



## A better carbon-water flux simulation in multiple vegetation types by data assimilation



Qiuyu Liu<sup>a,b</sup>, Tinglong Zhang<sup>a,\*</sup>, Mingxi Du<sup>c</sup>, Huanlin Gao<sup>a</sup>, Qingfeng Zhang<sup>a</sup>, Rui Sun<sup>d</sup>

<sup>a</sup> College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi, 712100, China

<sup>b</sup> Institute of Environment Sciences, Department of Biology Sciences, University of Quebec at Montreal, Montreal, H3C 3P8, Canada

<sup>c</sup> Department of Geography, McGill University, Montreal, Quebec, H3A 0B9, Canada

<sup>d</sup> Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China

### ARTICLE INFO

#### Keywords:

Biome-BGC model  
Leaf area index  
Evapotranspiration  
Net ecosystem CO<sub>2</sub> exchange  
Ensemble Kalman filter algorithm  
Unscented Kalman filter

### ABSTRACT

**Background:** The accurate estimation of carbon-water flux is critical for understanding the carbon and water cycles of terrestrial ecosystems and further mitigating climate change. Model simulations and observations have been widely used to research water and carbon cycles of terrestrial ecosystems. Given the advantages and limitations of each method, combining simulations and observations through a data assimilation technique has been proven to be highly promising for improving carbon-water flux simulation. However, to the best of our knowledge, few studies have accomplished both parameter optimization and the updating of model state variables through data assimilation for carbon-water flux simulation in multiple vegetation types. And little is known about the variation of the performance of data assimilation for carbon-water flux simulation in different vegetation types.

**Methods:** In this study, we assimilated leaf area index (LAI) time-series observations into a biogeochemical model (Biome-BGC) using different assimilation algorithms (ensemble Kalman filter algorithm (EnKF) and unscented Kalman filter (UKF)) in different vegetation types (deciduous broad-leaved forest (DBF), evergreen broad-leaved forest (EBF) and grassland (GL)) to simulate carbon-water flux.

**Results:** The validation of the results against the eddy covariance measurements indicated that, overall, compared with the original simulation, assimilating the LAI into the Biome-BGC model improved the carbon-water flux simulations ( $R^2$  increased by 35%, root mean square error decreased by 10%; the sum of the absolute error decreased by 8%) but more significantly, improved the water flux simulations ( $R^2$  increased by 31%, root mean square error decreased by 18%; the sum of the absolute error decreased by 16%). Among the different forest types, the data assimilation techniques (both EnKF and UKF) achieved the best performance towards carbon-water flux in EBF ( $R^2$  increased by 44%, root mean square error decreased by 24%; the sum of the absolute error decreased by 28%), and the performances of EnKF and UKF showed slightly different when simulating carbon fluxes.

**Conclusion:** We suggest that to reduce the uncertainty in global carbon-water flux quantification, forthcoming data assimilation treatment should consider the vegetation types where the data assimilation experiments are carried out, the simulated objectives and the assimilation algorithms.

### 1. Background

Accurate estimation of carbon-water flux is critical for understanding the feedback between terrestrial ecosystems and the atmosphere, as well as mitigation and adaptation to climate change (Stocker et al., 2013; Sándor et al., 2016). Carbon fixed by vegetation through photosynthesis and water evaporated and transpired from ecosystems are two key components of the terrestrial carbon and water cycles, respectively (Liu

et al., 2018). The accurate estimation of net ecosystem exchanges (NEE) and evapotranspiration (ET) is essential for assessing terrestrial carbon and water balances as well as energy exchange.

Observation and model simulation are two complementary tools for estimating terrestrial carbon (i.e., NEE) and water flux (i.e., ET) (Ito and Inatomi, 2012; Zhang et al., 2016). Modeling provides continuous spatial and temporal (both past and future) simulations. However, a model is just an abstraction and simplification of reality; the optimal values of model parameters are difficult to determine, and uncertainties

\* Corresponding author.

E-mail address: [dargon810614@126.com](mailto:dargon810614@126.com) (T. Zhang).

Abbreviations			
LAI	leaf area index	DHS	Dinghushan flux site
MODIS	Moderate Resolution Imaging Spectroradiometer	QYZ	Qianyanzhou flux site
KF	Kalman filter	DX	Dangxiong flux site
EnKF	ensemble Kalman filter	HB	Haibei flux site
UKF	unscented Kalman filter	RMSE	root mean square error
SEnKF	smoothed ensemble Kalman filter	SAE	the sum of the absolute error
DBF	deciduous broad-leaved forest	OS	original model without data assimilation treatment
EBF	evergreen broad-leaved forest	IS	the improved and optimized model without data assimilation treatment
GL	grassland	IEnKF	the improved and optimized model with data assimilation using the EnKF algorithm treatment
NEE	net ecosystem exchange	IUKF	the improved and optimized model with data assimilation using the UKF algorithm treatment
ET	evapotranspiration	NPP	net primary productivity
HF	Harvard Forest Environmental Monitoring flux Site		
CBS	Changbai Mountain flux site		

inevitably occur in the simulation process, especially when a model is applied at a large scale (Renard et al., 2010; Dong et al., 2016). As a notable example, in the North American Carbon Program Multi-Scale Synthesis work involving 26 models (both process-based and empirical models) and data from 39 flux tower sites, none of the associated models was able to predict gross primary productivity estimates within the observed range of uncertainties (Schaefer et al., 2012). In addition, despite the progress that has been made in remote sensing models (Mu et al., 2011; Wu et al., 2011), recent evidence has indicated that remotely sensed observation products (i.e., MOD16 and 17) may underestimate carbon fluxes in grasslands (Zhu et al., 2018a) and overestimate water fluxes in shrublands (Li et al., 2021); these results suggest that the uncertainties associated with these products vary among different vegetation types and climate conditions, especially in arid regions with limited water (Sims et al., 2008). Observations can provide high-quality and high-precision data; however, in some places and certain periods, observations are usually of poor quality or even missing. In addition, the regional-scale acquisition of field observation data is time-consuming and laborious (Kvamme, 1989). High-quality and high-precision observations are highly desirable. Such disadvantages of observations and model simulations could generate uncertainties when using either method alone for carbon-water flux estimation, especially at the regional scale. Given the rapid development of global observation networks such as FLUXNET and multi-sensor remote sensing observation data and the long-term accumulation of data through research network projects, the ecological research field has become a data-rich enterprise (Peng et al., 2011). It is therefore valuable and pressing to integrate models with satellite observations simultaneously to facilitate more effective studies of carbon and water cycles in terrestrial ecosystems (Peng et al., 2011; Zhang et al., 2016).

Data assimilation is the process of incorporating observations over a period of time into a prediction model to estimate the state of a system, and it has led to dramatic improvements in estimating and forecasting techniques (Rawlins et al., 2007; Luo et al., 2011; Peng et al., 2011; Dong et al., 2016). Modern data assimilation methods combine observations with models by updating model state variables and optimizing parameters, and/or selecting alternative model structures using optimization techniques (Luo et al., 2011). One of the most important state variables in terrestrial carbon and/or water flux models is LAI (Asaadi et al., 2018; Gilardelli et al., 2019; Hanes et al., 2019). LAI, defined as half the total green leaf area per unit ground area, can serve as an indicator in the exchange of energy and mass between the canopy and the atmosphere (Weiss et al., 2004) and is a critical parameter in many land surface models. The plant canopy is a locus of physical and biogeochemical processes in terrestrial ecosystems. Leaves are the point of exchange between plants and atmospheric CO<sub>2</sub> as well as water vapor. An increase

in LAI potentially enhances carbon uptake, albeit at the cost of a greater demand for water (Norby et al., 2003). Many previous studies demonstrated that the estimation of carbon-water flux could be improved by the better estimation of the LAI state variable of the model (van den Hurk et al., 2003; Albergel et al., 2010; Zhang et al., 2016; Post et al., 2018). For example, Vazifedoust et al. (2009) indicated that the better updating of the LAI state variable led to better water flux estimations and forecasts at the regional scale. Albergel et al. (2010) tested a land data assimilation system and showed that jointly assimilating the LAI positively impacted the carbon-water flux. An LAI assimilation framework was tested by Reville et al. (2013), and the results showed that compared with the simulations conducted without data assimilation, the cumulative NEE estimations were significantly improved, and the authors suggested that future developments should include assimilations of model state variables. Rüdiger et al. (2010) also found that the estimation skill of the employed land surface model in quantifying carbon-water fluxes between the vegetation and atmosphere could be greatly increased by assimilating the LAI. An accurate and efficient optimization of the LAI is therefore of key importance to use models in carbon-water flux simulations (van den Hurk et al., 2003; Li et al., 2017).

To date, optimization techniques can be classified into batch and sequential methods. Batch methods, such as the variational or adjoint method (Vukicevic et al., 2001), Levenburg-Marquardt method (Luo et al., 2003) and Markov chain Monte Carlo (MCMC) method (Metropolis and Ulam, 1949), assimilate all the data within a time interval at once and treat the cost function as a single function to be minimized over that window (Luo et al., 2011). One of the most widely used sequential methods is the Kalman filter (KF) method, which is a recursive data assimilation algorithm for estimating initial parameters and state variables of a system at each time using a state-space model from a series of heterogeneous observations (Kalman, 1960). The extended Kalman filter (EKF) is a popular method that is applied in the field of geomechanics for nonlinear state estimation. However, several alternative methods have emerged over the past several decades, namely, the ensemble Kalman filter (EnKF) and the unscented Kalman filter (UKF) (Hommels et al., 2009).

To date, the KF, as an advantageous technique for combining multi-source data, has been widely used in the estimation of carbon-water flux in terrestrial ecosystems and has become a frontier of this field (Dong et al., 2016; Yan et al., 2019; Wu et al., 2020). For example, Williams et al. (2005) used EnKF algorithms to link observations to carbon models and noted that the assimilation process reduced the uncertainty of using data or models alone, and NEE forecasts were statistically unbiased estimates in the EBF. Based on combining the EnKF with the kernel smoothing technique, a smoothed ensemble Kalman filter (SEnKF) was developed by Chen et al. (2008), and they demonstrated that the SEnKF could effectively reduce the error of the state variables, thus improving

the estimation accuracy of the carbon-water flux. Mo et al. (2008) designed sequential data assimilation with an EnKF to optimize the key parameters of the productivity model, showing a significant improvement in the simulation of ecosystem productivity. Based on the EnKF algorithm, Pipunic et al. (2008) assimilated remote sensing data into the CSIRO Biosphere Model (CBM) to improve the accuracy of the model's prediction of water and heat fluxes. Remotely sensed LAI data were assimilated into a biogeochemical model by Zhang et al. (2016) using the KF algorithms in EBF and DBF. After their treatments, the accuracy of carbon-water flux simulations was effectively improved. Zhu et al. (2017) studied the effects of assimilating multiscale soil moisture into hydrological models using the EnKF and demonstrated that coarse-scale soil moisture observations could also help identify the parameters and states of flow models. Yan et al. (2019) assimilated observed soil variables, including three-layer temperature and moisture data, into the biogeochemical model (Biome-BGC) by using the EnKF algorithm. Validation with eddy covariance flux measurements proved that the simulated ecosystem respiration, NEE and ET were improved.

These previous studies have successfully assimilated remotely sensed observations, such as vegetation indices, reflectance data and fraction of photosynthetically active radiation (FPAR) values, and ground measurements into the model for state variable updating and for executing and evaluating the data assimilation performance in specific forest ecosystems. Despite these advances, updating state variables through the use of data assimilation should be more tested and explored. Because most of the previous studies that aim to improve model performance mainly focus on optimizing static model parameters and structure. Compared with the static parameters and model structure, the value of model state variables usually changes as changes occur in the model inputs and as

forward simulations progress. Therefore, in order to update state variables, a robust dynamic optimization approach that can connect the model state variables with the observation is essential, but it is also difficult to achieve. In addition, selecting the appropriate model state variables to be updated and obtaining high-quality observations from simulated locations are also challenges. To reduce the uncertainty of model simulation, both optimization of state variables, model structure and state variable should be considered. However, to the best of our knowledge, few studies have accomplished both the optimization of parameters and the updating of the model state variables (i.e., LAI state variable) through data assimilation when performing carbon-water flux estimations (Gove and Hollinger, 2006; Zhang et al., 2016). And very few have focused on the comparison between different assimilation algorithms and their performance in different vegetation types to provide the research community with detailed insight for the use of data assimilation in carbon-water flux estimation.

