Simulating carbon and water fluxes using a coupled process-based terrestrial biosphere model and joint assimilation of leaf area index and surface soil moisture

Sinan Li [1,2], Li Zhang [1,3,*], Jingfeng Xiao [4,*], Rui Ma [5], Xiangjun Tian [6], Min Yan[1,3]

[1] Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, No. 9 Dengzhuang South Road, Beijing 100094, China.

[2] College of Resources and Environment, University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

[3] International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China

[4] Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, New Hampshire 03824, USA

[5] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

[6] International Center for Climate and Environment Sciences (ICCES), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

***** Correspondence: zhangli@aircas.ac.cn; and j.xiao@unh.edu

Abstract:

Reliable modeling of carbon and water fluxes is essential for understanding the terrestrial carbon and water cycles and informing policy strategies aimed at constraining carbon emissions and improving water use efficiency. We designed an assimilation framework (LPJ-Vegetation and soil moisture Joint Assimilation, or LPJ-VSJA) to improve gross primary production (GPP) and evapotranspiration (ET) estimates globally. The integrated model, LPJ-PM (LPJ-PT-JPL_{SM} Model) as the underlying model, was coupled from the Lund-Potsdam-Jena Dynamic Global Vegetation Model (LPJ-DGVM version 3.01) and a hydrology module (i.e., the updated Priestley–Taylor Jet Propulsion Laboratory model, PT-JPL_{SM}). Satellite-based soil moisture products derived from the Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active and Passive (SMAP) and leaf area index (LAI) from the global Land and Ground satellite (GLASS) product were assimilated into LPJ-PM to improve GPP and ET simulations using a Proper Orthogonal Decomposition-based ensemble four-dimensional variational assimilation method (PODEn4DVar). The joint assimilation framework LPJ-VSJA achieved the best model performance (with an R^2 of 0.91and 0.81 and an ubRMSD reduced by 40.3% and 29.9% for GPP and ET, respectively, compared with those of LPJ-DGVM at the monthly scale). The GPP and ET resulting from the assimilation demonstrated a better performance in the arid and semi-arid regions (GPP: R^2=0.73, ubRMSD=1.05 g C m^{-2} d^{-1}; ET: R^2=0.73, ubRMSD= 0.61 mm d^{-1}) than in the humid and sub-dry humid regions (GPP: R^2=0.61, ubRMSD=1.23 g C m^{-2} d^{-1}; ET: R^2=0.66; ubRMSD=0.67 mm d^{-1}). The ET simulated by LPJ-PM that assimilated SMAP or SMOS had a slight difference, and the SMAP soil moisture data performed better than that of SMOS data. Our global simulation modeled by LPJ-VSJA was compared with several global GPP and ET products (e.g., GLASS GPP, GOSIF GPP, GLDAS ET, GLEAM ET) using the triple collocation (TC) method. Our products, especially ET, exhibited advantages in the overall error distribution (estimated error (μ): 3.4 mm $month^{-1}$; estimated standard deviation of μ: 1.91 mm $month^{-1}$). Our research showed that the assimilation of multiple datasets could reduce model uncertainties, while the model performance differed across regions and plant functional types. Our assimilation framework (LPJ-VSJA) can improve the model simulation performance of daily GPP and ET globally, especially in water-limited regions.

Keywords: Data Assimilation; SMOS; SMAP; Gross primary production (GPP); evapotranspiration (ET); GLASS LAI

---

1. Introduction

Gross primary production (GPP) and evapotranspiration (ET) are essential components of the carbon and water cycles. Carbon and water fluxes are inherently coupled on multiple spatial and temporal scales (Law et al. 2002; Sun et al. 2019; Waring and Running 2010). Terrestrial biosphere models are the most sophisticated approach for providing a relatively detailed description of such interdependent relationships regarding water and carbon fluxes and understanding the response of terrestrial ecosystems to changes in atmospheric CO_2 and climate (Kaminski et al. 2017). The dynamic global vegetation models (DGVMs) are process-based dynamic terrestrial biosphere models, which can simulate water, carbon, and energy exchange between vegetation and the atmosphere under different conditions accounting for vegetation physiological processes, and are widely used to estimate carbon and water fluxes of terrestrial vegetation. However, there are still large uncertainties in carbon and water flux estimates at regional to global scales. Both diagnostic and prognostic models show substantial differences in the magnitude and spatiotemporal patterns of GPP and ET. For example, the global annual GPP estimates exhibited a large range (130–169 Pg C yr^{-1}) among 16 process-based terrestrial biosphere models (Anav et al. 2015). The global ET ranged from 70,000 to 75,000 km^3 yr^{-1}, and the uncertainty of regional or global ET estimates was up to 50% of the annual mean ET value, especially in the semi-arid regions (Miralles et al. 2016). These uncertainties mainly arise from the forcing datasets, simplification of mechanisms or imperfect assumptions in processes, and uncertain parameters in the processed models and assimilation methods (Xiao et al. 2019).

In the last two decades, remote sensing products have been assimilated into DGVM$_S$ to reduce the uncertainty in modeled carbon and water fluxes (MacBean et al. 2016; Scholze et al. (2017); Exbrayat et al. (2019)). Data assimilation (DA) is an effective approach to reduce uncertainties in terrestrial biosphere models by integrating satellite products with models to constrain related parameters or state variables. A DA system contains four main components: a set of observations, an observation operator, an underlying model, and an assimilation method. The assimilation method considers the errors from both models and observations, and reduces model uncertainties by minimizing a cost function. The Ensemble Kalman Filter (EnKF) has been widely applied in land surface process models for parameter optimization, which significantly improve simulations by periodically updating state variables (e.g., LAI and soil moisture) using remote sensing data without altering the model structure (Rahman et al. 2021; Bonan et al. 2020; Xu et al. 2021). Yet, the EnKF relies on the instantaneous observations to update the state variable at the current time, and gives the predicted value at the next time based on the forward integration of the updated state variable. The four-dimensional variational method (4DVar) assimilation method can obtain the dynamic balance of the estimation in the time window when it is applied to the long-series forecast model (Bateni et al. 2014; Xu et al. 2019). In particular, the Proper Orthogonal Decomposition (POD)-based ensemble 4DVAR assimilation method (referred to as PODEn4DVar) (Tian and Feng 2015) requires relatively less computation and can simultaneously assimilate the observations at different time intervals. Meanwhile, it maintains the structural information of the four-dimensional space. This method has a satisfactory performance in land DA for carbon and water variables (Tian et al. 2009; Tian et al. 2010) and can better estimate GPP and ET than EnKF (Ma et al. 2017).

Multiple sources of remote sensing data streams have been used to constrain models for assimilation. As a critical biophysical parameter of the land, leaf area index (LAI) is closely related to many land processes, such as photosynthesis, respiration, precipitation interception, ET, and surface energy exchange (Fang et al. 2019). LAI has a lot of impact on the simulation of carbon and water fluxes (Liu et al. 2018), and accurate LAI estimates can improve the simulations of the carbon and water fluxes (Bonan et al. 2014;; Mu et al. 2007). He et al. (2021) assimilated land surface temperature and LAI observations into the 4DVar framework and improved ET and GPP estimates. Soil moisture is a major driving factor affecting vegetation production in arid ecosystems, especially, in semi-arid areas (Liu et al. 2020). Introducing surface soil moisture (SSM) into the model can significantly improve GPP and ET simulation, particularly in water-limited areas (He et al. 2017; Li et al. 2020).

The advancement of earth observation, machine learning, inversion algorithms, and computer technology has improved the accuracy of global LAI products and boosted model-data fusion studies (Fang et al. 2019; Kganyago et al. 2020; Xiao et al. 2017). The Advanced Very High-Resolution Radiometer (AVHRR) generates global LAI products with the longest historic record (since the early 1980s). The GLASS LAI product has been verified to have a better accuracy than that of MODIS and CYCLOPES and is more temporally continuous and spatially complete (Xiao et al. 2013). Several recent studies showed that the assimilation of GLASS LAI into DGVMs enhanced the performance of the models in simulating carbon cycling (e.g., GPP, Net Ecosystem Exchange (NEE)) and hydrological (e.g., ET, SM) processes (Ling et al. 2019; Ma et al. 2017; Yan et al. 2016).

Microwave remote sensors are considered effective tools for measuring SM globally (Petropoulos et al. 2015). For example, SSM products have been derived from the Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active and Passive (SMAP) satellites equipped with an L-band microwave instrument. The products from these satellites have been evaluated against in-situ observations and other SSM products and overall have high accuracy(Burgin et al. 2017; Cui et al. 2018). Additionally, the SMAP performs better than SMOS and other SSM products (e.g., Advanced Scatterometer (ASCAT), Advanced Microwave Scanning Radiometer 2 (AMSR2)) with an overall lower error and a higher correlation based on the verification with in-situ SSM data from 231 sites (Cui et al. 2018; Kim et al. 2018). The assimilation of SMAP data can improve the simulation accuracy of carbon and water fluxes (He et al. 2017; Li et al. 2020) and hydrological variables (surface soil moisture, root-zone soil moisture (RZSM), and streamflow) (Blyverket et al. 2019; Koster et al. 2018; Reichle et al. 2017). In addition, the assimilation of SMAP data performed slightly better than that of SMOS and ESA CCI data (Blyverket et al. 2019).

In the nonlinear model or nonlinear observation operator, only simultaneous assimilation makes optimal use of observations (MacBean et al. 2016). Therefore, a joint assimilation of SM and LAI can make full use of the two variables. From site (Albergel et al. (2010);Rüdiger et al. (2010); Wu et al.,2018) to regional assimilation (Ines et al. (2013)), many studies showed that joint assimilation of vegetation parameters and SM can improve the simulation of the carbon and water cycles. Over small regions and at high spatial resolution, Xie et al. (2018) and Pan et al. (2019) showed that the joint assimilation of SM and LAI improved the accuracy of crop yield estimation using high-resolution satellites products from Sentinel-1 and -2. At a large regional scale, Bonan et al. (2020) assimilated LAI and SSM together into the Interactions between Soil, Biosphere and Atmosphere (ISBA) land model and improved the modeled GPP, ET, and runoff in the Mediterranean region. Rahman et al. (2022) jointly assimilates GLASS LAI and SMAP soil moisture to improve water and carbon flux simulations within the Noah-MP model over the Continental United States domain. Albergel et al.(2020) jointly assimilates the ASCAT soil moisture index (SMI) and LAI GEOV1 into ISBA through the Global Offline Land Data assimilation system LDAS-Monde to monitor extreme events such as drought and Heatwave events. In conclusion, Kalman Filter and its variant methods are mostly used to implement joint assimilation methods at regional scale, which requires many kinds of observation data and their accuracy directly affects the assimilation performance.

This study stems from the researches discussed above and further explored the potential of jointly assimilating satellite LAI and soil moisture products globally. Specifically, it was the first time that an updated LPJ-DGVM model was used to jointly assimilate GLASS LAI and SMAP soil moisture for simulating global water and carbon fluxes. The latest global soil moisture datasets (SMOS and SMAP) were used, and the assimilation performance of these two observations was analyzed. Since previous work showed the importance of surface soil moisture in the semi-arid and arid areas, one of the specific objectives of our study is to compare the assimilation effect in the humid and arid areas and improve the understanding of the effect of surface soil moisture on vegetation activity in wet and dry zones. In addition, compared with the assimilation methods in previous studies (mostly using Kalman Filter variants), the POD-En4DVar method is used, which greatly improves the computational efficiency.

2. LPJ-VSJA framework and assimilation strategy

2.1. Coupled- model (LPJ-PM) for assimilation

In this study, a coupled terrestrial biosphere model, LPJ-PM, was used to simulate daily GPP and ET by assimilating satellite-derived LAI and SSM. The LPJ-PM is coupled from LPJ-DGVM and PT-JPL_{SM}. The original input data in PT-JPL_{SM} were all inherited from LPJ-DGVM, with the exception of relative humidity (RH) and surface soil moisture (SMOS and SMAP), including the initial LAI calculated by the LPJ-DGVM or assimilated LAI obtained through the LAI assimilation scheme, canopy height, and the fraction of absorbed photosynthetic effective radiation (fAPAR). The detailed processes of the LPJ-PM have been described in Li et al. (2020), and the flow chart for the coupling is shown in Figure 1.

Table 1. Description of the models and outputs in this study

acronyms	Full name	Description	Output
LPJ-DGVM (Sitch et al. 2003)	Lund-Potsdam-Jena Dynamic Global Vegetation Model	This model is used as a model operator to simulated initial ET	GPP_{LPJ}, ET_{LPJ}
$PT\text{-}JPL_{SM}$ (Purdy et al. (2018))	Updated Priestley–Taylor Jet Propulsion Laboratory model	The model is used as a module of the LPJ-PM and establishes a connection between SMAP SM and ET	N/A
LPJ-PM (Li et al. (2020))	Lund-Potsdam-Jena and Updated Priestley–Taylor Jet Propulsion Laboratory coupled model	An integrated model corresponding to the coupling of the $PT\text{-}JPL_{SM}$ and LPJ-DGVM	GPP_{SM}, ET_{PM}
LPJ-VSJA (this study)	Lund-Potsdam-Jena Vegetation-Soil moisture-Joint -Assimilation system	A process-based assimilation framework for assimilating LAI and SSM jointly into LPJ-PM	GPP_{LAI}, ET_{LAI}; GPP_{SM}, ET_{SM}; GPP_{joint}; ET_{joint}

2.1.1 LPJ-DGVM

The LPJ-DGVM is a process-oriented dynamic model, which considers mutual interaction of carbon and water cycling and is designed to simulate vegetation distribution and carbon, soil and atmosphere fluxes (Sitch et al. 2003). For each plant functional type (PFT), the GPP is calculated by implementing coupled photosynthesis and water balance

The canopy GPP is updated daily:

$$GPP = \frac{(J_E + J_c - \sqrt{(J_E + J_c)^2 - 4\theta J_E J_c})}{2\theta} \quad (2.1)$$

where J_C is the Rubisco limiting rate of photosynthesis, J_E is the light limiting rate of photosynthesis, and the empirical parameter θ represents the common limiting effect between the two terms. J_E is related to APAR (absorbed photosynthetic radiation, product of FPAR and PAR), while J_C is related to Vcmax (canopy maximum carboxylation capacity, μ mol $CO^2/m^2/s$):

$$J_E = C_1 APAR \quad (2.2)$$

$$J_C = C_2 V_{C\max} \quad (2.3)$$

where C_1 and C_2 are determined by a variety of photosynthetic parameters and the intercellular partial pressure of CO_2, which is related to atmospheric CO_2 content and further altered by leaf stomatal conductance (Sitch et al. 2003). APAR and FPAR are directly related to LAI.

In the water cycle module, ET is calculated as the minimum of a plant- and soil-limited supply function (E_{supply}) and the atmospheric demand (E_{demand}) (Haxeltine and Prentice 1996; Sitch et al. 2003).The soil structure is simplified to a “two-layer bucket” model (the top soil layer at a 0-50 cm depth and the bottom layer at a 50-100 cm depth) .

$$E_S = Ep \times Wr_{20} \times (1 - fv) \quad (2.4)$$

In this module, it is assumed that the soil layer above 20 cm produces water through evaporation, and Wr_{20} is the relative water content of the soil above 20 cm, which is used as the only soil water limit for calculating vegetation transpiration and soil evaporation. In the evapotranspiration estimation, the over-simplification of soil structure and soil water limitation lead to a large error (Sitch et al. 2003), while LPJ-DGVM cannot directly assimilate surface soil water due to the limitation of soil layer stratification

, and therefore, the satellite-derived SSM cannot be assimilated into LPJ-DGVM directly. The oversimplified soil structure and single soil moisture limitation inevitably lead to sizeable uncertainty in ET simulation. Additionally, the monthly input caused a daily variation of the modeled SM, which was also not transmitted to the calculation of GPP and ET. Thus, the updated PT-JPL model (hereafter referred to as PT-JPL_{SM}) was coupled with LPJ-DGVM and the model structure was modified so that SSM can be directly assimilated into the coupled model at the daily time step.

2.1.2 PT-JPL_{SM}

In PT-JPL_{SM}, three ET components are modelled: soil evaporation (E), vegetation transpiration (T), and leaf evaporation (I). The PT-JPL_{SM} introduced a constraint (0–1, C_{RSM}) of SSM for T and E, which was used to avoid the implicit soil water control (represented by $f_{SM}=RH^{VPD}$) in the PT-JPL model.

Vegetation transpiration:

$$C_{\mathrm{RSM}} = (1 - RH^{4(1-VWC)(1-RH)})C_{SM} + (RH^{4(1-VWC)(1-RH)})C_{TRSM} \quad (2.5)$$

$$C_{TRSM} = 1 - \left(\frac{w_{CR} - w_{obs}}{w_{CR} - w_{pwp\_CH}}\right)^{\sqrt{CH}} \quad , \tag{2.6}$$

where w_{obs} is the SMAP SSM, w_{pwp} is the water content at the wilting point, and VWC is volumetric water content. W_{CR} is a crucial parameter in characterizing the extent of SM restriction on ET; w_{pwp_CH} is the canopy height (CH) adjusted surface soil moisture wilting point and is related to the potential of roots capturing water from deeper sources to limit the transpiration rate and characterize the SM availability (Purdy et al., 2018; Evensen 2003; Serraj et al., 1999). The specific formula is given in Purdy et al. (2018).

Soil evaporation:

$$C_{RSM} = \frac{w_{obs} - w_{pwp}}{w_{fc} - w_{pwp}} \tag{2.7}$$

The proportion of available water limits the soil evapotranspiration to the maximum available water. This scalar was formulated to represent the relatively accurate extractable water content for the vegetation, determined by soil properties and the water available for evaporation, which is estimated via surface water constraints.

The SMAP SSM was applied to model global ET using PT-JPL$_{SM}$ and the results demonstrated the largest improvements for ET estimates in dry regions (Purdy et al. 2018). Due to the limitation of soil stratification in LPJ-DGVM, the model was coupled with an updated remote-sensing ET algorithm in the PT-JPL$_{SM}$ that could better simulate ET in water-limited regions than in humid regions (Purdy et al. 2018).

2.2. Assimilation scheme and experiment procedure

To improve the prediction capability of LPJ-PM, we designed three assimilation schemes: assimilating LAI only (LAI-only, **output: ET_{LAI}, GPP_{LAI}**), assimilating SSM only (SSM-only, **output: GPP_{SM}, ET_{SM}**), and joint assimilation of LAI and SSM (Joint LAI and SSM assimilation, **output: ET_{joint}, GPP_{joint}**), i.e., LPJ-VSJA framework) to test the assimilation performance for simulating GPP and ET.

