We would like to thank both reviewers for their detailed and insightful comments. These comments have helped improve and clarify the submitted manuscript. Below we reply to each comment point by point, showing the reviewers' comments in black and our responses in blue. Changes to the original manuscript are highlighted in **bold blue**. Note that the line numbers in the response are updated based on the revised manuscript, which we provide with our response..

We note already here that we reran all our numerical experiments, in response to two comments of Reviewer #2, one on the processing of COS flux observations and one on the prior uncertainty specified for the parameter f_leaf and to one comment by reviewer # 1 on the size of the perturbation for the starting point of the twin experiments.

Reviewer #1

The paper by Zhu et al. presents an interesting study of data assimilation of carbonyl sulfide (COS) using the BEPS model. They used adjoint method to assimilate the COS fluxes as NUCAS v1.0. This is a new model tool to the modelling science and is useful for study of carbon cycle. The novelty of the model is that it assimilates COS flux to improve the model performance of GPP and other model parameters. Therefore, the research is within the scope of GMD and could be considered as publishable. However, there are some issues the authors should address before publication.

Response: We thank the reviewer for this comment. We will address these issues in order to make this paper publishable in GMD.

First of all, the adjoint code used in this paper is based on the automatic differentiation tool TAPENADE (Hascoët and Pascual, 2013). Yet, the authors did not validate the adjoint method or did not write it clearly. The question is: how do you justify that the adjoint codes will produce correct optimization?

Response: We extended the text as follows "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.**" (Line 125-127)

Secondly, the logic of the paper is lost in some places. Section 3.7 and 3.8 showed results of comparison and evaluation of simulated H and LE, and SWC. But it is unclear how data assimilation of COS flux can impact those parameters, and the performance is less satisfactory than evaluations of COS fluxes and GPP. The question: is there causality between assimilation of COS fluxes and H, LE, and SWC? What is your hypothesis that COS fluxes are linked to H, LE and SWC? Consider adding details in Section 2.

Response: Since the leaf exchange of COS, carbon dioxide (CO_2) and water vapor are tightly coupled though stomata, COS has been proved as a useful tracer of photosynthesis, stomatal conductance and transpiration (Sandoval-Soto et al., 2005; Wohlfahrt et al., 2012). Transpiration is closely linked to soil moisture because the water it dissipates originates from the soil (Berry et al., 2006). This process of water turning from liquid to vapor requires energy,

and that energy is a crucial part of the ecosystem latent heat (LE) (Gupta et al., 2018). The energy is obtained from the surrounding leaf cells, leading to a decrease in temperature within the leaf (so called "cooling effect") (Gates, 1968; Gupta et al., 2018). Thus, the sensible heat (H) can be linked to transpiration since the leaf-to-air temperature gradient is a key control factor of it (Monteith and Unsworth, 2013; Dong et al., 2017). Therefore, our hypothesis is that the assimilation of COS is expected to improve the modelling of LE, H and SWC due to the ability of COS to indicate transpiration and the mechanism of transpiration (i.e. the corresponding energy transfer, cooling effect and water source).

We have added detailed in Section 2.3: "**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).**" (Line 276-279)

Another recent paper By Cho et al. is worthy of a comparison and discussion: Cho, A., Kooijmans, L. M. J., 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, https://doi.org/10.5194/bg-20-2573-2023, 2023

Response: Based on previous studies on the temperature response of carbonic anhydrase (CA), Rubisco enzyme and LRU, Cho et al. (2023) proposed a new COS plant uptake scheme for CA with the argument that different enzymes have different physiological characteristics. Through data assimilation, they combined COS and GPP observations with the Simple Biosphere model (SiB4) simulations to optimize stomatal conductance parameters b0 and b1, empirical parameter a, and CA enzyme optimum temperature, and thus improved the model performance of stomatal conductance, 'interior' conductance, and COS leaf uptake. This study provides new insights into achieving accurate modeling of COS plant uptake, which is worthy of comparison and discussion.

Firstly, precise modeling of carbonyl sulfide (COS) is fundamental for the utility of COS observations in optimizing model parameters associated with COS. The remarkable contribution of Cho et al. (2023) to COS modeling would undoubtedly benefit the work in utilizing COS as a probe to explore the ecological processes such as water-carbon exchange and energy flow within ecosystems.

Secondly, while the study by Cho et al. (2023) focused on optimizing COS-related and stomatal-related parameters, our investigation concentrates on refining parameters associated with photosynthesis and soil hydrology. Although the parameters optimized in our study influence stomatal modeling, our results reveal that the optimization of transpiration-related variables (LE, H, SWC) is comparatively less successful than that of COS and GPP. The insights gained from Cho et al. (2023)'s work underscore the potential for achieving improved optimization of transpiration-related variables by utilizing COS to directly constrain parameters associated with stomatal conductance.

Thus, we extended the text as follows: "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.**" (Line 701-706)

Other minor comments:

Line 142: "For NUCAS, we use the same soil texture" to "we used the same soil texture."

Response: Corrected.

Line 185: the sites used in the study is better to be shown in a Figure to give a general idea of the locations of those sites.

Response: Thanks for your suggestion, we have added such a figure to our manuscript, as shown below.

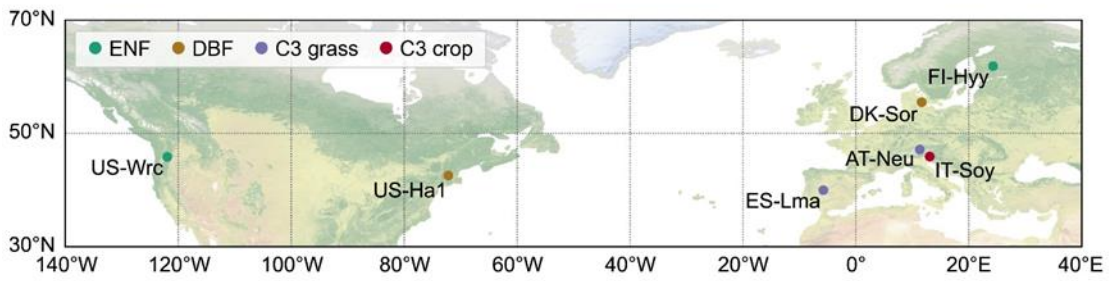


Figure 2. Locations of the 7 studied sites. Sites sharing the same plant function type are represented with consistent colors. The background map corresponds to the "Nature color I" map (https://www.naturalearthdata.com). ENF and DBF denote evergreen needleleaf forest and deciduous broadleaf forest,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

Line 197: "the CO_2 and COS mole fractions in the bulk air were assumed to be spatially invariant." What is the value of CO_2 and COS mole fractions in your case?

Response: Thanks for your comment. we extended the text as follows: "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 CO2 mole fraction data utilized 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.**" (Line 219-223)

Line 227: "in situ" to "*in situ*", and all elsewhere.

Response: Thanks for your reminder, we have changed to "*in situ*" throughout the manuscript.

Line 284: "For all cases where the PFT is evergreen needleleaf forest, a perturbation ratio of 0.2 was used. And for the remaining six single-site twin experiments, a perturbation rate of 0.4 was used." Please specify the reasons to those perturbation rate as 0.2 or 0.4.

Response: Thanks for your comment. The settings of the prior parameter uncertainties in this study refer to previous studies, e.g., Chen et al. (2022), Ryu et al. (2018). Now, the prior uncertainty of most model parameters was set to 25% of the prior value, while the prior uncertainty of f_leaf was estimated using the datasets provided by Ryu et al. (2018) and was about 7 % of the prior value. These studies also provide us with reference for understanding the degree of parameter variability and choosing the perturbation rate. Now, we chose a perturbation ratio (0.2) that falls between these two values (7 % and 25 %), but is closer to the prior uncertainty with most of the parameters, and reran all the twin experiments.

Line 440: "very reasonable". Is there another way to say "very"?

Response: Thank you for the suggestion. The relevant parts have been re-written in the revised manuscript.

Line 450: "very similar". The same as Line 440. And check all elsewhere.

Response: Thanks for your comment. The relevant parts have been re-written in the revised manuscript and we have checked all elsewhere.

Line 513: "assimilation using COS observations from multiple sites can also improve GPP simulations, and the assimilation is sometimes", it is vague to use sometimes to describe results.

Response: Thanks for your comment. We have reorganized the sentences to avoid vagueness.

Line 1165: Figure 4, it is not easy to see clearly the green and gray shading. Please consider better visualization.

Response: Thanks for your comment. We have remade our figures so that the results can be easily distinguished.

Line 1170 and 1175: Figure 5 and 6, why there are error bars for some sites but no error bars for other sites?

Response: Thanks for your comment. In this study, FI-Hyy and US-Ha1 are the only two sites with multi-year COS observations, which provides an opportunity to investigate the optimization results of COS-related parameters and the effectiveness of COS assimilation in different years. For these two sites, error bars were plotted to represent the maximum and minimum of the posterior parameter values. In contrast, no error bars were plotted for the other sites due to the lack of multi-year COS observations. We have described in the manuscript that we plotted error bars for sites with multiple years of COS observations. In response to your question, we have added a note to the figure legend of the revised manuscript: "**For those sites lacking multi-year COS observations, no error bars were plotted.**" (Line 1177-1178)

Line 1185: Figure 8. It is hard to see difference between green and gray. The dots in c and f are maybe too big.

Response: Thanks for your comment. We have reorganized the figure using smaller dots and

changed the colours for better visualization.

Review #2

Zhu et al. present a new assimilation model NUCAS v1.0 for simulating carbonyl sulfide (COS) fluxes at ecosystem scale. The model is a good addition to the COS modeling pool, but the study requires some modifications and the paper lacks important information and is in many places too ambiguous and inconsistent.

Response: We thank the reviewer for this comment. In response to this comment, we have refined the manuscript to enhance clarity and ensure consistency. The necessary information has been incorporated, rendering the manuscript comprehensive and informative.

General comments: The paper lacks consistency on terminology used throughout the paper. Examples: in Eq. 1 observation is marked with O and model with M while in Eq. 12 they are marked with c and s and in Eqs. 14-16 they are marked obs and sim, respectively. Soil moisture is sometimes marked with SWC and sometimes as Θ. Section 2.1.3 is full of examples (listed below in more detail). This makes the paper very difficult to follow for the reader.

Response: We thank the reviewer for this comment. To enhance readability, we have revised the manuscript to ensure consistency in terminology. In the revised manuscript, we have designated observations as 'O' and the model as 'M.' Soil moisture is identified by 'SWC.' Furthermore, to mitigate ambiguity with 'C' in Eq.1, we now use 'F' to represent the corrected COS fluxes. Additional details regarding the rationale for utilizing corrected COS data from the US-Wrc site have been elaborated below.

The authors model soil and plant COS fluxes separately but only report the total ecosystem flux. However, it would be interesting to see the simulated soil and plant fluxes separately and see how they compare with measured chamber COS fluxes from the different sites and also with e.g. other soil models.

Response: Thanks for this valuable comment. Actually, there are many difficulties in evaluating COS soil and plant fluxes separately for the sites used in this study. The five-year COS ecosystem flux data at FI-Hyy provided us an opportunity to investigate the difference of assimilation performance of COS. However, the soil COS flux data at FI-Hyy are only available in 2015, which makes it impossible for us to separately evaluate COS plant flux and soil flux for the vast majority of experiments conducted at FI-Hyy. In addition, Whelan et al. (2022)have evaluated the model performance at FI-Hyy in 2015 and US-Ha1 using a similar soil model. At US-Wrc, 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. Thus, there may be significant biases in our estimates of both plant and soil fluxes at US-Wrc. As for DK-Sor, ES-Lma and IT-Soy, a random forest regression model was trained for each site in order to simulate the soil COS exchange, and only the modelled COS soil fluxes are provided in Spielmann et al. (2019) while the observational data for COS soil flux is lacking. Overall, given the insufficient and inconsistent availability of separate COS soil and plant data, we face considerable obstacles in separately

assessing simulated COS soil and plant fluxes.

Additionally, in NUCAS, the resistance analog model of COS plant uptake and the empirical model of soil COS flux were embedded in the BEPS model, and the model performance of these COS models have been evaluated in numerous previous studies (Berry et al., 2013; Whelan et al., 2016; Kooijmans et al., 2021; Maignan et al., 2021; Whelan et al., 2022; Chen et al., 2023; Cho et al., 2023). These studies have demonstrated the usefulness and robustness of these models to simulate COS plant and soil fluxes, thus founded the basis for us to assimilate COS ecosystem flux in this study.

Last but not least, we do agree with your opinion and we also 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 component to provide a stronger constraint than using just their sum.

Therefore, we extended the text in the conclusion: "**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.**" (Line 739-743)

Some coefficients and uncertainty estimates used in the paper are very poorly explained. Where does a perturbation rate of 0.4 come for some sites while for others it is 0.2? How do the authors come up with an uncertainty of 1 pmol m^{-2} s^{-1} for the prior simulated COS flux (L275)? Section 2.1.3 is also filled with these coefficients, listed in more detail below.

Response: Thanks for your comment. Reviewer #1 asked a similar question about the choice of the perturbation size, please refer to our previous answer. Besides, we have changed the uncertainty of the prior simulated COS flux in twin experiments, and reperformed the experiments. Now, the uncertainty of the prior simulated COS flux was estimated as the standard deviation of the prior simulated COS fluxes within 24 hours around each simulation.

The benefit of the "multi-site" assimilation is unclear since it produces more or less similar results as the single-site assimilation. This is primarily due to using only two sites in this assimilation. The use of the word "multi" is thus exaggerated and I suggest leaving this part totally out of the paper, since it does not bring any notable improvement to the model. I understand that using only two sites is due to lack of in-situ COS flux measurements in similar ecosystems, but I don't really see a point doing a two-site assimilation since the results will be very similar to single-site assimilation.

Response: We appreciate the reviewer's understanding of the lack of *in situ* COS flux measurements in similar ecosystems. Therefore, we only performed a "multi-site" or "two-site" assimilation experiment at evergreen forest sites FI-Hyy and US-Wrc. 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. 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 did, however, replace the formulation "multi-site" by "two-site".

Also, we have extended the text in the conclusion: "**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.**" (Line 744-751)

I have several comments regarding the use of measured COS flux data:

- all sites: The authors do not specify any quality criteria used to filter the measured fluxes. Usually eddy covariance flux data are given a quality flag from 0 to 2; 2 indicating poor quality fluxes that should not be used, 1 indicating medium quality fluxes that are fine for budget calculations and 0 indicating the best quality that should be used for functional relationships and modelling. Please specify if you have used quality filtering in the data and if not, please give reasons why.

Response: Thanks for this comment. In the dataset for FI-Hyy (Vesala et al., 2022), No quality flags are provided, but measured COS fluxes as well as gap-filled COS fluxes are provided. In this study, only the measured COS fluxes are utilized and we have provided additional clarification on this (Line 260-261). For US-Ha1 and US-Wrc, no quality flag or gap-filled data is provided. At the remaining four sites, "COS filter" flag was provided to mark whether the COS observations are without flux detection limits. In this study, we do not use the detection limits to filter the COS flux data because such filtering would cause us to lose all values close to zero.

- US-Wrc: The dataset provided by Rastogi et al. 2018 does include the ready calculated gradient fluxes, and it is unclear why the authors are not using those fluxes but give a very ambiguous explanation of their own gradient flux parameter calculations. Moreover, since US-Wrc fluxes were calculated partly from the simulated COS fluxes in this study, this introduces a huge bias to these fluxes which gives even more reason not use this site in the "multi"-site assimilation.

Response: Thanks for this comment. The dataset (https://zenodo.org/records/1422820)

provided by Rastogi et al. (2018) **does lack** readily available gradient fluxes. Consequently, we implemented a bias correction to align the simulated and estimated COS fluxes for the US-Wrc site, drawing upon methodologies outlined in previous studies (Leung et al., 1999; Scholze et al., 2016). In addition, we have reached out to the corresponding authors via email to kindly request assistance in obtaining their readily-calculated flux data. Unfortunately, as of now, we have not received a response.

We acknowledge that the absence of precise COS flux data at US-Wrc poses challenges to our two-site assimilation experiments. Nevertheless, we maintain the importance of conducting two-site experiment, as detailed before.

- FI-Hyy: The dataset provided in Vesala et al. 2022 and Kohonen et al. 2022 already include storage corrected COS fluxes and it is not clear why the author have decided to do another storage correction for this site but not to other sites. In addition, this dataset includes gap-filled COS fluxes and it is not clear if the authors have used the gap-filled fluxes or the direct measured fluxes since the authors have not given any information on quality filtering.

Response: Thank you for pointing this out. We deleted the sentence: "We then corrected the COS fluxes from FI-Hyy using the storage-correction method (Kooijmans et al., 2017)." At FI-Hyy, only the direct measured COS flux data were utilized in the assimilation experiments, and we have clarified this (Line 260-261).

Simulation of sensible and latent heat fluxes as well as SWC seems quite out of place. Can you explain how COS fluxes should be related to sensible heat flux, and why assimilating COS fluxes should improve simulated sensible heat flux and soil moisture? Simulated sensible heat flux has even a different direction than the measured one. I suggest to leave this part out of the paper.

Response: Thanks for this comment. Reviewer #1 asked a similar question, please refer to our previous answer.

In this study, the diurnal variability of the simulated sensible heat fluxes using the BEPS model exhibited misalignment with observations, mainly at FI-Hyy. However, the simulated sensible heat showed good agreement with observations at the remaining sites. Moreover, the optimization of H was demonstrated successfully at FI-Hyy, despite the different direction of the simulated sensible heat and the measured one.

The abstract is too ambiguous and no concrete results are given. The authors use expressions "various processes" and "various ecosystems" without providing any details that would be useful for the reader.

Response: Thanks for your comment. We have deleted the expression "variable ecosystems" and listed the corresponding ecosystems of our study site in detail.

The authors need to mention in the method section if they use one-sided or all-sided LAI data, and if that applies everywhere in the paper or not. Also specify if negative fluxes mean uptake or emission. The word "significantly" is thrown around a lot, without any relation to statistical significance, it seems.

Response: Thanks for this comment. The leaf area index is commonly defined as half the total all-sided developed area of green leaves per unit ground surface area (Chen and Black, 1992; Liu et al., 2012; Xiao et al., 2016). In the publications listed in **Table 1**, only Kohonen et al. (2022) specified that the all-sided leaf area index (LAI) of FI-Hyy was ca. 8 m^2 m^{-2} during the measurement period (2013–2017). In this study, we followed the convention of using one-sided LAI (for broadleaves). We now have added "one-sided" (Line 99 and Line 1994) to account for this. In Sect. 2.4.3, we have specified positive values indicate COS uptake. Furthermore, we have corrected the inappropriate use of "significantly".

Section 2.1.3 needs to be rewritten, especially regarding the equations that are inconsistent and lacking information. Specifically:

- Where is Fcos,leaf used in the model? It is not present in any other equations after Eq. 3

Response: In eq.3, $F_{cos,leaf}$ represents the leaf-level COS uptake rate. For COS simulations, BEPS uses the leaf-level resistance analog model of COS (Berry et al., 2013) with a two-leaf upscaling scheme (Chen et al., 1999) from leaf to canopy.

- The authors need to explain where the different coefficients (e.g., 1.94 and 1.56 in Eq. 3; 1.4, 1.0, 5.33, -0.45 in Eq. 4; 0.437 and 0.0984 in Eq. 6; -0.00986, 0.197, -9.31 in Eq. 9; -0.119, 0.110, -1.18 in Eq. 10 , and 0.28 and 14.5 in Eq. 11) come from; what they represent and what is the reference.

Thanks for your comment, we have detailed the coefficients relevant to COS plant flux modeling (Eq. 3-6). For the COS soil model, we have updated them and detailed the coefficients currently used (please see Table S2 and Table S3 for details).

In NUCAS, the resistance analog model of COS plant uptake (Berry et al., 2013) were used. Such a model utilizes the COS mole fraction in the bulk air and the series conductance (conductance = 1/resistance) of the leaf system for COS (the terms in parentheses in Eq. 3) to calculate the flux of COS uptake. In the series conductance of the leaf system for COS, the stomatal conductance and laminar boundary layer conductance of COS are framed in reference to that of H_2O vapor. The greater mass and larger cross section of COS restricts its diffusion relative to H_2O in the stomatal pore by a factor of 1.94 and in the laminar boundary layer by 1.56 (Seibt et al., 2010; Stimler et al., 2010).

As for Eq. 5, we followed the modelling scheme of COS in the SiB (version 4.2) (Haynes et al., 2020), and we have provided additional clarification on this.

- What is fsw (how it is defined, is there an equation, what unit does it have and what kind of variation does it have) exactly.

Response: Thanks for your comment, we renamed it to f_w. In sect. 2.1.3, we mentioned f_w is a soil moisture stress factor describing the sensitivity of g_{sw} to soil water availability. We have added the definition of f_w to the appendix and also citations to the relevant literature, i.e. Ju et al. (2006).

- Vcmax; what is the unit and how do you get values (and which values) for it?

Response: The unit of V_{cmax} is μmol m^{-2} s^{-1}, we now added the detail calculation of V_{cmax} in the appendix.

- Fcos,biotic suddenly changes to FΘg in the switch from Eq. 7 to Eq. 8, if I got it right. Be consistent with the terms, as this is impossible to follow as a reader!! Also, where does Θi go in between these equations?? Is it switched to Θg?

Response: Thanks for this comment. To enhance readability, we have revised the manuscript to ensure consistency in terminology. In the soil COS model proposed by Whelan et al. (2016), The soil abiotic COS flux corresponding to a soil moisture of SWC_i can be calculated by Eq. 7 (Eq. 9 in the revised manuscript). In Eq. 7, SWC_{opt} denote the optimum soil moisture, at which soil abiotic COS flux reaches a maximum (F_{opt}), SWC_g denote a certain soil moisture, which is greater than SWC_{opt} and whose corresponding soil abiotic emissions are known. The last constant (a) that needs to be known in Eq. 7 can be calculated by Eq. 8 (Eq. 10 in the revised manuscript).

- How is "optimum soil moisture" defined? Optimum in terms of what?

Response: According to Whelan et al. (2016) and Whelan et al. (2022), there exists an optimum soil moisture at which the simulated biotic COS flux is maximized, i.e. optimum in terms of COS soil biotic uptake.

In general, there is lot of repetition throughout the paper and the text could certainly be condensed.

Response: Thank for your suggestion. We have thoroughly reviewed our manuscript and made refinements to the text.

Finally, I would like to see scatter plots in addition to the diurnal variation comparison, to better see how the model is able to simulate the COS fluxes and GPP.

Response: Thank for your suggestion. We now plotted the corresponding scatterplots and added them to the supplement.

Specific comments:

L19: "various processes" is too ambiguous

Response: Thanks for this comment. We have deleted the expression "variable ecosystems".

L25: "various ecosystems"; please specify which ecosystems

Response: we now specified the ecosystems, including evergreen needleleaf forest, deciduous broadleaf forest, C3 grass and C3 crop, respectively.

L26: "can significantly improve"; how much did it improve, which timescale, which ecosystem(s) etc?

Response: Thanks for this comment. Now we rewrite this sentence.

Comparing prior simulations with validation datasets, we found that the assimilation of COS can significantly improve the model performance in gross primary productivity,

sensible heat, latent heat and even soil moisture. (L26-L27)

L34: "carbon dioxide (CO2)" since this is the first time

Response: Corrected.

L 47-49: I don't really see a point in repeating the same references twice in the same sentence

Response: Thanks for this comment. We have revised the references in the manuscript.

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). (Line 47-51)

L55: Wohlfahrt et al 2012 and Kooijmans et al 2019 present an empirical model for leaf relative uptake (the uptake ratio of COS and CO2 at the leaf scale) but do not model COS flux itself

Response: Thanks for this comment. We now deleted these two references.

L58-60: This sentence is very unclear and I am not sure what the authors want to emphasize here.

Please rephrase

Response: Thanks for this comment. As mentioned earlier, a crucial hypothesis in this study is that the assimilation of COS is expected to improve the modelling of LE, H and SWC due to the ability of COS to indicate transpiration and the mechanism of transpiration. Therefore, here we would like to emphasize the second half of the sentence, i.e., only few experiments were conducted to systematically assessed the ability of COS to simultaneously constrain photosynthesis, transpiration and other related processes in ecosystem models. Of course, We also mentioned COS observations here (in the first half of the sentence). That is because the lack of COS measurements is for sure an essential limiting factor in examining the ability of COS to constrain ecosystem processes, such as photosynthesis and transpiration. At the same time, we also believe that the mention of observations here can also serve to pave the way for the introduction of data assimilation below. Therefore, we have rewritten the sentence while retaining the main content. The revised sentence now reads as: **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.** (Line 58-61)

L71-75: Please rephrase this sentence and preferably split it in two. At the moment it reads like Liu et al 1997 developed a model for simulating COS fluxes (which is not the case).

Response: Thank for this suggestion. We have split it in two:

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). (Line 72-76)

L78: Since you do not assimilate COS fluxes in all ecosystems existing, please specify which ecosystems you are talking about here

Response: Corrected.

L79: Controlling factors in which time scale of variability? E.g., in yearly scale temperature and radiation are for sure the most important drivers for carbon fluxes since they drive the seasonality, but this might not be the case in sub-daily time scales.

Response: Thanks for your comment. We have reorganized and revised that question and question one " What are the main changes in the parameters through the assimilation of COS flux and which processes are constrained?" The revised sentence reads as follows: **What parameters are the COS simulation sensitive to and how do these parameters change in the assimilation of ecosystem scale COS flux data?** (Line 78-79) **Which processes are constrained by the assimilation of COS and what are the mechanisms leading to adjustments of the corresponding process parameters?** (Line 82-83)

Response: Thanks for your comment.

L81: List the ecosystems

Response: Corrected.

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. (Line 85-88)

L96: all-sided or one-sided LAI?

Response: one-sided LAI.

L98: “phenology is driven by LAI” but isn’t it the other way around?