Here, we present a strategy of data assimilation for simulating carbon-water flux using an improved process-based biogeochemical model and the Moderate Resolution Imaging Spectroradiometer (MODIS) LAI product (MOD15A2) to mitigate uncertainties involved in the model simulation. Specifically, the EnKF and the UKF algorithms were used to assimilate remotely sensed LAI data into both the original and the improved Biome-BGC models at DBF, EBF and GL study sites (Zhang et al., 2011a, b; see Fig. 1 and Table 1). The objectives of this study were to evaluate the performance of data assimilation algorithms (EnKF and UKF) in different vegetation types to provide new insights into the accurate estimation of carbon-water flux estimation by using data assimilation.

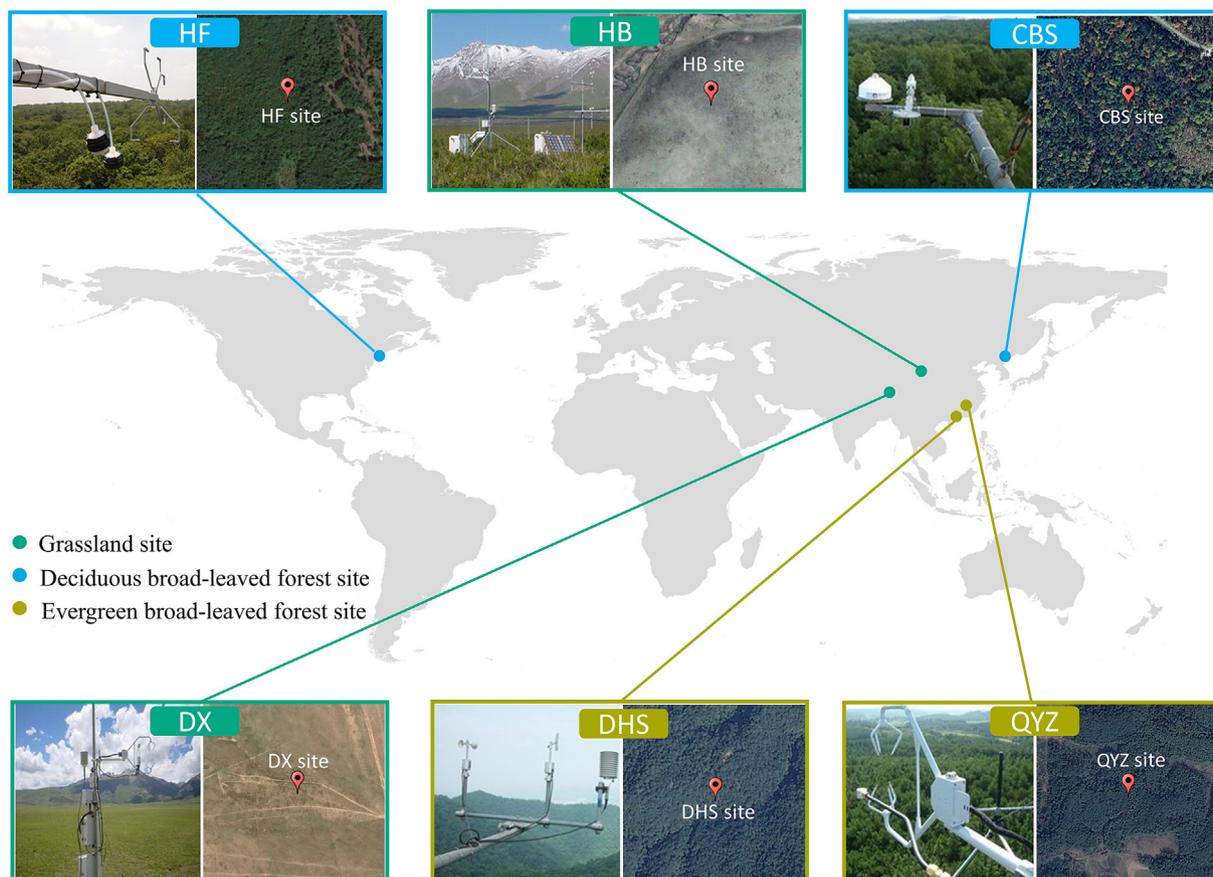


Fig. 1. Spatial distribution of six flux sites included in this study. Google Earth images represent 500 m by 500 m spatial coverage around each site. Landscape picture of each site comes from ChinaFLUX (<http://www.chinaflux.org/>) and AMERIFLUX (<https://ameriflux.lbl.gov/>).

**Table 1**  
Description of flux sites included in this study.

Sites	Lat. (°N)	Lon. (°E)	Alt. (m)	Plant functional type	Precipitation (mm)	Annual average temperature (°C)
HF	42.54	-72.17	340	DBF	1120	3.5
CBS	42.41	128.10	738	DBF	710	3.6
DHS	23.17	112.53	300	EBF	1960	20.9
QYZ	26.74	115.06	102	EBF	1540	17.9
DX	30.49	91.06	4250	GL	450	1.3
HB	37.62	101.32	3190	GL	540	-1.2

## 2. Methods

### 2.1. Study area

We evaluated the performance of the Biome-BGC model with the framework of data assimilation at 6 flux sites. As shown in Fig. 1 and Table 1, two DBF sites (Harvard Forest Environmental Monitoring flux Site, HF; Changbai Mountain flux site, CBS) (Zhang et al., 2006; Urbanski et al., 2007), two EBF (Dinghushan flux site, DHS; Qianyanzhou flux site, QYZ) (Wen et al., 2006; Zhang et al., 2006), and two GL sites (Dangxiong flux site, DX; Haibei flux site, HB) (Fu et al., 2006; Yu et al., 2006) were selected based on the availability of observed flux and micrometeorological records in this study.

### 2.2. Data

We collected eddy covariance measurements including NEE and ET recorded from 2003 to 2005 at the DHS, QYZ, CBS, HB and HF sites and two years of records (2004–2005) at the DX site. We also collected meteorology, latitude, topography, soil, vegetation, and other background data for these 6 flux sites for modeling input or initialization. Remotely sensed MODIS LAI (MOD15A2) data were collected from the DHS, QYZ, CBS, HB and HF sites for 3 years and 2 years at the DX site to assimilate into the model for carbon and water flux estimation. To alleviate the modeling uncertainties caused by the noise of LAI data, quality control treatments were conducted by using a quality control file.

### 2.3. Biome-BGC model

The Biome-BGC model (Running and Hunt, 1993), referred to as the original model in this study, is a biogeochemical process-based model that was developed from the FOREST-BGC model (Running and Gower, 1991). This model can simulate ecosystem carbon, nitrogen and water storage and flux in different ecosystems. The ecosystem gross primary production is calculated using the Farquhar photosynthesis routine separately for illuminated and shaded foliage (Farquhar et al., 1980). Autotrophic respiration is divided into growth respiration as a function of carbon, allocated to different plant compartments, and maintenance respiration, which is calculated proportional to the nitrogen content of the living tissue and adjusted for temperature (Farquhar et al., 1980). More detailed descriptions of the Biome-BGC model have been presented by Running and Hunt (1993). The Biome-BGC model is driven by three types of parameters: (1) the initialization information for the simulation site, including latitude, longitude, altitude, soil depth, soil particle composition, interannual atmospheric CO<sub>2</sub> concentration change, and vegetation type; (2) daily meteorological data, including maximum temperature, minimum temperature, average temperature, daily precipitation, daily vapor pressure deficit, and daily solar radiation; and (3) 44 physiological and ecological parameters.

#### 2.3.1. Improvement of the Biome-BGC model

The improved Biome-BGC model used in this study was optimized and improved by Zhang et al. (2011a, b). Specifically, based on observed

flux data and a simulated annealing algorithm, the major physiological and ecological parameters of the Biome-BGC model were optimized. More details about the optimization process are available in the literature (Zhang et al., 2011b). Due to the inadequate soil water balance module involved in the Biome-BGC model, the simulation processes of stomatal conductance under soil water stress, ET and soil water loss processes have been improved. More details about these improvements were presented by Zhang et al. (2011a). The value of major parameters used in Biome-BGC model in this study is presented in Table S1 (Supplementary Materials).

### 2.4. Data assimilation algorithm

#### 2.4.1. Unscented Kalman filter

The UKF algorithm is an extension of the traditional KF. The UKF is a recursive Bayesian estimation method based on unscented transformation (UT) and approximates the posterior probability density with a set of determined sample points to recurse and update the state and error covariance of the nonlinear model (Julier and Uhlmann, 1996; Wan and van Der Merwe, 2000). The UKF uses deterministic sampling to approximate the state distribution as Gaussian random variables. The sigma point was chosen and propagated through nonlinear systems to capture the true mean and covariance of the state distribution. These points are substituted into the nonlinear function to obtain the corresponding point set containing the nonlinear function values. Then, based on the observations, the positions of the sample points are adjusted to construct the sample. The mean and variance of the sample approximate the mean and variance, respectively, of the actual distribution with quadratic precision. The posterior mean and covariance were then calculated from the propagated sigma points. These calculations can be summarized in five steps: (1) giving the original value of the basic state variable; (2) extending the state algorithm to reduce noise; (3) calculating the sigma sampling point; (4) performing the backward propagation of sigma points; and (5) updating the observation information and calculating the predicted value and its variance. When the system obtains the new observed value, the algorithm enters a new filtering process. More detailed descriptions can be found in Text S1 (Supplementary Materials) and the literature by Wan et al. (2001).

#### 2.4.2. Ensemble Kalman filter

The EnKF is a sequential data assimilation algorithm that was proposed by Evensen (1994). To date, the EnKF, used mainly to forecast the error covariance of a model, is based on the Monte Carlo method (Evensen, 2003). The EnKF can sequentially assimilate multisource observations on the basis of the assumptions that system and measurement noises are both based on white and Gaussian distributions (Houtekamer and Mitchell, 2005; Yan et al., 2019). The key point of the EnKF is to generate an observation set at each update time by introducing noise. The mean of the noise distribution is zero, and the covariance is equal to the observation error covariance matrix. Otherwise, the updated ensemble will have a very low covariance. More detailed descriptions of the EnKF are presented by Text S2 (Supplementary Materials) and the literature by Evensen (1994) and Burgers et al. (1998).

### 2.5. Connection of assimilation algorithm with Biome-BGC model

The Biome-BGC model, as a process-based model, was developed based on the physiological mechanism. The processes involved in this model are closely linked and interact with each other; for instance, changes in photosynthesis and respiration affect the subsequent carbon allocation process. Therefore, the benefits from state variable updates in one process can be propagated throughout the simulation. The LAI is among the most critical state variables in the Biome-BGC model and is associated with model simulations of important physiological and ecological processes (e.g., photosynthesis, respiration, and evapotranspiration). For example, in the Biome-BGC model, the LAI is an important

variable considered in the calculations of net primary productivity (NPP) and canopy transpiration; this variable can thus significantly affect the carbon-water flux simulations (Running and Hunt, 1993). In our assimilation scheme (Fig. 2), the satellite observed LAI data were assimilated into the Biome-BGC model to update the modeled leaf carbon state variable ( $C$ , unit  $\text{kg C}\cdot\text{m}^{-2}$ ). The updated state variables are then used to simulate various ecological processes and ultimately affect the resulting carbon-water flux simulations. In the Biome-BGC model, the LAI is calculated by multiplying the specific leaf area (SLA, unit  $\text{m}^2\cdot\text{kg}^{-1}\text{C}$ ) by the leaf carbon state variable ( $C$ , unit  $\text{kg C}\cdot\text{m}^{-2}$ ) as follows:

$$\text{LAI} = C \times \text{SLA} \quad (1)$$

The complex ecosystem processes of carbon circulation and transformation are represented by the Biome-BGC model as the simulation of changes in leaf carbon (Running and Hunt, 1993). The amount of leaf carbon on that day ( $C_t$ ) is the sum of the amount of leaf carbon on the previous day ( $C_{t-1}$ ) and the current changes in leaf carbon ( $\Delta C$ ):

$$C_t = C_{t-1} + \Delta C_t \quad (2)$$

The calculation process of  $C_t$  includes the complex material migration and conversion processes within the ecosystem simulated by the Biome-BGC model, such as the amount of photosynthetic assimilation of organic matter in leaf tissue distribution and leaf tissue respiratory carbon consumption (Running and Hunt, 1993). Therefore, the dynamics of the state variable  $C_t$  can be represented in our data assimilation scheme as follows:

$$C_t = C_{t-1} + \text{BGC}(\Delta C_t) \quad (3)$$

where BGC represents the dynamic function of  $\Delta C_t$  in the Biome-BGC model.

