Environment-sensitivity functions for gross primary productivity in light use efficiency models

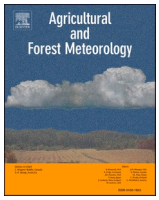
Shanning Bao, Thomas Wutzler, Sujan Koirala, Matthias Cuntz, Andreas Ibrom, Simon Besnard, Sophia Walther, Ladislav Šigut, Alvaro Moreno, Ulrich Weber, et al.

Agricultural and Forest Meteorology 312 (2022) 108708

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet



Environment-sensitivity functions for gross primary productivity in light use efficiency models

Shanning Bao [a,*], Thomas Wutzler [a], Sujan Koirala [a], Matthias Cuntz [b], Andreas Ibrom [c], Simon Besnard [a,d], Sophia Walther [a], Ladislav Šigut [a,e], Alvaro Moreno [f], Ulrich Weber [a], Georg Wohlfahrt [g], Jamie Cleverly [h], Mirco Migliavacca [a], William Woodgate [i,j], Lutz Merbold [k], Elmar Veenendaal [l], Nuno Carvalhais [a,m,*]

[a] Department of Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, Germany
[b] Université de Lorraine, AgroParisTech, INRAE, UMR Silva, 54000 Nancy, France
[c] Technical University of Denmark (DTU), Environmental Engineering, Kgs. Lyngby, Denmark
[d] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, the Netherlands
[e] Department of Matter and Energy Fluxes, Global Change Research Institute of the Czech Academy of Sciences, Bělidla 986/4a, 603 00 Brno, Czech Republic
[f] Image Processing Laboratory, University of Valencia, Valencia, Spain
[g] Department of Ecology, University of Innsbruck, Innsbruck, Austria
[h] Terrestrial Ecosystem Research Network, School of Life Sciences, University of Technology Sydney, Broadway, NSW, Australia
[i] The School of Earth and Environmental Sciences, The University of Queensland, St Lucia, 4067, Australia
[j] CSIRO, Land and Water, Canberra, 2601, Australia
[k] Agroscope, Research Division Agroecology and Environment, Reckenholzstrasse 191, 8046 Zurich, Switzerland
[l] Plant Ecology and nature Conservation Group, Wageningen University and Research centre, Wageningen, the Netherlands
[m] Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

ARTICLE INFO

Keywords:
Carbon assimilation
Radiation use efficiency
Model comparison
Model equifinality
Diffuse fraction
Sensitivity formulations
Randomly sampled sites
Temporal scales

ABSTRACT

The sensitivity of photosynthesis to environmental changes is essential for understanding carbon cycle responses to global climate change and for the development of modeling approaches that explains its spatial and temporal variability. We collected a large variety of published sensitivity functions of gross primary productivity (GPP) to different forcing variables to assess the response of GPP to environmental factors. These include the responses of GPP to temperature; vapor pressure deficit, some of which include the response to atmospheric CO_2 concentrations; soil water availability (W); light intensity; and cloudiness. These functions were combined in a full factorial light use efficiency (LUE) model structure, leading to a collection of 5600 distinct LUE models. Each model was optimized against daily GPP and evapotranspiration fluxes from 196 FLUXNET sites and ranked across sites based on a bootstrap approach. The GPP sensitivity to each environmental factor, including CO_2 fertilization, was shown to be significant, and that none of the previously published model structures performed as well as the best model selected. From daily and weekly to monthly scales, the best model's median Nash-Sutcliffe model efficiency across sites was 0.73, 0.79 and 0.82, respectively, but poorer at annual scales (0.23), emphasizing the common limitation of current models in describing the interannual variability of GPP. Although the best global model did not match the local best model at each site, the selection was robust across ecosystem types. The contribution of light saturation and cloudiness to GPP was observed across all biomes (from 23% to 43%). Temperature and W dominates GPP and LUE but responses of GPP to temperature and W are lagged in cold and arid ecosystems, respectively. The findings of this study provide a foundation towards more robust LUE-based estimates of global GPP and may provide a benchmark for other empirical GPP products.

* Corresponding authors.
E-mail addresses: sbao@bgc-jena.mpg.de (S. Bao), ncarvalhais@bgc-jena.mpg.de (N. Carvalhais).

https://doi.org/10.1016/j.agrformet.2021.108708
Received 19 May 2021; Received in revised form 24 September 2021; Accepted 1 November 2021
Available online 19 November 2021

1. Introduction

Gross photosynthetic rate, estimated by gross primary productivity (GPP), is a measure of carbon uptake by terrestrial ecosystems, which is driven by the availability of energy, water, atmospheric CO_2, and nutrients (Beer et al., 2010; Haverd et al., 2018; Ichii et al., 2005; Zhang et al., 2011). Due to complimentary and simultaneous dependencies on these factors, quantification of responses and sensitivities of GPP is often challenging (Anav et al., 2015; Ciais et al., 2014; Keenan et al., 2012; Zaehle et al., 2014).

Currently, there is a wide variety of GPP models developed for global purposes. Some models have been developed from local measurements for mechanistically representing plant function and physiology (Kaminski et al., 2013; Law et al., 2002; Lawrence et al., 2011; Vuichard et al., 2019), whereas other models are more empirical in nature (Jung et al., 2009; Teubner et al., 2019). Mechanistic models are usually robust under various climatic and environmental conditions (De Pury and Farquhar, 1997). However, their applications at a larger spatial scale are often limited due to model complexity, poorly known parameters, and divergences between them due to different embedded assumptions (Anav et al., 2015). In contrast, empirical models demand fewer drivers and parameters but are limited by data quality, representativeness, and generality (Jung et al., 2020). Generally, they lacked the theoretical links to underlying causes of productivity (Jung et al., 2011). Instead, the light use efficiency (LUE) model (Monteith, 1972), which originated from a purely empirical model, has been developed with the incorporation of some theoretical constraints, merging both advantages (Carvalhais et al., 2010a; Carvalhais et al., 2008; Mäkelä et al., 2008; Running et al., 2004).

The LUE model treated here defines GPP as a product of absorbed photosynthetically active radiation (APAR) and LUE (ε). APAR is the product of photosynthetically active radiation (PAR) and the fraction of absorbed PAR (FAPAR). ε denotes the maximum LUE (ε_{max}) scaled by a sensitivity function of environmental factors (fX). LUE models share the assumption that ε represents all the photosynthetic processes under certain light at the canopy scale and that fX denotes their responses and sensitivities to environmental factors. This assumption was later validated by Wang et al. (2017a), who derived ε and fX of P-model from detailed leaf-scale photosynthetic processes.

As a function of one or multiple environmental factors, fX differs between models. Temperature (T) is the driving factor in every LUE model due to its essential effect on enzyme activity and electron transport (Medlyn et al., 2002). The partial sensitivity function of T (fT) can be represented as a linearly peaked or sigmoidal function, indicating an increased photosynthetic rate until a plateau is reached (Mäkelä et al., 2008; Running et al., 2004). Alternatively, fT can be a bell-shaped function to account for a reduction if T exceeds a certain optimum (Horn and Schulz, 2011a; Potter et al., 1993; Xiao et al., 2004). It can also include a lagged function of T (Horn and Schulz, 2011a; Mäkelä et al., 2008) to represent delayed temperature effect on boreal and temperate vegetation (Jarvis et al., 2004; Mäkelä et al., 2004).

In addition to T, some models include vapor pressure deficit (VPD) (Horn and Schulz, 2011a; McCallum et al., 2013; Running et al., 2004; Wang et al., 2017a; Wang et al., 2015), soil water supply (W; Stocker et al., 2018), aridity index (Yan et al., 2017), precipitation index (Horn and Schulz, 2011a), land surface water index (LSWI; Xiao et al., 2004), or evaporative fraction (Horn and Schulz, 2011a; Yuan et al., 2007) as an indicator of water demand for plants. The partial sensitivity function of VPD, fVPD, is a decreasing function representing a limitation on stomatal conductance, which can be linear (Running et al., 2004) or sigmoidal (Horn and Schulz, 2011a; Kalliokoski et al., 2018; Mäkelä et al., 2008; Wang et al., 2018). In contrast, the partial sensitivity function for W (fW) is usually increasing linearly (Kalliokoski et al., 2018; Turner et al., 2006a; Xiao et al., 2004) or nonlinearly (Horn and Schulz, 2011a; Mäkelä et al., 2008; Stocker et al., 2020), since sufficient water supply can guarantee the stomatal openness and physiochemical reactions. Either fVPD or fW can include a lagged effect in arid areas (Horn and Schulz, 2011a), related to the root distribution and degree of aridity.

Some models embed a partial sensitivity function of APAR, called fL, into LUE models (Ibrom et al., 2008; Kalliokoski et al., 2018; Mäkelä et al., 2008; Propastin et al., 2012) to account for the light saturation effect (Medlyn, 1998), departing from traditionally linear LUE models.

Diffuse radiation has been found to increase total canopy ε since more light can be diffused into deeper canopy layers (Cheng et al., 2015; Farquhar and Roderick, 2003; Gu et al., 1999; Huang et al., 2014; Knohl and Baldocchi, 2008; Mercado et al., 2009). Therefore, the effect of the diffuse radiation ratio, represented by cloudiness index (CI), has been incorporated into some LUE models (He et al., 2013; Turner et al., 2006a; Wang et al., 2015; Wang et al., 2018; Yan et al., 2017), and is backed by the classical theory of radiative transfer (Ross, 1981) and adopted by mechanistic canopy models (Ibrom et al., 2006). The partial sensitivity function of CI (fCI) can be a summed (Turner et al., 2006a) or factorial form (Wang et al., 2018), both of which show a monotonically positive response to CI. Alternatively, a two-leaf LUE model structure, which separates ε_{max} and FAPAR of sunlit and shaded leaves, can account for the effect of diffuse radiation ratio on different parts of plants. fL has not been coupled with fCI in the current LUE models, possibly due to the correlation between them (Gu et al., 2002; Knohl and Baldocchi, 2008).

Another driving factor for photosynthesis is atmospheric CO_2 (Farquhar et al., 1980; Haverd et al., 2018; Kaminski et al., 2013) which can diffuse into leaves through stomata (Leuning, 1995). Nevertheless, most LUE models neglect it, with the exception of CFix (Veroustraete et al., 2002), P (Stocker et al., 2020), revised EC-LUE (Yuan et al., 2019), and PRELES models (Kalliokoski et al., 2018), meaning that traditional LUE models are not sensitive to the effect of increasing CO_2 concentrations. The effect of CO_2 variations is usually included in fVPD (Kalliokoski et al., 2018; Wang et al., 2017a) due to the dependence of stomatal conductance on VPD for controlling gas exchange. A CO_2 effect is also influenced by T, which affects the affinity of Rubisco for CO_2 molecules (Veroustraete et al., 2002; Wang et al., 2017a).

All the above differences between fXs have led to discrepancies between LUE models and GPP estimates. Although some of these models have been compared globally (Yuan et al., 2014; Zhang et al., 2015), or regionally (Carvalhais et al., 2010b; Mäkelä et al., 2006; Ruimy et al., 1999), and GPP estimates have been improved using the Bayesian average of multiple models (Zhang et al., 2021), the necessity of different environmental factors, their formulations and derived sensitivity of GPP dynamics have not been assessed.

Our study objective is to evaluate the environmental factors and partial sensitivity functions to find a generally robust LUE model for most eddy covariance (EC) sites from FLUXNET (Baldocchi et al., 2001). We hypothesized that global ecosystem-level variations in GPP and LUE would respond to independent changes in T, VPD, W, light saturation, CI, and CO_2, and that the responses (i.e., partial sensitivity functions) are nonlinear. To test these hypotheses, our study followed the workflow in Fig. 1.

2. Materials and methods

2.1. Light use efficiency model ensemble

For the first step, we built an ensemble of 5600 LUE models based on full factorial structure (Eq. (1)), which allows the most effective testing of individual environmental factor effects.

$$GPP = \varepsilon_{max} \cdot PAR \cdot FAPAR \cdot fT \cdot fVPD \cdot fW \cdot fL \cdot fCI \quad (1)$$

We selected T, VPD, W, light saturation, CI and CO_2 as the most dominant factors for governing GPP. Given the relationship between CO_2 diffusion and stomatal conductance controlled by VPD, the VPD and CO_2 effects were integrated into a single function, fVPD. We gathered

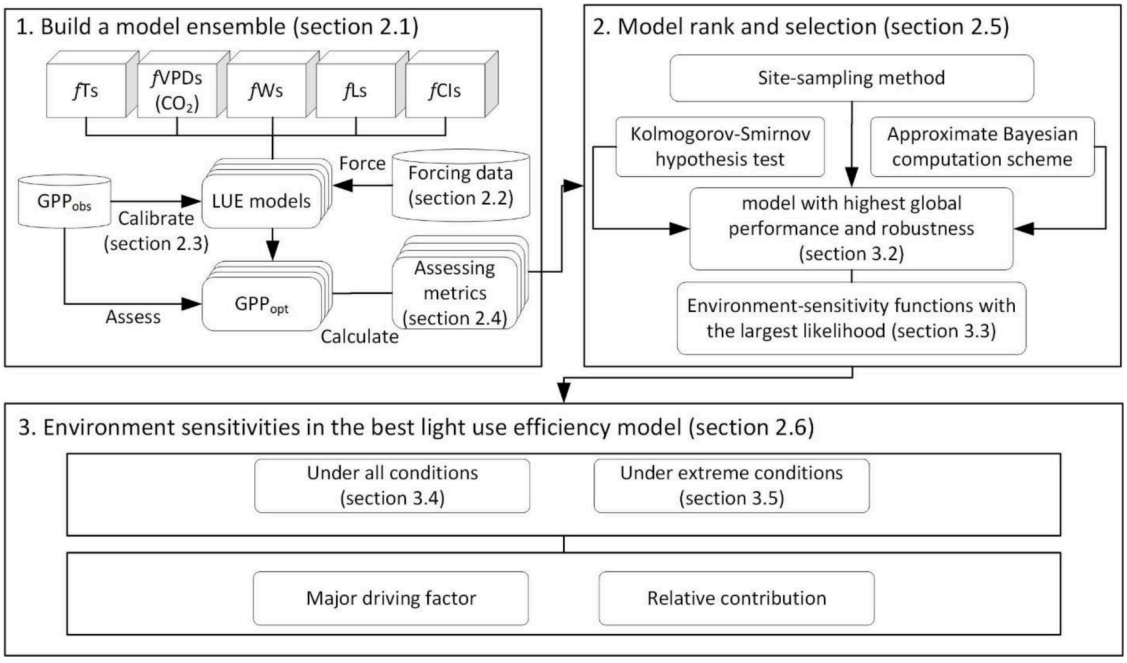


Fig. 1. The workflow of this study (GPP_{obs} and GPP_{opt} denote the GPP derived from eddy covariance measurement and simulated GPP using the LUE model, respectively).

different mathematical representations of the five kinds of independent partial sensitivity functions (six *f*Ts, six *f*VPDs, two of which include CO_2 effect, eight *f*Ws, one *f*L, and two *f*CIs, listed in Table 1) from the published LUE models (see references in Table 1). One of the *f*VPDs with a CO_2 effect, $fVPD_{P0}$, has fixed parameters for all climate and vegetation types (Stocker et al., 2020). We therefore supplemented an identical function with flexible parameters ($fVPD_P$). Additionally, we introduced a Weibull function for soil moisture ($fW_{Weibull}$), and an exponential function and a bell-shaped function for CI (fCI_{EXP} and fCI_{Horn}). We also set a contrast group for each factor (fT_{none}, $fVPD_{none}$, fW_{none}, fL_{none} and fCI_{none}, all equal to 1) to test if GPP is independent of these environmental factors. In summary, we combined seven *f*Ts, eight *f*VPDs, ten *f*Ws, two *f*Ls, and five *f*CIs (see function curves in Fig. 2, equations and references in Table 1) based on the structure in Eq. (2) to make an ensemble of 5600(=7 × 8 × 10 × 2 × 5) LUE models.

In the ensemble, eleven models have the same structures as published models (see their names and structures in supplementary information Table S1).

2.2. Forcing data and observational constraints

We forced the LUE models using ecosystem-scale observations from 196 FLUXNET EC sites (fluxnet.fluxdata.org; see distribution map in Fig. S1 and site information in Table S2). The forcing data, including PAR, FAPAR, T, VPD, W, CI, CO_2, and site elevation, were summarized in Table 2. We additionally collected GPP (GPP_{obs}), evapotranspiration (ET; ET_{obs}), and their uncertainties (see definitions in Table 2) to calibrate model parameters.

2.3. Model calibration

The purpose of model calibration is to find the best parameter vector for each LUE model at each site, and to reduce the model uncertainties associated with model parameters. We calibrated LUE model parameters, $\boldsymbol{\varepsilon_{max}}$ and parameters of *f*X (in **bold** in Table 1), and WAI parameters (see their definitions in Appendix A and the reason for jointly calibrating WAI and LUE parameters in Appendix C) using a stochastic and derivative-free evolutionary algorithm, CMAES (Hansen and Kern, 2004). CMAES, which is a reliable tool for global optimization (Hansen and Kern, 2004; Trautmann et al., 2018), searches the optimal parameters (maximum iterations=10^5 in this study) in a predefined range (Table S4) by minimizing a cost function. The definition of an appropriate cost function is, therefore, crucial to find the optimal parameters given certain constraints on model variables. We define a cost function (*cf*) as the sum of four parts (see equations of cost functions in Appendix D): GPP errors (cf_1) scaled by uncertainty for constraining LUE parameters, ET errors (cf_2) scaled by uncertainty for constraining WAI parameters, and *f*X constraints (cf_3 and cf_4).

To avoid the selection of a local optimum, we randomly selected nine additional parameter vectors using the reciprocal of *cf* as a weight in the last 10^3 iterations of the evolutionary parameter searching process (i.e., 10^3 parameter vectors) to cross-examine the optimal parameter vector given by CMAES. The ten parameter vectors were all used in model selection (Section 2.5). Parameter uncertainties were also quantified using the covariance matrix of parameters (Appendix E).

