Assimilation of Carbonyl Sulfide (COS) fluxes within the adjoint-based data assimilation system—Nanjing University Carbon Assimilation System (NUCAS v1.0)

Huajie Zhu[1], Mousong Wu[1]*, Fei Jiang[1,2,3,4], Michael Vossbeck[5], Thomas Kaminski[5], Xiuli Xing[1], Jun Wang[1], Weimin Ju[1], Jing M. Chen[6]

[1]International Institute for Earth System Science, Nanjing University, Nanjing, 210023, China
[2]Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, School of Geography and Ocean Science, Nanjing University, Nanjing, 210023, China
[3]Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing, 210023, China
[4]Frontiers Science Center for Critical Earth Material Cycling, Nanjing University, Nanjing, 210023, China
[5]The Inversion Lab, Hamburg, Germany
[6]Department of Geography and Program in Planning, University of Toronto, ON M5S 3G3, Canada

Correspondence: Mousong Wu (mousongwu@nju.edu.cn)

Abstract. Modeling and predicting changes in the function and structure of the terrestrial biosphere and its feedbacks to climate change strongly depends on our ability to accurately represent interactions of the carbon and water cycles, and energy exchange. However, carbon fluxes, hydrological status and energy exchange simulated by process-based terrestrial ecosystem models are subject to significant uncertainties, largely due to the poorly calibrated parameters. In this work, an adjoint-based data assimilation system (Nanjing University Carbon Assimilation System, NUCAS) was developed, which is capable of assimilating multiple observations to optimize process parameters of a satellite data driven ecosystem model—BEPS (Boreal Ecosystem Productivity Simulator). Data assimilation experiments were conducted to demonstrate the robustness and to investigate the feasibility and applicability of NUCAS on seven sites by assimilating the carbonyl sulfide (COS) fluxes, which were tightly related to the stomatal conductance and photosynthesis. Results showed that NUCAS is able to achieve a consistent fit to COS observations across various ecosystems, including evergreen needleleaf forest, deciduous broadleaf forest, C3 grass and C3 crop. Comparing prior simulations with validation datasets, we found that the assimilation of COS can notably improve the model performance in gross primary productivity, sensible heat, latent heat and even soil moisture. We also showed that the NUCAS is capable of constraining parameters from multiple sites simultaneously and achieving a good consistency to the single-site assimilation. Our results demonstrate that COS can provide strong constraints on parameters relevant to water, energy and carbon processes with the data assimilation system, and open new perspectives for better understanding of the ecosystem carbon, water and energy exchanges.

Keywords: Carbonyl sulfide; Data assimilation; Carbon cycle; Satellite-driven; Ecosystem model

1 Introduction

Overwhelmingly due to anthropogenic fossil fuel and carbonate emissions, as well as land use and land cover change (Arias et al., 2021), atmospheric carbon dioxide (CO_2) concentrations have increased at an unprecedented rate since the Industrial Revolution and the global climate has been profoundly affected. As a key component of earth system, the terrestrial biosphere has absorbed about 30% of anthropogenic CO_2 emissions since 1850 and has significantly mitigated climate change (Friedlingstein et al., 2022). However, in line with large-scale global warming, the structure and function of terrestrial

biosphere have changed rapidly (Grimm et al., 2013; Arias et al., 2021; Moore and Schindler, 2022). As a consequence, terrestrial carbon fluxes are subject to great uncertainty (Macbean et al., 2022).

Terrestrial ecosystem models have been an important tool to investigate the net effect of complex feedback loops between the global carbon cycle and climate change (Zaehle et al., 2005; Fisher et al., 2014; Fisher and Koven, 2020). Meanwhile, with the advancement of modern observational techniques, a rapidly increasing number of satellite- and ground-based observational data have played an important role in studying the spatiotemporal distribution and mechanisms of the terrestrial ecosystem carbon fluxes (Rodell et al., 2004; Quirita et al., 2016). Various observations (Scholze et al., 2017), such as sun-induced chlorophyll fluorescence (Schimel et al., 2015) and soil moisture (Wu et al., 2018), have been used to estimate or constrain carbon fluxes in terrestrial ecosystems. Recently, carbonyl sulfide (COS) has emerged as a promising proxy for understanding terrestrial carbon uptake and plant physiology (Montzka et al., 2007; Campbell et al., 2008) since it is taken up by plants through the same pathway of stomatal diffusion as CO_2 (Goldan et al., 1988; Sandoval-Soto et al., 2005; Seibt et al., 2010) and completely removed by hydrolysis without any back-flux in leaves under normal conditions (Protoschill-Krebs et al., 1996; Stimler et al., 2010).

Plants control the opening of leaf stomata in order to regulate the water and CO_2 transit during transpiration and photosynthesis (Daly et al., 2004). As an important probe for characterizing stomatal conductance, COS has shown great potential to constrain plant photosynthesis and transpiration and to improve understanding of the water-carbon coupling (Wohlfahrt et al., 2012). A number of empirical or mechanistic COS plant uptake models (Campbell et al., 2008; Wohlfahrt et al., 2012; Berry et al., 2013) and soil exchange models (Kesselmeier et al., 1999; Berry et al., 2013; Launois et al., 2015; Sun et al., 2015; Whelan et al., 2016; Ogée et al., 2016; Whelan et al., 2022) have been developed to simulate COS fluxes in order to more accurately estimate gross primary productivity (GPP) as well as other key ecosystem variables. However, with the lack of ecosystem-scale measurements of the COS flux (Brühl et al., 2012; Wohlfahrt et al., 2012; Kooijmans et al., 2021), only few studies were conducted to systematically assess the ability of COS to simultaneously constrain photosynthesis, transpiration and other related processes in ecosystem models.

Data assimilation is an approach that aims at producing physically consistent estimates of the dynamical behavior of a model by combining the information in process-based models and observational data (Liu and Gupta, 2007; Law et al., 2015). It has been widely applied in geophysics and numerical weather prediction (Tarantola, 2005). In the past few decades, substantial efforts have been put into the use of various satellite- (Knorr et al., 2010; Kaminski et al., 2012; Deng et al., 2014; Scholze et al., 2016; Norton et al., 2018; Wu et al., 2018) and ground-based (Knorr and Heimann, 1995; Rayner et al., 2005; Santaren et al., 2007; Kato et al., 2013; Zobitz et al., 2014) observational datasets to constrain or optimize the photosynthesis, transpiration and energy-related parameters and variables of terrestrial ecosystem models via data assimilation techniques. In particular, by applying data assimilation methods to process-based models, not only can the observed dynamics of ecosystems be more accurately portrayed, but also our understanding of ecosystem processes can be deepened, with respect to their responses to climate (Luo et al., 2011; Keenan et al., 2012; Niu et al., 2014).

In this study, we present the newly developed adjoint-based Nanjing University Carbon Assimilation System (NUCAS) v1.0. NUCAS v1.0 is designed to assimilate multiple observational data streams including COS flux data to improve the process-based Biosphere-atmosphere Exchange Process Simulator (BEPS) (Liu et al., 1997), which has been specifically extended for simulating the ecosystem COS flux with the advanced two-leaf model that is driven by satellite observations of leaf area index (LAI).

In this context, the main questions that we aim to answer in this paper are:

What parameters is the COS simulation sensitive to and how do these parameters change in the assimilation of ecosystem-scale COS flux data?

How effective is the assimilation of COS fluxes in improving the carbon, water and energy balance for different ecosystems (including Evergreen needleleaf forest, deciduous broadleaf forest, C3 grass and C3 crop)?

Which processes are constrained by the assimilation of COS and what are the mechanisms leading to adjustments of the corresponding process parameters?

How robust is the NUCAS when optimizing over single-site and over two sites simultaneously?

To achieve these objectives, COS observations across a wide range of ecosystems (including evergreen needleleaf forest, deciduous broadleaf forest, C3 grass and C3 crop) are assimilated into NUCAS to optimize the model parameters using the four-dimensional variational (4D-Var) data assimilation approach, and the optimization results are evaluated against *in situ* observations. Specifically, materials and methods used in our study are described in Sect. 2. In this section, the BEPS model and our new data assimilation system NUCAS are introduced, along with the data used and the parameters chosen to be optimized in this study. The results are presented in Sect. 3, including the fit of COS simulations to observations, the variation and impact of parameters on simulated COS, as well as the comparison and evaluation of model outputs. Sect. 4 discusses the impacts of the COS assimilation on parameters and processes related to the water-carbon cycle and energy exchange as well as the influence of uncertainty inputs, in particular of the LAI driving data on posterior parameters values. In addition, the caveats and implications of assimilating COS flux are summarized. Finally, the conclusions are laid out in Sect. 5.

2 Materials and Methods

2.1 NUCAS data assimilation system

2.1.1 NUCAS framework

NUCAS is built around the generic satellite data driven ecosystem model BEPS, and applies the 4D-Var data assimilation method (Talagrand and Courtier, 1987). The BEPS model uses satellite-derived one-sided LAI to drive the phenology dynamics and separates sunlit and shaded leaves in calculating canopy-level energy fluxes and photosynthesis. It further features detailed representations of water and energy processes (**Figure 1**). These features render BEPS more advanced in representing ecosystem processes than standard ecosystem models (Richardson et al., 2012) with less parameters to be calibrated owing to the LAI-driven phenology.

Data assimilation if performed in two sequential steps: First, an inversion step adjusts the values of parameters controlling photosynthesis, energy balance, hydrology and soil biogeochemical processes to match the observations. Second, the posterior parameters obtained in the first step are used as input data for the second step, in which the BEPS model is re-run to obtain the posterior model variables. The schematic of the system is shown in **Figure 1**.

Considering model and data uncertainties, NUCAS implements a probabilistic inversion concept (Talagrand and Courtier, 1987; Tarantola, 1987; Tarantola, 2005) by using Gaussian probability density functions to combine the dynamic model and observations to obtain an estimate of the true state of the system and model parameters (Talagrand, 1997; Dowd, 2007). Hereby, we minimize the following cost function:

$$J(x) = \frac{1}{2}\left[(M(x) - O)^T {C_O}^{-1}(M(x) - \mathrm{O}) + (x - x_0)^T {C_x}^{-1}(x - x_0)\right] \quad (1)$$

where O and M denote vectors of observations and their modelled counterparts, respectively; x and x_0 denotes the control parameter vector with current and prior values, respectively. C_O and C_x denote the uncertainty covariance matrices for observations and prior parameters. Both matrices are diagonal expressing the assumption that observation uncertainties and the parameter uncertainties to be independent (Rayner et al., 2005). This definition of the cost function contains both the

mismatch between modelled and observed COS fluxes and the mismatch between current and prior parameter values (Rayner et al., 2005).

To determine an optimal set of parameters which minimizes J, a gradient-based optimization algorithm (BFGS) performs an iterative search (Wu et al., 2020). In each iteration, the gradient of J is calculated by applying the adjoint of the model, where the model is run backward to efficiently compute the sensitivity of J and with respect to x (Rayner et al., 2005). The gradient of J is used to define a new search direction. The adjoint model is an efficient sensitivity analysis tool for calculating the parametric sensitivities of complex numerical model systems (An et al., 2016). The computational cost of it is independent of the number of parameters and is in the current case comparable to 3–4 evaluations of J. In this study, all derivative code is generated from the model code by the automatic differentiation tool TAPENADE (Hascoët and Pascual, 2013). The derivative with respect to each parameter was validated against finite differences of model simulations, which showed agreement within the accuracy of the finite difference approximation.

The minimization of the cost function is implemented in a normalized parameter space where the parameter values are measured in multiples of their respective standard deviation with Gaussian priors (Kaminski et al., 2012). The model parameters are the various constants that are not influenced by the model state. Therefore, while they may change between plant function types (PFT) to reflect different conditions and physiological mechanisms, they will not change in time (Rayner et al., 2005).

2.1.2 BEPS basic model

The BEPS model (Liu et al., 1997; Chen et al., 1999; Chen et al., 2012) is a process-based diagnostic model driven by remotely sensed vegetation data, including LAI, clumping index, and land cover type, as well as meteorological and soil data (Chen et al., 2019). With the consideration of coupling among terrestrial carbon, water, and nitrogen cycles (He et al., 2021), the BEPS model now consists of photosynthesis, energy balance, hydrological, and soil biogeochemical modules (Ju et al., 2006; Liu et al., 2015). It stratifies whole canopies into sunlit and shaded leaves to calculate carbon uptake and transpiration for these two groups of leaves separately (Liu et al., 2015). For each group of leaves, the GPP is calculated by scaling Farquhar's leaf biochemical model (Farquhar et al., 1980) up to canopy-level with a new temporal and spatial scaling scheme (Chen et al., 1999), and the stomatal conductance is calculated using a modified version of the Ball–Woodrow–Berry model (Ball et al., 1987; Ju et al., 2006). Evapotranspiration is calculated as the summation of sunlit leaf and shaded leaf transpirations, evaporation from soil and wet canopy, and sublimation from snow storage on the ground surface (Liu et al., 2003). The BEPS model stratifies the soil profile into multiple layers (five were used in this study), and simulates temperature and water content from each layer (Ju et al., 2006). The soil water content is then used to adjust stomatal conductance considering the water stress impacts (Ju et al., 2010; He et al., 2021). Over the last few decades, the BEPS model has been continuously improved and used for a wide variety of terrestrial ecosystems (Schwalm et al., 2010; Liu et al., 2015).

The previous version of BEPS considers a total of six PFTs as well as eleven soil textures (Chen et al., 2012). We use the same soil texture but added four PFTs to BEPS in order to better discriminate vegetation types, especially the C4 grass and crop. Detailed information on these ten PFTs and eleven soil textures is given in **Table S1**.

2.1.3 COS modelling

The ecosystem COS flux, $F_{cos_{ecosystem}}$, includes both plant COS uptake $F_{cos,plant}$ and soil COS flux exchange $F_{cos,soil}$ (Whelan et al., 2016). In this study, these two components were modelled separately. The canopy-level COS plant uptake $F_{cos,plant}$ (pmol m^{-2} s^{-1}) was calculated by upscaling the resistance analog model of COS uptake (Berry et al., 2013) with

the upscaling scheme (Chen et al., 1999). Specifically, considering the different responses of foliage to diffuse and direct solar radiation (Gu et al., 2002), $F_{cos,plant}$ is calculated as:

$$F_{cos,plant} = F_{cos,sunlit} LAI_{sunlit} + F_{cos,shaded} LAI_{shaded} \quad (2)$$

where LAI_{sunlit} and LAI_{shaded} are the LAI values ($m^2\ m^{-2}$) of sunlit and shaded leaves, respectively. $F_{cos,sunlit}$ and $F_{cos,shaded}$ are the leaf-level COS uptake rate (pmol $m^{-2}\ s^{-1}$) of sunlit and shaded leaves, respectively. The leaf-level COS uptake rate $F_{cos_{leaf}}$ is calculated as:

$$F_{cos,leaf} = cos_a * \left(\frac{1.94}{g_{sw}} + \frac{1.56}{g_{bw}} + \frac{1}{g_{COS}}\right)^{-1} \quad (3)$$

where COS_a is the COS mole fraction in the bulk air. g_{sw} and g_{bw} are the stomatal conductance and leaf laminar boundary layer conductance to water vapor (H_2O). The factors 1.94 and 1.56 account for the smaller diffusivity of COS with respect to H_2O (Seibt et al., 2010; Stimler et al., 2010). g_{COS} denotes the apparent conductance for COS uptake from the intercellular airspaces, combining the mesophyll conductance and the biochemical reaction rate of COS and carbonic anhydrase (CA). Independent studies indicate that both CA activity (Badger and Price, 1994) and mesophyll conductance (Evans et al., 1994) tend to scale with the photosynthetic capacity or the maximum carboxylation rate of Rubisco at 25℃.

$$g_{COS} = \alpha * V_{cmax25} \quad (4)$$

Where α is a parameter that is calibrated to observations of simultaneous measurements of COS and CO_2 uptake (Stimler et al., 2012). Analysis of these measurements yield estimates of α of ~1400 for C3 and ~7500 for C4 species. With reference the COS modelling scheme of the Simple biosphere model (version 4.2) (Haynes et al., 2020), g_{COS} can be calculated as:

$$g_{cos} = 1.4 * 10^3 * (1.0 + 5.33 * F_{C4}) * 10^{-6} * F_{APAR} * f_w * V_{cmax} \quad (5)$$

where F_{C4} denotes the C4 plant flag, which takes the value of 1 when the vegetation is C4 plants and 0 otherwise. f_w is a soil moisture stress factor describing the sensitivity of g_{sw} to soil water availability (Ju et al., 2006). F_{APAR} is the scaling factor for leaf radiation, calculated as:

$$F_{APAR} = 1 - e^{(-0.45*LAI)} \quad (6)$$

$F_{cos,soil}$ is taken as the combination of abiotic COS flux $F_{cos,abiotic}$ and biotic COS flux $F_{cos,biotic}$ (Whelan et al., 2016).

$$F_{cos,soil} = F_{cos,abiotic} + F_{cos,biotic} \quad (7)$$

$F_{cos,abiotic}$ is described as an exponential function of the temperature of soil T_{soil} (℃).

$$F_{cos,abiotic} = e^{(alpha + beta * T_{soil})} \quad (8)$$

Where $alpha$ (unitless) and $beta$ ($℃^{-1}$) are parameters determined using the least-squares fitting approach.