Response: The BEPS model (Liu et al., 1997; Chen et al., 1999) used in this study is a process-based diagnostic model driven by remotely sensed leaf area index (Chen et al., 2019). In BEPS, LAI is used as an indicator of the current state of vegetation within an ecosystem, and the plant phenology is driven by LAI. In contrast, in prognostic models, LAI is used as a dynamic variable that evolves over time, and the prognostic models allow researchers to make predictions about how LAI will change in response to varying environmental conditions and disturbances.

L103: remove one “of the”

Response: Corrected.

L148: “pmol/m2/s” -> should be pmol m^{-2} s^{-1} and units are not supposed to be written in italic. Check this everywhere in the paper, also with other units like m^2 m^{-2}

Response: Thank for this comment. we have corrected the units in this manuscript.

L153: “And the leaf-level” -> “The leaf-level”

Response: Corrected.

L156: Are the conductances different for shaded and sunlit leaves?

Uniform leaf laminar boundary layer conductance was applied to both shaded and sunlit leaves. However, BEPS takes into account radiation transmission processes (e.g., direction and scattering) within the canopy and calculates the amount of radiation received by the sunlit and shade leaves accordingly. Thus, the sunlit and shade leaves have different photosynthesis rates in theory due to the different radiation they receive, and in turn have different stomatal conductance (Ball et al., 1987; Ju et al., 2010).

L179: Do you perhaps mean Table S2?

Response: Yes, we have corrected the clerical error here.

L187-195: It is quite strange to cite here not the papers whose data you use but other papers from those same sites. Please cite the papers whose data you are using.

Response: Thanks for this comment. This arose from the fact that certain literature corresponding to the sites from which we obtained data lacked detailed site descriptions. We have addressed this by including references to the papers from which we sourced the data.

L189: ICOS is not defined (Integrated Carbon Observation System)

Response: Corrected.

L199-201: Specify that you use ecosystem scale eddy covariance (or gradient) flux measurements.

Response: Corrected.

Sect. 2.4.1: I don’t understand how the authors decided that the GLOBMAP LAI product was too low for the DK-Sor site but not for other sites. I did not find this information from Spielmann et al. 2019, as the authors claim. Please elaborate.

Response: Thanks for this comment. Mean LAI during the campaign of DK-Sor (referred to DBL in Spielmann et al. (2019)) was presented in Table S1 of the supplement in Spielmann et al. (2019).

L224-226: I am sure US-Ha1 site has some radiation data, at least PPFD data if not shortwave radiation, as well as air temperature and relative humidity. In-situ data is for sure better than the ERA5 data.

Thank you for your comment. We re-examined and collected the meteorological data of the

US-Ha1 site. As a FLUXNET site and an Ameriflux site, the meteorological data for the US-Ha1 can be found in both the Ameriflux and FLUXNET datasets, and both datasets does include some radiation data. However, the shortwave radiative data required by the BEPS model of US-Ha1 are only available at FLUXNET while only net radiation and PPFD data are available at Ameriflux. Considering the meteorological data of US-Ha1 provided by FLUXNET are only available in 1991-2012, we currently use FLUXNET data at US-Ha1 in 2012 and ERA5 shortwave radiation data with Ameriflux data in 2013 to drive the BEPS model.

L235: Table 1 does not list soil measurement information (and not the references either)

Thanks for your comment. Measurement information on COS soil fluxes already included in the literature we listed in Table 1 except for FI-Hyy. The reason we did not cite literature on soil COS flux observations at FI-Hyy (Sun et al., 2018) is that we assimilated ecosystem scale COS fluxes (Vesala et al., 2022) in this study. However, soil texture derived from the harmonized world soil database (Wieder et al., 2014) was used before. Now, we have updated the soil texture with *in situ* data and added relevant references (including Sun et al. (2018)).

L248-250: Now this is very confusing. In Kohonen et al 2020 the uncertainty is high with low absolute fluxes, the fact there is a stronger peak in negative fluxes is simply due to lack of observations of positive fluxes. In Kohonen et al. 2020 the negative fluxes are defined as uptake by the biosphere. In any case there should be no reason to remove either positive or negative fluxes, unless the quality criteria are not filled!

Thanks for your comment. Currently, we kept both positive and negative values of COS fluxes and re-ran the assimilation experiments.

L254: “gross primary productivity” -> “GPP”; “sensible heat” -> ”H”; “latent heat” -> “LE”

Response: Corrected.

L257: Cite Reichstein 2005 for the nighttime partitioning method

Response: Corrected.

L260: How is nighttime defined?

Response: In light of the extended daylight hours during the Northern Hemisphere summer and to prevent misclassification of actual daytime hours as nighttime due to discrepancies in local longitude and locally adopted time, we fit the equation for the relationship between respiration and temperature based only on data from 21:00 local time to 3:00 the following day.

L280: “And as a…” -> “As a..”

Response: Corrected.

L296: Do you really mean that only one set of model parameters is required, independent of the ecosystem type? I would assume e.g., Vcmax to be quite different for different ecosystems and PFTs.

Response: Thanks for your comment. We absolutely recognize that e.g., V_{cmax} varies greatly from ecosystem to ecosystem. In this study, we take the PFT- and texture-dependence of

parameters into consideration, thus the parameter number of one set of accurate and generalized model parameters is 76. In other words, the only one set of model parameters mentioned here, includes parameters that are specific to a PFT or texture but not to the point on the global that is populated by this PFT and characterized by this texture.

L307: I don't understand where the number 76 comes from. In table S2 there are 11 different parameters and their values are repeated as a constant value to get to 76, but there are certainly not 76 different parameters?

Response: The interdependence of parameters was considered in this study. Therefore, when counting the PFT-dependent parameters as well as the texture-dependent parameters, we multiply the number of PFTs and the number of textures considered in the BEPS model. This is how the number 76 is obtained.

L310: "correlation" -> "coefficient"

Response: Corrected.

L330: "dozens" -> please give an exact number

Response: We have modified this sentence with specific instructions.

L335-337: This sentence is too vague. Please be more specific.

Response: Thanks for your comment. We have reorganized the sentence: "**Corresponding to the PFT and soil texture of the experimental site, some PFT-dependent and texture-dependent parameters as well as global parameters showed different adjustments from others as they can affect the simulation of COS to different degrees.**"

L337-339: Where are these parameters used? Not in the COS model presented earlier

Response: We detailed how these parameters affect the simulation of COS in the appendix.

L353: 1.64% is very low, how do you explain that?

Response: As shown in the Figure 3j of the original manuscript, it is because the prior simulated COS at IT-Soy is already very close to the corresponding observations.

L357: Figure 3 comes in the text before Figure 2 is presented

Response: Corrected.

L360: Could this have something to do with the dry conditions and stomatal limitations, discussed in Vesala et al. 2022 regarding the low COS fluxes at FI-Hyy in July and August 2014?

Thanks for your comment. But according to Vesala et al. (2022), these months were not considered to be drought because the SWC remained at a normal level (well above 0.1 m^3 m^{-3}). However, the SWC observations as well as simulations in August 2014 are indeed noticeably lower than the other months, and are close to the optimum soil moisture for the COS abiotic flux modelling (see Figure S9 for details). As a result, the prior simulated COS for that month were significantly overestimated by 41.06 %, resulting in V_{cmax25} and VJ_slope being

considerable downward adjustments by -42.44 % and -41.03 % in the single-site experiments. 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. Regarding this, we have added the text in the manuscript: "**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.**" (Line 478-481)

L378: "for all experiments" -> not true for IT-Soy and US-Ha1!

Response: Corrected.

L385: Can this even be called an increase? In any case very low correlation coefficient.

Response: Yes, thus we say "R^2 remained almost unchanged by the optimizations".

L387: Why are the simulated nighttime fluxes unchanged?

Response: In the BEPS model, stomatal conductance was set to a constant value at night. Meanwhile, soil fluxes were small and less variable relative to the magnitude of plant COS flux.

L400: "due to high value of observation" or rather underestimation by simulation?

Response: Could, of course, be either, but according to Kooijmans et al. (2021), the air depleted in COS can then suddenly be captured by the EC system when turbulence is enhanced in the morning.

L412: I would not call two sites multiple sites….

Response: Now we changed our expression from 'multi-site' to 'two-site'.

L422: Can the ratio between PAR and SW really change that much? Why is it allowed to change so much?

Thanks for your comment. According to Ryu et al. (2018), the default f_leaf value in the BEPS model and the prior uncertainty of f_leaf in this study is overestimated. Thus, it tends to overshoot in the previous assimilation experiments. Now, we have computed the mean value of f_leaf with its standard deviation as an estimate of the error based on the MODIS PAR and SW data from 2012-2017 (Ryu et al., 2018) and re-ran the assimilation experiments.

L429: Either "In particular," or "Particularly"

Response: Corrected.

L444: "underestimated (by 55.72%), …"

Response: Corrected.

L444: "greatly increased"; how much?

Response: We have provided a quantitative description.

L445: "…simulations of COS flux at FI-Hyy.."

Response: Corrected.

L468: “forest sites (DK-Sor, FI-Hyy, US-Ha1, US-Wrc) compared to grassland and savanna (AT-Neu and ES-Lma)”

Response: Corrected.

L489-491: GPP cannot be observed directly, it is always a model!!

Response: Thanks for your comment. We know that GPP cannot be measured directly. In order to distinguish it from the modeled GPP of BEPS, we rephrase it to **GPP derived from EC measurements**.

L502: “excellent match” needs quantification

Response: Corrected.

L513-515: Not a very convincing result with the multi-site assimilation though

Thanks for your comment. Due to the lack of *in situ* COS observation data of the same PFT, we only conducted a two-site assimilation experiment. Therefore, we admit that the results of our experiments are not very convincing. More multi-site or two-site assimilation experiments would have helped us to get more statistically significant and plausible results, however we are faced with the challenge of lack of COS data.

L515-520: How would the results be without COS assimilation?

Response: the results be without COS assimilation, i.e., the prior simulation result can be found in Figure 4 and Figure 5 in the revised manuscript.

L523: It is not possible that there would not be sensible heat flux measured at a site where other eddy fluxes are measured, since it comes directly from the sonic anemometer used for wind measurements. If the authors have not published their sensible heat flux data, you can ask for it from the authors.

Response: Thanks for your suggestion. We have reached out to the corresponding authors via email to kindly request assistance in obtaining the sensible and latent heat flux data. With their assistance, we have conducted a thorough comparison and evaluation of H and LE simulations at the AT-Neu and IT-Soy sites. For the help they provided, we have added a note in the acknowledgements.

L525: “And the assimilation..” -> “The assimilation..”

Response: Corrected.

L536 & L554-556: Refer to the supplement figs

Response: Corrected.

L571: “not significant” by what metric? What is a “short period of time”?

Response: Thanks for your comment. Actually, this sentence is not necessary. We have therefore deleted it to avoid confusion.

L573: “almost no diurnal…” very vague, be more specific

Response: Thanks for your comment. We rewrite the sentence.

However, the simulated SWC exhibited a clear diurnal cycle whereas the observed SWC had almost no diurnal fluctuations. (Line 534-535)

L578-580: This is not really true, especially in the end of August (but other months are also underestimated)

Response: Thanks for your comment. We rewrote the sentence.

L583-585: Refer to the supplement figs

Response: Corrected.

L592: “COS fluxes of soil” -> “soil COS fluxes” or “COS fluxes from soil”

Response: Corrected.

Sect 4.1: Would it make sense to limit f_leaf and Vcmax25 variability to reasonable scales?

Response: Thanks for the comment. Since V_{cmax25} and f_leaf have their physical significance, the optimized values of both should be within certain ranges, e.g., greater than zero. Currently, both are within their physical significance, despite the huge relative change of them. The magnitude of the adjustment of f_leaf is expected to be limited by improving the estimation of its prior uncertainty. However, the prior uncertainty we set of the parameter V_{cmax25} is comparable to the existing dataset Chen et al. (2022). Furthermore, we have indeed refined the prior uncertainty of f_leaf and re-run the assimilation experiments.

L635: But since soil COS fluxes are low, wouldn’t that lead to higher change in the parameters, to compensate for low fluxes?

Response: Thanks for the comment. The optimized parameter values are the result of the trade-off between the two parts of the cost function. When the reduction in the discrepancy between observation and simulation resulting from the adjustment of the parameters is not sufficient to offset the increase in the discrepancy between the current and prior parameter values, the adjustment is not continued.

L652-655: Already mentioned in the previous section¨

Response: Removed.

L662: Could this be due to drought/ drier than normal conditions at FI-Hyy reported in Vesala et al. 2022?

Thanks for your comment. As shown in Table 3 of the original manuscript, f_leaf has been greatly downregulated after the assimilation of COS. We believe that this inappropriate parameter value is the main reason for the underestimation of posterior simulation. Now, we have refined the prior parameter uncertainty and re-ran the assimilation experiment.

L691: Table 1 perhaps?

Response: Yes, now we corrected this error.

L706: Which in-situ LAI data was used for FI-Hyy? Maybe the other one is all-sided and the other one-sided LAI?

According to Kohonen et al. (2022), the all-sided leaf area index (LAI) of FI-Hyy was ca. 8 m2 m−2 during the measurement period (2013–2017). In this study, we followed the convention of using one-sided LAI, so the LAI at FI-Hyy is 4 m^2 m^{-2}, as listed in **Table 1**.

L720: Start a new sentence "More laboratory…"

Response: Corrected.

L728: Why are the authors not already refining the uncertainty of prior values in this study?

Thanks for your comment. We have currently referred to the relevant literature and refined the prior uncertainty of the parameters (as mentioned before). Specifically, as the COS data utilized in this study range from 2012-2017, only the Moderate Resolution Imaging Spectroradiometer (MODIS) PAR and shortwave radiation (SW) data ranging from 2012-2017 was used to calculated the mean and standard deviation of f_leaf, and the prior uncertainty of f_leaf was estimated as the calculated standard deviation. The MODIS PAR and SW datasets are publicly available at: http://environment.snu.ac.kr.

L735-738: Given that this is already known, why is the COS concentration variation not already taken into account in this model?

Response: Continuous COS concentration data are a pre-condition for continuous COS flux simulations based on COS concentrations due to the linear relationship between the two (Stimler et al., 2011; Berry et al., 2013). However, similar to COS flux data, the *in situ* observed COS concentrations are not continuous in the whole assimilation windows. Therefore, in order to perform continuous simulations of COS flux based on a variable COS concentration, Kooijmans et al. (2021) used the surface COS mole fraction fields retrieved from an atmospheric transport inversion performed with TM5-4DVAR. We also think that modelling and assimilation of COS fluxes based on spatially and temporally varying COS concentrations is an aspect of the NUCAS system that can be further enhanced, and we will strive to combine the ecosystem model with atmospheric transport model to address this issue in our next steps. However, **with the lack of *in situ* COS mole fraction data**, COS mole fractions in the bulk air are currently assumed to be spatially invariant over the globe and to vary annually in NUCAS, which may introduce significant errors into the parameter calibration.

L749: Plants in lower rainfall conditions could also be e.g. CAM plants?

Response: Thanks for your comment. According to the summary of species information used in Yu et al. (2019), they do not include the crassulacean acid metabolism (CAM) plants in the study. However, the CAM plants are indeed commonly found in harsh environments such as arid and semi-arid regions (Amin et al., 2019), and the main feature of stomatal conductance patterns in CAM plants is nocturnal opening (Males and Griffiths, 2017).

Data availability section: Please include also citations to all datasets used

Response: Done.

Figure 1: How about mesophyll conductance? What does the dashed box represent?

Response: Thanks for your comment. In the resistance analog model of COS plant uptake (Berry et al., 2013), the apparent conductance for COS uptake from the intercellular airspaces (include the mesophyll conductance and the biochemical reaction rate of COS and carbonic anhydrase) is represented by g_{cos}. The dashed box includes the driver data of BEPS, and those data were utilized in both diagnostic process and prognostic process.

Figure 2: Are there any boundary values given to the parameters? How are these normalized? Add a similar plot from each site to same figure (as subplots) and put the figure to the supplementary material.

Response: We didn't set any boundary values for the parameters. Currently, they are normalized by their prior values. We have carefully considered showing the convergence trajectory through the parameter space from the starting point of the iterative procedure to the final point. In fact, this trajectory is to a large extent arbitrary, because branches depend on specifics of the floating-point arithmetic/rounding, which depend in turn on aspects like computing platform, compiler, or even compiler flags. What both technically and scientifically matters are the values of parameters, cost function and its gradient at the starting and end points of the minimization. These are now provided in Tables S5 for the twin experiments and S4 and 2 for the experiments with real data. We thus refrain from including the trajectory plots into the manuscript or its supplement, but provide the corresponding graphs and their presentation (requested by the reviewer) here:

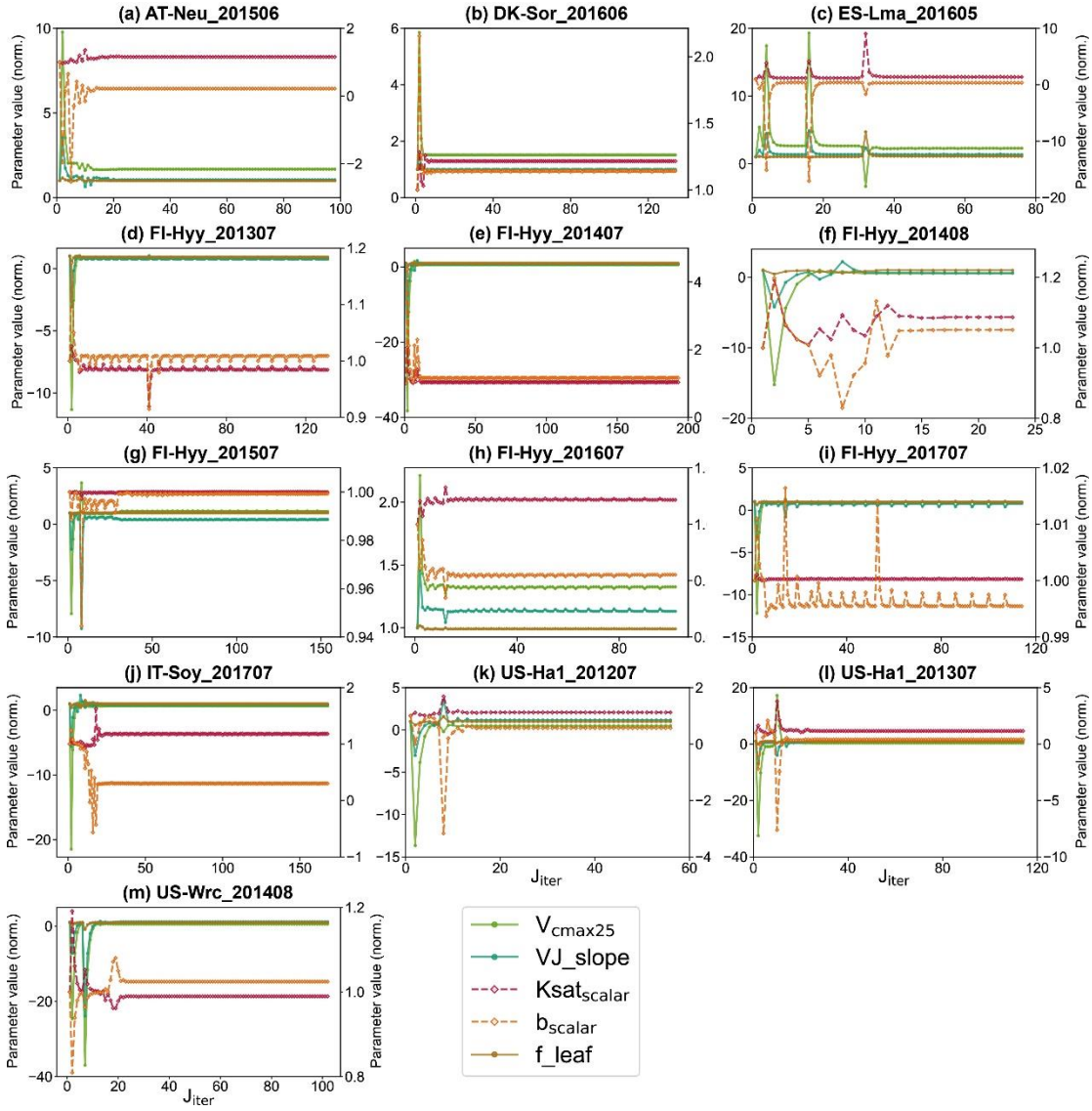


Figure 1. The evolution of model parameters with the number of iterations of cost function (J_{iter}) during the single-site experiments. Evolution (open carats and dashed lines) of soil texture dependent parameters is plotted on the right-hand y axis, evolution (filled circles and solid lines) of PFT-dependent parameters and global parameter is plotted on the left-hand y axis. Parameters are normalized by their prior values.

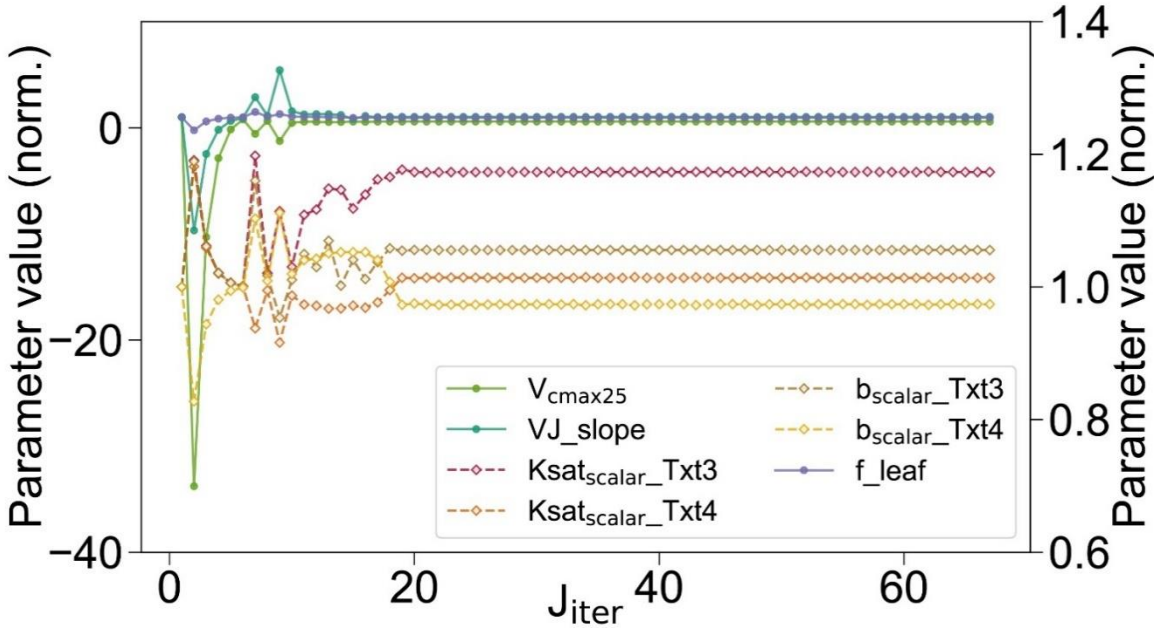


Figure 2. The evolution of model parameters with the number of iterations of cost function (J_{iter}) during the two-site experiment. Evolution (open carats and dashed lines) of soil texture (abbreviated as Txt) dependent parameters is plotted on the right-hand y axis, evolution (filled circles and solid lines) of PFT-dependent parameters and global parameter is plotted on the left-hand y axis. The texture-dependent parameters for FI-Hyy are denoted by "Txt3" and that of US-Wrc are denoted by "Txt4". Parameters are normalized by their prior values.

Corresponding to the PFT and soil texture of the experimental site, some PFT-dependent and texture-dependent parameters as well as global parameters showed different adjustments from others as they can affect the simulation of COS to different degrees. Those parameters are the maximum carboxylation rate at 25 °C (V_{cmax25}), the ratio of V_{cmax} to maximum electron transport rate J_{max} (VJ_slope), the scaling factors ($Ksat_{scalar}$ and (b_{scalar}) of saturated hydraulic conductivity (Ksat) and Campbell parameter (b), and the ratio of photosynthetically active radiation (PAR) to shortwave radiation (f_leaf). Particularly, as the soil textures at the FI-Hyy and US-Wrc are different, $Ksat_{scalar}$ and b_{scalar} corresponding to these two soil textures were both optimized in the two-site twin experiment.

Figure 3: I don't think these colors are color-blind friendly. Fig. 3 m: How is the RMSE in posterior lower, even though it looks worse than prior? Are the times presented here local time? For FI-Hyy the dataset is in local winter time (UTC +2). Please include the variability of the circle size (and what it means) to the figure legend. Why are you using mean instead of median diurnal variability?

Response: Thanks for your suggestion. We have modified the color scheme of our figures to make them easier to read for the color-blind. Certainly, the times presented here are local time. We have included the variability of the circle size in the legend in the revised manuscript. We use the mean because it is sensitive to all values.

Figure 4: I suggest to remove this fig with the whole “multi-site” analysis

Response: Thanks for your suggestion. For a detailed explanation of the need for two-site experiments we as well, refer to the previous section. Therefore, we’ve left the experiment in the main manuscript but changed to "two-site". Additionally, we also added the explanation of the need for two-site experiment in the revised manuscript. (Line 744-751)

Figure 5: Add in legend what the different colors mean. It is not clear from the caption what do the thick bars and the error bars represent.

Response: Corrected.

Figure 6: Same comments as for Fig. 5; you could combine these two figs in one as two different rows

Response: Thanks for your suggestion. We have combined these two figures in one as two different rows.

Figure 7: same comments as for Figure 3.

Response: Thanks for your suggestion. We will modify the color scheme of our figures to make them easier to read for the color-blind. Certainly, the times presented here are local time. We will include the variability of the circle size. We use the mean because it is sensitive to all values

Figure 8: Very weird pattern in simulated H. Solid and hollow circles are not distinguishable. I suggest to remove this fig with the analysis of H and LE.

Response: Thanks for this comment. The less effective simulation of H by the BEPS model compared to other variables, i.e. LE has been confirmed in previous studies (Ju et al., 2006). We acknowledge that the different direction of the simulated sensible heat and the measured one was observed at FI-Hyy. However, the optimization of H was demonstrated successfully, including at the FI-Hyy site. The connection between COS and latent and sensible heat, and the hypotheses of this paper have already been explained in the previous section and we have put the corresponding figures in the supplement.