In our data assimilation scheme,  $C_t$  is the state variable that is used in the assimilation process, and the observations to be assimilated are the MOD15A2 LAI data (which can be converted into  $C_t$  by Eq. (1)). During the actual data assimilation process, we convert the LAI observation representing time  $t$  to a  $C_t$  value of the corresponding time. Therefore, the measurement operator represents to be a  $1 \times 1$  identity matrix, and the model and observation error covariance matrix should be the model and observation error variance (Evensen, 1994; Burgers et al., 1998; Bouttier and Courtier, 2002). The updated state variable at time  $t$  ( $C_{t(\text{up})}$ ) can be

calculated using Eq. (4), where DA represents the data assimilation algorithms and is a function of observation ( $C_{t(\text{ob})}$ ) and modeled state variable ( $C_{t(\text{mo})}$ ).

$$C_{t(\text{up})} = \text{DA}(C_{t(\text{ob})}, C_{t(\text{mo})}) \quad (4)$$

The simulated state variable is then updated and used in the subsequent simulations.

$$C_{t(\text{mo})} = C_{t(\text{up})} \quad (5)$$

In our data assimilation scheme, the key process is to determine the corresponding model and observation errors, but this is also a difficult process. We first determined the errors of the modeled and observed LAI by referencing the methods reported by Fang et al. (2012) and Tian et al. (2002) by executing repeated data assimilation tests. The error of the  $C$  state variable was then reconverted to meet the requirements of the data assimilation process. In addition, the model background errors of the LAI were determined by comparing the measured data and the model simulation LAI results. The error of satellite observed LAI takes the same value, but the observation error of  $C$  converted from the LAI at the two flux points is different, and the model background error is also different. The Biome-BGC model is a daily time-step model, and the LAI product used in our assimilation scheme is an 8-day product. Therefore, the model assimilation window was 8 days, and the ensemble numbers of the EnKF were set to 200. More detailed parameter settings are shown in Table 2 and Table S1 (Supplementary Materials).

### 2.6. Evaluation of simulation framework

To analyze the performance of our simulation framework, we executed the following treatment: OS, original model (Running and Hunt, 1993) without data assimilation; IS, the improved model (Zhang et al., 2011a, b) without data assimilation; IEnKF, the improved model with data assimilation using the EnKF algorithm; and IUkf, the improved model with data assimilation using the UKF algorithm.

We calculated  $R^2$ , root mean square error (RMSE), and the sum of the absolute error (SAE) between the modeling results and the observed carbon-water flux records to evaluate the accuracy of the simulation framework. The RMSE and SAE calculations are as follows:

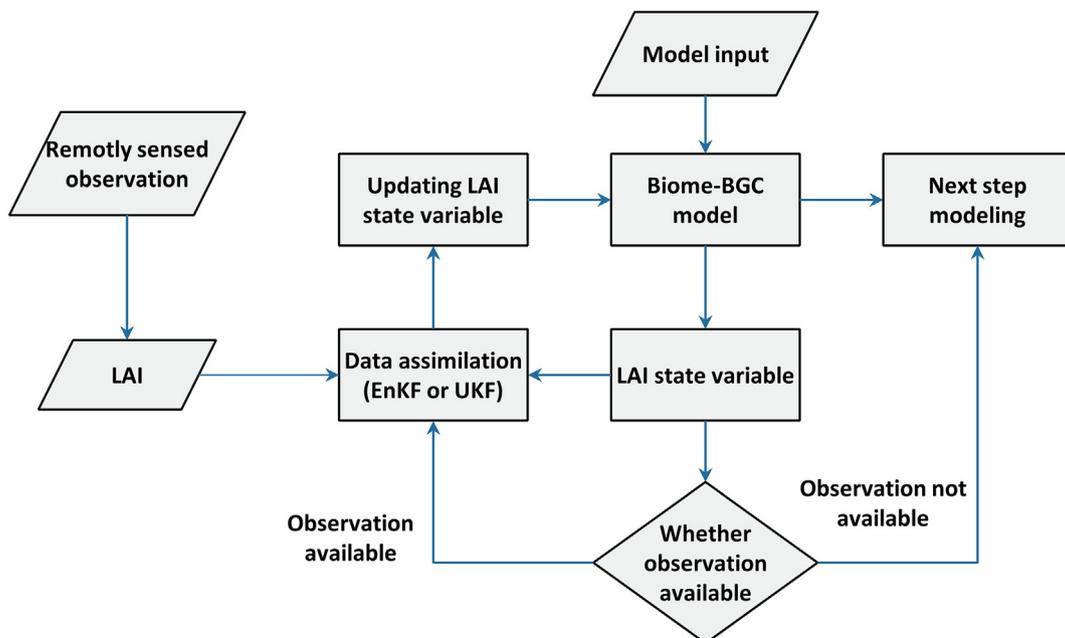


Fig. 2. Overall methodological flowchart for the model data assimilation scheme used in this study.

**Table 2**

The error variance settings of the model and observations at three vegetation type sites.

Forest function type	SLA (m <sup>2</sup> ·kg <sup>-1</sup> C)	Error variance of model LAI	Error variance of model Leaf.C	Error variance of observation LAI	Error variance of observation Leaf.C
DBF	30	1.44	0.0016	2.25	0.0025
EBF	12	1.44	0.01	2.25	0.016
GL	22.56	1.44	0.02	2.25	0.04

$$RMSE = \sqrt{\frac{1}{365} \sum_{i=1}^{365} |Model_i - Observations_i|^2} \quad (6)$$

$$SAE = \sum_{i=1}^{365} |Model_i - Observations_i| \quad (7)$$

where *i* represents the day of the year, and Model and Observations are the modeling results and the observed records from the flux site, respectively.

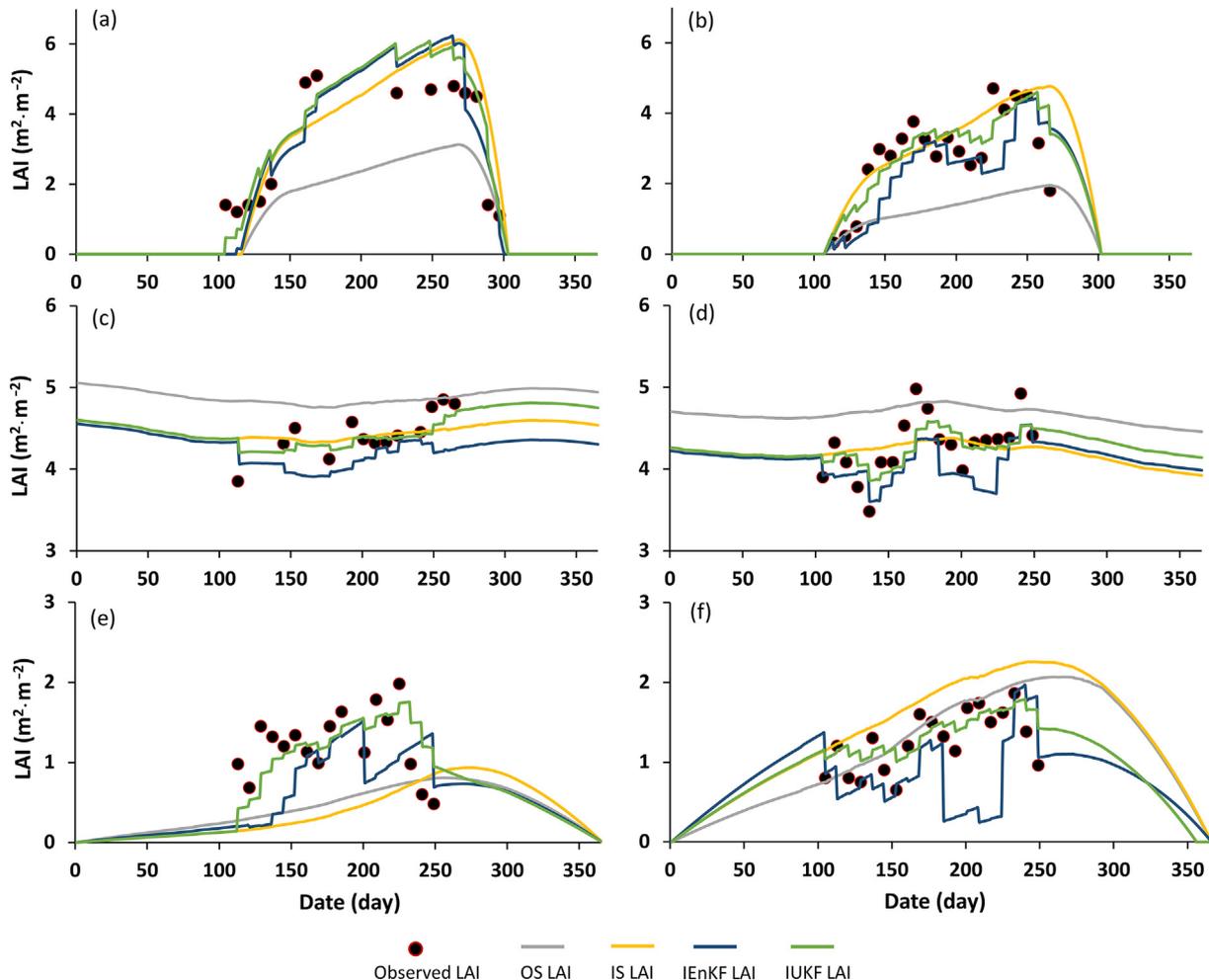
**3. Results**

Three years (2003–2005) of observed LAI from the HF, CBS, DHS, QY,

and HB sites and two years (2004–2005) of observed LAI from the DX site were assimilated into the Biome-BGC model. Although the performance of each assimilation treatment and its performance varied in different forest types. Overall, data assimilation discernably improved carbon-water flux estimations, showing improved matching between simulated carbon-water flux with observations over the simulated years. Here, the modeling results and observations from the first year of each site were used to evaluate the performance of the model under each treatment. The modeling results and evaluation results of the other years for each site are presented in Table S2.

**3.1. Evaluating modeled LAI under each treatment with observed LAI**

There was no significant difference in the annual trend between the observed LAI and the modeled LAI from each treatment, and it showed a tendency to first increase and then decrease during the growing season in DBF and GL (Figs. 3 and 4). In DBF (Fig. 3a and b), the modeled LAI from the model without improvement and the assimilation algorithms (OS LAI) was much lower than the LAI observation, especially during the mid-growing season (such as June, July, and August). The data assimilation treatments (both IEnKF and IUKF) significantly decreased RMSE and SAE and increased R<sup>2</sup> compared to those obtained under the OS and IS treatments (see Fig. 4a and b). In EBF (Fig. 3c and d), the modeled LAI from each treatment showed a relatively stable trend. As determined in the comparison of the OS and IS LAI, due to the assimilation algorithm, the trends of the IEnKF and IUKF LAI trends showed a jagged shape,



**Fig. 3.** Comparison of simulated LAI under each treatment and satellite observed LAI in each site. (a) HF station in 2003; (b) CBS station in 2003; (c) DHS station in 2003; (d) QY station in 2003. (e) DX station in 2004; (f) HB station in 2003.

showing fluctuations along with the LAI observations with lower RMSE and SAE and higher  $R^2$  (Fig. 4c and d). Similarly, data assimilation treatment improved the LAI estimation in GL showing closer trends for modeled LAI to the observations than OS and IS treatment (Fig. 3e and f and Fig. 4e and f). Overall, in the three vegetation types, due to the assimilation of the LAI observations, the LAIs simulated by EnKF and UKF were closer to the LAI observations. However, data assimilation is not just to close the observation, and it retains the dynamic change trend of the original simulated LAI.