$F_{cos_{biotic}}$is calculated according to Behrendt et al. (2014):

$$F_{cos,biotic} = F_{opt}\left(\frac{SWC}{SWC_{opt}}\right) * e^{-a\left(\frac{SWC}{SWC_{opt}} - 1\right)} \quad (9)$$

which can be rearranged to

$$a = ln\left(\frac{F_{opt}}{F_{SWC_g}}\right) * \left(ln\left(\frac{SWC_{opt}}{SWC_g}\right) + \left(\frac{SWC_g}{SWC_{opt}} - 1\right)\right)^{-1} \quad (10)$$

Here a is the curve shape constant, SWC is the soil moisture (percent volumetric water content). The maximum biotic COS uptake F_{opt} and the biotic COS uptake F_{SWC_g} are the COS fluxes ($pmol\ m^{-2}\ s^{-1}$) at optimum soil moisture SWC_{opt} and SWC_g, and $SWC_g > SWC_{opt}$. Here we use the parameterization scheme of soil COS modelling from Whelan et al. (2016) and Whelan et al. (2022), see **Table S2** and **Table S3** for details. Specifically, with reference of Abadie et al. (2022) and Whelan et al. (2022), the mean modelled SWC and temperature of the top 9 cm of the soil profile in BEPS were utilized to drive the

COS soil model in this study, and the mean modelled SWC and temperature were calculated through a weighted average considering the depth of each soil layer. A more detailed description about the soil hydrology and stomatal conductance modelling approach of BEPS is provided in the appendix.

Then ecosystem COS flux $F_{cos,ecosystem}$ can be calculated as the sum of COS plant uptake and the COS soil flux.

2.2 Model parameters

Here we optimized a total of 76 parameters belonging to BEPS. Of these parameters, some are global and others differentiated by PFT or soil texture class. The prior values of the parameters are taken as model defaults which have been tuned previous model in development and validation studies (Kattge et al., 2009; Chen et al., 2012). The prior uncertainty of parameters is set based on previous research (Chen et al., 2022; Ryu et al., 2018). For a more detailed description of these parameters, see **Table S4** in the supplement.

2.3 Site description

In this study, NUCAS was operated at seven sites distributed on the Eurasian and North American continents in boreal, temperate and subtropical regions (as illustrated in **Figure 2**) based on field observations collected from several studies. These sites were representative of different climate regions and land cover types (in the model represented by PFTs, and soil textures, as depicted in **Table 1**). They contained 4 of the 10 PFTs used in BEPS and 3 of the 11 soil textures. The sites comprise AT-Neu, located at an intensively managed temperate mountain grassland near the village of Neustift in the Stubai Valley, Austria (Hörtnagl et al., 2011; Spielmann et al., 2020); the Danish ICOS (Integrated Carbon Observation System) Research Infrastructure site (DK-Sor), which is dominated by European beech (Braendholt et al., 2018; Spielmann et al., 2019); the Las Majadas del Tietar site (ES-Lma) located in western Spain with a Mediterranean savanna ecosystem (El-Madany et al., 2018; Spielmann et al., 2019); the Hyytiälä forest Station (FI-Hyy), located in Finland and is dominated by Scots Pine (Bäck et al., 2012; Vesala et al., 2022); an agricultural soybean field measurement site (IT-Soy) located in Italy (Spielmann et al., 2019); the Harvard Forest Environmental Monitoring Site (US-Ha1) which is dominated by red oak and red maple in Petersham, Massachusetts, USA (Urbanski et al., 2007; Wehr et al., 2017); the Wind River Experimental Forest site (US-Wrc), located within the Gifford Pinchot National Forest in southwest Washington state, USA, with 478 ha of preserved old growth evergreen needleleaf forest (Rastogi et al., 2018). For further information on all sites, see publications listed in **Table 1**.

2.4 Data

The NUCAS system was driven by several temporally and spatially variant and invariant datasets. The CO_2 and COS mole fractions in the bulk air were assumed to be spatially invariant over the globe and to vary annually. The CO_2 mole fraction data in this study are taken from the Global Monitoring Laboratory (https://gml.noaa.gov/ccgg/trends/global.html). For the COS mole fraction, the average of the COS mole fraction observations from sites SPO (South Pole) and MLO (Mauna Loa, United States) was utilized to drive the model, the data are publicly available on line at: https://gml.noaa.gov/hats/gases/OCS.html. The other main inputs include a remotely sensed LAI dataset, a meteorological dataset and a soil dataset. Additionally, in order to conduct data assimilation experiments and to evaluate the effectiveness of the assimilation of COS fluxes, field observations including the ecosystem-scale (eddy-covariance or gradient-based) COS flux, GPP, sensible heat (H), latent heat (LE) and soil water content (SWC) at these sites collected at the sites were used.

2.4.1 LAI dataset

The LAI dataset used here are the GLOBMAP global leaf area index product (Version 3) (see GLOBMAP global Leaf Area Index since 1981 | Zenodo), the Global Land Surface Satellite (GLASS) LAI product (Version 3) (acquired from ftp://ftp.glcf.umd.edu/) and the level-4 MODIS global LAI product (see LP DAAC - MOD15A2H (usgs.gov)). The GLOBMAP LAI product represents Leaf area index at a spatial resolution of 8 km and a temporal resolution of 8-day (Liu et al., 2012). The GLASS LAI product is generated every 8 days at a spatial resolution of 1 km (Xiao et al., 2016). And the MODIS LAI is an 8-day composite dataset with 500 m pixel size. As default, we used GLOBMAP products for assimilation experiments as much as possible given its good performance in the BEPS applications to various cases (Chen et al., 2019). The other two LAI products were used to investigate the effect of the LAI products on the parameter optimization results. Also, according to Spielmann et al. (2019), the GLOBMAP product had considerably underestimated the LAI at the DK-Sor site in June 2016, and we noticed it was not consistent with the vegetation phenology at ES-Lma in May 2016. Therefore, GLASS LAI was used at these two sites and the GLOBMAP product was used at the remaining five sites. In addition, these 8-days LAI data were interpolated into daily values by the nearest neighbour method.

2.4.2 Meteorological dataset

Standard hourly meteorological data as input for BEPS including air temperature at 2 m, shortwave radiation, precipitation, relative humidity and wind speed were taken from the FLUXNET database (AT-Neu, DK-Sor, ES-Lma, FI-Hyy and US-Ha1 see https://fluxnet.org), the AmeriFlux database (US-Ha1, US-Wrc, see https://ameriflux.lbl.gov) and the ERA5 dataset (Site AT-Neu, IT-Soy, US-Ha1 see https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview), respectively. Since the experiments were conducted at the site scale, we used the FLUXNET and AmeriFlux data, which contains information about the downscaling of meteorological variables of the ERA-Interim reanalysis data product (Pastorello et al., 2020) as far as possible, and supplemented them with ERA5 reanalysis data. Particularly, although AT-Neu is a FLUXNET site, its FLUXNET meteorological data are only available for the years 2002-2012 while the measurement of COS was performed in 2015. Therefore, we first performed a linear fit of its ERA5-Land data and FLUXNET meteorological data for 2002-2012, and then corrected the ERA5 data for 2015 with the fitted parameters to obtain downscaling information for the meteorological variables. Additionally, for US-Ha1, we used the FLUXNET data in 2012, and Ameriflux data and ERA5 shortwave radiation data in 2013 to drive the BEPS model, due to the absence of FLUXNET data in 2013 and the lack of shortwave radiation data of Ameriflux.

2.4.3 Assimilation and evaluation datasets

The hourly ecosystem-scale COS flux observations were used to perform data assimilation experiments and to evaluate the assimilation results. They were taken from existing studies (listed in **Table 1**) and were available for at least a month. Most of the ecosystem COS flux observations were obtained using the eddy-covariance (EC) technique, with the exception US-Wrc, where the COS fluxes were derived with the gradient-based approach. The COS soil measurements were collected using soil chamber, except at US-Ha1, where a sub-canopy flux-gradient approach was used to calculate the soil COS flux. Detailed information about the COS measurements can be found in the publications listed in **Table 1**. Specifically, only the measured ecosystem COS flux data of FI-Hyy (Vesala et al., 2022) was utilized in this study.

Since only the raw COS concentration data at different altitudes are provided in Rastogi et al. (2018), while the values of the parameters needed to calculate the COS fluxes by the aerodynamic gradient method are not provided, there may be considerable biases in our estimates of COS fluxes at US-Wrc. Therefore, a bias correction scheme was implemented to match the simulated and estimated the ecosystem-scale COS fluxes for the US-Wrc site. The objectives of this correction scheme are

to obviate the need for accurate values of parameters relevant for COS flux calculations, and to retain as much useful information from the COS concentration measurements as possible (Leung et al., 1999; Scholze et al., 2016). This was done by using the mean and standard deviation of the simulated COS flux to correct the COS flux observations:

$$\mathrm{F} = \frac{\sigma_M(O - \bar{O})}{\sigma_O} + \bar{M} \tag{11}$$

where $\overline{O}$ and σ_O are mean and standard deviation of the observed COS flux series. F is the corrected observed COS flux, which is matched to the simulated COS flux. $\overline{M}$ and σ_M are mean and standard deviation of the COS simulations, calculated from the simulations using the prior parameters for the time period corresponding to the COS flux observations.

The standard deviation of the ecosystem COS fluxes within 24 hours around each observation was calculated as estimate of the observation uncertainty. For the case where there are no other observations within the surrounding 24 hours, the uncertainty was taken as the mean of the estimated uncertainties of the whole observation series.

Due to the coupling between leaf exchange of COS, CO_2 and H_2O, GPP and LE data are selected to evaluate the model performance of COS assimilation in this study. In addition, we further explored the ability of COS to constrain SWC as well as H simulations since the water dissipated in transpiration originates from the soil (Berry et al., 2006) and the transpiration contribute to a decrease in temperature within the leaf (so called "cooling effect") (Gates, 1968; Konarska et al., 2016). These data were taken from FLUXNET (DK-Sor, ES-Lma, FI-Hyy and US-Ha1), AmeriFlux (US-Ha1 and US-Wrc) and existing studies (Spielmann et al. (2020), Spielmann et al. (2019) and Rastogi et al. (2018)). As only CO_2 turbulent flux (FC) data are available for US-Ha1 in 2013 and only net ecosystem exchange (NEE) data are available for IT-Soy, a night flux partitioning model (Reichstein et al., 2005) was used to estimate ecosystem respiration (R_{eco}) and thus to calculate GPP. The model assumes that nighttime NEE represents ecosystem respiration, and thus partitions FC or NEE into GPP and R_{eco} based on the semi-empirical models of respiration, which use air temperature as a driver (Lloyd and Taylor, 1994; Lasslop et al., 2012).

2.5 Experimental design

Three groups of data assimilation experiments were conducted in this study: (1) 14 model-based twin experiments were performed to investigate the ability of NUCAS to assimilate COS flux data in different scenarios; (2) 13 single-site assimilation experiments were conducted at all seven sites to obtain the site-specific posterior parameters and the corresponding posterior model outputs based on COS flux observations; (3) one two-site assimilation experiment was carried out to refine one set of parameters over two sites simultaneously and to simulate the corresponding model outputs. Prior simulations using default parameters were also performed in order to investigate the effect of the COS flux assimilation. Moreover, due to the limitation of the COS observations, all of these experiments were conducted in a one-month time window at the peak of the growing season. Detailed information of these experiments is described in the following.

2.5.1 Twin experiment

Model-based twin experiments were performed to investigate the model performance of the data assimilation (Irrgang et al., 2017) at all seven sites considering single-site and two-site scenarios, and under different perturbation conditions. In each twin experiment, we first created a pseudo-observation sequence by NUCAS using the prior parameters. The pseudo-observation time series included the prior simulated ecosystem COS fluxes with its uncertainties, and the latter were estimated as the standard deviation of the prior simulated COS fluxes within 24 hours around each simulation. Then, a given perturbation ratio was applied to the prior parameters vector, as a starting point for the interactive adjustment of parameter values to match the COS flux pseudo-observations. The effectiveness of the data assimilation methodology of NUCAS can be validated if it successfully restores the control parameters from the pseudo-observations. As a gradient-based optimization algorithm is used

in NUCAS to tune the control parameters and minimize the cost function, the changes of cost function and gradient over assimilation processes can also be used to verify the assimilation performance of the system. In this work, a total of fourteen twin experiments were conducted, including thirteen single-site twin experiments and one two-site twin experiment. With reference the uncertainty of parameters, a perturbation size of 0.2 was utilized in all of the twin experiments.

2.5.2 Real data assimilation experiment

After the ability of NUCAS to assimilate COS flux data was confirmed by twin experiments, we could then use the system to conduct data assimilation experiments with real COS observations under single-site and multi-site conditions to optimize the control parameters and state variables of this model, and use the evaluation dataset to test the posterior simulations of the state variables. For the single-site case, a total of thirteen data assimilation experiments were conducted at all of these sites to investigate the assimilation effect of COS flux on optimizing key ecosystem variables. Detailed information about those single-site experiments is shown in **Table 2**.

Single-site assimilation can fully account for the site-specific information, and thus achieve accurate calibration. However, this assimilation approach often yields a range of different model parameters between sites. For large-scale model simulations, only one set of accurate and generalized model parameters is required (Salmon et al., 2022). Thus, a two-site assimilation experiment that can assimilate COS observations from two sites simultaneously is necessary to be conducted. Although both DK-Sor and US-Ha1 are dominated by deciduous broadleaved forest, and both AT-Neu and ES-Lma are dominated by C3 grass, none of the COS data from these two PFTs overlap in observation time. We therefore selected FI-Hyy and US-Wrc, which are both dominated by evergreen needleleaf forest, and conducted a two-site assimilation experiment with a one-month assimilation window in August 2014.

2.6 Model evaluation

For the purpose of demonstrating the process of control parameter vector being continuously adjusted in the normalized parameter space in twin experiment, and quantifying the deviation of the current control vector from the prior, the distance (D_x) between the parameter vector and the prior parameter vector was calculated.

$$D_x = \|x - x_0\| = \sqrt{\sum_{i=1}^{n} \left(x(i) - x_0(i)\right)^2} \tag{12}$$

where i denotes the *i* th parameter in the parameter vectors and n denotes the number of parameters in the parameter vector, and takes a value of 76.

With the aim of evaluating the performance of NUCAS in the real data assimilation experiments, we reran the model to obtain the posterior model outputs based on the posterior model parameters. Typical statistical metrics including mean bias (MB), root mean square error (RMSE) and coefficient of determination (R^2) are used to measure the difference between the simulations and *in situ* observations. They were calculated as:

$$MB = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i) = \overline{M} - \overline{O} \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2} \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(M_i - O_i)^2}{\sum_{i=1}^{N}\left(O_i - \overline{O}\right)^2} \tag{15}$$

where M_i denotes the simulation corresponding to the i th observation O_i and N is the total number of observations.

Additionally, in order to investigate the sensitivity of COS assimilation to the model parameters, we also calculated the sensitivity index (SI) for each parameter at the prior value based on the sensitivity information provided by the adjoint model. SI of i th parameter $x(i)$ of the parameter vector x was calculated as:

$$\mathrm{SI}\big(x(i)\big) = \frac{\partial J/\partial x(i)}{\|\partial J/\partial x\|} \tag{16}$$

where $\|\partial J/\partial x\|$ denotes the norm of the sensitivity vector of the cost function to the model parameters.

3 Results

3.1 Twin experiments

After averaging about 18 and 13 evaluations of the cost function and its gradients, each of the twin experiments was successfully performed. Details of those twin experiments are shown in **Table S5**. In summary, during those assimilations, the cost function values were substantially reduced by more than sixteen orders of magnitude, from greater than 50.75 to less than 5.09×10^{-13} and the respective gradient values also reduced from greater than 38.81 to less than 1.59×10^{-6}, which verified the ability of the data assimilation algorithm to correctly complete the assimilation.

The relative changes of the parameters with respect to the prior values at the ends of the experiments, as well as the initial values (D_{itial}) and the maximums (D_{max}) and the final values (D_{final}) of D_x are reported in **Table S5**. Results show that the relative differences of those parameters from the "true" values reached exceedingly small values at the ends of twin experiments, with the maximum of the absolute values of the relative changes below $8.55 * 10^{-9}$. D_x was also reduced to nearly zero with the maximum value below $6.60 * 10^{-8}$, which indicates that all parameters in the control parameter vectors were almost fully recovered from the pseudo observations. In conclusion, these results demonstrate that NUCAS has excellent data assimilation capability under various scenarios with different perturbations, and can effectively perform iterative computations to obtain reliable parameter optimization results during the assimilation process.

3.2 Single-site assimilation

With an average of approximately 113 cost function evaluations, all of the 13 single-site experiments were performed successfully. The experiments reduced cost function values substantially, with an average cost function reduction of 24.43 % (**Table 2**). However, the cost function reduction of the experiment varies considerably with PFT, site and assimilation window, ranging from 4.87 % to 69.05 %. The cost function decreased dramatically at US-Ha1, with an average decrease of 56.59 %. In contrast, at IT-Soy, the cost function reduction is only 4.87 %. With a same PFT (C3 grass), the cost function decreased by a similar degree at AT-Neu and ES-Lma, with the cost function reduction of 16.39 % and 15.70 %. The average cost function reduction at FI-Hyy was also similar to another evergreen needleleaf forest site, US-Wrc. However, the cost function reduction of FI-Hyy varied notably from year to year. In July and August 2014, the cost function reduction were as high as 40.59 % and 50.94 %, while in other years, the cost function reduction are much lower, ranging from 5.73 % to 18.94 %. Similar to the single-site twin experiments, only five parameters have been efficiently adjusted (**Table 2**).

The mean diurnal cycle and the scatterplots of observed and simulated COS fluxes are presented in **Figure 3** and **Figure S1**, respectively. Results show that the prior simulations can accurately reflect the magnitude of ecosystem COS fluxes and effectively capture the daily variation and the diurnal cycle of COS. On average across all sites, the prior simulated and observed ecosystem COS fluxes were remarkably close, with 20.60 $\mathrm{pmol\ m^{-2}\ s^{-1}}$ and 21.01 $\mathrm{pmol\ m^{-2}\ s^{-1}}$ respectively. However, there was substantial variability between sites and even between experiments at the same site. At ES-Lma, the prior

simulated COS fluxes were greatly underestimated by 63.38 %. In contrast, the prior simulated COS fluxes were overestimated at US-Ha1, with MBs of -10.01 pmol m^{-2} s^{-1} and -13.63 pmol m^{-2} s^{-1} in July 2012 and July 2013. In general, the MBs of COS fluxes are largely determined by the simulations and observations at daytime due to the larger magnitude (**Figure 3**). However, the model-observation differences at nighttime are also non-negligible. As shown in **Figure 3**, the underestimation is particularly evident at AT-Neu, ES-Lma and FI-Hyy.