2.4. Model assessing metrics

We evaluated model fitness at daily, weekly, monthly and annual scales using the good-quality observational data defined in Appendix B. First, we aggregated daily observed GPP (*GPP*) and simulated GPP ($\widehat{GPP}$) to weekly, monthly, and annual scales. Next, we calculated the Nash-Sutcliffe efficiency (NSE) to measure the global fitness (NSE_g), spatial fitness (NSE_{sp}) and site-level fitness (NSE_i) of models using Eqs. (2)-(4).

$$NSE_g = 1 - \frac{\sum_{i=1}^{N_{s0}}\sum_{t=1}^{N_{t,i}}\left(GPP_{i,t} - \widehat{GPP}_{i,t}\right)^2}{\sum_{i=1}^{Ns0}\sum_{t=1}^{Nt,i}\left(GPP_{i,t} - \overline{GPP}_{i,t}\right)^2} \quad (2)$$

$$NSE_{sp} = 1 - \frac{\sum_{i=1}^{N_{s0}}\left(\overline{GPP}_i - \overline{\widehat{GPP}}_i\right)^2}{\sum_{i=1}^{N_{s0}}\left(\overline{GPP}_i - \overline{\overline{GPP}}_i\right)^2} \quad (3)$$

Table 1
Equations of the partial sensitivity functions (the parameters are in bold). Parameter names in the individual equations are from the original publications. Some symbols might hence appear several times (e.g., α) but are actually completely independent.

fX	Equation	Reference
$f\mathrm{T}_{\mathrm{CASA}}$	$T_{\varepsilon 1}\cdot T_{\varepsilon 2}$, $T_{\varepsilon 1} = 0.8 + 0.02\cdot \boldsymbol{T_{opt}} - \boldsymbol{T}_{opt}^2$, $T_{\varepsilon 2} = 2\cdot(1+e^{\boldsymbol{T_a}\cdot(\boldsymbol{T_{opt}}-10-T)})^{-1}\cdot(1+e^{\boldsymbol{T_b}\cdot(-\boldsymbol{T_{opt}}-10+T)})^{-1}$	(Potter et al., 1993)
$f\mathrm{T}_{\mathrm{Horn}}$	$2\times e^{-(T_f-\boldsymbol{T_{opt}})/\boldsymbol{k_T}}/(1+e^{(-(T_f-\boldsymbol{T_{opt}})/\boldsymbol{k_T})^2})$, $T_{f_k} = (1-\boldsymbol{\alpha})\cdot T_k + \boldsymbol{\alpha}\cdot T_{f_{k-1}}$	(Horn and Schulz, 2011a)
$f\mathrm{T}_{\mathrm{MOD17}}$	$(T - \boldsymbol{TMIN_{min}})/(\boldsymbol{TMIN_{max}} - \boldsymbol{TMIN_{min}})$	(Running and Zhao, 2015)
$f\mathrm{T}_{\mathrm{P}}$	$\boldsymbol{a} + \boldsymbol{b}\cdot T - \boldsymbol{c}\cdot T^2$	(Stocker et al., 2020)
$f\mathrm{T}_{\mathrm{TAL}}$	$\min(S_k/\boldsymbol{S_{max}}, 1)$, $S_k = \max(X_k - \boldsymbol{X_0}, 0)$, $X_k = X_{k-1} + 1/\boldsymbol{\tau}(T_k - X_{k-1})$, $X_1 = T_1$	(Mäkelä et al., 2008)
$f\mathrm{T}_{\mathrm{VPM}}$	$\dfrac{(T-\boldsymbol{T_{min}})\cdot(T-\boldsymbol{T_{max}})}{((T-\boldsymbol{T_{min}})\cdot(T-\boldsymbol{T_{max}})-(T-\boldsymbol{T_{opt}})^2)}$	(Xiao et al., 2004)
$f\mathrm{VPD}_{\mathrm{Horn}}$	$1/(1+e^{\boldsymbol{kw}\cdot(VPD_f-\boldsymbol{W_1})})$, $VPD_{f_k} = (1-\boldsymbol{\alpha})\cdot VPD_k + \boldsymbol{\alpha}\times VPD_{f_{k-1}}$	(Horn and Schulz, 2011a)
$f\mathrm{VPD}_{\mathrm{MOD17}}$	$1 - (VPD - \boldsymbol{VPD_{min}})/(\boldsymbol{VPD_{max}} - \boldsymbol{VPD_{min}})$	(Running and Zhao, 2015)
$f\mathrm{VPD}_{\mathrm{P0}}$	$\mathrm{m}\sqrt{1-\left(\dfrac{c^*}{m}\right)^{\frac{2}{3}}}$, $\mathrm{m} = \dfrac{c_a-\Gamma^*}{c_a+2\Gamma^*+3\Gamma^*\sqrt{\dfrac{1.6\eta^* VPD}{\beta(\kappa+\Gamma^*)}}}$, no parameter is calibrated	(Stocker et al., 2020)
$f\mathrm{VPD}_{\mathrm{P}}$	$\mathrm{m}\sqrt{1-\left(\dfrac{c^*}{m}\right)^{\frac{2}{3}}}$, $\mathrm{m} = \dfrac{c_a-\Gamma^*}{c_a+2\Gamma^*+3\Gamma^*\sqrt{\dfrac{1.6\eta^* VPD}{\boldsymbol{\beta}(\kappa+\Gamma^*)}}}$, $\boldsymbol{c^*}$ and $\boldsymbol{\beta}$ are calibrated	as above
$f\mathrm{VPD}_{\mathrm{PRELES}}$	$\left(e^{\boldsymbol{\kappa}\cdot\left(\frac{\boldsymbol{C_{a0}}}{C_a}\right)^{\boldsymbol{c_\kappa}}\cdot VPD}\right)^{*1}\cdot(1+(C_a-\boldsymbol{C_{a0}})/(C_a-\boldsymbol{C_{a0}}+\boldsymbol{c_m}))^{*2}$	(Kalliokoski et al., 2018)
$f\mathrm{VPD}_{\mathrm{TAL}}$	$e^{\boldsymbol{\kappa}\cdot VPD}$	(Mäkelä et al., 2008)
$f\mathrm{VPD}_{\mathrm{Wang}}$	$1/(1+VPD/\boldsymbol{D_0})$	(Wang et al., 2018)
$f\mathrm{W}_{\mathrm{CFlux}}$	0, *if* $(W < \boldsymbol{WHC_{min}})$; $W/\boldsymbol{WHC_{max}}$, *if* $(\boldsymbol{WHC_{min}} \le W \le \boldsymbol{WHC_{max}})$; 1, *if* $(W > \boldsymbol{WHC_{max}})$	(Turner et al., 2006a)
$f\mathrm{W}_{\mathrm{Horn}}$	$1/(1+e^{\boldsymbol{kw}\cdot(W_f-\boldsymbol{W_1})})$, $W_{f_k} = (1-\boldsymbol{\alpha})\cdot W_k + \boldsymbol{\alpha} W_{f_{k-1}}$	(Horn and Schulz, 2011a)
$f\mathrm{W}_{\mathrm{P}}$	$\mathbf{q}(W-\boldsymbol{\theta}^*)^2+1$, if (W$\le \boldsymbol{\theta}^*$); 1, if (W$> \boldsymbol{\theta}^*$)	(Stocker et al., 2020)
$f\mathrm{W}_{\mathrm{PRELES}}$	$\min(1, W/\boldsymbol{\rho})$	(Kalliokoski et al., 2018)
$f\mathrm{W}_{\mathrm{TAL}}$	$(1+((1-W)/\boldsymbol{\alpha})^{\boldsymbol{r}})^{-1}$	(Mäkelä et al., 2008)
$f\mathrm{W}_{\mathrm{TAL1}}$	$1-e^{(-\boldsymbol{\alpha}\cdot W^*)}$	(Mäkelä et al., 2008)
$f\mathrm{W}_{\mathrm{VPM}}$	$(1+LSWI)/(1+\boldsymbol{LSWI_{max}})$	(Xiao et al., 2004)
$f\mathrm{W}_{\mathrm{Wang}}$	W	(Wang et al., 2018)
$f\mathrm{W}_{\mathrm{Weibull}}$	$\dfrac{\boldsymbol{k}}{\boldsymbol{\lambda}}\left(\dfrac{W}{\boldsymbol{\lambda}}\right)^{\boldsymbol{k}-1} e^{-\left(\frac{W}{\boldsymbol{\lambda}}\right)^{\boldsymbol{k}}}$	This study
$f\mathrm{L}_{\mathrm{TAL}}$	$1/(\boldsymbol{\gamma}(PAR\cdot FAPAR)+1)$	(Mäkelä et al., 2008)
$f\mathrm{CI}_{\mathrm{CFlux}}$	$1+\boldsymbol{\mu} CI_{nor}$ $CI_{nor}(CI - \min(CI_{Y_i}))/(\max(CI_{Y_i}) - \min(CI_{Y_i}))$, no parameter	(Turner et al., 2006a)
$f\mathrm{CI}_{\mathrm{EXP}}$	$CI^{\boldsymbol{\mu}}$	This study
$f\mathrm{CI}_{\mathrm{Horn}}$	$2\times e^{-(CI-\boldsymbol{R_{opt}})/\boldsymbol{k_C}}/(1+e^{(-(CI-\boldsymbol{R_{opt}})/\boldsymbol{k_C})^2})$	This study
$f\mathrm{CI}_{\mathrm{Wang}}$	$1-\boldsymbol{\mu}(1-CI_{nor})$, $CI_{nor} = (CI - \min(CI_{Y_i}))/(\max(CI_{Y_i}) - \min(CI_{Y_i}))$	(Wang et al., 2018)
$f\mathrm{T}_{\mathrm{none}}$ /$f\mathrm{VPD}_{\mathrm{none}}$ $f\mathrm{W}_{\mathrm{none}}$ /$f\mathrm{L}_{\mathrm{none}}$ /$f\mathrm{CI}_{\mathrm{none}}$	1	This study

T: Temperature.
$\mathrm{T_k}$: Lagged temperature on day k.
VPD: Vapor pressure deficiency.
Ca: Ambient CO_2 concentration.
Γ*: Photorespiratory compensation point.
η*: Water viscosity relative to its value at 25°C.
κ: Effective Michaelis-Menten coefficient of RuBisco.
W: Soil moisture.
CI: Cloudiness index.
$\mathrm{Y_i}$: Year i
*1: fVPD part in $f\mathrm{VPD}_{\mathrm{PRELES}}$, including part of CO_2 effect, which was constrained using cf_3 (see Appendix D)
*2: $f\mathrm{CO_2}$ part in $f\mathrm{VPD}_{\mathrm{PRELES}}$, which was not constrained

$$\mathrm{NSE}_i = 1 - \frac{\sum_{t=1}^{N_{t,i}}\left(GPP_{i,t} - \widehat{GPP}_{i,t}\right)^2}{\sum_{t=1}^{N_{t,i}}\left(GPP_{i,t} - \overline{GPP}_{i,t}\right)^2} \quad (4)$$

$GPP_{i,t}$ and $\widehat{GPP}_{i,t}$ stand for the observed GPP and simulated GPP at time step t at site i. N_{s0} and $N_{t,i}$ are the total site number and the total time steps at site i. $\overline{GPP}_i$ and $\widehat{\overline{GPP}}_i$ are the average observed GPP and simulated GPP at site i. The NSE of each site-year and across site-years (used in Section 2.5) are the same as Eqs. (4) and (2), but only account for one-year GPP. Similarly, we calculated coefficient of determination (R^2), and normalized root mean squared error (nRMSE).

2.5. Model rank and selection

Model selection usually depends on the methods, data information

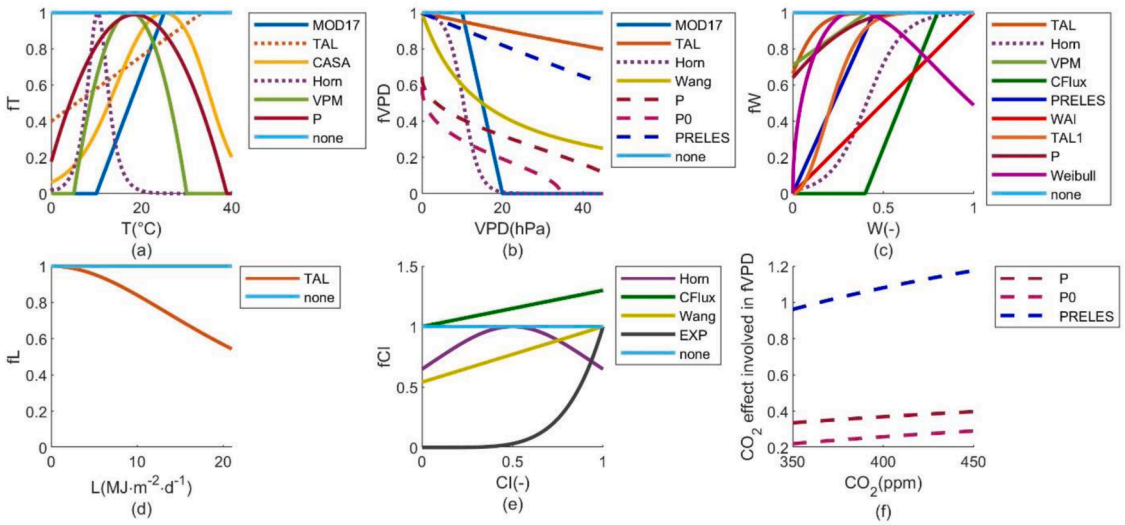


Fig. 2. Partial sensitivity functions in light use efficiency models describing the sensitivity of LUE to a) temperature (T; *f*T), b) vapor pressure deficit (VPD; *f*VPD), c) soil water availability (W; *f*W), d) light saturation (L; *f*L), e) cloudiness index (CI; *f*CI) and f) atmospheric CO_2 concentration in *f*VPD. The dotted line (…) denotes the function including a lag effect of the driving variable, the dashed line (–) denotes the function including CO_2 effect. All the response curves vary with parameters.

and assessment metrics (Burnham and Anderson, 2002; Burnham et al., 2011; Efron, 2014). Equifinality between models and between model parameters restricts the model selection for a complex ecosystem given limited observations (Beven, 1993; Fisher and Koven, 2020). Hence, it is challenging to select a robust model in a large model ensemble given a few observational sites, which are nearly 28 times fewer than the total number of models. Here, we adopted the following strategies to alleviate the influences of the model selection method, sites, assessment metrics and model equifinality.

First, to reduce uncertainties caused by using a unique ranking method, we ranked the models based on three methods: (1) a site-sampling method; (2) pair-wise Kolmogorov–Smirnov (KS) hypothesis tests (Appendix F; Lilliefors, 1967); and (3) following the approximate Bayesian computation (ABC) scheme (Appendix G; Toni and Stumpf, 2010). Among these three methods, both the site-sampling and ABC consider the equifinality between models and parameter vectors using a likelihood value to measure model robustness.

To avoid the selection of a model that is locally good for a specific group of sites, we used a site-sampling method, which defines model robustness using the likelihood of being the best model in multiple iterations of bootstrapped sites. The detailed workflow of the site-sampling method is summarized in Table 3.

The model likelihood, P_1, represents model goodness per site-year and can be regarded as the fraction of site-years in which a model is among the top 1% best models. In turn, the other model likelihood, P_2, represents model goodness across site-years, i.e., the overall goodness for all the site-years. With the model ranks R_1 and R_2, sorted according to P_1 and P_2, the model that is the locally best for most site-years and overall best can be selected.

To test if the selection of the metric affected the model selection, we repeated the above analysis with RMSE instead of NSE (*mef* at step 13 in Table 3, which corresponds to either RMSE or NSE in the respective scenarios). We also tested the other thresholds (step 21 in Table 3) to determine the best models (θ_0=2%, 3%, 4%, 5% and 10%). The results are shown in Fig. S3 and Fig. S4.

We repeated the above model selection process individually for nine climate-vegetation types: arid forest, arid grassland, boreal forest, boreal grassland, polar vegetation, temperate forest, temperate grassland, tropical forest, and tropical grassland (defined in Appendix H). Finally, we selected the global best model as the model in which the sum of R_1 and R_2 was smallest at all the sites and the sites of various climate-vegetation types; i.e., the best combination of *f*X.

To evaluate the strength of the global best LUE model, we compared it with a machine learning model (Appendix I) and an ensemble of site-best models; i.e., an ensemble of models with the highest NSE at a daily scale per site.

Furthermore, we calculated the likelihoods, P_1 and P_2, of various partial sensitivity functions and their combinations. The likelihoods were equal to the average ratio of a partial sensitivity function (e.g., $f\mathrm{T}_{\mathrm{MOD17}}$) or a combination of partial sensitivity functions (e.g., $f\mathrm{T}_{\mathrm{MOD17}}$ and $f\mathrm{VPD}_{\mathrm{MOD17}}$) that appeared in the top 1% best models of all the iterations to 1% of total number of models. We treated the partial sensitivity function and the combination with the largest sum of likelihoods as the best sensitivity function for each environmental factor or a combination of environmental factors.

2.6. Environment sensitivities in the best light use efficiency model

We analyzed the dominant controls for seasonal LUE dynamics using the partial sensitivity functions of the global best LUE model and site-best models in three approaches. In the first approach, we aimed to find the environmental factor to which LUE has the largest sensitivity. For this purpose, we calculated the average partial sensitivities and adopted the minimum as the dominant environmental factor at site *i* (X_i) as Eq. (5).

$$\mathrm{X}_i = \begin{cases} \mathrm{T},\ \text{if } \overline{\mathrm{fT}_i} = \min\limits_i\left(\overline{\mathrm{fT}_i}, \overline{f\mathrm{VPD}_i}, \overline{\mathrm{fW}_i}, \overline{\mathrm{fL}_i}, \overline{f\mathrm{CI}_i}\right) \\ \cdots \\ \mathrm{CI},\ \text{if } \overline{f\mathrm{CI}_i} = \min\limits_i\left(\overline{\mathrm{fT}_i}, \overline{f\mathrm{VPD}_i}, \overline{\mathrm{fW}_i}, \overline{\mathrm{fL}_i}, \overline{f\mathrm{CI}_i}\right) \end{cases} \quad (5)$$

Then we computed the fraction of each dominant factor ($Frac_{\mathrm{X}}$) per climate-vegetation type as Eq. (6), where N_{s0} denotes the total site number of a corresponding climate-vegetation type. We regarded the factor with the largest fraction as the major driving factor for the corresponding type.

$$Frac_{\mathrm{X}} = \frac{\sum_i(\mathrm{X}_i = \mathrm{X})}{N_{s0}},\ \mathrm{X} = \mathrm{T},\ \mathrm{VPD},\ \mathrm{W},\ \mathrm{L},\ \text{or CI} \quad (6)$$

In the second approach, we identified the relative contribution of each environmental factor (X) to the total LUE variation (C_X) per climate-vegetation type as Eqs. (7)-(8).

Table 2
Calculations and processings of the forcing data.

Abbreviation	Definition	Unit	Equation or source	Reference
Elev	Site elevation	m	FLUXNET websites and literature	Table S2
LE	Latent heat flux	$MJ{\cdot}m^{-2}{\cdot}day^{-1}$	EC observations	Table S2
NEE	Net ecosystem exchange	$gC{\cdot}m^{-2}{\cdot}d^{-1}$	EC observations	Table S2
Precip	Precipitation	mm	EC observations	Table S2
QA	Quality flags for all the variables from EC measurement	Unitless (0-1)	FLUXNET dataset	(Pastorello et al., 2020)
QC	Quality flags for all the reflectance of MCD43A4 product	Unitless	MCD43A2 quality assessment product	(Schaaf and Wang, 2015)
R_g	Global radiation	$MJ{\cdot}m^{-2}{\cdot}day^{-1}$	EC observations	Table S2
R_p	Potential radiation	$MJ{\cdot}m^{-2}{\cdot}day^{-1}$	EC observations	Table S2
R_n	Net radiation	$MJ{\cdot}m^{-2}{\cdot}day^{-1}$	EC observations	Table S2
r_{red}	Reflectance at red band	Unitless (0-1)	MCD43A4 version 6 Nadir BRDF-Adjusted Reflectance product	(Schaaf and Wang, 2015)
r_{nir}	Reflectance at near-infrared band	Unitless (0-1)	As above	As above
r_{swir}	Reflectance at short-wave infrared band (1230-1250 nm)	Unitless (0-1)	As above	As above
T	Air temperature	°C	EC observations	Table S2
VPD	Vapor pressure deficit	kPa	EC observations	Table S2
CI	Cloudiness index	Unitless (0-1)	$1 - R_g/R_p$	(Fu and Rich, 1999; Turner et al., 2006a)
CO_2	Atmospheric CO_2 concentration	ppm	Observations by NOAA/ESRL. The global annual mean atmospheric CO_2 concentration was converted to daily time steps using a linear interpolation function	www.esrl.noaa.gov/gmd/ccgg/trends/
ET_{obs}	Evapotranspiration	mm	converted from LE using a latent heat of vaporization changing with T	(Henderson-Sellers, 1984)
PET	Potential ET	mm	Estimated using Rn and T	(Priestley and Taylor, 1972)
GPP_{obs}	Gross primary productivity	$gC{\cdot}m^{-2}{\cdot}d^{-1}$	Estimated from NEE	(Reichstein et al., 2005)
LSWI	Land surface water index	Unitless (0-1)	$\frac{r_{nir} - r_{swir}}{r_{nir} + r_{swir}}$	(Xiao et al., 2004)
NDVI	Normalized difference vegetation index	Unitless (-1-1)	$\frac{r_{nir} - r_{red}}{r_{nir} + r_{red}}$	(Rouse et al., 1974)
PAR	Photosynthetically active radiation	$MJ{\cdot}m^{-2}{\cdot}day^{-1}$	$R_g \times 0.45$	
WAI	Water availability index	mm	Estimated using Precip and PET, with two site-calibrated parameters	See Appendix A and (Boese et al., 2019; Tramontana et al., 2016)
W	Soil water supply	Unitless (0-1)	$W = \min(1, WAI/AWC)$	-
σ_{LE}	Random uncertainty of ET	$MJ{\cdot}m^{-2}{\cdot}day^{-1}$	Standard deviation of LE	As above
σ_{NEE}	Random uncertainty of GPP	$gC{\cdot}m^{-2}{\cdot}d^{-1}$	Standard deviation of NEE	(Pastorello et al., 2020)
FAPAR	Fraction of absorbed PAR	Unitless (0-1)	$\begin{cases} = \text{NDVI (NDVI} > 0) \\ = 0 \text{ (NDVI} \le 0) \end{cases}$	(Myneni et al., 1997)

All the above variables are at the daily scale;
The gaps in the R_g, R_p, R_n, T, and VPD were filled using machine-learning-based downscaling (Besnard et al., 2019) of gridded product from CRUNCEP (Viovy, 2018a);
The linear relationship between FAPAR and NDVI was assumed according to Myneni et al., 1997.
QA was used to filter the reflectance data for good quality and snow-free conditions.
NDVI and LSWI were calculated at four QA-filtered pixels around EC sites and averaged across all pixels.
The time-series NDVI and LSWI were filtered by Savitzky-Golay filter (window size was eleven and polynomial order was three) (Savitzky and Golay, 1964).
Elevation and CO_2 were the particular forcing data for fVPD$_P$, fVPD$_{P0}$ and fVPD$_{PRELES}$. The other models only required PAR, T, VPD, W, CI, and FAPAR.
GPP_{obs} and σ_{NEE} were used to calibrate LUE model parameters, and ETobs and σ_{LE} were used to calibrate WAI parameters (see Appendix D).
Both QC and QA were applied to screen good-quality data used in cost functions and assessing metrics (Appendix B).

$$f\mathrm{X}_{\mathrm{nor},i} = \frac{1 - f\mathrm{X}_i}{\sum_{f\mathrm{X}}^{f\mathrm{T},f\mathrm{VPD},f\mathrm{W},f\mathrm{L},f\mathrm{CI}}(1 - f\mathrm{X}_i)},\ f\mathrm{X} = f\mathrm{T},\ f\mathrm{VPD},\ f\mathrm{W},\ f\mathrm{L},\ \text{or}\ f\mathrm{CI} \tag{7}$$

$$C_X = \sum_{i}^{N_{s0}} \left(\overline{f\mathrm{X}_{\mathrm{nor},i}}\right) \Big/ N_{s0} \tag{8}$$

$f\mathrm{X}_{\mathrm{nor},i}$ refers to the normalized partial sensitivity function at site i and $\overline{f\mathrm{X}_{nor,i}}$ is the average partial sensitivity function at site i. The C_X indicates the relative contribution of X from small to large (zero to one).