The proposed LPJ-VSJA framework consists of four main components: the model operator (the LPJ-PM), the observation operator (to establish the relation between the assimilation variable and the observed variable), the observation series (GLASS LAI and SMOS or SMAP products), and the assimilation algorithm (POD4DVar). With the surface soil moisture constraint in the PT-JPL_{SM}, the LPJ-VSJA corrects the output fluxes (GPP and ET in this study).

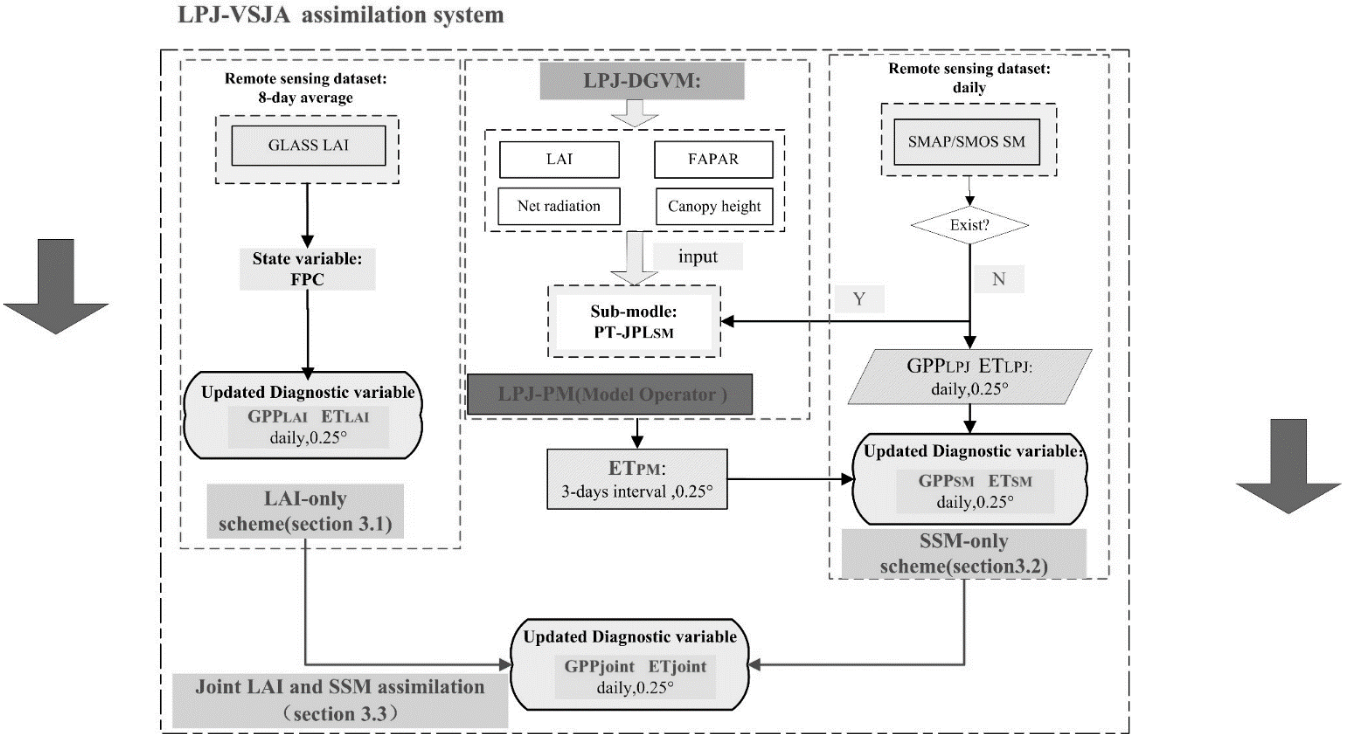


Figure 1. Flowchart of the LPJ-VSJA assimilation scheme: three assimilation schemes and the coupled model: LPJ-PM (adapted from Li et al., 2020). The abbreviation of model and assimilation framework is explained in Table 1.

The experiment consisted of six steps:

Step 1: initialize the LPJ-DGVM and output the reference state variables without assimilation over the experimental period (2010–2018), referred to as the “Control run” scenario.

Step 2: implement three assimilation schemes respectively, and the results represent the assimilation integration state (daily GPP and ET assimilation results are referred to as the “GPP_{LAI}” and “ET_{LAI}” in LAI-only scheme;“GPP_{SM}” and “ET_{SM}” in SSM-only scheme and “GPP_{joint} ” and “ET_{joint} ” in Joint LAI and SSM assimilation scheme. This scenario used the same input data and model parameter scheme with the "Control run" scenario.

Step 3:evaluate GPP and ET results (three schemes) by comparing the parameters, R^2 (correlation coefficient), BIAS, and ubRMSD (unbiased root mean square deviation), for conditions of without-DA ("Control run" scenario) and with-DA states, and assess the assimilation performance of separate assimilation and joint assimilation to determine the optimal assimilation scheme for GPP and ET, respectively.

Step 4: evaluate the in-situ GPP and ET resulting from the assimilation where the sites are located in wet or dry regions by dividing these validation sites into four parts (humid, sub-dry humid, semi-arid, and arid regions), and this step was designed to assess the superiority of the proposed assimilation scheme in water-limited areas.

Step 5: compare the ET assimilation performance by assimilating the SMOS data with that by assimilating the SMAP data.

Step 6: evaluate the simulated GPP and ET maps based on the optimal assimilation scheme against existing global flux products.

2.2.1 LAI-only assimilation scheme

In the LAI-only assimilation scheme, the observation operator determines the relationship between LAI and foliage projective cover (FPC) in the process model (equation 2.1), and the assimilated LAI will be propagated by energy transmission and ecosystem processes (e.g. photosynthesis, transpiration of vegetative process) in the dynamic model to improve GPP and ET simulations (Bonan et al. 2014; Mu et al. 2007). FPC, the vertically projected percentage of the land covered by foliage, regulates the rate of photosynthate conversion and transpiration. In this study, the GLASS LAI with 8-day interval for the period 2010–2018 was selected as the observation dataset for assimilation, and the FPC state variable was updated daily through running the LPJ-PM (FPC_{DA}, GPP_{LAI}, ET_{LAI} in this study) as shown below:

$$FPC = 1 - e^{-0.5LAI} \qquad (2.1)$$

We set the model and observation errors at a given time as 20% and 10% (scale factor) of the LAI value and the observed LAI value, respectively. By verifying the assimilation performance (R^2, RMSD, BIAS) for different scale factors(f) of model simulation and observations in the range of 0.05 to 0.40, taking a step size of 0.05 (a total of 64 combinations), the optimal scale factors (0.2 and 0.1) were determined (Bonan et al., 2020).The model and observation errors was the LAI value multiplied by f. The model integration generation method described by Pipunic et al. (2008) was used to determine the minimum number of ensemble members required to achieve maximum efficiency, and the number of sets was 20.

2.2.2 SSM-only assimilation scheme

In this scheme, the SSM products (SMOS or SMAP) were assimilated into LPJ-PM to obtain more accurate ET (ET_{SM}) estimates in water-limited areas. The observation series was the SMOS or SMAP SSM product, and the observation operator was the PT-JPL_{SM} model. The ET_{PM} (see Table 1) was estimated by the coupled model (LPJ-PM) introducing SSM as a diagnostic variable. The ET values resulting from the assimilation was applied to compute the top layer SM (50 cm) at the next time step (a nonlinear soil water availability function described by Zhao et al. (2013), providing feedback for subsequent hydrologic and carbon cycle processes. Then, the updated SM values regulated the GPP simulation (output: GPP_{SM}). Different from other "constant" ET observations, the ET_{PM} ("observation") at each time *t* were adjusted by absorbing intermediate variables updated after assimilation at time *t*-1. The ET_{PM} was shown to be better than ET simulated by LPJ-DGVM but not as good as that simulated by the model with SMAP SSM assimilated (Li et al. 2020). Thus, it is also proven that this SSM assimilation schemes could improve the accuracy of ET simulations (Li et al. 2020).

All assimilation simulations were conducted between January 2010 and December 2018. Between January 2010 and April 2015, SMOS data were used for assimilation; and after May 2015, both SMOS and SMAP data were used for assimilation. An assimilation scheme was conducted when RH and SMOS or SMAP SSM data existed simultaneously; otherwise, the original simulation of the LPJ-DGVM was conducted directly without adjustment of assimilation.

Similar to the LAI assimilation scheme, the model and observation errors were set as 15% and 5% of ET_{LPJ} and ET_{PM}, respectively (LPJ-PM was adopted before assimilation). The number of ensemble members was set to 50.The ET_{PM} must be rescaled to the ET_{LPJ} distribution via their corresponding cumulative probabilities using the cumulative distribution function (CDF) matching to avoid introducing any BIAS in the LPJ-VSJA system (Li et al. 2020).

2.2.3 Joint LAI and SSM assimilation scheme

In this scheme, both LAI from GLASS and SSM from SMOS or SMAP were the observation datasets. The GLASS LAI was assimilated to obtain the FPC_{DA} and ET_{LAI}, and then the FPC_{DA} served as input to LPJ-PM to simulate optimized ET_{PM}, and the ET_{joint} was generated using ET_{LAI} and ET_{PM}. Then, the SM (referred to as SM_{CO} in Figure S1) updated by ET_{joint} and the FPC_{DA} were used as input to correct GPP (ET_{joint}).

Here, we applied the error regulation in the LAI-only scheme and maintained the error setting of the LAI observation and model simulation. Considering the transmission of integrated model error, we recalculated the model error of LPJ-PM after the LAI assimilation and set model and observation errors of ET_{LAI} and ET_{PM} to be 15 and 10%, respectively.

2.3. POD-Based Ensemble 4D Variational Assimilation Method

The Proper Orthogonal Decomposition (POD)-based ensemble four-dimensional variational (4DVar) assimilation method (referred to as PODEn4DVar) (Tian and Feng 2015) has the advantage of avoiding the calculation of adjoint patterns as its incremental analysis field, which can be represented linearly by the POD base (Transformed OP (Observing Perturbation) and MP (Model Perturbation)). Moreover, the PODEn4DVar can simultaneously assimilate multiple-time observation data and provide flow-dependent (the flow-dependent is the ensembles of forecasting statistical characteristics in the t time) error estimates of the background errors. It has shown advantages in terrestrial assimilation, Tan-Tracker system (a

Chinese carbon cycle data-assimilation system; in Chinese, “Tan” means carbon), and Radar assimilation (Tian et al. 2010; Tian et al. 2009; Tian et al. 2014; Zhang and Weng 2015).

By minimizing the following initial incremental format of the cost function in the 4DVar algorithm, an analysis field can be obtained:

$$\boldsymbol{J(x')=\frac{1}{2}(x')B^{-1}(x')+\frac{1}{2}[y'(x')-y'_{obs}]^{T}R^{-1}[y'(x')-y'_{obs}]}$$

Here, the $x' = x - x_b$, $y'(x') = y(x' + x_b) - y(x_b)$, $y'_{obs} = y_{obs} - y(x_b)$, $y = H[M_{to\rightarrow tk}(x)]$. $x'(x'_1, x'_2, \ldots\ldots, x'_N)$ is the model perturbation (MP) matrix and $y'(y'_1, y'_2, \ldots\ldots, y'_N)$ is the observation perturbation (OP) matrix with N samples. Following Rüdiger et al. (2010), the LAI perturbation was set to a fraction (0.001) of the LAI itself. The perturbation of ET_{PM} and ET_{LPJ} conforms to a Gaussian distribution with a mean of 0 and a specified covariance (10 and 5% of the ET_{PM} and ET_{LPJ} at time t). The subscript b represents the background field, the superscript T represents a transpose, H is the observation operator of the LAI-only assimilation scheme as described in section 2.2.1, and the SSM-only assimilation scheme is the $PT\text{-}JPL_{SM}$ (described in 2.1.2). M is the forecast model (LPJ-PM in this study), B is the background error covariance, R is the observation error covariance, and obs denotes observation.

Assuming the approximately linear relationship between OP(y’) and MP(x’), POD decomposition and transformation were successively conducted for OP and MP. The transformed OP samples ($\Phi_y =$

$y'_1, y'_2, \ldots\ldots, y'_n$) are orthogonal and independent, and the transformed MP samples ($\Phi_x = x'_1, x'_2, \ldots., x'_n$) are orthogonal to the corresponding OP samples, where n is the number of POD modes.

The manifestation of the background error covariance is the same as the Ensemble Kalman filter (EnKF, Evensen (2004)), and the incremental analysis x'_a was expressed by the $\Phi_{x,n}$, and $\widetilde{\Phi}_y(\widetilde{\Phi}_y = \left[(n-1)I_{n\times n} + \Phi_{y,n}^T R^{-1}\Phi_{y,n}\right]^{-1}\Phi_{y,n}^T R^{-1})$. Finally, the optimal analysis x_a is calculated as $x_a = x_b + \Phi_{x,n}\widetilde{\Phi}_y y'_{obs}$. The detailed derivation process of the algorithm is described by a previous study (Tian et al. 2011).

In the ensemble-based method (Evensen et al.,2004), the number of ensemble members is usually fewer than that of the observation data and the degrees of freedom of the model variables, and spurious long-range correlations occur between observation locations and model variables. A practical method, the localization technique, is applied to address this issue (Mitchell et al. 2002). The final incremental analysis is rewritten as:

$$x'_a = \Phi_{x,n}\widetilde{\Phi}_y y'_{obs} C_0(\frac{d_h}{d_{h,0}}) \cdot C_0(\frac{d_v}{d_{v,0}})$$

where d_h and d_v are the horizontal and vertical distances between the spatial positions of state and observed variables, respectively; and $d_{h,0}$ and $d_{v,0}$ are the horizontal and vertical covariance localization Schur radii, respectively. The filtering function C_0 is expressed as:

$$C_0(r)=\begin{cases}-\frac{1}{4}r^5+\frac{1}{2}r^4+\frac{5}{8}r^3-\frac{5}{3}r^2+1, & 0\le r\le 1,\\ \frac{1}{12}r^5-\frac{1}{2}r^4+\frac{5}{8}r^3+\frac{5}{3}r^2-5r+4-\frac{2}{3}r^{-1}, & 1\le r\le 2,\\ 0, & 2<r\end{cases}$$

where r is the radius of the filter.

The assimilation algorithm is mainly divided into two steps: (1) prediction: run LPJ-PM in the current assimilation window and generate simulation results and background field vectors; (2) update: the algorithm is used to calculate the optimal assimilation increment x'_a and analysis solution x_a, and the simulation results and the initial conditions of the model in the current window are updated using the analysis solution. The updated initial conditions were applied for model LPJ-PM prediction, and the above process was repeated.

2.4. Validation method for assimilation performance

The R^2 (coefficient of determination), BIAS, and ubRMSD (unbiased root mean square deviation) between simulation and tower-based observations were applied for evaluation. In addition, a Taylor chart was also used to demonstrate the performance of two ET estimations with different SSM observations in terms of R, ubRMSD, and Normalized Standard Deviation (NSD) on 2D plots, to display how closely the datasets matched observations in one diagram (Taylor 2001). In the Taylor diagram, NSD represents the radial distance from the origin point and the correlation with the site observations as an angle in the polar plot. The ubRMSD is the distance between the observation and the model and is represented in the Taylor chart as a green semi-circular arc with point A as the center of the circle. The closer the model point to the reference point (Point A), the better the performance. This diagram is convenient and visual in evaluating multiple aspects of various models.

The error variance of GPP and ET products was estimated using the triple collocation (TC) approach (Stoffelen 1998) to validate the global simulation in this study. The method has been extensively applied in the study of hydrology and oceanography (Caires and Sterl 2003; Khan et al. 2018; O’Carroll et al. 2008; Stoffelen 1998), particularly in SM studies (Chan et al. 2016; Kim et al. 2018). The TC provides a reliable platform for comparison of spatial assimilation results and in-situ measurements. In this experiment, no calculation was performed on the non-vegetated areas where the correlation was lower than 0.2 to have independent datasets and avoid correlated errors (crucial assumptions in TC) (Yilmaz and Crow 2014).

In this study, the five products were divided into three product categories, including satellite product (MODIS, GOSIF GPP), reanalysis product (GLASS, GLDAS) and data assimilation product (GLEAM ET, LPJ-VSJA) (Li et al.,2018). One product in each category was selected to form a group to calculate their error. The LPJ-VSJA product was set as the reference data.

For GPP products, GOSIF, GLASS, and LPJ-VSJA were treated as a group, and MODIS, GLASS and LPJ-VSJA were treated as another group to calculate the errors; the final errors were determined by the average of these two.

Similarly, to calculate the errors for ET, GLEAM, GLASS, and MODIS were chosen as a group; LPJ-VSJA, GLDAS, and MODIS were treated as a group; LPJ-VSJA, GLASS and MODIS were considered as a group. In order to reduce the influence of orthogonality hypothesis of error, the first and third groups are for indirect and effective comparison between LPJ-VSJA product and GLEAM product.

3. Experiment sites and data

3.1. Description of flux tower sites

We screened over 300 EC flux sites across the globe from the FLUXNET2015 (https://fluxnet.fluxdata.org/data/fluxnet2015-dataset/), AmeriFlux (http://public.ornl.gov/ameriflux), and the HeiHe river basin (Liu et al. (2018), http://www.heihedata.org)) for the evaluation of assimilation performance over the period from January 2010 to December 2018. The in-situ half-hourly LE and GPP data from the sites were aggregated into daily data. The daily gap-filled data were excluded if the percentage of gap-filled half-hourly values was more than 20%. Then we corrected the data of energy non-closure by using the Bowen ratio closure method (Twine et al. 2000) to improve the energy closure rate (Huang et al. 2015; Yang et al. 2020). The data were selected to cover the 2010–2018 period with at least one year of reliable data, and the result from the error of assimilation is relative to the LE value and seasonal variation (Purdy et al. 2018; Zou et al. 2017). It is essential to have available data every month during a one-year period, and only days with less than 25% missing data were processed per month (Feng et al. 2015). In addition, for flux tower data, the data were also excluded for the analysis if the SMAP/SMOS SSM data were not of good quality.

Finally, we identified a total of 105 sites across the globe encompassing five major biomes: grassland (18 for GPP and 19 for ET), savanna (11), shrubland (4), forest (49 and 53), and cropland (13 and 14). In the comparative analysis of the performance for simulating ET by assimilating SMOS and SMAP SSM data separately, we selected 46 AmeriFlux sites (Figure S3) with at least one year of reliable data from 2015 to 2018 based on the simultaneous availability of SMAP and SMOS data, including grassland (19), savanna (11), shrubland (5), forest (23), and cropland (7). Figure S2 and S3 illustrate the location and distribution of the 105 and 46 EC flux tower sites, respectively. A more detailed description is summarized in the Supporting Information Table S1.