Figure 9: Suggest to remove or move to supplement.

Response: Thanks for this comment. The connection between COS and SWC, and the hypotheses of this paper have already been carefully explained in the previous section, and we have put the corresponding figures in the supplement.

Figure 10: Not cited in the results section. What are “four LAI data”?

Response: Thanks for this comment. We have cited this figure in the results section and specified these four types of LAI data.

Table 1: Better reference to FI-Hyy would in this case be Vesala et al. 2022, since that paper presents the COS fluxes while Kohonen et al 2022 is about GPP.

Response: Thanks for this comment. We've changed the reference.

Table 4: Suggest to remove.

Response: Thanks for this suggestion. The necessity of conducting two-site experiment, we have already explained in detail above in this response and now also provide the explanation in the revised manuscript on lines 744-751.

Table S2: Not clear why the constant parameter values are repeated so many time

Response: Thanks for your comment. This is due to the fact that we take into account the interdependence of parameters, and we actually optimize the scaling factor of Ksat and b in this study. Regarding this, we have modified the table (**Table S4** in the revised supplement) and restated the description of the parameters.

References

Amin, A. B., Rathnayake, K. N., Yim, W. C., Garcia, T. M., Wone, B., Cushman, J. C., and Wone, B. W.: Crassulacean acid metabolism abiotic stress-responsive transcription factors: a potential genetic engineering approach for improving crop tolerance to abiotic stress, Frontiers in Plant Science, 10, 129, 2019.

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,

Berry, J., Wolf, A., Campbell, J. E., Baker, I., Blake, N., Blake, D., Denning, A. S., Kawa, S. R., Montzka, S. A., and Seibt, U.: 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, 2013.

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

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, 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.

Chen, B., Wang, P., Wang, S., Ju, W., Liu, Z., and Zhang, Y.: Simulating canopy carbonyl sulfide uptake of two forest stands through an improved ecosystem model and parameter optimization using an ensemble Kalman filter, Ecological Modelling, 475, 110212, 2023.

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. and Black, T.: Defining leaf area index for non‐flat leaves, Plant, Cell & Environment, 15, 421-429, 1992.

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., 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.

Dong, N., Prentice, I., Harrison, S. P., Song, Q., and Zhang, Y.: Biophysical homoeostasis of leaf temperature: A neglected process for vegetation and land‐surface modelling, Global Ecology and Biogeography, 26, 998-1007, 2017.

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.

Gupta, S., Ram, J., and Singh, H.: Comparative study of transpiration in cooling effect of tree species in the atmosphere, Journal of Geoscience and Environment Protection, 6, 151-166, 2018.

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.

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.

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., Maseyk, K., Seibt, U., Sun, W., Vesala, T., Mammarella, I., Kolari, P., Aalto, J., Franchin, A., and Vecchi, R.: Canopy uptake dominates nighttime carbonyl sulfide fluxes in a boreal forest, Atmospheric Chemistry and Physics, 17, 11453-11465, 2017.

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.

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., 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., 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.

Maignan, F., Abadie, C., Remaud, M., Kooijmans, L. M., Kohonen, K.-M., Commane, R., Wehr, R., Campbell, J. E., Belviso, S., and Montzka, S. A.: Carbonyl sulfide: comparing a mechanistic representation of the vegetation uptake in a land surface model and the leaf relative uptake approach, Biogeosciences, 18, 2917-2955, 2021.

Males, J. and Griffiths, H.: Stomatal Biology of CAM Plants Plant Physiology, 174, 550-560, 10.1104/pp.17.00114, 2017.

Monteith, J. and Unsworth, M.: Principles of environmental physics: plants, animals, and the atmosphere, Academic Press2013.

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.

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.

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.

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.

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.

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.

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.

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.

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.

Stimler, K., Berry, J. A., Montzka, S. A., and Yakir, D.: Association between carbonyl sulfide uptake and 18Ď 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., 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.

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.

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.

Wieder, W., Boehnert, J., Bonan, G., and Langseth, M.: Regridded harmonized world soil database v1. 2, ORNL DAAC, 2014.

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.

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.

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~~ related to various processes.~~. 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 ~~significantly~~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)~~, which makes~~. 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 ~~(Sandoval-Soto et al., 2005; Montzka et al., 2007; Campbell et al., 2008; Seibt et al., 2010; Stimler et al., 2010; Stimler et al., 2011)~~(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 ~~condition~~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 ~~with~~ 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 (~~Sandoval-Soto et al., 2005;~~ Campbell et al., 2008; Wohlfahrt et al., 2012; Berry et al., 2013~~; Kooijmans et al., 2019~~) 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, ~~due to~~with the lack of ecosystem-scale measurements of the COS flux (Brühl et al., 2012; Wohlfahrt et al., 2012; Kooijmans et al., 2021), ~~little experiments~~only few studies were conducted to systematically assess the ~~added value~~ability of COS ~~in~~to simultaneously ~~constraining~~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 ~~data assimilation system NUCAS (~~Nanjing University Carbon Assimilation System~~), that~~ (NUCAS) v1.0. NUCAS v1.0 is designed to assimilate multiple observational data streams including ~~the recently promising~~ COS flux data to improve the process-based ~~model Boreal Ecosystem Productivity~~Biosphere-atmosphere Exchange Process Simulator (BEPS) ~~(Liu et al., 1997),~~(Liu et al., 1997), which has been specifically ~~developed~~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~~ as follows~~:

What ~~are the main changes in the~~ parameters ~~through~~is the ~~assimilation of~~ COS ~~flux~~simulation sensitive to and ~~which processes are constrained~~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)?

~~What are the controlling factors of variability of carbon, water and energy exchange?~~

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 ~~multiple~~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 ~~make~~features render BEPS more advanced in representing ecosystem processes than standard ecosystem models (Richardson et al., 2012) ~~and~~ with less parameters to be calibrated ~~given~~owing to the LAI-driven phenology~~ is driven by LAI.~~.

~~By assimilating the observed data, NUCAS can achieve the optimization of the model process parameters and the model state variables of BEPS~~Data assimilation if performed in two sequential steps: First, ~~the BEPS model is run with default parameters and the model output is combined with COS flux observations to optimize the~~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~~ 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:

$$J(x) = \frac{1}{2}\left[(M_{cos}(x) - O_{cos})^T C_{cos}^{-1}(M_{cos}(x) - O_{cos}) + (x - x_0)^T C_x^{-1}(x - x_0)\right] \quad (1)$$

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

where MO and O denotes model and observationM denote vectors of observations and their modelled counterparts, respectively; x and x_0 denotes the control parameter vector with current and the prior control parameter vector; C denotesvalues, respectively. C_O and C_x denote the uncertainty covariance matrices for observations and prior parameters, and both. Both matrices are diagonal as we suppose 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 the prior and 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), and. 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.

Additionally, theThe minimization of the cost function is implemented in a normalized parameter space where the parameter values are specifiedmeasured 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 in spacebetween 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 ~~plant function types (PFTs) as well as eleven soil textures (see https://github.com/JChen-UToronto/BEPS_hourly_site). For NUCAS, we~~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 $F_{cos,plant}$ (~~$pmol/m^2/s$~~ 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:

$$F_{cos,plant} = F_{cos,sunlit} LAI_{sunlit} + F_{cos,shaded} LAI_{shaded} \tag{2}$$

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

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

~~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 H_2O vapor. 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. It can be calculated as:~~

$$\cancel{g_{cos} = 1.4 * 10^3 * (1.0 + 5.33 * F_{C4}) * 10^{-6} * \left(1 - e^{(-0.45*LAI)}\right) * f_{sw} * V_{cmax}} \tag{4}$$

~~where F_{C4} denotes the C4 plant flag, which takes the value of 1 when the vegetation is C4 plants and 0 otherwise. f_{sw} is a parameter describing the soil water stress on stomatal conductance.V_{cmax} denotes the maximum carboxylation rate.~~

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

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℃.

$$g_{COS} = \alpha * V_{cmax25} \tag{4}$$

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:

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

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).

$$F_{cos,soil} = F_{cos,abiotic} + F_{cos,biotic} \quad (5)$$

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

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

$$F_{cos,abiotic} = 0.437 * e^{0.0984*T_{soil}} \quad (6)$$

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

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):

$$F_{cos,biotic} = F_{opt}\left(\frac{\theta_t}{\theta_{opt}}\right) * e^{-a\left(\frac{\theta_t}{\theta_{opt}}-1\right)} \quad (7)$$

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

which can be rearranged to

$$a = ln\left(\frac{F_{opt}}{F_{\theta_g}}\right) * \left(ln\left(\frac{\theta_{opt}}{\theta_g}\right) + \left(\frac{\theta_g}{\theta_{opt}} - 1\right)\right)^{-1} \quad (8)$$

Here a is the curve shape constant, θ_t is the soil moisture (percent volumetric water content). The maximum biotic COS uptake F_{opt} and the biotic COS uptake F_{θ_g} are the COS fluxes ($pmol/m^2/s$) at optimum soil moisture θ_{opt} and θ_g, and can be calculated from T_{soil} using eqs. (9) and (10) respectively.

$$F_{opt} = -0.00986 * T_{soil}^{\ 2} + 0.197 * T_{soil} - 9.32 \quad (9)$$

$$F_{\theta_g} = -0.119 * T_{soil}^{\ 2} + 0.110 * T_{soil} - 1.18 \quad (10)$$

θ_g is assumed to be a constant 0.35, and θ_{opt} is assumed to be a first order function of T_{soil}.

$$\theta_{opt} = 0.28 * T_{soil} + 14.5 \quad (11)$$

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

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 $SWC_g \geq SWC_{opt}$. 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

~~In this study, we optimized a total of 76 parameters belonging to BEPS, the parameters are described in **Table S3**. 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 with efforts from previous model development and validation, and the prior uncertainty of parameters is set as 25% of the prior values.~~

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

~~The NUCAS was evaluated at seven sites distributed on the Eurasian and North American continents in boreal, temperate and subtropical regions based on field observations collected from several studies. Those 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 5 of the 10 PFTs used in BEPS and 5 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); the Danish ICOS RI site (DK-Sor), which is dominated by European beech (Braendholt et al., 2018); the Las Majadas del Tietar site (ES-Lma) located in western Spain with a Mediterranean savanna ecosystem (El-Madany et al., 2018); the Hyytiälä forest Station (FI-Hyy), located in Finland and is dominated by Scots Pine (Bäck et al., 2012); an agricultural soybean field measurement site (IT-Soy) located in Italy~~ 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. Those 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)~~(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. ~~And the~~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 ~~moisture~~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. ~~Overall~~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). ~~And all of the three~~The other two LAI products were used to ~~drive the model to~~ investigate the effect of the LAI products on the parameter optimization results. ~~According~~Also, according to Spielmann et al. (2019), the GLOBMAP product had ~~significantly~~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 ~~is available through~~were taken from the FLUXNET database (AT-Neu, DK-Sor, ES-Lma ~~and~~, 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. ~~In addition,~~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 ~~US-Ha1 were also derived from ERA5 since there are no in situ~~FLUXNET data in 2013 and the lack of shortwave radiation ~~measurements at this site.~~ 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. ~~We then corrected the COS fluxes from FI-Hyy using the storage-correction method (Kooijmans et al., 2017).~~ 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 ~~on the~~

~~observations of COS can be found in the publications listed in **Table 1**~~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 ~~significant~~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:

$$\cancel{C = \frac{\sigma_s(c - m_c)}{\sigma_c} + m_s \quad (12)}$$

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

where ~~c denotes the COS flux observations (converted to $pmol/m^2/s$). m_c~~$\bar{O}$ and ~~σ_c~~σ_O are mean and standard deviation of the observed COS flux series. ~~C~~F is the corrected observed COS flux, which is matched to the simulated COS flux. ~~m_s~~$\bar{M}$ and ~~σ_s~~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.

~~Considering that COS soil fluxes are much lower than the anticipated plant fluxes in general (positive values indicate COS uptake) and that the relative uncertainty in COS fluxes is very large at low values, especially when negative (Kohonen et al., 2020), we first removed the negative values of the ecosystem COS fluxes. Then, the~~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.

~~In order to evaluate the assimilation results, gross primary productivity, sensible heat, latent heat and volumetric soil water content (SWC) observations were also taken from FLUXNET (DK-Sor, ES-Lma and FI-Hyy), AmeriFlux (US-Ha1 and US-Wrc),~~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. (2019)~~Spielmann et al. (2020) and Spielmann et al. (2019) for AT-Neu and IT-Soy). As GPP is only available for FLUXNET sites, and CO_2 turbulent flux (FC) or net ecosystem exchange (NEE) data are available for other sites, 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~~ (Reichstein et al., 2005), and thus partitions FC or NEE into GPP and R_{eco}~~, 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 ~~multi~~two-site assimilation experiment was carried out to refine one set of parameters over ~~multiple~~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 ~~multi~~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 ~~sequence~~time series included the prior simulated ecosystem COS fluxes with its uncertainties, and the latter were ~~set to a constant~~estimated as the standard deviation of ~~1 $(pmol/m^2/s)$.~~the prior simulated COS fluxes within 24 hours around each simulation. Then, a given perturbation ratio was applied to the prior parameters vector, ~~and a perturbed ecosystem COS simulation sequence could be obtained based on the perturbed~~as a starting point for the interactive adjustment of parameter ~~vector. Finally, the data assimilation experiments were performed to minimize the discrepancy between the prior parameters and the perturbed parameters, and thus the discrepancy between~~values to match the COS flux pseudo-observations~~ and the perturbed ecosystem COS simulations~~. The effectiveness of the data assimilation methodology of NUCAS can be validated if it successfully restores the control parameters from the pseudo-observations. ~~And as~~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 ~~multi~~two-site twin experiment. ~~For all cases where~~With reference the ~~PFT is evergreen needleleaf forest~~uncertainty of parameters, a perturbation ratio of 0.2 was ~~used. And for~~utilized in all of the ~~remaining six single-site~~ twin experiments~~, a perturbation rate of 0.4 was used~~.

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. ~~In the diagnostic processes, no perturbation was applied to the default parameters, except for the experiment conducted at the FI-Hyy site in July 2017, where a perturbation ratio of 0.2 was applied.~~ Detailed information about those single-site experiments is shown in **Table ~~3~~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, ~~multi~~a two-site assimilation experiment that can assimilate COS observations from ~~multiple~~two sites simultaneously is necessary to be conducted. ~~Across the seven sites,~~ Although both DK-Sor and US-Ha1 are ~~both~~ dominated by deciduous broadleaved forest, ~~while there is no overlap in the timing~~and both AT-Neu and ES-Lma are dominated by C3 grass, none of the ~~observations for their~~ 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 ~~multi~~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.

$$\cancel{D_x = \|x - x_0\| = \sqrt{\sum_{i=1}^{n}\left(x(i) - x_0(i)\right)^2} \quad (13)}$$

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

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 ~~correlation~~coefficient of determination (R^2) are used to measure the difference between the simulations and *in situ* observations. They were calculated as:

$$\cancel{MB = \frac{1}{N}\sum_{i=1}^{N}(obs_t - sim_t) = \overline{obs} - \overline{sim} \quad (14)}$$

$$\cancel{RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(obs_t - sim_t)^2} \quad (15)}$$

$$\cancel{R^2 = 1 - \frac{\sum_{i=1}^{N}(obs_t - sim_t)^2}{\sum_{i=1}^{N}\left(obs_t - \overline{obs}\right)^2} \quad (16)}$$

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

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

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

where ~~"*obs*" and "*sim*" denote the observations and simulations, respectively. sim_t~~M_i denotes the simulation corresponding to the *i* th observation ~~obs_t. The terms $\overline{obs}$ and $\overline{sim}$ are the mean of observations and the mean of simulations corresponding to the observations.~~ O_i and N is the total number of observations.

~~Given the large variation in the magnitudes of simulations and observations across experiments, the coefficient of variation of RMSE (CV(RMSE)) was employed to compare the assimilation results between different experiments, and it was calculated by normalizing the RMSE using the mean of observations.~~

$$CV(RMSE) = \frac{RMSE}{\overline{obs}} \quad (17)$$

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

$$\Phi(var) = \frac{\partial J/\partial x_0(var)}{\|\partial J/\partial x_0\|} \quad (18)$$

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

where $\|\partial J/\partial x_0\|$ denote$\|\partial J/\partial x\|$ denotes the norm of the sensitivity vector of the cost function to the model parameters at the prior values.

3 Results

3.1 Twin experiments

After dozens ofaveraging 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 2S5**. In summary, during those assimilations, the cost function values were significantlysubstantially reduced by more than sixteen orders of magnitude, from greater than 4.58×10^{3} 50.75 to less than $3.50$5.09 $\times 10^{-13}$ and the respective gradient values also reduced from greater than 3.94×10^{3}38.81 to less than $2.79$1.59 $\times 10^{-4}$$10^{-6}$, which verified the ability of the data assimilation algorithm to correctly complete the assimilation.

Corresponding to the PFT and soil texture of the experimental site, some PFT-dependent and texture-dependent parameters as well as global parameters showed different adjustments from others as they can affect the simulation of COS to different degrees. Those parameters are the maximum carboxylation rate at 25 °C ($V_{cmaxz25}$), the ratio ofThe relative changes of the V_{cmax} to maximum electron transport rate J_{max} (VJ_slope), saturated hydraulic conductivity (Ksat), Campbell parameter (b), and the ratio of photosynthetically active radiation (PAR) to shortwave radiation (f_leaf). Particularly, as the soil textures at the FI-Hyy and US-Wrc are different, Ksat and b corresponding to these two soil textures were both optimized in the multi-site twin experiment. The relative changes of those 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 S3S5**. Results show that the relative differences of those parameters from the "true" values reached veryexceedingly small values at the ends of twin experiments, with the maximum of the absolute values of the relative changes below $2$8.55 $* 10^{-8}$. 10^{-9}. D_x was also reduced to nearly zero with the maximum value below $2$6.60 $* 10^{-7}$$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 118113 cost function evaluations, all of the 13 single-site experiments were performed successfully. The experiments reduced cost function values significantlysubstantially, with an average cost function reduction of 33.7824.43 % (**Table 32**). However, the minimization efficiency cost function reduction of the experiment varies considerably with PFT, site and assimilation window, ranging from 1.644.87 % to 64.9269.05 %. The single-site assimilations tend to achieve greater minimization efficiencycost function decreased dramatically at the deciduous broadleaf forest sites

~~and the evergreen needleleaf forest sites~~US-Ha1, with ~~mean minimization efficiency of 42.74% and 42.39%, respectively. For the other three PFTs, i.e. grass, crop and shrub, the minimization efficiencies were quite small, ranging from 1.64% to 10.48%, as the simulations of COS using the default parameters at these three sites are already very close to the corresponding observations (**Figure 3**). We found that for different sites with the same PFT, their~~an average ~~minimization efficiencies of the assimilation are in good agreement. However, for the same site, the minimization efficiencies varied considerably~~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~~, yet were very similar for the same year. For example, at FI-Hyy, the cost function reduction in~~ . In July and August 2014~~ were almost identical, with 62.23% and 64.92% respectively, both much greater than~~, the cost function reduction ~~rates~~were as high as 40.59 % and 50.94 %, while in other years~~.~~

~~For all single-site experiments, the model parameters were continuously adjusted during the assimilation and eventually stabilized.~~, 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~~. **Figure 2** illustrates the evolution of the values of those parameters during the single-site assimilation experiment at the DK-Sor site in June 2016. At the beginning of the assimilation, each parameter had a great adjustment. As the iterations continued, the parameters gradually stabilized and the minimization was eventually completed. Specifically, V_{cmax25}, VJ_slope and f_leaf varied over a very large range during the assimilation, up to 47.92 in the normalized model parameter space. In contrast, the texture-dependent parameter Ksat and b, varied in a very small range between 3.99 and 4.01.~~ (**Table 2**).

~~**Figure 3** illustrates the~~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 ~~very~~remarkably close, with ~~21.92 $pmol/m^2/s$~~ 20.60 pmol m^{-2} s^{-1} and 21.~~88 $pmol/m^2/s$,~~01 pmol m^{-2} s^{-1} respectively. However, there was substantial variability between sites and even between experiments at the same site. At ~~DK-Sor~~ES-Lma, the prior simulated COS fluxes were greatly underestimated by ~~55.72~~63.38 %. In contrast, the prior simulated COS fluxes were overestimated at ~~FI-Hyy, while the overestimation is only significant in 2014~~US-Ha1, with MBs of ~~11.59 $pmol/m^2/s$~~ -10.01 pmol m^{-2} s^{-1} and ~~8.34 $pmol/m^2/s$~~ -13.63 pmol m^{-2} s^{-1} in July 2012 and ~~August respectively.~~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 ~~simulated COS fluxes during nighttime were almost constant and lower than the observations for all experiments. Moreover, the~~ underestimation is particularly evident ~~in~~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.~~69 $pmol/m^2/s$~~49 pmol m^{-2} s^{-1} in the prior case to 13.~~64 $pmol/m^2/s$~~86 pmol m^{-2} s^{-1} in the posterior case. ~~And similar~~Similar to the ~~values of~~ 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 ~~significantly~~ decreased from 6.94 ~~$pmol/m^2/s$~~ pmol m^{-2} s^{-1} to 3.~~84 $pmol/m^2/s$.~~09 pmol m^{-2} s^{-1}. In contrast, R^2 remained almost unchanged by the optimizations, with its mean value ~~increasing slightly from~~of 0.~~2956 to 0.3037. In addition,~~2967 in the prior case and 0.2970 in the posterior case. Our results also ~~demonstrate~~showcase that the ~~assimilation~~ model-observation differences of COS ~~mainly optimizes the simulated COS fluxes~~were effectively reduced at daytime~~, while~~. However, the ~~simulated nighttime COS~~

~~fluxes~~remarkable differences between COS observations and simulations at nightime, are ~~almost unchanged.~~ not effectively corrected in a number of assimilation experiments (i.e., the experiment conducted at FI-Hyy in July 2013, see **Figure 3d**). ~~The impacts of the assimilation of COS in improving the COS posterior simulations were particularly evident at forest sites, where the prior simulated COS often deviated significantly from the observations, and less evident at low-stature vegetation (including grass, crop and shrub) sites, as the model using prior parameters already performed very well in the simulations. This result is very reasonable since a similar pattern was also found in the cost function reductions at these sites. For example, with the largest cost function reduction, the assimilation of COS significantly corrected the overestimation of the COS simulations at FI-Hyy in August 2014, with RMSE decrease from 16.13 $pmol/m^2/s$ to 10.11 $pmol/m^2/s$. In contrast, with a reduction in the cost function of only 2.08%, the assimilation of COS had little effect at the IT-Soy site, where the RMSE of simulated and observed COS only decreased from 12.23 $pmol/m^2/s$ to 12.10 $pmol/m^2/s$. In addition, the performance of the assimilation of COS at these sites was evaluated utilizing CV(RMSE). Results showed that the three experiments with the smallest CV(RMSE)s all were carried out at the FI-Hyy site, in July 2013, 2016 and 2017 respectively, with a mean value of CV(RMSE) of 0.51. While at AT-Neu and US-Wrc, the CV(RMSE)s were much larger, with 0.90 and 0.85 respectively. For AT-Neu, in addition to the large model-observation biases during nighttime (**Figure 3a**), there were also significant deviations between observations and simulations in the morning due to the high values of observations.~~

3.3 ~~Multi~~Two-site assimilation

FI-Hyy and US-Wrc have different soil textures, with ~~loamy sand and silty~~sandy loam and loam, respectively. In the ~~multi~~two-site assimilation experiment, NUCAS took this difference into account and successfully minimized the cost function from ~~703.36~~495.94 to ~~370.44~~365.63 after ~~146~~67 evaluations of cost function. The cost function reduction for the experiment ~~is very reasonable, with~~has a value of ~~47.33~~28.29 %, comparable to the cost function reductions for corresponding single-site assimilation experiments at FI-Hyy and US-Wrc (~~64.92~~50.94 % and ~~44.65~~27.71 %). Furthermore, corresponding to these two soil textures, the texture-dependent parameters ~~Ksat~~$Ksat_{scalar}$ and ~~b~~b_{scalar} yielded two different posterior parameter values, respectively, so that a total of seven parameters were optimized in the ~~multi~~two-site experiment (**Table ~~4~~). ~~Table 4~~ ~~shows~~3).** It can be seen that ~~with~~ the ~~exception of~~two-site optimized results of V_{cmax25}. VJ_slope~~, the multi-site posterior parameters~~ and f_leaf are ~~all very~~ similar to ~~those~~that of the single-site ~~experiments in both the sign of~~optimized results at US-Wrc, as most of the observations of the two-site experiment originated from US-Wrc. As for the ~~change (increase or decrease) and~~ 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 ~~minimization efficiencies~~cost function reduction and the parameter optimization results of the ~~multi~~two-site assimilation experiments were ~~very~~ similar to the corresponding single-site experiments, demonstrating the ability of NUCAS to correctly perform joint data assimilation from COS observations at ~~multiple~~two sites simultaneously.

The posterior simulations of COS flux using the ~~multi~~two-site posterior parameters, also demonstrated the ability of NUCAS to correctly assimilate ~~multi~~two-site COS fluxes simultaneously~~.~~ **(Figure 4** and **Figure S2).** As shown in **Figure ~~4~~4a**, the prior COS simulations for both the FI-Hyy site and US-Wrc site show overestimation compared to the observations. ~~However, after~~After the ~~multi~~two-site COS assimilation, the discrepancies between COS simulations and observations were ~~significantly~~ reduced in both FI-Hyy and US-Wrc, with RMSE reductions of ~~36.86~~24.75 % and ~~9.27~~3.39 %, achieving similar results to the simulations using the single-site posterior parameters.

3.4 Parameter change

As mentioned before, there were ~~only~~ five parameters that have been ~~significantly changed~~adjusted during the assimilation of COS flux observations by the NUCAS system, whether in twin, single-site or ~~multi~~two-site experiments. They are the maximum carboxylation rate at 25 °C (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 ~~provide~~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.