In the last approach, we assessed the dominant controls under extreme conditions, defined as a condition of T, VPD or APAR larger than the 90th percentile, or T, W or CI smaller than the 10th percentile values. We aimed to identify which factor most affected GPP when an environmental condition is limiting. We calculated the site fraction of major driving factors and the relative contribution of each environmental factor per climate-vegetation type under extreme conditions.

To distinguish the sensitivities of GPP and LUE and cross-validate the dominant controls, we compared the above variables with the Pearson correlation coefficients (R) between GPP_{obs} and daily forcing variables per site, and between LUE_{obs}, GPP_{obs} by APAR, and the same forcings.

Table 3
Workflow of model ranking based on site-sampling method.

Steps	Pseudocode
1	**Requirement**: *ModelSet* = all the LUE models *ParSet* = ten parameter vectors for each LUE model at each site *SiteSet* = all the EC sites
2	**Initialization**: set total model number N_m=5600, total parameter vector number N_p=10, total bootstrap times N_{itr} =200, bootstrap site number N_s=150, bootstrap site-year number N_y=2, bootstrap time l=1
3	**While** $1 \leq l \leq N_{itr}$
4	*SiteSample* = generate a bootstrap sample N_s sites from *SiteSet*
5	*SiteYearSample* = generate bootstrap sample of N_y years of data from each site in *SiteSample*
6	**For** $m = 1$ **to** N_m
7	**For** $j = 1$ **to** $N_s \bullet N_y$
8	**For** $k = 1$ **to** N_p
9	Simulate GPP ($\widehat{GPP}_{j,m,k}$) using model m and parameter vector $Par_{m,k}$
10	Store GPP in a global dataset across all site-years ($\widehat{GPP}_{m,k}$)
11	**For** time scale *ts*=daily, weekly, monthly, annual scale
12	Aggregate simulated GPP and observed GPP to time scale *ts* ($\widehat{GPP}_{j,m,k,ts}$, $\widehat{GPP}_{m,k,ts}$, $GPP_{j,m,k,ts}$ and $GPP_{m,k,ts}$)
13	Estimate model fitness, which refers to NSE if *ts* is daily, weekly or monthly scale, and Root Mean Squared Error (RMSE) if *ts* is annual scale, for each site-year j ($mef_{j,m,k,ts}$) and across all site-years ($mef_{0,m,k,ts}$)
14	Normalize $mef_{j,m,k,ts}$ and $mef_{0,m,k,ts}$, $nmf_{j,m,k,ts} = \frac{mef_{j,m,k,ts} - \min_m(mef_{j,m,k,ts})}{\max_m(mef_{j,m,k,ts}) - \min_m(mef_{j,m,k,ts})}$ $nmf_{0,m,k,ts} = \frac{mef_{0,m,k,ts} - \min_m(mef_{0,m,k,ts})}{\max_m(mef_{0,m,k,ts}) - \min_m(mef_{0,m,k,ts})}$
15	Calculate two weights of M_i ($w_{1,m,ts,l}$ and $w_{2,m,ts,l}$) using the average $nmf_{j,m,k,ts}$, and using average $nmf_{0,m,k,ts}$, $w_{1,m,ts,l} = \sum_j \frac{\sum_{k=1}^{N_p} nmf_{j,m,k,ts}}{N_p} / (N_s \cdot N_y)$ $w_{2,m,ts,l} = \frac{\sum_{k=1}^{N_p} nmf_{0,m,k,ts}}{N_p}$
16	**End For**
17	**End For**
18	**End For**
19	**End For**
20	Rank models according to $w_{1,m,ts,l}$ and $w_{2,m,ts,l}$ and calculate the average rank of all time scales ($R_{1,m,l}$ and $R_{2,m,l}$)
21	Calculate model likelihoods $P_{1,m,l}$ and $P_{2,m,l}$ according to $R_{1,m,l}$ and $R_{2,m,l}$ relative to a threshold θ_0 (=1% of N_m), $P_{1,m,l} = \begin{cases} 1, & R_{1,m,l} \leq \theta_0 \\ 0, & R_{1,m,l} > \theta_0 \end{cases}$ $P_{2,m,l} = \begin{cases} 1, & R_{2,m,l} \leq \theta_0 \\ 0, & R_{2,m,l} > \theta_0 \end{cases}$
22	$l = l$+1, go to step 3
23	Calculate the average likelihoods across all bootstrap iterations ($P_{1,m}$ and $P_{2,m}$), sort models according to $P_{1,m}$ and $P_{2,m}$, and get the model rank $R_{1,m}$ and $R_{2,m}$, respectively

3. Results

3.1. Overall model performance

On a global scale, most calibrated models had good performance across different time scales (Fig. 3). However, none of the models was selected as the best one at each site or site-year. The site-best model ensemble included 166 different models for 196 sites in total. No single model was among the best 1% models at all sites (reflected by P_1). Considering each site and site-year equally (reflected by P_2), some models ranked at the top 1% regardless of the sites considered. P_1 and P_2 (Fig. 4) had maximum values of 0.82 and 0.83, respectively, which is better than published models (P_1, P_2=0). Therefore, there were novel LUE models, i.e., fX combinations, that were equally good for all sites and are better than published models.

3.2. Model with the highest global performance and robustness

According to the site-sampling method, the global best model (model #1, with $f\mathrm{T}_{\mathrm{Horn}}$, $f\mathrm{VPD}_{\mathrm{PRELES}}$, $f\mathrm{W}_{\mathrm{Horn}}$, $f\mathrm{L}_{\mathrm{TAL}}$ and $f\mathrm{CI}_{\mathrm{EXP}}$) with the highest rank (i.e., the smallest sum of R_1 and R_2) was found to be Eq. (9).

$$\mathrm{T}_{\mathrm{f}t} = (1 - \boldsymbol{\alpha}) \cdot \mathrm{T}_t + \boldsymbol{\alpha} \cdot \mathrm{T}_{\mathrm{f}t-1} \tag{10}$$

$$\mathrm{W}_{\mathrm{f}t} = (1 - \boldsymbol{\alpha}) \cdot \mathrm{W}_t + \boldsymbol{\alpha} \cdot \mathrm{W}_{\mathrm{f}t-1} \tag{11}$$

$\boldsymbol{\varepsilon_{max}}$, $\boldsymbol{T_{opt}}$, $\boldsymbol{k_T}$, $\boldsymbol{\kappa}$, $\boldsymbol{c_\kappa}$, $\boldsymbol{C_{a0}}$, $\boldsymbol{c_m}$, $\boldsymbol{k_W}$, $\boldsymbol{W_I}$, $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$, and $\boldsymbol{\alpha}$ denote the model parameters, and the others were forcing data. T at day t includes a lag effect scaled by $\boldsymbol{\alpha}$ of the previous day t-1 (Eq. (10)) in polar, boreal, and temperate areas. The same lag effect was included for W in arid areas (Eq. (11)). In the other areas, $\boldsymbol{\alpha}$=0, and $\mathrm{W_f}$ and $\mathrm{T_f}$ were equal to T and W.

$$\mathrm{GPP} = \boldsymbol{\varepsilon_{max}} \cdot \mathrm{PAR} \cdot \mathrm{FAPAR} \cdot \left(\frac{2 \times \mathrm{e}^{-\left(\mathrm{T_f} - \boldsymbol{T_{opt}}\right)/\boldsymbol{k_T}}}{1 + \mathrm{e}^{\left(-\left(\mathrm{T_f} - \boldsymbol{T_{opt}}\right)/\boldsymbol{k_T}\right)^2}}\right) \cdot \left(\left(\mathrm{e}^{\boldsymbol{\kappa} \cdot \left(\frac{\boldsymbol{C_{a0}}}{\mathrm{C_a}}\right)^{\boldsymbol{c_\kappa}} \cdot \mathrm{VPD}}\right) \cdot \left(1 + \frac{\mathrm{C_a} - \boldsymbol{C_{a0}}}{\mathrm{C_a} - \boldsymbol{C_{a0}} + \boldsymbol{c_m}}\right)\right) \cdot \left(\frac{1}{1 + \mathrm{e}^{\boldsymbol{k_W} \cdot \left(\mathrm{W_f} - \boldsymbol{W_I}\right)}}\right) \cdot \left(\frac{1}{\boldsymbol{\gamma}(\mathrm{PAR} \cdot \mathrm{FAPAR}) + 1}\right) \cdot (\mathrm{CI}^{\boldsymbol{\mu}}) \tag{9}$$

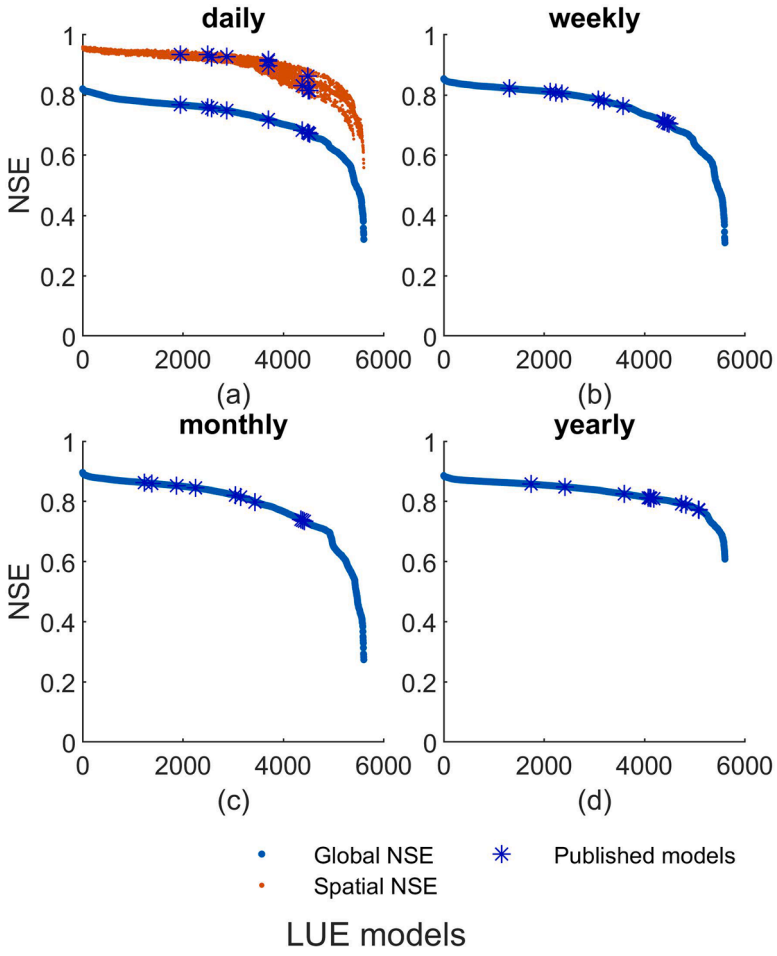


Fig. 3. The distribution of Nash-Sutcliffe model Efficiency (NSE) of LUE models at a) daily, b) weekly, c) monthly, and d) yearly scale. Global NSE represents the NSE calculated using GPP of all sites, which is the same one in Figure 6. Spatial NSE represent the NSE calculated using site-averaged GPP. Blue asterisks represent the published models collected in this study

Model #1 had the highest sum of ranks for all sites and climate-vegetation types. It was most frequently involved among the best 1% models even when the sites were bootstrapped and different climate-vegetation types were separated (Fig. 5). The likelihood, P_1 and P_2, was lower in arid grasslands, arid forests, polar vegetation, and temperate forests than other climate-vegetation types, whereas model #1 was not significantly different from the best model at the site subsets. Site NSE values from daily to annual scales were also not statistically different according to a KS test with a significance level of 0.05. Model #1 was also the best model using RMSE as the assessment metric (refers to the *mef* at step 13 in Table 3) instead of NSE (Fig. S3) and using other thresholds (θ_0 =2%, 3%, 4%, 5%, or 10% at step 21 in Table 3, Fig. S4). It was the 13th- and second-best model according to the hypothesis test (Fig. S5) and ABC (Fig. S6), respectively. The best models based on the two methods, which had a different fW ($f\mathrm{W}_{\mathrm{CFlux}}$) and fCI ($f\mathrm{CI}_{\mathrm{Wang}}$) and a different fCI ($f\mathrm{CI}_{\mathrm{Wang}}$) from model #1, were the 42nd- and the third-best models of the site-sampling method. However, the KS test at a significance level of 0.05 did not show a difference between model #1 and these two models at site-level, across which fWs ($f\mathrm{W}_{\mathrm{CFlux}}$ and $f\mathrm{W}_{\mathrm{Horn}}$) and fCIs ($f\mathrm{CI}_{\mathrm{EXP}}$ and $f\mathrm{CI}_{\mathrm{Wang}}$) were quite similar (Section 3.3). The rankings of the models with $f\mathrm{T}_{\mathrm{Horn}}$, $f\mathrm{VPD}_{\mathrm{PRELES}}$, and $f\mathrm{L}_{\mathrm{TAL}}$, and various fWs and fCIs, were higher than other models according to P_2 (zoom-in windows in Fig. 4). Hence, model #1 was a robust model across the climate-vegetation types.

Compared with the published models and all the other models in the ensemble, the global best model, model #1, could simulate the global GPP with good performance (global NSE=0.82,0.85,0.89, and 0.88, global nRMSE=40%,34%,29%, and 19% at daily, weekly, monthly and annual scales, Fig. 6). It could also reflect the spatial variability of site-averaged GPP (spatial NSE=0.96, spatial nRMSE=11%). At the site-level, model #1 could fit the temporal GPP variations of 85% sites at daily and weekly scales, and 73% sites at the monthly scale, when using NSE larger than 0.5 as a criterion. Simulated GPP at the sites with the largest, median, 25th percentile, and the smallest site NSE (= 0.96, 0.73, 0.58 and -0.04) at the daily scale were exemplified in Fig. S7a-d. The performance of model #1 was close to the site-best models (Fig. 7), which indicated generally good representativeness of the global best model at a site-level. However, the performance of model #1 for inter-annual variations was poor. Only 26% of site NSE and 41% of site R^2 were larger than 0.5 at annual scales. This is also apparent in the GPP estimates using site-best models and $\mathrm{ML}_{\mathrm{best}}$ (Fig. S8), which reflected the common limitations in LUE and machine learning models to capture interannual variability of GPP given current forcing data. Although $\mathrm{ML}_{\mathrm{best}}$ generally performed better than LUE models, $\mathrm{ML}_{\mathrm{best}}$ tended to overfit the GPP and result in considerable error over the period without good-quality forcing data (Fig. S9 a and b). In turn, LUE models did not have the same issue due to functional constraints.

3.3. Environment-sensitivity functions with the largest likelihood

No $f\mathrm{T}_{\mathrm{none}}$, and few $f\mathrm{VPD}_{\mathrm{none}}$, $f\mathrm{W}_{\mathrm{none}}$ and $f\mathrm{CI}_{\mathrm{none}}$ appeared in the best models for all sites (Fig. 8) or various climate-vegetation types (Fig. S10). Although some of the environmental factors and partial sensitivity functions were correlated (e.g., APAR and CI, fL and fCI, see

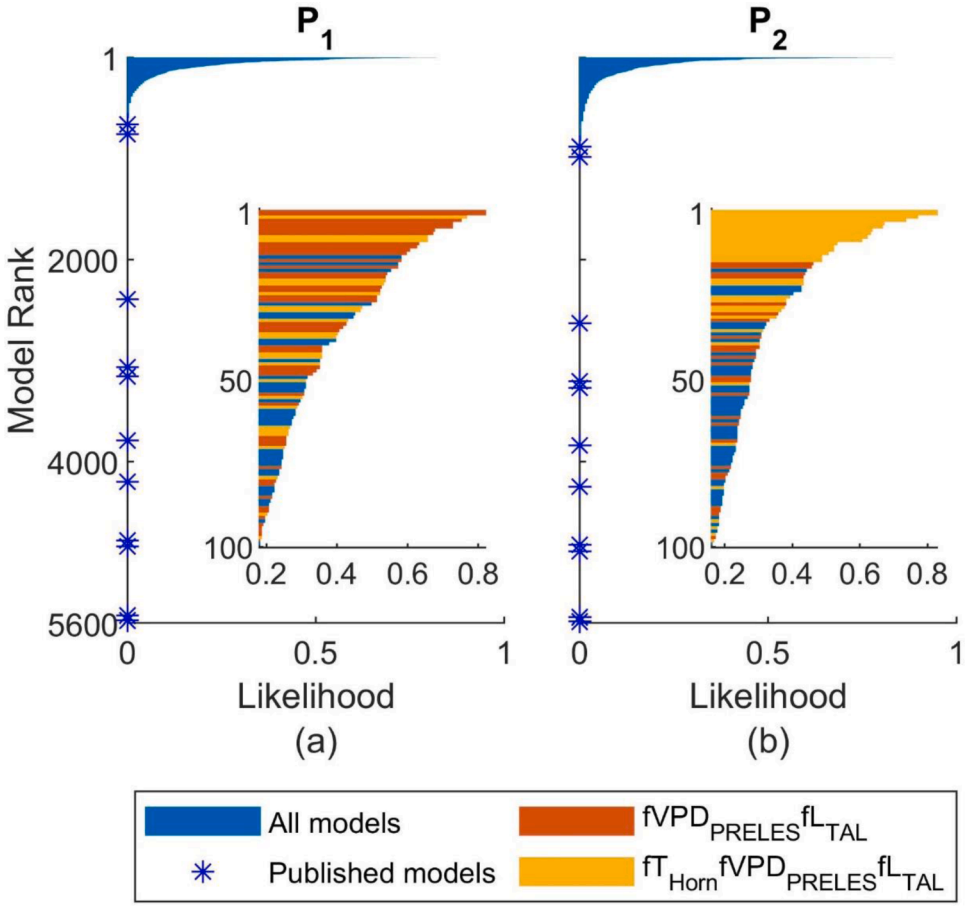


Fig. 4. Likelihoods of all the LUE models based on site-sampling method: (a) P_1 is the likelihood of a model that is among the best 1% models for each site-year, which represents a local goodness and can be regarded as a site fraction at which a model is among the best 1% models; (b) P_2 is the likelihood of a model among the best 1% models across site-years, which represents an overall goodness. The two zoom-in windows are the best 100 models sorted according to P_1 and P_2. The models with the structures of $f\mathrm{VPD}_{\mathrm{PRELES}}$ and $f\mathrm{L}_{\mathrm{TAL}}$ (in red color and yellow color) typically have a higher rank than other models (in blue color). Among these models, the models with the structure of $f\mathrm{T}_{\mathrm{Horn}}$, $f\mathrm{VPD}_{\mathrm{PRELES}}$ and $f\mathrm{L}_{\mathrm{TAL}}$ (in yellow color) have higher ranks according to P_2. The likelihoods of the published models (blue asterisk) are all equal to zero, which represent that these published models were never among the best 1% models.

Fig. S11), model improvement was apparent when the variability of each factor was considered (Fig. S12). Independent responses to saturated light and CO_2 fertilization (included in $f\mathrm{VPD}_{\mathrm{PRELES}}$, $f\mathrm{VPD}_{\mathrm{P}}$ and $f\mathrm{VPD}_{\mathrm{P0}}$) could be ignored in some best models for some specific climate-vegetation types, for example, their likelihoods in arid (Fig. S10b and g, d and i) and polar areas (Fig. S10v and x) were not superior. However, they were important for boreal and temperate forests (Fig. S10l and aa, n and ac) with much larger likelihoods than other partial sensitivity functions. Thus, considering light saturation and CO_2 fertilization is necessary for a global LUE model.

Some partial sensitivity functions were significantly better than others for representing the response of LUE to changes in environmental factors globally. $f\mathrm{VPD}_{\mathrm{PRELES}}$, a joint VPD and CO_2 effect, and $f\mathrm{L}_{\mathrm{TAL}}$, a light saturation effect, were substantially better when compared to the other fVPDs (Fig. 8b and d) and $f\mathrm{L}_{\mathrm{none}}$ (Fig. 8g and i) for more site-years (indicated by P_1) and across site-years (indicated by P_2). $f\mathrm{T}_{\mathrm{Horn}}$, a bell-shaped function with a lag effect in cold climates (i.e., polar, boreal and temperate), was superior to the other fTs (Fig. 8a and f). As for fW and fCI, several functions performed similarly (Fig. 8c,e,h and j), except for the poor performance of $f\mathrm{W}_{\mathrm{WAI}}$, in which the slope responding to the variations in W was fixed. Thus, many models with $f\mathrm{T}_{\mathrm{Horn}}$, $f\mathrm{VPD}_{\mathrm{PRELES}}$ and $f\mathrm{L}_{\mathrm{TAL}}$ and different fWs and fCIs were among the best models (Fig. 4, yellow color). The best models selected according to KS test and ABC, which had the same fT, fVPD and fL but different fW and fCI, were also similar to model #1.