### 3.2. Evaluating modeled carbon-water flux under each treatment with observations

Overall, assimilation algorithms discernably improved model performance toward ET and NEE (Figs. 5 and 6). In DBF, the simulated ET values from OS and IS were much higher than the observations. After data assimilation, although the simulated ET value was still higher than the observed value overall, it was closer to the actual observed value than was the OS (Fig. 5a and b). Compared with the OS treatment, the RMSE of the simulated ET from the IEnKF and IUKF treatments decreased by 13% and 15%, respectively, the SAE decreased by 11% and 13% (Fig. 7a), and the  $R^2$  increased by 6% and 15%, respectively (see Fig. 8). Compared with the IS treatment, the RMSE obtained for the simulated ET under the IEnKF and IUKF treatments decreased by 11% and 13%, respectively, the

SAE decreased by 8% and 10%, respectively (Fig. 7c), and the  $R^2$  remained almost unchanged (see Fig. 8). For the carbon flux (Fig. 6a and b) in DBF, compared to the OS treatment, the RMSE of the simulated NEE from the IEnKF and IUKF treatments decreased by 25% and 21%, respectively, the SAE decreased by 21% and 11% (Fig. 7b), respectively, and the  $R^2$  increased 14% and 14%, respectively (see Fig. 9). Compared to the IS treatment, the RMSE obtained under the IEnKF and IUKF treatments decreased by 13% and 7%, respectively, the SAE decreased by 12% and 1%, respectively (Fig. 7d), and the  $R^2$  increased by 18% and 18%, respectively (see Fig. 9).

Similar to DBF, simulated carbon-water flux from data assimilation treatment (IEnKF and IUKF) in EBF was closer to the observations than the OS and IS treatment (Fig. 5c and d; Fig. 6c and d). As Fig. 7a shows, compared with OS treatment, for ET estimation, the RMSE of the simulated ET from the IEnKF and IUKF treatments decreased by 25% and 24%, respectively, the SAE decreased by 25% and 24%, respectively, and the  $R^2$  increased by 38% and 54% (Fig. 8), respectively. In comparison with the IS treatment, the RMSE of IEnKF and IUKF decreased by 17% and 16%, respectively, the SAE decreased by 16% and 15%, respectively (Fig. 7c), and the  $R^2$  increased by 18% and 32%, respectively (Fig. 8). For the carbon flux, in comparison to the OS treatment, the RMSE of the simulated NEE from the IEnKF and IUKF treatments decreased by 21% and 26%, respectively, the SAE decreased by 24% and 39%, respectively (see Fig. 7b), and the  $R^2$  increased by 50% and 33% (Fig. 9), respectively.

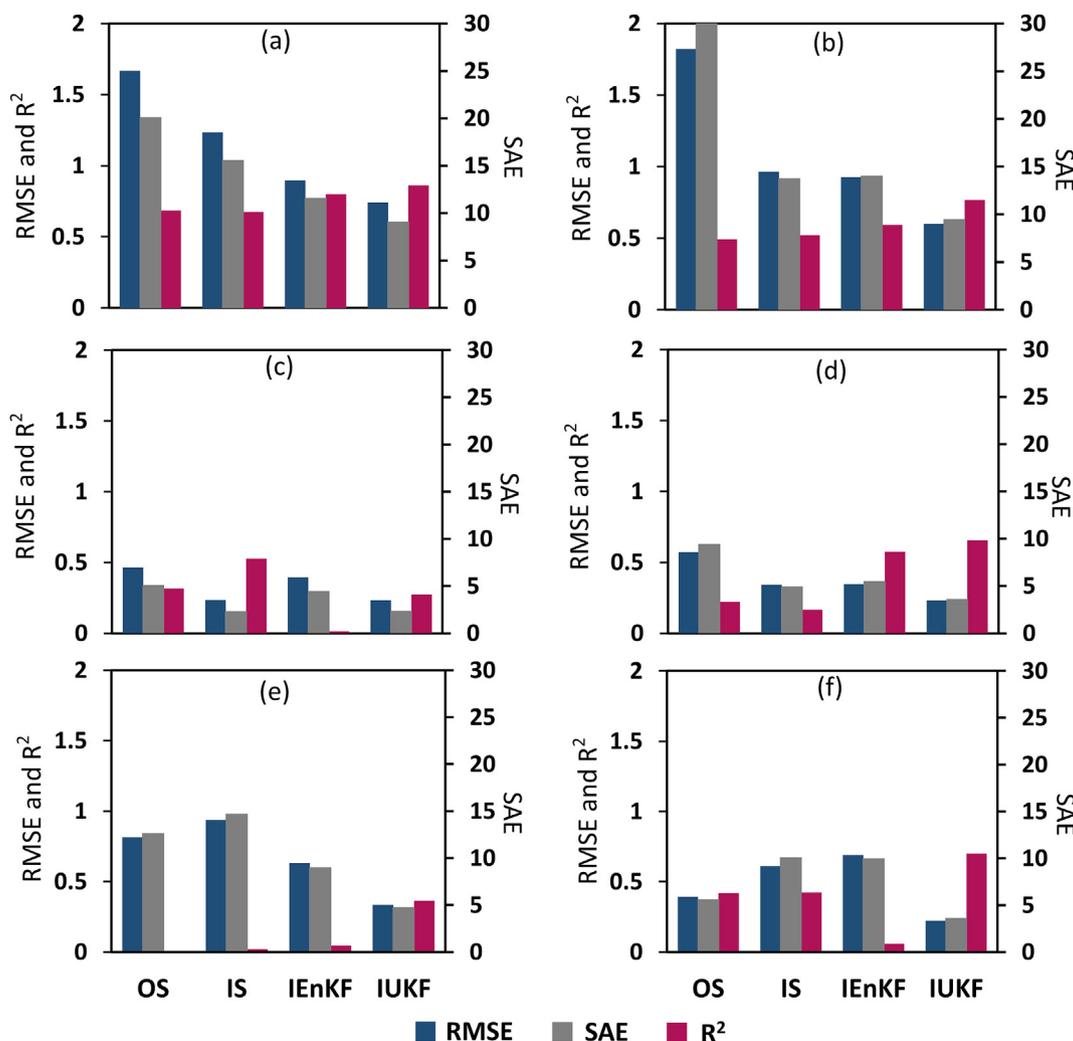
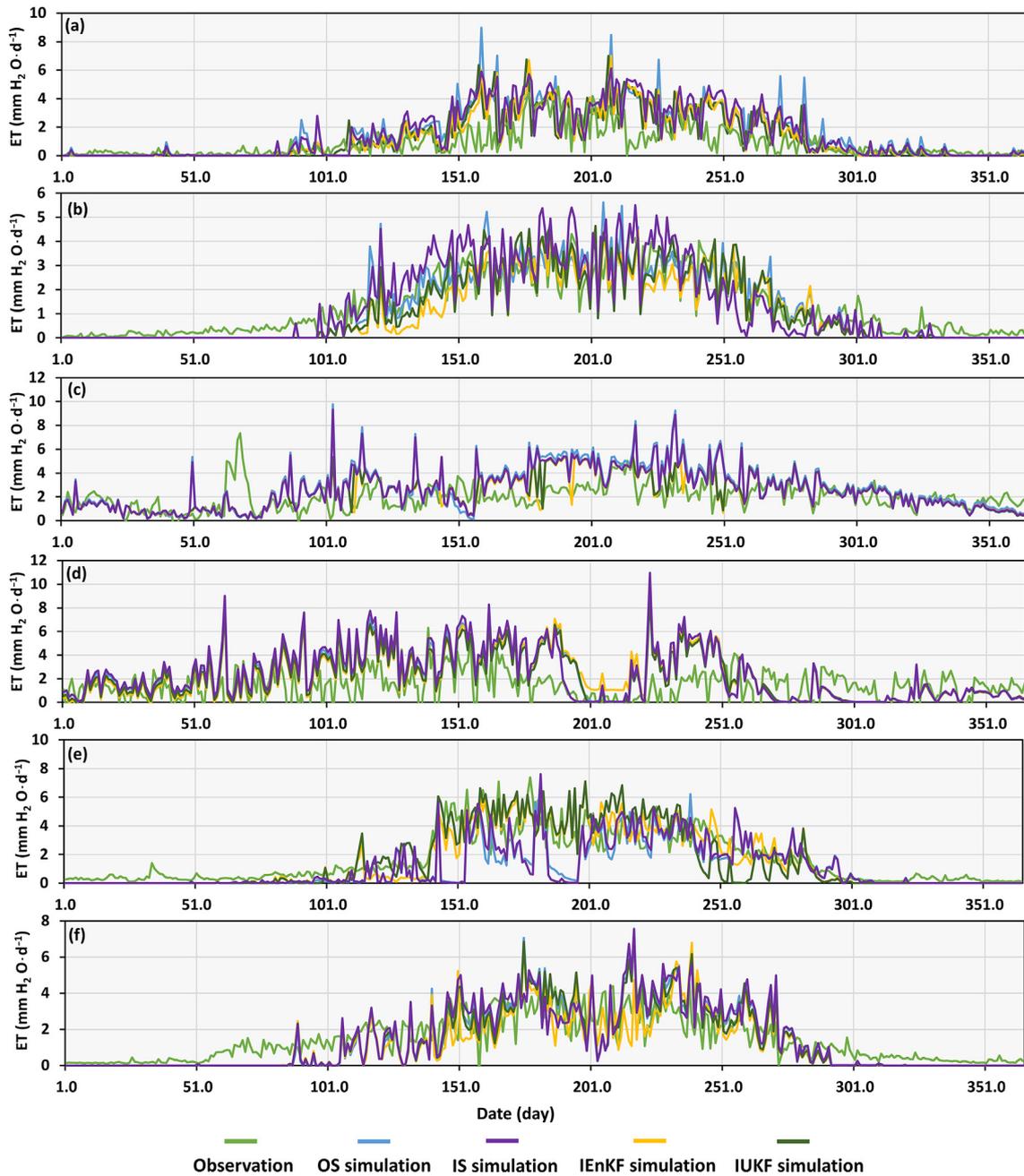


Fig. 4. The performance of each treatment for LAI state variable updating in each site. (a) HF station in 2003; (b) CBS station in 2003; (c) DHS station in 2003; (d) DX station in 2003. (e) DX station in 2004; (f) HB station in 2003. RMSE: root mean square error; SAE: the sum of the absolute error.



**Fig. 5.** Simulated ET (mm H<sub>2</sub>O·d<sup>-1</sup>) under each treatment and observed ET in each site. (a) HF station in 2003; (b) CBS station in 2003; (c) DHS station in 2003; (d) QY station in 2003; (e) DX station in 2004; (f) HB station in 2003.

Compared to the IS treatment, the IEnKF and IUKF simulations also achieved better performances (Fig. 7d), showing decreased in RMSE (by 13% and 19%, respectively), and SAE (by 10% and 25%, respectively) and increased  $R^2$  (by 29% and 14%, respectively).

For the GL vegetation type, the IEnKF and IUKF treatments performed better than both the OS and IS treatments when simulating the water flux but did not perform as well when simulating the carbon flux (Fig. 5e and f; Fig. 6e and f). Compared with the OS treatment, the RMSE of the simulated ET from the IEnKF and IUKF treatments decreased by 20% and 13%, respectively, the SAE decreased by 9% and 10%, respectively, and the  $R^2$  increased by 50% and 48%, respectively (see Figs. 7a and 8). Compared with IS treatment, the RMSE for IEnKF and IUKF decreased by 21% and

14%, respectively, and SAE for IEnKF and IUKF decreased by 9% and 10%, respectively (Fig. 7c); the  $R^2$  of IEnKF and IUKF treatments increased by 53% and 51% (Fig. 8), respectively. However, compared with the OS treatment, the simulated NEE from the data assimilation treatment had lower accuracy, showing 80% and 6% increases in the RMSE, and 80% and 15% increases in the SAE, respectively (see Fig. 7b). As determined through a comparison with the IS treatment outputs, the NEE estimation was not improved under the IEnKF treatment; both the RMSE and SAE values increased by approximately 60% (Fig. 7d). However, the IUKF treatment performed better than the IEnKF treatment, showing decreased RMSE and SAE by 11% and 6% (Fig. 7d), respectively, and an increased  $R^2$  value by over 54% (Fig. 9).