~~For both single-site and multi-site experiments, the changes of those five parameters exhibited different characteristics: The texture-dependent parameters Ksat and b had a very little relative change, while the PFT-specific parameters (V_{cmax25} and VJ_slope) and f_leaf changed dramatically (**Figure 5**). In particularly, the experiment with the largest relative change of Ksat and b performed in July 2017 at FI-Hyy, showed the corresponding relative change of only 1.33% and -2.08% respectively. For other experiments, the relative changes of Ksat and b were much smaller, on average 0.09% and 0.14%, respectively of their absolute values. In contrast, the other three parameters varied considerably after the assimilations, in particular f_leaf, which decreased by 31.55% on average in the single-site experiments. However, among these posterior parameters, V_{cmax25} has the greatest variability, with relative changes ranging from -60.64% to 113.45%.~~

~~Across all single-site experiments, there were significant differences in the results of parameter optimization between sites. We found that for those sites where the prior simulations of COS were already very close to COS observations, such as AT-Neu, ES-Lma and IT-Soy, there are still some parameters that varied significantly in the assimilation experiments. For example, in the experiment conducted at AT-Neu, although the cost function reduction of this experiment was only 1.64%, both V_{cmax25} and VJ_slope were changed significantly, with the relative changes of 45.54% and -45.42% respectively. With the opposite directions and similar magnitudes, the relative changes in V_{cmax25} and VJ_slope are very reasonable, and reflect the trade-off of the assimilation system for the parameters which ensured the posterior simulated COS fluxes are still close to the COS observations. For those sites where the prior COS simulations deviated considerably from the observations, the relative changes of the posterior parameters were relatively larger. At DK-Sor, where the prior simulations of COS were significantly underestimated by 55.72%, both V_{cmax25}, VJ_slope and f_leaf have been greatly increased in the assimilation. In response to the apparent overestimation in the prior simulations of COS at FI-Hyy, the posterior COS plant uptake related parameters showed an overall decrease, especially f_leaf.~~

~~In the multi-site experiment, corresponding to the different soil textures of FI-Hyy and US-Wrc, two different posterior parameter values were obtained for the texture-dependent parameters Ksat and b respectively, while only one posterior parameter value was obtained for each of other parameters. The results show that the posterior values of V_{cmax25} and txt-dependent parameters obtained from the multi-site optimization are very similar to those from the single-site optimization both in terms of the sign and the magnitude of adjustments. However, with a relative change of 30.72% and -63.64% in the multi-site experiment, the posterior VJ_slope and f_leaf were significantly larger and smaller than those in the single-site experiments, respectively.~~

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 ~~f_leaf and~~ VJ_slope~~, while Ksat and b do not influence the assimilation results significantly (**Figure 6**).~~. With absolute ~~sensitivity coefficients~~SIs ranging from ~~89.06~~88.47 % to ~~97.39% except at IT-Soy,~~96.41 %, the mean absolute ~~sensitivity coefficient~~SI of V_{cmax25} is more than three times that of VJ_slope~~ and f_leaf~~, which are ~~24.71% and 28.76% respectively.~~ 27.67 %. In contrast, ~~for the texture-dependent parameter Ksat and b, their~~ the average absolute ~~sensitivity coefficients were only 0.01%~~SIs of b_{scalar}, f_leaf and ~~0.02%,~~$Ksat_{scalar}$ are much lower, with 11.13 %, 8.30 % and 2.96 % respectively.

Unlike the great variability of the posterior ~~COS plant uptake related parameters~~V_{cmax25} and VJ_slope, the ~~sensitivities~~SIs of ~~the cost function to those~~these two parameters are ~~very~~ stable~~ (except IT-Soy),~~ , especially at the same site. At US-Ha1, for example, the difference between the ~~sensitivity coefficients~~SIs of V_{cmax25}~~,~~ and VJ_slope ~~and f_leaf~~ in its two experiments were all smaller than 0.~~57%. Among the three parameters,~~ 54 %. Furthermore, V_{cmax25} has the smallest magnitude of variation in ~~sensitivity coefficient (except IT-Soy), only about half that of VJ_slope and f_leaf, although its sensitivity coefficients~~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. ~~As for Ksat and b, despite the small values of their sensitivity coefficients,~~With the ~~relative variability is large, with sensitivity coefficients~~SIs ranging from ~~-0.05~~20.62 % to 33.78 % and 4.17 % to ~~0.04 and from -0.03% to 0.07% respectively.~~

~~Our results also suggest that~~11.99 % (with the ~~parameters related to light reaction (~~exception of DK-Sor), VJ_slope and f_leaf~~),~~ ~~tend to~~ also play ~~more~~ important roles in the ~~COS assimilation at the forest sites compared to AT-Neu and ES-Lma, while V_{cmax25} does the opposite. However, the smallest absolute $\Phi_{V_{cmax25}}$ was found at the agricultural site IT-Soy with a value of only 23.76%, yet its sensitivity coefficient of f_leaf is as high as 94.97%.~~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 ~~very~~ well in modelling the daily variation and diurnal cycle of GPP, with mean R^2 of 0.~~76~~80 and 0.~~75~~78, respectively~~.~~ (**Figure 7** and **Figure S3**). The discrepancy between simulations and observations was ~~significantly~~substantially reduced by the assimilation of COS, from mean RMSE of ~~8.22 $\mu mol/m^2/s$~~7.43 umol m^{-2} s^{-1} in the prior case to ~~6.38 $\mu mol/m^2/s$~~5.34 umol m^{-2} s^{-1} in the posterior case ~~(**Figure 7**). The mean bias between the observed and simulated GPP was also corrected with the reduction in mean absolute MB from 4.82 $\mu mol/m^2/s$ to 3.14 $\mu mol/m^2/s$.~~

. Similar to COS ~~flux~~, the mean of prior simulated GPP is also generally larger than the observed. ~~We found that~~With the assimilation of COS, the ~~tuning directions of the GPP simulations and the COS simulations were consistent for almost all single-site experiments (12/13). The only exception occurred at AT-Neu, with the simulated COS increasing by 10.32% while the simulated GPP decreasing by 15.24%. Such results also reflect that the sensitivity of COS exchange and photosynthesis to the model parameters differs due to the different physiological mechanisms.~~bias between the observed and simulated GPP was effectively corrected, with the reduction in mean absolute MB from 4.31 umol m^{-2} s^{-1} to 2.28 umol m^{-2} s^{-1}.

In general, the GPP performance was improved for most of the single-site experiments (~~9~~12 of 13), with RMSE reductions ranging from ~~9.41% to 59.83%, while for the other 4 experiments, the posterior RMSEs were slightly higher than the prior by 0.84% to 23.96%. More specifically, across~~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 ~~37.92~~42.00 %. At the ~~sites that were dominated by~~ 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 ~~11.99~~20.95%. However, for experiments conducted on ~~other~~ low-stature vegetation types (including C3 grass~~,~~ and C3 crop ~~and shrub), the~~), 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 ~~larger~~lower than the prior~~. Nevertheless~~, with reduction ratios of 8.60 % and 3.81 %, respectively. At AT-Neu, the ~~posterior simulations of~~addition of COS observation shifted the GPP ~~for these three sites also achieved a consistent fit to~~simulations away from the GPP derived from EC observations, with ~~their CV(~~the RMSE~~)s all smaller than the averaged CV(RMSE) of all posterior simulations in single-site experiments. Moreover, for AT-Neu and~~ increasing from 3.48 umol m^{-2} s^{-1} to 5.97 umol m^{-2} s^{-1}. ~~IT-Soy, the GPP observations exhibited significant fluctuations even at night, suggesting that they may have large uncertainties, which is to be considered in the evaluations of our GPP simulations.~~

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~~ overestimated GPP~~, by ~~almost the same degree, 30.58~~21.26 % and ~~34.58~~42.02 % respectively~~, while~~. With the assimilation of COS, the ~~corresponding posterior simulated GPP differs considerably.~~modelled COS exhibited substantial decreases. In ~~July 2012~~parallel, the model~~ using~~ observation difference also reduced, by 12.36 % and 24.46 %, respectively. However, the posterior ~~parameters performed very well in GPP simulations, with MB of only 0.20 $\mu mol/m^2/s$. In contrast, the posterior GPP simulations in July 2013 were significantly~~simulated GPP appeared to be underestimated~~, with MB of -6.38 $\mu mol/m^2/s$.~~. 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 ~~$\mu mol/m^2/s$~~ umol m^{-2} s^{-1} to 9.15 ~~$\mu mol/m^2/s$, while~~

umol m^{-2} s^{-1}. In August 2014, the ~~mean for August of 6.43 $\mu mol/m^2/s$ was~~ GPP observations were noticeably lower than that in July~~.~~, with a mean of 6.43 umol m^{-2} s^{-1}. As for simulations, the ~~prior simulations tend~~model tends to overestimate GPP, with MBs ranging from ~~3.76 $\mu mol/m^2/s$~~2.79 umol m^{-2} s^{-1} to ~~6.61 $\mu mol/m^2/s$. However,~~ 5.25 umol m^{-2} s^{-1}. After the ~~posterior GPP differs considerably, in some experiments achieving excellent match with the observations and other experiments yielding very low simulated GPP. In July 2013, 2015 and 2016, the model using posterior parameters performs well in simulating GPP and achieves the smallest CV(RMSE)s~~assimilation of ~~all single-site experiments~~COS, the overestimation of the COS simulation for FI-Hyy were effectively corrected, with ~~CV(RMSE)s ranging from 0.39 to 0.42. In contrast, as the observed COS is lower than the prior simulated COS by 39.64% and 39.32% in July and~~ the mean absolute MBs of 1.01 umol m^{-2} s^{-1}. However, with a low SWC in August 2014, ~~f_leaf and V_{cmax} were dramatically adjusted downwards in July and~~ 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 ~~respectively,~~2014, ultimately resulting in ~~notable~~the underestimation ~~in~~of the single-site posterior simulated GPP~~, with MBs of -6.27 $\mu mol/m^2/s$ and -2.57 $\mu mol/m^2/s$. In addition, a dramatic reduction of f_leaf was also reported in July 2017 and resulted in an underestimation of posterior simulated GPP.~~ .

In the ~~multi~~two-site experiment, the ~~posterior~~ model-observation differences ~~for~~of GPP ~~were reduced~~ for both FI-Hyy and US-Wrc were reduced by the assimilation of COS, with RMSE reductions of ~~45.85~~39.90 % and ~~55.71~~42.69 %, respectively. These RMSE reductions are even higher than those in the corresponding single-site experiments, by ~~20.34~~55.08 % for FI-Hyy and ~~7.84~~16.31 % for US-Wrc. These results suggest that simultaneous assimilation using COS observations from ~~multiple~~two sites can also improve GPP simulations, and the assimilation ~~is sometimes~~can be more ~~effective~~robust than the single-site assimilation because the possibility of over-fit local noise is reduced.

Overall, the assimilation of ecosystem COS flux data ~~can improve~~improved the simulation of GPP in both single-site ~~assimilation~~experiments and ~~multi~~the two-site ~~assimilation~~experiment. However, the assimilation effects vary considerably for different sites and even for different periods within the same site. ~~The~~ Our results suggest the assimilation of COS ~~degrades the fit~~is able to ~~observed GPP at~~ 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 ~~sites~~type (including ~~AT-Neu, ES-Lma~~C3 grass and ~~IT-Soy) where the prior COS simulations perform well. By contrast, for the single-site experiments conducted at forest sites,~~ C3 crop) sites, the assimilation ~~can always improve the simulation of GPP, although the optimizations were sometimes affected by the over-tuning of V_{cmax25} and f_leaf.~~ of COS is less effective in constraining the GPP simulations.

3.7 Comparison and evaluation of simulated ~~H~~LE and ~~LE~~ H

In order to verify the impact of COS assimilation on stomatal conductance and energy balance, observations of ~~latent heat~~LE and ~~sensible heat~~H were compared to the prior and posterior model outputs. ~~Due to the lack of observations at AT-Neu and IT-Soy, the validation was carried out at the remaining five sites only.~~ Results showed that the assimilation of COS is generally able to improve both latent and sensible heat, whether in single-site ~~experiment~~experiments or ~~multi~~the two-site experiment~~.~~ ~~And the~~ (**Figure S4-S7**). The assimilation is more effective in ~~improving~~reducing the ~~simulation~~model-observation difference of LE, with the average RMSE decreasing from ~~94.69 W/m^2~~89.55 W m^{-2} to ~~79.69 W/m^2,~~73.94 W m^{-2}, while for H, the average RMSE only decreased from ~~101~~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 ~~W/m^2 to 96.29 W/m^2.~~ to 0.64.

Results show that the BEPS model can simulate the daily variations of ~~H~~LE and ~~LE~~H as well as the diurnal cycle of LE ~~very~~ 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 ~~41.88 W/m^2 (**Figure 8** and **Figure S1**)~~31.60 W m^{-2} while the prior simulated H is underestimated by ~~39.92 W/m^2 (**Figure 8** and **Figure S1**).~~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). ~~In addition, at~~At FI-Hyy and US-Wrc, the model-observation biases are more pronounced for H, with an averaged MB of -~~62.13 W/m^2~~66.36 W m^{-2} than for LE with the averaged MB of ~~41.78 W/m^2. These results indicate that the BEPS model may underestimate the solar radiation absorbed by the evergreen needleleaf forest ecosystem.~~51.09 W m^{-2}. For the deciduous broadleaf forest sites DK-Sor and US-Ha1, the prior simulations of H ~~are very close to~~both fit well with the observations, with a maximum absolute MB of only ~~16.18W/m^2.~~17.88 W m^{-2}. However, ~~similar to evergreen needleleaf forests, its~~ the prior simulations ~~also~~ tend to overestimate LE~~, with MB ranging from 17.92 W/m^2 to 61.34 W/m^2. With a shrub PFT, ES-Lma is the only site where the prior simulations overestimate both H and LE~~ at US-Ha1, with a mean MB of ~~22.00 $W/m^2$and 50.06 W/m^2 respectively, which poses a significant challenge for the simultaneous optimization of H and LE.~~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 ~~mainly affected the daytime. Moreover, the correction was particularly effective for the evergreen needleleaf forest sites, where the mean values of the simulations of H and LE were increased by 30.95 W/m^2 and decreased by 31.04 W/m^2 respectively. With a mean RMSE reduction of -25.56%, the improvements of LE are also larger than the improvements of H. For the deciduous broadleaf forest sites, the optimization results for LE and H show considerable inconsistency. At US-Ha1, the model overestimated the absorbed solar radiation energy both in July 2012 and 2013. And the assimilation of COS significantly corrected the overestimation of LE, with RMSE reduction of 25.63% in July 2012 and 28.90% in July 2013. In contrast to the reduction of LE, the H was increased by 21.40 and 54.40 W/m^2, in the respective period. At DK-Sor, the simulations of H and LE using the default parameters of the BEPS model already performed very well, and little improvement is needed. However, as the prior simulated COS was much lower than observed COS, parameters including V_{cmaxZS}, VJ_slope and f_leaf were increased after the assimilation. As a result, the model output using the posterior parameters overestimated LE and underestimated H. As for ES-Lma, where the prior model output overestimated both H and LE, the posterior simulated LE was overestimated yet stronger, while the overestimation of H was partially corrected~~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 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 ~~multi~~two-site assimilation ~~greatly~~of COS effectively corrected the overestimation of LE and the underestimation of H in the prior simulations~~ during the daytime~~, with RMSE reductions of ~~26.57~~17.58 % for LE and ~~32.99~~22.33 % for H, ~~achieving almost identical effect to~~ 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 ~~multi~~two-site posterior parameters ~~were reduced by about one third compared~~are also superior to the prior simulations at FI-Hyy, ~~which allowed the overestimation of the prior simulation during the first half of the month to be effectively corrected (**Figure 8a**). Meanwhile, the model-observation differences of H were also remarkably reduced at FI-Hyy,~~ with ~~MB~~the RMSE reductions of ~~-63.44 W/m^2~~19.34 % for ~~the prior case~~LE and ~~-39.93 W/m^2~~5.90 % for ~~the posterior case~~H.

Overall, the BEPS model performed well in simulating the daily variations and diurnal cycle of ~~H~~LE and ~~LE~~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 ~~multiple~~two

sites simultaneously, and this improvement was particularly noticeable for ~~the simulation of~~ LE. We also found 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 ~~effectiveness~~influence of COS assimilation ~~in improving soil moisture simulations~~on the modelling of SWC was assessed by comparing hourly ~~soil water content~~SWC observations with hourly simulations of ~~soil moisture using prior parameters, single-site and multi-site posterior parameters.~~SWC. The assessments were carried out at all sites except US-Ha1, where no soil water observations were available. ~~We found that~~ 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 ~~improved~~suggested that the assimilation of COS is able to improve the simulation of ~~soil moisture~~SWC and this improvement ~~was~~is closely linked to the improved simulation of LE. However, ~~the improvement of soil moisture was not significant in a short period of time~~with the considerable adjustment of soil hydrology related parameters, the posterior simulated SWC also deviated noticeably from observations at several sites, i.e., the AT-Neu site.

Results show that the model can roughly follow the soil moisture trend (**Figure ~~9 and Figure S3~~S8**). However, the simulated ~~soil water content (~~SWC~~)~~ exhibited a clear ~~cycle of~~ diurnal ~~variation~~cycle whereas the observed SWC had almost no diurnal fluctuations. ~~Generally, in~~In response to the overestimation of LE at the ENF sites, the prior simulations ~~tended to overestimate the rate of decline in SWC. After~~underestimated the ~~assimilation of COS,~~SWC in most (6/7) of the single-site experiments conducted at ENF sites. As the overestimation of ~~the decline rate of SWC was significantly corrected and the posterior SWC simulations were more closely aligned with observations in terms of state and trend. For example, during the first half month of August 2014 at FI-Hyy, the prior simulations greatly overestimated~~ LE ~~(**Figure 8a**), such that the corresponding simulated SWC dropped rapidly to the wilting point and then remained constant (**Figure 9c**). In contrast, with the simulated LE being notably corrected~~ was effectively corrected by the assimilation of COS, the ~~simulated SWC was also effectively corrected~~decline in soil moisture slowed down, leading to the ~~level of the observations.~~

~~However, the effect of the assimilation of COS on the optimization of SWC simulations varied considerably from site to site. Little difference was found between the prior and the~~ posterior ~~simulations of SWC for those sites (AT-Neu, ES-Lma, IT-Soy) where there the GPP simulations also changed little after the assimilations of COS. The model significantly overestimated the rate of soil moisture decline at US-Wrc and DK-Sor, with the posterior simulated LE~~SWC simulation being ~~about 169% and 78% larger~~higher than the ~~observed. In contrast, the assimilation of COS remarkably improved the SWC simulations at FI-Hyy, with an average RMSE reduction of 24.86%. Yet, at FI-Hyy site,~~prior in the majority (6/7) of experiments. This conclusion was confirmed by the experiment ~~results (**Figure 9**)~~ 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 ~~showed there is still a large mismatch of observed and simulated decline rate of SWC during inter-storm periods~~show rapid declines and ~~of the effect of precipitation on SWC~~deviated from observations.

4 Discussion

4.1 Parameter changes

As ~~we~~ mentioned before, our results show V_{cmax25} was tuned the ~~texture-dependent parameters Ksat~~most in both the single-site experiments and ~~b had a very small~~the two-site experiments, with the mean absolute relative change ~~in the assimilation~~ of ~~COS, while the parameters related to PFT (V_{cmax25}~~44.59 %, followed by b_{scalar} and VJ_slope~~) and f_leaf varied dramatically~~. 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 ~~texture-dependent~~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. ~~Among the three COS plant uptake related parameters, it was found that the posterior V_{cmax25} had the largest change relative to the prior, with the relative change ranging from -60.64% to 113.45%, followed by f_leaf and VJ_slope~~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.

~~Although the posterior f_leaf has significant variability, f_leaf varies little in reality and is usually between 41% and 53% on an annual mean scale (Ryu et al., 2018). Considering that f_leaf is set to 0.5 in our model, it should remain about the same or be slightly reduced after the optimization. Certainly, the relative change rate of f_leaf is very reasonable in some experiments, such as the single-site experiments conducted at~~ 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.

~~FI-Hyy in August 2014 and July 2015, with relative changes of -14.18% and -13.29% respectively. However, the posterior f_leaf was also reduced dramatically by more than 60% in some single-site experiments conducted at FI-Hyy and US-Ha1, which suggested that the assimilation of COS may lead to over-tuning of f_leaf in some cases.~~ 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. ~~As mentioned in Sect 3.2, the prior simulated COS fluxes were larger than the observed ones at FI-Hyy and US-Ha1. Therefore, the assimilation of COS resulted in down-regulations of f_leaf in the single-site experiments performed at FI-Hyy and US-Ha1.~~ 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 %.

~~In addition to f_leaf, V_{cmax25} was also over-adjusted in a few assimilation experiments, particularly at~~ We noticed remarkably different optimization results for photosynthesis-related parameters in the experiments conducted in ~~August 2014. For example, at US-Wrc, V_{cmax25} was dramatically down-regulated by a similar degree in the single-site~~July 2015 and ~~multi-site experiment, with~~ July 2017 at FI-Hyy, especially for V_{cmax25} and VJ_slope. In these two experiments, the difference in the relative change ~~of -50.63% and -44.64% respectively, whereas the posterior VJ_slope~~in V_{cmax25} is more than 20%, and ~~f_leaf are significantly different.~~that in VJ_slope is as high as 37.04%. However, ~~with~~ such different ~~posterior~~sets of parameters~~,~~ caused similar effects on the ~~posterior simulated~~ COS ~~are very similar (**Figures 4b**).~~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 ~~widely~~ 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 ~~significantly~~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)~~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 and b,~~$Ksat_{scalar}$ and b_{scalar}~~.~~ they also play ~~a~~an important role in the assimilation of COS since the SWC simulations of BEPS are sensitive to ~~the two~~Ksat and b (Liu et al., 2011)~~. But since~~, and SWC is the primary factor for COS soil biotic flux modelling (Whelan et al., 2016). However, the soil COS exchange is generally much smaller than COS plant uptake (Whelan et al., 2018)~~ and they have less impact on the simulation of GPP (Novick et al., 2022), the assimilation results are not significantly affected by these two parameters.~~, and the parameter scheme provided by Whelan et al. (2022) sets different empirical parameter values (See **Table S3** for details) depending on the PFTs. Thus, 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 parameters related to light reaction (VJ_slope and 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). ~~At IT-Soy, **Figure 3j** showed that the assimilation of COS observations mainly changes the COS simulation in the early evening to minimize the cost function. Thus, it is reasonable that~~

~~f_leaf is the most influential parameter for that experiment as photosynthesis is very sensitive to radiation under such low light condition and f_leaf is an essential parameter for the calculation of PAR.~~

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. ~~And it~~It was confirmed by our single-site and ~~multi~~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, ~~especially if the prior simulations of COS are already very close to the observations. Moreover, the assimilation of COS could sometimes lead to overshooting of photosynthesis-related parameters, such as f_leaf, and thus result in considerable errors in the GPP simulations. In our experiments, those significant overshoots of f_leaf all occurred at well-vegetated forest sites (FI-Hyy and US-Ha1). This is also very reasonable as f_leaf is relevant to the calculation of PAR and light can become a limiting factor for photosynthesis, in particular when plants grow in dense vegetation (Demarsy et al., 2018).~~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. ~~In our experiments, posterior simulation results consistent with this mechanism were obtained in that although the posterior GPP simulations significantly deviate from reality due to parameter overshooting, the posterior LE does not. An example is the experiment conducted at FI-Hyy in July 2014, in which the posterior simulated GPP was substantially underestimated by 68.77%, while the posterior simulated LE was only 19.57% lower than the observations.~~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 ~~average of the absolute values of the~~ relative ~~changes~~change of -0.~~17%~~0026 % for $Ksat_{scalar}$ and -0.~~28% across all of the data assimilation experiments,~~0717 % for b_{scalar}, hardly caused any change in leakage. Therefore, our results indicate that the assimilation of COS can ~~significantly~~markedly improve the modelling of stomatal conductance and transpiration and finally improve soil moisture. However, our results also show that there are ~~large uncertainties in~~ remarkable discrepancies between the ~~BEPS~~

~~model for the simulation of the decline rate of SWC during inter-storm periods~~ecosystem COS flux simulations and ~~of the effect of precipitation on SWC, although in some cases the model using the posterior parameters has already achieved an excellent simulation of LE. This result suggests that there may still be significant errors in the soil texture-related parameters~~observations, and that ~~these errors~~discrepancies cannot be effectively ~~corrected by the assimilation of COS due to the weak connection between ecosystem COS fluxes~~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 ~~soil hydrological processes.~~ 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** ~~**2**~~**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. ~~Almost all~~All experiments were successfully performed, ~~with the exception of a few at the DK-Sor and IT-Soy sites,~~ and the results were shown in ~~in~~ **Figure** ~~**10**~~**7** and **Figure** ~~**S4**~~**S9**.

We found that the posterior V_{cmax25} significantly correlated ~~best~~ with the LAI (R^2 = 0.~~23~~17, P < 0.01~~), followed by VJ_slope (R^2 = 0.14, P < 0.05) and f_leaf (R^2 = 0.09, P < 0.1). Whilst~~) 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~~, to some extent~~, 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 ~~and~~, $Ksat_{scalar}$, b_{scalar}, f_leaf) weakly, although ~~the latter~~they also considerably changed by the assimilation.