After the single-site optimizations, both the daily variation and diurnal cycle of COS simulations were improved. This was reflected in the reduction of mean RMSE between the simulated and the observed COS fluxes from 16.49 pmol m^{-2} s^{-1} in the prior case to 13.86 pmol m^{-2} s^{-1} in the posterior case. Similar to the cost function, the RMSEs were also reduced in all single-site experiments. Moreover, the assimilation of COS observations also effectively corrected the bias between prior simulations and observations, with mean absolute MB decreased from 6.94 pmol m^{-2} s^{-1} to 3.09 pmol m^{-2} s^{-1}. In contrast, R^2 remained almost unchanged by the optimizations, with its mean value of 0.2967 in the prior case and 0.2970 in the posterior case. Our results also showcase that the model-observation differences of COS were effectively reduced at daytime. However, the remarkable differences between COS observations and simulations at nighttime, are not effectively corrected in a number of assimilation experiments (i.e., the experiment conducted at FI-Hyy in July 2013, see **Figure 3d**).

3.3 Two-site assimilation

FI-Hyy and US-Wrc have different soil textures, with sandy loam and loam, respectively. In the two-site assimilation experiment, NUCAS took this difference into account and successfully minimized the cost function from 495.94 to 365.63 after 67 evaluations of cost function. The cost function reduction for the experiment has a value of 28.29 %, comparable to the cost function reductions for corresponding single-site assimilation experiments at FI-Hyy and US-Wrc (50.94 % and 27.71 %). Furthermore, corresponding to these two soil textures, the texture-dependent parameters $Ksat_{scalar}$ and b_{scalar} yielded two different posterior parameter values, respectively, so that a total of seven parameters were optimized in the two-site experiment (**Table 3**). It can be seen that the two-site optimized results of V_{cmax25}, VJ_slope and f_leaf are similar to that of the single-site optimized results at US-Wrc, as most of the observations of the two-site experiment originated from US-Wrc. As for the texture-dependent parameters, they had the same signs and comparable magnitudes of the adjustments to that of the corresponding single-site experiment at FI-Hyy and were minutely adjusted at US-Wrc as in the corresponding single-site experiment. Overall, both the cost function reduction and the parameter optimization results of the two-site assimilation experiments were similar to the corresponding single-site experiments, demonstrating the ability of NUCAS to correctly perform joint data assimilation from COS observations at two sites simultaneously.

The posterior simulations of COS flux using the two-site posterior parameters, also demonstrated the ability of NUCAS to correctly assimilate two-site COS fluxes simultaneously (**Figure 4** and **Figure S2**). As shown in **Figure 4**, the prior COS simulations for both the FI-Hyy site and US-Wrc site were overestimated in the daytime compared to the observations. After the two-site COS assimilation, the discrepancies between COS simulations and observations were reduced in both FI-Hyy and US-Wrc, with RMSE reductions of 24.75 % and 3.39 %, achieving similar results to the simulations using the single-site posterior parameters.

3.4 Parameter change

As mentioned before, there were five parameters that have been adjusted during the assimilation of COS flux observations by the NUCAS system, whether in twin, single-site or two-site experiments. They are the maximum carboxylation rate at 25 ℃ (V_{cmax25}), the ratio of V_{cmax} to maximum electron transport rate J_{max} (VJ_slope), the scaling factor ($Ksat_{scalar}$ and b_{scalar}) of saturated hydraulic conductivity (Ksat) and Campbell parameter (b), and the ratio of PAR to shortwave radiation (f_leaf).

These parameters are strongly linked to the COS exchange processes and it is therefore reasonable that they could be optimized by the assimilation of COS flux. Furthermore, these parameters are also closely linked to processes such as photosynthesis, transpiration and soil water transport, and therefore the assimilation of COS flux provides an indirect constraint for improving the simulation of GPP, LE, H and soil moisture based on the assimilation of COS flux.

In both single-site and the two-site experiments, V_{cmax25} has been considerably adjusted, with average absolute relative change of 45.09 % and 41.36 %, respectively (**Figure 5a**). b_{scalar} and VJ_slope also varied greatly in the single-site experiments, with mean absolute relative changes of 30.92 % and 21.00 %, respectively. However, in the two-site experiment, their mean absolute changes were much smaller, at 4.08 % and 2.96 %. The relative changes of $Ksat_{scalar}$ are modest in both single-site and two-site experiments, with mean absolute values of 11.65 % and 9.34 %, respectively. As for f_leaf, the average absolute relative changes are even smaller than that of $Ksat_{scalar}$, at 3.67 % and 6.28 % in the single-site and the two-site experiments. In addition, we found that the parameters can be tuned considerably in cases where the prior simulations are close to the observations. For example, at IT-Soy, where the prior simulations agree well with the observations and the cost function only decrease 4.87 % in the experiment, both V_{cmax25} and b_{scalar} were remarkably tuned, with relative change of 32.55 % and -44.72 %.

Across all single-site experiments, there are notable differences in the results of parameter optimization, especially in V_{cmax25}. For the single-site experiment at US-Ha1 in July 2013, the posterior value of V_{cmax25} is 62.08 % lower than the prior. In contrast, the posterior V_{cmax25} is 127.80 % higher than the prior at ES-Lma. In addition to V_{cmax25}, The relative changes of b_{scalar} and VJ_slope also vary considerably, ranging from -78.13 % to 16.84 % and -58.23 % to 35.18 %, respectively. On the contrary, the posterior values of f_leaf show less variability, and do not differ from the prior value by more than 10.05%.

3.5 Parameter sensitivity

The adjoint-based sensitivity analysis results of the parameters are illustrated in **Figure 5b**. Our results suggest that V_{cmax25} has a critical impact on the assimilation results, followed by VJ_slope. With absolute SIs ranging from 88.47 % to 96.41 %, the mean absolute SI of V_{cmax25} is more than three times that of VJ_slope, which are 27.67 %. In contrast, the average absolute SIs of b_{scalar}, f_leaf and $Ksat_{scalar}$ are much lower, with 11.13 %, 8.30 % and 2.96 % respectively.

Unlike the great variability of the posterior V_{cmax25} and VJ_slope, the SIs of these two parameters are stable, especially at the same site. At US-Ha1, for example, the difference between the SIs of V_{cmax25} and VJ_slope in its two experiments were all smaller than 0.54 %. Furthermore, V_{cmax25} has the smallest magnitude of variation in SIs among the five parameters with the standard deviation of the SIs of 2.25 %, despite its SIs are of a much larger order of magnitude. With the SIs ranging from 20.62 % to 33.78 % and 4.17 % to 11.99 % (with the exception of DK-Sor), VJ_slope and f_leaf also play important roles in the modelling of COS. As for $Ksat_{scalar}$ and b_{scalar}, their SIs varied considerably across sites and even across experiments at the same site. For example, the absolute SIs of b_{scalar} are as high as 30.80 % and 34.04 % at the C3 grass sites AT-Neu and ES-Lma. On the contrary, the mean absolute SI of b_{scalar} is only 1.95 % at FI-Hyy. Yet, the absolute SIs of b_{scalar} of FI-Hyy varies considerably across the experiments, ranging from 0.07 % to 7.99 %.

Our results also suggest that f_leaf tends to play a more important role in the COS assimilation at the forest sites (except DK-Sor) compared to the low-stature vegetation type sites (including AT-Neu, ES-Lma and IT-Soy), with the mean absolute SIs about two times than that of the latter. With the absolute SIs ranging from 93.00 % to 96.41 %, V_{cmax25} is also observed to be more sensitive at the forest sites. Specifically, the largest SI of V_{cmax25} was observed at DK-Sor, while the SIs of VJ_slope and f_leaf of DK-Sor are noticeably lower than that of other sites, at 12.05 % and 0.94 %, respectively.

3.6 Comparison and evaluation of simulated GPP

For single-site experiments, both the prior and posterior GPP simulations performed well in modelling the daily variation and diurnal cycle of GPP, with mean R^2 of 0.80 and 0.78, respectively (**Figure 6** and **Figure S3**). The discrepancy between simulations and observations was substantially reduced by the assimilation of COS, from mean RMSE of 7.43 $\mu mol\ m^{-2}\ s^{-1}$ in the prior case to 5.34 $\mu mol\ m^{-2}\ s^{-1}$ in the posterior case. Similar to COS, the mean of prior simulated GPP is also generally larger than the observed. With the assimilation of COS, the bias between the observed and simulated GPP was effectively corrected, with the reduction in mean absolute MB from 4.31 $\mu mol\ m^{-2}\ s^{-1}$ to 2.28 $\mu mol\ m^{-2}\ s^{-1}$.

In general, the GPP performance was improved for most of the single-site experiments (12 of 13), with RMSE reductions ranging from 3.81 % to 64.27 %. Across all single-site experiments performed at evergreen needleleaf forest sites, the posterior GPP simulations were remarkably improved, with an averaged RMSE reduction of 42.00 %. At the deciduous broadleaf forest sites (DK-Sor and US-Ha1), the posterior simulated GPP also achieved a better fit with the GPP derived from EC observations, with an averaged RMSE reduction of 20.95 %. However, for experiments conducted on low-stature vegetation types (including C3 grass and C3 crop), the assimilation of COS is less effective in constraining the modelled GPP. At ES-Lma and IT-Soy, the RMSEs of the posterior simulated GPP are slightly lower than the prior, with reduction ratios of 8.60 % and 3.81 %, respectively. At AT-Neu, the addition of COS observation shifted the GPP simulations away from the GPP derived from EC observations, with the RMSE increasing from 3.48 $\mu mol\ m^{-2}\ s^{-1}$ to 5.97 $\mu mol\ m^{-2}\ s^{-1}$.

Covering different years or months, the single-site experiments performed at FI-Hyy and US-Ha1 provided an opportunity to analyze inter-annual and seasonal variation in the simulated and observed GPP. At US-Ha1, the prior simulations overestimated GPP in both July 2012 and July 2013, by 21.26 % and 42.02 % respectively. With the assimilation of COS, the modelled COS exhibited substantial decreases. In parallel, the model-observation difference also reduced, by 12.36 % and 24.46 %, respectively. However, the posterior simulated GPP appeared to be underestimated. At FI-Hyy, a total of six single-site experiments were conducted between 2013 and 2017, five of them in July and one in August 2014. The observed GPP shows little inter-annual variation in July from 2013 to 2017, with the mean ranging from 8.30 $\mu mol\ m^{-2}\ s^{-1}$ to 9.15 $\mu mol\ m^{-2}\ s^{-1}$. In August 2014, the GPP observations were noticeably lower than that in July, with a mean of 6.43 $\mu mol\ m^{-2}\ s^{-1}$. As for simulations, the model tends to overestimate GPP, with MBs ranging from 2.79 $\mu mol\ m^{-2}\ s^{-1}$ to 5.25 $\mu mol\ m^{-2}\ s^{-1}$. After the assimilation of COS, the overestimation of the COS simulation for FI-Hyy were effectively corrected, with the mean absolute MBs of 1.01 $\mu mol\ m^{-2}\ s^{-1}$. However, with a low SWC in August 2014, the prior simulated COS were obviously overestimated by 41.06 %, which led to remarkable downward adjustments of V_{cmax25} as well as VJ_slope. Thus, the simulated GPP were also markedly downgraded by 53.54 % in August 2014, ultimately resulting in the underestimation of the single-site posterior simulated GPP (**Figure 6f**).

In the two-site experiment, the model-observation differences of GPP for both FI-Hyy and US-Wrc were reduced by the assimilation of COS, with RMSE reductions of 39.90 % and 42.69 %, respectively. These RMSE reductions are even higher than those in the corresponding single-site experiments, by 55.08 % for FI-Hyy and 16.31 % for US-Wrc. These results suggest that simultaneous assimilation using COS observations from two sites can also improve GPP simulations, and the assimilation can be more robust than the single-site assimilation because the possibility of over-fit local noise is reduced.

Overall, the assimilation of ecosystem COS flux data improved the simulation of GPP in both single-site experiments and the two-site experiment. However, the assimilation effects vary considerably for different sites and even for different periods within the same site. Our results suggest the assimilation of COS is able to provide strong constrain to the modelling of GPP at forest sites, with an average RMSE reduction of 36.62 %. In contrast, at the low-stature vegetation type (including C3 grass and C3 crop) sites, the assimilation of COS is less effective in constraining the GPP simulations.

3.7 Comparison and evaluation of simulated LE and H

In order to verify the impact of COS assimilation on stomatal conductance and energy balance, observations of LE and H were compared to the prior and posterior model outputs. Results showed that the assimilation of COS is generally able to improve both latent and sensible heat, whether in single-site experiments or the two-site experiment (**Figure S4-S7**). The assimilation is more effective in reducing the model-observation difference of LE, with the average RMSE decreasing from 89.55 W m^{-2} to 73.94 W m^{-2}, while for H, the average RMSE only decreased from 103.10 W m^{-2} to 98.02 W m^{-2}. However, the average $R^2$of the simulated H increased noticeably from 0.39 in the prior case to 0.46 in the posteriori case, while that of LE slighted decreased from 0.65 to 0.64.

Results show that the BEPS model can simulate the daily variations of LE and H as well as the diurnal cycle of LE well, while the diurnal cycle of H is relatively poorly simulated. The prior simulation tends to overestimate LE during the daytime, and to exhibit short-time fluctuations in H that is not present in the observations. On average across all experiments, the prior simulated LE is overestimated by 31.60 W m^{-2} while the prior simulated H is underestimated by 37.28 W m^{-2}. The overestimation of LE and the underestimation of H are particularly apparent at the evergreen needleleaf forest sites (FI-Hyy and US-Wrc). At FI-Hyy and US-Wrc, the model-observation biases are more pronounced for H, with an averaged MB of -66.36 W m^{-2} than for LE with the averaged MB of 51.09 W m^{-2}. For the deciduous broadleaf forest sites DK-Sor and US-Ha1, the prior simulations of H both fit well with the observations, with a maximum absolute MB of only 17.88 W m^{-2}. However, the prior simulations tend to overestimate LE at US-Ha1, with a mean MB of 47.18 W m^{-2}.

In general, the single-site assimilation of COS effectively corrected the biases in the prior simulations of H and LE, and the correction are primarily reflected at daytime. Moreover, the correction was particularly effective for the evergreen needleleaf forest sites. On average across the ENF sites, the overestimation of LE and the underestimation of H were effectively corrected through the assimilation of COS, by 19.71 W m^{-2} and 18.38 W m^{-2}, respectively. At the DBF site US-Ha1, the simulation of LE increased by 38.07 W m^{-2} after the assimilation of COS, which considerably corrected the overestimation of the prior LE simulation. In contrast, the modelled H decreased by an average of 37.56 W m^{-2}, and deviated from the H observations in July 2013.

At US-Wrc, the two-site assimilation of COS effectively corrected the overestimation of LE and the underestimation of H in the prior simulations, with RMSE reductions of 17.58 % for LE and 22.33 % for H, which is even larger than that of the single-site optimization and confirms the robustness of the two-site assimilation. Similar to US-Wrc, the LE and H simulations obtained with the two-site posterior parameters are also superior to the prior simulations at FI-Hyy, with the RMSE reductions of 19.34 % for LE and 5.90 % for H.

Overall, the BEPS model performed well in simulating the daily variations and diurnal cycle of LE and H, while it tended to overestimate LE during the daytime and underestimate H around midday and sunset. Generally, the assimilation of COS could effectively improve the simulation of LE and H, whether the assimilation was conducted at single site or at two sites simultaneously, and this improvement was particularly noticeable for LE. We also observed that the simulated LE was always adjusted in the same direction as the COS, while H was adjusted in the opposite direction.

3.8 Comparison and evaluation of simulated SWC

The influence of COS assimilation on the modelling of SWC was assessed by comparing hourly SWC observations with hourly simulations of SWC. The assessments were carried out at all sites except US-Ha1, where no soil water observations were available. Results show the impact of COS assimilation on the modelling of SWC varies considerably by site and by period at the same site (**Figure S8**). Our results also suggested that the assimilation of COS is able to improve the simulation of SWC and this improvement is closely linked to the improved simulation of LE. However, with the considerable adjustment of soil

hydrology related parameters, the posterior simulated SWC also deviated noticeably from observations at several sites, i.e., AT-Neu.

Results show that the model can roughly follow the soil moisture trend (**Figure S8**). However, the simulated SWC exhibited a clear diurnal cycle whereas the observed SWC had almost no diurnal fluctuations. In response to the overestimation of LE at the ENF sites, the prior simulations underestimated the SWC in most (6/7) of the single-site experiments conducted at ENF sites. As the overestimation of LE was effectively corrected by the assimilation of COS, the decline in soil moisture slowed down, leading to the posterior SWC simulation being higher than the prior in the majority (6/7) of experiments. This conclusion was confirmed by the experiment at FI-Hyy in July 2015, in which the soil hydrology-related parameters $Ksat_{scalar}$ and b_{scalar} were adjusted as low as -0.0026 % and -0.0717 %, respectively. On the contrary, the soil hydrology-related parameters were considerably adjusted in the single-site experiment at FI-Hyy in July 2016, with relative changes of 18.13 % and -69.86 % for $Ksat_{scalar}$ and b_{scalar}, respectively. As a result, the corresponding posterior soil moisture simulations declined rapidly and deviated markedly from observations. Similar adjustment results for soil hydrology-related parameters were also observed at the C3 grass sites (AT-Neu and ES-Lma), with mean relative changes in $Ksat_{scalar}$ and b_{scalar} at these two sites of 26.32 % and 71.73 %, respectively. Accordingly, the posterior SWC simulations also show rapid declines and deviated from observations.