The combination of paired partial sensitivity functions did not display apparent universal features except two of them. One was the $f\mathrm{VPD}_{\mathrm{PRELES}}$ and $f\mathrm{L}_{\mathrm{TAL}}$, with a much larger likelihood than the other combinations of fVPD and fL (Fig. S13). The other was the $f\mathrm{T}_{\mathrm{Horn}}$ and $f\mathrm{L}_{\mathrm{TAL}}$. The likelihoods were 0.32 and 0.50 according to P_1 and P_2. The combinations of other partial sensitivity functions, e.g., fVPD and fW or fT and fCI, were all very close in terms of the likelihood shown in the best models.

The best partial sensitivity functions differed among various climate-vegetation types (Fig. S10). For example, $f\mathrm{T}_{\mathrm{Horn}}$ was more relevant for the polar vegetation, boreal vegetation, temperate grasslands and tropical forests, but less important for vegetation in an arid climate and temperate forests. Instead, the linearly peaked function, $f\mathrm{T}_{\mathrm{MOD17}}$, was the best for arid vegetation and temperate forests. $f\mathrm{VPD}_{\mathrm{PRELES}}$, including a CO_2 effect, was superior for all climate-vegetation types except the arid forest and arid grassland, where $f\mathrm{VPD}_{\mathrm{Horn}}$ with a delayed VPD function was better, and polar vegetation and tropical forest. By contrast, $f\mathrm{W}_{\mathrm{Horn}}$ involving a lag effect of W outperformed in the arid forests and grasslands. $f\mathrm{W}_{\mathrm{Weibull}}$, a bell-shaped or a decreasing function, was more suitable for temperate and tropical grasslands. Furthermore, fCIs, $f\mathrm{CI}_{\mathrm{EXP}}$, $f\mathrm{CI}_{\mathrm{Wang}}$ and $f\mathrm{CI}_{\mathrm{Horn}}$ particularly, were similar for every climate-vegetation type. The $f\mathrm{CI}_{\mathrm{none}}$ only appeared with very low likelihood. $f\mathrm{L}_{\mathrm{none}}$ occurred in the best models for every climate-vegetation type and was almost as good as $f\mathrm{L}_{\mathrm{TAL}}$ for arid forests, arid grasslands and polar vegetation. In general, $f\mathrm{VPD}_{\mathrm{PRELES}}$ and $f\mathrm{L}_{\mathrm{TAL}}$ were necessary for most climate-vegetation types, whereas $f\mathrm{T}_{\mathrm{Horn}}$ was superior for cold sites, and $f\mathrm{VPD}_{\mathrm{Horn}}$ and $f\mathrm{W}_{\mathrm{Horn}}$ in arid sites. The other fWs and fCIs were similar among various climate-vegetation types.

3.4. Dominant controls of various climate-vegetation types

The global best model, model #1, showed a clear pattern of environmental effects changing with climate-vegetation types. T was the major driving factor (defined in Section 2.6) for all polar and boreal sites, along with 54% and 78% of the temperate forests and grasslands, respectively (Fig. 9a). The relative contribution (defined in Section 2.6) of T was also the largest at the polar, boreal and temperate sites but was

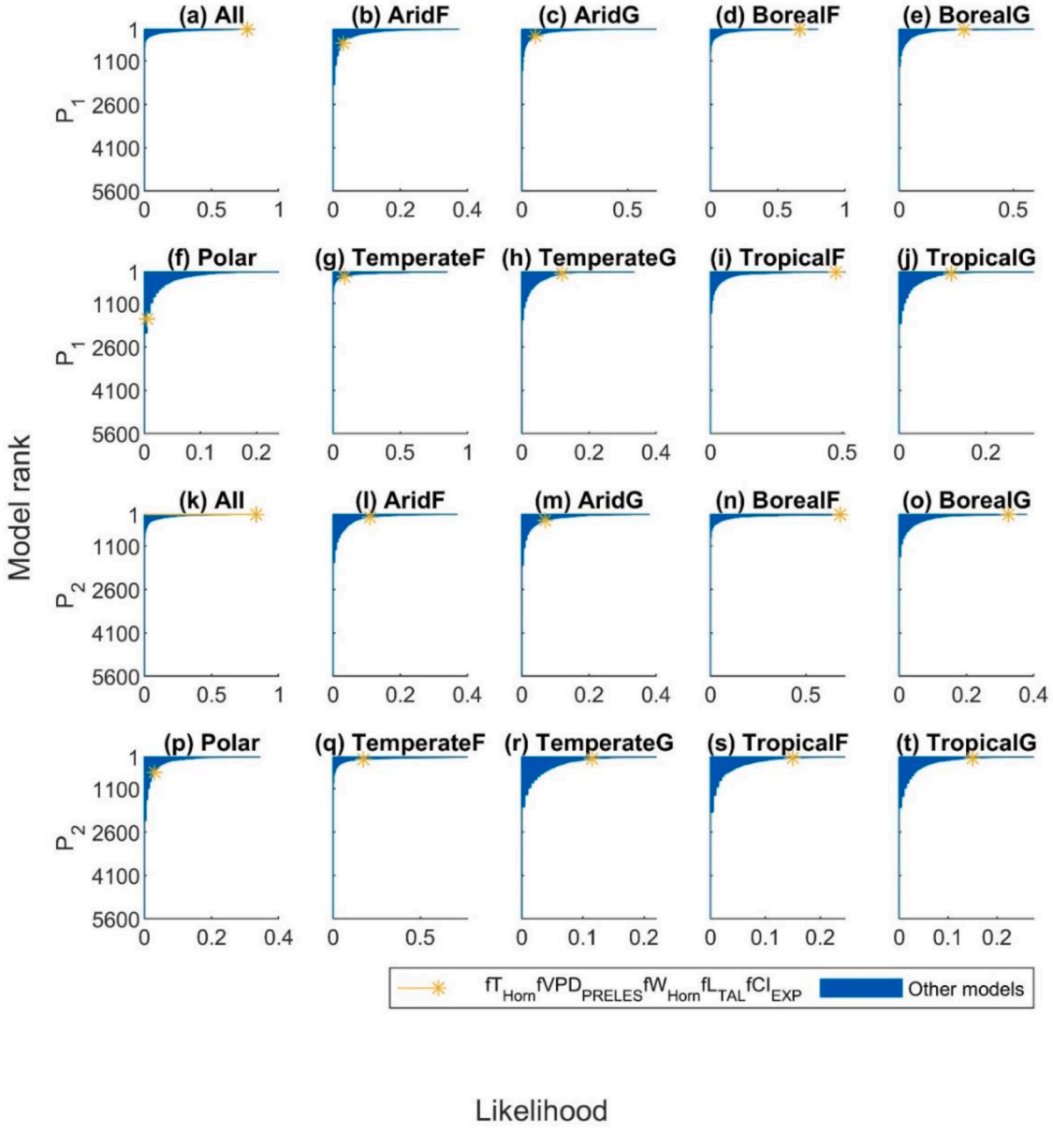


Fig. 5. Ranks of all the LUE models sorted based on site-sampling method for all sites ('All') and different climate-vegetation types (arid forest='AridF', arid grassland='AridG', boreal forest='BorealF', boreal grassland=BorealG, polar vegetation='Polar', temperate forest='TemperateF', temperate grassland='TemperateG', tropical forest='TropicalF' and tropical grassland='TropicalG'). P_1(a-j) represents a local goodness for each site-year, and P_2 (k-t) corresponds to global goodness across all site-years that belong to corresponding climate-vegetation types. The yellow line with asterisk refers to the model with the highest sum of ranks for all sites and different climate-vegetation types, which is the global best model selected based on site-sampling method. The other models are in blue.

rarely higher than 50% (Fig. 9b). W was the major driving factor for 63% of arid forests and for 57% arid grasslands. Similarly, the relative contribution of W was the largest for arid sites. CI and VPD, including CO_2, were the respective major driving factors for 40% of tropical forest and for 100% of tropical grassland sites. The contributions of CI and VPD, including CO_2, were the largest to tropical forests and tropical grasslands, respectively. However, light saturation effect was less observed compared with other factors for most sites. In particular, it only had a small contribution to arid grassland and polar sites.

The pattern of environmental effects on LUE shown by model #1 was similar to the site-best models (Fig. S14). The biggest difference between them was in the tropical grasslands, CI was the major driving factor and had the largest contribution instead of VPD. Moreover, R^2 between the *f*T, *f*VPD, *f*W, *f*L and *f*CI derived from the global best model and site-best models was above 0.7 at 73%, 66%, 56%, 87% and 93% sites. Thus, the global best model could generally reflect the local responses of LUE to the environmental factors.

The pattern of environmental effects on modeled LUE was however not consistent with the correlations between environmental factors and GPP_{obs} or LUE_{obs}. The factor that correlated most strongly with GPP_{obs} for most climate-types was APAR instead of T or CI (Fig. S15a). W and T were also strongly correlated with GPP_{obs} in arid sites and tropical sites, and in temperate and cold sites, respectively. The same was observed in the average R^2 between environmental factors and GPP_{obs} (Fig. S15b). By contrast, CI was correlated most strongly with LUE_{obs} (Fig. S15c) at cold sites. The average R between CI and LUE_{obs} was also the largest in cold sites (Fig. S15d). Nevertheless, the correlation of T with LUE_{obs} was extremely low compared with its correlation to GPP_{obs}. Thus, GPP was more correlated with APAR, whereas LUE was more correlated with CI. Correlations alone could not reflect the dominant controls of the different environmental conditions on GPP or LUE. The delayed effects of T and W variations in cold and arid sites could not be captured using

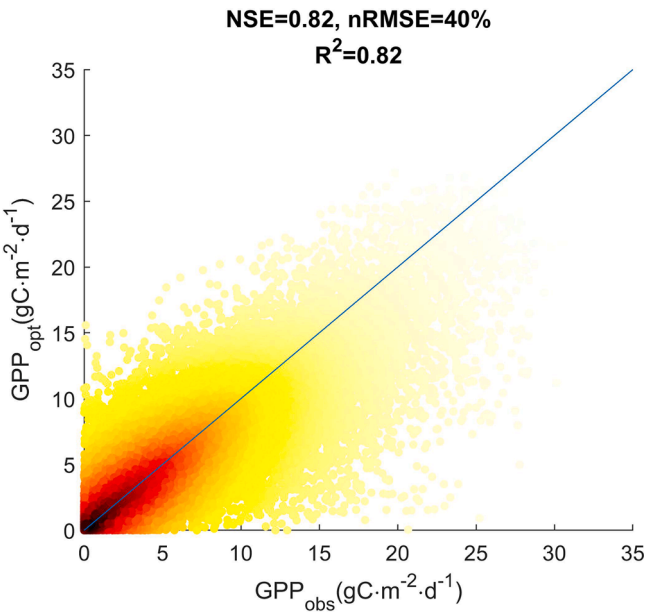


Fig. 6. Scatter of GPP estimated using EC data (GPP_{obs}) against GPP simulated by the global best LUE model (GPP_{opt}). The scatter color from black to yellow represents the scatter density (from large to small), which is related to the magnitude of GPP. NSE, nRMSE, and R^2 denotes the Nash-Sutcliffe model efficiency, root mean square error normalized by average GPP_{obs}, and coefficient of determination calculated using all GPP modeled by the good quality data. The blue line is a 1:1 line.

the correlation of instantaneous variables.

3.5. Dominant controls under extreme conditions

Intuitively, under extreme conditions of a given environmental factor, LUE should be dominated by such a factor. For example, when T was below the 10th percentile, it was the major driving factor for most sites and had the largest relative contribution compared with other factors except in arid forests and tropical vegetation (the sixth row in Fig. 10). Similarly, once W is below the 10th percentile, it could also directly affect the arid vegetation and tropical grassland (the fourth rows in Fig. 10a and b).

However, we see that LUE variability could be dominated by an environmental factor that is not under extreme conditions co-occuring when another factor is under extreme conditions. While APAR was above the 90th percentile (the third rows in Fig. 10a-i), light saturation was the major driving factor of LUE at some arid forests, boreal vegetation, temperate vegetation and tropical vegetation sites. However, the relative contribution of T, VPD, W or CI was often larger. In contrast, CI was the major driving factor and gave the largest contribution to the boreal forest and temperate forest sites (the fifth columns in Fig. 10c and f) when one of the other factors was under extreme conditions (T$\geq$90th percentile, VPD$\geq$90th percentile, W$\leq$10th percentile, APAR$\geq$90th percentile, or CI$\leq$10th percentile). In the arid sites, W was always the major driving factor and had the largest contribution.

When the environmental factors were all under extreme conditions simultaneously, which refers to T, VPD, APAR were above the 90th percentile and CI was below the 10th percentile, the major driving factor was T for most sites ('Co-occur' in Fig. 10). However, the relative contribution of W in arid vegetation and CI in other climate-vegetation types were generally more considerable than T. Moreover, a similar pattern could be observed from the site-best models (Fig. S16), which demonstrated the robustness of the above spatial pattern of these environmental effects on LUE.

4. Discussion

Through a comprehensive comparison between various combinations of environmental factors and their sensitivity functions, our results support the initial hypothesis that global ecosystem-level variations in GPP and LUE would respond to changes in T, VPD, W, L, CI, and CO_2. These environment-sensitivity functions reflect underlying mechanisms of the responses of GPP to these factors (Section 4.1). The robustness and uncertainties of the selected model and environment-sensitivity functions that might affect the derived environment effects are discussed in Section 4.2. A general evaluation of the proposed approach and its utility for data-driven global GPP estimation is presented in Section 4.3.

4.1. Responses of gross primary productivity to environmental factors

The initial hypothesis was that the responses of GPP to T, VPD, W, L, CI, and CO_2 were all significant and nonlinear even though the variables are correlated to some extent both temporally and spatially. The underlying mechanisms of such responses are discussed in Sections 4.1.1-4.1.4.

4.1.1. Temperature

Given certain light, photosynthetic capacity can be facilitated by increasing T but reduced when T exceeds a certain optimum (Bernacchi et al., 2003). The limitation of high temperature on GPP is related to high VPD and low W (Fu et al., 2020; Sims and Bradford, 2001). However, a bell-shaped function (e.g., $f\Gamma_{Horn}$) can better represent the global response of GPP to T than a sigmoidal or linearly peaked function, reflecting that the reduction in GPP at high temperatures cannot be completely compensated by considering other limitations (e.g., high VPD or low W) in an LUE model (Bernacchi et al., 2003; Potter et al., 1993; Xiao et al., 2004). The response of LUE to T can be adequately represented by a linearly peaked function, $f\Gamma_{MOD17}$, in only arid areas (Fig. S10f), because other limiting factors (e.g., low W) dominate in hot climates (Fig. 10a and b; Guo et al., 2015; Zhang et al., 2017a). Hence, considering only a positive response of GPP and LUE to T can lead to underestimation of the sensitivity of GPP to high temperature even if combined with the other factors (Horn and Schulz, 2011a; Hwang et al., 2008; Turner et al., 2006b).

A more flexible $f\Gamma$, which can be either a peaked or bell-shaped function (e.g., $f\Gamma_{CASA}$), cannot represent the global response of GPP to T due to lack of a lag function for T in boreal and temperate climates. The lagged T effect on GPP and LUE has been observed (Pelkonen and Hari, 1980) and attributed to the incomplete recovery of reduced photosynthesis capacity in needleleaves as a result of damage caused by low winter temperature (Bergh et al., 1998; Tanja et al., 2003; Yang et al., 2020), thawing of frozen soil (Jarvis and Linder, 2000), and occurrence of night frosts in spring (Bergh and Linder, 1999). Our results demonstrate that a lagged T effect is non-negligible for coniferous (Mäkelä et al., 2004) and deciduous forests (Jarvis et al., 2004) in boreal and temperate climate (i.e., cold ecosystems; Horn and Schulz, 2011a). The effect indicated by the lag parameter of $f\Gamma_{Horn}$ in the global best model follows Mäkelä et al's (2008) study instead of Horn et al's study (2011). Polar and boreal vegetation showed an apparent delayed response, whereas vegetation from a temperate climate reacted faster to T changes (Fig. S17a), possibly due to a warmer spring and different plant species (Yang et al., 2020). Although the frost-induced reduction in photosynthesis capacity also happens during autumn (Bergh et al., 1998; Hollinger et al., 1999), which was not considered in this study, the loss of GPP due to severe autumn frosts is not as considerable as in spring (Bergh et al., 1998; Liu et al., 2018).

4.1.2. Vapor pressure deficit and soil water supply

Vegetation responds to both VPD and W limitations by controlling stomatal closure to reduce water loss, and avoid critically negative xylem water potentials and embolism. Across the models, in general, it is

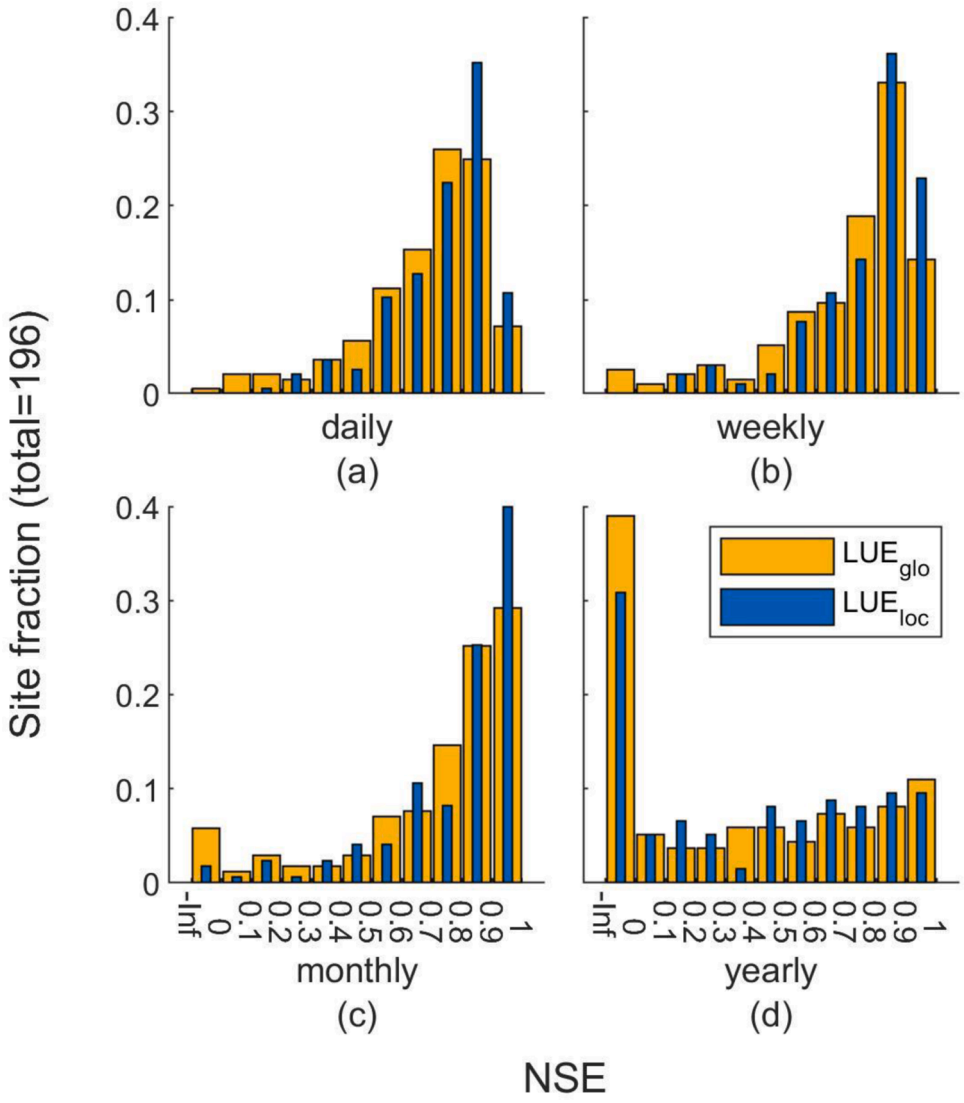


Fig. 7. Distribution of site NSE of the global best LUE model (LUE_{glo}) and site-best LUE models (LUE_{loc}) at daily (a), weekly (b), monthly (c) and annual scales (d).

commonly considered as a single factor encompassing the effects of water availability (McCallum et al., 2013; Running et al., 2004; Wang et al., 2017a; Wang et al., 2015; Yan et al., 2017; Yuan et al., 2007). However, the responses of LUE or GPP to VPD and W are not simultaneous (see correlation matrix in Fig. S11) and depend on the isohydricity or anisohydricity of the vegetation (Roman et al., 2015; Tardieu and Simonneau, 1998). The soil water supply also affects the sensitivity of GPP to variations in VPD; e.g., a well-watered plant can endure higher VPD before closing stomata, whereas a drought-stressed plant are shown to close stomata earlier (Tardieu and Simonneau, 1998). Our results demonstrate that both W and VPD effects should be considered, which agrees with some other (Bagnara et al., 2015; Mäkelä et al., 2006; Mäkelä et al., 2008; Stocker et al., 2020; Turner et al., 2006a; Wang et al., 2018) studies (Bagnara et al., 2015; Mäkelä et al., 2008; Stocker et al., 2020; Turner et al., 2006a; Wang et al., 2018).

