.5194/gmd-8-1139-2015

Pinnington, E., Quaife, T., & Black, E. (2018). Impact of remotely sensed soil moisture and precipitation on soil moisture prediction in a data assimilation system with the JULES land surface model. *Hydrology and Earth System Sciences*, *22*(4), 2575–2588. https://doi.org/10.5194/hess-22-2575-2018

Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2002). Hydrologic data assimilation with the Ensemble Kalman filter. *Monthly Weather Review*, *130*(1), 103–114. https://doi.org/10.1175/1520-0493(2002)130<0103:hdawte>2.0.co;2

Rinaldi, M. (2003). Variation of specific leaf area for sugar beet depending on sowing date and irrigation. *Italian Journal of Agronomy*, *7*(1), 23–32.

Sabater, J. M., Rüdiger, C., Calvet, J. C., Fritz, N., Jarlan, L., & Kerr, Y. (2008). Joint assimilation of surface soil moisture and LAI observations into a land surface model. *Agricultural and Forest Meteorology*, *148*(8–9), 1362–1373. https://doi.org/10.1016/j.agrformet.2008.04.003

Saigusa, N., Yamamoto, S., Murayama, S., Kondo, H., & Nishimura, N. (2002). Gross primary production and net ecosystem exchange of a cool-temperate deciduous forest estimated by the eddy covariance method. *Agricultural and Forest Meteorology*, *112*(3–4), 203–215. https://doi.org/10.1016/s0168-1923(02)00082-5

Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth System Modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*(24), 12396–12417. https://doi.org/10.1002/2017gl076101

Sellers, P. J. (1997). Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, *275*, 502–509. https://doi.org/10.1126/science.275.5299.502

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., et al. (2007). *A description of the advanced research WRF version 2* (p. 100). NCAR Tech. Note NCAR/TN – 468+STR.

Sun, Y., Frankenberg, C., Jung, M., Joiner, J., Guanter, L., Köhler, P., & Magney, T. (2018). Overview of Solar-Induced chlorophyll Fluorescence (SIF) from the orbiting carbon observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP. *Remote Sensing of Environment*, *209*, 808–823. https://doi.org/10.1016/j.rse.2018.02.016

Tian, S., Tregoning, P., Renzullo, L. J., Dijk, van, A. I. J. M., Walker, J. P., Pauwels, V. R. N., & Allgeyer, S. (2017). Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals. *Water Resources Research*, *53*(3), 1820–1840. https://doi.org/10.1002/2016wr019641

Verma, M., Schimel, D., Evans, B., Frankenberg, C., Beringer, J., Drewry, D. T., et al. (2017). Effect of environmental conditions on the relationship between solar-induced fluorescence and gross primary productivity at an OzFlux grassland site. *Journal of Geophysical Research: Biogeosciences*, *122*, 716–733. https://doi.org/10.1002/2016JG003580

Wang, H., Zhu, Y., Li, W., Cao, W., & Tian, Y. (2014). Integrating remotely sensed leaf area index and leaf nitrogen accumulation with RiceGrow model based on particle swarm optimization algorithm for rice grain yield assessment. *Journal of Applied Remote Sensing*, *8*(1), 083674. https://doi.org/10.1117/1.jrs.8.083674

Williams, I. N., Lu, Y., Kueppers, L. M., Riley, W. J., Biraud, S. C., Bagley, J. E., & Torn, M. S. (2016). Land-atmosphere coupling and climate prediction over the U.S. Southern Great Plains. *Journal of Geophysical Research: Atmospheres*, *121*(20), 12125–12144. https://doi.org/10.1002/2016jd025223

Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T., et al. (2016). *Evaluation of JULES-crop performance against site observations of irrigated maize from Mead, Nebraska*. https://doi.org.10.5194/gmd-2016-252

Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T., et al. (2017). Evaluation of JULES-crop performance against site observations of irrigated maize from Mead, Nebraska. *Geoscientific Model Development*, *10*, 1291–1320. https://doi.org/10.5194/gmd-10-1291-2017

Wilson, P. J., Thompson, K., & Hodgson, J. G. (1999). Specific leaf area and leaf dry matter content as alternative predictors of plant strategies. *New Phytologist*, *143*, 155–162. https://doi.org/10.1046/j.1469-8137.1999.0042710.1046/j.1469-8137.1999.00427.x

Wright, I. J., & Westoby, M. (1999). Differences in seedling growth behavior among species: Trait correlations across species, and trait shifts along nutrient compared to rainfall gradients. *Journal of Ecology*, *87*(1), 85–97. https://doi.org/10.1046/j.1365-2745.1999.00330.x

Xia, Y., Hao, Z., Shi, C., Li, Y., Meng, J., Xu, T., et al. (2019). Regional and global land data assimilation systems: Innovations, challenges, and prospects. *Journal of Meteorological Research*, *33*(2), 159–189. https://doi.org/10.1007/s13351-019-8172-4

Xie, Y., Wang, P., Bai, X., Khan, J., Zhang, S., Li, L., & Wang, L. (2017). Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using Landsat imagery and the CERES-Wheat model. *Agricultural and Forest Meteorology*, *246*, 194–206. https://doi.org/10.1016/j.agrformet.2017.06.015

