Addressing Challenges in Simulating Inter–annual Variability of Gross Primary Production

Ranit De[1,2], Shanning Bao[1,3], Sujan Koirala[1], Alexander Brenning[2,4], Markus Reichstein[1,4], Torbern Tagesson[5], Michael Liddell[6], Andreas Ibrom[7], Sebastian Wolf[8], Ladislav Šigut[9], Lukas Hörtnagl[8], William Woodgate[10,11], Mika Korkiakoski[12], Lutz Merbold[13], T. Andrew Black[14], Marilyn Roland[15], Anne Klosterhalfen[16], Peter D. Blanken[17], Sara Knox[18,19], Simone Sabbatini[20,21], Bert Gielen[15], Leonardo Montagnani[22], Rasmus Fensholt[23], Georg Wohlfahrt[24], Ankur R. Desai[25], Eugénie Paul-Limoges[26], Marta Galvagno[27], Albin Hammerle[24], Georg Jocher[28,9], Borja Ruiz Reverter[29], David Holl[30], Jiquan Chen[31], Luca Vitale[32], M. Altaf Arain[33], and Nuno Carvalhais[1,4,34]

[1]Department for Biogeochemical Integration, Max Planck Institute for Biogeochemistry, 07745, Jena, Germany
[2]Friedrich Schiller University Jena, Department of Geography, Löbdergraben 32, 07743, Jena, Germany
[3]National Space Science Center, Chinese Academy of Sciences, 100190, Beijing, China
[4]ELLIS Unit Jena, Jena, Germany
[5]Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden
[6]Centre for Tropical, Environmental, and Sustainability Sciences, James Cook University, Cairns, Queensland, Australia
[7]Department of Environment and Resource Engineering, Technical University of Denmark (DTU), Bygningstorvet 115, 2800 Kgs. Lyngby, Denmark
[8]Department of Environmental Systems Science, ETH Zürich, 8092, Zürich, Switzerland
[9]Global Change Research Institute of the Czech Academy of Sciences, Belidla 4a, 60300, Brno, Czech Republic
[10]School of the Environment, The University of Queensland, St Lucia 4072, Australia
[11]CSIRO, Space and Astronomy, Kensington, 6151, WA, Australia
[12]Finnish Meteorological Institute, Climate System Research Unit, P.O. Box 503, 00101 Helsinki, Finland
[13]Integrative Agroecology Group, Agroscope, 8046, Zürich, Switzerland
[14]Faculty of Land and Food Systems, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada
[15]Plants and Ecosystems (PLECO), Department of Biology, University of Antwerp, B-2610 Wilrijk, Belgium
[16]Bioclimatology, University of Goettingen, 37077, Goettingen, Germany
[17]Department of Geography, University of Colorado Boulder, Boulder, CO 80309, USA
[18]Department of Geography, McGill University, Montreal, Quebec H3A 0B9, Canada
[19]Department of Geography, The University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada
[20]CMCC Foundation - Euro-Mediterranean Center on Climate Change, 73100, Lecce, Italy
[21]Department for Innovation in Biological, Agro-food and Forest Systems (DIBAF), University of Tuscia, 01100, Viterbo, Italy
[22]Faculty of Agricultural, Environmental and Food Sciences, Free University of Bozen-Bolzano, 39100, Bolzano, Italy
[23]Department of Geosciences and Natural Resource Management, University of Copenhagen, Øster Voldgade 10, DK-1350 Copenhagen, Denmark
[24]Universität Innsbruck, Institut für Ökologie, Sternwartestr. 15, 6020 Innsbruck, Austria
[25]Dept of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI 53706 USA
[26]Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), 8903, Birmensdorf, Switzerland
[27]Environmental Protection Agency of Aosta Valley, Climate Change Unit, (ARPA Valle d'Aosta), 11020, Saint-Christophe AO, Italy
[28]Thünen Institute of Climate-Smart Agriculture, Bundesallee 65, 38116, Braunschweig, Germany
[29]Departamento de Química e Física, Universidade Federal da Paraíba - Campus II, 58397-000 Areia, Paraíba, Brazil
[30]University Hamburg, Institute of Soil Science, 20146, Hamburg, Germany
[31]Department of Geography, Environment, and Spatial Sciences, Michigan State University, MI 48823, USA
[32]Institute for Agriculture and Forestry Systems in the Mediterranean (ISAFoM), P.le Enrico Fermi 1, 80055, Portici, Italy
[33]School of Earth, Environment and Society, McMaster University, Hamilton, Ontario L8S 4K1, Canada
[34]CENSE, Departamento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, Caparica, Portugal

Key Points:

- We investigated the limitations of biogeochemical models in simulating inter–annual variability (IAV) of gross primary production (GPP).
- Capturing year–to–year variability of model parameters and diurnal GPP peaks can be key to understanding IAV.
- Variability in model performance is majorly influenced by model types, parameterization strategies, and site characteristics.