4 Discussion

4.1 Parameter changes

As mentioned before, our results show V_{cmax25} was tuned the most in both the single-site experiments and the two-site experiments, with the mean absolute relative change of 44.59 %, followed by b_{scalar} and VJ_slope. This is because COS plant fluxes are much larger than COS fluxes of soil in general (Whelan et al., 2016; Whelan et al., 2018) and the soil hydrology-related parameters cannot directly influence the COS plant uptake. Therefore, the assimilation of the COS flux mainly changed the parameters related to COS plant uptake rather than texture-dependent parameters that relate to soil COS flux to minimize the cost function. However, the adjustment of soil hydrology related parameters should not be neglected as well, as they play an important role in minimizing the discrepancy between COS simulations and observations.

As shown in **Figure 3**, the prior simulations underestimated COS fluxes at nighttime for many sites, i.e., FI-Hyy. On the one hand, this is due to the substantial gap between current modelled COS soil fluxes and observations (Whelan et al., 2022). On the other hand, this also stems from the fact that the nighttime stomatal conductance was set to a low and constant value (1 $\mathrm{mmol\,m^{-2}\,s^{-1}}$) in the BEPS model. As a result, the discrepancy between nighttime ecosystem COS simulations and observations could not be reduced by adjusting photosynthesis-related parameters to have an effect on stomatal conductance modelling. Thus, soil hydrology-related parameters were adjusted to compensate for the differences in both soil and plant components simultaneously. In this study, the COS soil model proposed by Whelan et al. (2016) and Whelan et al. (2022) was utilized, in which the optimal SWC for soil COS biotic uptake was set to 12.5 (%) for both grass and needleleaf forest. Such an optimal SWC value is much lower than the prior simulated SWC, as shown in **Figure S8**. Therefore, the soil hydrology-related parameters were considerably tuned, resulting in a rapid decline in the posterior SWC simulation to a level comparable to the optimum SWC.

COS plant uptake is governed by the reaction of COS destruction (Wohlfahrt et al., 2012) by carbonic anhydrase though it can also be destroyed by other photosynthetic enzymes, e.g., RuBisCo (Lorimer and Pierce, 1989), and the reaction is not dependent on light (Stimler et al., 2011; Whelan et al., 2018). Yet, given that stomatal conductance is simulated from net photosynthetic rate with a modified version (Woodward et al., 1995; Ju et al., 2010) of the Ball-Woodrow-Berry (BWB) model

(Ball et al., 1987) in BEPS, the adjustment of light reaction related parameters (VJ_slope and f_leaf) can therefore indirectly affect the simulation of COS plant uptake by influencing the calculation of stomatal conductance. According to Ryu et al. (2018), f_leaf varies little in reality and is usually between 41 % and 53 % on an annual mean scale. In our assimilation experiments, the optimized f_leaf values were distributed between 42.50 % and 51.28 %, consistent with this study. In contrast, the other light reaction related parameter VJ_slope, has a much wider range of variation, with relative changes ranging from -58.23 % to 35.18 %.

We noticed remarkably different optimization results for photosynthesis-related parameters in the experiments conducted in July 2015 and July 2017 at FI-Hyy, especially for V_{cmax25} and VJ_slope. In these two experiments, the difference in the relative change in V_{cmax25} is more than 20%, and that in VJ_slope is as high as 37.04%. However, these different adjustments to the parameter set caused similar impact on COS simulations, leading to the latter being reduced by 12.51 % and 10.43 % in July 2015 and July 2017, respectively. These results revealed the 'equifinality' (Beven, 1993) of the inversion problem at hand, i.e. the fact that different combinations of parameter values can achieve a similar fit to the COS observations. Assimilation of further observational data streams is expected to reduce the level of equifinality by differentiating between such combinations of parameter values that achieve a similar fit to COS observations.

4.2 Parameter sensitivity

It has been proved that photosynthetic capacity simulated by terrestrial ecosystem models is highly sensitive to V_{cmax}, J_{max}, and light conditions (Zaehle et al., 2005; Bonan et al., 2011; Rogers, 2014; Sargsyan et al., 2014; Koffi et al., 2015; Rogers et al., 2017). Therefore, it is expected that V_{cmax25}, VJ_slope, and f_leaf would markedly affect the optimization results, as these parameters ultimately have an impact on the simulation of plant COS uptake by influencing the estimation of photosynthesis capacity and stomatal conductance. Specifically, results of Wang et al. (2004), Verbeeck et al. (2006), Staudt et al. (2010), Han et al. (2020) and Ma et al. (2022) showed that the simulated photosynthetic capacity was generally more sensitive to J_{max} and light conditions than to V_{cmax}. However, due to the differences in the physiological mechanisms of COS plant uptake and photosynthesis, e.g., the hydrolysis reaction of COS by carbonic anhydrase is not dependent on light, the sensitivities of the two processes with respect to the model parameters may differ considerably although they are tightly coupled. Indeed, our adjoint sensitivity results suggest that the same change of V_{cmax25} is capable to influence the assimilation results to a greater extent than of VJ_slope and f_leaf. This result can be attributed to the model structure that V_{cmax25} not only affects the estimation of stomatal conductance through photosynthesis, but is also used to characterize mesophyll conductance and CA activity due to their linear relationships with V_{cmax} (Badger and Price, 1994; Evans et al., 1994; Berry et al., 2013). In addition, such a large sensitivity of V_{cmax25} also indicates the importance of accurate modelling of the apparent conductance of COS for ecosystem COS flux simulation.

As for $Ksat_{scalar}$ and b_{scalar}, they also play an important role in the assimilation of COS since the SWC simulations of BEPS are sensitive to Ksat and b (Liu et al., 2011), and SWC is the primary factor for COS soil biotic flux modelling (Whelan et al., 2016). However, as the soil COS exchange is generally much smaller than COS plant uptake (Whelan et al., 2018) and the parameter scheme provided by Whelan et al. (2022) sets different empirical parameter values (See **Table S3** for details) depending on the PFTs, the SIs of $Ksat_{scalar}$ and b_{scalar} differs considerably across PFTs, and are overall lower than those of photosynthesis related parameters.

In Sect 3.5, we mentioned that the light reaction related parameter f_leaf tend to play more essential roles in the assimilation of COS at the forest sites. Actually, similar features were found in the sensitivity of photosynthesis to radiation, i.e. the simulated GPP was more sensitive to radiation at forested vegetation types and less sensitive at low-stature vegetation types

(Sun et al., 2019). Particularly, the simulated GPP was also found to be highly sensitive to variations of radiation at low radiation conditions (Koffi et al., 2015).

4.3 Impacts of COS assimilation on ecosystem carbon, energy and water cycles

Due to the physiological basis that COS is taken up by plants through the same pathway of stomatal diffusion as CO_2, the assimilation of COS was expected to optimize the simulation of GPP. It was confirmed by our single-site and the two-site experiments conducted in a variety of ecosystems, including evergreen needleleaf forest, deciduous broadleaf forest, C3 grass and C3 crop. However, limited by many factors, such as the observation errors of the COS fluxes, the assimilation of COS does not always improve the simulation of GPP, i.e., at AT-Neu site.

Similar to the photosynthesis, the transpiration is also coupled with the COS plant uptake through stomatal conductance. But the difference is that after CO_2 is transported to the chloroplast surface, it continues its journey inside the chloroplast, and is eventually assimilated in the Calvin cycle (Wohlfahrt et al., 2012; Kohonen et al., 2022). Based on the BWB model, photosynthesis-related parameters only indirectly influence the calculation of stomatal conductance through photosynthesis in our model. Thus, the transpiration related variable LE, was not optimized as dramatically as GPP in the assimilation of COS. In comparison, the RMSEs of GPP simulations were reduced by an average of 25.37 % within the assimilation of COS, while that of LE were reduced by 16.27 %. Moreover, as transpiration rate and leaf temperature change show a linear relationship (Kümmerlen et al., 1999; Prytz et al., 2003) and surface-air temperature difference is a key control factor for sensible heat fluxes (Campbell and Norman, 2000; Arya, 2001; Jiang et al., 2022), the optimization for transpiration can therefore improve the simulation of leaf temperature and consequently improve the simulation of sensible heat flux.

Driven by the difference in water potential between the atmosphere and the substomatal cavity (Manzoni et al., 2013), the water is taken up by the roots, flows through the xylem, and exits through the leaf stomata to the atmosphere in the soil-plant-atmosphere continuum (Daly et al., 2004). Thus, when plants transpire, the water potential next to the roots decreases, driving water from bulk soil towards roots (Carminati et al., 2010) and reducing soil moisture. Certainly, soil moisture dynamics are also influenced by soil evaporation and leakage during inter-storm periods under ideal conditions (Daly et al., 2004). However, studies have shown that transpiration represents 80 to 90 percent of terrestrial evapotranspiration (Jasechko et al., 2013) and evaporation is typically a small fraction of transpiration for well-vegetated ecosystems (Scholes and Walker, 1993; Daly et al., 2004). Based on current knowledge of leakage, for example the relationship between leakage and the behavior of hydraulic conductivity (Clapp and Hornberger, 1978), extremely small adjustments of Ksat and b, i.e., with relative changes of -0.0026 % for $Ksat_{scalar}$ and -0.0717 % for b_{scalar} in July 2015 at FI-Hyy, hardly caused any change in leakage. Therefore, our results indicate that the assimilation of COS not only can markedly improve the modelling of stomatal conductance and transpiration, but it can also ultimately improve SWC predictions. However, our results also show that there are remarkable discrepancies between the ecosystem COS flux simulations and observations, and that discrepancies cannot be effectively reduced by the adjustment by the photosynthesis related parameters duo to the simplification of BPES for nighttime stomatal conductance modelling. As a result, it was also observed that the soil hydrology related parameters were drastically adjusted to minimize the discrepancy of COS simulations and observations, which instead biased the SWC simulations away from observations.

4.4 Impacts of leaf area index data on parameter optimization

As an essential input data of the BEPS model, LAI products have been demonstrated to be a source of uncertainty in the simulation of carbon and water fluxes (Liu et al., 2018). Therefore, it is necessary to investigate the influence of LAI on our parameter optimization results, as the LAI is directly related to the simulation of COS and the discrepancy between COS simulations and COS observations is an essential part of the cost function. Here we collected three widely used satellite-derived

LAI products (GLOBMAP, GLASS and MODIS) and the means of *in situ* LAI during the growing seasons or during the COS measurement periods for these sites (see **Table 1**). These *in situ* LAI means were used to drive the BEPS model along with the other three satellite-derived LAI products, with the assumption that they are representative of the LAI values during the assimilation periods. The configurations of those assimilation experiments were the same as those listed in **Table 2**, so that a total of 52 single-site experiments were conducted. All experiments were successfully performed, and the results were shown in **Figure 7** and **Figure S9.**

We found that the posterior V_{cmax25} significantly correlated with the LAI (R^2 = 0.17, P < 0.01) whilst there was no apparent relationship between the optimization results of the other three parameters and the LAI. As mentioned before, the LAI is directly related to the simulation of COS and thus influences the optimal values of the parameters. Therefore, the correlations of LAI with these parameters reflects the robustness of the constraint abilities of COS assimilation with respect to them. These results suggest that the assimilation of COS is able to provide strong constraints on V_{cmax25}, while it constrains other parameters (VJ_slope, $Ksat_{scalar}$, b_{scalar}, f_leaf) weakly, although they also considerably changed by the assimilation. In conclusion, our results suggest that the uncertainty in satellite-derived LAI not only can exert large impacts on the modelling of water-carbon fluxes (Wang et al., 2021), but also is an important source of the uncertainty in the parameter optimization results when performing data assimilation experiments with ecosystem models driven by LAI.

4.5 Caveats and implications

In general, we found that the assimilation of COS can improve the model performance for GPP, LE and H for both single-site assimilation and two-site assimilation. Nonetheless, there are currently limitations that affect the use of COS data for the optimization of parameters, processes and variables related to water-carbon cycling and energy exchange in terrestrial ecosystem models. For SWC, there is a mixed picture. Affected by the substantial downward adjustment of soil moisture to the optimal soil moisture at individual sites (i.e., AT-Neu), the RMSE of soil moisture simulations did not improve on average. However, in some experiments (especially those where soil hydrological parameters do not change much, such as the experiment conducted at FI-Hyy in July 2015), SWC simulations did improve with the assimilation of COS.

The assimilation of COS fluxes relies on the availability and quality of field observations. As both COS plant uptake and COS soil exchange are modelled within NUCAS and the data assimilation was performed at the ecosystem scale, a large number of accurate measurements of both COS soil flux and COS plant flux are essential for COS assimilation and model evaluation. However, at present, we face a serious lack of COS measurements (Brühl et al., 2012; Wohlfahrt et al., 2012). More laboratory and field measurements are needed for better understanding of mechanistic processes of COS. Besides, the existing COS flux data were calculated based on different measurement methods and data processing steps, which poses considerable challenges for comparing COS flux measurements across sites. Standardization of measurement and processing techniques of COS (Kohonen et al., 2020) is therefore urgently needed.

In this study, the prior uncertainty of observation was estimated by the standard deviation of ecosystem COS fluxes within 24 hours with the assumption of a normal distribution. However, Hollinger and Richardson (2005) suggested that flux measurement error more closely follows a double exponential than a normal distribution. Kohonen et al. (2020) showed that the overall uncertainty in the COS flux varies with the sign (uptake or release) as well as the magnitude of the COS flux. Furthermore, there is a lack of understanding of the prior uncertainty for certain model parameters, such as VJ_slope, which makes the uncertainty estimates subject to potentially large errors. In conclusion, we should be more careful in considering the distribution and the magnitude of the prior uncertainty of observations and parameters.

The spatial and temporal variation in atmospheric COS concentrations has a considerable influence on the COS plant uptake (Ma et al., 2021) due to the linear relationship between the two (Stimler et al., 2010). The typical seasonal amplitude of

atmospheric COS concentrations is ~ 100–200 parts per trillion (ppt) around an average of ~ 500 ppt (Montzka et al., 2007; Kooijmans et al., 2021; Hu et al., 2021; Ma et al., 2021; Belviso et al., 2022). However, in NUCAS, COS mole fractions in the bulk air are currently assumed to be spatially invariant over the globe and to vary annually, which may introduce substantial errors into the parameter calibration. Kooijmans et al. (2021) has confirmed that modifying the COS mole fractions to vary spatially and temporally markedly improved the simulation of ecosystem COS flux. Thus, we suggest to take into account the variation in COS concentration and their interaction with surface COS fluxes at high spatial and temporal resolution in order to achieve better parameter calibration.

Currently, there are still uncertainties in the simulation of COS fluxes by BEPS particularly for nighttime COS fluxes. As the nighttime COS plant uptake is driven by stomatal conductance (Kooijmans et al., 2021), the nighttime COS fluxes can therefore be used to test the accuracy of the model settings for nighttime stomatal conductance (g_n). In the BEPS model, a low and constant value (1mmol m^{-2} s^{-1}) of g_n was set for all PFTs. Our simulations of nighttime COS flux indicate that in BEPS, g_n is underestimated to different degrees for different sites. This result is also proved by Resco De Dios et al. (2019), which found that the median g_n in the global dataset was 40 mmol m^{-2} s^{-1}. Therefore, utilizing COS to directly optimize stomatal related parameters should be perused. Cho et al. (2023) has proven the effectiveness of optimizing the minimum stomatal conductance as well as other parameters by the assimilation of COS. Besides, with the argument that different enzymes have different physiological characteristics, Cho et al. (2023) proposed a new temperature function for the CA enzyme and showcase the considerate difference in temperature response of enzymatic activities of CA and RuBisCo enzyme, which also provided valuable insights into the modelling and assimilation of COS. In addition, soil COS exchange is an important source of uncertainty in the use of COS as carbon-water cycle tracer since carbonic anhydrase activity occurs in the soil as well (Kesselmeier et al., 1999; Smith et al., 1999; Ogée et al., 2016; Meredith et al., 2019). Kaisermann et al. (2018) showed that COS hydrolysis rates were linked to microbial C biomass, whilst COS production rates were linked to soil N content and mean annual precipitation (MAP). Interestingly, MAP was also suggested to be the best predictor of g_n in Yu et al. (2019), who found that plants in locations with lower rainfall conditions had higher g_n. Therefore, using the global microbial C biomass, soil N content and MAP datasets and the relationships between these variables and the associated COS exchange processes is expected to achieve more accurate modelling of terrestrial ecosystem COS fluxes, increase the understanding of the global COS budget and facilitate the assimilation of COS fluxes.

5 Conclusions

Over the past decades, considerable efforts have been made to obtain field observations of COS ecosystem fluxes and to describe empirically or mechanistically COS plant uptake and soil exchange, which offers the possibility of investigating the ability of assimilating ecosystem COS flux to optimize parameters and variables related to the water and carbon cycles and energy exchange. In this study, we introduced the NUCAS system, which has been developed based on the BEPS model and was designed to have the ability to assimilate ecosystem COS flux data. In NUCAS, a resistance analog model of COS plant uptake and an empirical model of soil COS flux were embedded in the BEPS model to achieve the simulation of ecosystem COS flux, and a gradient-based 4D-Var data assimilation algorithm was implemented to optimize the internal parameters of BEPS.

Fourteen twin experiments, thirteen single-site experiments and one two-site experiment covering the period from 2012 to 2017, were conducted to investigate the data assimilation capability and the optimization effect of parameters and variables of NUCAS for COS flux observations over a range of ecosystems that contains four PFTs and three soil textures. Our results show that NUCAS has the ability to optimize parameter vectors, and the assimilation of COS can constrain parameters affecting the simulation of carbon and water cycles and energy exchange and thus effectively improve the performance of the BEPS

model. We found that there is a tight link between the assimilation of COS and the optimization of LE, which demonstrates the role of COS as an indicator of stomatal conductance and transpiration. The improvement of transpiration can further improve the model performance for H and SWC, although the propagation of the optimization effect is subject to some limitations. These results highlight the broad perspective of COS as a tracer for improving the simulation of variables related to stomatal conductance. Furthermore, we demonstrated that COS can provide a strong constraint on V_{cmax25}, whereas the adjustment of parameters related to the soil hydrology appears to compensate for weaknesses in the model, i.e., the nighttime stomatal conductance set in BEPS model. We also proved the strong impact of LAI on the parameter optimization results, emphasizing the importance of developing more accurate LAI products for models driven by observed LAI. In addition, we made a number of recommendations for future improvement of the assimilation of COS. Particularly, we flagged the need for more observations of COS, suggested better characterisation of observational and prior parameter uncertainties, the use of varying COS concentrations and the refinement of the model for COS fluxes of soil. Specifically, with the lack of separate COS plant and soil flux data, the ecosystem-scale COS flux observations were utilized in this study. However, we believe that assimilating the component fluxes of COS individually should be pursued in the future as this assimilation approach would provide separate constraints on different parts of the model. We expect the observational information on the partitioning between the two flux components to provide a stronger constraint than using just their sum.