3.2. Remote sensing datasets: LAI and SSM

The GLASS LAI product with an 8-day time step (8-day average) and 5 km resolution was derived from MODIS and CYCLOPES surface reflectance and ground observations using general regression neural networks (GRNNs) (Xiao et al. 2013; Xiao et al. 2016). The verification of the product using the mean values of high-resolution LAI maps showed that the GLASS LAI values were closer to these high-resolution LAI maps (RMSD= 0.78 and R^2= 0.81) (Xiao et al. 2016; Liang et al. 2013). Therefore, the GLASS LAI product has satisfactory performance and can be assimilated into terrestrial biosphere models.

The SMAP mission (Entekhabi et al. 2010) and SMOS mission (Jacquette et al. 2010), the two dedicated soil moisture satellites currently in orbit equipped with L-band microwave instruments, provide SSM retrievals. We chose the SMOS-L2 product and the SMAP-L3-Enhanced product, which both provide global coverage every three days for soil depth of 5 cm. Only good-quality SMAP and SMOS data were used. The grid cells with water areas larger than 10% and those with less than 50% good-quality data in one year were masked out, which alleviates the undesirable model simulations caused by the decrease in SMAP retrieval accuracy (Chan et al. 2016; O'Neill et al. 2010). We only adopted the data with an uncertainty below 0.1 m^3 m^{-3}, in the actual range (0.00–0.6 m^3 m^{-3}), and the temperature of the LSM observation layer (the second layer) was higher than 2 °C (Blyverket et al. 2019).

The GLASS LAI, SMOS and SMAP observations were resampled to 9 km for site simulation and 0.25° for regional simulation. The 8- day average of GLASS LAI were assimilated for each day, and the SMAP or SMOS SSM was assimilated every 3 days.

3.3. Model-forcing and validation datasets

In this study, the meteorological, soil property, and CO_2 concentration datasets were used to drive the LPJ-PM. The climate-driven datasets used for the initialization of the LPJ-DGVM are the atmospheric CO_2 concentrations (1901-2018) of ice-core measurements and atmospheric observations at the Mauna Loa Observatory and CRU TS4.03 version Climate data from 1901 to 1930 provided by the Climatic Research Unit (CRU) of the Climate Laboratory, University of East Anglia, UK, including monthly precipitation, surface temperature, cloud cover and wet day. In the simulation period of 2010-2018, the Modern Era Retrospective-Analysis for Research and Applications Version 2 (MERRA-2) was adopted, and the variables used included precipitation, temperature, cloud cover and relative humidity. Soil properties (including limited water content of vegetation at wilting points, field capacity and Soil porosity) from Harmonized World Soil Database (HWSD) V1.2 dataset (Wieder et al. 2014) were selected as inputs to the PT-JPLSM model. Table 2 provides the spatial and temporal characteristics of the model-forcing datasets in the LPJ-PM (submodule: LPJ-DGVM and PT-JPL$_{SM}$).

The GLASS LAI product, SMOS-L2 product and the SMAP-L3-Enhanced product were assimilated to simulate GPP and ET.For site simulation, in order to maintain consistency with the SMAP Enhanced 3 Level product (Entekhabi et al. 2010), model-forcing data were resampled to a 9 km spatial resolution based on EASE-2 projection grid. In the global spatial simulation, the model-forcing datasets were resampled to 0.25° based on the bilinear method to ensure the consistency of spatial representation.

Table 2. List of the selected forcing and remote-sensing datasets used in this study

Datasets	Variable	Period	Spatial resolution	References
CRU TS v4.1[a]	Cloud cover, temperature, precipitation, wet day	1901-1930	0.5°× 0.5°	New et al. (2000), https://crudata.uea.ac.uk/cru/data/hrg/
Ice-core measurements and atmospheric observations at the Mauna Loa Observatory [a]	Atmospheric CO_2 concentrations	1901-2018	NA	(Etheridge et al. (1996); Keeling et al. (1995)) , https://scrippsco2.ucsd.edu/da

				ta/atmospheric_co2/
MERRA-2 [a]	Precipitation, surface temperature, cloud fraction, relative humidity	2010-2018	0.5°× 0.625°	Rienecker et al. (2011) （https://www.esrl.noaa.gov/psd/）
HWSD (v121)[b]	Soil texture data	NA	1 km×1 km	Wieder et al. (2014) （http://daac.ornl.gov）
SPL3SMP_E[b]	Surface soil moisture	2015.4–present	9 km×9 km	Entekhabi et al. (2010), (https://smap.jpl.nasa.gov/)
GLASS LAI [a,b]	Leaf area index	2010-2018	5 km×5 km	Xiao et al. (2016), (http://www.glass.umd.edu/Download.

				html)
SMOS_L3 CATDS[b]	Surface soil moisture	2010-present	25km×25 km	Jacquette et al. (2010),(https: //earth.esa.int /eogateway/ missions/smos)

[a]: forcing dataset for LPJ-DGVM

[b]: external input dataset for PT-JPL_{SM}

We used four global ET products and three global GPP products (Li et al. 2018; Li and Xiao 2019; Wang et al. 2017) that was resample to 0.25° to evaluate the performance of the model with the joint assimilation scheme. Table 3 shows the details of these GPP and ET products.

Table 3. Global GPP and ET products for comparison in this study

Product	Dataset	Temporal resolution	Spatial resolution	Retrieval algorithm	References

MOD17A2	GPP and ET	8-day average	1 km × 1 km	GPP: Based on the light use efficiency (LUE) model ET: Improved Penman formula	Running et al. (2004)
GLASS	GPP and ET	8-day average	5 km × 5 km	GPP: EC-LUE model ET: Combining five Bayesian averages based on process models (BMA)	Yuan et al. (2010)
GOSIF GPP	GPP	8-day average	0.05° × 0.05°	Estimated from solar-induced chlorophyll fluorescence with GPP-SIF relationships	Li and Xiao (2019)
GLDAS ET	ET	daily	0.25°× 0.25°	Processed model assimilation	Fang et al. (2009)
GLEAM v3a ET	ET	daily	0.25°× 0.25°	Processed model assimilation	Martens et al. (2017)

4. Results

4.1. Performance of LPJ-PM for simulating GPP and ET with the assimilation of LAI and soil moisture

4.1.1 Accuracy assessment of GPP for separate and joint assimilation

In general, the R^2 between GPP_{LPJ} and GPP_{OBS} was above 0.4 at most of the sites (62 sites) and were relatively weak for some sites. The LAI assimilation improved the simulations at most sites (R^2 value increased at 82 sites), particularly for sites in the U.S. and Europe (Figure S4). The R^2 improvement from the LAI assimilation (LAI-only assimilation) was superior to that from the SSM assimilation (Figure S4- (b) and (c)). The performance of the joint assimilation was similar to that of LAI-only assimilation. Sites (Figure S5 (a)) showed positive BIAS (GPP_{OBS}-GPP_{LPJ}) were mainly distributed in the humid and sub-dry humid forest, grassland, and arid cropland regions, showing an underestimation for GPP_{OBS}. The assimilation improved the accuracy for overestimated sites, but there was no significant improvement for underestimated sites. The ubRMSD implied that the SSM assimilation alone had a better performance than the LAI assimilation alone, especially for sites in arid areas(Figure 2). The analysis of the above three statistical measures (R^2, BIAS, and ubRMSD) indicated that the accuracy of joint assimilation was much better than that of separate assimilation.

At the seasonal scale, all three assimilation schemes corrected the model trajectory and significantly improved the growing season simulations, especially for peak values (IT-Tor, US-NR1, US-NE1) (Figure 3). In addition, the linear fitting of GPP_{joint} and GPP_{OBS} on a monthly scale was closer to 1:1 ($y= 0.92x + 21.66$ $p < 0.001$) than that of GPP_{LAI} ($y= 0.89x + 28.3$, $p < 0.001$) and GPP_{SM} ($y= 0.86x + 41.70$, $p < 0.001$) (Figure S9). The results in Table S2 support the above analysis, and the joint assimilation showed advantages in overall accuracy in both arid and humid areas.

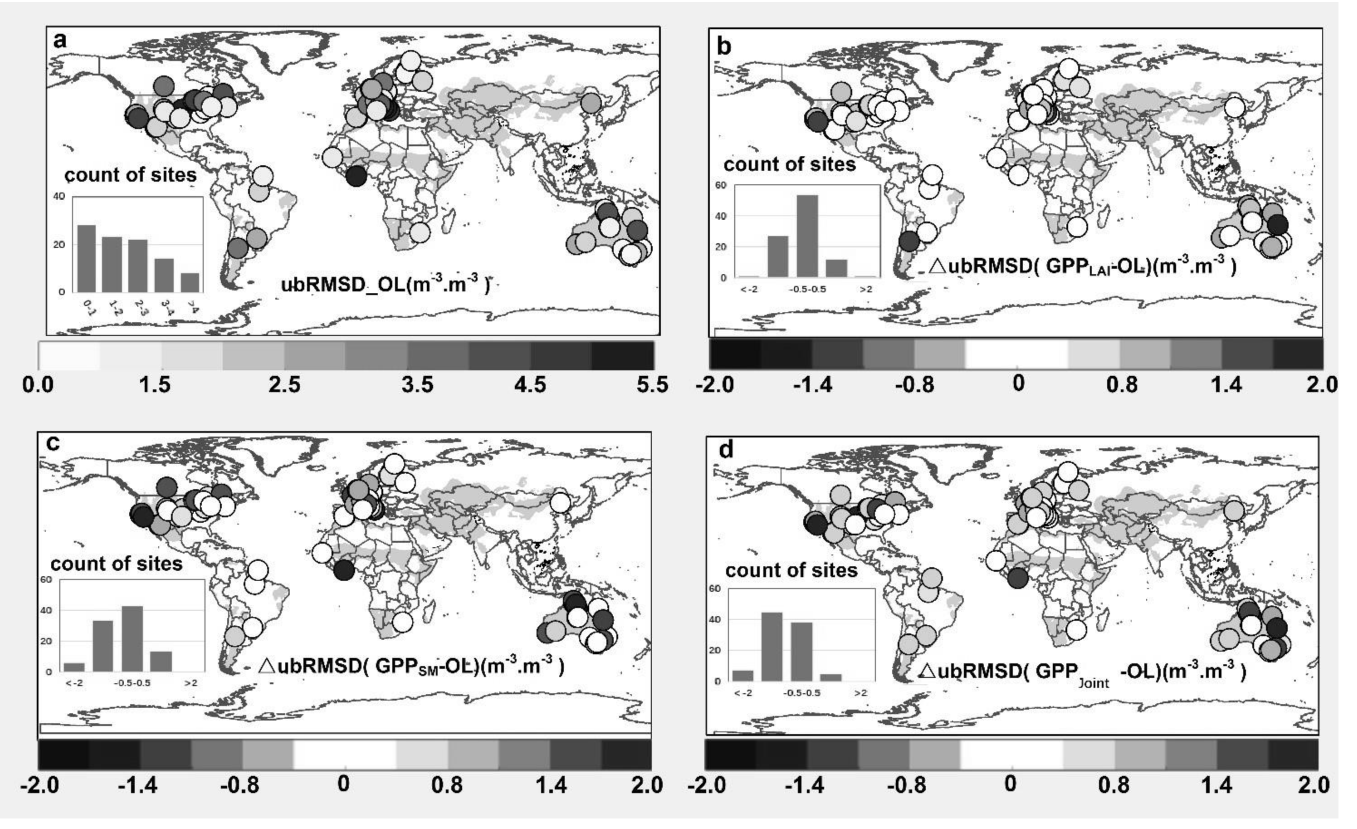
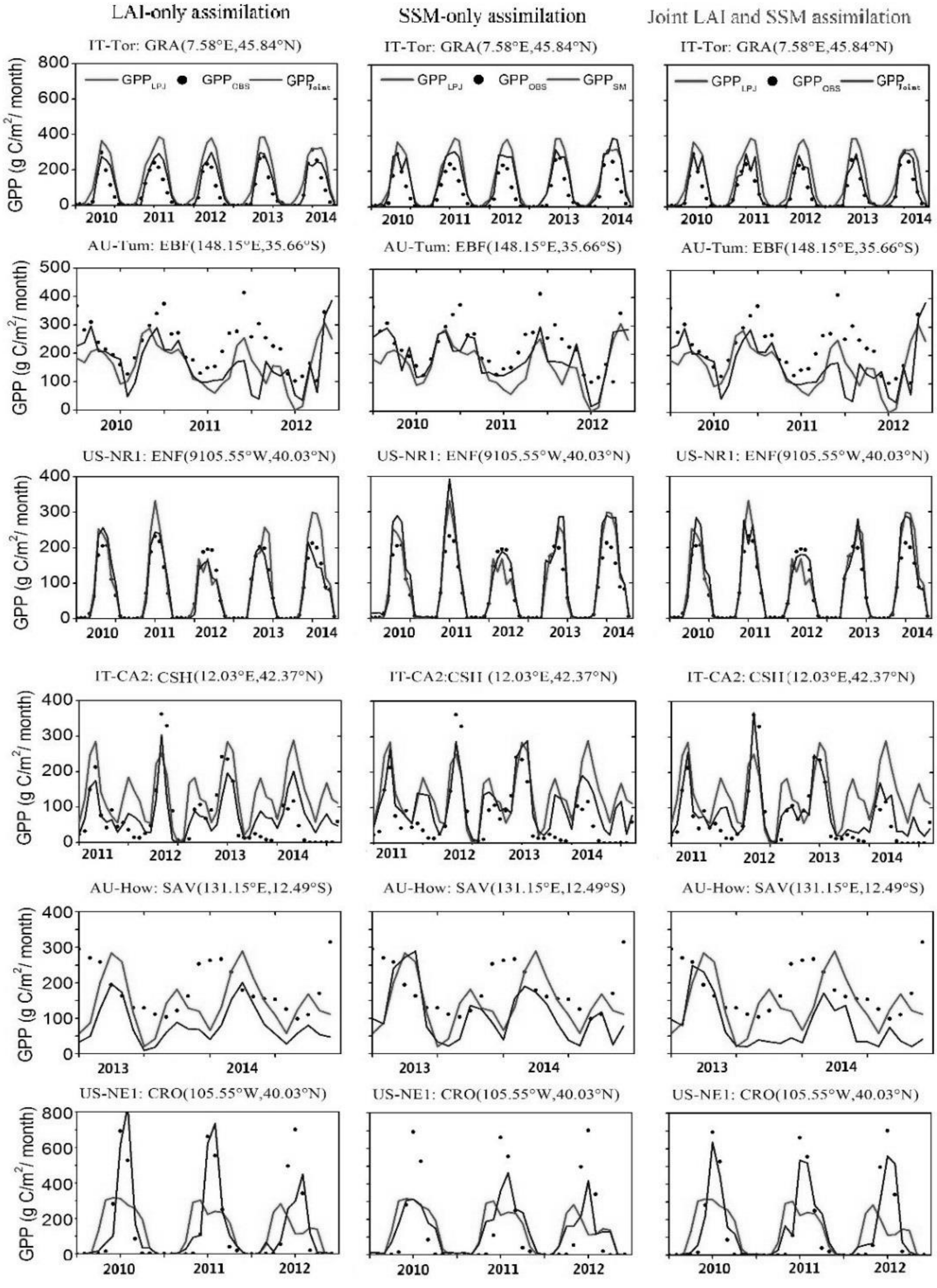


Figure 2 (a)The Unbiased Root Mean Square Error (ubRMSD) between the GPP_{LPJ} and the site observations, the yellow/blue indicating low/high ubRMSD; (b)△ubRMSD（GPP_{LAI}- GPP_{LPJ}）;(c)△ubRMSD（GPP_{SM}- GPP_{LPJ}）; (d)△ubRMSD（GPP_{Joint}- GPP_{LPJ}） , blue/red represent positive/negative values.


LAI-only assimilation
SSM-only assimilation
Joint LAI and SSM assimilation
IT-Tor: GRA(7.58°E,45.84°N)
AU-Tum: EBF(148.15°E,35.66°S)
US-NR1: ENF(9105.55°W,40.03°N)
IT-CA2: CSH(12.03°E,42.37°N)
AU-How: SAV(131.15°E,12.49°S)
US-NE1: CRO(105.55°W,40.03°N)
GPP (g C/m2/ month)
GPP_LPJ
GPP_OBS
GPP_Joint
GPP_SM

473

Figure 3. Seasonal cycles of tower GPP and simulated gross primary productivity (GPP) from Lund-Potsdam-Jena (LPJ), LAI-only assimilation, SSM-only assimilation and joint assimilation for six sites representing six PFTs.

The residual analysis indicated that the three assimilation schemes for GPP (Figure S11 (left)) were different. For the assimilation results, most of the errors were distributed around –70 ~ 60 g C m^{-2} $month^{-1}$. The high GPP_{OBS} values were considerably underestimated. The maximum negative error reached 100 g C m^{-2} $month^{-1}$. The error distribution of GPP_{SM} was more dispersed than that of GPP_{LAI} and GPP_{joint}. Among the residuals of these three schemes, GPP_{SM} significantly overestimated the GPP_{OBS}, mainly distributed in the 0–200 g C m^{-2} $month^{-1}$ range. GPP_{LAI} showed significant improvement in the overestimation of GPP_{OBS} compared with GPP_{joint}. In general, the GPP_{joint} with the most concentrated error distribution had significant improvement.