Increasing VPD reduced LUE with changing sensitivity, thus the nonlinear functions (e.g., $fVPD_{Horn}$, $fVPD_{TAL}$, and $fVPD_{Wang}$) performed better than a linear one (e.g., $fVPD_{MOD17}$) for representing the response of LUE to VPD (Fig. 8). We observed that the VPD effect was lagged in arid climates, especially for grasslands (Fig. S17c), as the canopy has to redevelop green tissue after a dry period (Horn and Schulz, 2011a). Woody plants can typically extract water from deeper soils than grasses (Kulmatiski and Beard, 2013), resulting in different lags. VPD lags were nevertheless not as apparent as those for W (Fig. S17b), both of which could reflect the acclimation of plants to water stress (Flexas et al., 2009), but could also reflect a too fast response of the modeled WAI to precipitation and ET, or a model limitation of not representing deeper water access (Dorigo et al., 2017).

LUE responds to increasing W monotonically, but with a different sensitivity across sites. Thus, fW_{WAI} with a fixed slope cannot represent the response of GPP to W at the global scale. LSWI is a better indicator than WAI for the vegetation-relevant moisture state of boreal vegetation, possibly due to its advantage of capturing instant variations in leaf water content (Chandrasekar et al., 2010). The lower likelihood of non-monotonic or decreasing fW, $fW_{Weibull}$, reflects no constraints of high W on photosynthesis, which is opposite of expectations (e.g., Stocker et al., 2018). This might be due to lack of water-logged site-years on consecutive days in the current FLUXNET dataset.

4.1.3. *Light saturation and diffuse radiation*

Intense light is usually coupled with clear skies, high temperature, high atmospheric water demand (Fig. S11) and low soil water supply (Piao et al., 2020). Thus, any or a combination of these factors can obscure the effect of light saturation in a model, indicating possible effects of model over-parameterization in relation to the information given in the data (Ibrom et al., 2006). Moreover, the high correlation

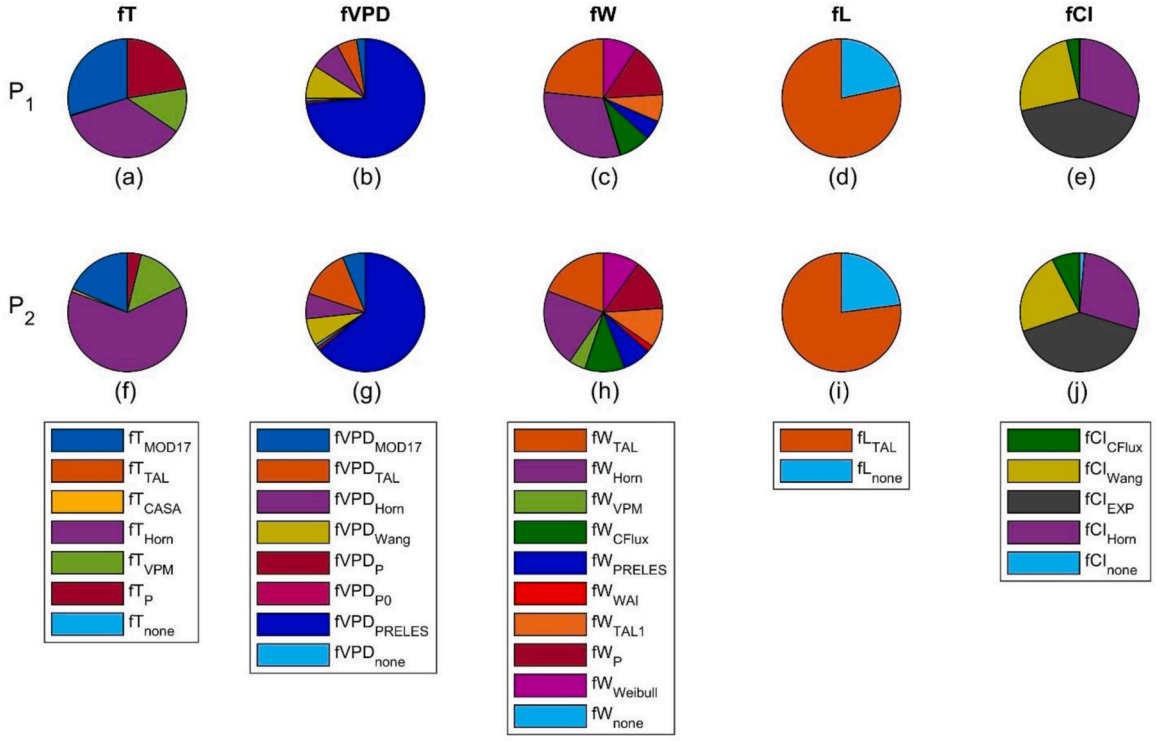


Fig. 8. Likelihood of partial sensitivity functions, which is equal to its ratio among the best 1% models derived from P_1 (a-e) and P_2 (f-j). Two fTs ($f\mathrm{T_{TAL}}$ and $f\mathrm{T_{Horn}}$) and one fW ($f\mathrm{W_{Horn}}$) include a lag effect in boreal and arid climate, and three fVPDs ($f\mathrm{VPD_{PRELES}}$, $f\mathrm{VPD_P}$, $f\mathrm{VPD_{P0}}$) include CO_2 effect.

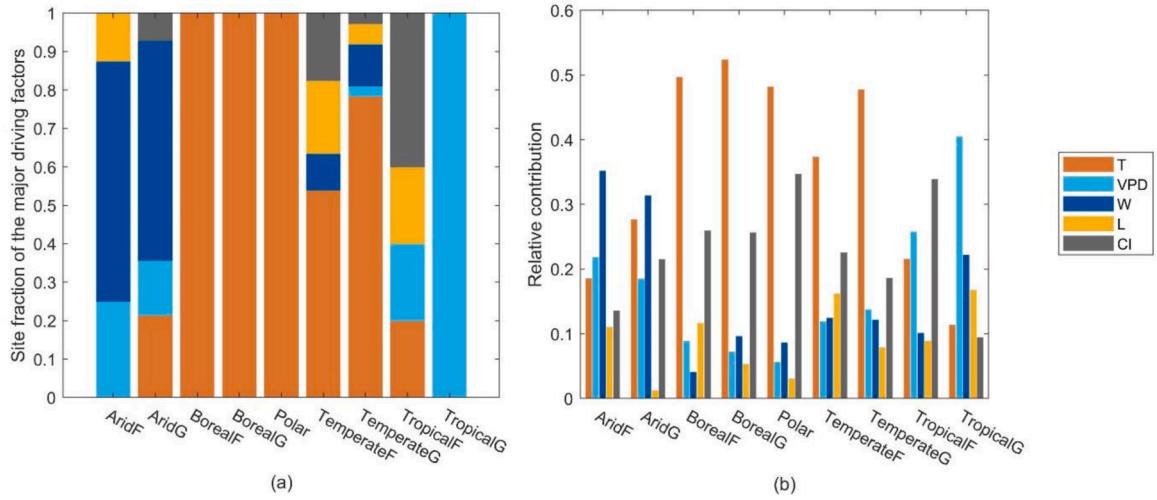


Fig. 9. Environment sensitivities in the global best model across different climate-vegetation types: (a) site fraction of the major driving factor and (b) relative contribution of each environmental factor. T (orange) is temperature, VPD (sky blue) is vapor pressure deficit, W (dark blue) is soil water supply, L(yellow) is light saturation and CI (gray) is cloudiness index in both figures. 'Major driving factor' represents the environmental factor that has the strongest effect on LUE at site-level. 'Relative contribution' is a measure of the contribution of an environmental factor to the total limitation on LUE at site-level and is averaged per climate-vegetation type.

between parameters of fL and fCI (R=-0.68, on average) indicates the temporal and spatial simultaneity between these two effects. The reason is that when CI is large, the same amount of light can be distributed across more leaves and light saturation of total canopy can be relaxed (Gu et al., 2002; Ibrom et al., 2006; Knohl and Baldocchi, 2008; Roderick et al., 2001). Due to the interactions between these factors, light saturation seems not as significant as other driving factors in the LUE models as it may be. Part of its influence may be mapped to parameters of other response functions and vice versa, especially on models with many functions and parameters. This phenomenon may reduce the predictability of the models.

Intense light however had the strongest effect on GPP at high NDVI (Fig. S18), such as some sites in tropical forests (Propastin et al., 2012), temperate forests and temperate grasslands. Within the scope of having more climate-vegetation types in a global model, it is necessary to consider both effects of light saturation (Mäkelä et al., 2008; Medlyn, 1998) and CI (Wang et al., 2018). To stand for the response of LUE to saturated light, an inverse function of APAR is appropriate since another study has derived a similar form (Ibrom et al., 2008).

Increasing CI can improve LUE since it represents a larger ratio of

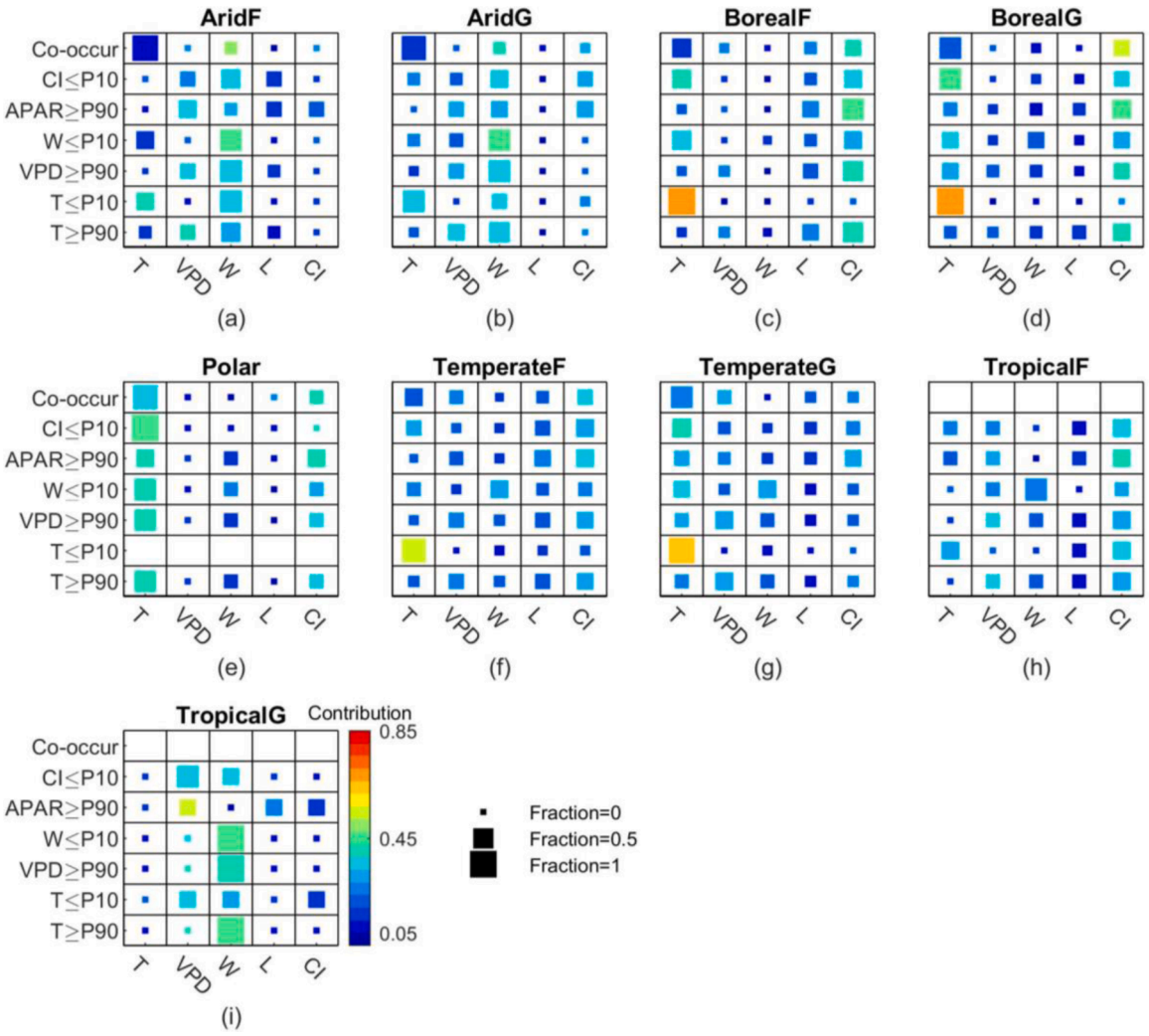


Fig. 10. Environment sensitivities of LUE in the global best model under extreme conditions across different climate-vegetation types: site fraction of each 'major driving factor' (indicated by the size of the squares) and 'relative contribution' of each environmental factor (indicated by the color of the squares). X-axis represents the environmental factors: temperature (T), vapor pressure deficit (VPD), soil water supply (W), light saturation (L), cloudiness (CI). 'Co-occur' represents the environmental condition at site-level when T≥90^{th} percentile, VPD≥90^{th} percentile, W≤10^{th} percentile, APAR≥90^{th} percentile, and CI≤10^{th} percentile. There are no 'Co-occur' points at tropical sites (h and i). CI≤P10, APAR≥P90, W≤P10, VPD≥P90, T≤P10 and T≥P90 represent the environmental conditions that CI≤10^{th} percentile (clear skies), APAR≥90^{th} percentile (intense light), W≤10^{th} percentile (low soil water availability), VPD≥90^{th} percentile (high atmospheric water demand), T≤10^{th} percentile (low temperature), and T≥90^{th} percentile (high temperature), respectively. The larger a square is, the more sites at which an environmental factor has the strongest effect on LUE. The more bright (yellow and red color) a square is, the larger the relative contribution of an environmental factor to the total limitation on LUE is. For example, W is the major driving factor of LUE and has the largest relative contribution when CI≤P10 in arid forests (a).

diffuse radiation when total radiation does not change. However, the importance of CI cannot be identified in GPP due to the negative correlation between CI and PAR, which might offset the positive effect of CI. However, a more flexible function, e.g., $f\mathrm{CI}_{\mathrm{Horn}}$ which can be monotonically increasing, decreasing or bell-shaped function, cannot better represent the response of GPP to CI at the global scale. This indicates that a monotonic and positive relationship between LUE and CI across climate-vegetation types is the most appropriate, in contrast to Alton's study (2008) but in agreement with other studies (Cheng et al., 2015; Gu et al., 2002).

4.1.4. CO_2

As the fuel of photosynthesis, elevating CO_2 concentration fertilizes GPP at a slow speed (Smith et al., 2016; Wenzel et al., 2016). CO_2 fertilization can be heterogeneous both temporally and spatially (Liu et al., 2016). Although the effect was not considered in most LUE models, our results reflect the advantages of including a CO_2 effect in an LUE model. It would translate into a further larger spectrum of LUE applications considering the large increase in atmospheric CO_2 in the last few decades.

Since the degree of the CO_2 fertilization is related to VPD, which affects stomata closure and gas exchange, whether an LUE model can capture the CO_2 effect depends on the function form of fVPD. Model performance can be restricted if the function is inappropriate (e.g., $f\mathrm{VPD}_{\mathrm{P}}$ is not as good as $f\mathrm{VPD}_{\mathrm{PRELES}}$), and the modeled responses to elevated CO_2 could be divergent (Haverd et al., 2018; Smith et al., 2016).

Furthermore, using a spatially varying CO_2 forcing in an LUE model (Sun et al., 2018; Worden et al., 2021) can be worthy for exploring the effect of CO_2 fertilization on the spatial variability of GPP.

4.2. Model robustness and uncertainties

In the following section, we will discuss model robustness tested using various model selection methods, groups of randomly selected sites, and model assessing metrics (Section 4.2.1), possible uncertainties

of the selected model resulting from model parameter estimation, model structure (Section 4.2.2), observational data (Section 4.2.3), and other potential drivers (Section 4.2.4).

4.2.1. Model selection methods, sites, and model assessing metrics

Model assessments are usually sensitive to the model selection methods, the given data information and the assessment metrics (Burnham and Anderson, 2002; Burnham et al., 2011; Efron, 2014). For example, VPM model has different performances in Zhang et al. (2017b) (global R^2=0.74 and RMSE= 2.08 $gC{\cdot}m^{-2}{\cdot}d^{-1}$, 113 sites) and in Yuan et al. (2014) (spatial R^2=0.44, 157 sites). In our study, the best model varied with selection methods and the bootstrapped sites (represented by site-years) but not with the assessment metrics or thresholds to identify the best models. The ranks of the best 20 models are more sensitive to sites and site-years (variance in ranking =134) than model selection methods (variance in ranking = 96), highlighting the importance of the bootstrap approach in selecting a robust model across sites, thereby, on a global scale. The best model selected here using two assessment metrics, six thresholds and three independent model selection methods, was among the best 1% models of 77% of different site-years groups (200 random groups in total) and among the best 10% models of 100% of different site-years groups. Furthermore, it was not statistically different from any other model ranked at the first place in any site-years group using KS test (significance level = 0.05) at daily and weekly scales. Thus, the selected model and partial sensitivity functions are robust for the current FLUXNET dataset.

4.2.2. Model parameters and structures

Another source of model uncertainty is its parameters. To avoid the selection of local optima in parameter calibration, we used a stochastic and derivative-free evolutionary global-search algorithm (Hansen and Kern, 2004). Multiple optimized parameter vectors were considered both in site-sampling-based and ABC-based model selection so that a robust model that is insensitive to different parameter vectors could be selected (Toni and Stumpf, 2010). However, it is possible to find uncertain parameters due to the functional form, correlations among driving factors (Fig. S11) and among parameters (see Section 4.1.3). An experiment of parameter prediction using site-specific biotic and abiotic properties (Section S1) showed large biases in parameter prediction. The parameter extrapolation in an area without EC measurement could be therefore problematic. Nevertheless, it was reported that the parameters of Horn's model could be extrapolated using the site-specific properties (Horn and Schulz, 2011b). Site-averaged fT and fW, the least correlated environmental factors and sensitivity functions (Fig. S11) among all others, could also be predicted (Fig. S19), demonstrating that the pattern of GPP response to these environmental factors is controlled by the biophysical and environmental characteristics and emphasizing the importance of reducing parameter correlation.

The global best model selected in this study has the lowest uncertainty in the application at the global scale if neglecting the uncertainty from parameters. At the local scale or site-specific scale, the biome-best model or site-best model might have a lower uncertainty than the global best model. Besides, the model uncertainty can come from the model structure that combines partial environmental sensitivity functions. For example, according to Liebig's law, partial sensitivity functions can be combined in a minimal form (Minunno et al., 2015; Tian et al., 2020; Yuan et al., 2007). They can also be combined in a summed (Horn and Schulz, 2011a) or more complex mathematical form (McCallum et al., 2013; Oliphant et al., 2011; Wang et al., 2017a). Furthermore, whether a two-leaf model structure (He et al., 2013; Yan et al., 2017), splitting absorbed light into sunlit and shaded portions, can represent the effect of diffuse radiation on GPP better than big-leaf model structure is still unknown.

4.2.3. Eddy covariance and remote sensing data uncertainty

In this study, model selection and analysis of environment sensitivities rely on the assessment using EC data. Although the quality control variables can filter for good data quality, systematic and random errors due to filtering methods remain in EC measurements (Aubinet et al., 2012), thereby influencing the training and evaluation of models (Jung et al., 2020). Furthermore, the partitioning method of fluxes causes data uncertainties. On the one hand, the estimation algorithm of GPP from NEE still needs to improve. For example, nocturnal estimates of ecosystem respiration neglect the inhibition of respiration by plants during daytime (Keenan et al., 2019), thus creating error and uncertainty in GPP (Reichstein et al., 2005). On the other hand, standard partitioning methods can produce physically unrealistic estimates of GPP under the influence of photodegradation in an arid region (Cleverly et al., 2016). We note that the LUE models, in general, can produce reasonable estimates of GPP in the absence of unbiased EC data. However, even at some sites where these biases are small at a daily scale, the errors can propagate to the annual scale, albeit only at a few sites. Thus, the evaluation of modeled interannual variability, trends, and magnitudes of annual GPP for the sites with these issues can be problematic. Moreover, the sparse distribution of EC towers globally creates gaps for some climate and vegetation types (Alton, 2020), thus affecting the applicability of our models in regions with extreme environmental conditions.