Our two-site setup constitutes a challenge for the assimilation system, the model and the observations. In this setup, the assimilation system has to determine a parameter set that achieves a fit to the observations at both sites, and NUCAS passes this important test. It should be noted that the NUCAS was designed as a platform that integrates multiple data streams to provide a consistent map of the terrestrial carbon cycle although only ecosystem COS flux data were used to evaluate the performance of NUCAS in this study. The "two-site" assimilation experiment conducted in this study gives us more confidence that the calibrated model will provide a reasonable parameter set and posterior simulation throughout the plant functional type. In other words, what we present here is a pre-requisite for applying the model and assimilation system at regional to global scales.

We noticed the optimization of model parameters faced the challenge of 'equifinality' due to the complexity of the model and the limited observation data. However, the 'equifinality' can be avoided by imposing additional observational constraints (Beven, 2006). Indeed, using several different data streams to simultaneously (Kaminski et al., 2012; Schürmann et al., 2016; Scholze et al., 2016; Wu et al., 2018; Scholze et al., 2019) or step-wise (Peylin et al., 2016) to constrain multiple processes in the carbon cycle is becoming a focus area in carbon cycle research. Therefore, it is necessary to combine COS with other observations to constrain different ecosystem processes and/or exploit multiple constraints on the same processes in order to achieve better modelling and prediction of the ecosystem water-carbon cycle and energy exchange.

Code availability. The source code for BEPS is publicly available at https://zenodo.org/doi/10.5281/zenodo.8288750, the adjoint code for BEPS is available upon request to the correspondence author (mousongwu@nju.edu.cn).

Data availability. Measured eddy covariance Carboy sulfide fluxes data can be found at https://zenodo.org/records/3993111 for AT-Neu, https://zenodo.org/record/3406990 for DK-Sor, ES-Lma and IT-Soy, https://zenodo.org/record/6940750 for FI-Hyy, and from the Harvard Forest Data Archive under record HF214 (https://portal.edirepository.org/nis/mapbrowse?packageid=knb-lter-hfr.214.4) for US-Ha1.The raw COS concentration data of US-Wrc can be obtained at https://zenodo.org/record/1422820. The meteorological data can be obtained from the FLUXNET database (https://fluxnet.org/) for AT-Neu, DK-Sor, ES-Lma, FI-Hyy and US-Ha1; from the AmeriFlux database (https://ameriflux.lbl.gov/) for US-Ha1 (except shortwave radiation data) and US-Wrc; from the ERA5 dataset

(https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview) for AT-Neu, IT-Soy and US-Ha1. The evaluation data can be obtained from the FLUXNET database for DK-Sor, ES-Lma, FI-Hyy and US-Ha1; from the AmeriFlux database for US-Ha1 and US-Wrc; from https://zenodo.org/records/3993111 for AT-Neu, from https://zenodo.org/record/6940750 for IT-Soy and from https://zenodo.org/record/1422820 for US-Wrc. The H and LE data of AT-Neu and IT-Soy are provided by Felix M. Spielmann and Georg Wohlfahrt. The GLOBMAP LAI is available at https://zenodo.org/record/4700264#.YzvSYnZBxD8%2F, the GLASS LAI is available at ftp://ftp.glcf.umd.edu/, and the MODIS LAI product is available at https://lpdaac.usgs.gov/products/mod15a2hv006/. All datasets used in this study and the model outputs are available upon request.

Author contributions: MW designed the experiments and developed the model, MV and TK developed the data assimilation layer including the adjoint code for the ecosystem model, HZ wrote the original manuscript and made the analysis. All the authors contributed to the writing of the manuscript.

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

Acknowledgements: This study was supported by the National Key Research and Development Program of China (2020YFA0607504, 2016YFA0600204), the National Natural Science Foundation of China (42141005, 41901266), the Research Funds for the Frontiers Science Center for Critical Earth Material Cycling, Nanjing University (Grant No: 090414380031). We thank Felix M. Spielmann and Georg Wohlfahrt for providing H and LE data for AT-Neu and IT-Soy. MV and TK thank Laurent Hascoët for supporting this activity. The authors thank two anonymous reviewers for highly valuable comments.

References

Abadie, C., Maignan, F., Remaud, M., Ogée, J., Campbell, J. E., Whelan, M. E., Kitz, F., Spielmann, F. M., Wohlfahrt, G., and Wehr, R.: Global modelling of soil carbonyl sulfide exchanges, Biogeosciences, 19, 2427-2463, 2022.

An, X. Q., Zhai, S. X., Jin, M., Gong, S., and Wang, Y.: Development of an adjoint model of GRAPES–CUACE and its application in tracking influential haze source areas in north China, Geoscientific Model Development, 9, 2153-2165, 2016.

Arias, P., Bellouin, N., Coppola, E., Jones, R., Krinner, G., Marotzke, J., Naik, V., Palmer, M., Plattner, G.-K., and Rogelj, J.: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; Technical Summary, 2021.

Arya, P. S.: Introduction to micrometeorology, Elsevier2001.

Bäck, J., Aalto, J., Henriksson, M., Hakola, H., He, Q., and Boy, M.: Chemodiversity of a Scots pine stand and implications for terpene air concentrations, Biogeosciences, 9, 689-702, 2012.

Badger, M. R. and Price, G. D.: The role of carbonic anhydrase in photosynthesis, Annual review of plant biology, 45, 369-392, 1994.

Ball, J. T., Woodrow, I. E., and Berry, J. A.: A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, Progress in photosynthesis research: volume 4 proceedings of the VIIth international congress on photosynthesis providence, Rhode Island, USA, august 10–15, 1986, 221-224,

Behrendt, T., Veres, P. R., Ashuri, F., Song, G., Flanz, M., Mamtimin, B., Bruse, M., Williams, J., and Meixner, F. X.: Characterisation of NO production and consumption: new insights by an improved laboratory dynamic chamber technique, Biogeosciences, 11, 5463-5492, 10.5194/bg-11-5463-2014, 2014.

Belviso, S., Remaud, M., Abadie, C., Maignan, F., Ramonet, M., and Peylin, P.: Ongoing Decline in the Atmospheric COS Seasonal Cycle Amplitude over Western Europe: Implications for Surface Fluxes, Atmosphere, 13, 812, 2022.

Berry, J., Wolf, A., Campbell, J. E., Baker, I., Blake, N., Blake, D., Denning, A. S., Kawa, S. R., Montzka, S. A., Seibt, U., Stimler, K., Yakir, D., and Zhu, Z.: A coupled model of the global cycles of carbonyl sulfide and CO2: A possible new window on the carbon cycle, Journal of Geophysical Research: Biogeosciences, 118, 842-852, https://doi.org/10.1002/jgrg.20068, 2013.

Berry, S. L., Farquhar, G. D., and Roderick, M. L.: Co ‐ evolution of climate, soil and vegetation, Encyclopedia of hydrological sciences, 2006.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Advances in water resources, 16, 41-51, 1993.
Beven, K.: A manifesto for the equifinality thesis, Journal of hydrology, 320, 18-36, 2006.
Bonan, G. B.: A biophysical surface energy budget analysis of soil temperature in the boreal forests of interior Alaska, Water Resources Research, 27, 767-781, 1991.
Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., and Swenson, S. C.: Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, Journal of Geophysical Research: Biogeosciences, 116, 2011.
Braendholt, A., Ibrom, A., Larsen, K. S., and Pilegaard, K.: Partitioning of ecosystem respiration in a beech forest, Agricultural and Forest Meteorology, 252, 88-98, 2018.
Brühl, C., Lelieveld, J., Crutzen, P., and Tost, H.: The role of carbonyl sulphide as a source of stratospheric sulphate aerosol and its impact on climate, Atmospheric Chemistry and Physics, 12, 1239-1253, 2012.
Campbell, G. S. and Norman, J. M.: An introduction to environmental biophysics, Springer Science & Business Media2000.
Campbell, J. E., Carmichael, G. R., Chai, T., Mena-Carrasco, M., Tang, Y., Blake, D., Blake, N., Vay, S. A., Collatz, G. J., and Baker, I.: Photosynthetic control of atmospheric carbonyl sulfide during the growing season, Science, 322, 1085-1088, 2008.
Carminati, A., Moradi, A. B., Vetterlein, D., Vontobel, P., Lehmann, E., Weller, U., Vogel, H.-J., and Oswald, S. E.: Dynamics of soil water content in the rhizosphere, Plant and soil, 332, 163-176, 2010.
Chen, J., Liu, J., Cihlar, J., and Goulden, M.: Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications, Ecological modelling, 124, 99-119, 1999.
Chen, J. M., Ju, W., Ciais, P., Viovy, N., Liu, R., Liu, Y., and Lu, X.: Vegetation structural change since 1981 significantly enhanced the terrestrial carbon sink, Nature communications, 10, 4259, 2019.
Chen, J. M., Mo, G., Pisek, J., Liu, J., Deng, F., Ishizawa, M., and Chan, D.: Effects of foliage clumping on the estimation of global terrestrial gross primary productivity, Global Biogeochemical Cycles, 26, 2012.
Chen, J. M., Wang, R., Liu, Y., He, L., Croft, H., Luo, X., Wang, H., Smith, N. G., Keenan, T. F., and Prentice, I. C.: Global datasets of leaf photosynthetic capacity for ecological and earth system research, Earth System Science Data, 14, 4077-4093, 2022.
Cho, A., Kooijmans, L. M., Kohonen, K.-M., Wehr, R., and Krol, M. C.: Optimizing the carbonic anhydrase temperature response and stomatal conductance of carbonyl sulfide leaf uptake in the Simple Biosphere model (SiB4), Biogeosciences, 20, 2573-2594, 2023.
Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water resources research, 14, 601-604, 1978.
Commane, R., Meredith, L. K., Baker, I. T., Berry, J. A., Munger, J. W., Montzka, S. A., Templer, P. H., Juice, S. M., Zahniser, M. S., and Wofsy, S. C.: Seasonal fluxes of carbonyl sulfide in a midlatitude forest, Proceedings of the National Academy of Sciences, 112, 14162-14167, 2015.
Daly, E., Porporato, A., and Rodriguez-Iturbe, I.: Coupled dynamics of photosynthesis, transpiration, and soil water balance. Part I: Upscaling from hourly to daily level, Journal of Hydrometeorology, 5, 546-558, 2004.
Deng, F., Jones, D., Henze, D., Bousserez, N., Bowman, K., Fisher, J., Nassar, R., O'Dell, C., Wunch, D., and Wennberg, P.: Inferring regional sources and sinks of atmospheric CO 2 from GOSAT XCO 2 data, Atmospheric Chemistry and Physics, 14, 3703-3727, 2014.
Dowd, M.: Bayesian statistical data assimilation for ecosystem models using Markov Chain Monte Carlo, Journal of Marine Systems, 68, 439-456, 2007.
El-Madany, T. S., Reichstein, M., Perez-Priego, O., Carrara, A., Moreno, G., Martín, M. P., Pacheco-Labrador, J., Wohlfahrt, G., Nieto, H., and Weber, U.: Drivers of spatio-temporal variability of carbon dioxide and energy fluxes in a Mediterranean savanna ecosystem, Agricultural and Forest Meteorology, 262, 258-278, 2018.
Evans, J. R., Caemmerer, S., Setchell, B. A., and Hudson, G. S.: The relationship between CO2 transfer conductance and leaf anatomy in transgenic tobacco with a reduced content of Rubisco, Functional Plant Biology, 21, 475-495, 1994.
Farquhar, G. D., von Caemmerer, S. v., and Berry, J. A.: A biochemical model of photosynthetic CO 2 assimilation in leaves of C 3 species, planta, 149, 78-90, 1980.
Fisher, J. B., Huntzinger, D. N., Schwalm, C. R., and Sitch, S.: Modeling the terrestrial biosphere, Annual Review of Environment and Resources, 39, 91-123, 2014.
Fisher, R. A. and Koven, C. D.: Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems, Journal of Advances in Modeling Earth Systems, 12, e2018MS001453, https://doi.org/10.1029/2018MS001453, 2020.
Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C., Hauck, J., Le Quéré, C., Peters, G. P., Peters, W., and Pongratz, J.: Global carbon budget 2021, Earth System Science Data, 14, 1917-2005, 2022.
Gates, D. M.: Transpiration and leaf temperature, Annual Review of Plant Physiology, 19, 211-238, 1968.
Goldan, P. D., Fall, R., Kuster, W. C., and Fehsenfeld, F. C.: Uptake of COS by growing vegetation: A major tropospheric sink, Journal of Geophysical Research: Atmospheres, 93, 14186-14192, 1988.
Grimm, N. B., Chapin III, F. S., Bierwagen, B., Gonzalez, P., Groffman, P. M., Luo, Y., Melton, F., Nadelhoffer, K., Pairis, A., and Raymond, P. A.: The impacts of climate change on ecosystem structure and function, Frontiers in Ecology and the Environment, 11, 474-482, 2013.

Gu, L., Baldocchi, D., Verma, S. B., Black, T., Vesala, T., Falge, E. M., and Dowty, P. R.: Advantages of diffuse radiation for terrestrial ecosystem productivity, Journal of Geophysical Research: Atmospheres, 107, ACL 2-1-ACL 2-23, 2002.
Han, T., Zhu, G., Ma, J., Wang, S., Zhang, K., Liu, X., Ma, T., Shang, S., and Huang, C.: Sensitivity analysis and estimation using a hierarchical Bayesian method for the parameters of the FvCB biochemical photosynthetic model, Photosynthesis research, 143, 45-66, 2020.
Hascoët, L. and Pascual, V.: The Tapenade automatic differentiation tool: Principles, model, and specification, ACM Trans. Math. Softw., 39, Article 20, 10.1145/2450153.2450158, 2013.
Haynes, K., Baker, I., and Denning, S.: Simple biosphere model version 4.2 (SiB4) technical description, Colorado State University: Fort Collins, CO, USA, 2020.
He, Q., Ju, W., Dai, S., He, W., Song, L., Wang, S., Li, X., and Mao, G.: Drought risk of global terrestrial gross primary productivity over the last 40 years detected by a remote sensing‐driven process model, Journal of Geophysical Research: Biogeosciences, 126, e2020JG005944, 2021.
Hollinger, D. and Richardson, A.: Uncertainty in eddy covariance measurements and its application to physiological models, Tree physiology, 25, 873-885, 2005.
Hörtnagl, L., Bamberger, I., Graus, M., Ruuskanen, T. M., Schnitzhofer, R., Müller, M., Hansel, A., and Wohlfahrt, G.: Biotic, abiotic, and management controls on methanol exchange above a temperate mountain grassland, Journal of Geophysical Research: Biogeosciences, 116, 2011.
Hu, L., Montzka, S. A., Kaushik, A., Andrews, A. E., Sweeney, C., Miller, J., Baker, I. T., Denning, S., Campbell, E., and Shiga, Y. P.: COS-derived GPP relationships with temperature and light help explain high-latitude atmospheric CO2 seasonal cycle amplification, Proceedings of the National Academy of Sciences, 118, e2103423118, 2021.
Irrgang, C., Saynisch, J., and Thomas, M.: Utilizing oceanic electromagnetic induction to constrain an ocean general circulation model: A data assimilation twin experiment, Journal of Advances in Modeling Earth Systems, 9, 1703-1720, 2017.
Jasechko, S., Sharp, Z. D., Gibson, J. J., Birks, S. J., Yi, Y., and Fawcett, P. J.: Terrestrial water fluxes dominated by transpiration, Nature, 496, 347-350, 2013.
Jiang, K., Pan, Z., Pan, F., Wang, J., Han, G., Song, Y., Zhang, Z., Huang, N., Ma, S., and Chen, X.: Influence patterns of soil moisture change on surface-air temperature difference under different climatic background, Science of the Total Environment, 822, 153607, 2022.
Ju, W., Gao, P., Wang, J., Zhou, Y., and Zhang, X.: Combining an ecological model with remote sensing and GIS techniques to monitor soil water content of croplands with a monsoon climate, Agricultural Water Management, 97, 1221-1231, 2010.
Ju, W., Chen, J. M., Black, T. A., Barr, A. G., Liu, J., and Chen, B.: Modelling multi-year coupled carbon and water fluxes in a boreal aspen forest, Agricultural and Forest Meteorology, 140, 136-151, 2006.
Kaisermann, A., Ogée, J., Sauze, J., Wohl, S., Jones, S. P., Gutierrez, A., and Wingate, L.: Disentangling the rates of carbonyl sulfide (COS) production and consumption and their dependency on soil properties across biomes and land use types, Atmospheric Chemistry and Physics, 18, 9425-9440, 2018.
Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R., and Mathieu, P.-P.: Consistent assimilation of MERIS FAPAR and atmospheric CO 2 into a terrestrial vegetation model and interactive mission benefit analysis, Biogeosciences, 9, 3173-3184, 2012.
Kato, T., Knorr, W., Scholze, M., Veenendaal, E., Kaminski, T., Kattge, J., and Gobron, N.: 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, 2013.
Kattge, J., Knorr, W., Raddatz, T., and Wirth, C.: Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global‐scale terrestrial biosphere models, Global Change Biology, 15, 976-991, 2009.
Keenan, T. F., Davidson, E., Moffat, A. M., Munger, W., and Richardson, A. D.: Using model‐data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling, Global Change Biology, 18, 2555-2569, 2012.
Kesselmeier, J., Teusch, N., and Kuhn, U.: Controlling variables for the uptake of atmospheric carbonyl sulfide by soil, Journal of Geophysical Research: Atmospheres, 104, 11577-11584, 1999.
Knorr, W. and Heimann, M.: Impact of drought stress and other factors on seasonal land biosphere CO2 exchange studied through an atmospheric tracer transport model, Tellus B, 47, 471-489, 1995.
Knorr, W., Kaminski, T., Scholze, M., Gobron, N., Pinty, B., Giering, R., and Mathieu, P. P.: Carbon cycle data assimilation with a generic phenology model, Journal of Geophysical Research: Biogeosciences, 115, 2010.
Koffi, E., Rayner, P., Norton, A., Frankenberg, C., and Scholze, M.: Investigating the usefulness of satellite-derived fluorescence data in inferring gross primary productivity within the carbon cycle data assimilation system, Biogeosciences, 12, 4067-4084, 2015.
Kohonen, K.-M., Kolari, P., Kooijmans, L. M., Chen, H., Seibt, U., Sun, W., and Mammarella, I.: Towards standardized processing of eddy covariance flux measurements of carbonyl sulfide, Atmospheric Measurement Techniques, 13, 3957-3975, 2020.
Kohonen, K.-M., Dewar, R., Tramontana, G., Mauranen, A., Kolari, P., Kooijmans, L. M., Papale, D., Vesala, T., and Mammarella, I.: Intercomparison of methods to estimate gross primary production based on CO 2 and COS flux measurements, Biogeosciences, 19, 4067-4088, 2022.