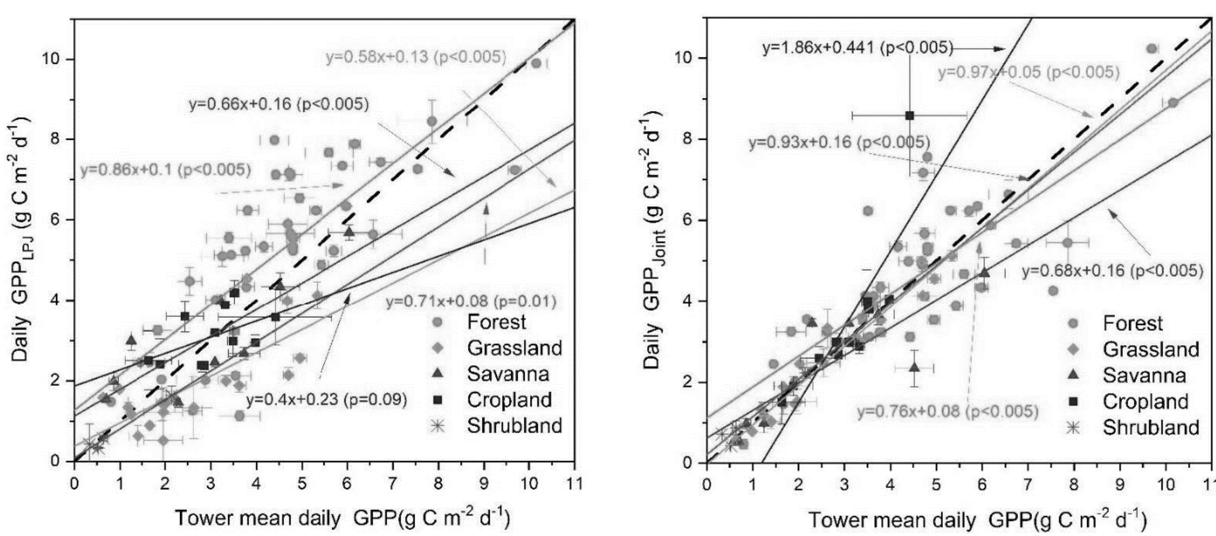


Figure 4. Scatterplots of daily GPP_{LPJ} (left) and GPP_{joint} (right) versus tower GPP for different PFTs.

After determining the optimal assimilation scheme (Joint LAI and SSM assimilation scheme), we evaluated the GPP_{LPJ} and GPP_{joint} at the site level (Figure.4). The results showed that GPP_{joint} performed better (R^2= 0.83, ubRMSD= 1.15 g C m^{-2} d^{-1}) than GPP_{LPJ} (R^2= 0.69, ubRMSD= 1.91 g C m^{-2} d^{-1}). The noticeable underestimation in all PFTs and overestimation at most forest sites for GPP_{LPJ} were corrected by joint assimilation (GPP_{joint}). Our joint assimilation methods had better performance in forests, shrublands, and grasslands than in croplands and savannas. Except for the cropland, the linear fitting results of other types were all below the 1:1 line, showing the overall underestimation. Superior performance in both original simulation and assimilation occurred at shrubland (R^2= 0.93, ubRMSD= 0.89 g C m^{-2} d^{-1}) and grassland (R^2= 0.97, ubRMSD= 0.83 g C m^{-2} d^{-1}) sites. However, the standard deviation of GPP_{joint} and GPP_{OBS} at savanna sites was relatively large, and the GPP_{joint} at several savanna sites was significantly underestimated.

4.1.2 Accuracy assessment of ET for separate and joint assimilation

In general, the coefficient of determination (R^2) between ET_{LPJ} and ET_{OBS} was generally over 0.4 (the simulations were superior to GPP_{LPJ}) (Figure S6). ET_{LAI} showed slightly higher R^2, while some sites showed reduced values (41 sites). The ET_{SM} and ET_{joint} were significantly improved compared with the ET_{LAI}. The R^2 increased considerably in Australia but declined at some sites in the United

States after assimilation. For ubRMSD, ET_{joint} performed better than ET_{SM} and ET_{LAI}. The SSM assimilation improved more in humid regions, while the ubRMSD of ET_{SM} was slightly higher in South America (Figure 5). In the original LPJ-DGVM simulation, the sites with a negative BIAS were mostly located in the humid and sub-dry humid regions, while most of the sites in arid and semi-arid regions had underestimation (Figure. S7- (a), Table S3). The assimilation improved ET at some of the overestimated sites, but the underestimation over these sites showed little improvement.

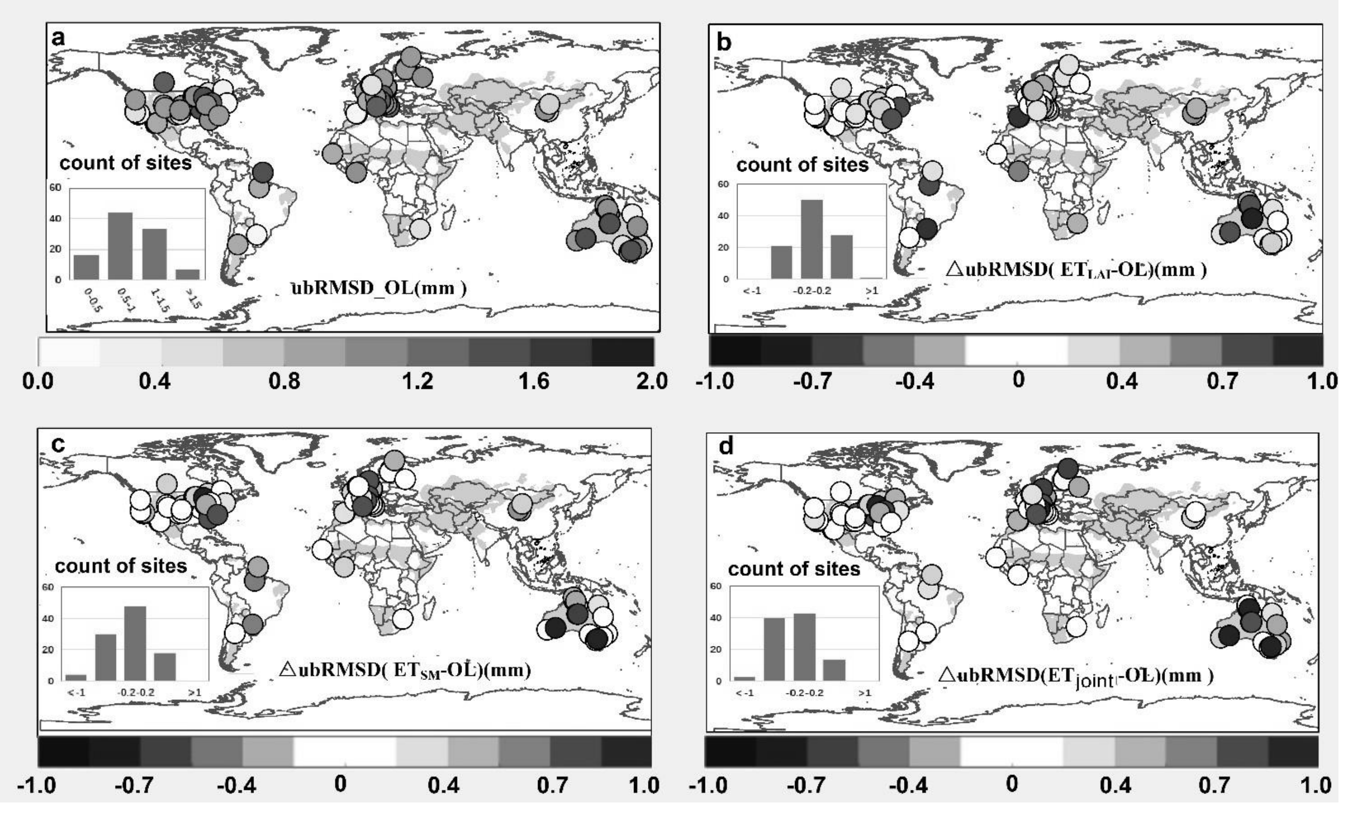


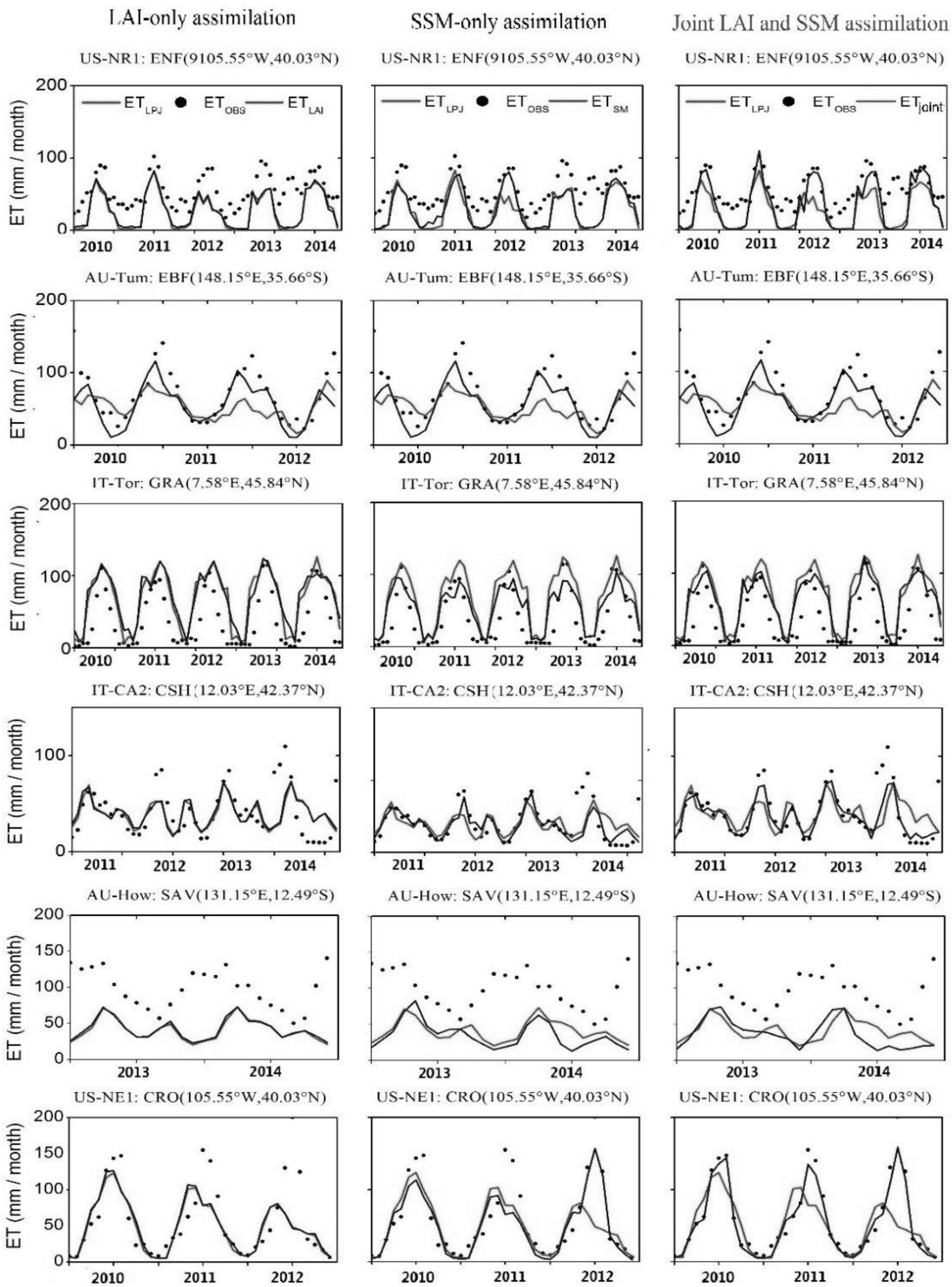
Figure 5 (a) The Unbiased Root Mean Square Error (ubRMSE) between the ET simulated by the LPJ-DGVM and the site observations, with yellow/blue indicating low/high ubRMSD; (b)△ubRMSD（ET_{LAI}- ET_{LPJ}）;(c) △ubRMSD（ET_{SM}- ET_{LPJ}）; (d)△ubRMSD（ET_{Joint}- ET_{LPJ}）, blue/red represent positive/negative value.

511

At the seasonal scale, the model simulations were able to capture the temporal trend of ET_{OBS}, and joint assimilation significantly improved the simulation in the growing season (US-NR1, US-NE1); overall underestimation was observed for ET_{OBS}, especially in winter (Figure 6). Overall, the linear fitting of monthly ET_{joint} and ET_{OBS} was closer to 1:1 than that of ET_{LAI} and ET_{SM} (Figure S6). The simulation accuracy of joint assimilation was better than that of separate assimilation, and the performance of the SSM assimilation was better than that of the LAI assimilation.

The ET residual analysis (Figure S11 (right)) indicated that the three assimilation scheme errors showed underestimation for ET_{OBS}. In general, the error distribution of separate assimilations was more dispersed than that of the joint assimilation. Similar to the assimilation performance of GPP, ET_{joint} and ET_{SM} significantly improved the overestimation of ET_{OBS}, but did not significantly improve the underestimation. For the ET_{joint}, most of the errors were distributed around -30–18 mm month^{-1}. The region with high ET_{OBS} was considerably underestimated, and the maximum negative error reached –57 mm month^{-1}.

We also evaluated the ET assimilation results at the PFT scale (Figure 7). The results showed that our ET values resulting from the assimilation performed better at the site level (R^2= 0.77, ubRMSD= 0.65 mm d^{-1}) than that of ET_{LPJ} (R^2= 0.67, ubRMSD=0.95 mm d^{-1}). Joint assimilation significantly reduced the errors of those shrubland sites with overestimation for ET_{OBS}, and the site distribution was closer to the 1:1 line. Our assimilation methods had better performance in forest, savanna, and grassland ecosystems than in cropland and shrubland (Table S3). The linear fitting results of grassland and shrubland were all above the 1:1 line, showing overall overestimation. Although the original simulation and assimilation performance were superior at savanna sites (R^2= 0.95, ubRMSD= 0.78 mm d^{-1}), the standard deviations of ET_{joint} and ET_{OBS} at savanna sites were relatively large, which was similar to the GPP results at savanna sites.


LAI-only assimilation
SSM-only assimilation
Joint LAI and SSM assimilation
US-NR1: ENF(9105.55°W,40.03°N)
AU-Tum: EBF(148.15°E,35.66°S)
IT-Tor: GRA(7.58°E,45.84°N)
IT-CA2: CSH(12.03°E,42.37°N)
AU-How: SAV(131.15°E,12.49°S)
US-NE1: CRO(105.55°W,40.03°N)
ET (mm / month)
ET_{LPJ}
ET_{OBS}
ET_{LAI}
ET_{SM}
ET_{joint}

535

Figure 6. Seasonal cycles of tower-based and simulated ET from Lund-Potsdam-Jena (LPJ), LAI-only assimilation, SSM-only assimilation and joint assimilation for the six sites representing six PFTs during the study period.

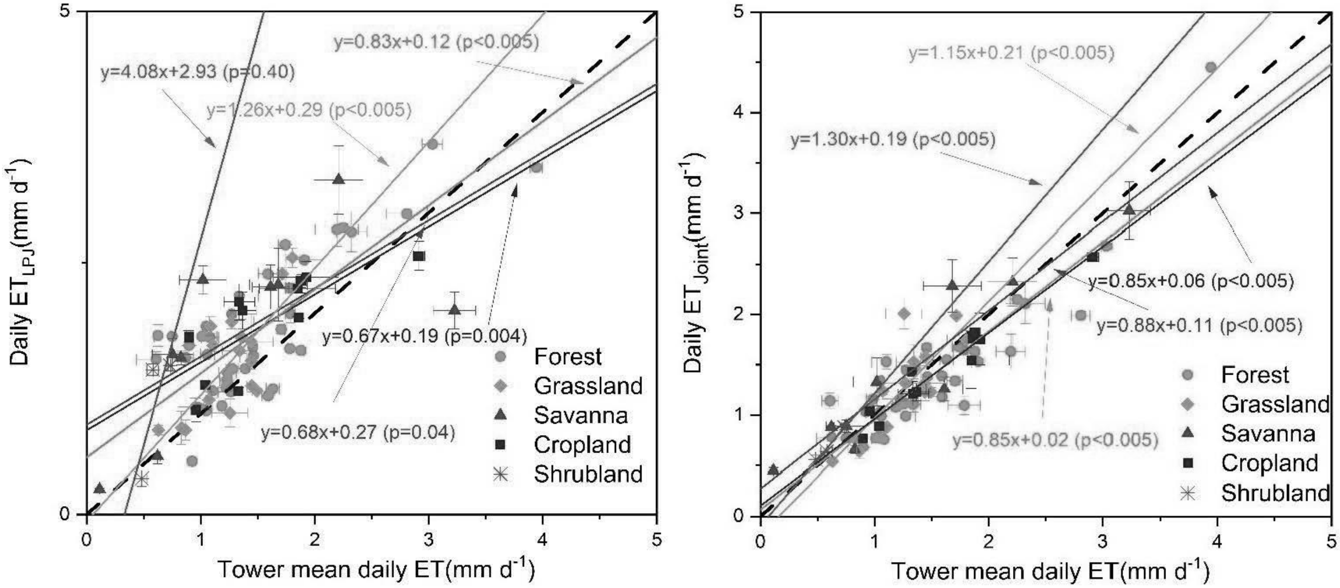
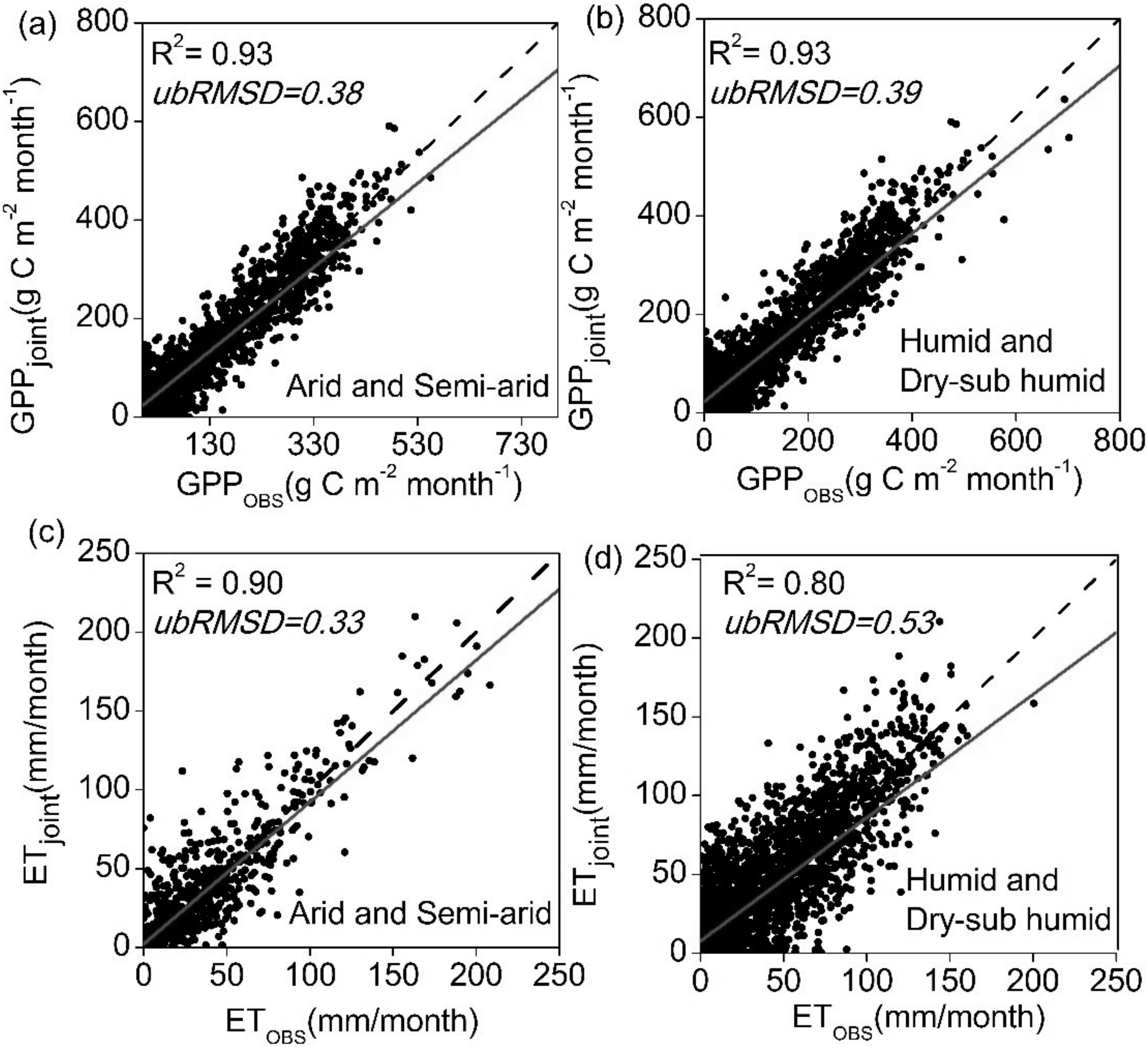


Figure 7. Scatter plots of daily ET_{joint} versus tower ET under different PFTs.