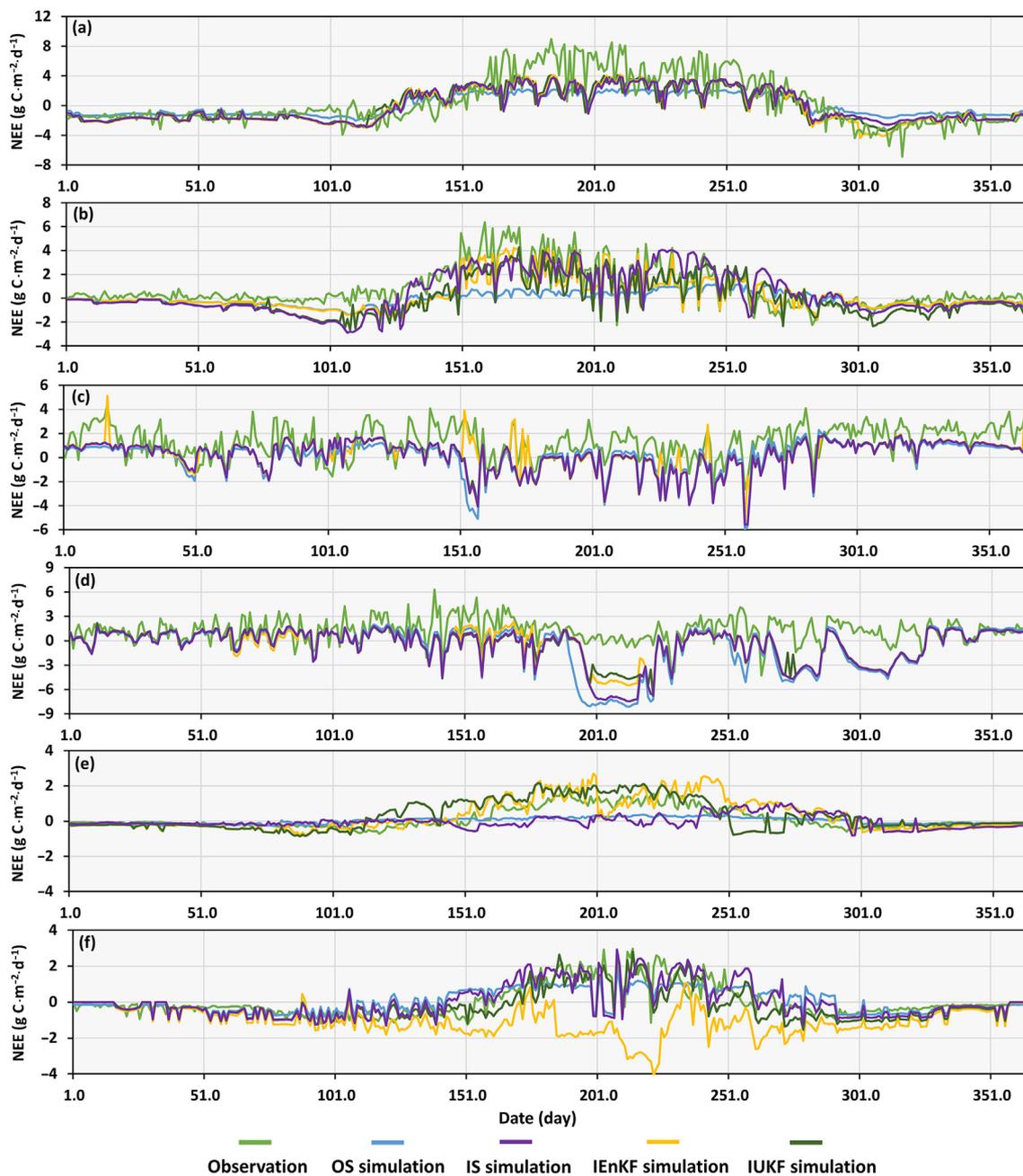


Fig. 6. Simulated NEE ( $\text{g C}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ ) under each treatment and observed NEE in each site. (a) HF station in 2003; (b) CBS station in 2003; (c) DHS station in 2003; (d) QY station in 2003; (e) DX station in 2004; (f) HB station in 2003.

### 3.3. Evaluating the performance of the IEnKF and IUKF treatments

Although overall, the data assimilation scheme greatly improved the simulation accuracy of the carbon-water flux, the performance of the IEnKF and IUKF treatments in the different vegetation types as well as in simulating carbon-water flux simulations was different (Figs. 5–7). In DBF, the IUKF treatment performed better than the IEnKF in water flux simulation. However, the carbon flux simulation from the IEnKF treatment was closer to the observation. In EBF, these two assimilation treatments had similar performance in water flux simulation, but the IEnKF treatment made the carbon flux simulation closer to the observation. Different from DBF and EBF, although the IUKF and IEnKF treatments improved the water flux simulation, they increased the error in the carbon flux simulation.

## 4. Discussion

### 4.1. LAI state variable updating

Most uncertainties in carbon-water flux modeling are introduced by the model architecture and model parameters (e.g., input parameters, state variables), resulting in biased modeling results (Huang et al., 2015). The LAI influences many biological and physical processes of vegetation, such as photosynthesis, respiration, transpiration, and light and rain interception, and thus plays an important role in the water, carbon and energy cycles of terrestrial ecosystems (Asrar et al., 1984; Chen and Cihlar, 1996; Zhu et al., 2018b). Modeling carbon-water flux without considering real variations in the LAI over time may induce uncertainties in the modeling results (Houborg et al., 2015). The importance of LAI

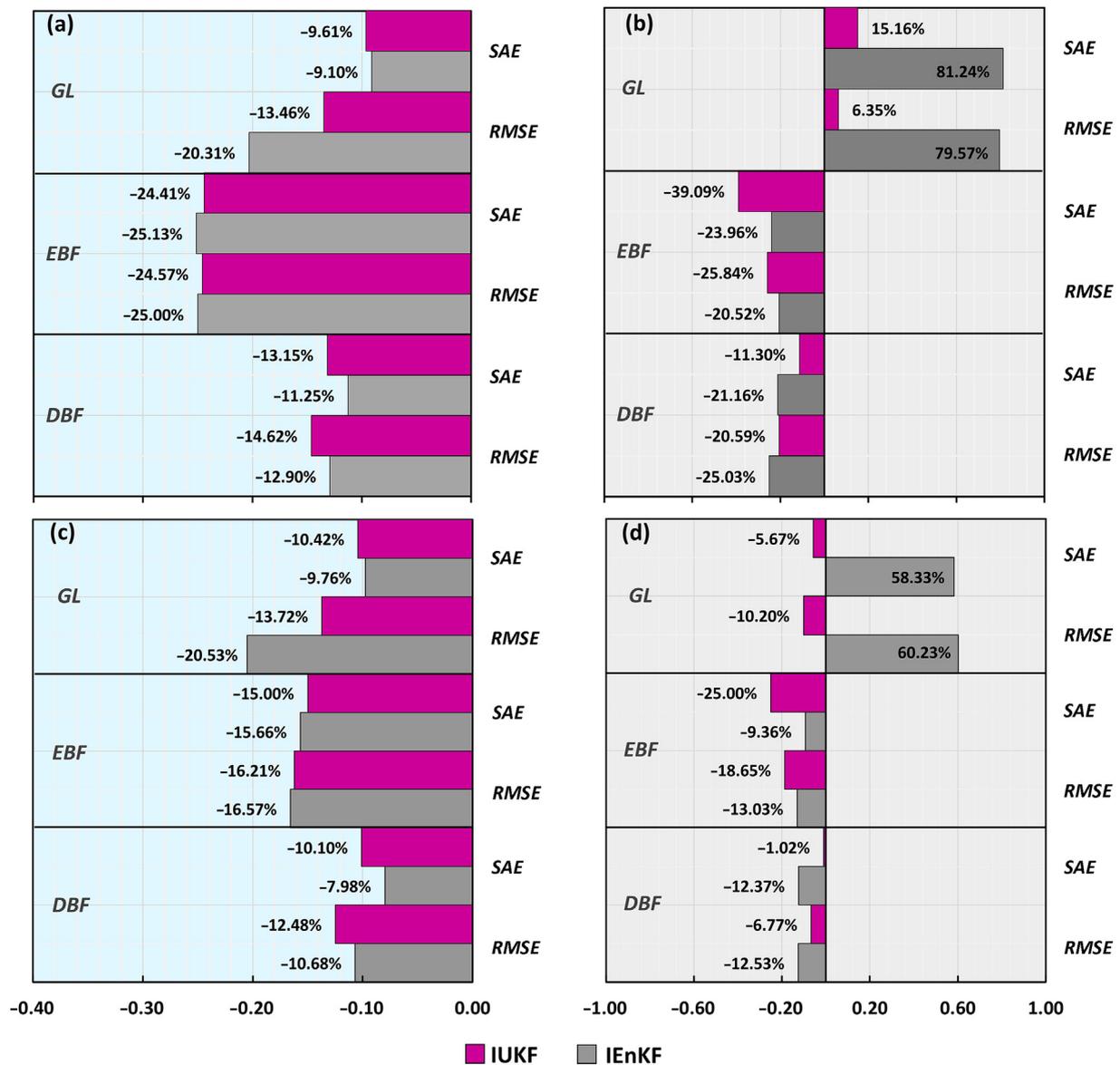


Fig. 7. The performance of the IES and IUKF treatment for carbon-water flux simulation compared to the OS treatment (a and b) and IS treatment (c and d) in DBF, EBF and GL. a: compared to the OS treatment for ET simulation; b: compared to the OS treatment for NEE simulation; c: compared to the IS treatment for ET simulation; d: compared to the IS treatment for NEE simulation. DBF: deciduous broad-leaved forest; EBF: evergreen broad-leaved forest; GL: grassland. Slight blue background represents the comparison of ET simulation and gray background represents the comparison of NEE simulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

updating in carbon-water flux simulation has been highlighted by numerous previous studies.

In this study, satellite observed LAI data were used to improve the LAI state variable updating through the use of data assimilation (both EnKF and UKF). The results showed that the estimated LAI from the IEEnKF and IUKF treatments were closer to the observations than those estimated under the OS and IS treatments were. In DBF, the original model significantly underestimated the LAI, while in EBF, it overestimated the LAI. Interestingly, in GL, the estimated LAI from the original model varied by study site, showing underestimation at the DX site and overestimation at the HB site (Fig. 3). Although the estimated LAI from the IS treatment showed higher accuracy than the OS treatment did, the variations in the LAI over time were not well captured by either treatment. After data assimilation, the estimated LAI showed a jagged appearance that was accompanied by jumping, showing higher consistency with the fluctuations of the observed LAI. A more accurate estimation of the LAI

state variable due to data assimilation corrections triggers a more robust representation of some physiological processes, such as respiration and photosynthetic activity, eventually leading to improvements in carbon-water flux estimations. Despite such advantages involved in the assimilation of observed LAI for carbon-water flux simulations, a proper understanding of the uncertainties associated with remotely sensed LAI observations is critical (Kala et al., 2014; Liu et al., 2018). As they represent modeling results, the MODIS LAI products used in this study inevitably have uncertainties (Tian et al., 2002). Such uncertainties can be generated from utilized retrieval algorithms, the presence of snow or clouds, or through atmospheric effects (van den Hurk et al., 2003). However, during the past two decades, the performance of the MODIS algorithm has been extensively tested and further optimized (Myneni et al., 2002; Yang et al., 2006). MODIS LAI products have been widely used in various fields since they were first publicly released in the 2000s (Yan et al., 2021). Given the wide application of MODIS LAI products in