Konarska, J., Uddling, J., Holmer, B., Lutz, M., Lindberg, F., Pleijel, H., and Thorsson, S.: Transpiration of urban trees and its cooling effect in a high latitude city, International journal of biometeorology, 60, 159-172, 2016.
Kooijmans, L. M. J., Cho, A., Ma, J., Kaushik, A., Haynes, K. D., Baker, I., Luijkx, I. T., Groenink, M., Peters, W., Miller, J. B., Berry, J. A., Ogée, J., Meredith, L. K., Sun, W., Kohonen, K. M., Vesala, T., Mammarella, I., Chen, H., Spielmann, F. M., Wohlfahrt, G., Berkelhammer, M., Whelan, M. E., Maseyk, K., Seibt, U., Commane, R., Wehr, R., and Krol, M.: Evaluation of carbonyl sulfide biosphere exchange in the Simple Biosphere Model (SiB4), Biogeosciences, 18, 6547-6565, 10.5194/bg-18-6547-2021, 2021.
Kümmerlen, B., Dauwe, S., Schmundt, D., and Schurr, U.: Thermography to measure water relations of plant leaves, Handbook of computer vision and applications, 3, 763-781, 1999.
Lasslop, G., Migliavacca, M., Bohrer, G., Reichstein, M., Bahn, M., Ibrom, A., Jacobs, C., Kolari, P., Papale, D., and Vesala, T.: On the choice of the driving temperature for eddy-covariance carbon dioxide flux partitioning, Biogeosciences, 9, 5243-5259, 2012.
Launois, T., Peylin, P., Belviso, S., and Poulter, B.: A new model of the global biogeochemical cycle of carbonyl sulfide–Part 2: Use of carbonyl sulfide to constrain gross primary productivity in current vegetation models, Atmospheric Chemistry and Physics, 15, 9285-9312, 2015.
Law, K., Stuart, A., and Zygalakis, K.: Data assimilation, Cham, Switzerland: Springer, 214, 52, 2015.
Leung, L. R., Hamlet, A. F., Lettenmaier, D. P., and Kumar, A.: Simulations of the ENSO hydroclimate signals in the Pacific Northwest Columbia River basin, Bulletin of the American Meteorological Society, 80, 2313-2330, 1999.
Liu, J., Chen, J., and Cihlar, J.: Mapping evapotranspiration based on remote sensing: An application to Canada's landmass, Water resources research, 39, 2003.
Liu, J., Chen, J., Cihlar, J., and Park, W.: A process-based boreal ecosystem productivity simulator using remote sensing inputs, Remote sensing of environment, 62, 158-175, 1997.
Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water resources research, 43, 2007.
Liu, Y., Liu, R., and Chen, J. M.: Retrospective retrieval of long‐term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data, Journal of Geophysical Research: Biogeosciences, 117, 2012.
Liu, Y., Xiao, J., Ju, W., Zhou, Y., Wang, S., and Wu, X.: Water use efficiency of China's terrestrial ecosystems and responses to drought, Scientific reports, 5, 13799, 2015.
Liu, Y., Xiao, J., Ju, W., Zhu, G., Wu, X., Fan, W., Li, D., and Zhou, Y.: 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, 2018.
Liu, Z., Zhou, Y., Ju, W., and Gao, P.: Simulation of soil water content in farm lands with the BEPS ecological model, Transactions of the Chinese Society of Agricultural Engineering, 27, 67-72, 2011.
Lloyd, J. and Taylor, J.: On the temperature dependence of soil respiration, Functional ecology, 315-323, 1994.
Lorimer, G. and Pierce, J.: Carbonyl sulfide: an alternate substrate for but not an activator of ribulose-1, 5-bisphosphate carboxylase, Journal of Biological Chemistry, 264, 2764-2772, 1989.
Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., Clark, J. S., and Schimel, D. S.: Ecological forecasting and data assimilation in a data‐rich era, Ecological Applications, 21, 1429-1442, 2011.
Ma, J., Kooijmans, L. M., Cho, A., Montzka, S. A., Glatthor, N., Worden, J. R., Kuai, L., Atlas, E. L., and Krol, M. C.: Inverse modelling of carbonyl sulfide: implementation, evaluation and implications for the global budget, Atmospheric Chemistry and Physics, 21, 3507-3529, 2021.
Ma, R., Xiao, J., Liang, S., Ma, H., He, T., Guo, D., Liu, X., and Lu, H.: Pixel-level parameter optimization of a terrestrial biosphere model for improving estimation of carbon fluxes with an efficient model–data fusion method and satellite-derived LAI and GPP data, Geoscientific Model Development, 15, 6637-6657, 2022.
MacBean, N., Bacour, C., Raoult, N., Bastrikov, V., Koffi, E., Kuppel, S., Maignan, F., Ottlé, C., Peaucelle, M., and Santaren, D.: Quantifying and reducing uncertainty in global carbon cycle predictions: lessons and perspectives from 15 years of data assimilation studies with the ORCHIDEE Terrestrial Biosphere Model, Global Biogeochemical Cycles, 36, e2021GB007177, 2022.
Manzoni, S., Vico, G., Katul, G., Palmroth, S., Jackson, R. B., and Porporato, A.: Hydraulic limits on maximum plant transpiration and the emergence of the safety–efficiency trade‐off, New Phytologist, 198, 169-178, 2013.
Medlyn, B. E., Badeck, F. W., De Pury, D., Barton, C., Broadmeadow, M., Ceulemans, R., De Angelis, P., Forstreuter, M., Jach, M., and Kellomäki, S.: Effects of elevated [CO2] on photosynthesis in European forest species: a meta‐analysis of model parameters, Plant, Cell & Environment, 22, 1475-1495, 1999.
Meredith, L. K., Ogée, J., Boye, K., Singer, E., Wingate, L., von Sperber, C., Sengupta, A., Whelan, M., Pang, E., and Keiluweit, M.: Soil exchange rates of COS and CO18O differ with the diversity of microbial communities and their carbonic anhydrase enzymes, The ISME journal, 13, 290-300, 2019.
Montzka, S., Calvert, P., Hall, B., Elkins, J., Conway, T., Tans, P., and Sweeney, C.: On the global distribution, seasonality, and budget of atmospheric carbonyl sulfide (COS) and some similarities to CO2, Journal of Geophysical Research: Atmospheres, 112, 2007.
Moore, J. W. and Schindler, D. E.: Getting ahead of climate change for ecological adaptation and resilience, Science, 376, 1421-1426, 2022.

Niu, S., Luo, Y., Dietze, M. C., Keenan, T. F., Shi, Z., Li, J., and III, F. S. C.: The role of data assimilation in predictive ecology, Ecosphere, 5, 1-16, 2014.
Norton, A. J., Rayner, P. J., Koffi, E. N., and Scholze, M.: Assimilating solar-induced chlorophyll fluorescence into the terrestrial biosphere model BETHY-SCOPE v1. 0: model description and information content, Geoscientific Model Development, 11, 1517-1536, 2018.
Ogée, J., Sauze, J., Kesselmeier, J., Genty, B., Van Diest, H., Launois, T., and Wingate, L.: A new mechanistic framework to predict OCS fluxes from soils, Biogeosciences, 13, 2221-2240, 2016.
Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., and Humphrey, M.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific data, 7, 1-27, 2020.
Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., Koffi, E., Kane, A., Maignan, F., and Chevallier, F.: A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle, Geoscientific Model Development, 9, 3321-3346, 2016.
Protoschill-Krebs, G., Wilhelm, C., and Kesselmeier, J.: Consumption of carbonyl sulphide (COS) by higher plant carbonic anhydrase (CA), Atmospheric Environment, 30, 3151-3156, 1996.
Prytz, G., Futsaether, C. M., and Johnsson, A.: Thermography studies of the spatial and temporal variability in stomatal conductance of Avena leaves during stable and oscillatory transpiration, New Phytologist, 158, 249-258, 2003.
Quirita, V. A. A., da Costa, G. A. O. P., Happ, P. N., Feitosa, R. Q., da Silva Ferreira, R., Oliveira, D. A. B., and Plaza, A.: A new cloud computing architecture for the classification of remote sensing data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 409-416, 2016.
Rastogi, B., Berkelhammer, M., Wharton, S., Whelan, M. E., Itter, M. S., Leen, J. B., Gupta, M. X., Noone, D., and Still, C. J.: Large uptake of atmospheric OCS observed at a moist old growth forest: Controls and implications for carbon cycle applications, Journal of Geophysical Research: Biogeosciences, 123, 3424-3438, 2018.
Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), Global biogeochemical cycles, 19, 2005.
Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., and Granier, A.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Global change biology, 11, 1424-1439, 2005.
Resco de Dios, V., Chowdhury, F. I., Granda, E., Yao, Y., and Tissue, D. T.: Assessing the potential functions of nocturnal stomatal conductance in C3 and C4 plants, New Phytologist, 223, 1696-1706, 2019.
Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., Chen, J. M., Ciais, P., Davis, K. J., and Desai, A. R.: Terrestrial biosphere models need better representation of vegetation phenology: results from the N orth A merican C arbon P rogram S ite S ynthesis, Global Change Biology, 18, 566-584, 2012.
Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., and Bosilovich, M.: The global land data assimilation system, Bulletin of the American Meteorological society, 85, 381-394, 2004.
Rogers, A.: The use and misuse of V c, max in Earth System Models, Photosynthesis research, 119, 15-29, 2014.
Rogers, A., Medlyn, B. E., Dukes, J. S., Bonan, G., Von Caemmerer, S., Dietze, M. C., Kattge, J., Leakey, A. D., Mercado, L. M., and Niinemets, Ü.: A roadmap for improving the representation of photosynthesis in Earth system models, New Phytologist, 213, 22-42, 2017.
Ryu, Y., Jiang, C., Kobayashi, H., and Detto, M.: MODIS-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000, Remote Sensing of Environment, 204, 812-825, 2018.
Salmon, E., Jégou, F., Guenet, B., Jourdain, L., Qiu, C., Bastrikov, V., Guimbaud, C., Zhu, D., Ciais, P., and Peylin, P.: Assessing methane emissions for northern peatlands in ORCHIDEE-PEAT revision 7020, Geoscientific Model Development, 15, 2813-2838, 2022.
Sandoval-Soto, L., Stanimirov, M., Von Hobe, M., Schmitt, V., Valdes, J., Wild, A., and Kesselmeier, J.: Global uptake of carbonyl sulfide (COS) by terrestrial vegetation: Estimates corrected by deposition velocities normalized to the uptake of carbon dioxide (CO 2), Biogeosciences, 2, 125-132, 2005.
Santaren, D., Peylin, P., Viovy, N., and Ciais, P.: Optimizing a process‐based ecosystem model with eddy‐covariance flux measurements: A pine forest in southern France, Global Biogeochemical Cycles, 21, 2007.
Sargsyan, K., Safta, C., Najm, H. N., Debusschere, B. J., Ricciuto, D., and Thornton, P.: Dimensionality reduction for complex models via Bayesian compressive sensing, International Journal for Uncertainty Quantification, 4, 2014.
Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., Miller, C., Frankenberg, C., Hibbard, K., and Cox, P.: Observing terrestrial ecosystems and the carbon cycle from space, Global Change Biology, 21, 1762-1776, 2015.
Scholes, R. J. and Walker, B. H.: An African savanna: synthesis of the Nylsvley study, Cambridge University Press1993.
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, 2017.
Scholze, M., Kaminski, T., Knorr, W., Blessing, S., Vossbeck, M., Grant, J., and Scipal, K.: Simultaneous assimilation of SMOS soil moisture and atmospheric CO2 in-situ observations to constrain the global terrestrial carbon cycle, Remote sensing of environment, 180, 334-345, 2016.

Scholze, M., Kaminski, T., Knorr, W., Voßbeck, M., Wu, M., Ferrazzoli, P., Kerr, Y., Mialon, A., Richaume, P., and Rodríguez-Fernández, N.: Mean European carbon sink over 2010–2015 estimated by simultaneous assimilation of atmospheric CO2, soil moisture, and vegetation optical depth, Geophysical Research Letters, 46, 13796-13803, 2019.
Schürmann, G. J., Kaminski, T., Köstler, C., Carvalhais, N., Voßbeck, M., Kattge, J., Giering, R., Rödenbeck, C., Heimann, M., and Zaehle, S.: Constraining a land-surface model with multiple observations by application of the MPI-Carbon Cycle Data Assimilation System V1. 0, Geoscientific Model Development, 9, 2999-3026, 2016.
Schwalm, C. R., Williams, C. A., Schaefer, K., Anderson, R., Arain, M. A., Baker, I., Barr, A., Black, T. A., Chen, G., and Chen, J. M.: A model‐data intercomparison of CO2 exchange across North America: Results from the North American Carbon Program site synthesis, Journal of Geophysical Research: Biogeosciences, 115, 2010.
Seibt, U., Kesselmeier, J., Sandoval-Soto, L., Kuhn, U., and Berry, J.: A kinetic analysis of leaf uptake of COS and its relation to transpiration, photosynthesis and carbon isotope fractionation, Biogeosciences, 7, 333-341, 2010.
Shaw, D. C., Franklin, J. F., Bible, K., Klopatek, J., Freeman, E., Greene, S., and Parker, G. G.: Ecological setting of the Wind River old-growth forest, Ecosystems, 7, 427-439, 2004.
Smith, K. S., Jakubzick, C., Whittam, T. S., and Ferry, J. G.: Carbonic anhydrase is an ancient enzyme widespread in prokaryotes, Proceedings of the National Academy of Sciences, 96, 15184-15189, 1999.
Spielmann, F., Wohlfahrt, G., Hammerle, A., Kitz, F., Migliavacca, M., Alberti, G., Ibrom, A., El‐Madany, T. S., Gerdel, K., and Moreno, G.: Gross primary productivity of four European ecosystems constrained by joint CO2 and COS flux measurements, Geophysical research letters, 46, 5284-5293, 2019.
Spielmann, F. M., Hammerle, A., Kitz, F., Gerdel, K., and Wohlfahrt, G.: Seasonal dynamics of the COS and CO 2 exchange of a managed temperate grassland, Biogeosciences, 17, 4281-4295, 2020.
Staudt, K., Falge, E., Pyles, R. D., Paw U, K. T., and Foken, T.: Sensitivity and predictive uncertainty of the ACASA model at a spruce forest site, Biogeosciences, 7, 3685-3705, 2010.
Stimler, K., Berry, J. A., and Yakir, D.: Effects of carbonyl sulfide and carbonic anhydrase on stomatal conductance, Plant Physiology, 158, 524-530, 2012.
Stimler, K., Berry, J. A., Montzka, S. A., and Yakir, D.: Association between carbonyl sulfide uptake and 18D during gas exchange in C3 and C4 leaves, Plant physiology, 157, 509-517, 2011.
Stimler, K., Montzka, S. A., Berry, J. A., Rudich, Y., and Yakir, D.: Relationships between carbonyl sulfide (COS) and CO2 during leaf gas exchange, New Phytologist, 186, 869-878, 2010.
Sun, W., Maseyk, K., Lett, C., and Seibt, U.: A soil diffusion–reaction model for surface COS flux: COSSM v1, Geoscientific Model Development, 8, 3055-3070, 2015.
Sun, W., Kooijmans, L. M., Maseyk, K., Chen, H., Mammarella, I., Vesala, T., Levula, J., Keskinen, H., and Seibt, U.: Soil fluxes of carbonyl sulfide (COS), carbon monoxide, and carbon dioxide in a boreal forest in southern Finland, Atmospheric Chemistry and Physics, 18, 1363-1378, 2018.
Sun, Z., Wang, X., Zhang, X., Tani, H., Guo, E., Yin, S., and Zhang, T.: Evaluating and comparing remote sensing terrestrial GPP models for their response to climate variability and CO2 trends, Science of the total environment, 668, 696-713, 2019.
Talagrand, O.: Assimilation of observations, an introduction (gtspecial issueltdata assimilation in meteology and oceanography: Theory and practice), Journal of the Meteorological Society of Japan. Ser. II, 75, 191-209, 1997.
Talagrand, O. and Courtier, P.: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, Quarterly Journal of the Royal Meteorological Society, 113, 1311-1328, 1987.
Tarantola, A.: Inverse problem theory : methods for data fitting and model parameter estimation, 1987.
Tarantola, A.: Inverse problem theory and methods for model parameter estimation, SIAM2005.
Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czikowsky, M., and Munger, J.: Factors controlling CO2 exchange on timescales from hourly to decadal at Harvard Forest, Journal of Geophysical Research: Biogeosciences, 112, 2007.
Verbeeck, H., Samson, R., Verdonck, F., and Lemeur, R.: Parameter sensitivity and uncertainty of the forest carbon flux model FORUG: a Monte Carlo analysis, Tree physiology, 26, 807-817, 2006.
Vesala, T., Kohonen, K.-M., Kooijmans, L. M., Praplan, A. P., Foltýnová, L., Kolari, P., Kulmala, M., Bäck, J., Nelson, D., and Yakir, D.: Long-term fluxes of carbonyl sulfide and their seasonality and interannual variability in a boreal forest, Atmospheric Chemistry and Physics, 22, 2569-2584, 2022.
Wang, J., Jiang, F., Wang, H., Qiu, B., Wu, M., He, W., Ju, W., Zhang, Y., Chen, J. M., and Zhou, Y.: Constraining global terrestrial gross primary productivity in a global carbon assimilation system with OCO-2 chlorophyll fluorescence data, Agricultural and Forest Meteorology, 304, 108424, 2021.
Wang, K.-Y., Kellomäki, S., Zha, T., and Peltola, H.: Component carbon fluxes and their contribution to ecosystem carbon exchange in a pine forest: an assessment based on eddy covariance measurements and an integrated model, Tree Physiology, 24, 19-34, 2004.
Wehr, R., Commane, R., Munger, J. W., McManus, J. B., Nelson, D. D., Zahniser, M. S., Saleska, S. R., and Wofsy, S. C.: Dynamics of canopy stomatal conductance, transpiration, and evaporation in a temperate deciduous forest, validated by carbonyl sulfide uptake, Biogeosciences, 14, 389-401, 2017.
Whelan, M. E., Hilton, T. W., Berry, J. A., Berkelhammer, M., Desai, A. R., and Campbell, J. E.: Carbonyl sulfide exchange in soils for better estimates of ecosystem carbon uptake, Atmospheric Chemistry and Physics, 16, 3711-3726, 2016.
Whelan, M. E., Shi, M., Sun, W., Vries, L. K. d., Seibt, U., and Maseyk, K.: Soil carbonyl sulfide (OCS) fluxes in terrestrial ecosystems: an empirical model, Journal of Geophysical Research: Biogeosciences, 127, e2022JG006858, 2022.