In ~~Sect 3.4, we have noted that the posterior V_{cmax25} and f_leaf were sometimes over-tuned, which significantly influenced the posterior simulation of GPP. Here, by comparing the posterior parameters obtained with different LAI data, we further found that the over-tuning of those parameters could be partly attributed to the uncertainty of the LAI. For example, in the experiment conducted at FI-Hyy in July 2017, driven by the GLOBMAP LAI which were on average 41% greater than the in situ LAI, the posterior f_leaf value was significantly reduced, with a decrease rate of 78.09%. However, when the GLASS LAI, which is only 4% larger than the in situ LAI, is used to drive the model, the percentage decrease in f_leaf is significantly reduced to only 43.12%. Such~~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~~, H~~ and ~~SWC~~H for both single-site assimilation and ~~multi~~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 assimilation. However, at present, we face a serious lack of ecosystem-scale field measurements (Brühl et al., 2012; Wohlfahrt et al., 2012)~~, more~~. 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 ~~significant~~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. ~~Furthermore, the prior uncertainty of the parameters was simply set to 25% of the prior values in this study, which could certainly be refined. In conclusion, we should be more careful in considering the distribution and the magnitude of the~~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~~ in NUCAS~~, which may introduce ~~significant~~substantial errors into the parameter calibration. Kooijmans et al. (2021) has confirmed that modifying the COS mole fractions to vary spatially and temporally ~~significantly~~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~~a low and constant value (~~1 $mmol/m^2/s$~~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 ~~in~~to different degrees ~~in BEPS~~ 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 ~~$mmol/m^2/s$.~~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)~~ which~~, 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 ~~further~~ 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~~ first~~ 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, ~~the~~a resistance analog model of COS plant uptake and ~~the~~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 ~~multi~~two-site experiment ~~within~~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 ~~five~~four PFTs and ~~five~~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 ~~light reaction of photosynthesis~~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. ~~As shown here, the optimization of model parameters often faces~~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://github.com/yongguangzhang/BEPS-SIF-model,~~https://zenodo.org/records/8288751, 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 ~~AT-Neu,~~ 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~~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~~Lma, FI-Hyy; from the AmeriFlux database for US-Ha1 and US-Wrc; ~~and~~ from https://zenodo.org/records/3993111 for AT-Neu and from https://zenodo.org/record/6940750 ~~for~~IT-Soy. 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.
~~Demarsy, E., Goldschmidt-Clermont, M., and Ulm, R.: Coping with 'dark sides of the sun'through photoreceptor signaling, Trends in plant science, 23, 260-271, 2018.~~
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.
~~Kooijmans, L. M., Sun, W., Aalto, J., Erkkilä, K. M., Maseyk, K., Seibt, U., Vesala, T., Mammarella, I., and Chen, H.: Influences of light and humidity on carbonyl sulfide-based estimates of photosynthesis, Proceedings of the National Academy of Sciences, 116, 2470-2475, 2019.~~
~~Kooijmans, L. M., Maseyk, K., Seibt, U., Sun, W., Vesala, T., Mammarella, I., Kolari, P., Aalto, J., Franchin, A., and Vecchi, R.: Canopy uptake dominates nighttime carbonyl sulfide fluxes in a boreal forest, Atmospheric Chemistry and Physics, 17, 11453-11465, 2017.~~
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.
~~Novick, K. A., Ficklin, D. L., Baldocchi, D., Davis, K. J., Ghezzehei, T. A., Konings, A. G., MacBean, N., Raoult, N., Scott, R. L., and Shi, Y.: Confronting the water potential information gap, Nature Geoscience, 15, 158-164, 2022.~~
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.

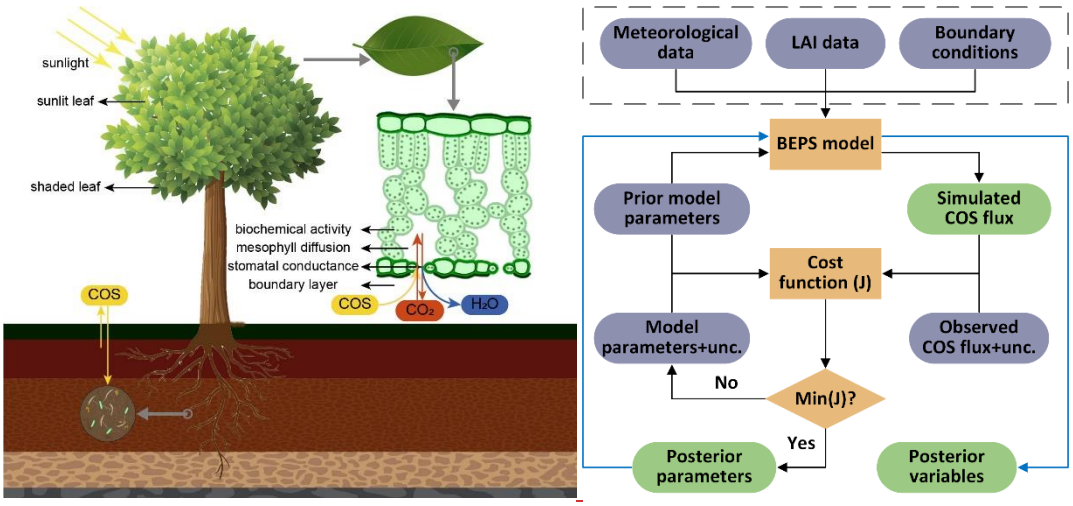
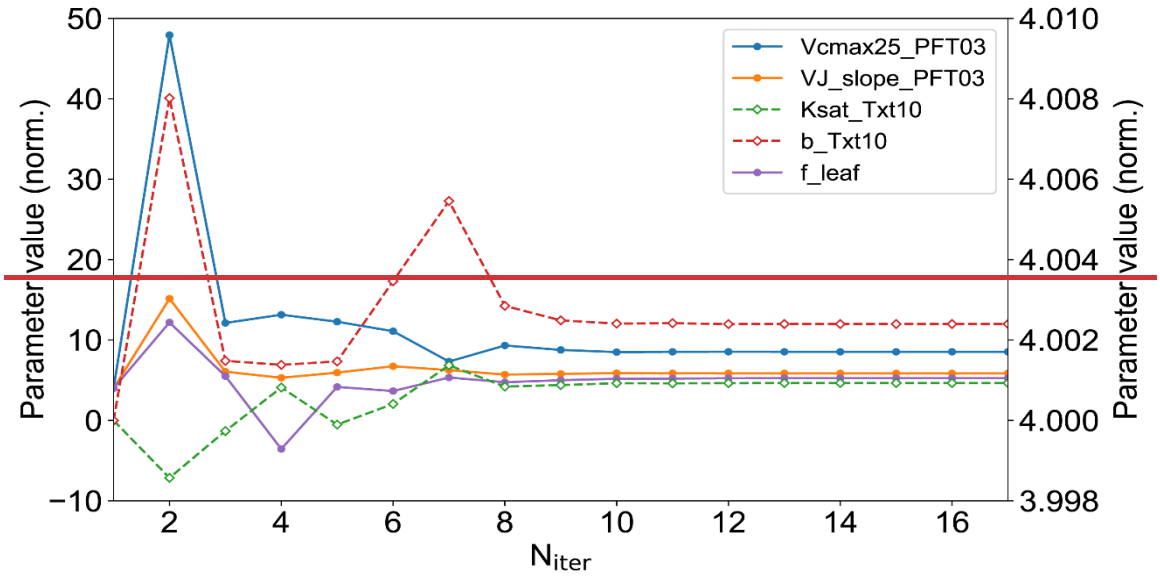
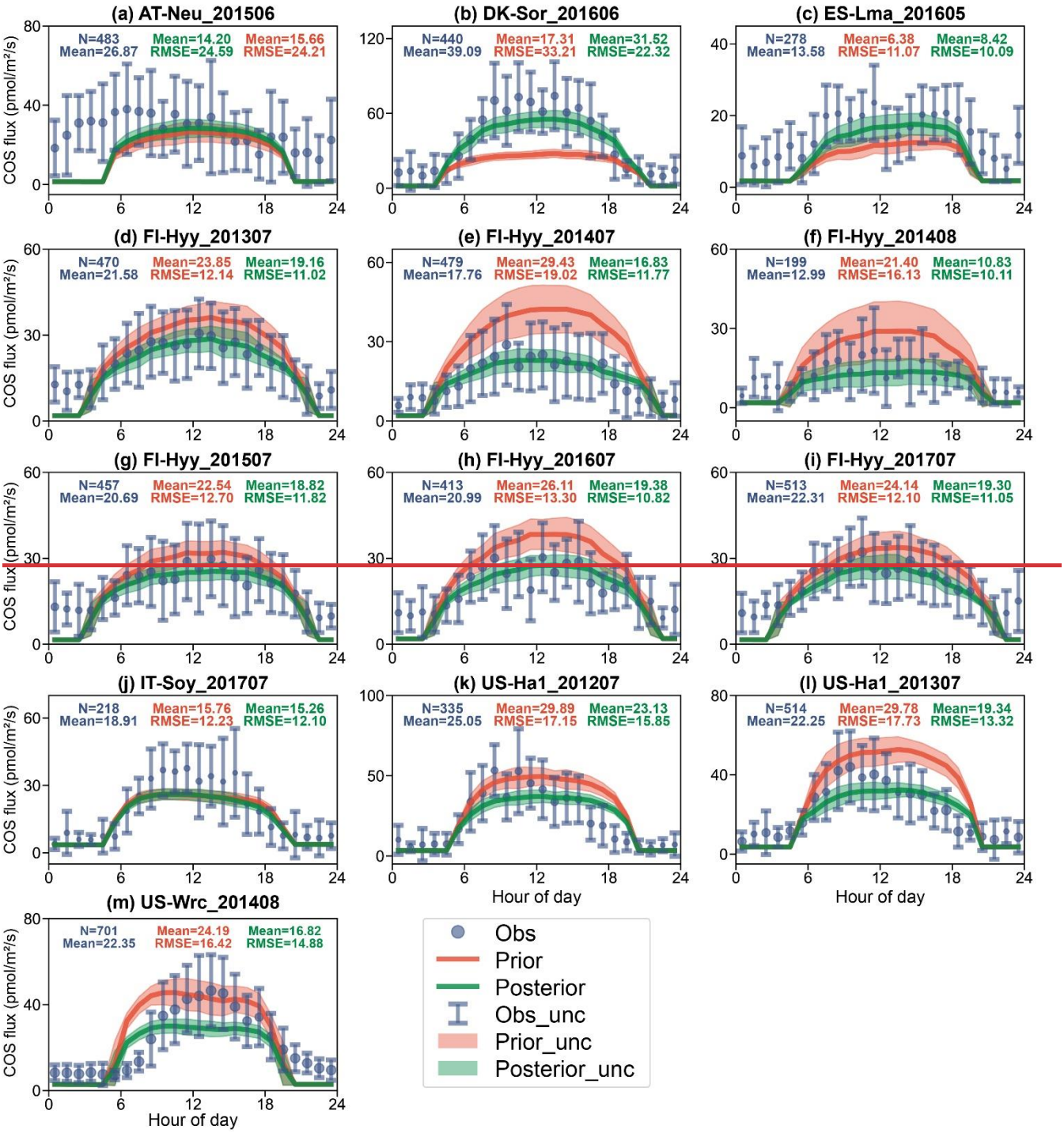


Figure 1. Schematic of the Nanjing University Carbon Assimilation System (NUCAS). Left: illustration of a two-leaf model coupling stomatal conductance, photosynthesis, transpiration and COS uptake, and an empirical model for simulating soil COS fluxes in NUCAS. Right: data assimilation flowchart of NUCAS. 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. The evolution of model parameters with the number of iterations of cost function (N_{iter}) in the normalized parameter space during the single-site experiment at the DK-Sor site in June 2016. Evolution (open carats and dashed lines) of soil texture (abbreviated as Txt) dependent parameters is plotted on the right hand y axis, evolution (filled circles and solid lines) of PFT-dependent parameters and global parameter is plotted on the left hand y axis.**~~


(a) AT-Neu_201506
N=483 Mean=26.87 Mean=14.20 RMSE=24.59 Mean=15.66 RMSE=24.21
(b) DK-Sor_201606
N=440 Mean=39.09 Mean=17.31 RMSE=33.21 Mean=31.52 RMSE=22.32
(c) ES-Lma_201605
N=278 Mean=13.58 Mean=6.38 RMSE=11.07 Mean=8.42 RMSE=10.09
(d) FI-Hyy_201307
N=470 Mean=21.58 Mean=23.85 RMSE=12.14 Mean=19.16 RMSE=11.02
(e) FI-Hyy_201407
N=479 Mean=17.76 Mean=29.43 RMSE=19.02 Mean=16.83 RMSE=11.77
(f) FI-Hyy_201408
N=199 Mean=12.99 Mean=21.40 RMSE=16.13 Mean=10.83 RMSE=10.11
(g) FI-Hyy_201507
N=457 Mean=20.69 Mean=22.54 RMSE=12.70 Mean=18.82 RMSE=11.82
(h) FI-Hyy_201607
N=413 Mean=20.99 Mean=26.11 RMSE=13.30 Mean=19.38 RMSE=10.82
(i) FI-Hyy_201707
N=513 Mean=22.31 Mean=24.14 RMSE=12.10 Mean=19.30 RMSE=11.05
(j) IT-Soy_201707
N=218 Mean=18.91 Mean=15.76 RMSE=12.23 Mean=15.26 RMSE=12.10
(k) US-Ha1_201207
N=335 Mean=25.05 Mean=29.89 RMSE=17.15 Mean=23.13 RMSE=15.85
(l) US-Ha1_201307
N=514 Mean=22.25 Mean=29.78 RMSE=17.73 Mean=19.34 RMSE=13.32
(m) US-Wrc_201408
N=701 Mean=22.35 Mean=24.19 RMSE=16.42 Mean=16.82 RMSE=14.88
COS flux (pmol/m²/s)
Hour of day
Obs
Prior
Posterior
Obs_unc
Prior_unc
Posterior_unc

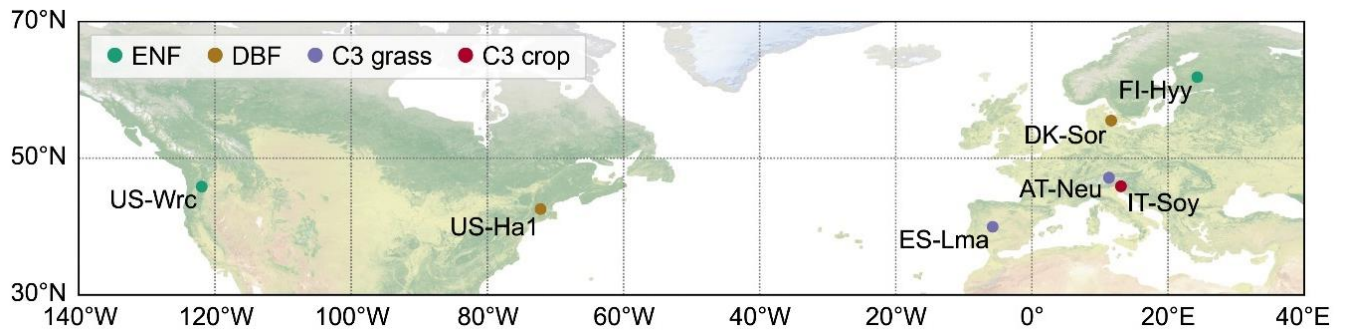

ENF
DBF
C3 grass
C3 crop
FI-Hyy
DK-Sor
AT-Neu
IT-Soy
ES-Lma
US-Wrc
US-Ha1
70°N
50°N
30°N
140°W
120°W
100°W
80°W
60°W
40°W
20°W
0°
20°E
40°E

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.

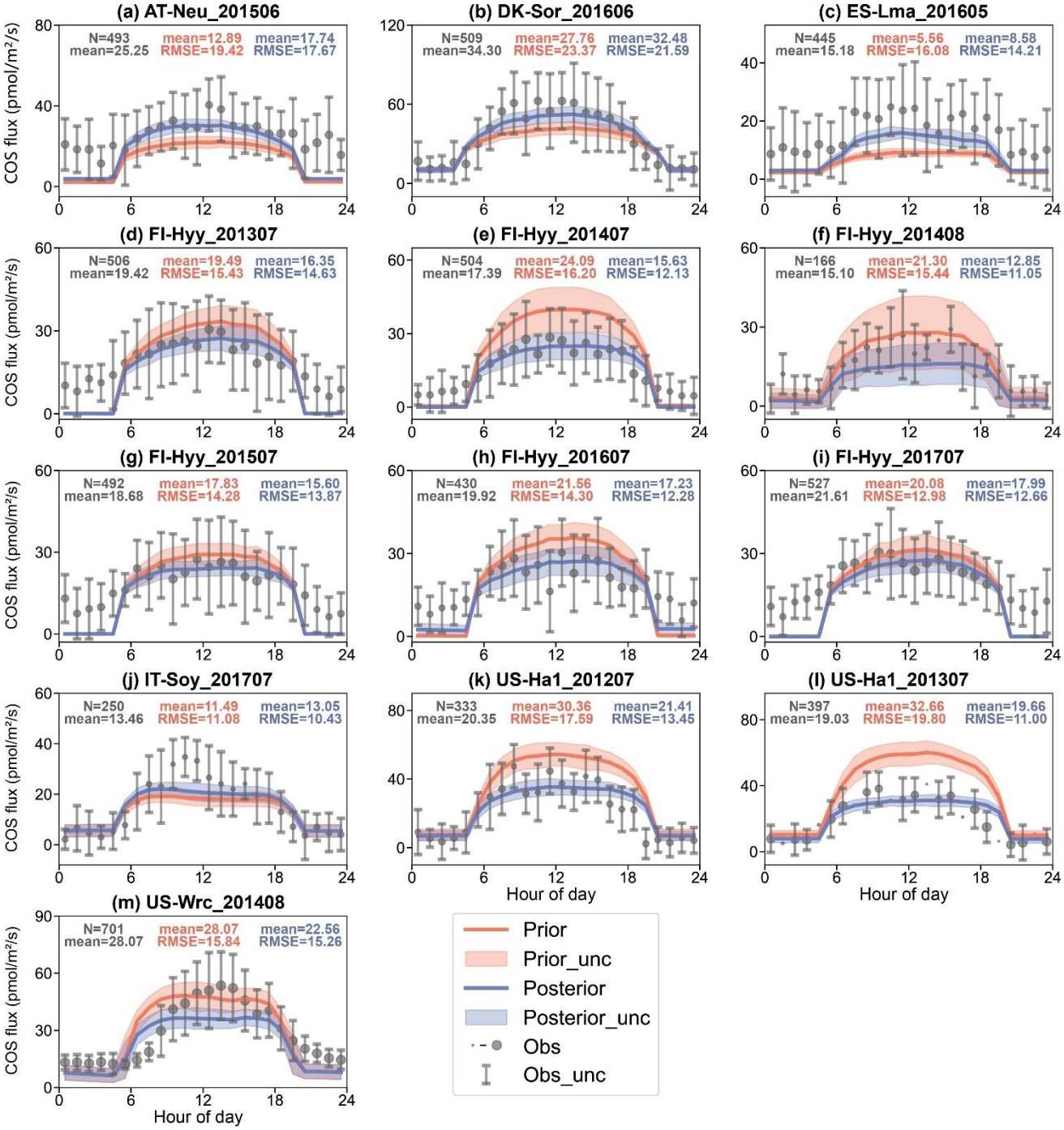


Figure 3. The mean diurnal cycle of observed (blue) and simulated COS flux using prior parameters (red) and single-site posterior parameters (~~green~~blue). The size of the circle indicates the number of observations (ranging from 1 to 31) within each circle, and the error bars depict the standard deviations 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.

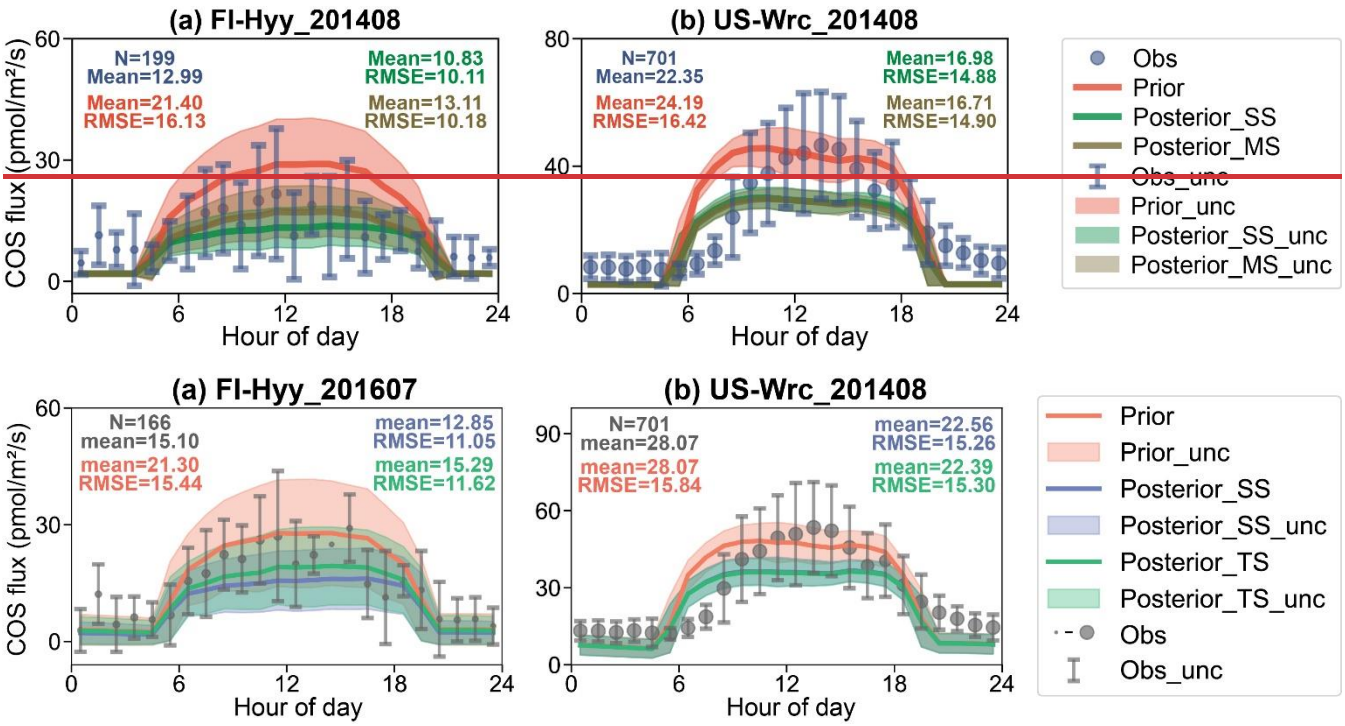
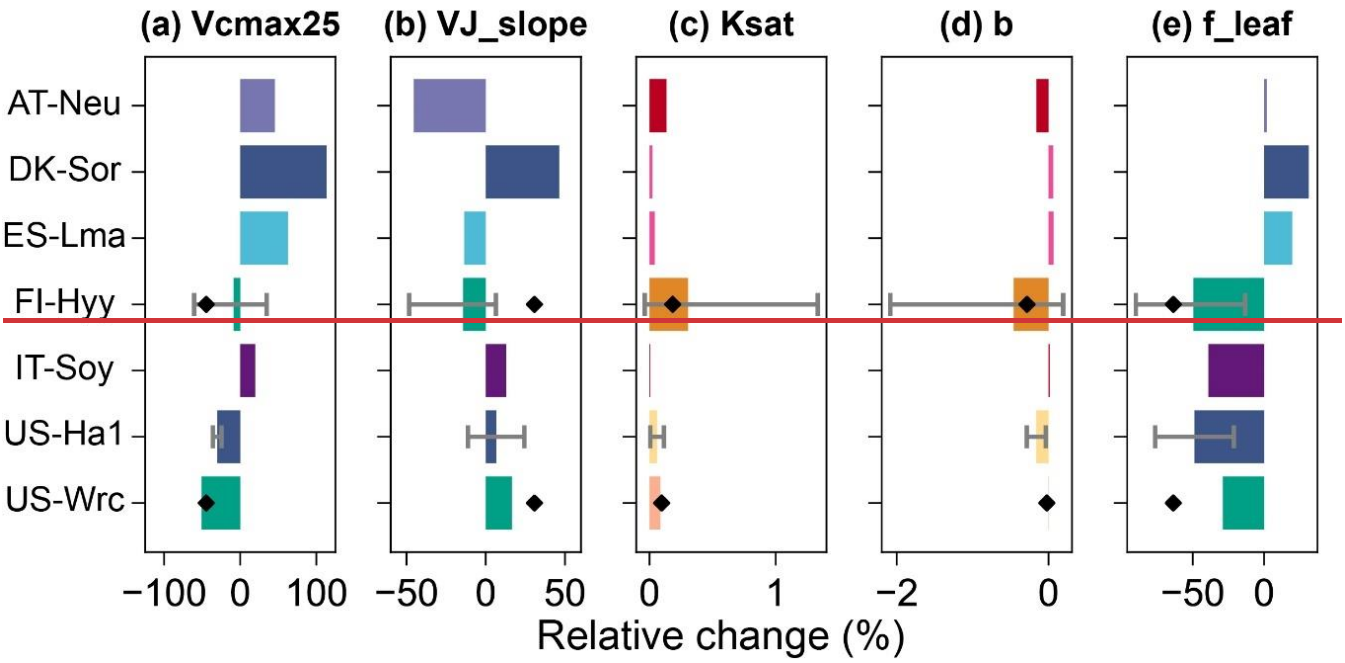


Figure 4. The diurnal cycle of observed (blue) and simulated COS flux using prior parameters (red), single-site (~~green~~blue) and ~~multi~~two-site (~~brown~~green) posterior parameters. The size of the circle indicates the number of observations (ranging from 1 to 31) within each circle, and the error 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.