4.2. Comparison of assimilation performance in semi-arid and arid regions with that in humid and sub-dry humid regions



(a)
R^2= 0.93
ubRMSD=0.38
Arid and Semi-arid
GPP$_{joint}$(g C m^{-2} month^{-1})
GPP$_{OBS}$(g C m^{-2} month^{-1})
(b)
R^2= 0.93
ubRMSD=0.39
Humid and
Dry-sub humid
(c)
R^2 = 0.90
ubRMSD=0.33
Arid and Semi-arid
ET$_{joint}$(mm/month)
ET$_{OBS}$(mm/month)
(d)
R^2= 0.80
ubRMSD=0.53
Humid and
Dry-sub humid

543

Figure 8. Scatter plots of daily tower GPP and ET versus GPP_{joint} and ET_{joint} under arid and humid sites: (a) and (c) are the fitting results of GPP and ET in arid and semi-arid regions, respectively; (b) and (d) are the fitting results of GPP and ET in humid and dry sub-humid zone, respectively.

During the period 2010–2014, monthly GPP_{joint} and ET_{joint} performed differently in humid and sub-dry humid regions and semi-arid and arid regions (Figure 8, Table S2,3). Overall, the GPP and ET simulations had good consistency with the tower data in the two regions. For GPP_{joint}, there was no significant difference in the correlation and fitting coefficients between the two regions. As for ET_{joint}, the fitting results and R^2 values in the semi-arid and arid regions performed better than those in the humid and sub-dry humid regions, which also suggested the importance of SSM for ET estimation in water-limited areas.

On the daily scale, the original GPP simulations (GPP_{LPJ}) performed better in the semi-arid and arid regions than in the humid and sub-dry humid regions with higher R^2 and lower ubRMSD (Table S2). the R^2 and BIAS implied that the LAI assimilation alone had a better performance than the SSM assimilation alone. However, for sites in arid and semi-arid areas, the ubRMSD showed that the GPP_{SM} improved better than GPP_{LAI}, which both demonstrated SSM data are essential in water-limited regions. For GPP_{joint}, the shrubland in the semi-arid and arid regions had the lowest R^2 values and the second lowest ubRMSD. The forest in the semi-arid and arid regions had the largest improvement after assimilation. In the humid and sub-dry humid regions, the GPP_{joint} of the savanna and cropland showed the largest improvement (R^2 increased by 64.7% and 71.1%, respectively; ubRMSD decreased by 47.0% and 31.8%, respectively). The grassland in the semi-arid and arid regions had the highest R^2, and the savanna by combining all indicators had the best assimilation results compared to other types in both regions.

Similar to ET_{joint}, the ET_{LPJ} in the semi-arid and arid regions was better than that in humid and sub-dry humid regions in terms of four evaluation indicators (ubRMSD decreased by 34.4% in semi-arid and arid regions and the ubRMSD decreased by 30.9% in humid and sub-dry humid regions compared with ET_{LPJ}). The R^2 and ubRMSD implied that the SSM assimilation alone had a better performance than the LAI assimilation alone, especially for sites in arid areas. and the BIAS showed that the ET_{LAI} improved better than ET_{SM} for sites in humid and sub-dry humid areas. The performance of the original simulation and assimilation of grassland sites in the semi-arid and arid regions was the best among all five PFTs.

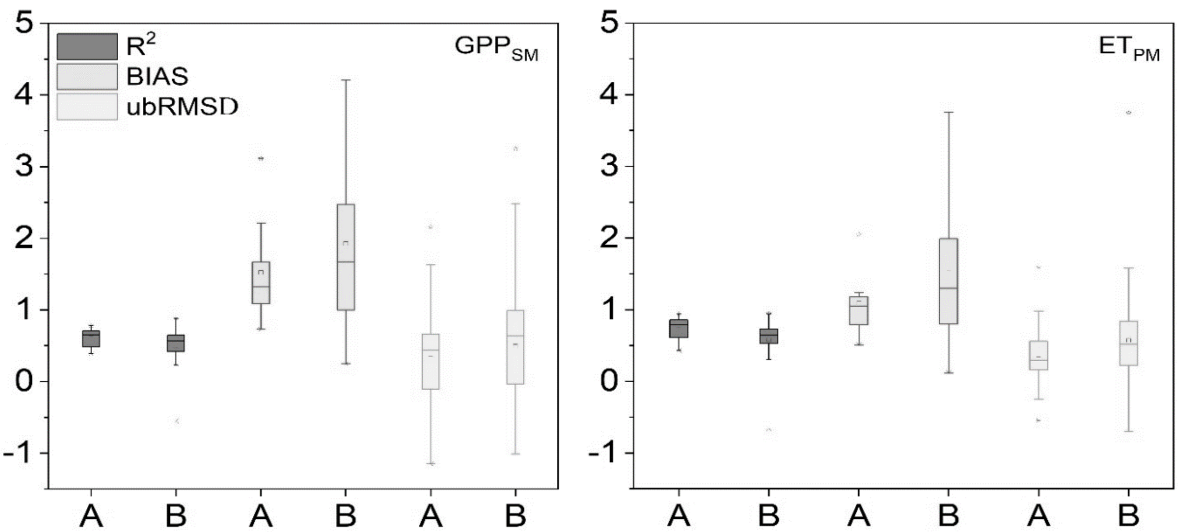


Figure 9. Boxplots of R^2, ubRMSD and BIAS for GPP_{SM} (left) and ET_{PM} (right). A represents the sites in arid and semi-arid areas, and B represents the sites in humid and dry sub-humid areas.

To investigate the reasons for better assimilation performance in water-limited regions, we evaluated the GPP and ET simulated by the LPJ-PM according to R^2, ubRMSD, and BIAS (Figure 7). Compared with the semi-arid and arid regions, the humid and sub-dry humid region had smaller R^2 mean, larger BIAS, and no significant difference in mean ubRMSD for GPP_{SM}. In general, the evaluation results of joint assimilation for ET_{PM} were generally consistent with those for GPP_{SM} and GPP_{SM}. ET_{PM} showed underestimation, which was consistent with the underestimation in SSM assimilation. These results indicated that, both GPP and ET modeled by LPJ-PM with joint assimilation were less stable and had a lower performance in the humid and sub-dry regions than in the semi-arid and arid regions.

4.3. Comparison of assimilation performance in assimilating SMOS and SMAP soil moisture data

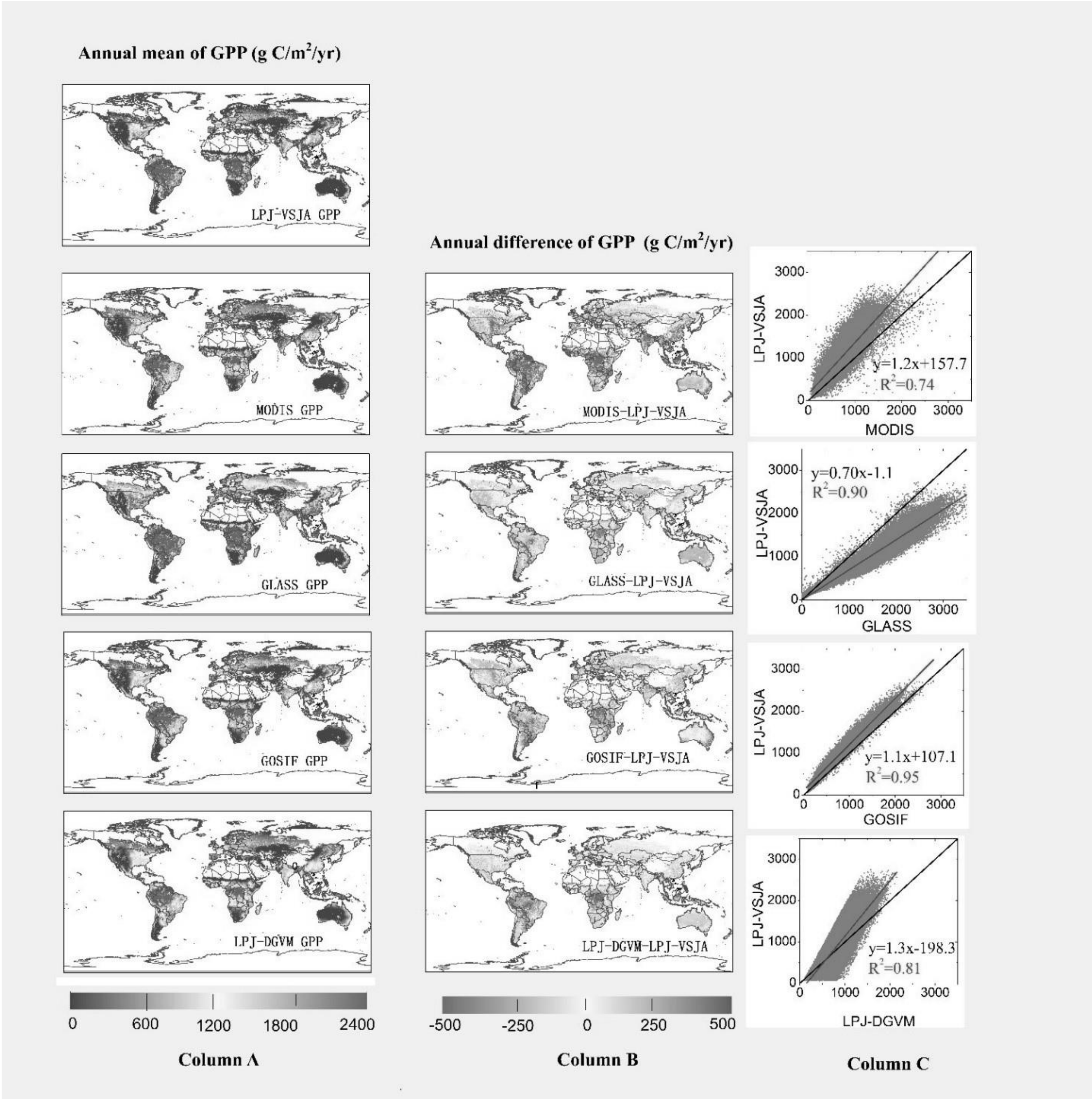


Figure 10. Taylor diagram (left) comparing ET simulations with observations at all 46 AmeriFlux sites at the daily time step between April 2015 and December 2018. Blue dots represent results based on

assimilation with SMAP SSM only and red dots represent results based on assimilation with SMOS SSM only. Reference points A and B-F correspond to the vegetation functional types (PFTs). The grid diagram (right) compares the evaluation indices of ET simulations with those of the observed values at all 46 AmeriFlux sites with different wet and dry zones at the daily time step; the yellow cells indicate that ET_{SMAP} performs better in the metric, and green cells indicate that ET_{SMOS} performs better in the metric.

The Taylor chart was used to compare the assimilation performance of ET_{SMAP} and ET_{SMOS} at 46 AmeriFux sites (Figure 10-left). The results showed that ET_{SMAP} performed better than ET_{SMOS} for most PFTs, except forest. Both ET_{SMAP} and ET_{SMOS} performed well for grassland (closer to point A), and there was little difference between R^2 and ubRMSD. The NSD of ET_{SMAP} in grassland was 0.88, which was closer to 1 than that of ET_{SMOS}. The assimilation of ET in the forest had a lower R^2 and higher ubRMSD (0.7-0.8) than those of other PFTs, and the NSD of cropland and shrubland was lower than that of other PFTs (0.6-0.8), indicating that the assimilation for cropland and shrubland could not reproduce the variations in ET effectively. However, ET_{SMAP} showed significant improvement in R^2 compared with ET_{SMOS} for shrubland and cropland. The assimilation performance of ET_{SMAP} and ET_{SMOS} for savanna showed the greatest difference. In general, the ET_{SMAP} and ET_{SMOS} were slightly different, and the ET_{SMAP} was more improved than ET_{SMOS}.

Figure 10 (right) shows the assimilation accuracy of ET_{SMOS} and ET_{SMAP} in different humid and arid regions. The ET_{SMAP} had significant advantages for the four indicators. The R^2 of ET_{SMAP} was higher than that of ET_{SMOS} in all the areas. However, ET_{SMOS} in some evaluation indicators showed a better performance than ET_{SMAP} (BIAS in the humid region; ubRMSD in the sub-dry humid region). This may be due to the overall more humid nature of SMOS SSM than the SMAP SSM. Moreover, the sensitivity of deep soil moisture contributed more to the ET in humid areas than in the water-limited areas.

4.4. Global simulations of GPP and ET with joint assimilation of LAI and soil moisture data

To assess the spatial scalability of the LPJ-VSJA assimilation scheme, we simulated the global daily GPP and ET for 2010–2018 with a spatial resolution of 0.25°. The original results simulated by the LPJ-DGVM and LPJ-VSJA were referred to as LPJ-DGVM GPP(ET) and LPJ-VSJA GPP(ET), respectively. We compared the annual spatial GPP and ET values and the error standard deviation of the LPJ-VSJA with several existing flux products.

Figures 11 and 12 depict the spatial distribution of the annual mean and the differences between our simulation results and the global independent satellite-based products. The developed LPJ-VSJA GPP was the closest to GOSIF GPP (Li and Xiao 2019) in most regions with the lowest spatial mean deviation (LPJ-VSJA-GOSIF) (27.9 g C m^{-2} yr^{-1}), followed by GLASS GPP (51.2 g C m^{-2} yr^{-1}) (Yuan et al. 2010), LPJ-DGVM (-73.4 g C m^{-2} yr^{-1}), and MODIS GPP (93.1 g C m^{-2} yr^{-1}). LPJ-VSJA had higher GPP values than GOSIF GPP in tropical regions, such as Amazonia, Central Africa, and Southeast Asia. In general, the annual mean and differences between MODIS, GOSIF GPP, LPJ-DGVM, and our LPJ-VSJA were in broad agreement (with higher R^2 ranging from 0.74 to 0.95).

LPJ-VSJA ET was the closest to GLEAM ET on the spatial average with the least spatial average deviation (-13.9 mm yr^{-1}) and highest R^2 (0.88), followed by GLASS ET (-23.1 mm $yr^{-1}$and 0.82), GLDAS ET (-34.7 mm yr^{-1} and 0.73), LPJ-DGVM (-48.7 and 0.66 mm yr^{-1}), and MODIS ET (-122.1and 0.54 mm yr^{-1}).

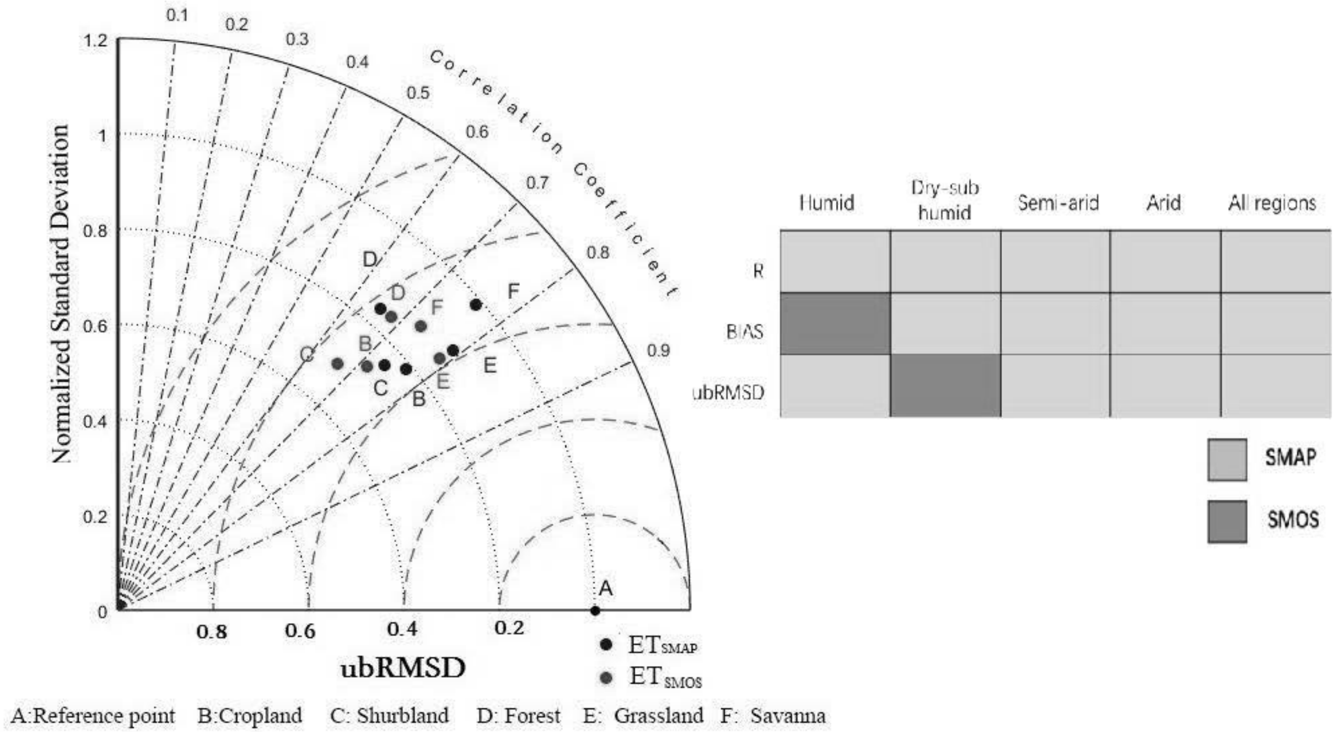


Figure 11. Column A: Spatial distribution of annual LPJ-VSJA GPP and other independent satellite-based datasets (a: MODIS GPP; b: GLASS GPP; c: GOSIF GPP; e: LPJ-DGVM). Column B: Spatial distribution of the difference between annual LPJ-VSJA GPP and other independent satellite-based datasets. Column C: Scatter plots between these products. Black lines show the 1:1-line, red lines show the regression fit.