The remote-sensing-derived vegetation index, NDVI, related to leaf chlorophyll content (Rouse et al., 1974; Zhang et al., 2018), was used in this study to represent FAPAR which has a linear relationship with NDVI (Myneni et al., 1997). Consequently, the calibrated $\boldsymbol{\varepsilon_{max}}$ values can be smaller than using FAPAR since FAPAR was usually overestimated (Kanniah et al., 2009) but should have a similar spatial pattern with calibrated $\boldsymbol{\varepsilon_{max}}$ using FAPAR (Horn and Schulz, 2011a). However, the mismatch between the footprints of remote sensing data and EC measurements might result in a lack of representativeness for pixel data (i.e., remote sensing image pixel) relative to point data (i.e., EC tower), especially at an EC site with extremely fragmented landscape (Cescatti et al., 2012). Consequently, landscape heterogeneity might affect the reliability of the modeled environment-sensitivities in a few sites (Migliavacca et al., 2015).

4.2.4. Other drivers of gross primary productivity

It is known that stand age and slowly changing environmental factors (e.g., N deposition, CO_2 fertilization, climate change and management) affect GPP responses to the local environment (He et al., 2012; Ma et al., 2019; Medlyn, 1998). Given a long time series of data, this effect can be demonstrated by calibrating the model parameters for every year and analyzing the time series of their values (Wu et al., 2013; Wu et al., 2012). In the FLUXNET dataset, there are several sites with long time series that exhibit long-term trends (Fernández-Martínez et al., 2017; Pilegaard and Ibrom, 2020). Some of this variability might be captured with partial sensitivity functions that include, e.g., the CO_2 effect (Section 4.1.4). With the increasing availability of long-term flux datasets (Baldocchi, 2020), similar studies might reveal ways to include these long-term dynamics in LUE models.

4.3. General evaluation of the approach and prospects for future application

We showed that a large variety of LUE models are currently being used. Many of those were developed from predecessors (e.g., He et al., 2013; Wang et al., 2015). To our knowledge, this is the first study to evaluate the potential of all possible LUE modeling approaches, considering every combination of environmental factors and their partial sensitivity functions, which resulted in testing 5600 different LUE models on their global and site-specific performances, and in considering various site groups in the model evaluation.

The results confirmed that no single model can be used with the highest performance to simulate all sites and years, however, a global best model, model #1, performed surprisingly well in its robustness. The

selected model constituted a not as yet published model structure, which confirmed both our initial hypothesis and the usefulness of a systematic model development and benchmarking approach as presented in this study. Although LUE models do not present a full physiological or ecological description of GPP responses, model #1 reflects a representation required to explain the variability in the current FLUXNET dataset. At the local scale or site-specific scale, a site-best model with specific features (i.e., the necessary environmental factors and the appropriate sensitivity functions) might be more effective but underperform at other sites where such features may be redundant while others may be missing. At the global scale, model #1 captures the features that are essential across various ecosystems, suggesting that model #1 could be a reference for future model selection towards a robust global GPP estimation.

A global, or regional, application of model #1 requires necessarily a larger scale availability of forcing data and understanding of changes in model parameters, as in any other semi-empirical or mechanistic modeling approach. The global best LUE model has the potential to be applied both globally and regionally since the forcing data are available at various spatial scales (e.g., CRUNCEP; Viovy, 2018a). However, model parameters are unknown for locations without observational data, which requires assumptions on factors controlling its variability, such as plant functional type (Running et al., 2004), or in addition vegetation properties and climatic conditions (Carvalhais et al., 2010b), or assuming a single parameter set optimized across all sites (Yuan et al., 2007). Parameter extrapolation is an underlying challenge for LUE models, and for all other kinds of parametric approaches, from empirical to mechanistic models (Luo and Schuur, 2020). As such, to apply the best model, or any other model developed locally, at global scales, demands parameter extrapolation approaches to be further studied.

5. Conclusion

To identify the environmental drivers of GPP and the appropriate partial sensitivity functions, our study focused on a comprehensive and exhaustive assessment of an ensemble (n=5600) of previously reported and new LUE models based on a bootstrap approach using 196 eddy covariance CO_2 flux sites. A robust model with good performance at the global scale and across various climate-vegetation types, which has not been published before, was selected. Our results demonstrated that the responses of GPP to temperature, VPD, soil water supply, light saturation, cloudiness and CO_2 are all significant and mostly nonlinear. The effects of temperature and soil water supply on GPP are lagged in cold and arid regions, respectively, due to the recovery of photosynthetic capacity. According to the selected model, temperature and soil moisture are the dominant factors in cold and arid environments, respectively. The effects of cloudiness index and light saturation are observed across various ecosystems.

Our study confirms that the light saturation, diffuse fraction, and CO_2 fertilization, which are usually ignored in global empirical models, should be considered for the estimation of GPP at the global scale. In addition, non-arid regions reveal a nonlinear and bell-shaped response of GPP to temperature while cold and arid ecosystems show lagged effects of temperature and soil moisture on GPP. The future use of the best model at global scales, necessarily entails further studies on the extrapolation of model parameters. For a justified evaluation and calibration of both mechanistic and data-driven models, the recommendation from this study is to consider the representativeness of the observational data across various ecosystems and the equifinality between model structures and parameters.

Declaration of Competing Interest

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

Acknowledgments

This work is supported by Sino-German (CSC-DAAD) Postdoc Scholarship Program and the European Space Agency (ESA) via the Earth System Data Lab (ESDL) project. We sincerely appreciate Martin Jung for his constructive comments from the beginning of the study, Jacob Nelson for his help in collecting the remote sensing data, ancillary data and contacting PIs, and Tina Trautmann for sharing her code of CMAES. We also gratefully acknowledge MEDSPEC (PTDC/AAG-MAA/3699/2014) and SINDBAD for sharing data for this analysis. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster and the AmeriFlux Management Project (with support by European Union H2020 projects and U.S. Department of Energy Office of Science, respectively), with contributions from the Carbon Dioxide Information Analysis Center, ICOS Ecosystem Thematic Centre, and OzFlux, ChinaFlux, and AsiaFlux offices. We express our gratitude to all the FLUXNET investigators generously providing the dataset.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2021.108708.

Appendix A. Water availability indicator

We calculated water availability indicator (WAI, in mm) to represent soil water supply conditions (Tramontana et al., 2016). It assumes the soil water storage as a single bucket, and the precipitation (or melted snow) and ET are the only influx and outflux. The ET, called ET_{sim}, is estimated from PET, or a water decay velocity multiplies with the water deficit (i.e., bucket size minus WAI at last time step) if it is smaller than PET. The bucket size and the water decay velocity are controlled by available water content (AWC, in mm) and a decay coefficient for ET (θ), and vary across ecosystems (Boese et al., 2019). We assume these two parameters were different at each site, relative to the soil properties and climate condition. Thus, these two parameters were calibrated by constraining the errors in ET_{sim} (see Appendix D). Furthermore, we compared different approaches to calibrate WAI parameters, which denote fixing them, calibrating them independently or jointly with LUE model parameters, to find the best approach that has the most good-performance LUE models (Appendix C). According to the comparison result, we calibrated AWC and θ jointly with LUE model parameters, ε_{max} and parameters of fX.

Appendix B. Data screening strategy

We screened the data quality used in the cost functions and model assessment. We defined the criteria that the climate data (PAR, T, VPD, W and CI) was bad-quality if the corresponding quality flag (QA), acquired from FLUXNET dataset, was smaller than 0.80. If the forcing data was estimated from observational data, we used the QA of the observational data instead (e.g., use QA of R_g for PAR, minimum QA of Precip, T and Rn for W). We regarded the remote sensing data as bad-quality if the average quality flags (QC) of the 4 pixels around the EC sites, acquired from MCD43A2, were smaller than 0.5. In the cost functions (see Appendix D), we excluded all the bad quality data, the filled data (see gap-filling method in Table 2) and the data violates the corresponding variable definition (i.e., GPP<0) or makes the cost function infinite (i.e., σ_{NEE}<0.02 or σ_{LE}<0.02). Besides, we removed the site-years without LE observations and sites with less than 160 effective points that meet the above quality screening criteria. We defined the data used in the cost functions as good-quality data.

We also set different criteria in model assessment at different time scales. At a daily scale, we calculated the statistics only using the good-quality data. At the weekly and monthly scale, we calculated the weekly or monthly QA based on the average minimum of the daily QA and QC of all the forcing data. If GPP or ET at a daily scale is a filled value, or GPP<0, QA at weekly and monthly was set to zero. We only used the data that weekly or monthly QA was larger than 0.70 in more than four weeks or months. At the annual scale, we only considered the data that the average QA of GPP was larger than 0.5 and the effective observation period was longer than 4 years (136 sites in total).

Appendix C. Comparison of WAI parameters calibration

The WAI parameters, controlling the simulated water availability dynamics in our study, affect the model result. Here we set six experiments (see descriptions in Table S3) to assess the best approach for calibrating WAI parameters: optimize WAI parameters jointly with LUE model parameters ('Full'); fix WAI parameters but optimize LUE model parameters ('Fix'); optimize WAI parameters and LUE model parameters individually ('IDP') under all conditions; the same three optimizing approaches under the supply-limited condition ('Sup', 'SupFix' and 'SupIDP') which is defined by ET<PET. In these experiments, we applied the trust-region-reflective algorithm (Moré and Sorensen, 1983), a nonlinear least squares solver integrated into the optimization toolbox of MATLAB R2019a, to search the optimal parameters. Then we assessed the model fitness using the site NSE (NSE_i), global NSE (NSE_g) and spatial NSE (NSE_{sp}) at daily, weekly, monthly and annual scales defined in Section 2.4. As our study aimed to find the best model, we selected the calibration approach according to the number of good-performance (global NSE>=0.75) models.

The result showed that the 'Sup' had the most good-performance models (256 models) at a daily scale, and the 'Full' had the most good-performance models (849, 1210, and 1922 models) at weekly, monthly and annual scales (Fig. S2). The largest global NSE of the 'Full' (0.78, 0.81, 0.85 and 0.87) and the 'Sup' (0.78, 0.82, 0.85, and 0.87) were all greater than the 'Fix', 'SupFix', 'IDP' and 'SupIDP' at daily, weekly, monthly and annual scales. However, the site NSE of the best-performance model in 'Full' was not different from the best model in 'Sup' significantly according to the Kolmogorov–Smirnov (KS) hypothesis test with a significance level of 0.05. Since the 'Full' had the most good-performance models at more time scales than the others, we adopt it as the best approach for calibrating WAI parameters. In other words, we optimized the WAI parameters jointly with LUE model parameters to calibrate the LUE models.

Appendix D. Cost functions

We define the cost function (*cf*) as the sum of the GPP errors (cf_1, equation D.2), the ET errors (cf_2, equation D.3), and the *f*X constraints (cf_3 and cf_4).

$$cf = cf_1 + cf_2 + (cf_3 + cf_4) \tag{D.1}$$

$$cf_1 = \sum_{t=1}^{N_t} \sqrt{\left(GPP_t - \widehat{GPP_t}\right)^2 \cdot {\sigma_{NEE_t}}^{-2}} \tag{D.2}$$

$$cf_2 = \sum_{t=1}^{N_t} \sqrt{\left(ET_t - \widehat{ET_t}\right)^2 \cdot {\sigma_{LE_t}}^{-2}} \tag{D.3}$$

Here, we compared GPP_{obs} (*GPP*) and ET_{obs} (*ET*), against simulated GPP using the calibrated LUE models ($\widehat{GPP}$) and ET_{sim} (defined in S1; $\widehat{ET}$), respectively. *t* represents the time step of GPP and ET, ranging from 1 to the total days (N_t). Due to the uncertainties in observation and estimation and the different units of GPP and ET, we weighted the squared differences between *GPP* and $\widehat{GPP}$, *ET* and $\widehat{ET}$ using the estimated uncertainty of *GPP* (σ_{NEE}) and *ET* (σ_{LE}), respectively. The parameter vector that minimizes the sum of cf_1 and cf_2 is the best for the LUE model and WAI, respectively.

We follow the concept of ecological and dynamic constraints (EDC, by Bloom and Williams, 2015) to regularize the inversion approach via two additional constraints: cf_3 (equation D.4) and cf_4 (equation D.5).

$$cf_3 = ((1 - \max(f\mathrm{T}_r)) + (1 - \max(f\mathrm{VPD}_r)) + (1 - \max(f\mathrm{W}_r)) + (1 - \max(f\mathrm{L}_r)))\cdot c \tag{D.4}$$

$$cf_4 = \left(\sum_r \left(f\mathrm{T}_r(\mathrm{T} < 0^{\circ}\mathrm{C}) > \theta_{f\mathrm{T}}\right) + \sum_r \left(f\mathrm{VPD}_r(\mathrm{VPD} > 2\mathrm{kPa}) > \theta_{f\mathrm{VPD}}\right) + \sum_r \left(f\mathrm{W}_r(\mathrm{W} < 0.01) > \theta_{f\mathrm{W}}\right)\right)\cdot c \tag{D.5}$$

These impose constraints on the simulated *f*X based on two assumptions: the instantaneous $\boldsymbol{\varepsilon}$ of vegetation can reach its potential, $\boldsymbol{\varepsilon_{max}}$, under some specific environmental condition (cf_3) and is inhibited under a non-ideal growing condition (cf_4). These two cost functions were calculated independently from cf_1 and cf_2, using an artificial input (PAR=0-20 $MJ{\cdot}m^{-2}{\cdot}day^{-1}$, FAPAR=0-1, T=-10-40°C, VPD=0-2 kPa, W=0-1 and CI=0-1). *cf*3 is to set the maximum of *f*T, *f*VPD, *f*W, and *f*L ($\leq$1, see Fig. 2 and Section 2.1) to one, which implies that the corresponding environmental factor does not limit $\boldsymbol{\varepsilon}$ at a certain point within the given ranges of PAR$\in$ [0,20] (in $MJ{\cdot}m^{-2}{\cdot}day^{-1}$), FAPAR$\in$[0,1], T$\in$[-10,40] (in°C), VPD$\in$[0,2] (in kPa), W$\in$[0,1] and CI$\in$[0,1], represented by *r* in equations D.4-D.5 (e.g., $\max(f\mathrm{T}_r)$ represents the maximum *f*T when temperature is ranging between -10 and 40°C).

The constraint of *cf*3 was not applied on $f\mathrm{CI}_{\mathrm{CFlux}}$, $f\mathrm{VPD}_{\mathrm{P}}$, and $f\mathrm{VPD}_{\mathrm{P0}}$ and applied in a different way on $f\mathrm{VPD}_{\mathrm{PRELES}}$, since they can be larger than one

in theory. The fVPDs include the effect of CO_2 variations, which can have fertilization on GPP (Haverd et al., 2018) and increase $\boldsymbol{\varepsilon_{max}}$. The increasing CI can also enhance $\boldsymbol{\varepsilon_{max}}$ (Wang et al., 2015), but the range of fCI can be different depending on the assumption about $\boldsymbol{\varepsilon_{max}}$. Here, the $f\mathrm{CI}_{\mathrm{CFlux}}$ is larger than one, assuming $\boldsymbol{\varepsilon_{max}}$ is the maximum LUE under all sky conditions and the other fCIs are smaller than one, assuming $\boldsymbol{\varepsilon_{max}}$ is the maximum LUE under totally diffuse radiation condition (see equations in Table 1). Furthermore, we separated $f\mathrm{VPD}_{\mathrm{PRELES}}$ into fVPD and $f\mathrm{CO}_2$ parts (*1 and *2 in Table 1) and only restricted the fVPD part using cf_3.

Another constraint, cf_4, is to guarantee the fT, fVPD, and fW smaller than a threshold ($\theta_{f\mathrm{T}}$, $\theta_{f\mathrm{VPD}}$, and $\theta_{f\mathrm{W}}$) under a non-ideal condition (T<0°C, VPD>2 kPa, or W<0.01). Here the thresholds ($\theta_{f\mathrm{T}}$=0.2, $\theta_{f\mathrm{VPD}}$=0.9, and $\theta_{f\mathrm{W}}$=0.2) were estimated according to the normalized ratio of GPP to APAR at all sites. The other non-ideal conditions were not included since they vary across sites. The c in equations D.4-D.5 denotes a penalty term (=10^4, an empirical value) to coordinate the scales of cf_1, cf_2, cf_3, and cf_4.

Appendix E. Parameter uncertainty estimation

We calculated the standard errors and correlation coefficient of the optimized parameters based on the Jacobian matrix of cost function against parameters (Omlin and Reichert, 1999) of each model at each site. We normalized the standard errors by the optimized parameters to unify the units of different parameters. If the optimized parameter approximates zero, the normalized standard error is equal to the original standard error. We also calculated the spatial correlation coefficient between the parameters across sites. We utilized the normalized standard error to measure the parameter uncertainties used in Section S1.

Appendix F. Model selection using the pair-wise hypothesis test

The two-sample Kolmogorov–Smirnov (KS) hypothesis test (Lilliefors, 1967) is a non-parametric test commonly used to compare two samples. It compares the empirical cumulative distribution functions of two samples and tests if they differ (two-tailed) or one is larger than the other (one-tailed) at a significance level. Therefore, it is sensitive to the differences in both shape and location of the cumulative distribution functions. Here we adopt the test to compare every pair of the LUE models based on their site NSE at different time scales. We sorted the models according to the workflow as below.

Requirement: *ModelSet* = all the LUE models
ParSet = one optimized parameter vector for each LUE model at each site
SiteSet = all the EC sites
Initialization: set total model number N_m=5600, total site number N_{s0}=196
For time scale *ts*=daily, weekly, monthly, annual scale
For $m = 1$ **to** N_m
For $m_2 = 1$ **to** N_m
If $m \neq m_2$
For $i = 1$ **to** N_{s0}
Simulate GPP using model m and m_2 ($\widehat{GPP}_{m,i,ts}$ and $\widehat{GPP}_{m_2,i,ts}$)
Store GPP in a global dataset across sites ($\widehat{GPP}_{m,ts}$).
Estimate model fitness at site-level, NSE, of m and m_2 ($mef_{1,i}$ and $mef_{2,i}$)
End For
Compare NSE vector of m and m_2 ($\boldsymbol{mef_1}$ and $\boldsymbol{mef_2}$): test two hypothesis H_1 ($\boldsymbol{mef_1} < \boldsymbol{mef_2}$) and H_2 ($\boldsymbol{mef_1} > \boldsymbol{mef_2}$)
Store larger times of m_1:

$$L_{m,m_2,ts} = \begin{cases} 1, & \text{if } H_1 \text{ is rejected and } H_2 \text{ is accepted} \\ 0, & \text{if } H_1 \text{ is rejected and } H_2 \text{ is rejected} \\ -1, & \text{if } H_1 \text{ is accepted and } H_2 \text{ is rejected} \end{cases}$$

Calculate global NSE and spatial NSE of m ($\mathrm{NSE}_{g,m}$ and $\mathrm{NSE}_{sp,m}$, as defined in Section 2.4)
End If
End For
Calculate the frequency of m larger than the other models:
$L_{m,ts} = \sum_{m=1,m\neq m_2}^{N_m} L_{m,m_2,ts}/N_m$
End For
End For
Calculate the average frequency of m larger than the other models at different time scales
$L_m = \sum_{ts} L_{m,ts}/4$
Get the rank R_m of m according to L_m, $\mathrm{NSE}_{g,m}$ and $\mathrm{NSE}_{sp,m}$

All the LUE models were sorted according to R_m (Fig. S5). The percentages of $L_{m,m_2,ts}$ is positive (red color) and is negative (blue color) represent the goodness and weakness of m, respectively. On the other hand, the blank area stands for the percentage of models equal to m. The KS test showed that the best model has the structure of $f\mathrm{T}_{\mathrm{Horn}}$, $f\mathrm{VPD}_{\mathrm{PRELES}}$, $f\mathrm{W}_{\mathrm{CFlux}}$, $f\mathrm{L}_{\mathrm{TAL}}$ and $f\mathrm{CI}_{\mathrm{Wang}}$ (site NSE medians were 0.72, 0.78, 0.81, 0.25 at daily, weekly, monthly and annual scales). The global best model, model #1, was the 13$^{\mathrm{th}}$ best model according to R_m, which is not different from the best model significantly at daily, weekly, monthly scales but slightly worse at the annual scale (site NSE medians were 0.73, 0.79, 0.82, 0.22 at daily, weekly, monthly and annual scales).

Appendix G. Model selection using approximate Bayesian computation scheme

The approximate Bayesian computation (ABC) scheme (Beaumont et al., 2002) refers to an evolutionary algorithm successfully used for parameter estimation and model selection (Toni and Stumpf, 2010; Toni et al., 2009). ABC randomly sample the models and parameter vectors from the whole model and parameter spaces and screen them by rejecting the poor models or models with improper parameter vectors according to their posterior

distribution of the model fitness. The model that survives has the best performance and the most applicable parameter vectors.