(a) Vcmax25
(b) VJ_slope
(c) Ksat
(d) b
(e) f_leaf
AT-Neu
DK-Sor
ES-Lma
FI-Hyy
IT-Soy
US-Ha1
US-Wrc
−100 0 100
−50 0 50
0 1
−2 0
−50 0
Relative change (%)

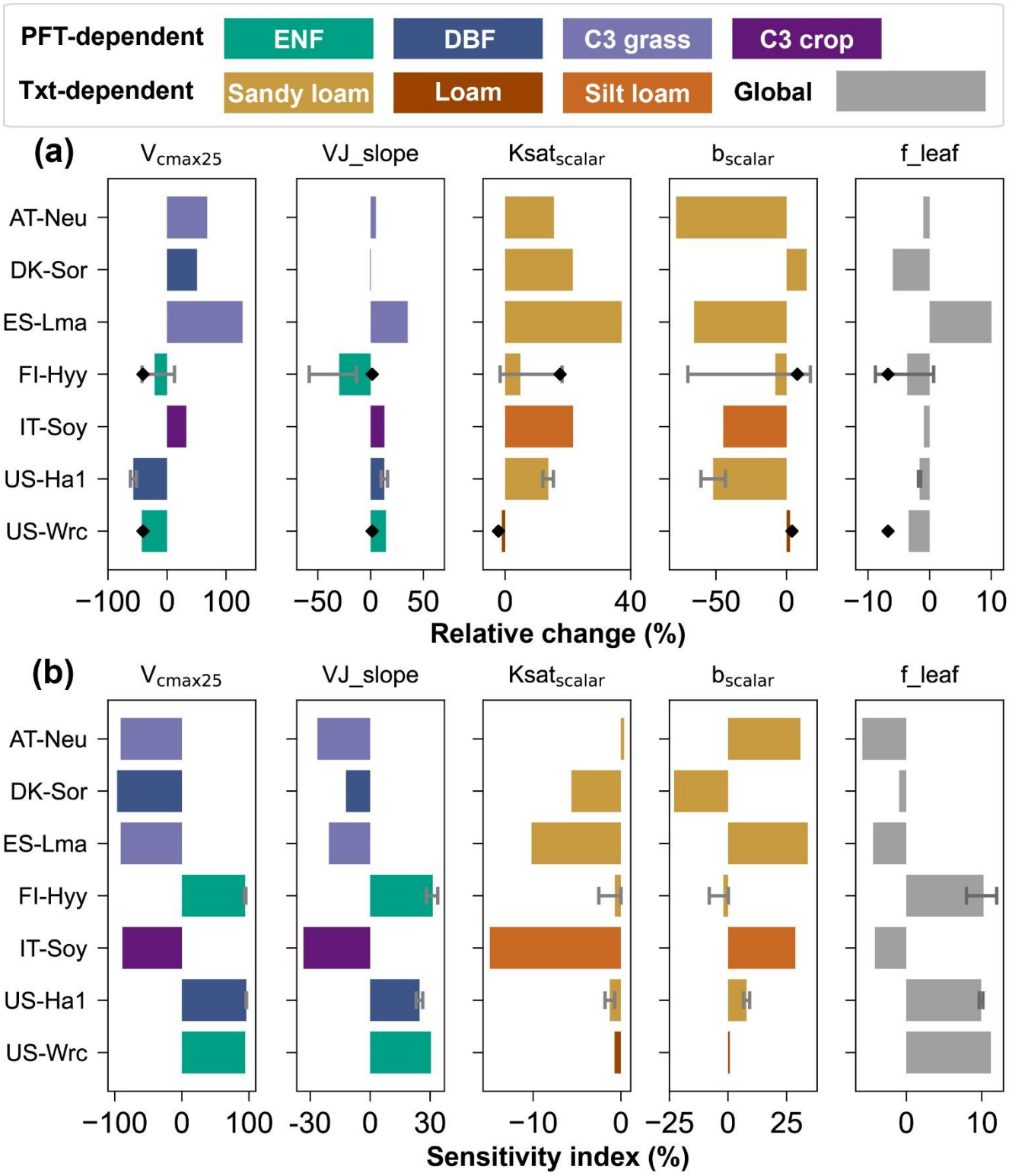
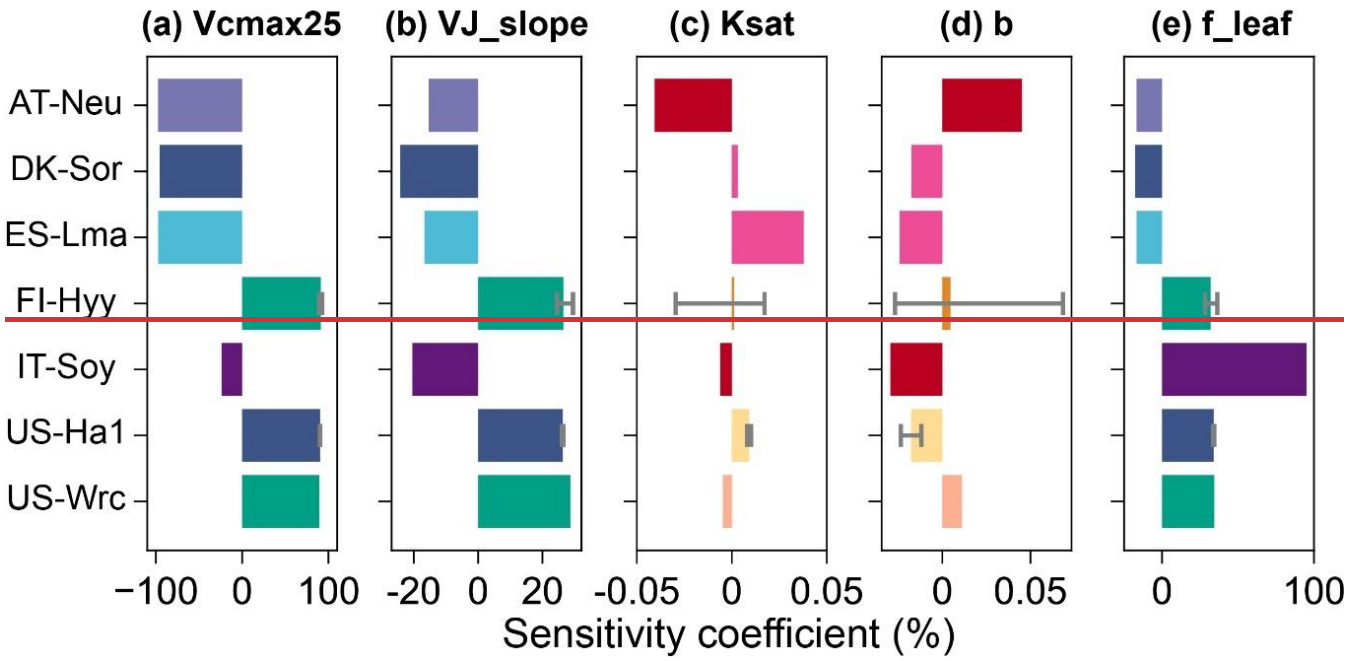
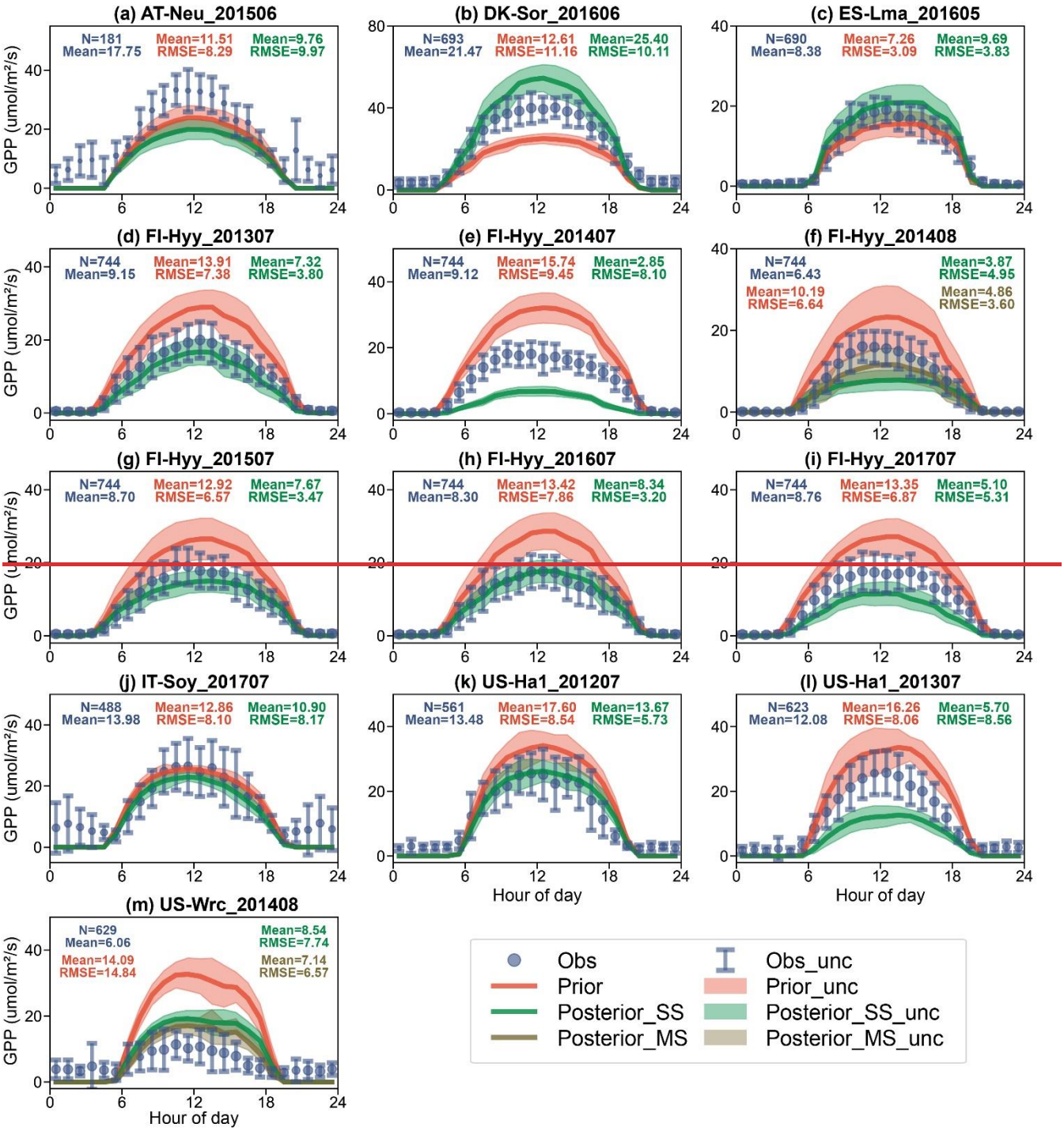


Figure 5. (a) Relative changes of parameters for single-site experiments (bars) and the multi-site experiment (diamond points). (b) Sensitivity indexes of parameters at prior values. For sites where multiple single-site experiments were conducted, the ends of the error bars and the bar indicate the maximum,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



~~**Figure 6. Sensitivity coefficients of parameters at default values. The ends of the error bars and the bar indicate the maximum, minimum and mean of the sensitivity coefficients of the parameters, respectively. For sites with the same PFT or soil texture, the same colors were used for their PFT-dependent and texture-dependent parameters, and f_leaf was plotted using the same color scheme as the PFT-dependent parameters.**~~


(a) AT-Neu_201506
N=181
Mean=17.75
Mean=11.51
RMSE=8.29
Mean=9.76
RMSE=9.97
(b) DK-Sor_201606
N=693
Mean=21.47
Mean=12.61
RMSE=11.16
Mean=25.40
RMSE=10.11
(c) ES-Lma_201605
N=690
Mean=8.38
Mean=7.26
RMSE=3.09
Mean=9.69
RMSE=3.83
(d) FI-Hyy_201307
N=744
Mean=9.15
Mean=13.91
RMSE=7.38
Mean=7.32
RMSE=3.80
(e) FI-Hyy_201407
N=744
Mean=9.12
Mean=15.74
RMSE=9.45
Mean=2.85
RMSE=8.10
(f) FI-Hyy_201408
N=744
Mean=6.43
Mean=10.19
RMSE=6.64
Mean=3.87
RMSE=4.95
Mean=4.86
RMSE=3.60
(g) FI-Hyy_201507
N=744
Mean=8.70
Mean=12.92
RMSE=6.57
Mean=7.67
RMSE=3.47
(h) FI-Hyy_201607
N=744
Mean=8.30
Mean=13.42
RMSE=7.86
Mean=8.34
RMSE=3.20
(i) FI-Hyy_201707
N=744
Mean=8.76
Mean=13.35
RMSE=6.87
Mean=5.10
RMSE=5.31
(j) IT-Soy_201707
N=488
Mean=13.98
Mean=12.86
RMSE=8.10
Mean=10.90
RMSE=8.17
(k) US-Ha1_201207
N=561
Mean=13.48
Mean=17.60
RMSE=8.54
Mean=13.67
RMSE=5.73
(l) US-Ha1_201307
N=623
Mean=12.08
Mean=16.26
RMSE=8.06
Mean=5.70
RMSE=8.56
(m) US-Wrc_201408
N=629
Mean=6.06
Mean=14.09
RMSE=14.84
Mean=8.54
RMSE=7.74
Mean=7.14
RMSE=6.57
GPP (umol/m²/s)
Hour of day
Obs
Prior
Posterior_SS
Posterior_MS
Obs_unc
Prior_unc
Posterior_SS_unc
Posterior_MS_unc

~~Figure 7.~~

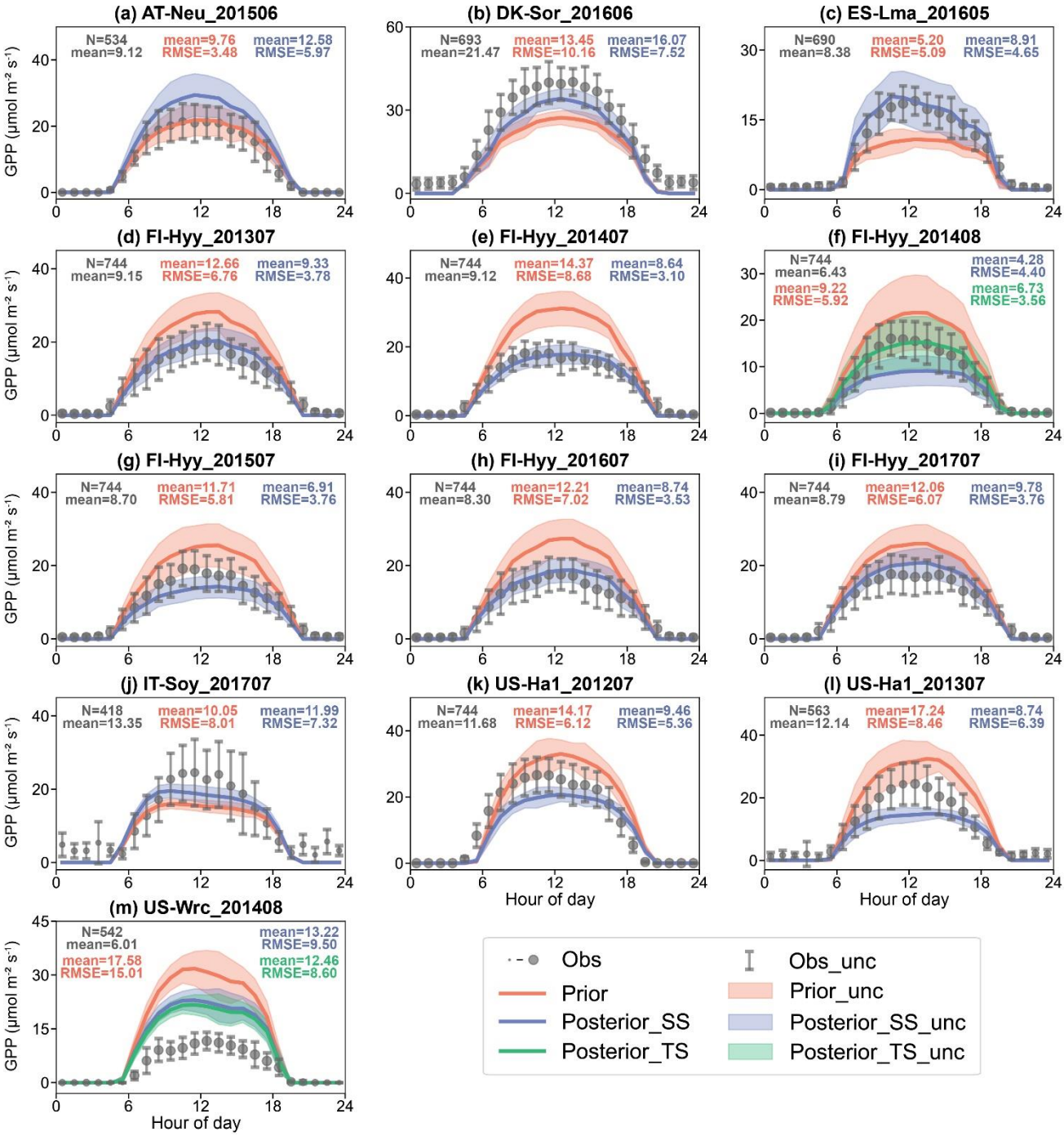
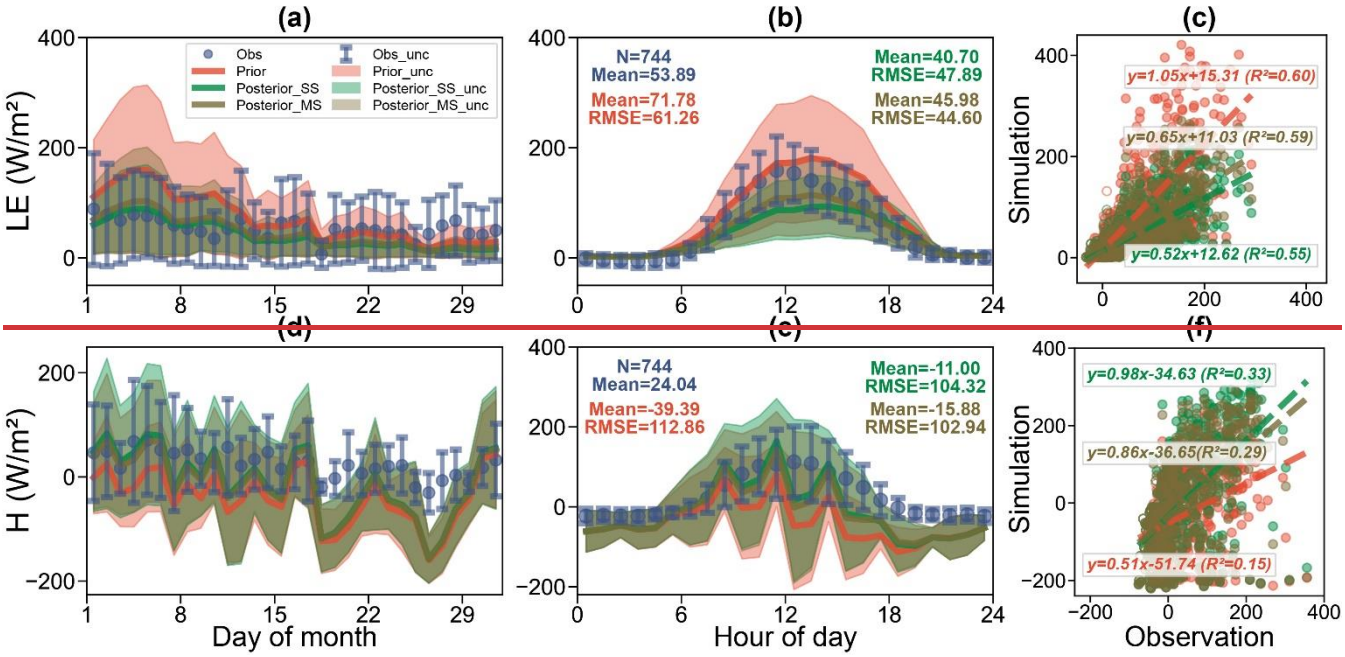


Figure 6. The diurnal cycle of observed (blue) and simulated GPP using prior parameters (red), single-site (green) and multi-site (brown) posterior parameters. The size of the circle indicates the number of observations within each circle~~,~~ (ranging from 1 to 31), and the error bars depict the standard 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.



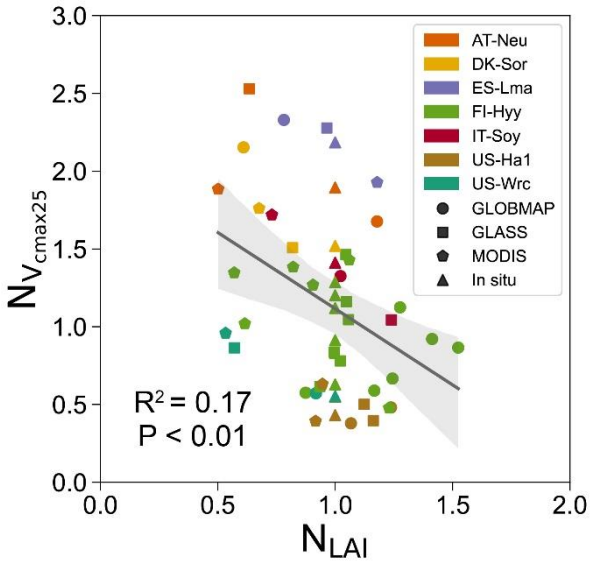
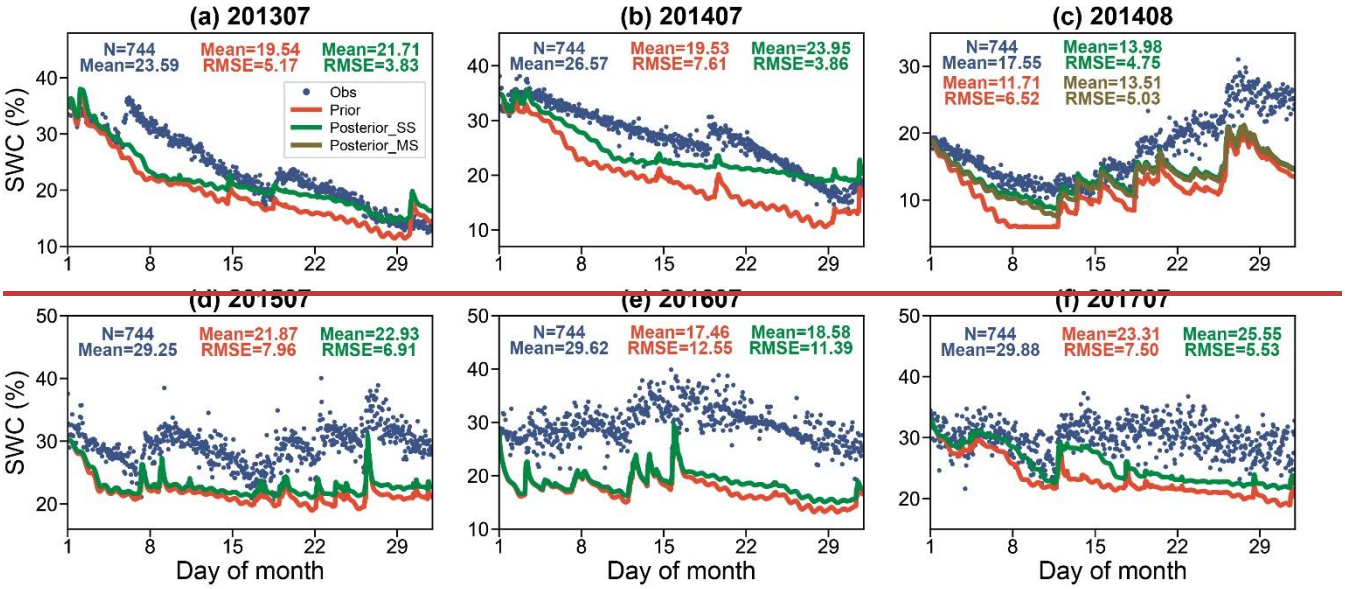
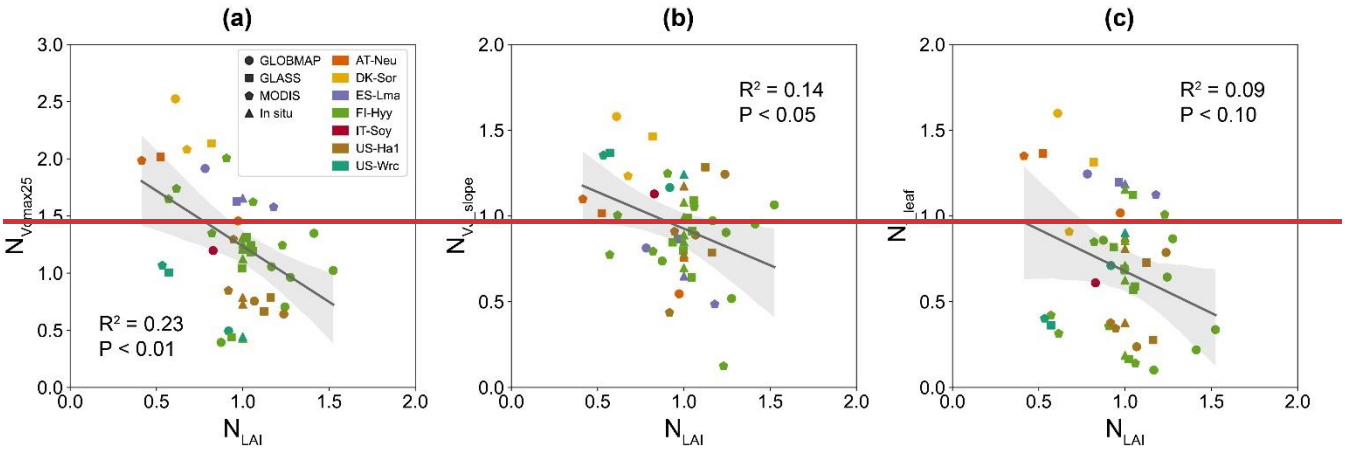


Figure ~~8. Daily variation (a and d), diurnal cycle (b and e) and scatter (c and f) plots of LE and H at FI-Hyy in August 2014. Observations (blue) are compared to simulations using prior (red) parameters, single-site (green) and multi-site (brown) posterior parameters. In the daily variation and diurnal plots, the size of the circle indicates the number of observations 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. And in the scatter plots, the daytime data (6:00-18:00LT) and nighttime data (18:00-6:00LT) are represented as solid and hollow circles respectively.~~



~~**Figure 9. Observed (blue point) and simulated SWC (%) at FI-Hyy. Results show SWC simulated using prior parameters (red line), single-site (green line) and multi-site (brown line) posterior parameters.**~~



~~**Figure 10**~~**7.** Influence of LAI on the posterior V_{cmax25} ~~**(a), the posterior VJ_slope (b) and the posterior f_leaf (c)**~~ 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}~~, the posterior VJ_slope and the posterior f_leaf~~ and the LAI were represented by their normalized values ~~$N_{Vcmax25}$, N_{VJ_slope}, N_{f_leaf}~~$N_{V_{cmax25}}$ 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 ~~lines~~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 ~~intervals~~interval spread around the ~~lines~~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~~, **shrub**~~ and C3 crop; Soil texture covered by the sites are ~~**silty clay**~~sandy loam, ~~**silty clay, loamy sand,**~~slit loam and ~~**silty**~~ loam.