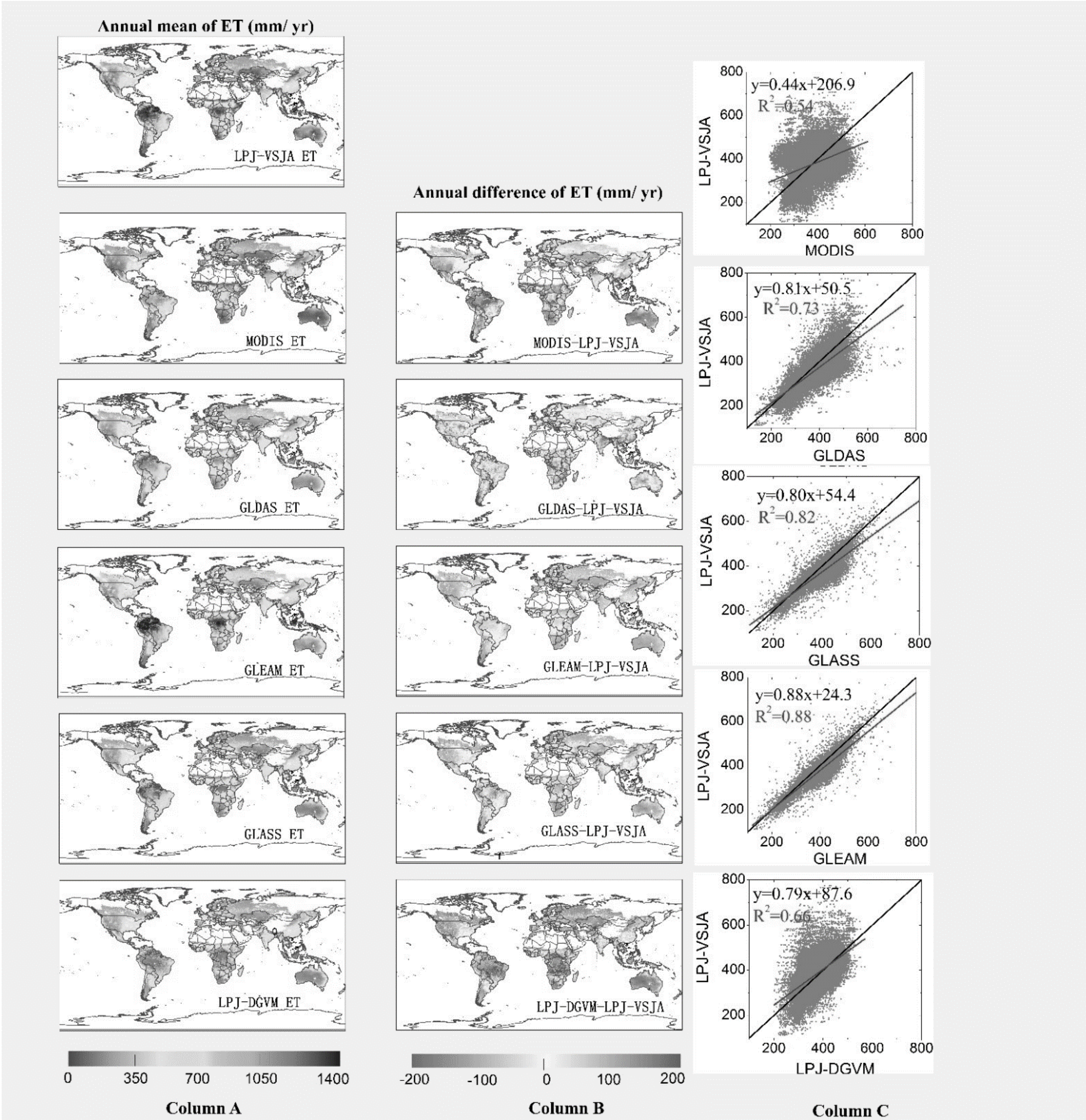


Figure 12. Column A: Spatial distribution of annual LPJ-VSJA ET and other independent satellite-

based datasets (a: MODIS GPP; b: GLDAS ET; c: GLEAM ET; d: GLASS ET; e: LPJ-DGVM ET). Column B: Spatial distribution of the difference between annual LPJ-VSJA ET and other independent satellite-based datasets. Column C: Scatter plots between these products are provided on the right of the difference maps. Black lines show the 1:1-line, red lines show the regression fit.

Figure 13 (a)–(e) represent the spatial error standard deviation (σ) distribution of MODIS, GLASS, GOSIF, and LPJ-VSJA GPP, respectively. The graphs on the right side depict the corresponding histograms. The σ of the MODIS GPP was evenly distributed between 30 and 60 g C m^{-2} $month^{-1}$, while the average σ of other products was concentrated in 0–20 g C m^{-2} $month^{-1}$ (90%). The high errors of all products were concentrated in the high temperature and humid areas of southern North America, eastern South America, humid and dry sub-humid areas of South Asia, and the savannas of Africa and Australia. The error histogram of GOSIF GPP and LPJ-DGVM GPP were in line with the normal distribution, with an average value of 8.3 g C m^{-2} $month^{-1}$ and 22.4 g C m^{-2} $month^{-1}$. The GLASS GPP product had the lowest mean value (3.6 g C m^{-2} $month^{-1}$), followed by LPJ-VSJA (4.7 g C m^{-2} $month^{-1}$), but the error variance of the LPJ-VSJA product was the lowest, indicating a stability of the regional error (Table S4). Compared to the LPJ-DGVM, the joint assimilation results showed improvement in all regions (the average error reduced by 17.7 g C m^{-2} $month^{-1}$), especially in the humid regions of South Asia, Australia, and the United States. Our LPJ-VSJA GPP was generally proven to have high accuracy and stability for spatial analysis and could provide a reference for other model products.

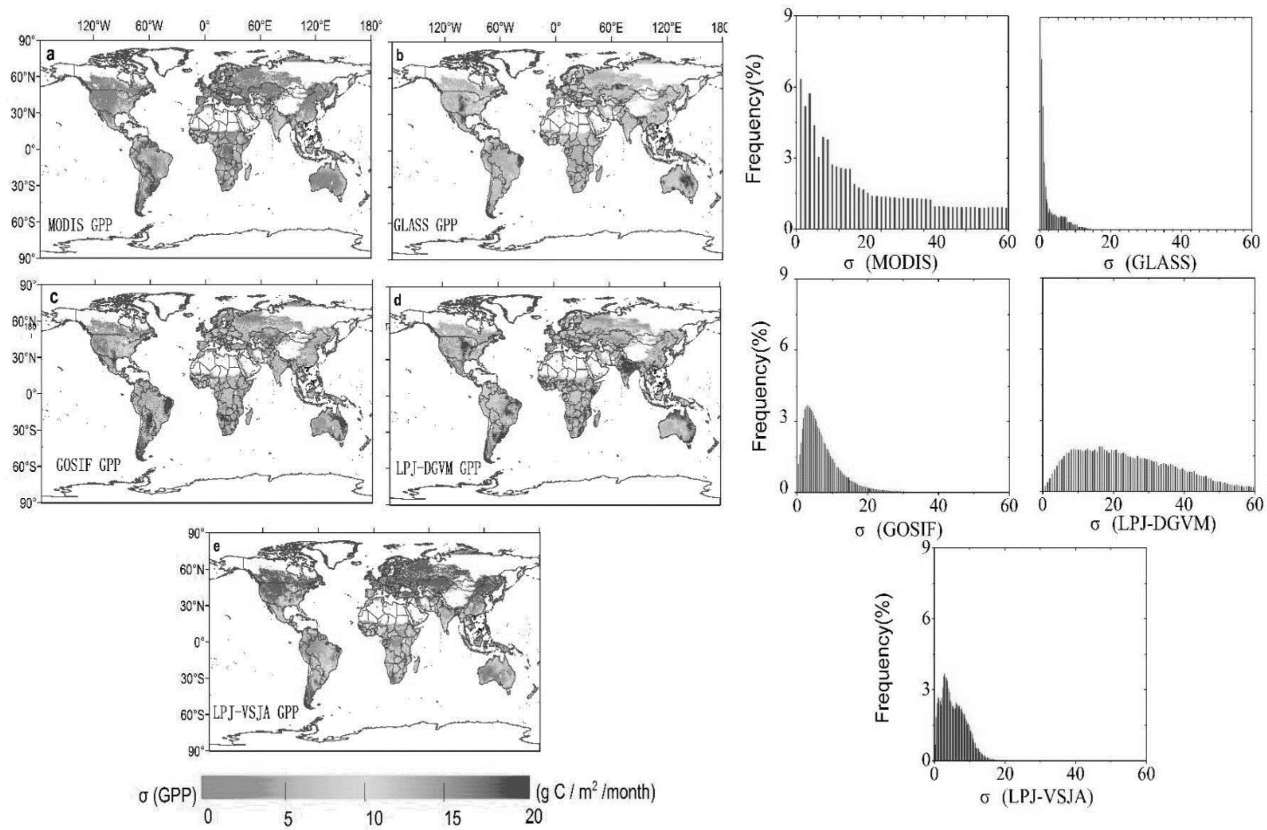


Figure 13. Spatial distribution and histograms of error standard deviation (σ) for global GPP products: MODIS (a), GOSIF (b), GLASS (c), LPJ-DGVM (d), and LPJ-VSJA (e).

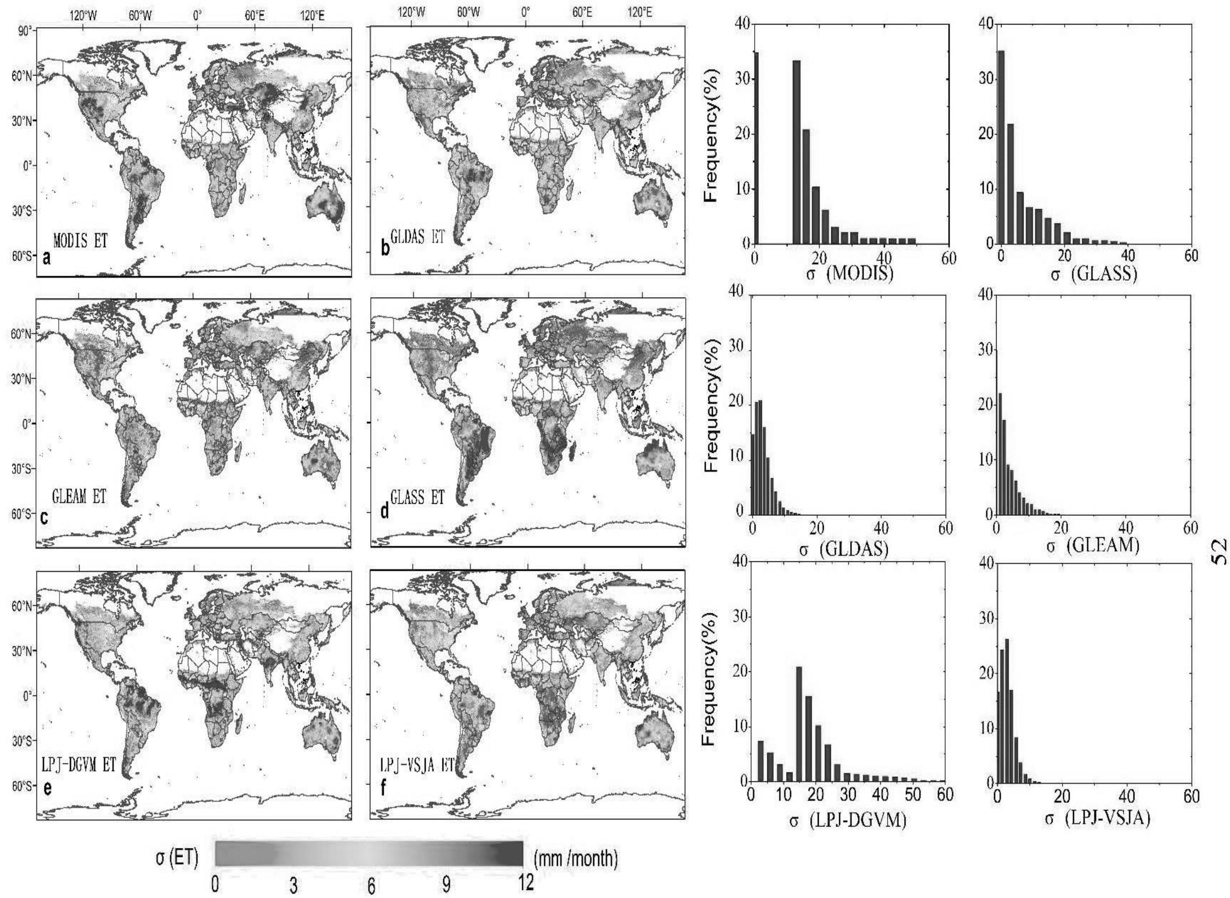


Figure14. Spatial distribution and histograms of error standard deviation (σ) for global ET products: MODIS (a),GLDAS (b),GLEAM (c), GLASS (d), LPJ-DGVM (e), and LPJ-VSJA (f).

Figures 14 (a)–(f) show the σ of MODIS, GLDAS, GLEAM, GLASS, and LPJ-VSJA ET (the units are mm/month), and the right graphs are the corresponding histograms. The σ values of GLDAS and LPJ-VSJA represented a normal distribution trend. Except for MODIS, GLASS, and LPJ-DGVM (0–60 mm $month^{-1}$), the σ of other products was generally between 0-20 mm $month^{-1}$. The simulation error was relatively smaller in the Northern Hemisphere than in the Southern Hemisphere, especially for GLASS ET and GLDAS ET. Significant improvements in joint assimilation were observed in the northern hemisphere (especially in the semi-arid areas of the western United States and savanna and cropland areas of central India) and African savanna areas, and the average error was reduced by 15.1 mm $month^{-1}$. In general, the error mean and variance of LPJ-VSJA and GLEAM products were relatively low (Table S4), and there was no apparent extremely high value region in the error distribution. Among the five products, LPJ-VSJA had the lowest error mean and variance and the highest accuracy.

5. Discussion

5.1 Advantage of joint assimilation for GPP and ET

The benefit of employing multiple data flows in an assimilation system is the complementarity of the data, which enables constraints on different components of the underlying process-based terrestrial biosphere model. Due to the interaction and feedback between the internal components of the model, the assimilation of multiple observations has a synergistic effect, and the integrated constraints are greater than the individual constraint (Kato et al. (2013)). The advantage of our joint assimilation is that it can improve the simulation accuracy of both GPP and ET, especially ET, in arid and semi-arid regions.

In the GPP assimilation experiment, the performance of the LAI assimilation was better than that of the SSM assimilation possibly for two reasons: (1) the LPJ-VSJA is more controlled by LAI data because the ratio of assimilated LAI (daily input) to SSM observations (3-day interval input) is approximately 3:1, which makes the likelihood function biased to LAI data; (2) the SM directly influences the simulation of ET, and the corresponding time function (computes the top layer SM (50 cm)) used here by Zhao et al. (2013)will result in the error of the updated top SM and propagating the error to the GPP_{SM}. In addition, the 8-day interval LAI has the capability to capture the temporal variability of phenology.

Current studies on terrestrial water and carbon flux assimilation mostly focus on the assimilation between a single model framework and observation results, lacking the fusion and comparison between multiple models. The processed models used in DA are simplifications and approximations of reality, and different models focus on different ecological processes. In this study, the updated ET module was integrated to compensate for the simplification of soil stratification and the lack of SM information in the hydrological module of the LPJ-DGVM. Therefore, the integration of multiple types of models and multi-source observation data (remotely sensed data, ecological inventory data (National Ecological Observatory Network, NEON (Keller et al. 2008)), and other measurements (Desai et al. 2011; Hayes et al. 2012) is expected to more objectively and effectively simulate the real state of ecosystems.

5.2 Comparison of joint assimilation (LPJ-VSJA) and other models for GPP and ET across regions and vegetation types

Global GPP and ET for different products were calculated by multiplying the global mean GPP density flux with the global vegetation area (122.4 million km^2) originated from the MODIS land cover product (Friedl et al. 2010). The mean global GPP of the LPJ-VSJA (130.2 Pg C yr^{-1}) was approximately 12% lower than that of PML-V2 (145.8 Pg C yr^{-1}) and 18% higher than that of GLASS and MODIS, respectively (Table S6). The GPP values of LPJ-VSJA and GOSIF were the most similar. The GOSIF GPP was developed from gridded SIF using simple linear relationships between SIF and GPP. Our global LPJ-VSJA GPP estimates were within the currently most plausible 110–150 Pg C/yr range.

As for ET, our results were similar to those of GLEAM ET and lower than those of PML-V2, GLDAS-2, and GLASS ET (~72000 km^2 yr^{-1}). Joint assimilation improved the overestimation of LPJ-DGVM ET. At the daily scale, the estimation accuracy of PML-V2 and GLDAS-2 products, calibrated with flux tower data, was better than that of our estimates, which suggests an underestimation of LPJ-

VSJA ET in wet regions. It is likely because the SSM of SMAP or SMOS was underestimated in the wet region or the influence of deep SM was under-represented. According to Seneviratne et al. (2010), satellite-based ET estimation approaches often overestimate ET in areas of arid and semi-arid climatic regimes in the magnitude of 0.50 to 3.00 mm d^{-1}. The poor performance of these models can largely be attributed to the lack of constraints of SSM or RZSM and more accurate vegetation parameters (Gokmen et al. 2012; Pardo et al. 2014). For instance, the monthly estimated ET modeled by the Penman-Monteith-Leuning (PML) model agreed with flux tower data well (R^2 = 0.77; BIAS = − 9.7%, approximately 0.2 mm d^{-1}). Our annual ET simulations were lower than other products and slightly underestimated tower ET with a BIAS of 0.19 mm d^{-1} (ET_{OBS}- ET_{joint}).