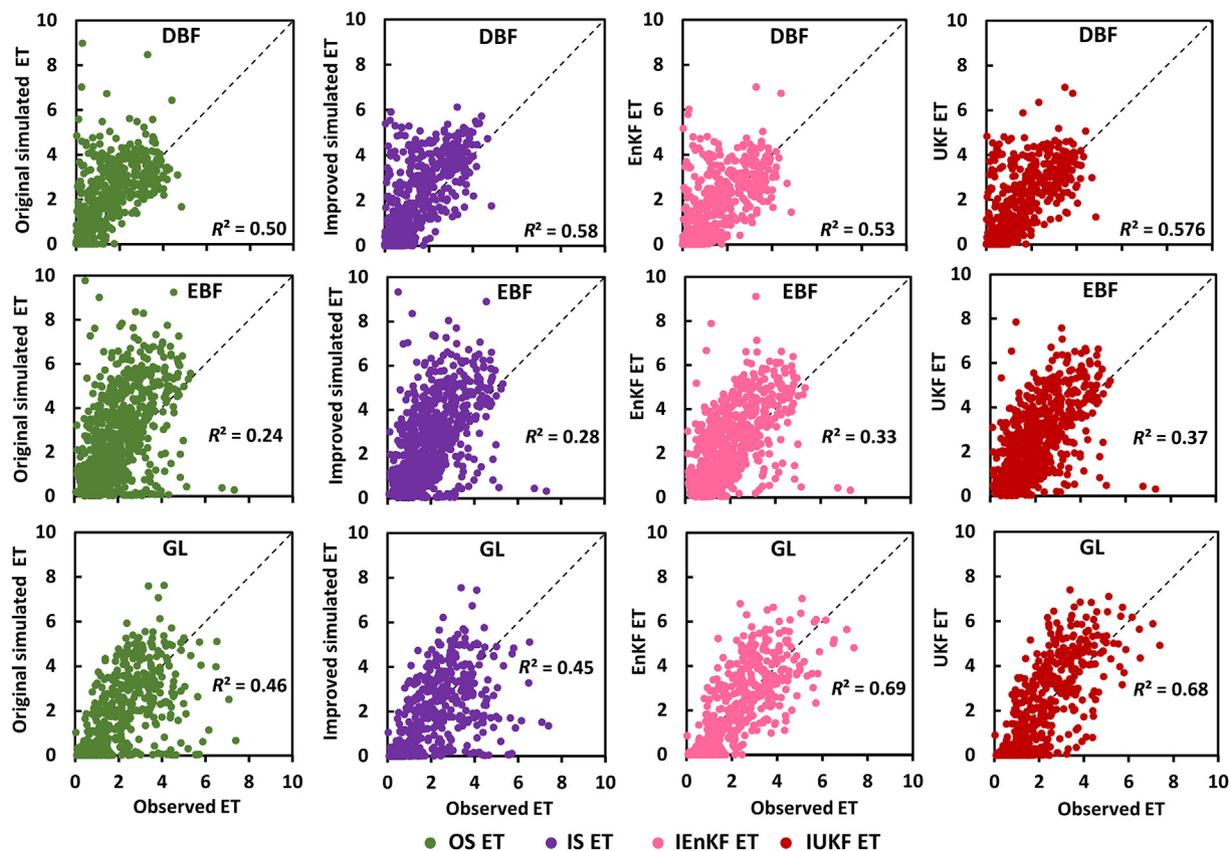


Fig. 8. Comparison of simulated ET ( $\text{mm H}_2\text{O}\cdot\text{d}^{-1}$ ) and observed ET ( $\text{mm H}_2\text{O}\cdot\text{d}^{-1}$ ) under each treatment in each vegetation type. DBF: deciduous broad-leaved forest; EBF: evergreen broad-leaved forest; GL: grassland. OS ET, IS ET, IEnKF ET and IUKE ET represent simulated ET from OS, IS, IEnKF and IUKE treatment, respectively.

various research fields and the quality control treatment in our study, we are confident in the results obtained from the LAI assimilation treatments conducted herein.

#### 4.2. The impact of data assimilation on carbon-water flux simulation

After data assimilation, the model performed differently in each forest type. In summary, data assimilation improved the ET simulation more significantly than the NEE simulation (Fig. 7). Generally, the assimilation of the LAI-simulated daily NEE matched the magnitude of the observed values more closely than the OS and IS treatments did. The original and improved models significantly underestimated NEE, especially during the vegetation growth season. In DBF and EBF, after data assimilation, the overall representation of the simulated NEE was improved, indicating that a more accurate estimated LAI state variable was an efficient way to better simulate carbon NEE. The LAI state variable was reasonably constrained by observations through assimilation during simulation to yield results that approximate reality as closely as possible. These findings were similar to those reported in many previous studies (e.g. Williams et al., 2005; Revill et al., 2013; Zhang et al., 2016). However, in GL, data assimilation enhanced the deviation between the simulated NEE and the observations. The original and improved models underestimated NEE, and although the IEnKF and IUKE treatments increased the simulated NEE (except for the IEnKF treatment at the HB site, see Fig. 7), the simulated values were higher than the observed values, which may have resulted from several aspects. First, state-of-the-art biogeochemical models are known to be impacted by several aspects, resulting in uncertainties in carbon flux simulation, especially in GL ecosystems (Stocker et al., 2013; Sándor et al., 2016). This result is partly because the functioning of GL ecosystems is highly dependent on hydrology and soil processes (Nagy et al., 2010); therefore, the simulations are highly sensitive to errors in any of the governing environmental factors and

disturbances (Hidy et al., 2012). In addition, according to the climate observations, during the time period when the simulated NEE deviates most from the observed value, the rainfall, daily mean maximum temperature, and vapor pressure deficit all correspond to the extreme values of the simulated year. To alleviate the uncertainties involved in carbon flux simulation, further improvements, such as model parameter and model structure optimization, need to be made to model carbon dynamics in GL ecosystems (Sándor et al., 2016).

The OS and IS treatments at five out of six sites overestimated ET. After data assimilation, the simulated daily ET matched the magnitude of the observed values more closely than the OS and IS treatments did. Similarly, a more accurate estimated LAI state variable resulting from data assimilation efficiently improved ET simulations. Similarly, Albergel et al. (2010) indicated that the joint assimilation of the LAI had a great impact on the water flux simulation. Pan et al. (2008) and Qin et al. (2008) applied data assimilation and a model to assimilate observations and showed that this process improved the efficiency of ET simulation. In DBF and EBF, Zhang et al. (2016) assimilated remotely sensed LAI into a biogeochemical model and found that data assimilation significantly improved ET simulation. Such studies could support the findings of our research. However, other studies have also shown that data assimilation in water flux estimation should be further investigated. For example, Xie and Zhang (2010) suggested that more efforts must be made to overcome the bottleneck involved in model parameters when constructing data assimilation relationships.

#### 4.3. The variation of the performance of data assimilation in different forest types

In the EnKF, an ensemble of possible state vectors, which are randomly generated using a Monte Carlo approach, represents the statistical properties of the state vector. The algorithm does not require a

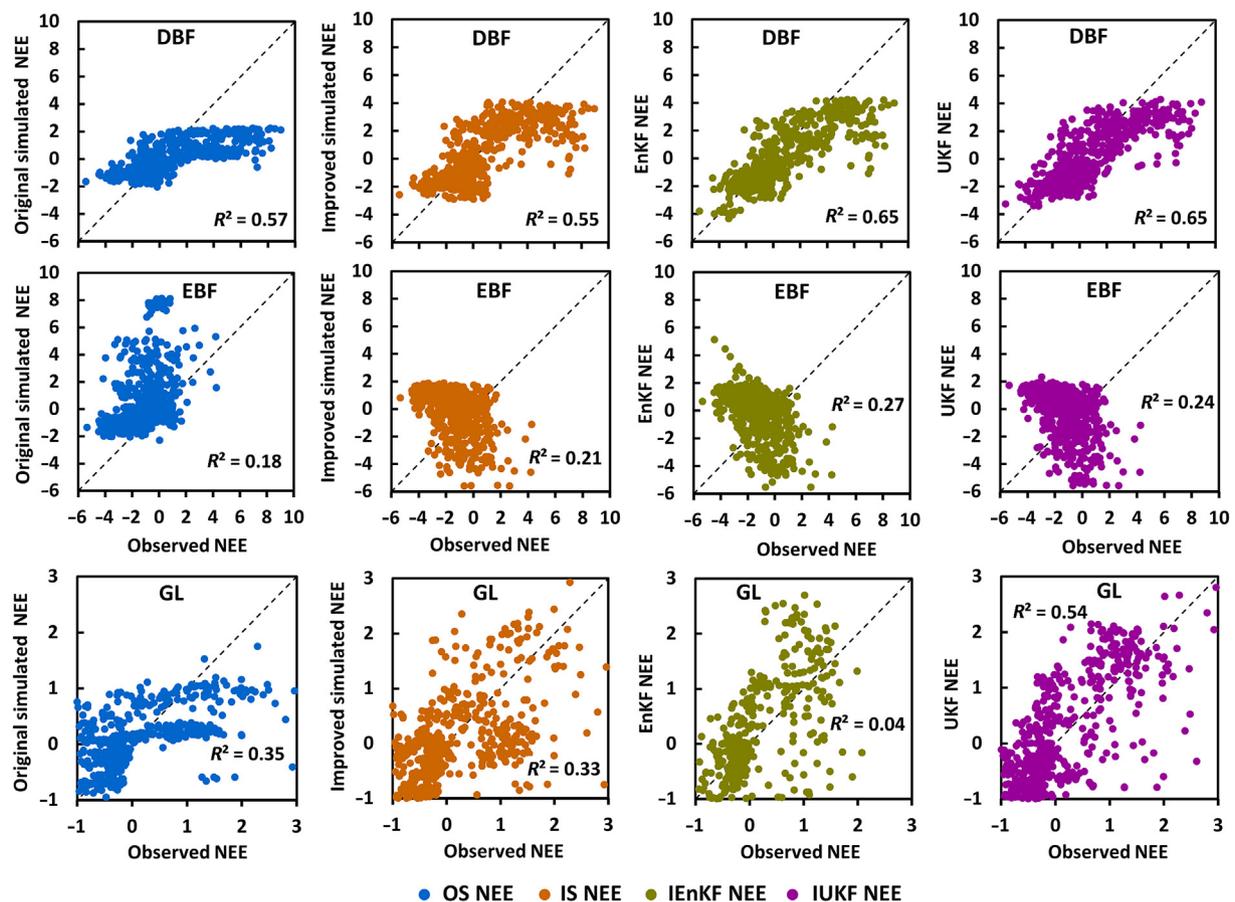


Fig. 9. Comparison of simulated NEE ( $\text{g C}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ ) and observed NEE ( $\text{g C}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ ) under each treatment in each vegetation type. DBF: deciduous broad-leaved forest; EBF: evergreen broad-leaved forest; GL: grassland. OS NEE, IS NEE, IEnKF NEE and IUKF NEE represent simulated NEE from OS, IS, IEnKF and IUKF treatment, respectively.

tangent linear model. For the UKF, instead of linearizing the functions as is done in the EnKF, the UT uses a set of points and propagates these points through the actual nonlinear function (Hommels et al., 2009). These differences led to slight differences in the performance of the EnKF and UKF in estimating the carbon-water flux for each forest type. For example, for ET simulation, the EnKF performed better than the UKF in EBF and GL, while the UKF was better in DBF (Fig. 7). The IUKF achieved better performance for NEE simulation in EBF and GL but had lower accuracy in DBF (Fig. 7). Neither approach showed universal superiority for carbon-water flux simulations. Due to its easy implementation, computational efficiency and optimum performance, the EnKF is widely used in data assimilation for carbon-water flux simulations (Ines et al., 2013). Compared with the EnKF, the UKF has received less attention in the research community. To some extent, our results enriched the application of the UKF in assimilating the LAI to simulate water-carbon fluxes and proved that the UKF could also be an efficient way to improve carbon-water flux simulations. The performance of data assimilation (both EnKF and UKF) in different vegetation types showed great differences. Specifically, data assimilation achieved the best performance in EBF for ET and NEE simulations (Fig. 7). Several reasons may have contributed to these results. First, evergreen trees, such as evergreen conifer trees, have developed a unique leaf regeneration form in which needles are usually retained for more than a year. Therefore, only a small number of needles are shed each year, resulting in the limited seasonality of EBF vegetation (Wang et al., 2019). The satellite observed EBF LAI values used for the data assimilation in this study are, therefore, more likely to be stable and involve fewer uncertainties throughout the year

than the data obtained for other vegetation types. In addition, the performance of the Biome-BGC model varies slightly among different vegetation types owing to different climate conditions and uncertainties in the model structure. For instance, Zhang et al. (2016) indicated that the uncertainties in the soil-water balance submodule of the Biome-BGC model may be negligible because precipitation at the simulated site was abundant and the soil moisture content was high. This higher stability and fewer uncertainties associated with the observed LAI data and model may have led to better EBF carbon-water flux simulations compared to the simulations obtained for other vegetation types. In DBF and GL, data assimilation showed similar performance in simulating ET (Fig. 7). However, there were negative impacts on NEE simulation in GL using the data assimilation technique. The high sensitivity of GL ecosystems to errors in any of the governing environmental factors and disturbances may contribute to this result (Hidy et al., 2012). Although the data assimilation performances were different among different vegetation types, overall, the improvement was significant. Considering that recent evidence has revealed that terrestrial carbon-water flux estimations remain uncertain (Zhu et al., 2018a; McCabe et al., 2019; Ryu et al., 2019; Li et al., 2021), our results provide new insights that could improve the quantification of carbon-water fluxes.