To select a robust model that is insensitive to different parameter vectors, we collected the optimized parameter vectors of each model at each site in 10% of the first half (using a uniform sampling with a step of 10) and 100% of the last half iterations of the parameter searching process ($5.5*10^4$ parameter vectors in total). Notwithstanding that a generic algorithm can be used to cut down the time, searching the best model and parameter vector in a whole parameter space of 5600 models at 196 sites is not practical. For this reason, we randomly select 5000 parameter vectors for each model at each site using the normalized reciprocal of *cf* (defined in Appendix D) as a weight. A model with a set of parameter vectors for all sites was regarded as a particle. All the particles (5600*5000) were compared simultaneously using a model distance (d) compared to a genetic threshold (θ). The models with the parameter vector that meet the threshold in the last iteration were selected as the best models. The workflow of ABC method is defined below.

Requirement: *ModelSet* = all the LUE models;
ParSet = 5000 optimized parameter vectors for each LUE model at each site;
SiteSet = all the EC sites;
Initialization: set total model number N_m=5600, total parameter vector number N_p=5000, total site bootstrap times N_{itr} =10, bootstrap site number N_s=150, total threshold update times N_θ=20, set site bootstrap time l=1
While $1 \le l \le N_{itr}$
For $n = 1$ **to** N_θ
For $m = 1$ **to** N_m
For $k = 1$ **to** N_p
For $j = 1$ **to** N_s
Simulate GPP ($\widehat{GPP}_{j,m,k}$) using model m and parameter vector $Par_{m,k}$
For time scale *ts*=daily, weekly, monthly, annual scale
Aggregate simulated GPP and observed GPP to time scale *ts* ($\widehat{GPP}_{j,m,k,ts}$ and $GPP_{j,m,k,ts}$)
Estimate model fitness at site j, NSE, of m ($mef_{j,m,k,ts}$)
End For
End For
Divide NSE vector $\boldsymbol{mef_{m,k,ts}}$into twelve bins (-∞ - -1.0, -1.0-0.0, 0.0-0.1, 0.1-0.2, 0.2-0.3, 0.3-0.4, 0.4-0.5, 0.5-0.6, 0.6-0.7, 0.7-0.8, 0.8-0.9, 0.9-1.0) and statistic the counts of site NSE in each bins (N_b)
Calculate the cumulative sum of normalized N_b (Cum_b) and distance of model m ($d_{m,k,ts}$), which is used to reject a particle (a model with a parameter vector).

$$Cum_b = \sum_{b=1}^{12}(N_b/N_s + Cum_{b-1})$$

$$Cum_1 = N_1/N_s$$

$$d_{m,k,ts} = \sum_{b=1}^{12} Cum_b$$

End For
End For
Update threshold $\theta_{l,ts,n}$, where $\theta_{l,ts,1}$ is the 10th percentile of distance vector, $\boldsymbol{d_{m,k,ts}}$ and $\theta_{l,ts,20}$ is the minimum in $\boldsymbol{d_{m,k,ts}}$

$$\theta_{l,ts,n} = n\cdot(\theta_{l,ts,1} - \theta_{l,ts,20})/N_\theta$$

Compare $\boldsymbol{d_{m,k,ts}}$ with $\theta_{l,ts}$ and calculate model weight

$$w_{m,k,ts,l,n} = \begin{cases} 1, & if\ d_{i,k,ts} \le \theta_{l,ts,n} \\ 0, & if\ d_{i,k,ts} > \theta_{l,ts,n} \end{cases}$$

$$w_{m,l,n} = \sum_{ts}(\sum_{k=1}^{Np} w_{m,k,ts,l,n}/N_p)/4$$

End For
Calculate the average model likelihood

$$P_{m,l} = \sum_{n=16}^{20} w_{m,l,n}/5$$

l=l+1, go to 3)
Calculate the average likelihood of all the iterations, P_m, and derive the model rank R_m

---

The top model sorted according to R_m (Fig. S6) was selected as the best model according to ABC, which has $f\mathrm{T}_{\mathrm{Horn}}$, $f\mathrm{VPD}_{\mathrm{PRELES}}$, $f\mathrm{W}_{\mathrm{Horn}}$, $f\mathrm{L}_{\mathrm{TAL}}$ and $f\mathrm{CI}_{\mathrm{Wang}}$ in the model structure (site NSE were 0.73, 0.78, 0.83 and 0.22 at daily, weekly, monthly and annual scales). Model #1 was the second-best model, which was not different from the best model according to ABC at daily, weekly, monthly and annual scales (KS test, significant level=0.05).

Appendix H. Ancillary data

We gathered the Plant Functional Type (PFT) classified according to IGBP global vegetation classification scheme from the site information published on the FLUXNET network websites and literature (Table S2). Then we classified the vegetation to the forest (=EBF, ENF, DBF, DNF, MF, WSA, OSH and CSH) or grassland (=GRA, CRO, WET and SAV) based on the PFT per site. We also extracted Koeppen-Geiger climate types (Clim) from the re-analyzed Koeppen-Geiger map (Rubel et al., 2017), which had the spatial resolution of five arc minutes, according to the nearest neighboring pixel. We determined the main climate type according to the first character of KG classification: tropical (=A), arid (=B), temperate (=C), boreal (=D), and polar (=E). Finally, we defined nine climate-vegetation types as tropical forest (TropicalF, five sites), tropical grassland (TropicalG, three sites), arid forest (AridF, eight sites), arid grassland (AridG, 14 sites), temperate forest (TemperateF, 63 sites), temperate grassland (TemperateG, 37 sites), boreal forest (BorealF, 41 sites), boreal grassland (BorealG, 20 sites) and polar vegetation (Polar, five sites).

Moreover, we collected the site-specific biotic and abiotic traits from different sources. We calculated the nineteen bioclimatic variables (BIO1-BIO19 in Table S5) using CRUNCEP products based on the ANUCLIM algorithms (Xu and Hutchinson, 2013). Besides, we computed eleven bioclimatic vegetation indexes using enhanced vegetation index (EVI) according to the algorithms of the first eleven bioclimatic variables (VIF1-VIF11 in Table S5). We also collected the leaf dry matter content (LDMC), specific leaf area (SLA), leaf N and P concentrations (LNC and LPC) derived from remote sensing data (Moreno-Martínez et al., 2018) and soil properties (Table S6) of the surface soil (Hengl et al., 2017). We extracted the average total atmospheric nitrogen ($\mathrm{N}_{\mathrm{depNHX}}$ and $\mathrm{N}_{\mathrm{depNOY}}$, in $\mathrm{g{\bullet}m^{-2}{\bullet}yr^{-1}}$) and phosphorus deposition ($\mathrm{P}_{\mathrm{dep}}$, in $\mathrm{g{\bullet}m^{-2}{\bullet}yr^{-1}}$) from a global estimation (Wang et al., 2017b) using the mean of three nearest inversed distance weighted pixels for each site. Furthermore, we estimated the annual aridity index ($\mathrm{AI}_{\mathrm{ann}}$) and monthly AI's seasonality ($\mathrm{AI}_{\mathrm{seas}}$) using mean annual Precip divided by mean annual PET, and standard deviation of mean monthly aridity

index (= mean monthly Precip divided by mean monthly PET) divided by its mean, respectively.

Furthermore, we set another two plant types according to the PFT. The first was cropland, forest, grassland, savanna, shrubland and wetland (PT1). The second (PT) was equal to the PFT with a merged evergreen forest (=ENF and EBF), deciduous forest (=DNF and DBF), savanna (=WSA and SAV) and shrubland (=OSH and CSH). Based on all the above covariates and site-specific elevation (Table S2), we predicted the parameters and partial sensitivity functions using a random forest method (Section S1).

Appendix I. Model assessment using machine learning models

Another method to assess the LUE model was based on its distance to the machine learning model. The machine learning algorithms are non-parametric and more flexible for handling multi-dimensional data. Machine learning models should generally perform better because they can flexibly harvest the information in the forcing data, while the LUE models are constrained to their functional forms. We assume they can achieve either the same or the higher model fitness at each site. Accordingly, we simulate the GPP using the same forcing data based on the machine learning tool provided by simpleR (Lázaro-Gredilla et al., 2013), which includes ten different algorithms. We trained the ten learners using 70% of the good-quality data and assess the fitness using NSE at a daily scale. We only compared the machine learning models with the largest NSE per site (ML_{best} models) among the ten trained models per site with the site-best and global best LUE models.

However, our results show that the site-best LUE models reached values close to the ML_{best} models at most sites (see Fig. S8). NSE values at only 50 (in 196) sites were lower than the ML_{best} apparently (NSE difference larger than 0.2) at a daily scale. The global best LUE model underperformed slightly compared with the site-best LUE models. NSE values at 77 sites were apparently lower than the ML_{best} at a daily scale. At weekly and monthly scales, the NSE differences were even smaller. Notwithstanding, ML_{best} models were worse than LUE models at 72 (in 136) sites at annual scale (see the first bar in Fig. S8d). Furthermore, ML_{best} models have a considerable error in the period without good-quality forcing data for training (Fig. S9a and b), in contrast to the LUE models which did not have the same issue due to functional constraints.

References

Alton, P., 2008. Reduced carbon sequestration in terrestrial ecosystems under overcast skies compared to clear skies. Agric. For. Meteorol. 148, 1641–1653.

Alton, P.B., 2020. Representativeness of global climate and vegetation by carbon-monitoring networks; implications for estimates of gross and net primary productivity at biome and global levels. Agric. For. Meteorol., 108017

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N.C., Peylin, P., 2015. Spatiotemporal patterns of terrestrial gross primary production: a review. Rev. Geophys. 53, 785–818.

Aubinet, M., Vesala, T., Papale, D., 2012. Eddy Covariance: a Practical Guide to Measurement and Data Analysis. Springer Atmospheric Sciences, Springer Netherlands, Dordrecht, p. 438. XXII.

Bagnara, M., Sottocornola, M., Cescatti, A., Minerbi, S., Montagnani, L., Gianelle, D., Magnani, F., 2015. Bayesian optimization of a light use efficiency model for the estimation of daily gross primary productivity in a range of Italian forest ecosystems. Ecol. Modell. 306, 57–66.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Meteorol. Soc. 82, 2415–2434.

Baldocchi, D.D., 2020. How eddy covariance flux measurements have contributed to our understanding of Global Change Biology. Global Change Biol. 26, 242–260.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. Genetics 162, 2025–2035.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M.A., Baldocchi, D., Bonan, G.B., 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science 329, 834–838.

Bergh, J., Linder, S., 1999. Effects of soil warming during spring on photosynthetic recovery in boreal Norway spruce stands. Global Change Biol. 5, 245–253.

Bergh, J., McMurtrie, R.E., Linder, S., 1998. Climatic factors controlling the productivity of Norway spruce: a model-based analysis. Forest Ecol. Manag. 110, 127–139.

Bernacchi, C., Pimentel, C., Long, S.P., 2003. In vivo temperature response functions of parameters required to model RuBP-limited photosynthesis. Plant, Cell Environ. 26, 1419–1430.

Besnard, S., Carvalhais, N., Arain, M.A., Black, A., Brede, B., Buchmann, N., Chen, J., Clevers, J.G.W., Dutrieux, L.P., Gans, F., 2019. Memory effects of climate and vegetation affecting net ecosystem CO2 fluxes in global forests. PLoS One 14, e0211510.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Adv. Water Res. 16, 41–51.

Bloom, A., Williams, M., 2015. Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological" common sense" in a model–data fusion framework. Biogeosciences 12, 1299–1315.

Boese, S., Jung, M., Carvalhais, N., Teuling, A.J., Reichstein, M., 2019. Carbon–water flux coupling under progressive drought. Biogeosciences 16, 2557–2572.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach. Springer, New York.

Burnham, K.P., Anderson, D.R., Huyvaert, K.P., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav. Ecol. Sociobiol. 65, 23–35.

Carvalhais, N., Reichstein, M., Ciais, P., Collatz, G.J., Mahecha, M.D., Montagnani, L., Papale, D., Rambal, S., Seixas, J., 2010a. Identification of vegetation and soil carbon pools out of equilibrium in a process model via eddy covariance and biometric constraints. Global Change Biol. 16, 2813–2829.

Carvalhais, N., Reichstein, M., Collatz, G., Mahecha, M., Migliavacca, M., Neigh, C., Tomelleri, E., Benali, A., Papale, D., Seixas, J., 2010b. Deciphering the components of regional net ecosystem fluxes following a bottom-up approach for the Iberian Peninsula. Biogeosciences 7, 3707–3729.

Carvalhais, N., Reichstein, M., Seixas, J., Collatz, G.J., Pereira, J.S., Berbigier, P., Carrara, A., Granier, A., Montagnani, L., Papale, D., 2008. Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. Glob. Biogeochem. Cycles 22.

Cescatti, A., Marcolla, B., Santhana Vannan, S.K., Pan, J.Y., Román, M.O., Yang, X., Ciais, P., Cook, R.B., Law, B.E., Matteucci, G., Migliavacca, M., Moors, E., Richardson, A.D., Seufert, G., Schaaf, C.B., 2012. Intercomparison of MODIS albedo retrievals and in situ measurements across the global FLUXNET network. Remote Sens. Environ. 121, 323–334.

Chandrasekar, K., Sesha Sai, M., Roy, P., Dwevedi, R., 2010. Land surface water index (LSWI) response to rainfall and NDVI using the MODIS vegetation index product. Int. J. Remote Sens. 31, 3987–4005.

Cheng, S.J., Bohrer, G., Steiner, A.L., Hollinger, D.Y., Suyker, A., Phillips, R.P., Nadelhoffer, K.J., 2015. Variations in the influence of diffuse light on gross primary productivity in temperate ecosystems. Agric. For. Meteorol. 201, 98–110.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., 2014. Carbon and other biogeochemical cycles, climate change 2013: the physical science basis. In: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 465–570.

Cleverly, J., Eamus, D., Van Gorsel, E., Chen, C., Rumman, R., Luo, Q., Coupe, N.R., Li, L., Kljun, N., Faux, R., Yu, Q., Huete, A., 2016. Productivity and evapotranspiration of two contrasting semiarid ecosystems following the 2011 global carbon land sink anomaly. Agric. For. Meteorol. 220, 151–159.

De Pury, D., Farquhar, G., 1997. Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. Plant, Cell Environ. 20, 537–557.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., 2017. ESA CCI soil moisture for improved earth system understanding: state-of-the art and future directions. Remote Sens. Environ. 203, 185–215.

Efron, B., 2014. Estimation and accuracy after model selection. J. Am. Statist. Assoc. 109, 991–1007.

Farquhar, G.D., Roderick, M.L., 2003. Pinatubo, diffuse light, and the carbon cycle. Science 299, 1997–1998.

Farquhar, G.D., von Caemmerer, S.v., Berry, J., 1980. A biochemical model of photosynthetic CO 2 assimilation in leaves of C 3 species. Planta 149, 78–90.

Fernández-Martínez, M., Vicca, S., Janssens, I.A., Ciais, P., Obersteiner, M., Bartrons, M., Sardans, J., Verger, A., Canadell, J., Chevallier, F., 2017. Atmospheric deposition, CO_2, and change in the land carbon sink. Sci. Rep. 7, 1–13, 9632.

Fisher, R.A., Koven, C.D., 2020. Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. J. Adv. Model. Earth Syst. 12, e2018MS001453.

Flexas, J., Barón, M., Bota, J., Ducruet, J.-M., Gallé, A., Galmés, J., Jiménez, M., Pou, A., Ribas-Carbó, M., Sajnani, C., 2009. Photosynthesis limitations during water stress acclimation and recovery in the drought-adapted Vitis hybrid Richter-110 (V. berlandieri× V. rupestris). J. Exp. Bot. 60, 2361–2377.

Fu, P., Rich, P.M., 1999. Design and implementation of the Solar Analyst: an ArcView extension for modeling solar radiation at landscape scales. In: Proceedings of the nineteenth annual ESRI user conference. San Diego USA, pp. 1–31.