Site ~~Name~~name	AT-Neu~~Lat (°N)~~	DK-Sor~~Lon (°E)~~	~~PFT~~ES-Lma	FI-Hyy~~Soil texture~~	IT-Soy~~LAI (m^2/m^2)*~~	~~year~~US-Ha1	US-Wrc~~References~~
Lat (°N)~~AT-Neu~~	47.12	55.49~~11.32~~	~~grass~~39.94	~~sandy clay~~61.85	~~4.7~~45.87	42.54~~2015~~	~~Spielmann et al. (2019)~~45.82
Lon (°E)~~DK-Sor~~	11.32~~55.49~~	11.64	~~DBF~~-5.77	~~clay~~24.29	~~5.0~~13.08	~~2016~~-72.17	~~Spielmann et al. (2019)~~-121.95
~~ES-Lma~~PFT	~~39.94~~C3 grass	~~-5.77~~DBF	~~shrub~~C3 grass	~~clay~~ENF	C3 crop~~1.82~~	~~2016~~DBF	~~Spielmann et al. (2019)~~ENF

FI-HyySoil texture	61.85Sandy loam	24.29Sandy loam	ENFSandy loam	sandySandy loam	4.0Slit loam	2013-2017Sandy loam	Kohonen et al. (2022)Loam
IT-SoyLAI*	45.873.88	13.085.0	1.82crop	sandy clay4.0	2.3	2017 5.0	Spielmann et al. (2019)8.7
US-Ha1Year	2015 42.54	-72.172016	DBF2016	silty loam2013-2017	2017 5.0	2012-2013	2014Wehr et al. (2017)
ReferencesUS-Wre	45.82(Spielmann et al., 2020)	-121.95(Spielmann et al., 2019)	ENF(Spielmann et al., 2019)	sandy clay loam(Sun et al., 2018; Vesala et al., 2022; Kohonen et al., 2022)	8.7(Spielmann et al., 2019; Abadie et al., 2022)	2014(Commane et al., 2015; Wehr et al., 2017)	Rastogi et al. (2018)(Shaw et al., 2004; Rastogi et al., 2018)

** Mean one-sided LAI (m² m⁻²) during the experimental period*

Table 2. Configuration and assimilation result of each twin experiment. $J_{initial}$ and J_{final} denote the initial value and the final value of the cost function $J(x)$ respectively, $G_{initial}$ and G_{final} denote the initial value and the final value of the gradient respectively.

Site	Assimilation window	Perturbation	$J_{initial}$	J_{final}	$G_{initial}$	G_{final}
AT-Neu	June 2015	0.4	2.31E+04	2.70E-14	1.91E+04	3.14E-05
DK-Sor	June 2016	0.4	3.20E+04	2.34E-16	2.54E+04	8.28E-05
ES-Lma	May 2016	0.4	4.58E+03	1.63E-18	3.94E+03	1.22E-06
FI-Hyy	July 2013	0.2	1.05E+04	4.99E-16	1.66E+04	2.77E-05
	July 2014	0.2	1.56E+04	1.51E-16	2.44E+04	6.41E-05
	August 2014	0.2	7.76E+03	1.87E-18	1.20E+04	1.49E-06
	July 2015	0.2	7.95E+03	4.01E-19	1.33E+04	8.42E-07
	July 2016	0.2	1.20E+04	1.01E-14	1.92E+04	2.18E-04
	July 2017	0.2	9.27E+03	8.35E-16	1.55E+04	1.48E-04
IT-Soy	July 2017	0.4	1.72E+04	3.50E-13	1.42E+04	2.79E-04
US-Ha1	July 2012	0.4	6.85E+04	1.61E-14	5.48E+04	8.54E-05
	July 2013	0.4	7.76E+04	8.21E-16	6.23E+04	2.65E-05
US-Wre	August 2014	0.2	1.13E+04	6.90E-15	1.78E+04	6.69E-05
Multi-site	August 2014	0.2	1.70E+04	3.17E-14	2.68E+04	1.41E-04

Table 3.**Table 2.** The configuration and the relative changes (%) of the parameters for each single-site assimilation experiment. The minimization efficiency 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 $1 - J_{final}/J_{initial}$ $1 - J_{final}/J_{initial}$, and N_{cos} N_{COS} denotes the number of ecosystem COS flux observations.

Site name	Assimilation window	N_{cos} N_{COS}	Cost function reduction (%)	Relative change (%) of parameters				
				V_{cmax25}	VJ_slope	$Ksat$ $Ksat_{scalar}$	b b_{scalar}	f_leaf
AT-Neu	June 2015	483493	1.6416.39	45.5467.69	-45.425.10	0.134715.57	-0.158378.13	-1.7701
DK-Sor	June 2016	440509	42.179.46	113.4550.77	46.37-0.47	0.023321.54	0.060014.23	31.35-5.97
ES-Lma	May 2016	278445	10.4815.70	62.60127.80	-13.4935.18	0.041237.08	0.0669-65.33	19.6510.05
FI-Hyy	July 2013	470506	21.434.87	2.2832.55	6.4813.15	0.006721.60	-0.030544.72	-66.260.94
	July 2014	479504	62.237.74	5.60-13.42	-2.7925.48	0.0399-1.58	-0.085990	-89.938.80
	August 2014	199166	64.9240.59	-60.6441.09	-26.2819.10	0.22234.02	-0.370416.84	-14.186.21

	July 2015	457492	14.7450.94	-3.7442.44	-48.2241.03	-0.03748.65	0.19395.07	-13.291.66
	July 2016	413430	35.025.73	-29.5912.45	-9.6558.23	0.268900	-0.377307	-350.65
	July 2017	513527	53.7118.94	34.79-33.32	-4.6613.48	1.332918.13	-2.084569.86	-78.091.60
IT-Soy	July 2017	218250	2.086.35	19.69-7.88	12.81-21.20	0.004903	-0.015745	-39.004.14
US-Ha1	July 2012	335333	27.9644.14	-35.9251.89	24.3116.08	0.006012.05	-0.035843.31	-21.311.44
	July 2013	514397	58.1069.05	-24.5462.08	-11.1510.00	0.113715.39	-0.286460.58	-76.311.82
US-Wrc	August 2014	701	44.6527.71	-50.6342.77	1614.52	0.0860-1.04	0.00602.45	-28.923.39

Table 43. The configuration and the relative changes (%) of the parameters for the multi-site assimilation experiment at FI-Hyy and US-Wrc site. N_{cos} N_{COS} denotes the total number of ecosystem COS flux observations.

Site name	Assimilation window	N_{cos} N_{COS}	Cost function reduction (%)	Relative change (%) of parameters				
				V_{cmax25}	VJ_slope	Ksat $Ksat_{scalar}$	b b_{scalar}	f_leaf
FI-Hyy	August 2014	900867	47.3328.29	-44.6441.36	30.722.96	0.183717.32	-0.284115.56	-63.646.28
US-Wrc						0.09631.36	-0.02252.60	

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):

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

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

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:

$$k = G(\theta)\Omega\, cos(\theta) \tag{A9}$$

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):

$$J_{max} = VJ\_slope\, V_{cmax} - 14.2 \tag{A10}$$

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

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

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 °C) 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:

$$w_i = \frac{R_i f_{w,i}}{\sum_{i=1}^{n} R_i f_{w,i}} \tag{A12}$$

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:

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

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:

$$f_w = \sum_{i=1}^{n} f_{w,i} w_i \tag{A14}$$

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

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

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:

$$Ksat_i = Ksat_{scalar} Ksat_{df,i} \tag{A16}$$

$$b_i = b_{scalar} b_{df,i} \tag{A17}$$

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

Supplement of

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

Huajie Zhu et al.

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

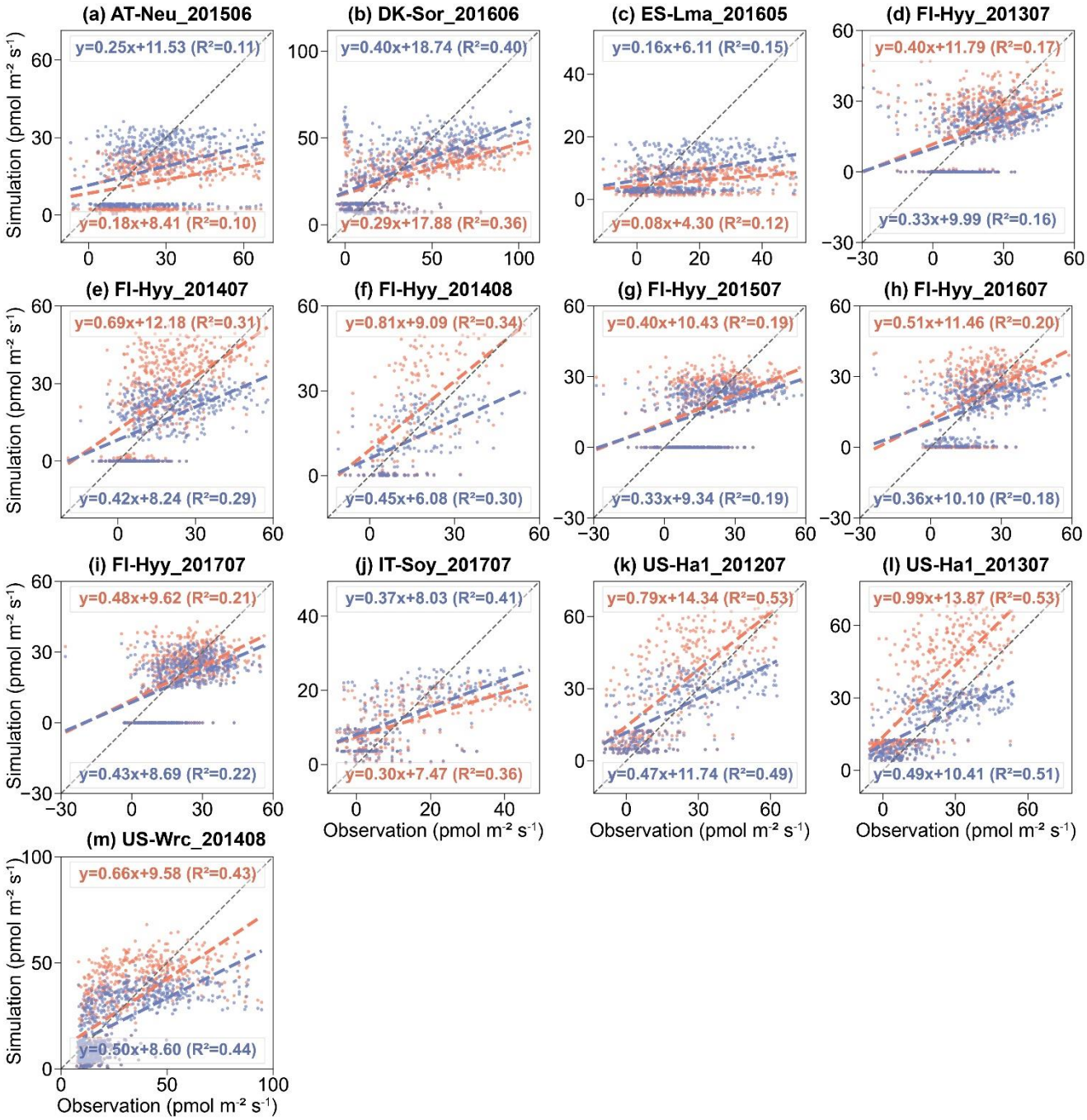


Figure S1. Scatterplots of observed versus simulated hourly COS flux using prior (red) and single-site posterior (blue) parameters.

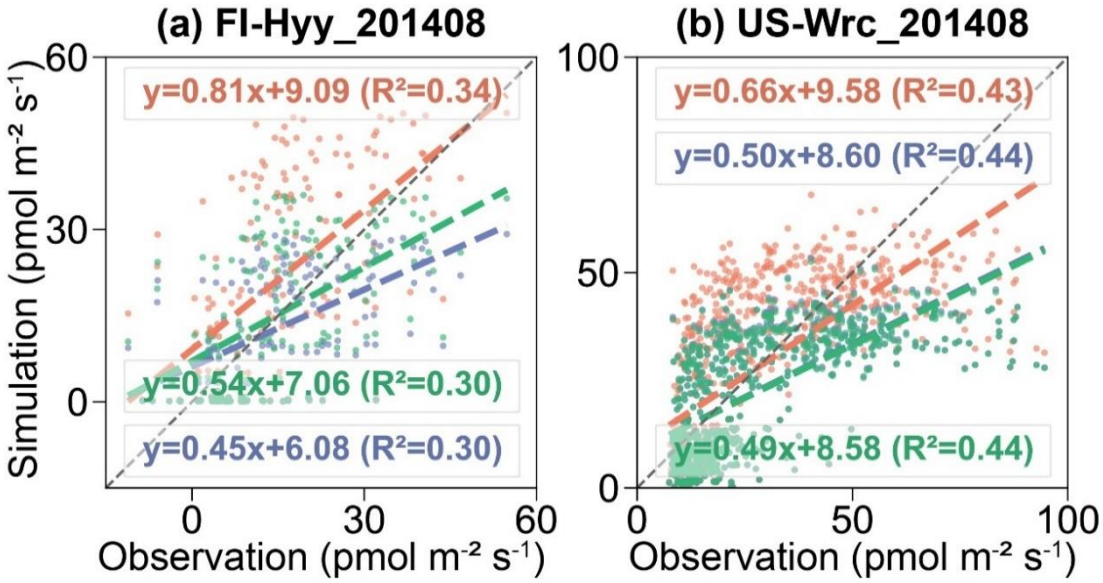


Figure S2. Hourly scatterplots of observed versus simulated hourly COS flux using prior (red), single-site (blue) and two-site (green) posterior parameters.

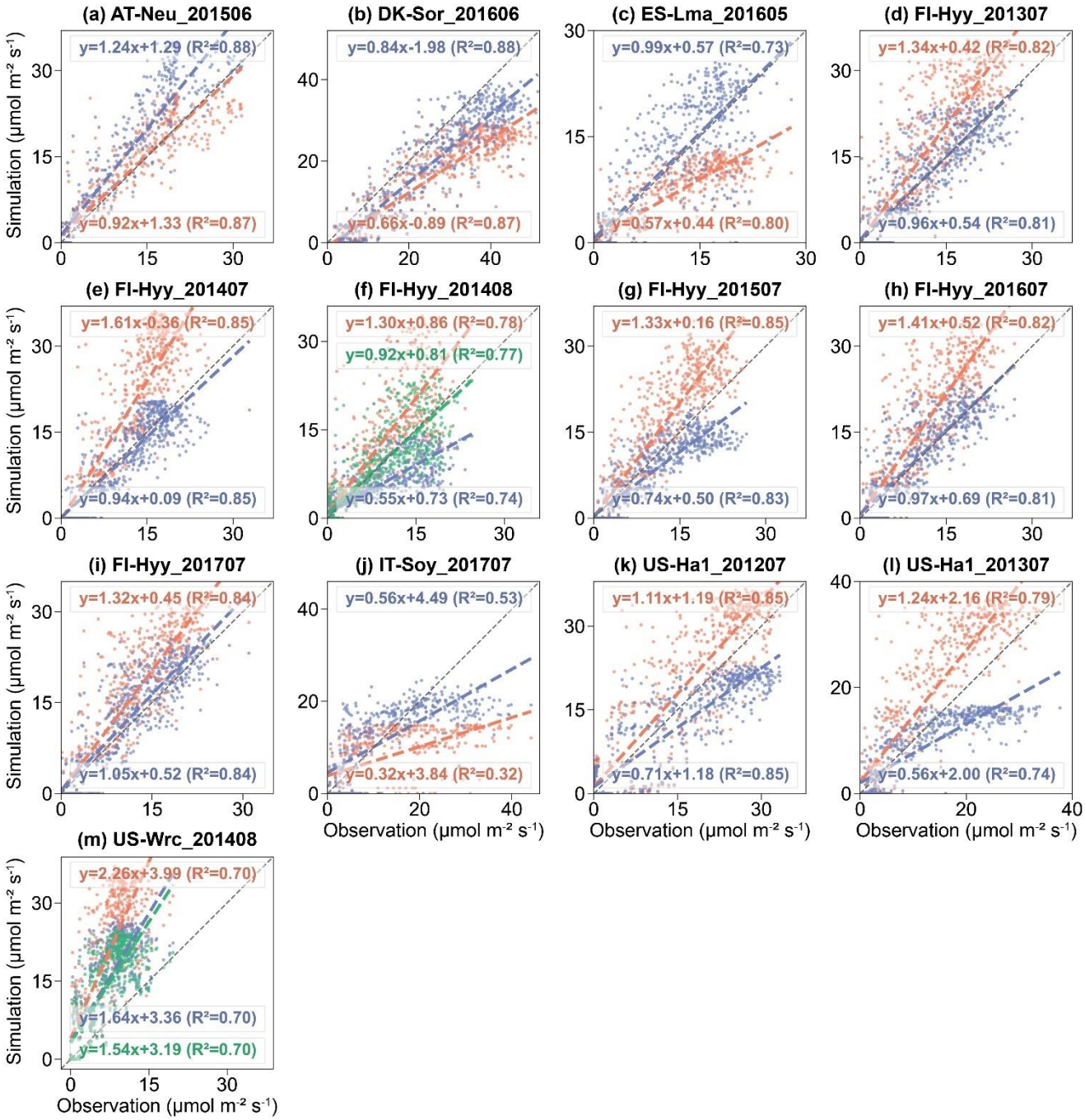


Figure S3. Hourly scatterplots of observed versus simulated hourly GPP using prior (red), single-site (blue) and two-site (green) posterior parameters.

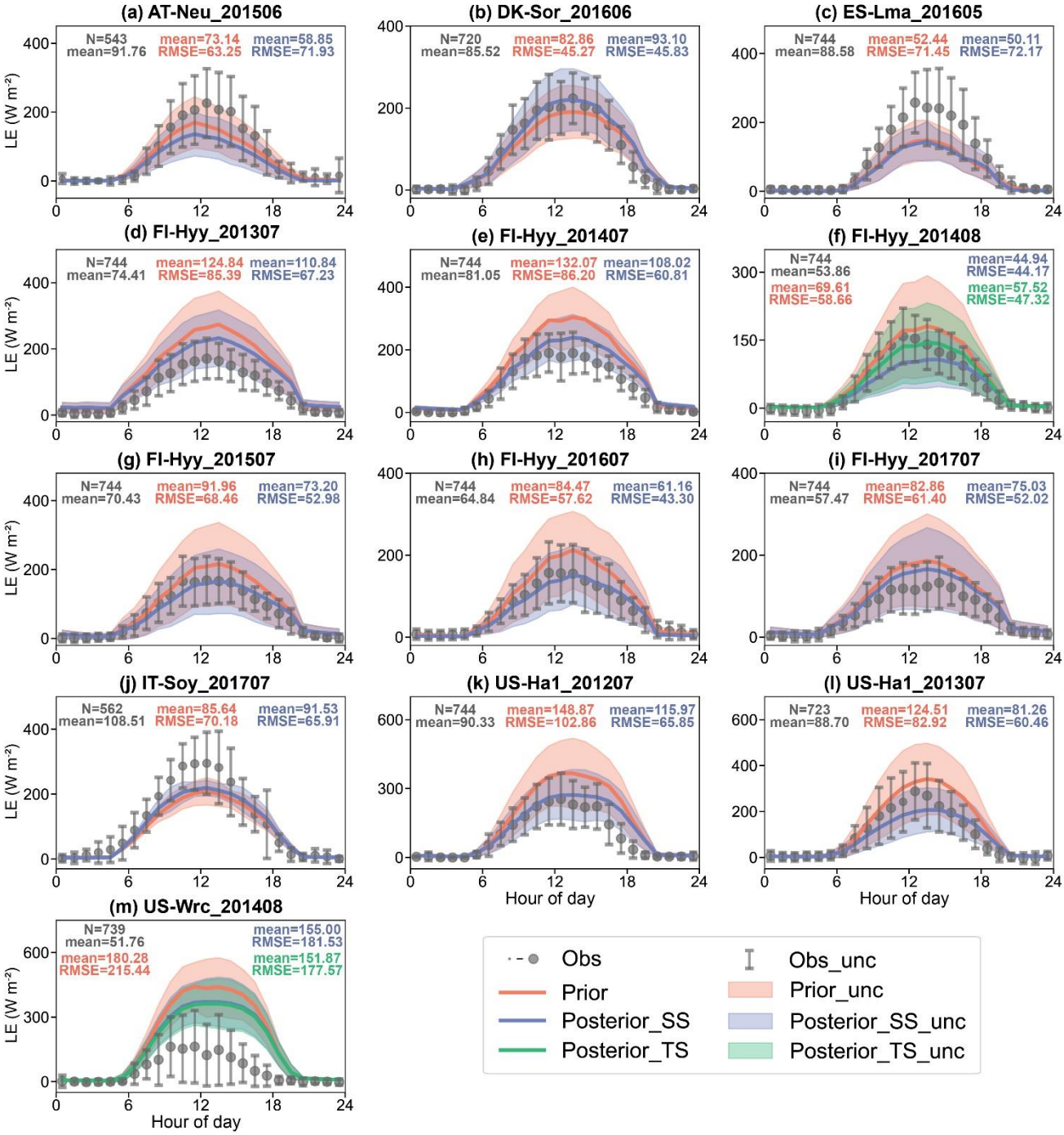


Figure S4. The diurnal cycle of observed (black) and simulated LE using prior parameters (red), single-site (blue) and two-site (green) posterior parameters. The size of the circle indicates the number of observations within each circle (ranging from 1 to 31), and the error bars depict the standard deviations 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.

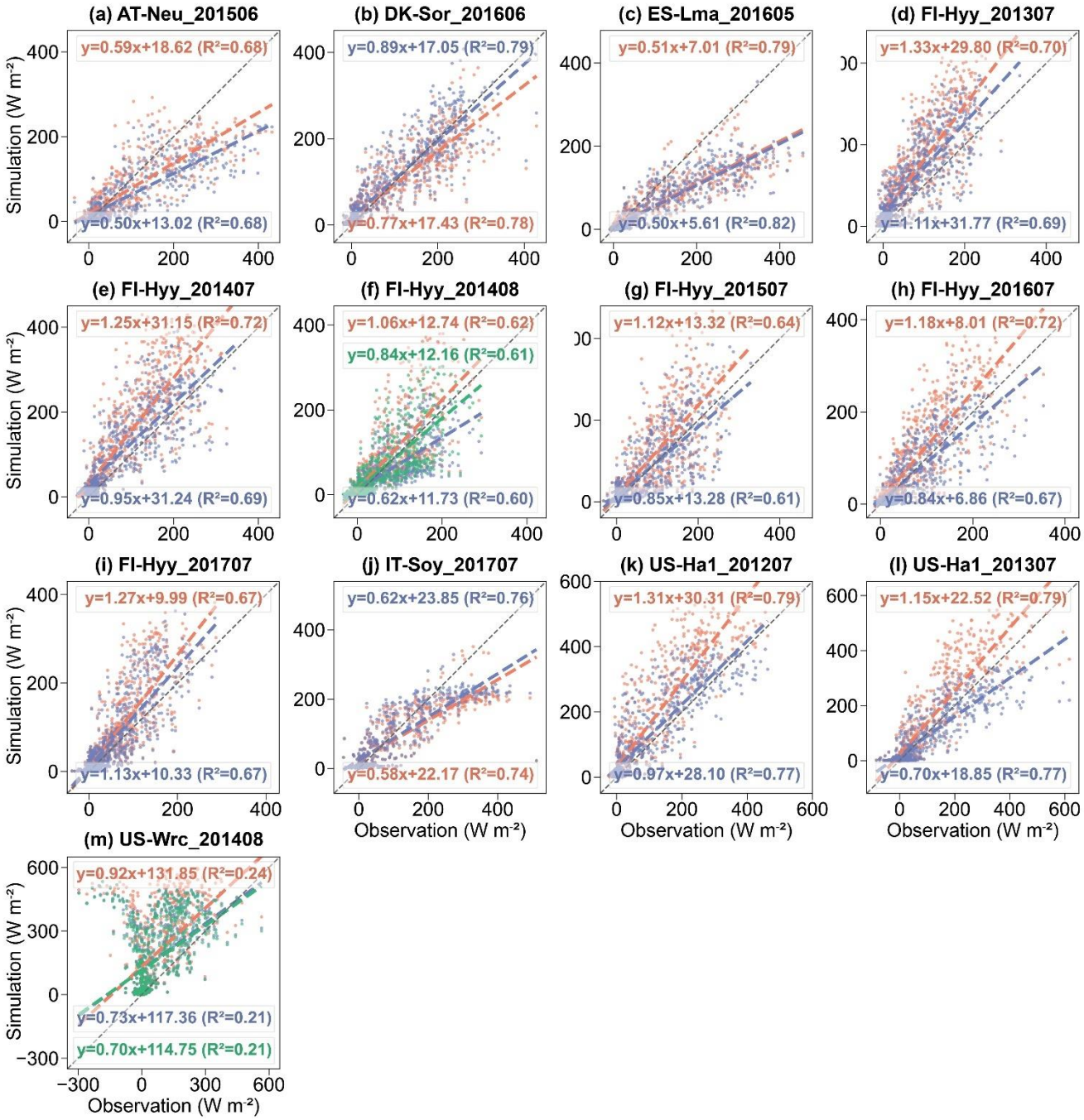


Figure S5. Scatterplots of observed versus simulated hourly LE using prior (red), single-site (blue) and two-site (green) posterior parameters.