In general, GPP and ET had better assimilation performance in arid and semi-arid regions than in humid and sub-dry humid regions likely because of the following reasons. First, the incorporation of SSM is more important for vegetation growth in water-limited areas. The module PT-JPL_{SM} has been proven to have better performance in semi-arid and arid regions (Purdy et al. 2018). Our integrated model LPJ-PM also performed better in semi-arid and arid regions by assimilating SMAP soil moisture (Li et al. 2020). Second, the input performance, including SMOS and SMAP SSM products, is better in arid and temperate regions than in cold and humid regions (Zhang et al. 2019). Third, the vegetation types in humid regions are more complex and relatively less accurately simulated by the LPJ-DGVM within a single grid cell. For comparison, Zhang et al. (2020) used a data-driven upscaling approach to estimate GPP and ET in global semi-arid regions. This data-driven approach (R^2 = 0.79, RMSD = 1.13 g C m^{-2} d^{-1}) had slightly higher performance in estimating GPP than our LPJ-VSJA (R^2 =0.73 and RMSD= 1.14 g C m^{-2} d^{-1}) and the data-driven method (R^2 = 0.72 and RMSD = 0.72mm d^{-1}) had identical performance for estimating ET with our LPJ-VSJA(R^2 =0.73 and RMSD= 0.72 mm d^{-1}).

Our assimilation performance varied with PFT. The GPP and ET assimilation results of savanna sites performed well in both dry and wet regions, and those of shrubland sites showed the most remarkable improvement for simulations of LPJ-DGVM. The original simulation and assimilation performance of grassland sites in the semi-arid and arid regions were the best for all five PFTs. Consistent with our research, previous studies also showed better GPP or ET simulations for grassland, savannas, and shrublands biomes. For instance, Feng et al. (2015) validated five satellite-based ET algorithms for semi-arid ecosystems and concluded that all the models produced acceptable and relatively better results for most grassland, savanna, and shrubland sites. Yang et al. (2017) demonstrated that he GLEAM ET had a superior performance for the grassland sites. The GOSIF GPP demonstrated better simulation for grassland and woody savannas sites at 8-day time steps with higher R^2 (0.77 and 0.83, respectively) and lower RMSD (1.48 g C m^{-2} d^{-1} and 1.1 g C m^{-2} d^{-1}) (Li and Xiao 2019). In contrast, our LPJ-VSJA GPP showed an R^2 of 0.87 for grassland and 0.75 for savannas and an RMSD of 1.11 g C m^{-2} d^{-1} and 1.1 g C m^{-2} d^{-1}, respectively, in semi-arid and arid regions.

5.3 Uncertainty analysis of joint assimilation

Our validation results at both site and regional scales indicated that uncertainty existed in LPJ-VSJA daily GPP and ET estimates. The errors from the tower EC observations, model-driven data, model structure, error of satellite-based observations (e.g., LAI and SSM), and the spatial scale mismatch between the ground observed footprint size and satellite-derived footprint size were the vital factors affecting assimilation performance.

First, recent studies have revealed errors in the GLASS LAI and SMOS or SMAP SSM compared with ground measurements. By computing the RMSD and R^2 of each product, the GLASS LAI accuracy was clearly superior to that of MODIS and Four-Scale Geometric Optical Model based LAI (FSGOM) in forests and GLASS and FSGOM led to in much higher annual GPP and ET estimates compared to MCD15(Liu et al. 2018). The vegetation type (or land cover) misclassification caused 15–50% differences in LAI retrieval (Fang and Liang 2005; Gonsamo and Chen 2011). Yan et al. (2016) calculated a RMSD of 0.18 for the GLASS LAI over a range of HeiHe river basin sites and used the error to improve the simulation of LAI and fluxes by assimilating GLASS LAI data. Previous studies reported an improvement in the performance of the SMOS and SMAP products (Lievens et al. 2015; Miernecki et al. 2014), which both provide an accuracy of 0.04 m^3 m^{-3} (Zhang et al. 2019). However, the actual observation error of these two products typically depends on the spatial location and time of the year (RMSD varying between 0.035 and 0.056 m^3 m^{-3} for several retrieval configurations) (Brocca et al. 2012). According to Purdy et al. (2018), the ET simulated by PT-JPL_{SM} using the 9 km SM_L3_P_E data showed an inferior agreement (R^2= 0.47) but a relatively low RMSD (0.77 mm d^{-1}), due to the SMAP errors in the grid cell with soil heterogeneity and the climatological differences between model SM forecasts and SMAP SSM (Reichle and Koster 2004). We rescaled the ET_{PM} to the probability distribution of the ET_{LPJ} through a cumulative distribution function (CDF) to correct the potential seasonal biases of ET_{PM} before assimilation.

Second, there is large uncertainty in the influence of RZSM as the source of water available to plants (Albergel et al. 2008; Bonan et al. 2020). Our GPP results of irrigated sites were largely influenced by US-Ne1, an irrigate site. This site maintained high annual GPP in 2012 despite the drought (Figure S4). However, the SMOS SSM in 2012 had a lower SSM annual mean than the site observations likely because the detected soil layer (0-50 cm) of the site observation is deeper than that of the satellite retrieval and the cumulative deep soil moisture due to the regular irrigation was higher than the SSM that could easily be vaporized during the drought period (Figure S4). Therefore, the influence of deep SM of some cropland sites during the drought years induced large simulation errors and unsatisfactory assimilation performance. Moreover, some deep-rooted forests maintain a high LAI during drought by absorbing deep SM (>2 m) and groundwater (Zhang et al. 2016). Thus, joint assimilation of the LAI and SSM may eliminate a portion of the underestimation of GPP of such vegetation in drought periods. Therefore, further research is needed on how to optimally utilize satellite SM data for improving GPP and ET simulations.

Third, the problem of mixed pixels and mismatches in the observation footprints may also have an influence on the accuracy of estimated GPP and ET. The 5 km spatial resolution of the GLASS LAI ,9 km of SMAP, and 25 km of SMOS products cannot capture the sub-grid-scale condition, especially in grid cells for complex land surfaces or strong soil heterogeneity. To ensure the consistency of the grid-cell representativeness for the LAI and SSM, the interpolation result in errors that propagate through the modeling and assimilation, causing the accumulation of output errors (Nijssen and Lettenmaier 2004). Moreover, the shrubland in the LPJ-DGVM was most likely simulated as C4 grassland in the hydrothermal condition of semi-arid and arid regions. In contrast, the shrubland tended to be hybrid vegetation types (grassland mixed with other types of forest vegetation) in the hydrothermal condition of humid and sub-dry humid regions, and the simulated canopy height is closer to the real condition of shrubland. This might also be the reason for the superior performance of ET_{LPJ} and assimilation results of shrubland sites in humid and sub-dry humid regions.

When assimilating multiple data streams, all data streams could be in the same optimization (simultaneous assimilation) or use a sequential (step-by-step) approach. Mathematically, simultaneous optimization is optimal because strong parametric connections are maintained between different processes. However, complications may arise due to computational constraints related to the inversion of large matrices or the requirement of numerous simulations, particularly for global datasets (e.g. Peylin et al.,2016), and due to the "weight" of different data streams in the optimization (e.g. Wutzler and Carvalhais, 2014). This is particularly true when considering a regional-to-global-scale, multiple site optimization of a complex model that contains many parameters, and which typically takes on the order of minutes to an hour to run a one-year simulation. In practice, it is very difficult to define a probability distribution that properly characterizes the model structural uncertainty and observation errors accounting for biases and non-Gaussian distributions. Nevertheless, a step-wise assimilation may be useful in dealing with possible inconsistencies on a temporary basis, since parameter error covariance matrix must be propagated at each step. It's worth noting that the deviation between the model and observational data should be solved in the process of step-wise assimilation, such as the joint assimilation in this study, the satellite observations and model simulation were fitting through the CDF method so that the first step assimilation will strongly constrain the uncertainty of parameters related to phenology and carbon flux and propagate to the second step . Alternative solutions were found for water -related parameters through soil moisture, providing a better fit for all data streams. The sequence of assimilation is essential in the step-wise assimilation, and if the first observation contains a strong BIAS, then the associated error correlation will also propagate through the first assimilation. If the autocorrelation in the observation error, or the correlation between the data stream errors is not considered, it is likely that the posterior simulation has been overturned. That is, we overestimate the reduction in parametric uncertainty. If two observational data are less uncertainty (i.e., high precision of observation data), and the model of deviation is smaller (depend on the spatial scale and inversion method). Moreover, the correlation of these observations is stronger, and contain enough spatio-temporal information to limit all the parameters optimization accurately, the step-wise assimilation performance is basically the same as that of simultaneous assimilation.

6. Conclusions

We developed an assimilation system LPJ-VSJA that integrates GLASS LAI, SMOS SSM, and SMAP SSM data to improve GPP and ET estimates globally. The system was designed to assimilate two SSM products (SMOS and SMAP) into the integrated model - LPJ-PM for both dry and humid regions through separate and joint assimilation. The results show that the joint constraints provided by vegetation and soil variable strategies improve model simulations. Both the original and joint assimilation results for GPP and ET in semi-arid and arid regions performed better than those in humid and sub-dry humid regions, and the LPJ-PM that emphasized the SSM information is more suitable for the water-limited regions. For

ET assimilation, the different SSM products influence assimilation performance, and SMAP SSM possesses a slight advantage in most vegetation types and in both dry and humid regions. Our global LPJ-VSJA GPP and ET products have relatively higher accuracy than other products, especially in water-limited regions with lower ET values.

Data availability

The LPJ-DGVM v4.1 version code (LPJ-ML) and example configurations are public available via the project homepage (https://github.com/PIK-LPJmL/LPJmL). We used the 3.01 version of LPJ-DGVM, which removed the agricultural management module. The access of all the input and validation dataset of assimilation system have been described in article. The assimilation method code configurated by Fortran platform could be provided by contacting the X.T co-author. The modified code of LPJ-PM model and the underlying and global LPJ-VSJA GPP and ET data can be obtained by contacting the lead author of this manuscript.

Author contributions

S.L. and L.Z. designed the experiment and wrote the paper with support from all coauthors. S.L. and R.M. implemented the codes necessary for the experiments. J.X. contributed to the structure of the article and comparison of assimilation performance between the SMOS and SMAP experiments. X.T provided the POD-En4DVAR method and the code. M.Y contributed to the validation and analysis of the results. All the authors contributed to the synthesis of results and key conclusions.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Financial support

S.L., L.Z., R.M., and M.Y. were funded by the National Natural Science Foundation of China (Grant No. 41771392; PI Li Zhang) and (Grant No. 41901364; PI Min Yan).

References

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., & Martin, E. (2008). From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. *Hydrology and Earth System Sciences*, *12*, 1323-1337

Albergel, C., Calvet, J.-C., Mahfouf, J.-F., Rüdiger, C., Barbu, A. L., Lafont, S., Roujean, J.-L., Walker, J. P., Crapeau, M., and Wigneron, J.-P.: Monitoring of water and carbon fluxes using a land data assimilation system: a case study for southwestern France, Hydrol. Earth Syst. Sci., 14, 1109–1124, https://doi.org/10.5194/hess-14-1109-2010, 2010.

Albergel, C., Zheng, Y., Bonan, B., Dutra, E., Rodríguez-Fernández, N., Munier, S., Draper, C., de Rosnay, P., Muñoz-Sabater, J., Balsamo, G., Fairbairn, D., Meurey, C., and Calvet, J.-C.: Data assimilation for continuous global assessment of severe conditions over terrestrial surfaces, Hydrol. Earth Syst. Sci., 24, 4291–4316, https://doi.org/10.5194/hess-24-4291-2020, 2020.

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray‐Tortarolo, G., Papale, D., Parazoo, N.C., & Peylin, P. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, *53*, 785-818

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 Resour. Res. 50, 8420–8440. https://doi.org/10.1002/2013WR014573

Blyverket, J., Hamer, P.D., Bertino, L., Albergel, C., Fairbairn, D., & Lahoz, W.A. (2019). An Evaluation of the EnKF vs. EnOI and the Assimilation of SMAP, SMOS and ESA CCI Soil Moisture Data over the Contiguous US. *Remote Sensing, 11*, 478

Bonan, B., Albergel, C., Zheng, Y., Barbu, A.L., Fairbairn, D., Munier, S., & Calvet, J.-C. (2020). An ensemble square root filter for the joint assimilation of surface soil moisture and leaf area index within the Land Data Assimilation System LDAS-Monde: application over the Euro-Mediterranean region. *Hydrology and Earth System Sciences, 24*, 325-347

Bonan, G., Williams, M., Fisher, R., & Oleson, K. (2014). Modeling stomatal conductance in the earth system: linking leaf water-use efficiency and water transport along the soil–plant–atmosphere continuum. *Geoscientific Model Development, 7*, 2193-2222

Brocca, L., Tullo, T., Melone, F., Moramarco, T., & Morbidelli, R. (2012). Catchment scale soil moisture spatial–temporal variability. *Journal of hydrology, 422*, 63-75

Burgin, M.S., Colliander, A., Njoku, E.G., Chan, S.K., Cabot, F., Kerr, Y.H., Bindlish, R., Jackson, T.J., Entekhabi, D., & Yueh, S.H. (2017). A comparative study of the SMAP passive soil moisture product with existing satellite-based soil moisture products. *IEEE Transactions on Geoscience and Remote Sensing, 55*, 2959-2971

Caires, S., & Sterl, A. (2003). Validation of ocean wind and wave data using triple collocation. *Journal of geophysical research: oceans, 108*

Chan, S.K., Bindlish, R., O'Neill, P.E., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., & Piepmeier, J. (2016). Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and Remote Sensing, 54*, 4994-5007

Cui, C., Xu, J., Zeng, J., Chen, K.-S., Bai, X., Lu, H., Chen, Q., & Zhao, T. (2018). Soil moisture mapping from satellites: An intercomparison of SMAP, SMOS, FY3B, AMSR2, and ESA CCI over two dense network regions at different spatial scales. *Remote Sensing, 10*, 33

Desai, A.R., Moore, D.J., Ahue, W.K., Wilkes, P.T., De Wekker, S.F., Brooks, B.G., Campos, T.L., Stephens, B.B., Monson, R.K., & Burns, S.P. (2011). Seasonal pattern of regional carbon balance in the central Rocky Mountains from surface and airborne measurements. *Journal of Geophysical Research: Biogeosciences, 116*

Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E., & Wagner, W. (2011). Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. *Hydrology and Earth System Sciences, 15*, 3829-3841

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., & Johnson, J. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE, 98*, 704-716

Etheridge, D.M., Steele, L., Langenfelds, R.L., Francey, R.J., Barnola, J.M., & Morgan, V. (1996). Natural and anthropogenic changes in atmospheric CO2 over the last 1000 years from air in Antarctic ice and firn. *Journal of Geophysical Research: Atmospheres, 101*, 4115-4128

Evensen, G. (2004). Sampling strategies and square root analysis schemes for the EnKF. *Ocean dynamics, 54*, 539-560

Exbrayat, JF., Bloom, A.A., Carvalhais, N. et al. Understanding the Land Carbon Cycle with Space Data: Current Status and Prospects. Surv Geophys 40, 735–755 (2019). https://doi.org/10.1007/s10712-019-09506-2

Fang, H., Baret, F., Plummer, S., & Schaepman‐Strub, G. (2019). An overview of global leaf area index (LAI): Methods, products, validation, and applications. *Reviews of Geophysics, 57*, 739-799

Fang, H., Beaudoing, H.K., Rodell, M., Teng, W.L., & Vollmer, B.E. (2009). Global Land data assimilation system (GLDAS) products, services and application from NASA hydrology data and information services center (HDISC). In, *ASPRS 2009 Annual Conference, Baltimore, Maryland* (pp. 8-13)

Fang, H., & Liang, S. (2005). A hybrid inversion method for mapping leaf area index from MODIS data: Experiments and application to broadleaf and needleleaf canopies. *Remote Sensing of Environment*, *94*, 405-424

Feng, F., Chen, J., Li, X., Yao, Y., Liang, S., Liu, M., Zhang, N., Guo, Y., Yu, J., & Sun, M. (2015). Validity of five satellite-based latent heat flux algorithms for semi-arid ecosystems. *Remote Sensing, 7*, 16733-16755

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, *114*, 168-182

Gokmen, M., Vekerdy, Z., Verhoef, A., Verhoef, W., Batelaan, O., & Van der Tol, C. (2012). Integration of soil moisture in SEBS for improving evapotranspiration estimation under water stress conditions. *Remote Sensing of Environment, 121*, 261-274

Gonsamo, A., & Chen, J.M. (2011). Evaluation of the GLC2000 and NALC2005 land cover products for LAI retrieval over Canada. *Canadian Journal of Remote Sensing*, *37*, 302-313

Haxeltine, A., & Prentice, I.C. (1996). BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. *Global biogeochemical cycles, 10*, 693-709

Hayes, D.J., Turner, D.P., Stinson, G., McGuire, A.D., Wei, Y., West, T.O., Heath, L.S., De Jong, B., McConkey, B.G., & Birdsey, R.A. (2012). Reconciling estimates of the contemporary North American carbon balance among terrestrial biosphere models, atmospheric inversions, and a new approach for estimating net ecosystem exchange from inventory‐based data. *Global Change Biology*, *18*, 1282-1299

He, L., Chen, J.M., Liu, J., Bélair, S., & Luo, X. (2017). Assessment of SMAP soil moisture for global simulation of gross primary production. *Journal of Geophysical Research: Biogeosciences*, *122*, 1549-1563

He, Xinlei, Xu, T., Bateni, S.M., Ki, S.J., Xiao, J., Liu, S., Song, L., He, Xiangping, 2021. Estimation of Turbulent Heat Fluxes and Gross Primary Productivity by Assimilating Land Surface Temperature and Leaf Area Index. Water Res 57. https://doi.org/10.1029/2020WR028224

Huang, C., Li, Y., Gu, J., Lu, L., & Li, X. (2015). Improving estimation of evapotranspiration under water-limited conditions based on SEBS and MODIS data in arid regions. *Remote Sensing*, *7*, 16795-16814