Fu, Z., Ciais, P., Bastos, A., Stoy, P.C., Yang, H., Green, J.K., Wang, B., Yu, K., Huang, Y., Knohl, A., 2020. Sensitivity of gross primary productivity to climatic drivers during the summer drought of 2018 in Europe. Philos. Trans. R. Soc. B 375, 20190747.
Gu, L., Baldocchi, D., Verma, S.B., Black, T., Vesala, T., Falge, E.M., Dowty, P.R., 2002. Advantages of diffuse radiation for terrestrial ecosystem productivity. J. Geophys. Res.: Atmospheres 107. ACL 2-1-ACL 2-23.
Gu, L., Fuentes, J.D., Shugart, H.H., Staebler, R.M., Black, T.A., 1999. Responses of net ecosystem exchanges of carbon dioxide to changes in cloudiness: results from two North American deciduous forests. J. Geophys. Res.: Atmospheres 104, 31421–31434.
Guo, Q., Hu, Z., Li, S., Yu, G., Sun, X., Zhang, L., Mu, S., Zhu, X., Wang, Y., Li, Y., 2015. Contrasting responses of gross primary productivity to precipitation events in a water-limited and a temperature-limited grassland ecosystem. Agric. For. Meteorol. 214, 169–177.
Hansen, N., Kern, S., 2004. Evaluating the CMA evolution strategy on multimodal test functions. In: International Conference on Parallel Problem Solving from Nature. Springer, pp. 282–291.
Haverd, V., Smith, B., Nieradzik, L., Briggs, P.R., Woodgate, W., Trudinger, C.M., Canadell, J.G., Cuntz, M., 2018. A new version of the CABLE land surface model (Subversion revision r4601) incorporating land use and land cover change, woody vegetation demography, and a novel optimisation-based approach to plant coordination of photosynthesis. Geosci. Model Dev. 11, 2995–3026.
He, L., Chen, J.M., Pan, Y., Birdsey, R., Kattge, J., 2012. Relationships between net primary productivity and forest stand age in US forests. Glob. Biogeochem. Cycles 26.
He, M., Ju, W., Zhou, Y., Chen, J., He, H., Wang, S., Wang, H., Guan, D., Yan, J., Li, Y., 2013. Development of a two-leaf light use efficiency model for improving the calculation of terrestrial gross primary productivity. Agric. For. Meteorol. 173, 28–39.
Henderson-Sellers, B., 1984. A new formula for latent heat of vaporization of water as a function of temperature. Q. J. R. Meteorolog. Soc. 110, 1186–1190.
Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12, e0169748.
Hollinger, D.Y., Goltz, S.M., Davidson, E.A., Lee, J.T., Tu, K., Valentine, H.T., 1999. Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest. Global Change Biol. 5, 891–902.
Horn, J., Schulz, K., 2011a. Identification of a general light use efficiency model for gross primary production. Biogeosciences 8, 999.
Horn, J., Schulz, K., 2011b. Spatial extrapolation of light use efficiency model parameters to predict gross primary production. J. Adv. Model. Earth Syst. 3.
Huang, K., Wang, S., Zhou, L., Wang, H., Zhang, J., Yan, J., Zhao, L., Wang, Y., Shi, P., 2014. Impacts of diffuse radiation on light use efficiency across terrestrial ecosystems based on eddy covariance observation in China. PLoS One 9, e110988.
Hwang, T., Kang, S., Kim, J., Kim, Y., Lee, D., Band, L., 2008. Evaluating drought effect on MODIS Gross Primary Production (GPP) with an eco-hydrological model in the mountainous forest, East Asia. Glob. Change Biol. 14, 1037–1056.
Ibrom, A., Jarvis, P.G., Clement, R., Morgenstern, K., Oltchev, A., Medlyn, B.E., Wang, Y. P., Wingate, L., Moncrieff, J.B., Gravenhorst, G., 2006. A comparative analysis of simulated and observed photosynthetic CO2 uptake in two coniferous forest canopies. Tree Physiol. 26, 845–864.
Ibrom, A., Oltchev, A., June, T., Kreilein, H., Rakkibu, G., Ross, T., Panferov, O., Gravenhorst, G., 2008. Variation in photosynthetic light-use efficiency in a mountainous tropical rain forest in Indonesia. Tree Physiol. 28, 499–508.
Ichii, K., Hashimoto, H., Nemani, R., White, M., 2005. Modeling the interannual variability and trends in gross and net primary productivity of tropical forests from 1982 to 1999. Global Planet. Change 48, 274–286.
Jarvis, A.J., Stauch, V.J., Schulz, K., Young, P.C., 2004. The seasonal temperature dependency of photosynthesis and respiration in two deciduous forests. Global Change Biol. 10, 939–950.
Jarvis, P., Linder, S., 2000. Constraints to growth of boreal forests. Nature 405, 904–905.
Jung, M., Reichstein, M., Bondeau, A., 2009. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. Biogeosciences 6, 2001–2013.
Jung, M., Reichstein, M., Margolis, H.A., Cescatti, A., Richardson, A.D., Arain, M.A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J.Q., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B.E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E.J., Papale, D., Sottocornola, M., Vaccari, F., Williams, C., 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J. Geophys. Res.-Biogeosci. 116, G00j07.
Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D.S., Haverd, V., Kohler, P., Ichii, K., Jain, A.K., Liu, J.Z., Lombardozzi, D., Nabel, J., Nelson, J.A., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rodenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., Reichstein, M., 2020. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. Biogeosciences 17, 1343–1365.
Kalliokoski, T., Makela, A., Fronzek, S., Minunno, F., Peltoniemi, M., 2018. Decomposing sources of uncertainty in climate change projections of boreal forest primary production. Agric. For. Meteorol. 262, 192–205.
Kaminski, T., Knorr, W., Schürmann, G., Scholze, M., Rayner, P., Zaehle, S., Blessing, S., Dorigo, W., Gayler, V., Giering, R., 2013. The BETHY/JSBACH carbon cycle data assimilation system: experiences and challenges. J. Geophys. Res.: Biogeosciences 118, 1414–1426.
Kanniah, K.D., Beringer, J., Hutley, L.B., Tapper, N.J., Zhu, X., 2009. Evaluation of collections 4 and 5 of the MODIS gross primary productivity product and algorithm improvement at a tropical savanna site in northern Australia. Remote Sens. Environ. 113, 1808–1822.
Keenan, T., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., Dragoni, D., Gough, C.M., Grant, R., Hollinger, D., 2012. Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO2 exchange. Global Change Biol. 18, 1971–1987.
Keenan, T.F., Migliavacca, M., Papale, D., Baldocchi, D., Reichstein, M., Torn, M., Wutzler, T., 2019. Widespread inhibition of daytime ecosystem respiration. Nat. Ecol. Evol. 3, 407–415.
Knohl, A., Baldocchi, D.D., 2008. Effects of diffuse radiation on canopy gas exchange processes in a forest ecosystem. J. Geophys. Res.: Biogeosciences 113.
Kulmatiski, A., Beard, K.H., 2013. Woody plant encroachment facilitated by increased precipitation intensity. Nat. Clim. Chang. 3, 833–837.
Law, B.E., Falge, E., Gu, L., Baldocchi, D.D., Bakwin, P., Berbigier, P., Davis, K., Dolman, A.J., Falk, M., Fuentes, J.D., Goldstein, A., Granier, A., Grelle, A., Hollinger, D., Janssens, I.A., Jarvis, P., Jensen, N.O., Katul, G., Mahli, Y., Matteucci, G., Meyers, T., Monson, R., Munger, W., Oechel, W., Olson, R., Pilegaard, K., Paw, K.T., Thorgeirsson, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2002. Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation. Agric. For. Meteorol. 113, 97–120. Pii s0168-1923(02)00104-1.
Lawrence, D.M., Oleson, K.W., Flanner, M.G., Thornton, P.E., Swenson, S.C., Lawrence, P.J., Zeng, X., Yang, Z.L., Levis, S., Sakaguchi, K., 2011. Parameterization improvements and functional and structural advances in version 4 of the community land model. J. Adv. Model. Earth Syst. 3.
Lázaro-Gredilla, M., Titsias, M.K., Verrelst, J., Camps-Valls, G., 2013. Retrieval of biophysical parameters with heteroscedastic Gaussian processes. IEEE Geosci. Remote Sens. Lett. 11, 838–842.
Leuning, R., 1995. A critical appraisal of a combined stomatal-photosynthesis model for C3 plants. Plant, Cell Environ. 18, 339–355.
Lilliefors, H.W., 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. J. Am. Statist. Assoc. 62, 399–402.
Liu, Q., Piao, S., Janssens, I.A., Fu, Y., Peng, S., Lian, X., Ciais, P., Myneni, R.B., Peñuelas, J., Wang, T., 2018. Extension of the growing season increases vegetation exposure to frost. Nat. Commun. 9, 1–8.
Liu, S., Zhuang, Q., He, Y., Noormets, A., Chen, J., Gu, L., 2016. Evaluating atmospheric CO2 effects on gross primary productivity and net ecosystem exchanges of terrestrial ecosystems in the conterminous United States using the AmeriFlux data and an artificial neural network approach. Agric. For. Meteorol. 220, 38–49.
Luo, Y., Schuur, E.A., 2020. Model parameterization to represent processes at unresolved scales and changing properties of evolving systems. Global Change Biol. 26, 1109–1117.
Ma, J., Xiao, X., Miao, R., Li, Y., Chen, B., Zhang, Y., Zhao, B., 2019. Trends and controls of terrestrial gross primary productivity of China during 2000–2016. Environ. Res. Lett. 14, 084032.
Mäkelä, A., Hari, P., Berninger, F., Hänninen, H., Nikinmaa, E., 2004. Acclimation of photosynthetic capacity in Scots pine to the annual cycle of temperature. Tree Physiol. 24, 369–376.
Mäkelä, A., Kolari, P., Karimäki, J., Nikinmaa, E., Perämäki, M., Hari, P., 2006. Modelling five years of weather-driven variation of GPP in a boreal forest. Agric. For. Meteorol. 139, 382–398.
Mäkelä, A., Pulkkinen, M., Kolari, P., Lagergren, F., Berbigier, P., Lindroth, A., Loustau, D., Nikinmaa, E., Vesala, T., Hari, P., 2008. Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer sites in Europe. Global Change Biol. 14, 92–108.
McCallum, I., Franklin, O., Moltchanova, E.V., Merbold, L., Schmullius, C., Shvidenko, A., Schepaschenko, D., Fritz, S., 2013. Improved light and temperature responses for light-use-efficiency-based GPP models. Biogeosciences 10, 6577–6590.
Medlyn, B., Dreyer, E., Ellsworth, D., Forstreuter, M., Harley, P., Kirschbaum, M., Le Roux, X., Montpied, P., Strassemeyer, J., Walcroft, A., 2002. Temperature response of parameters of a biochemically based model of photosynthesis. II. A review of experimental data. Plant, Cell Environ. 25, 1167–1179.
Medlyn, B.E., 1998. Physiological basis of the light use efficiency model. Tree Physiol. 18, 167–176.
Mercado, L.M., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M., Cox, P.M., 2009. Impact of changes in diffuse radiation on the global land carbon sink. Nature 458, 1014–1U87.
Migliavacca, M., Reichstein, M., Richardson, A.D., Mahecha, M.D., Cremonese, E., Delpierre, N., Galvagno, M., Law, B.E., Wohlfahrt, G., Andrew Black, T., 2015. Influence of physiological phenology on the seasonal pattern of ecosystem respiration in deciduous forests. Global Change Biol. 21, 363–376.
Minunno, F., Peltoniemi, M., Launiainen, S., Aurela, M., Lindroth, A., Lohila, A., Mammarella, I., Minkkinen, K., Mäkelä, A., 2015. A simplified gross primary production and evapotranspiration model for boreal coniferous forests–is a generic calibration sufficient? Geosci. Model Dev. Discussions 8, 5089–5137.

Monteith, J., 1972. Solar radiation and productivity in tropical ecosystems. J. Appl. Ecol. 9, 747–766.
Moré, J.J., Sorensen, D.C., 1983. Computing a trust region step. SIAM J. Sci. Stat. Comput. 4, 553–572.
Moreno-Martínez, Á., Camps-Valls, G., Kattge, J., Robinson, N., Reichstein, M., van Bodegom, P., Kramer, K., Cornelissen, J.H.C., Reich, P., Bahn, M., 2018. A methodology to derive global maps of leaf traits using remote sensing and climate data. Remote Sens. Environ. 218, 69–88.
Myneni, R.B., Ramakrishna, R., Nemani, R., Running, S.W., 1997. Estimation of global leaf area index and absorbed PAR using radiative transfer models. IEEE Trans. Geosci. Remote Sens. 35, 1380–1393.
Oliphant, A., Dragoni, D., Deng, B., Grimmond, C., Schmid, H.-P., Scott, S., 2011. The role of sky conditions on gross primary production in a mixed deciduous forest. Agric. For. Meteorol. 151, 781–791.
Omlin, M., Reichert, P., 1999. A comparison of techniques for the estimation of model prediction uncertainty. Ecol. Modell. 115, 45–59.
Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci. Data 7, 1–27.
Pelkonen, P., Hari, P., 1980. The dependence of the springtime recovery of CO2 uptake in Scots pine on temperature and internal factors. Flora 169, 398–404.
Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J.G., Ciais, P., Friedlingstein, P., Sitch, S., 2020. Interannual variation of terrestrial carbon cycle: Issues and perspectives. Global Change Biol. 26, 300–318.
Pilegaard, K., Ibrom, A., 2020. Net carbon ecosystem exchange during 24 years in the Sorø Beech Forest–relations to phenology and climate. Tellus B: Chem. Phys. Meteorol. 72, 1–17.
Potter, C.S., Randerson, J.T., Field, C.B., Matson, P.A., Vitousek, P.M., Mooney, H.A., Klooster, S.A., 1993. Terrestrial ecosystem production: a process model based on global satellite and surface data. Glob. Biogeochem. Cycles 7, 811–841.
Priestley, C.H.B., Taylor, R., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev. 100, 81–92.
Propastin, P., Ibrom, A., Knohl, A., Erasmi, S., 2012. Effects of canopy photosynthesis saturation on the estimation of gross primary productivity from MODIS data in a tropical forest. Remote Sens. Environ. 121, 252–260.
Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Global Change Biol. 11, 1424–1439.
Roderick, M.L., Farquhar, G.D., Berry, S.L., Noble, I.R., 2001. On the direct effect of clouds and atmospheric particles on the productivity and structure of vegetation. Oecologia 129, 21–30.
Roman, D.T., Novick, K.A., Brzostek, E.R., Dragoni, D., Rahman, F., Phillips, R.P., 2015. The role of isohydric and anisohydric species in determining ecosystem-scale response to severe drought. Oecologia 179, 641–654.
Ross, J., 1981. The Radiation Regime and Architecture of Plant Stands. Dr W. Junk Pulishers, The Hague.
Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W. and Harlan, J.C., 1974. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. NASA/GSFC Type III Final Report, Greenbelt, Md, 371.
Rubel, F., Brugger, K., Haslinger, K., Auer, I., 2017. The climate of the European Alps: shift of very high resolution Köppen-Geiger climate zones 1800–2100. Meteorol. Z. 26, 115–125.
Ruimy, A., Kergoat, L., Bondeau, A., Intercomparison, T.P.O.T.P.N.M., 1999. Comparing global models of terrestrial net primary productivity (NPP): Analysis of differences in light absorption and light-use efficiency. Global Change Biol. 5, 56–64.
Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., Hashimoto, H., 2004. A continuous satellite-derived measure of global terrestrial primary production. Bioscience 54, 547–560.
Running, S.W., Zhao, M., 2015. Daily GPP and annual NPP (MOD17A2/A3) products NASA Earth Observing System MODIS land algorithm. MOD17 User's Guide, 2015. MODIS Land Team, Washington, DC.
Savitzky, A., Golay, M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.
Schaaf, C. and Wang, Z., 2015. MCD43A4 MODIS/Terra+ Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global-500m V006 [Data set]. NASA EOSDIS Land Processes DAAC.
Sims, P.L., Bradford, J., 2001. Carbon dioxide fluxes in a southern plains prairie. Agric. For. Meteorol. 109, 117–134.
Smith, W.K., Reed, S.C., Cleveland, C.C., Ballantyne, A.P., Anderegg, W.R.L., Wieder, W. R., Liu, Y.Y., Running, S.W., 2016. Large divergence of satellite and Earth system model estimates of global terrestrial CO2 fertilization. Nat. Clim. Chang. 6, 306–310.
Stocker, B.D., Wang, H., Smith, N.G., Harrison, S.P., Keenan, T.F., Sandoval, D., Davis, T., Prentice, I.C., 2020. P-model v1. 0: an optimality-based light use efficiency model for simulating ecosystem gross primary production. Geosci. Model Dev. 13, 1545–1581.
Stocker, B.D., Zscheischler, J., Keenan, T.F., Prentice, I.C., Peñuelas, J., Seneviratne, S.I., 2018. Quantifying soil moisture impacts on light use efficiency across biomes. New Phytol. 218, 1430–1449.
Sun, Z., Wang, X., Yamamoto, H., Tani, H., Zhong, G., Yin, S., Guo, E., 2018. Spatial pattern of GPP variations in terrestrial ecosystems and its drivers: climatic factors, CO2 concentration and land-cover change, 1982–2015. Ecol. Inform. 46, 156–165.
Tanja, S., Berninger, F., Vesala, T., Markkanen, T., Hari, P., Mäkelä, A., Ilvesniemi, H., Hänninen, H., Nikinmaa, E., Huttula, T., 2003. Air temperature triggers the recovery of evergreen boreal forest photosynthesis in spring. Global Change Biol. 9, 1410–1426.
Tardieu, F., Simonneau, T., 1998. Variability among species of stomatal control under fluctuating soil water status and evaporative demand: modelling isohydric and anisohydric behaviours. J. Exp. Bot. 419–432.
Teubner, I.E., Forkel, M., Camps-Valls, G., Jung, M., Miralles, D.G., Tramontana, G., van der Schalie, R., Vreugdenhil, M., Mosinger, L., Dorigo, W.A., 2019. A carbon sink-driven approach to estimate gross primary production from microwave satellite observations. Remote Sens. Environ. 229, 100–113.
Tian, X.L., Minunno, F., Cao, T.J., Peltoniemi, M., Kalliokoski, T., Makela, A., 2020. Extending the range of applicability of the semi-empirical ecosystem flux model PRELES for varying forest types and climate. Global Change Biol. 26, 2923–2943.
Toni, T., Stumpf, M.P., 2010. Simulation-based model selection for dynamical systems in systems and population biology. Bioinformatics 26, 104–110.
Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P., 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc., Interface 6, 187–202.
Tramontana, G., Jung, M., Schwalm, C.R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M.A., Cescatti, A., Kiely, G., 2016. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. Biogeosciences 13, 4291–4313.
Trautmann, T., Koirala, S., Carvalhais, N., Eicker, A., Fink, M., Niemann, C., Jung, M., 2018. Understanding terrestrial water storage variations in northern latitudes across scales. Hydrol. Earth Syst. Sci. 22, 4061–4082.
Turner, D., Ritts, W., Styles, J., Yang, Z., Cohen, W., Law, B., Thornton, P., 2006a. A diagnostic carbon flux model to monitor the effects of disturbance and interannual variation in climate on regional NEP. Tellus B: Chem. Phys. Meteorol. 58, 476–490.
Turner, D.P., Ritts, W.D., Cohen, W.B., Gower, S.T., Running, S.W., Zhao, M., Costa, M. H., Kirschbaum, A.A., Ham, J.M., Saleska, S.R., 2006b. Evaluation of MODIS NPP and GPP products across multiple biomes. Remote Sens. Environ. 102, 282–292.
Veroustraete, F., Sabbe, H., Eerens, H., 2002. Estimation of carbon mass fluxes over Europe using the C-Fix model and Euroflux data. Remote Sens. Environ. 83, 376–399.
Viovy, N., 2018a. CRUNCEP version 7–atmospheric forcing data for the community land model. In: Research Data Archive at the National Center for Atmospheric Research, 10. Computational and Information Systems Laboratory.
Vuichard, N., Messina, P., Luyssaert, S., Guenet, B., Zaehle, S., Ghattas, J., Bastrikov, V., Peylin, P., 2019. Accounting for carbon and nitrogen interactions in the global terrestrial ecosystem model ORCHIDEE (trunk version, rev 4999): multi-scale evaluation of gross primary production. Geosci. Model Dev. 12, 4751–4779.
Wang, H., Prentice, I.C., Keenan, T.F., Davis, T.W., Wright, I.J., Cornwell, W.K., Evans, B. J., Peng, C., 2017a. Towards a universal model for carbon dioxide uptake by plants. Nat. Plants 3, 734–741.
Wang, R., Goll, D., Balkanski, Y., Hauglustaine, D., Boucher, O., Ciais, P., Janssens, I., Penuelas, J., Guenet, B., Sardans, J., 2017b. Global forest carbon uptake due to nitrogen and phosphorus deposition from 1850 to 2100. Global Change Biol. 23, 4854–4872.
Wang, S., Huang, K., Yan, H., Yan, H., Zhou, L., Wang, H., Zhang, J., Yan, J., Zhao, L., Wang, Y., 2015. Improving the light use efficiency model for simulating terrestrial vegetation gross primary production by the inclusion of diffuse radiation across ecosystems in China. Ecol. Complex. 23, 1–13.
Wang, S., Ibrom, A., Bauer-Gottwein, P., Garcia, M., 2018. Incorporating diffuse radiation into a light use efficiency and evapotranspiration model: an 11-year study in a high latitude deciduous forest. Agric. For. Meteorol. 248, 479–493.
Wenzel, S., Cox, P.M., Eyring, V., Friedlingstein, P., 2016. Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO2. Nature 538, 499-+.
Worden, J., Saatchi, S., Keller, M., Bloom, A., Fu, R., Worden, S., Liu, J., Parazoo, N., Fisher, J.B., Worden, H., 2021. Satellite Observations of the Tropical Terrestrial Carbon Balance and Interactions with the Water Cycle During the 21st Century. Wiley Online Library.
Wu, J., Jansson, P.-E., van der Linden, L., Pilegaard, K., Beier, C., Ibrom, A., 2013. Modelling the decadal trend of ecosystem carbon fluxes demonstrates the important role of functional changes in a temperate deciduous forest. Ecol. Modell. 260, 50–61.
Wu, J., Linden, L., Lasslop, G., Carvalhais, N., Pilegaard, K., Beier, C., Ibrom, A., 2012. Effects of climate variability and functional changes on the interannual variation of the carbon balance in a temperate deciduous forest. Biogeosciences 9, 13–28.
Xiao, X., Hollinger, D., Aber, J., Goltz, M., Davidson, E.A., Zhang, Q., Moore B., III, 2004. Satellite-based modeling of gross primary production in an evergreen needleleaf forest. Remote Sens. Environ. 89, 519–534.
Xu, T., Hutchinson, M.F., 2013. New developments and applications in the ANUCLIM spatial climatic and bioclimatic modelling package. Environ. Model. Softw. 40, 267–279.
Yan, H., Wang, S.Q., Yu, K.L., Wang, B., Yu, Q., Bohrer, G., Billesbach, D., Bracho, R., Rahman, F., Shugart, H.H., 2017. A novel diffuse fraction-based two-leaf light use efficiency model: an application quantifying photosynthetic seasonality across 20 AmeriFlux flux tower sites. J. Adv. Model. Earth Syst. 9, 2317–2332.
Yang, Q., Blanco, N.E., Hermida-Carrera, C., Lehotai, N., Hurry, V., Strand, Å., 2020. Two dominant boreal conifers use contrasting mechanisms to reactivate photosynthesis in the spring. Nat. Commun. 11, 1–12.
Yuan, W., Cai, W., Xia, J., Chen, J., Liu, S., Dong, W., Merbold, L., Law, B., Arain, A., Beringer, J., 2014. Global comparison of light use efficiency models for simulating terrestrial vegetation gross primary production based on the LaThuile database. Agric. For. Meteorol. 192, 108–120.
Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L.L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A.H., Goulden, M.L., 2007. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. Agric. For. Meteorol. 143, 189–207.

Yuan, Wenping, Zheng, Yi, Piao, Shilong, et al., 2019. Science Advances 5 (8). https://doi.org/10.1126/sciadv.aax1396. https://www.science.org/doi/10.1126/sciadv.aax1396.

Zaehle, S., Medlyn, B.E., De Kauwe, M.G., Walker, A.P., Dietze, M.C., Hickler, T., Luo, Y., Wang, Y.P., El-Masri, B., Thornton, P., 2014. Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate F ree-A ir CO 2 E nrichment studies. New Phytol. 202, 803–822.

Zhang, B., Tan, X., Wang, S., Chen, M., Chen, S., Ren, T., Xia, J., Bai, Y., Huang, J., Han, X., 2017a. Asymmetric sensitivity of ecosystem carbon and water processes in response to precipitation change in a semi-arid steppe. Funct. Ecol. 31, 1301–1311.

Zhang, J., Wang, X., Ren, J., 2021. Simulation of gross primary productivity using multiple light use efficiency models. Land 10, 329.

Zhang, L.-X., Zhou, D.-C., Fan, J.-W., Hu, Z.-M., 2015. Comparison of four light use efficiency models for estimating terrestrial gross primary production. Ecol. Modell. 300, 30–39.

Zhang, Q., Wang, Y., Pitman, A., Dai, Y., 2011. Limitations of nitrogen and phosphorous on the terrestrial carbon uptake in the 20th century. Geophys. Res. Lett. 38.

Zhang, S., Zhang, J., Bai, Y., Koju, U.A., Igbawua, T., Chang, Q., Zhang, D., Yao, F., 2018. Evaluation and improvement of the daily boreal ecosystem productivity simulator in simulating gross primary productivity at 41 flux sites across Europe. Ecol. Modell. 368, 205–232.

Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., Dong, J., 2017b. A global moderate resolution dataset of gross primary production of vegetation for 2000–2016. Sci. data 4, 170165.