Whelan, M. E., Lennartz, S. T., Gimeno, T. E., Wehr, R., Wohlfahrt, G., Wang, Y., Kooijmans, L. M., Hilton, T. W., Belviso, S., and Peylin, P.: Reviews and syntheses: Carbonyl sulfide as a multi-scale tracer for carbon and water cycles, Biogeosciences, 15, 3625-3657, 2018.
Wohlfahrt, G., Brilli, F., Hörtnagl, L., Xu, X., Bingemer, H., Hansel, A., and Loreto, F.: Carbonyl sulfide (COS) as a tracer for canopy photosynthesis, transpiration and stomatal conductance: potential and limitations, Plant, cell & environment, 35, 657-667, 2012.
Woodward, F. I., Smith, T. M., and Emanuel, W. R.: A global land primary productivity and phytogeography model, Global biogeochemical cycles, 9, 471-490, 1995.
Wu, M., Scholze, M., Kaminski, T., Voßbeck, M., and Tagesson, T.: Using SMOS soil moisture data combining CO2 flask samples to constrain carbon fluxes during 2010–2015 within a Carbon Cycle Data Assimilation System (CCDAS), Remote Sensing of Environment, 240, 111719, 2020.
Wu, M., Scholze, M., Voßbeck, M., Kaminski, T., and Hoffmann, G.: Simultaneous assimilation of remotely sensed soil moisture and FAPAR for improving terrestrial carbon fluxes at multiple sites using CCDAS, Remote Sensing, 11, 27, 2018.
Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., and Song, J.: 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, 2016.
Yu, K., Goldsmith, G. R., Wang, Y., and Anderegg, W. R.: Phylogenetic and biogeographic controls of plant nighttime stomatal conductance, New Phytologist, 222, 1778-1788, 2019.
Zaehle, S., Sitch, S., Smith, B., and Hatterman, F.: Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics, Global Biogeochemical Cycles, 19, 2005.
Zierl, B.: A water balance model to simulate drought in forested ecosystems and its application to the entire forested area in Switzerland, Journal of Hydrology, 242, 115-136, 2001.
Zobitz, J., Moore, D. J., Quaife, T., Braswell, B. H., Bergeson, A., Anthony, J. A., and Monson, R. K.: 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, 2014.

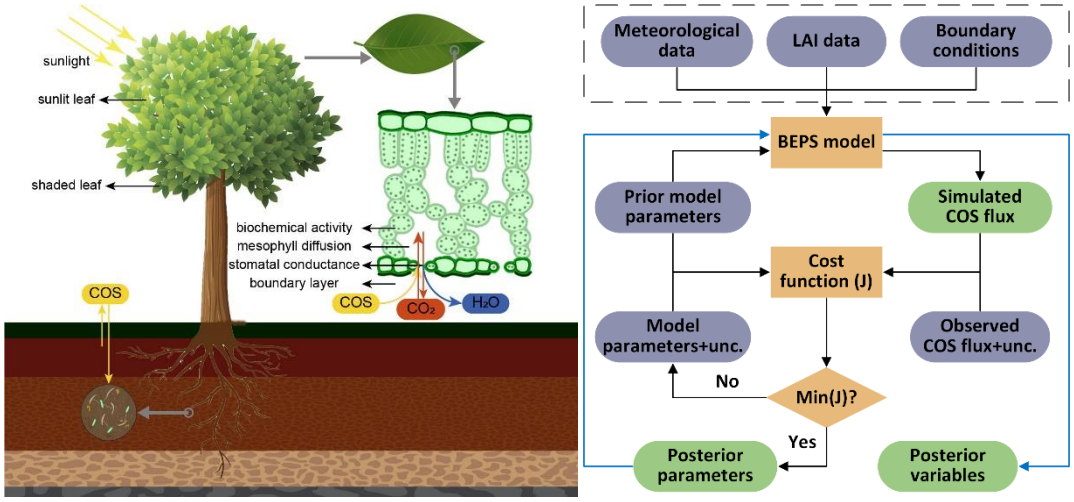


Figure 1. Schematic of the Nanjing University Carbon Assimilation System (NUCAS). Left: illustration of a two-leaf model coupling stomatal conductance, photosynthesis, transpiration and COS uptake, and an empirical model for simulating soil COS fluxes in NUCAS. Right: data assimilation flowchart of NUCAS. Ovals represent input (blue-grey) and output data (green). Boxes and the rhombi represent the calculation and judgement steps. The solid black line represents the diagnostic process, the solid blue line represents the prognostic process, and the input datasets of BEPS (in the dashed box) are used in both diagnostic process and prognostic process.

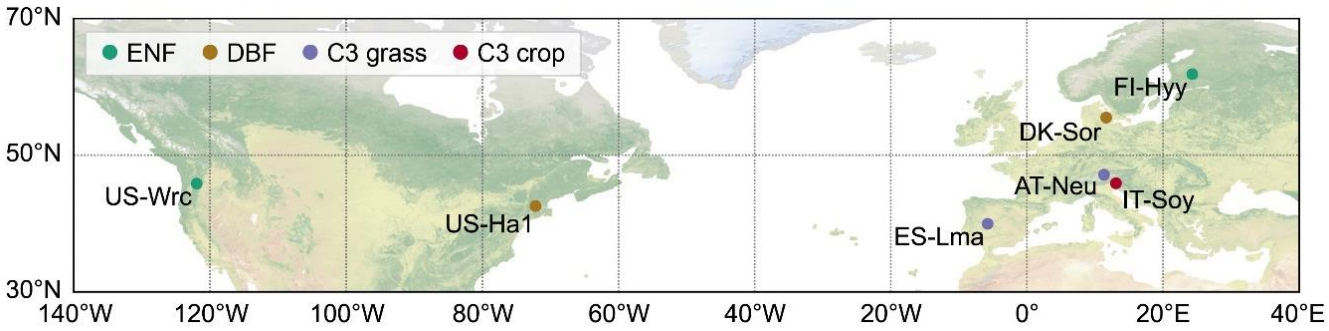


Figure 2. Locations of the 7 studied sites. Sites sharing the same plant function type are represented with consistent colors. The background map corresponds to the "Nature color I" map (https://www.naturalearthdata.com). ENF and DBF denote evergreen needleleaf forest and deciduous broadleaf forest, respectively.

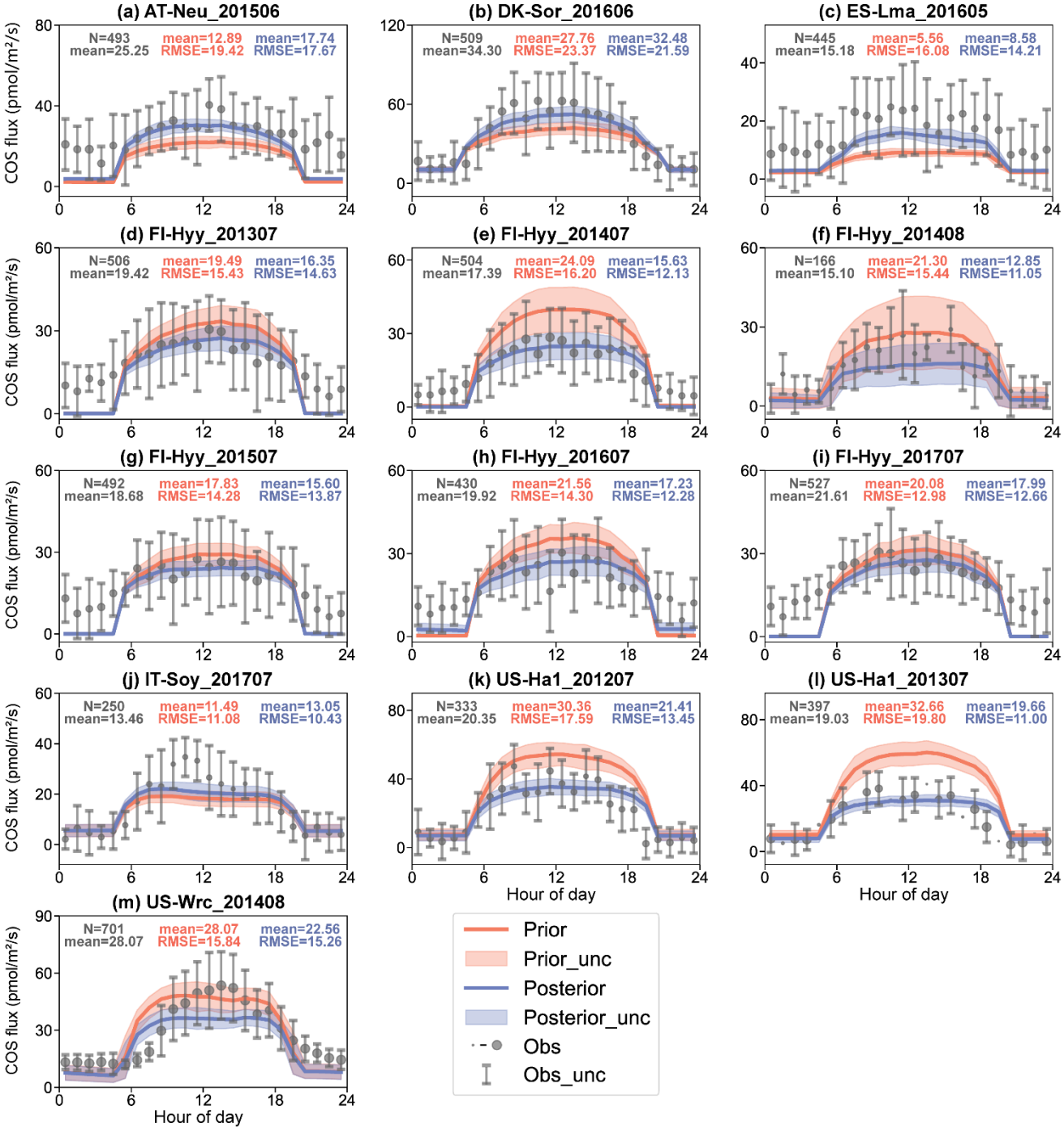


Figure 3. The mean diurnal cycle of observed (blue) and simulated COS flux using prior parameters (red) and single-site posterior parameters (blue). The size of the circle indicates the number of observations (ranging from 1 to 31) within each circle, and the error bars depict the standard deviations in the mean of observations from the variability within each circle if the number of corresponding observations is greater than three. Lines connect the mean values of simulations and pale bands depict the standard deviation in the mean of simulations from the variability within each bin.

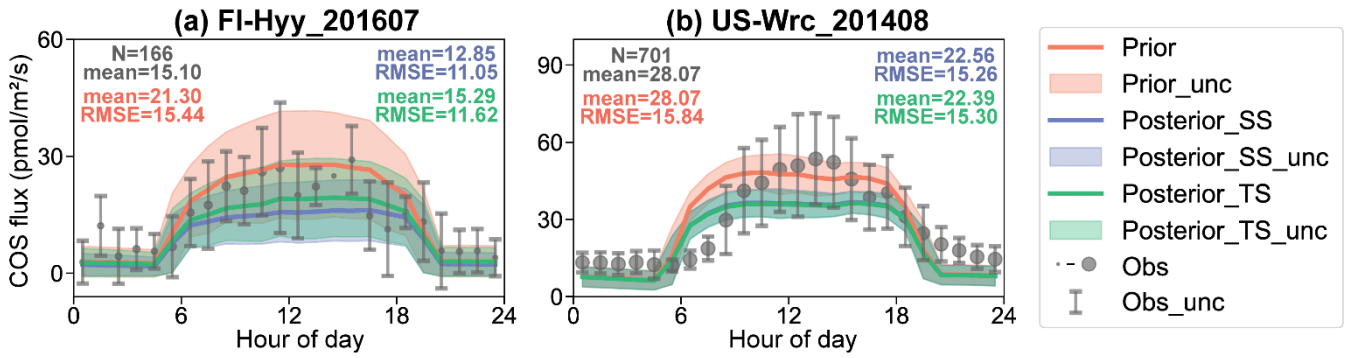


Figure 4. The diurnal cycle of observed (blue) and simulated COS flux using prior parameters (red), single-site (blue) and two-site (green) posterior parameters. The size of the circle indicates the number of observations (ranging from 1 to 31) within each circle, and the error bars depict the standard deviations in the mean of observations from the variability within each circle if the number of corresponding observations is greater than three. Lines connect the mean values of simulations and pale bands depict the standard deviation in the mean of simulations from the variability within each bin.