Ines, A.V., 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

Jacquette, E., Al Bitar, A., Mialon, A., Kerr, Y., Quesney, A., Cabot, F., & Richaume, P. (2010). SMOS CATDS level 3 global products over land. In, *Remote Sensing for Agriculture, Ecosystems, and Hydrology XII* (p. 78240K): International Society for Optics and Photonics

Kaminski, T., Scholze, M., Vossbeck, M., Knorr, W., Buchwitz, M., & Reuter, M. (2017). Constraining a terrestrial biosphere model with remotely sensed atmospheric carbon dioxide. *Remote Sensing of Environment*, *203*, 109-124

Kato, T., Knorr, W., Scholze, M., Veenendaal, E., Kaminski, T., Kattge, J., & Gobron, N. (2013). Simultaneous assimilation of satellite and eddy covariance data for improving terrestrial water and carbon simulations at a semi-arid woodland site in Botswana. *Biogeosciences, 10*, 789-802

Keeling, C.D., Whorf, T.P., Wahlen, M., & Van der Plichtt, J. (1995). Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980. *Nature, 375*, 666-670

Keller, M., Schimel, D.S., Hargrove, W.W., & Hoffman, F.M. (2008). A continental strategy for the National Ecological Observatory Network. *Frontiers in Ecology and the Environment, 6*, 282-284

Kganyago, M., Mhangara, P., Alexandridis, T., Laneve, G., Ovakoglou, G., & Mashiyi, N. (2020). Validation of sentinel-2 leaf area index (LAI) product derived from SNAP toolbox and its comparison with global LAI products in an African semi-arid agricultural landscape. *Remote Sensing Letters, 11*, 883-892

Khan, M.S., Liaqat, U.W., Baik, J., & Choi, M. (2018). Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach. *Agricultural and Forest Meteorology, 252*, 256-268

Kim, H., Parinussa, R., Konings, A.G., Wagner, W., Cosh, M.H., Lakshmi, V., Zohaib, M., & Choi, M. (2018). Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products. *Remote Sensing of Environment, 204*, 260-275

Koster, R.D., Crow, W.T., Reichle, R.H., & Mahanama, S.P. (2018). Estimating basin‐scale water budgets with SMAP soil moisture data. *Water resources research, 54*, 4228-4244

Law, B., Falge, E., Gu, L.v., Baldocchi, D., Bakwin, P., Berbigier, P., Davis, K., Dolman, A., Falk, M., & Fuentes, J. (2002). Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation. *Agricultural and Forest Meteorology, 113*, 97-120

Lee, H., Seo, D.-J., & Koren, V. (2011). Assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states. *Advances in water resources, 34*, 1597-1615

Li, B., & Rodell, M. (2013). Spatial variability and its scale dependency of observed and modeled soil moisture over different climate regions. *Hydrology and Earth System Sciences, 17*, 1177-1188

Li C, Tang G, Hong Y. Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: Applicability of the Triple Collocation method across Mainland China[J]. Journal of Hydrology, 2018, 562: 71-83.

Li, S., Wang, G., Sun, S., Chen, H., Bai, P., Zhou, S., Huang, Y., Wang, J., & Deng, P. (2018). Assessment of multi-source evapotranspiration products over china using eddy covariance observations. *Remote Sensing, 10*, 1692

Li, S., Zhang, L., Ma, R., Yan, M., & Tian, X. (2020). Improved ET assimilation through incorporating SMAP soil moisture observations using a coupled process model: A study of US arid and semiarid regions. *Journal of hydrology, 590*, 125402

Li, X., Cheng, G., Liu, S., Xiao, Q., Ma, M., Jin, R., Che, T., Liu, Q., Wang, W., & Qi, Y. (2013). Heihe watershed allied telemetry experimental research (HiWATER): Scientific objectives and experimental design. *Bulletin of the American Meteorological Society, 94*, 1145-1160

Li, X., Mao, F., Du, H., Zhou, G., Xu, X., Han, N., Sun, S., Gao, G., & Chen, L. (2017). Assimilating leaf area index of three typical types of subtropical forest in China from MODIS time series data based on the integrated ensemble Kalman filter and PROSAIL model. *ISPRS Journal of Photogrammetry and Remote Sensing, 126*, 68-78

Li, X., & Xiao, J. (2019). A global, 0.05-degree product of solar-induced chlorophyll fluorescence derived from OCO-2, MODIS, and reanalysis data. *Remote Sensing, 11*, 517

Liang, S., Zhao, X., Liu, S., Yuan, W., Cheng, X., Xiao, Z., Zhang, X., Liu, Q., Cheng, J., & Tang, H. (2013). A long-term Global LAnd Surface Satellite (GLASS) data-set for environmental studies. *International Journal of Digital Earth, 6*, 5-33

Lievens, H., Tomer, S.K., Al Bitar, A., De Lannoy, G.J., Drusch, M., Dumedah, G., Franssen, H.-J.H., Kerr, Y.H., Martens, B., & Pan, M. (2015). SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sensing of Environment, 168*, 146-162

Ling, X.-L., Fu, C.-B., Yang, Z.-L., & Guo, W.-D. (2019). Comparison of different sequential assimilation algorithms for satellite-derived leaf area index using the Data Assimilation Research Testbed (version Lanai). *Geoscientific Model Development, 12*, 3119-3133

Liu, L., Gudmundsson, L., Hauser, M., Qin, D., Li, S., & Seneviratne, S.I. (2020). Soil moisture dominates dryness stress on ecosystem production globally. *Nature communications, 11*, 1-9

Liu, Y., Xiao, J., Ju, W., Zhu, G., Wu, X., Fan, W., Li, D., & Zhou, Y. (2018). Satellite-derived LAI products exhibit large discrepancies and can lead to substantial uncertainty in simulated carbon and water fluxes. *Remote Sensing of Environment, 206*, 174-188

Ma, H., Huang, J., Zhu, D., Liu, J., Su, W., Zhang, C., & Fan, J. (2013). Estimating regional winter wheat yield by assimilation of time series of HJ-1 CCD NDVI into WOFOST–ACRM model with Ensemble Kalman Filter. *Mathematical and Computer Modelling, 58*, 759-770

Ma, R., Zhang, L., Tian, X., Zhang, J., Yuan, W., Zheng, Y., Zhao, X., & Kato, T. (2017). Assimilation of remotely-sensed leaf area index into a dynamic vegetation model for gross primary productivity estimation. *Remote Sensing, 9*, 188

MacBean, N., Peylin, P., Chevallier, F., Scholze, M., & Schürmann, G. (2016). Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. *Geoscientific Model Development, 9*, 3569-3588

Martens, B., Miralles, D.G., Lievens, H., Schalie, R.v.d., De Jeu, R.A., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., & Verhoest, N.E. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development, 10*, 1903-1925

Miernecki, M., Wigneron, J.-P., Lopez-Baeza, E., Kerr, Y., De Jeu, R., De Lannoy, G.J., Jackson, T.J., O'Neill, P.E., Schwank, M., & Moran, R.F. (2014). Comparison of SMOS and SMAP soil moisture retrieval approaches using tower-based radiometer data over a vineyard field. *Remote Sensing of Environment, 154*, 89-101

Miralles, D.G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M., Hirschi, M., Martens, B., Dolman, A.J., & Fisher, J.B. (2016). The WACMOS-ET project–Part 2: Evaluation of global terrestrial evaporation data sets. *Hydrology and Earth System Sciences, 20*, 823-842

Mitchell, H.L., Houtekamer, P.L., & Pellerin, G. (2002). Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Monthly weather review, 130*, 2791-2808

Mu, Q., Zhao, M., Heinsch, F.A., Liu, M., Tian, H., & Running, S.W. (2007). Evaluating water stress controls on primary production in biogeochemical and remote sensing based models. *Journal of Geophysical Research: Biogeosciences, 112*

New, M., Hulme, M., & Jones, P. (2000). Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *Journal of climate, 13*, 2217-2238

Nijssen, B., & Lettenmaier, D.P. (2004). Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. *Journal of Geophysical Research: Atmospheres, 109*

O'Neill, P., Entekhabi, D., Njoku, E., & Kellogg, K. (2010). The NASA soil moisture active passive (SMAP) mission: Overview. In, *2010 IEEE International Geoscience and Remote Sensing Symposium* (pp. 3236-3239): IEEE

O'Carroll, A.G., Eyre, J.R., & Saunders, R.W. (2008). Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *Journal of atmospheric and oceanic technology, 25*, 1197-1207

Pan, H.; Chen, Z.; de Wit, A.; Ren, J. Joint Assimilation of Leaf Area Index and Soil Moisture from Sentinel-1 and Sentinel-2 Data into the WOFOST Model for Winter Wheat Yield Estimation. Sensors 2019, 19, 3161.

Pardo, N., Sánchez, M.L., Timmermans, J., Su, Z., Pérez, I.A., & García, M.A. (2014). SEBS validation in a Spanish rotating crop. *Agricultural and Forest Meteorology, 195*, 132-142

Petropoulos, G.P., Ireland, G., & Barrett, B. (2015). Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth, Parts A/B/C, 83*, 36-56

Pipunic, R., Walker, J., & Western, A. (2008). Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study. *Remote Sensing of Environment, 112*, 1295-1305

Purdy, A.J., Fisher, J.B., Goulden, M.L., Colliander, A., Halverson, G., Tu, K., & Famiglietti, J.S. (2018). SMAP soil moisture improves global evapotranspiration. *Remote Sensing of Environment, 219*, 1-14

Rahman, A., Zhang, X., Houser, P., Sauer, T., Maggioni, V., 2022. Global Assimilation of Remotely Sensed Leaf Area Index: The Impact of Updating More State Variables Within a Land Surface Model. Front. Water 3, 789352. https://doi.org/10.3389/frwa.2021.789352

Rahman, A.; Maggioni, V.; Zhang, X.; Houser, P.; Sauer, T.; Mocko, D.M. The Joint Assimilation of Remotely Sensed Leaf Area Index and Surface Soil Moisture into a Land Surface Model. Remote Sens. 2022, 14, 437. https://doi.org/10.3390/rs14030437

Rüdiger, C., Albergel, C., Mahfouf, J.F., Calvet, J.C., & Walker, J.P. (2010). Evaluation of the observation operator Jacobian for leaf area index data assimilation with an extended Kalman filter. *Journal of Geophysical Research: Atmospheres, 115*Reichle, R.H., De Lannoy, G.J., Liu, Q., Koster, R.D., Kimball, J.S., Crow, W.T., Ardizzone, J.V., Chakraborty, P., Collins, D.W., & Conaty, A.L. (2017). Global assessment of the SMAP level-4 surface and root-zone soil moisture product using assimilation diagnostics. *Journal of Hydrometeorology, 18*, 3217-3237

Reichle, R.H., & Koster, R.D. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters, 31*

Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., & Kim, G.-K. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of climate, 24*, 3624-3648

Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *Bioscience, 54*, 547-560

Scholze, M., Buchwitz, M., Dorigo, W., Guanter, L., and Quegan, S.: Reviews and syntheses: Systematic Earth observations for use in terrestrial carbon cycle data assimilation systems, Biogeosciences, 14, 3401–3429, https://doi.org/10.5194/bg-14-3401-2017, 2017.

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., & Teuling, A.J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews, 99*, 125-161

Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J.O., Levis, S., Lucht, W., & Sykes, M.T. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology, 9*, 161-185

Stoffelen, A. (1998). Toward the true near‐surface wind speed: Error modeling and calibration using triple collocation. *Journal of geophysical research: oceans, 103*, 7755-7766

Sun, P., Wu, Y., Xiao, J., Hui, J., Hu, J., Zhao, F., Qiu, L., & Liu, S. (2019). Remote sensing and modeling fusion for investigating the ecosystem water-carbon coupling processes. *Science of the total environment, 697*, 134064

Taylor, K.E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres, 106*, 7183-7192

Tian, S., Renzullo, L.J., Van Dijk, A.I., Tregoning, P., & Walker, J.P. (2019). Global joint assimilation of GRACE and SMOS for improved estimation of root-zone soil moisture and vegetation response. *Hydrology and Earth System Sciences, 23*, 1067-1081

Tian, X., & Feng, X. (2015). A non-linear least squares enhanced POD-4DVar algorithm for data assimilation. *Tellus A: Dynamic Meteorology and Oceanography, 67*, 25340

Tian, X., Xie, Z., Dai, A., Jia, B., & Shi, C. (2010). A microwave land data assimilation system: Scheme and preliminary evaluation over China. *Journal of Geophysical Research: Atmospheres, 115*

Tian, X., Xie, Z., Dai, A., Shi, C., Jia, B., Chen, F., & Yang, K. (2009). A dual‐pass variational data assimilation framework for estimating soil moisture profiles from AMSR‐E microwave brightness temperature. *Journal of Geophysical Research: Atmospheres, 114*

Tian, X., Xie, Z., Liu, Y., Cai, Z., Fu, Y., Zhang, H., & Feng, L. (2014). A joint data assimilation system (Tan-Tracker) to simultaneously estimate surface CO_2 fluxes and 3-D atmospheric CO_2 concentrations from observations. *Atmospheric Chemistry and Physics, 14*, 13281-13293

Tian, X., Xie, Z., & Sun, Q. (2011). A POD-based ensemble four-dimensional variational assimilation method. *Tellus A: Dynamic Meteorology and Oceanography, 63*, 805-816

Twine, T.E., Kustas, W., Norman, J., Cook, D., Houser, P., Meyers, T., Prueger, J., Starks, P., & Wesely, M. (2000). Correcting eddy-covariance flux underestimates over a grassland. *Agricultural and Forest Meteorology, 103*, 279-300

Wang, L., Zhu, H., Lin, A., Zou, L., Qin, W., & Du, Q. (2017). Evaluation of the latest MODIS GPP products across multiple biomes using global eddy covariance flux data. *Remote Sensing, 9*, 418

Waring, R.H., & Running, S.W. (2010). *Forest ecosystems: analysis at multiple scales*. Elsevier

Wieder, W., Boehnert, J., Bonan, G., & Langseth, M. (2014). Regridded harmonized world soil database v1. 2. *ORNL DAAC*

Wu, M.; Scholze, M.; Voßbeck, M.; Kaminski, T.; Hoffmann, G. Simultaneous Assimilation of Remotely Sensed Soil Moisture and FAPAR for Improving Terrestrial Carbon Fluxes at Multiple Sites Using CCDAS. Remote Sens. 2019, 11, 27. https://doi.org/10.3390/rs11010027

Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J.A., Huete, A.R., Ichii, K., Ni, W., Pang, Y., & Rahman, A.F. (2019). Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment, 233*, 111383

Xiao, Z., Liang, S., & Jiang, B. (2017). Evaluation of four long time-series global leaf area index products. *Agricultural and Forest Meteorology, 246*, 218-230

Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., & Song, J. (2013). Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing, 52*, 209-223

Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., & Song, J. (2016). Long-time-series global land surface satellite leaf area index product derived from MODIS and AVHRR surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing, 54*, 5301-5318

Xie, Y.; Wang, P.; Sun, H.; Zhang, S.; Li, L. Assimilation of Leaf Area Index and Surface Soil Moisture With the CERES-Wheat Model for Winter Wheat Yield Estimation Using a Particle Filter Algorithm. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2017, 10, 1303–1316.

Xu, T., He, X., Bateni, S.M., Auligne, T., Liu, S., Xu, Z., Zhou, J., Mao, K., 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., Chen, F., He, Xinlei, Barlage, M., Zhang, Z., Liu, S., He, Xiangping, 2021. Improve the Performance of the Noah - MP - Crop Model by Jointly Assimilating Soil Moisture and Vegetation Phenology Data. J Adv Model Earth Syst 13. https://doi.org/10.1029/2020MS002394

Yan, M., Tian, X., Li, Z., Chen, E., Wang, X., Han, Z., & Sun, H. (2016). Simulation of forest carbon fluxes using model incorporation and data assimilation. *Remote Sensing, 8,* 567

Yang, W., Wang, Y., Liu, X., Zhao, H., Shao, R., & Wang, G. (2020). Evaluation of the rescaled complementary principle in the estimation of evaporation on the Tibetan Plateau. *Science of the total environment, 699*, 134367

Yang, X., Yong, B., Ren, L., Zhang, Y., & Long, D. (2017). Multi-scale validation of GLEAM evapotranspiration products over China via ChinaFLUX ET measurements. *International Journal of Remote Sensing, 38,* 5688-5709

Yilmaz, M.T., & Crow, W.T. (2014). Evaluation of assumptions in soil moisture triple collocation analysis. *Journal of Hydrometeorology, 15*, 1293-1302

Yuan, W., Liu, S., Yu, G., Bonnefond, J.-M., Chen, J., Davis, K., Desai, A.R., Goldstein, A.H., Gianelle, D., & Rossi, F. (2010). Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. *Remote Sensing of Environment, 114*, 1416-1431

Zhang, D.-H., Li, X.-R., Zhang, F., Zhang, Z.-S., & Chen, Y.-L. (2016). Effects of rainfall intensity and intermittency on woody vegetation cover and deep soil moisture in dryland ecosystems. *Journal of hydrology, 543*, 270-282

Zhang, F., & Weng, Y. (2015). Predicting hurricane intensity and associated hazards: A five-year real-time forecast experiment with assimilation of airborne Doppler radar observations. *Bulletin of the American Meteorological Society, 96*, 25-33

Zhang, L., Xiao, J., Zheng, Y., Li, S., & Zhou, Y. (2020). Increased carbon uptake and water use efficiency in global semi-arid ecosystems. *Environmental Research Letters, 15*, 034022

Zhang, X., Huang, X.-Y., Liu, J., Poterjoy, J., Weng, Y., Zhang, F., & Wang, H. (2014). Development of an efficient regional four-dimensional variational data assimilation system for WRF. Journal of atmospheric and oceanic technology, 31, 2777-2794

Zhang, R., Kim, S., & Sharma, A. (2019). A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes. *Remote Sensing of Environment, 223*, 82-94

Zhao, L., Xia, J., Xu, C.-y., Wang, Z., Sobkowiak, L., & Long, C. (2013). Evapotranspiration estimation methods in hydrological models. *Journal of Geographical Sciences, 23*, 359-369

Zobitz, J., Moore, D.J., Quaife, T., Braswell, B.H., Bergeson, A., Anthony, J.A., & Monson, R.K. (2014). Joint data assimilation of satellite reflectance and net ecosystem exchange data constrains ecosystem carbon fluxes at a high-elevation subalpine forest. *Agricultural and Forest Meteorology*, *195*, 73-88

Zou, L., Zhan, C., Xia, J., Wang, T., & Gippel, C.J. (2017). Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman filter. *Journal of hydrology*, *549*, 685-702