## 5. Conclusions

In this study, we used the Biome-BGC model to simulate carbon-water flux by incorporating assimilated remotely sensed LAI data derived from MODIS using the EnKF and UKF algorithms in DBF, EBF and GL. With the

assimilation of the MODIS LAI, the carbon-water flux simulation improved. The best carbon-water flux simulation results were obtained in EBF, and the EnKF and UKF achieved similar performance in this forest type, demonstrating the significant superiority of applying data assimilation in EBF. The assimilation of the LAI improved the water flux simulation more significantly than the carbon flux simulation. These results revealed that the improved updating of the LAI state variable during the simulation process by data assimilation could remarkably improve carbon-water flux simulations. Although both the EnKF and UKF improved carbon-water flux estimation, the performances of EnKF and UKF showed slight differences at each study site, suggesting that forthcoming global carbon-water flux estimation based on data assimilation should consider the vegetation types where the data assimilation experiments are carried out, the simulated objectives and the assimilation algorithms. Our results highlight that terrestrial carbon-water flux estimation can be improved through data assimilation techniques and can provide an opportunity to lead to a better understanding of the terrestrial carbon and water cycles.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

### Funding

This study was financially supported by the National Natural Science Foundation of China (No. 41301451).

### Authors' contributions

QL conceived, designed and performed the experiments, contributed the data collection, analyzed the data and prepared the manuscript. TZ conceived, designed experiments. MD assisted in data collection and analysis and prepared the manuscript. HG assisted in data computation. QZ and RS assisted in manuscript writing. All authors read and approved the final manuscript.

### Declaration of competing interest

All authors have no conflict of interest.

### Acknowledgements

The authors are grateful to the team of J. William Munger of the School of Engineering and Applied Sciences at Harvard University for providing the flux data of Harvard Forest Environmental Monitoring Station.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fecs.2022.100013>.

### References

Albergel, C., Calvet, J.C., Mahfouf, J.F., Rudiger, C., Barbu, A.L., Lafont, S., Roujean, J.L., Walker, J.P., Crapeau, M., Wigneron, J.P., 2010. 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>.

Asaadi, A., Arora, V.K., Melton, J.R., Bartlett, P., 2018. An improved parameterization of leaf area index (LAI) seasonality in the Canadian Land Surface Scheme (CLASS) and Canadian Terrestrial Ecosystem Model (CTEM) modelling framework. *Biogeosciences* 15, 6885–6907. <https://doi.org/10.5194/bg-15-6885-2018>.

Asrar, G., Fuchs, M., Kanemasu, E.T., Hatfield, J.L., 1984. Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat 1. *Agron. J.* 76, 300–306. <https://doi.org/10.2134/agronj1984.00021962007600020029x>.

Bouttier, F., Courtier, P., 2002. Data assimilation concepts and methods March 1999. Meteorol train course Lect Ser ECMWF 6–7. [http://msi.ttu.ee/&Eelken/Assim\\_concepts.pdf](http://msi.ttu.ee/&Eelken/Assim_concepts.pdf). (Accessed 15 March 2021).

Burgers, G., Van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* 126, 1719–1724. [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2).

Chen, M., Gihlar, J., 1996. Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote Sens. Environ.* 55, 153–162. [https://doi.org/10.1016/0034-4257\(95\)00195-6](https://doi.org/10.1016/0034-4257(95)00195-6).

Chen, M., Liu, S., Tieszen, L.L., Hollinger, D.Y., 2008. An improved state-parameter analysis of ecosystem models using data assimilation. *Ecol. Model.* 219, 317–326. <https://doi.org/10.1016/j.ecolmodel.2008.07.013>.

Dong, Q., Zhan, C., Wang, H., Wang, F.Y., Zhu, M.C., 2016. A review on evapotranspiration data assimilation based on hydrological models. *J. Geogr. Sci.* 26, 230–242. <https://doi.org/10.1007/s11442-016-1265-4>.

Evensen, G., 1994. Inverse methods and data assimilation in nonlinear ocean models. *Phys. Nonlinear Phenom.* 77, 108–129. [https://doi.org/10.1016/0167-2789\(94\)90130-9](https://doi.org/10.1016/0167-2789(94)90130-9).

Evensen, G., 2003. The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dynam.* 53, 343–367. <https://doi.org/10.1007/s10236-003-0036-9>.

Fang, H., Wei, S., Liang, S., 2012. Validation of MODIS and CYCLOPES LAI products using global field measurement data. *Remote Sens. Environ.* 119, 43–54. <https://doi.org/10.1016/j.rse.2011.12.006>.

Farquhar, G.D., von Caemmerer, S., von Berry, J.A., 1980. A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C3 species. *Planta* 149, 78–90. <https://doi.org/10.1007/BF00386231>.

Fu, Y.L., Yu, G.R., Sun, X.M., Li, Y.N., Wen, X.F., Zhang, L.M., Li, Z.Q., Zhao, L., Hao, Y.B., 2006. Depression of net ecosystem CO<sub>2</sub> exchange in semi-arid Leymus chinensis steppe and alpine shrub. *Agric. For. Meteorol.* 137, 234–244. <https://doi.org/10.1016/j.agrformet.2006.02.009>.

Gilardelli, C., Stella, T., Confalonieri, R., Ranghetti, L., Campos-Taberner, M., Garcia-Haro, F.J., Boschetti, M., 2019. Downscaling rice yield simulation at sub-field scale using remotely sensed LAI data. *Eur. J. Agron.* 103, 108–116. <https://doi.org/10.1016/j.eja.2018.12.003>.

Gove, J.H., Hollinger, D.Y., 2006. Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. *J. Geophys. Res. Atmos.* <https://doi.org/10.1029/2005JD006021>.

Hanes, C.C., Wang, X., Jain, P., Parisien, M.A., Little, J.M., Flannigan, M.D., 2019. Fire-regime changes in Canada over the last half century. *Can. J. For. Res.* 49, 256–269. <https://doi.org/10.1139/cjfr-2018-0293>.

Hidy, D., Barcza, Z., Haszpra, L., Churkina, G., Pinter, K., Nagy, Z., 2012. Development of the Biome-BGC model for simulation of managed herbaceous ecosystems. *Ecol. Model.* 226, 99–119. <https://doi.org/10.1016/j.ecolmodel.2011.11.008>.

Hommels, A., Murakami, A., Nishimura, S., 2009. A comparison of the ensemble Kalman filter with the unscented Kalman filter: application to the construction of a road embankment. *Geotechniek* 13, 52–54.

Houborg, R., McCabe, M.F., Cescatti, A., Gitelson, A.A., 2015. Leaf chlorophyll constraint on model simulated gross primary productivity in agricultural systems. *Int. J. Appl. Earth Obs. Geoinf.* 43, 160–176. <https://doi.org/10.1016/j.jag.2015.03.016>.

Houtekamer, P.L., Mitchell, H., 2005. Ensemble Kalman filtering. *Q. J. R. Meteorol. Soc.* 131, 3269–3289. <https://doi.org/10.1256/qj.05.135>.

Huang, J., Tian, L., Liang, S., Ma, H.Y., Becker-Reshef, I., Huang, Y.B., Su, W., Zhang, X.D., Zhu, D.H., Wu, W.B., 2015. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agric. For. Meteorol.* 204, 106–121. <https://doi.org/10.1016/j.agrformet.2015.02.001>.

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

Ito, A., Inatomi, M., 2012. Water-use efficiency of the terrestrial biosphere: a model analysis focusing on interactions between the global carbon and water cycles. *J. Hydrometeorol.* 13, 681–694. <https://doi.org/10.1175/JHM-D-10-05034.1>.

Julier, S., Uhlmann, J., 1996. A General Method for Approximating Nonlinear Transformations of Probability Distributions. Technical Report, Department of Engineering Science; Oxford University, Oxford, UK.

Kala, J., Decker, M., Exbrayat, J.F., Pitman, A.J., Carouge, C., Evans, J.P., Abramowitz, G., Mocko, D., 2014. Influence of leaf area index prescriptions on simulations of heat, moisture, and carbon fluxes. *J. Hydrometeorol.* 15, 489–503. <https://doi.org/10.1175/JHM-D-13-063.1>.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *J. Fluids Eng. Trans. ASME* 82, 35–45. <https://doi.org/10.1115/1.3662552>.

Kvamme, K.L., 1989. Geographic information systems in regional archaeological research, data management. In: Schiffer, M.B. (Ed.), *Archaeological Method and Theory*. University of Arizona Press, Tucson, pp. 139–203.