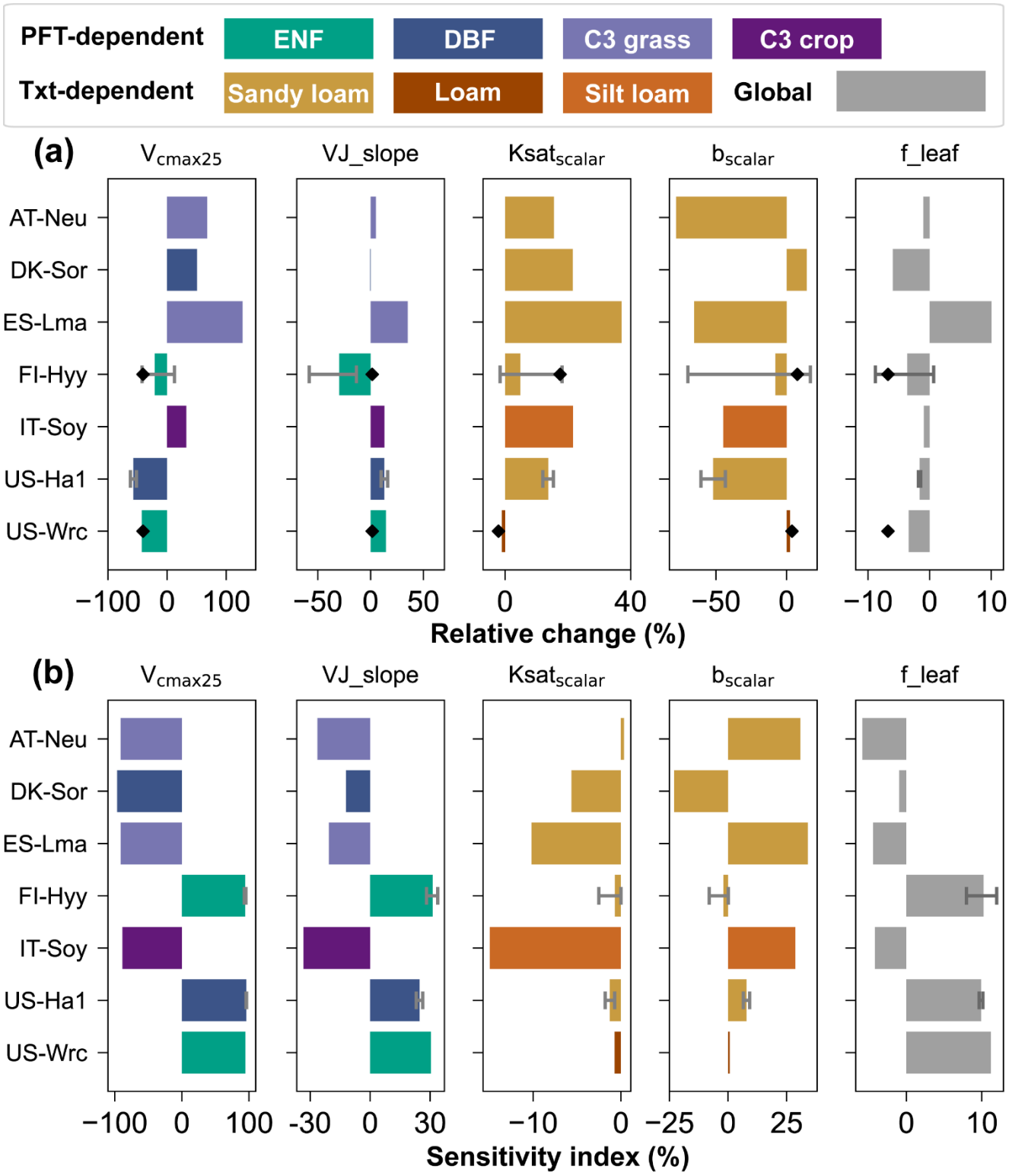
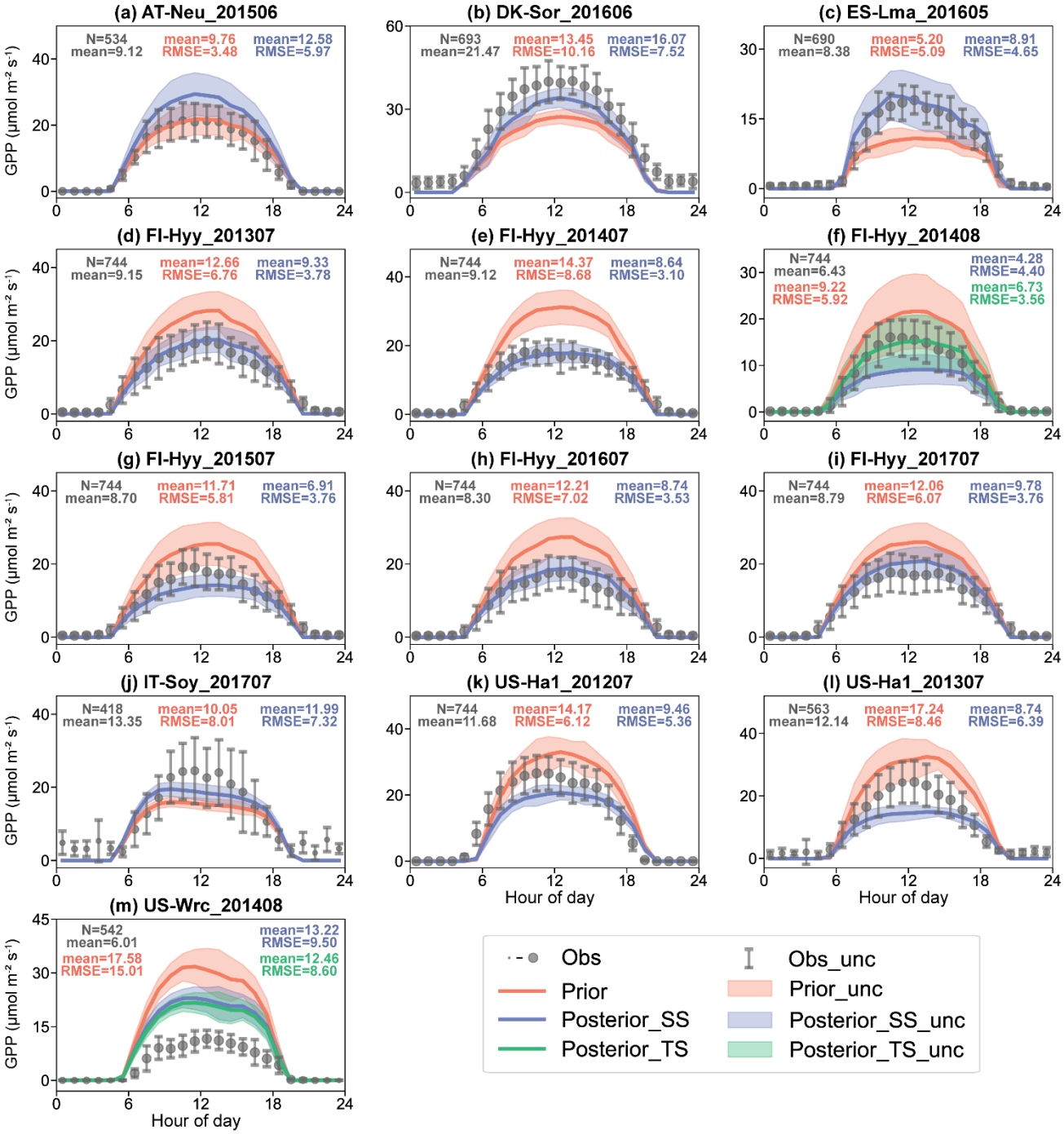


Figure 5. (**a**) Relative changes of parameters for single-site experiments (bars) and the multi-site experiment (diamond points). (**b**) Sensitivity indexes of parameters at prior values. For sites where multiple single-site experiments were conducted, the ends of the error bars and the bar indicate the maximum, minimum and mean of the relative changes of the parameters, respectively. For those sites lacking multi-year COS observations, no error bars were plotted. The color of bar is drawn according to PFT/texture.



1180 **Figure 6.** The diurnal cycle of observed (blue) and simulated GPP using prior parameters (red), single-site (green) and multi-site (brown) posterior parameters. The size of the circle indicates the number of observations within each circle (ranging from 1 to 31), and the error bars depict the standard deviations in the mean of observations from the variability within each circle. Lines connect the mean values of simulations and pale bands depict the standard deviation in the mean of simulations from the variability within each bin.

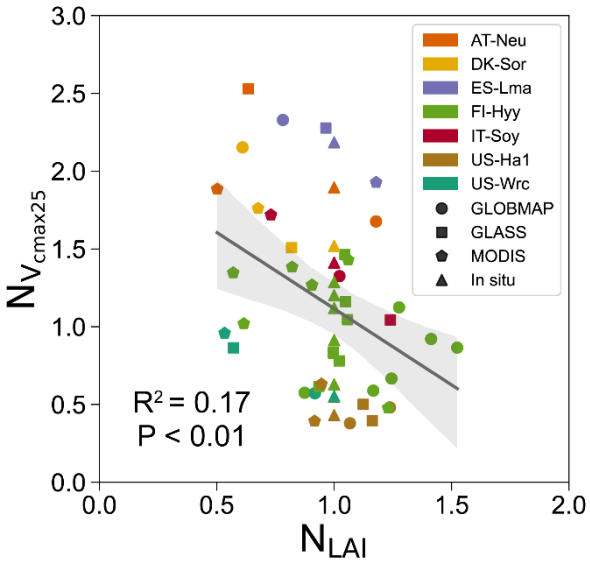


Figure 7. Influence of LAI on the posterior V_{cmax25} obtained by the single-site experiments conducted at seven sites and driven by four LAI data (GLOBMAP, GLASS, MODIS and *in situ*). The posterior V_{cmax25} and the LAI were represented by their normalized values $N_{V_{cmax25}}$ and N_{LAI}, respectively. The posterior parameters were normalized by their prior values and the LAI were normalized by the *in situ* values. The linear regression fit line of the posterior parameters obtained based on the satellite-derived LAI (GLOBMAP, GLASS and MODIS) with the corresponding LAI data is shown, with 95% confidence interval spread around the line.

Table 1. Site characteristics. Site identification includes the country initials and a three-letter name for each site; locations of the sites are provided by the latitude (Lat) and longitude (Lon); PFTs covered by the sites are evergreen needleleaf forest (ENF), deciduous broadleaf forest (DBF), C3 grass and C3 crop; Soil texture covered by the sites are sandy loam, slit loam and loam.

Site name	AT-Neu	DK-Sor	ES-Lma	FI-Hyy	IT-Soy	US-Ha1	US-Wrc
Lat (°N)	47.12	55.49	39.94	61.85	45.87	42.54	45.82
Lon (°E)	11.32	11.64	-5.77	24.29	13.08	-72.17	-121.95
PFT	C3 grass	DBF	C3 grass	ENF	C3 crop	DBF	ENF
Soil texture	Sandy loam	Sandy loam	Sandy loam	Sandy loam	Slit loam	Sandy loam	Loam
LAI*	3.88	5.0	1.82	4.0	2.3	5.0	8.7
Year	2015	2016	2016	2013-2017	2017	2012-2013	2014
References	(Spielmann et al., 2020)	(Spielmann et al., 2019)	(Spielmann et al., 2019)	(Sun et al., 2018; Vesala et al., 2022; Kohonen et al., 2022)	(Spielmann et al., 2019; Abadie et al., 2022)	(Commane et al., 2015; Wehr et al., 2017)	(Shaw et al., 2004; Rastogi et al., 2018)

** Mean one-sided LAI* ($m^2\ m^{-2}$) *during the experimental period*

Table 2. The configuration and the relative changes (%) of the parameters for each single-site assimilation experiment. The cost function reduction of each experiment is indicated by the reduction rate between the initial value of cost function ($J_{initial}$) and the final value of cost function (J_{final}), defined as $1 - J_{final}/J_{initial}$, and N_{COS} denotes the number of ecosystem COS flux observations.

Site name	Assimilation window	N_{COS}	Cost function reduction (%)	Relative change (%) of parameters				
				V_{cmax25}	VJ_slope	$Ksat_{scalar}$	b_{scalar}	f_leaf
AT-Neu	June 2015	493	16.39	67.69	5.10	15.57	-78.13	-1.01
DK-Sor	June 2016	509	9.46	50.77	-0.47	21.54	14.23	-5.97
ES-Lma	May 2016	445	15.70	127.80	35.18	37.08	-65.33	10.05
FI-Hyy	July 2013	506	4.87	32.55	13.15	21.60	-44.72	-0.94
	July 2014	504	7.74	-13.42	-25.48	-1.58	0.90	-8.80
	August 2014	166	40.59	-41.09	-19.10	4.02	16.84	-6.21
	July 2015	492	50.94	-42.44	-41.03	8.65	5.07	-1.66
	July 2016	430	5.73	12.45	-58.23	0.00	-0.07	0.65
	July 2017	527	18.94	-33.32	-13.48	18.13	-69.86	-1.60
IT-Soy	July 2017	250	6.35	-7.88	-21.20	0.03	-0.45	-4.14

US-Ha1	July 2012	333	44.14	-51.89	16.08	12.05	-43.31	-1.44
	July 2013	397	69.05	-62.08	10.00	15.39	-60.58	-1.82
US-Wrc	August 2014	701	27.71	-42.77	14.52	-1.04	2.45	-3.39

Table 3. The configuration and the relative changes (%) of the parameters for the multi-site assimilation experiment at FI-Hyy and US-Wrc. N_{COS} denotes the total number of ecosystem COS flux observations.

Site name	Assimilation window	N_{COS}	Cost function reduction (%)	Relative change (%) of parameters				
				V_{cmax25}	VJ_slope	$Ksat_{scalar}$	b_{scalar}	f_leaf
FI-Hyy	August 2014	867	28.29	-41.36	2.96	17.32	5.56	-6.28
US-Wrc						1.36	-2.60	

Appendix: Stomatal conductance and soil hydrology modelling in BEPS, including parameters to be optimised

In the BEPS model, the leaf stomatal conductance to water vapor (g_{sw} in mol m^{-2} s^{-1}) is estimated using a modified version of Ball-Berry (BB) empirical model (Ball et al., 1987) following Woodward et al. (1995):

$$g_{sw} = b_{H_2O} + \frac{m_{H_2O}\, A\ R_h\, f_w}{C_a} \tag{A1}$$

where b_{H_2O} is the intercept of the BB model, representing the minimum g_{sw} (mol m^{-2} s^{-1}), m_{H_2O} is the empirical slope parameter in the BB model (unitless), R_h is the relative humidity at the leaf surface (unitless), f_w is a soil moisture stress factor describing the sensitivity of g_{sw} to soil water availability (Ju et al., 2006), C_a is the atmospheric CO_2 concentration (μmol mol^{-1}), and the net photosynthesis rate (A) is calculated using the Farquhar model (Farquhar et al., 1980; Chen et al., 1999):

$$A = \min(A_i, A_j) - R_d \tag{A2}$$

$$A_c = V_{cmax}\, \frac{C_i - \Gamma_i^*}{C_i + K_c\left(1 + \frac{O_i}{K_o}\right)} \tag{A3}$$

$$A_j = J\, \frac{C_i - \Gamma_i^*}{4(C_i - 2\Gamma_i^*)} \tag{A4}$$

where A_i and A_j are Rubisco-limited and RuBP-limited gross photosynthetic rates (μmol m^{-2}s^{-1}), respectively. R_d is leaf dark respiration (μmol m^{-2}s^{-1}). V_{cmax} is the maximum carboxylation rate of Rubisco (μmol m^{-2}s^{-1}); J is the electron transport rate (μmol m^{-2}s^{-1}); Ci and Oi are the intercellular carbon dioxide (CO_2) and oxygen (O_2) concentrations (mol mol^{-1}), respectively; Kc and Ko are Michaelis–Menten constants for CO_2 and O_2 (mol mol^{-1}), respectively.

The electron transport rate, J, is dependent on incident photosynthetic photon flux density (PPFD, μmol m^{-2}s^{-1}) as:

$$\mathrm{J} = \frac{J_{max}\, I}{I + 2.1 J_{max}} \tag{A5}$$

where J_{max} is the maximum electron transport rate (μmol m^{-2}s^{-1}), I is the incident PPFD calculated from the incident shortwave radiation R_{SW} (W m^{-2}):

$$I = \beta\, R_{SW}\, f\_leaf \tag{A6}$$

where $\beta = 4.55$ is the energy – quanta conversion factor (μmol J^{-1}), f_leaf is the ratio of photosynthesis active radiation to the shortwave radiation (unitless).

The maximum carboxylation rate of Rubisco V_{cmax} was calculated according the Arrhenius temperature function and the maximum carboxylation rate of Rubisco at 25 °C (V_{cmax25}). V_{cmax} is generally proportional to leaf nitrogen content. Considering both the fractions of sunlit and shaded leaf areas to the total leaf area and the leaf nitrogen content vary with the depth into the canopy, the V_{cmax} values of sunlit ($V_{cmax,sun}$) and shaded ($V_{cmax,sh}$) leaves can be obtained through vertical integrations with respect to leaf area index (Chen et al., 2012):

$$V_{cmax,sunlit} = V_{cmax}\chi_n N_{leaf}\frac{k\left[1-e^{(k_n+k)LAI_{sunlit}}\right]}{(k_n+k)(1-e^{-kLAI_{sunlit}})} \quad \text{(A7)}$$

$$V_{cmax,shaded} = V_{cmax}\chi_n N_{leaf}\frac{\frac{1}{k_n}[1-e^{-K_nL}]-\frac{1}{k_n+k}\left[1-e^{(k_n+k)LAI_{shaded}}\right]}{LAI_{shaded}-\frac{1}{k}(1-e^{-kLAI_{shaded}})} \quad \text{(A8)}$$

where χ_n ($m^2\ g^{-1}$) is the relative change of V_{cmax} to leaf nitrogen content; N_{leaf} ($g\ m^{-2}$) is the leaf nitrogen content at the top of the canopy; k_n (unitless) is the leaf nitrogen content decay rate with increasing depth into the canopy, taken as 0.3; k is calculated as:

$$k = G(\theta)\Omega\cos(\theta) \quad \text{(A9)}$$

where G(θ) is the projection coefficient, taken as 0.5, Ω is the clumping index, and θ is the is the solar zenith angle.

After V_{cmax} values for the representative sunlit and shaded leaves are obtained, the maximum electronic transport rate for the sunlit and shaded leaves are obtained from Medlyn et al. (1999):

$$J_{max} = VJ\_slope\ V_{cmax} - 14.2 \quad \text{(A10)}$$

Soil water availability factor $f_{w,i}$ in each layer *i* is calculated as:

$$f_{w,i} = \frac{1.0}{f_i(\psi_i)f_i(T_{s,i})} \quad \text{(A11)}$$

where $f_i(\psi_i)$ is a function of matrix suction ψ_i (m) (Zierl, 2001), $f_i(T_{s,i})$ is a function describing the effect of soil temperature ($T_{s,i}$ in ℃) on soil water uptake (Bonan, 1991).

To consider the variable soil water potential at different depths, the scheme of Ju et al. (2006) was employed to calculate the weight of each layer (w_i) to f_w:

$$w_i = \frac{R_i f_{w,i}}{\sum_{i=1}^{n} R_i f_{w,i}} \quad \text{(A12)}$$

where *n* is the number of soil layer (five were used in this study) of the BEPS model, R_i is the root fraction in layer *i*, calculated as:

$$R_i = \begin{cases} 1 - {r_{decay}}^{100cd_i} & i = 1 \\ {r_{decay}}^{100cd_{i-1}} - {r_{decay}}^{100cd_i} & 1 < i < n \\ {r_{decay}}^{100cd_{i-1}} & i = n \end{cases} \quad \text{(A13)}$$

where cd_i is the cumulative depth (m) of layer *i*. In this study, each soil layer depth (from top to bottom) of the BEPS model is 0.05 m, 0.10 m, 0.20 m, 0.40 m and 1.25 m, respectively.

The overall soil water availability f_w is then calculated as:

$$f_w = \sum_{i=1}^{n} f_{w,i} w_i \quad \text{(A14)}$$

The hydraulic conductivity of each soil layer K_i ($m\ s^{-1}$) is expressed as:

$$K_i = Ksat_i\left(\frac{swc_i}{\theta_{s,i}}\right)^{2b_i+3} \quad \text{(A15)}$$

where $Ksat_i$ is the saturated hydrological conductivity of soil layer *i* ($m\ s^{-1}$); SWC_i is the volumetric liquid soil water content of soil layer *i* ($m\ s^{-1}$); $\theta_{s,i}$ is the porosity of soil layer *i* (unitless); b_i is the Campbell parameter for soil layer *i*, determining the change rate of hydraulic conductivity with SWC (unitless). In this study, $Ksat_i$ and b_i are expressed as:

$$Ksat_i = Ksat_{scalar}Ksat_{df,i} \quad \text{(A16)}$$

$$b_i = b_{scalar}b_{df,i} \quad \text{(A17)}$$

where $Ksat_{df,i}$ and $b_{df,i}$ are the default values of $Ksat_i$ and b_i respectively.