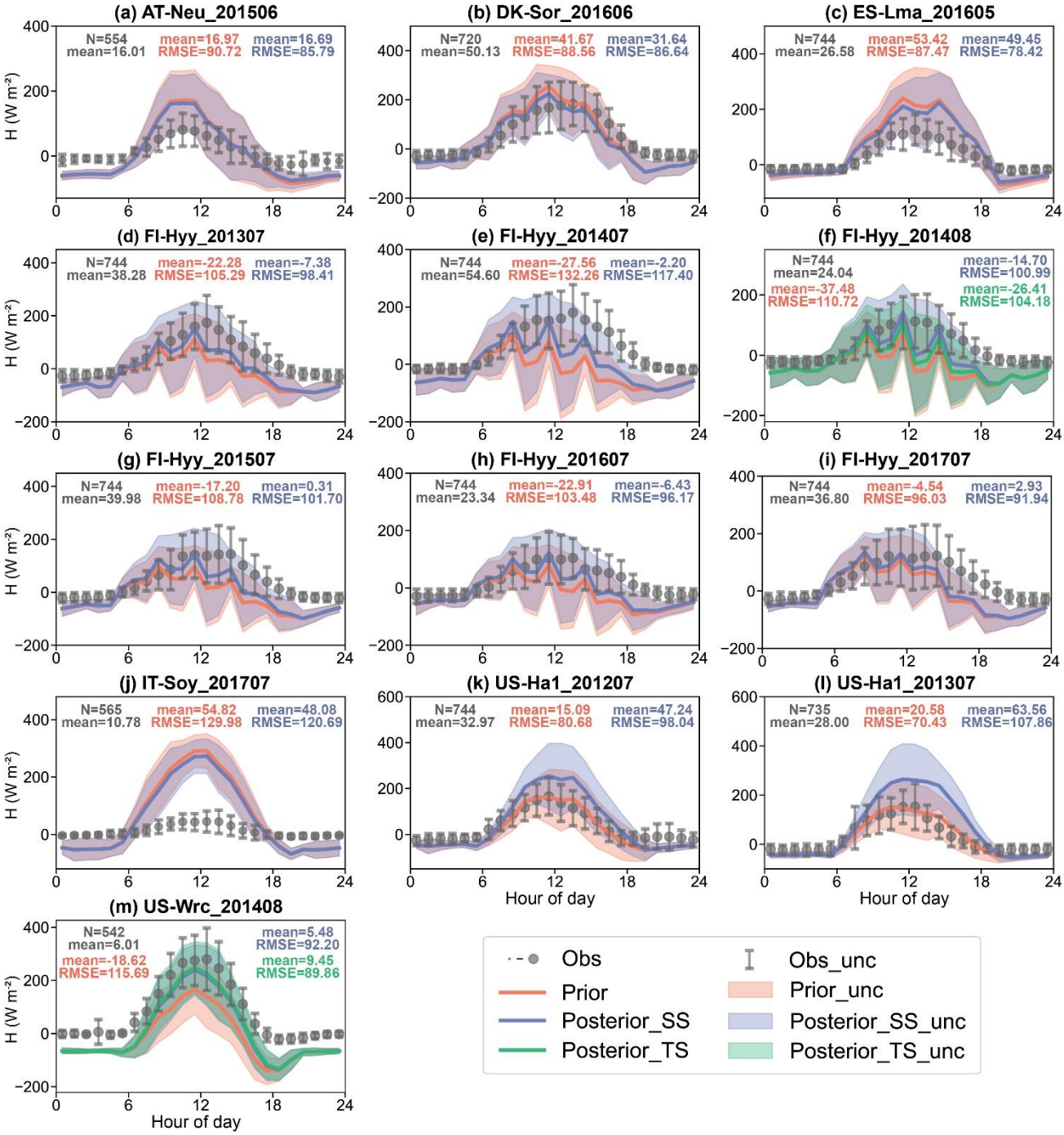


Figure S6. The diurnal cycle of observed (black) and simulated H using prior parameters (red), single-site (blue) and two-site (green) posterior parameters. The size of the circle indicates the number of observations within each circle (ranging from 1 to 31), and the error bars depict the standard deviations 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.

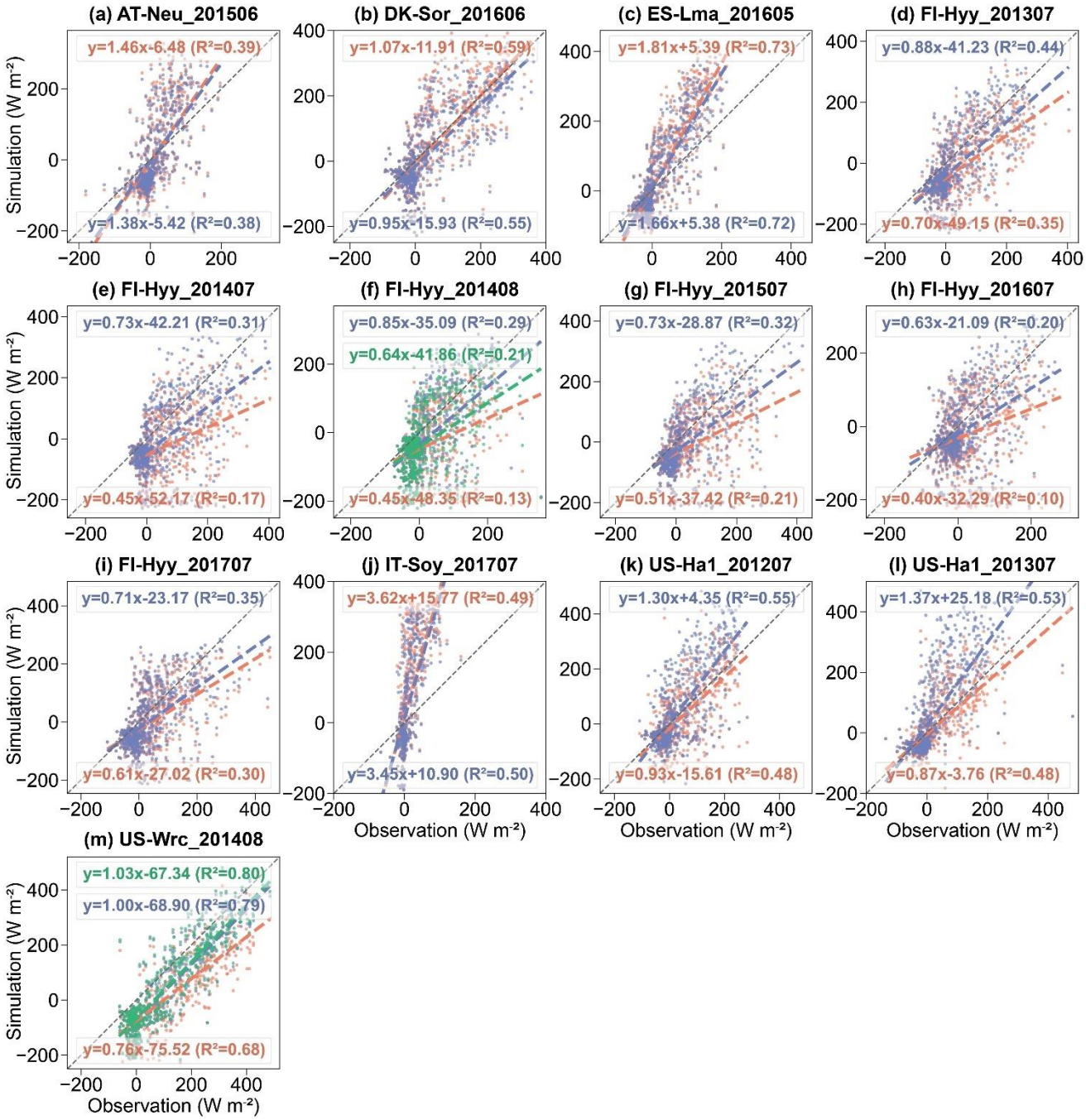


Figure S7. Hourly scatterplots of observed versus simulated hourly H using prior (red), single-site (blue) and two-site (green) posterior parameters.

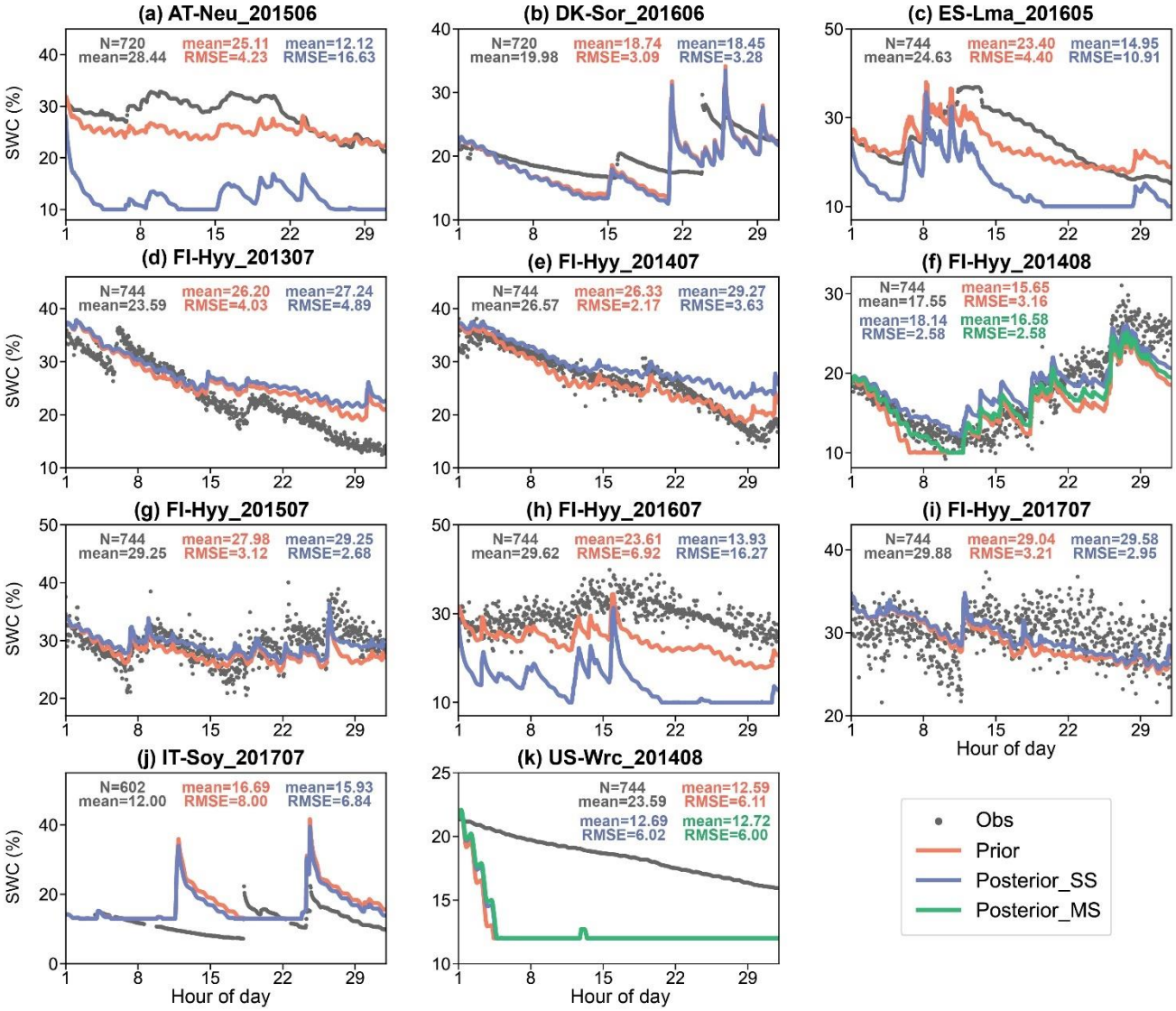


Figure S8. Observed (black point) and simulated SWC (%). Results show SWC simulated using prior parameters (red line), single-site (blue line) and two-site (green line) posterior parameters.

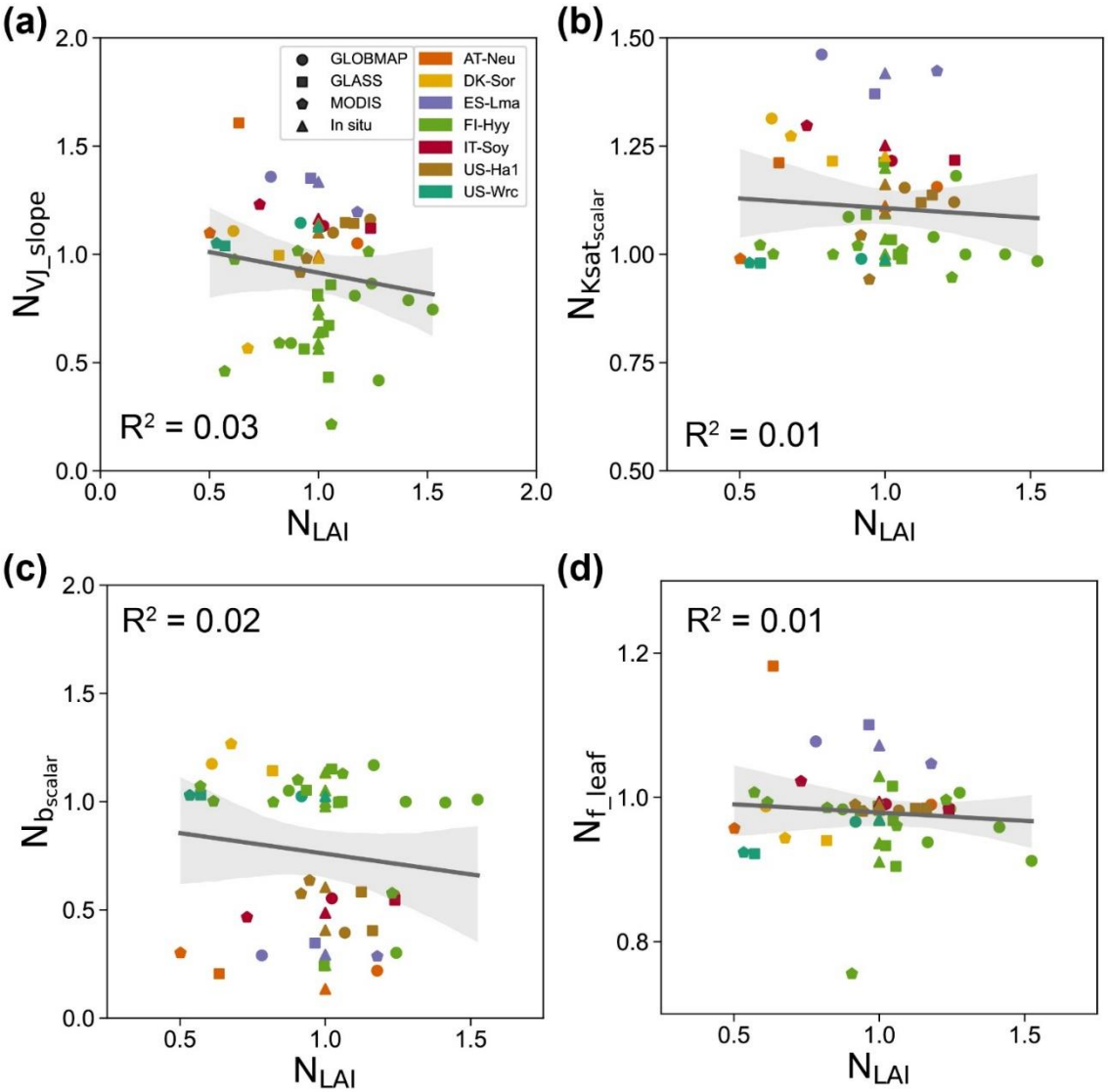


Figure S9. Influence of LAI on the posterior VJ_slope, $\boldsymbol{Ksat_{scalar}}$, $\boldsymbol{b_{scalar}}$ and f_leaf obtained by the single-site experiments conducted at seven sites and driven by four LAI data (GLOBMAP, GLASS, MODIS and *in situ*). The posterior VJ_slope, $\boldsymbol{Ksat_{scalar}}$, $\boldsymbol{b_{scalar}}$, f_leaf and the LAI were represented by their normalized values $\boldsymbol{N_{VJ\_slope}}$, $\boldsymbol{N_{Ksat_{scalar}}}$, $\boldsymbol{N_{b_{scalar}}}$, $\boldsymbol{N_{f\_leaf}}$ and $\boldsymbol{N_{LAI}}$, 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 S1. PFT and Soil Texture descriptions in BEPS model.

PFT No.	Descriptions
1	Evergreen needleleaf forest
2	Deciduous needleleaf forest

3	Deciduous broadleaf forest
4	Evergreen broadleaf forest
5	Mixed forest
6	Shrub
7	C3 grass
8	C3 crop
9	C4 grass
10	C4 crop

Soil texture No.	Description
1	Sand
2	Loamy sand
3	Sandy loam
4	Loam
5	Silt loam
6	Sandy clay loam
7	Clay loam
8	Silty clay loam
9	Sandy clay
10	Silty clay
11	Clay

Table S2. $alpha$ and $beta$ parameters for COS production term.

Site name	PFT in BEPS	PFT in Whelan et al. (2016)	$alpha$ (unitless)	$beta$ (℃$^{-1}$)
AT-Neu	C3 grass	Savanna	-9.54	0.108
ES-Lma	C3 grass	Savanna	-9.54	0.108
DK-Sor	Deciduous broadleaf forest	Temperate forest	-7.77	0.119
US-Ha1	Deciduous broadleaf forest	Temperate forest	-7.77	0.119
FI-Hyy	Evergreen needleleaf forest	Temperate forest	-7.77	0.119
US-Wrc	Evergreen needleleaf forest	Temperate forest	-7.77	0.119
IT-Soy	C3 crop	Soy field	-6.12	0.096

Table S3. Parameters for COS uptake term.

PFT in BEPS	PFT in Whelan et al. (2022)	SWC_{opt} (%)	F_{opt} (pmol m^{-2} s^{-1}) with temperature (℃) at SWC_{opt}	SWC_g (%)	F_{opt} (pmol m^{-2} s^{-1}) with temperature (℃) at SWC_g
C3 grass	Grassland	12.5	F_{opt}: -4.5 F_{T_g}: -1.5 T_{opt}: -10.9	26.9	F_{opt}: -2.3 F_{T_g}: -1.3 T_{opt}: -14.8

			T_g: -25		T_g: -25
Deciduous broadleaf forest	Forest - Temperate or broadleaf	24.6	12.6	51	-0.18T+0.48
Evergreen needleleaf forest	Forest – Boreal or needleleaf	12.5	F_{opt}: -18 F_{T_g}: -12 T_{opt}: 28 T_g: 35	19.3	F_{opt}: -5.9 F_{T_g}: -3.8 T_{opt}: 28 T_g: 35
C3 crop	Agricultural	17.7	-9.7	22	-5.36

Table S4. Description of parameters used for optimizations within the Nanjing University Carbon Assimilation System (NUCAS). Parameters are either specified per PFT, per soil texture, or globally, i.e., all PFTs and textures share one value, as indicated in column 3.

No.	Parameter	Dependent	Unit	Description	Prior Value	Prior Uncertainty
1	V_{cmax25}	PFT	μmol m^{-2} s^{-1}	maximum carboxylation rate at 25℃	62.5	15.625
2					39.1	9.775
3					57.7	14.425
4					29	7.25
5					66	16.5
6					57.85	14.4625
7					48	12
8					84.5	21.125
9					30	7.5
10					30	7.5
11	VJ_slope	PFT	unitless	Slope of the V_{cmax} and J_{max} (maximum electron transport rate) relationship	2.39	0.5975
12					2.39	0.5975
13					2.39	0.5975
14					2.39	0.5975
15					2.39	0.5975
16					2.39	0.5975
17					2.39	0.5975
18					2.39	0.5975
19					2.39	0.5975
20					2.39	0.5975
21	Q10	PFT	unitless	Soil respiration temperature factor	0.046	0.0115
22					0.046	0.0115
23					0.046	0.0115
24					0.046	0.0115
25					0.046	0.0115
26					0.046	0.0115
27					0.046	0.0115

28					0.046	0.0115
29					0.046	0.0115
30					0.046	0.0115
31	SIF_alpha	PFT	$W\ m^{-2}$	Quadratic term coefficient for the relationship between additional heat dissipation under light adapted conditions and relative reduction of photochemical yield	6.2473	1.561825
32					6.2473	1.561825
33					6.2473	1.561825
34					6.2473	1.561825
35					6.2473	1.561825
36					6.2473	1.561825
37					6.2473	1.561825
38					6.2473	1.561825
39					6.2473	1.561825
40					6.2473	1.561825
41	SIF_beta	PFT	$W\ m^{-2}$	Primary term coefficient for the relationship between additional heat dissipation under light adapted conditions and relative reduction of photochemical yield	0.5994	0.14985
42					0.5994	0.14985
43					0.5994	0.14985
44					0.5994	0.14985
45					0.5994	0.14985
46					0.5994	0.14985
47					0.5994	0.14985
48					0.5994	0.14985
49					0.5994	0.14985
50					0.5994	0.14985
51	$Ksat_{scalar}$	texture	unitless	Scaling factor of saturated hydraulic conductivity (Ksat)	1	0.25
52					1	0.25
53					1	0.25
54					1	0.25
55					1	0.25
56					1	0.25
57					1	0.25
58					1	0.25
59					1	0.25
60					1	0.25
61					1	0.25
62	b_{scalar}	texture	unitless	Scaling factor of Campbell parameter b (the exponential parameter of Campbell's soil moisture retention model)	1	0.25
63					1	0.25
64					1	0.25
65					1	0.25
66					1	0.25
67					1	0.25
68					1	0.25

69					1	0.25
70					1	0.25
71					1	0.25
72					1	0.25
73	f_leaf	global	unitless	The ratio of photosynthetically active radiation to shortwave radiation	0.5	0.125
74	kc25	global	μbar	Michaelis–Menten constants for CO_2 in 25℃	274.6	68.65
75	ko25	global	mbar	Michaelis–Menten constants for O_2 in 25℃	419.8	104.95
76	tau25	global	unitless	The CO_2/O2 specificity factor, which reflects the carbon assimilation efficiency of Rubisco	2904.12	726.03

Table S5. Summary of configurations of twin experiments. $\boldsymbol{J_{initial}}$ and $\boldsymbol{J_{final}}$ denote the initial value and the final value of the cost function $\boldsymbol{J(x)}$ respectively; $\boldsymbol{G_{initial}}$ and $\boldsymbol{G_{final}}$ denote the initial value and the final value of the gradient respectively; $\boldsymbol{D_{initial}}$ and $\boldsymbol{D_{final}}$ denote the initial value and the final value of the respectively. $\boldsymbol{D_{final}}$ denote the final value of the distance ($\boldsymbol{D_x}$) between the parameter vector and the prior parameter vector. The initial value ($\boldsymbol{D_{initial}}$) of $\boldsymbol{D_x}$ for all twin experiments is 7.48, due to an identical perturbation size (0.2) being applied.

Site name	Data duration	$J_{initial}$	J_{final}	$G_{initial}$	G_{final}	D_{final}	Relative changes of parameters (%)				
							V_{cmax25}	VJ_slope	$Ksat_{scalar}$	b_{scalar}	f_leaf
AT-Neu	June 2015	55.08	6.52E-16	48.09	6.65E-07	1.48E-07	-8.13E-10	-3.16E-09	-6.88E-10	-1.68E-09	1.24E-09
DK-Sor	June 2016	77.13	7.45E-16	77.01	1.30E-06	1.70E-08	1.55E-09	-8.85E-10	-2.82E-09	-1.08E-09	-1.80E-09
ES-Lma	May 2016	53.01	3.34E-15	51.59	1.55E-06	8.80E-10	-1.06E-09	1.88E-09	8.54E-09	7.58E-09	4.26E-11
FI-Hyy	July 2013	73.44	2.02E-17	70.43	1.10E-06	2.57E-08	1.29E-10	3.66E-10	-9.30E-11	4.46E-10	-2.01E-10
	July 2014	77.59	1.06E-17	76.83	2.97E-07	4.74E-09	3.18E-10	-6.80E-10	-2.08E-11	-1.96E-10	-1.56E-10
	August 2014	74.09	9.27E-18	70.00	4.63E-07	1.02E-09	-7.33E-11	1.22E-10	5.99E-10	4.59E-10	2.20E-10
	July 2015	72.76	1.19E-16	70.07	7.93E-07	7.58E-10	-1.16E-10	-4.87E-10	1.14E-11	7.20E-10	1.07E-09
	July 2016	75.89	1.13E-18	73.35	2.12E-07	4.53E-08	-9.64E-11	1.08E-10	3.16E-11	3.95E-11	-5.55E-12
	July 2017	73.94	8.47E-17	73.64	7.18E-07	2.45E-08	8.68E-11	7.31E-10	3.69E-12	2.01E-10	8.47E-10
IT-Soy	July 2017	50.75	5.09E-13	38.82	4.94E-07	6.98E-08	2.86E-09	-7.41E-09	2.74E-09	-5.89E-09	-5.70E-10
US-Ha1	July 2012	66.15	1.93E-19	59.66	2.05E-07	1.63E-07	-6.01E-12	7.29E-11	1.35E-11	7.87E-11	-5.81E-12
	July 2013	66.50	1.61E-17	60.25	9.99E-07	2.36E-08	4.42E-09	7.44E-10	-9.77E-11	4.07E-10	-3.52E-11
US-Wrc	August 2014	58.97	3.28E-18	46.87	1.45E-07	2.84E-08	-1.16E-10	4.40E-10	1.22E-10	-7.50E-11	6.04E-11
FI-Hyy*	August 2014	108.04	3.95E-15	119.27	1.28E-06	2.01E-08	-1.16E-10	4.40E-10	1.22E-10	-7.50E-11	6.04E-11
US-Wrc*									-3.41E-10	4.63E-10	