- Li, Y., Guo, Q., Su, Y., Tao, S.L., Zhao, K.G., Xu, G.C., 2017. Retrieving the gap fraction, element clumping index, and leaf area index of individual trees using single-scan data from a terrestrial laser scanner. ISPRS J. Photogrammetry Remote Sens. 130, 308–316. <https://doi.org/10.1016/j.isprsjprs.2017.06.006>.
- Li, Y., Li, L., Dong, J., Bai, J., 2021. Assessing MODIS carbon and water fluxes in grasslands and shrublands in semiarid regions using eddy covariance tower data. Int. J. Rem. Sens. 42, 595–616. <https://doi.org/10.1080/01431161.2020.1811915>.
- Liu, Y., Xiao, J., Ju, W., Zhu, G.L., Wu, X.C., Fan, W.L., Li, D.Q., Zhou, Y.L., 2018. Satellite-derived LAI products exhibit large discrepancies and can lead to substantial uncertainty in simulated carbon and water fluxes. Remote Sens. Environ. 206, 174–188. <https://doi.org/10.1016/j.rse.2017.12.024>.
- Luo, Y., Ogle, K., Tucker, C., Fei, S.F., Gao, C., LaDeau, S., Clark, J.S., Schimel, D.S., 2011. Ecological forecasting and data assimilation in a data-rich era. Ecol. Appl. 21, 1429–1442. <https://doi.org/10.1890/09-1275.1>.
- Luo, Y., White, L.W., Canadell, J.G., DeLucia, E.H., Ellsworth, D.S., Finzi, A.C., Lichten, J., Schlesinger, W.H., 2003. Sustainability of terrestrial carbon sequestration: a case study in Duke Forest with inversion approach. Global Biogeochem. Cycles 17, 1–13. <https://doi.org/10.1029/2002GB001923>.
- McCabe, M.F., Miralles, D.G., Holmes, T.R., Fisher, J., 2019. Advances in the remote sensing of terrestrial evaporation. Rem. Sens. 11, 1138–1146. <https://doi.org/10.3390/rs11091138>.
- Metropolis, N., Ulam, S., 1949. The Monte Carlo method. J. Am. Stat. Assoc. 44, 335–341.
- Mo, X., Chen, J.M., Ju, W., Black, T.A., 2008. Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble Kalman filter. Ecol. Model. 217, 157–173. <https://doi.org/10.1016/j.ecolmodel.2008.06.021>.
- Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115, 1781–1800. <https://doi.org/10.1016/j.rse.2011.02.019>.
- Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P., Nemani, P.R., Running, S.W., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens. Environ. 83, 214–231. [https://doi.org/10.1016/S0034-4257\(02\)00074-3](https://doi.org/10.1016/S0034-4257(02)00074-3).
- Nagy, Z., Barcza, Z., Horváth, L., 2010. Measurements and estimations of biosphere-atmosphere exchange of greenhouse gases—Grasslands. In: Haszpra, L. (Ed.), Atmospheric Greenhouse Gases: The Hungarian Perspective. Springer, Dordrecht Heidelberg London New York, 2011.
- Norby, R.J., Sholtis, J.D., Gunderson, C.A., Jawdy, S.S., 2003. Leaf dynamics of a deciduous forest canopy: no response to elevated CO<sub>2</sub>. Ecosyst. Ecol. <https://doi.org/10.1007/s00442-003-1296-2>.
- Pan, M., Wood, E.F., Wójcik, R., McCabe, M.F., 2008. Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation. Remote Sens. Environ. 112, 1282–1294. <https://doi.org/10.1016/j.rse.2007.02.039>.
- Peng, C., Guiot, J., Wu, H., Jiang, H., Luo, Y., 2011. Integrating models with data in ecology and palaeoecology: advances towards a model-data fusion approach. Ecol. Lett. 14, 522–536. <https://doi.org/10.1111/j.1461-0248.2011.01603.x>.
- Pipunic, R.C., Walker, J.P., Western, A., 2008. Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: a comparative synthetic study. Remote Sens. Environ. 112, 1295–1305. <https://doi.org/10.1016/j.rse.2007.02.038>.
- Post, H., Hendricks Franssen, H.-J., Han, X., Baatz, R., Montzka, C., Schmidt, M., Vereecken, H., 2018. Evaluation and uncertainty analysis of regional scale CLM4.5 net carbon flux estimates. Biogeosci. Discuss. 15, 1–37. <https://doi.org/10.5194/bg-2016-540>.
- Qin, C., Jia, Y., Su, Z., Zhou, Z., Qiu, Y., Suhui, S., 2008. Integrating remote sensing information into a distributed hydrological model for improving water budget predictions in large-scale basins through data assimilation. Sensors 8, 4441–4465. <https://doi.org/10.3390/s8074441>.
- Rawlins, F.B., Allard, S.P., Bovis, K.J., Clayton, A.M., Li, D., Inverarity, G.W., Lorenc, A.C., Payne, T.J., 2007. The Met office global four-dimensional variational data assimilation scheme. Q. J. R. Meteorol. Soc. 133, 347–362. <https://doi.org/10.1002/qj.32>.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S., 2010. Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. Water Resour. Res. 46, 1–22. <https://doi.org/10.1029/2009WR008328>.
- Reville, A., Sus, O., Barrett, B., Williams, M., 2013. Carbon cycling of European croplands: a framework for the assimilation of optical and microwave Earth observation data. Remote Sens. Environ. 137, 84–93. <https://doi.org/10.1016/j.rse.2013.06.002>.
- 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. J. Geophys. Res. Atmos. 115, 1–10. <https://doi.org/10.1029/2009JD012912>.
- Running, S.W., Gower, S.T., 1991. FOREST-BGC, a general model of forest ecosystem processes for regional applications. II. Dynamic carbon allocation and nitrogen budgets. Tree Physiol. 9, 147–160. <https://doi.org/10.1093/treephys/9.1.2.147>.
- Running, S.W., Hunt, E.R., 1993. Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models. In: Ehleringer, J.R., Field, C.B. (Eds.), Scaling Physiological Processes. Leaf to Globe. Academic Press, San Diego, California, USA.
- Ryu, Y., Berry, J.A., Baldocchi, D.D., 2019. What is global photosynthesis? History, uncertainties and opportunities. Remote Sens. Environ. 223, 95–114. <https://doi.org/10.1016/j.rse.2019.01.016>.
- Sándor, R., Barcza, Z., Hidy, D., Lellei-Kovacs, E., Ma, S., Bellocchi, G., 2016. Modelling of grassland fluxes in Europe: evaluation of two biogeochemical models. Agric. Ecosyst. Environ. 215, 1–19. <https://doi.org/10.1016/j.agee.2015.09.001>.
- Schaefer, K., Schwalm, C.R., Williams, C., Arain, M.A., Barr, A., Chen, J.M., Davis, K.J., Dimitrov, D., Hilton, T.W., Hollinger, D.Y., Humphreys, E., Poulter, B., Raczka, B.M., Richardson, A.D., Sahoo, A., Thornton, P., Vargas, R., Verbeeck, H., Anderson, R., Baker, I., Black, T.A., Bolstad, P., Chen, J., Curtis, P.S., Desai, A.R., Dietze, M., Dragoni, D., Gough, C., Grant, R.F., Gu, L., Jain, A., Kucharik, C., Law, B., Liu, S., Lokipitiya, E., Margolis, H.A., Matamala, R., McCaughey, J.H., Monson, R., Munger, J.W., Oechel, W., Peng, C., Price, D.T., Ricciuto, D., Riley, W.J., Roulet, N., Tian, H., Tonitto, C., Torn, M., Weng, E., Zhou, X., 2012. A model-data comparison of gross primary productivity: results from the north American carbon program site synthesis. J. Geophys. Res. Biogeosci. 117, 1–15. <https://doi.org/10.1029/2012JG001960>.
- Sims, D.A., Rahman, A.F., Cordova, V.D., El-Masri, B.Z., Baldocchi, D.D., Bolstad, P.V., Flanagan, L.B., Goldstein, A.H., Hollinger, D.Y., Misson, L., Monson, R.K., Oechel, W.C., Schmid, H.P., Wofsy, S.C., Xu, L., 2008. A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. Remote Sens. Environ. 112, 1633–1646. <https://doi.org/10.1016/j.rse.2007.08.004>.
- Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., 2013. Technical summary. In: Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, pp. 33–115.
- Tian, Y., Woodcock, C.E., Wang, Y., Privette, J.L., Shabanov, N.V., Zhou, L.M., Zhang, Y., Buermann, W., Dong, J., Veikkanen, B., Hame, T., Andersson, K., Ozdogan, M., Knyazikhin, Y., Myneni, R.B., 2002. Multiscale analysis and validation of the MODIS LAI product I. Uncertainty assessment. Remote Sens. Environ. 83, 414–430. [https://doi.org/10.1016/S0034-4257\(02\)00047-0](https://doi.org/10.1016/S0034-4257(02)00047-0).
- Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czirkowsky, M., Munger, J.W., 2007. Factors controlling CO<sub>2</sub> exchange on timescales from hourly to decadal at Harvard Forest. J. Geophys. Res. Biogeosci. 112, 1–25. <https://doi.org/10.1029/2006JG000293>.
- van den Hurk, B., Viterbo, P., Los, S.O., 2003. Impact of leaf area index seasonality on the annual land surface evaporation in a global circulation model. J. Geophys. Res. Atmos. <https://doi.org/10.1029/2002JD002846>.
- Vazifedoust, M., van Dam, J.C., Bastiaanssen, W.G.M., Feddes, R.A., 2009. Assimilation of satellite data into agrohydrological models to improve crop yield forecasts. Int. J. Rem. Sens. 30, 2523–2545. <https://doi.org/10.1080/01431160802552769>.
- Vukicevic, T., Braswell, B.H., Schimel, D., 2001. A diagnostic study of temperature controls on global terrestrial carbon exchange. Tellus Ser. B Chem. Phys. Meteorol. 53, 150–170. <https://doi.org/10.3402/tellusb.v53i2.16570>.
- Wan, E.A., van Der Merwe, R., 2000. The unscented Kalman filter for nonlinear estimation. In: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, Lake Louise, AB, pp. 153–158. Canada, 1–4 October 2000.
- Wan, E.A., van Der Merwe, R., Haykin, S., 2001. The unscented Kalman filter. Kalman Filter neural networks, 5, pp. 221–280. <https://doi.org/10.1002/0471221546.ch7>.
- Wang, R., Chen, J.M., Luo, X., Black, A., Arain, A., 2019. Seasonality of leaf area index and photosynthetic capacity for better estimation of carbon and water fluxes in evergreen conifer forests. Agric. For. Meteorol. 279, 107708. <https://doi.org/10.1016/j.agrformet.2019.107708>.
- Weiss, M., Baret, F., Smith, G.J., Jonckheere, I., Coppin, P., 2004. Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. Agric. For. Meteorol. 121, 37–53. <https://doi.org/10.1016/j.agrformet.2003.08.001>.
- Wen, X.F., Yu, G.R., Sun, X.M., Li, Q.K., Liu, Y.F., Zhang, L.M., Ren, C.Y., Fu, Y.L., Li, Z.Q., 2006. Soil moisture effect on the temperature dependence of ecosystem respiration in a subtropical *Pinus* plantation of southeastern China. Agric. For. Meteorol. 137, 166–175. <https://doi.org/10.1016/j.agrformet.2006.02.005>.
- Williams, M., Schwarz, P.A., Law, B.E., Irvine, J., Kurpius, M.R., 2005. An improved analysis of forest carbon dynamics using data assimilation. Global Change Biol. 11, 89–105. <https://doi.org/10.1111/j.1365-2486.2004.00891.x>.
- Wu, C., Chen, J.M., Huang, N., 2011. Predicting gross primary production from the enhanced vegetation index and photosynthetically active radiation: evaluation and calibration. Remote Sens. Environ. 115, 3424–3435. <https://doi.org/10.1016/j.rse.2011.08.006>.
- Wu, M., Scholze, M., Kaminski, T., Vossbeck, M., Tagesson, T., 2020. Using SMOS soil moisture data combining CO<sub>2</sub> flask samples to constrain carbon fluxes during 2010–2015 within a Carbon Cycle Data Assimilation System (CCDAS). Remote Sens. Environ. 240, 111719. <https://doi.org/10.1016/j.rse.2020.111719>.
- Xie, X., Zhang, D., 2010. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Adv. Water Resour. 33, 678–690. <https://doi.org/10.1016/j.advwatres.2010.03.012>.
- Yan, K., Zou, D., Yan, G., Fang, H., Weiss, M., Rautiainen, M., Knyazikhin, Y., Myneni, R.B., 2021. A bibliometric visualization review of the MODIS LAI/FPAR products from 1995 to 2020. J. Remote Sens. <https://doi.org/10.34133/2021/7410921>.
- Yan, M., Li, Z., Tian, X., Zhang, L., Zhou, Y., 2019. Improved simulation of carbon and water fluxes by assimilating multi-layer soil temperature and moisture into process-based biogeochemical model. For. Ecosyst. 6, 1–15. <https://doi.org/10.1186/s40663-019-0171-5>.
- Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N.V., Wang, Y., Privette, J.L., Member, I.E.E.E., Huemmrich, K.F., Fensholt, R., Sandholt, I., Weiss, M., Ahl, D.E., Gower, S.T., Nemani, R.R., Knyazikhin, Y., Myneni, R.B., 2006. MODIS leaf area index products: from validation to algorithm improvement. IEEE Trans. Geosci. Rem. Sens. 44, 1885–1896. <https://doi.org/10.1109/TGRS.2006.871215>.

- Yu, G.R., Wen, X.F., Sun, X.M., Tanner, B.D., Lee, X.H., Chen, J.Y., 2006. Overview of ChinaFLUX and evaluation of its eddy covariance measurement. *Agric. For. Meteorol.* 137, 125–137. <https://doi.org/10.1016/j.agrformet.2006.02.011>.
- Zhang, L.M., Yu, G.R., Sun, X.M., Wen, X.F., Ren, C.Y., Fu, Y.L., Li, Q.K., Li, Z.Q., Liu, Y.F., Guan, D.X., Yan, J.H., 2006. Seasonal variations of ecosystem apparent quantum yield ( $\alpha$ ) and maximum photosynthesis rate ( $P_{\max}$ ) of different forest ecosystems in China. *Agric. For. Meteorol.* 137, 176–187. <https://doi.org/10.1016/j.agrformet.2006.02.006>.
- Zhang, T.L., Sun, R., Zhang, R.H., Zhang, L., 2011a. Simulation of water and carbon fluxes in Harvard forest area by using improved Biome-BGC model. *Chin. J. Appl. Ecol.* 30, 2099–2106. <https://doi.org/10.3724/SP.J.1011.2011.00110>.
- Zhang, T.L., Sun, R., Hu, B., Feng, L., 2011b. Using simulated annealing algorithm to optimize the parameters of Biome-BGC model. *Chin. J. Ecol.* 30, 408–414. <https://doi.org/10.3724/SP.J.1011.2011.00110>.
- Zhang, T.L., Sun, R., Peng, C.H., Zhou, G.Y., Wang, C.L., Zhu, Q.A., Yang, Y.Z., 2016. Integrating a model with remote sensing observations by a data assimilation approach to improve the model simulation accuracy of carbon flux and evapotranspiration at two flux sites. *Sci. China Earth Sci.* 59, 337–348. <https://doi.org/10.1007/s11430-015-5160-0>.
- Zhu, P., Shi, L., Zhu, Y., Zhang, Q.R., Huang, K., Williams, M., 2017. Data assimilation of soil water flow via ensemble Kalman filter: infusing soil moisture data at different scales. *J. Hydrol.* 555, 912–925. <https://doi.org/10.1016/j.jhydrol.2017.10.078>.
- Zhu, X., Pei, Y., Zheng, Z., Dong, J.W., Zhang, Y., Wang, J.B., Chen, L.J., Doughty, R.B., Zhang, G.L., Xiao, X.M., 2018a. Underestimates of grassland gross primary production in MODIS standard products. *Rem. Sens.* <https://doi.org/10.3390/rs10111771>.
- Zhu, X., Skidmore, A.K., Wang, T., Liu, J., Darvishzadeh, R., Shi, Y.F., Premier, J., Heurich, M., 2018b. Improving leaf area index (LAI) estimation by correcting for clumping and woody effects using terrestrial laser scanning. *Agric. For. Meteorol.* 263, 276–286. <https://doi.org/10.1016/j.agrformet.2018.08.026>.