Xu, T., Bateni, S. M., Neale, C. M. U., Auligne, T., & Liu, S. (2018). Estimation of turbulent heat fluxes by assimilation of land surface temperature observations from GOES satellites into an Ensemble Kalman smoother framework. *Journal of Geophysical Research: Atmospheres*, *123*(5), 2409–2423. https://doi.org/10.1002/2017jd027732

Xu, T., Guo, Z., Liu, S., He, X., Meang, Y., Xu, Z., et al. (2018). Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale. *Journal of Geophysical Research: Atmospheres*, *123*, 8674–8690. https://doi.org/10.1029/2018JD028447

Xu, T., He, X., Bateni, S. M., Auligne, T., Liu, S., Xu, Z., et al. (2019). Mapping regional turbulent heat fluxes via variational assimilation of land surface temperature data from polar orbiting satellites. *Remote Sensing of Environment*, *221*, 444–461. https://doi.org/10.1016/j.rse.2018.11.023

Xu, T., Liang, S., & Liu, S. (2011). Estimating turbulent fluxes through assimilation of geostationary operational environmental satellites data using ensemble Kalman filter. *Journal of Geophysical Research*, *116*(D9). https://doi.org/10.1029/2010jd015150

Xu, T., Liu, S., Xu, Z., Liang, S., & Xu, L. (2015). A dual-pass data assimilation scheme for estimating surface fluxes with FY3A-VIRR land surface temperature. *Science China Earth Sciences*, *58*(2), 211–230. https://doi.org/10.1007/s11430-014-4964-7

Xu, X., Chen, F., Barlage, M., Gochis, D., Miao, S., & Shen, S. (2019). Lessons learned from modeling irrigation from field to regional scales. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2428–2448. https://doi.org/10.1029/2018ms001595

Yan, H., Moradkhani, H., & Zarekarizi, M. (2017). A probabilistic drought forecasting framework: A combined dynamical and statistical approach. *Journal of Hydrology*, *548*, 291–304. https://doi.org/10.1016/j.jhydrol.2017.03.004

Yang, K., Watanabe, T., Koike, T., Li, X., Fujii, H., Tamagawa, K., et al. (2007). Auto-calibration system developed to assimilate AMSR-E data into a Land Surface Model for estimating soil moisture and the surface energy budget. *Journal of the Meteorological Society of Japan*, *85A*, 229–242. https://doi.org/10.2151/jmsj.85a.229

Yang, Z. L., Niu, G. Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *Journal of Geophysical Research*, *116*(D12). https://doi.org/10.1029/2010jd015140

Zhang, G., Chen, F., & Gan, Y. (2016). Assessing uncertainties in the Noah-MP ensemble simulations of a cropland site during the Tibet Joint International Cooperation program field campaign. *Journal of Geophysical Research: Atmospheres*, *121*(16), 9576–9596. https://doi.org/10.1002/2016jd024928

Zhang, Y., Guanter, L., Berry, J. A., Joiner, J., van der Tol, C., Huete, A., et al. (2014). Estimation of vegetation photosynthetic capacity from space-based measurements of chlorophyll fluorescence for terrestrial biosphere models. *Global Change Biology*, *20*(12), 3727–3742. https://doi.org/10.1111/gcb.12664

Zhang, Y., Guanter, L., Berry, J. A., van der Tol, C., Yang, X., Tang, J., & Zhang, F. (2016). Model-based analysis of the relationship between sun-induced chlorophyll fluorescence and gross primary production for remote sensing applications. *Remote Sensing of Environment*, *187*, 145–155. https://doi.org/10.1016/j.rse.2016.10.016

Zhang, Y., Joiner, J., Alemohammad, S. H., Zhou, S., & Gentine, P. (2018). A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. *Biogeosciences*, *15*(19), 5779–5800. https://doi.org/10.5194/bg-15-5779-2018

Zhang, Y., Suyker, A., & Paustian, K. (2018). Improved crop canopy and water balance dynamics for agroecosystem modeling using DayCent. *Agronomy Journal*, *110*(2), 511–524. https://doi.org/10.2134/agronj2017.06.0328

Zhang, Y., Xiao, X., Jin, C., Dong, J., Zhou, S., Wagle, P., et al. (2016). Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America. *Remote Sensing of Environment*, *183*, 154–169. https://doi.org/10.1016/j.rse.2016.05.015

Zhang, Z., Barlage, M., Chen, F., Li, Y., Helgason, W., Xu, X., et al. (2020). Joint modeling of crop and irrigation in the Central United States using the Noah-MP land surface model. *Journal of Advances in Modeling Earth Systems*, *12*. https://doi.org/10.1029/2020ms002159

Zhuo, L., Dai, Q., Han, D., Chen, N., & Zhao, B. (2019). Assessment of simulated soil moisture from WRF Noah, Noah-MP, and CLM land surface schemes for landslide hazard application. *Hydrology and Earth System Sciences*, *23*, 4199–4218. https://doi.org/10.5194/hess-23-4199-2019