Corresponding author: Ranit De, rde@bgc-jena.mpg.de or de.ranit19@gmail.com

Corresponding author: Nuno Carvalhais, ncarvalhais@bgc-jena.mpg.de

Abstract

A long-standing challenge in studying the global carbon cycle has been understanding the factors controlling inter–annual variation (IAV) of carbon fluxes related to vegetation photosynthesis and respiration, and improving their representations in existing biogeochemical models. Here, we compared an optimality-based mechanistic model and a semi-empirical light use efficiency model to understand how current models can be improved to simulate IAV of gross primary production (GPP). Both models simulated hourly GPP and were parameterized for (1) each site–year, (2) each site with an additional constraint on IAV ($Cost^{IAV}$), (3) each site, (4) each plant–functional type, and (5) globally. This was followed by forward runs using calibrated parameters, and model evaluations at different temporal scales across 198 eddy covariance sites. Both models performed better on hourly scale than annual scale for most sites. Specifically, the mechanistic model substantially improved when drought stress was explicitly included. Most of the variability in model performances was due to model types and parameterization strategies. The semi-empirical model produced statistically better hourly simulations than the mechanistic model, and site–year parameterization yielded better annual performance for both models. Annual model performance did not improve even when parameterized using $Cost^{IAV}$. Furthermore, both models underestimated the peaks of diurnal GPP in each site–year, suggesting that improving predictions of peaks could produce a comparatively better annual model performance. GPP of forests were better simulated than grassland or savanna sites by both models. Our findings reveal current model deficiencies in representing IAV of carbon fluxes and guide improvements in further model development.

Plain Language Summary

Terrestrial vegetation assimilates and releases carbon dioxide through photosynthesis and respiration, respectively, and their net magnitude determines if vegetation can be a sink or source of carbon. We are interested in understanding what controls the inter–annual variability (IAV) of gross primary production (GPP) which represents photosynthesis, in a given location and how their representations can be improved in models simulating GPP. Here, we considered a mechanistic model that can be applied equally well globally, and a data-driven semi-empirical model. We found both models better simulated diurnal and seasonal cycles of GPP than IAV. Such differences probably stem from model parameters, as critical ecosystem functions they represent may not be well-constrained or model structures may lack critical representations via inaccurate simulation of peak diurnal GPP and drought stress. The IAV of GPP was comparatively better simulated if model parameters were fine-tuned with data from specific years. Another challenge is that IAV of GPP can also be observed due to disturbances, such as forest fire, and human management besides natural causes, which were also not represented in models. Our results suggest that learning the variability of model parameters over the years can be key to better simulation of the IAV of GPP.

1 Introduction

The global carbon cycle is an important biogeochemical cycle, which affects the climate on Earth (Schimel, 2001). Terrestrial vegetation, which covers a large part of the land area, assimilates atmospheric carbon dioxide (CO_2) through photosynthesis. Simultaneously, CO_2 of similar magnitude is released into the atmosphere during terrestrial ecosystem respiration (TER). The net balance of these two fluxes determines if terrestrial vegetation acts as a sink or source of carbon (Ruehr et al., 2023). Terrestrial gross primary production (GPP) can be defined as 'apparent' photosynthesis, i.e., the rate at which the vegetation assimilates carbon through photosynthesis minus the loss of carbon only through photorespiration (Plummer, 2006; Wohlfahrt & Gu, 2015). GPP can be estimated directly using gas exchange measurements at the leaf and canopy scales (Jez et al., 2021), and indirectly through measurements of net ecosystem exchange (NEE) using the eddy covariance (EC) method at the ecosystem or landscape scale (D. D. Baldocchi, 2003). Though the GPP estimated using the EC method represents 'apparent' photosynthesis, its magnitude can be closer to 'true' photosynthesis which is the actual amount of carbon assimilated due to overestimation of daytime mitochondrial respiration in flux–partitioning algorithm (Reichstein et al., 2005; Wohlfahrt & Gu, 2015). Furthermore, a large variety of biogeochemical models have been developed to simulate and upscale carbon fluxes from local to regional or global scales to better describe the global carbon cycle (Xiao et al., 2014; Burton et al., 2023; Dannenberg et al., 2023; Nelson et al., 2024).

Biogeochemical models that simulate GPP can be of different types and complexities. On the one hand, process-based models, such as the models used in the Trends in Net Land-Atmosphere Carbon Exchange (TRENDY) project, mechanistically describe the physiological processes involved in photosynthesis or plant respiration (Sitch et al., 2015). The ability of these process-based models to capture a certain process largely depends on the underlying model structure and calibration of model parameters (Anav et al., 2015). Similar, but simpler than fully mechanistic approaches are the models constructed on the concept of light use efficiency (LUE), which treat a canopy as one big leaf, but where the GPP is calculated as the product of the absorbed photosynthetically active radiation (aPAR) and LUE (Monteith, 1972). These models are semi-empirical as they combine both the simplicity of empirical models and the theoretical mechanisms that underpin process-based models (Running et al., 2000; Yuan et al., 2007; J. Chen, 2021). On the other hand, empirical models are largely based on learning regression functions to establish a general relation between input data, such as meteorology and ecosystem properties, and the desired output, such as GPP. At the site level, the ability of such data–driven models (Jung et al., 2011, 2020) to accurately simulate the GPP fluxes generally outperforms mechanistic approaches, but they are largely reliant on good quality training data and generally lack comprehensive representations of long-term forcing functions, such as CO_2 fertilization effect, i.e., increased GPP with the increase in atmospheric CO_2 concentration (Schimel et al., 2015).

Considering the methodological diversity and differences in GPP estimates, various model benchmarking and model–data integration experiments have been designed to compare approaches, but also to unveil drivers of ecosystem functioning for various bioclimatic and vegetation types, across spatial and temporal scales. A long-standing challenge, and still a key area of interest, lies in understanding the factors controlling inter–annual variability (IAV) of the various carbon fluxes (D. Baldocchi et al., 2018). The challenge presents itself from the mechanistic to the more data–driven approaches and contests the dominant role of meteorology in determining the IAV of ecosystem fluxes (Richardson et al., 2007). At the local ecosystem level, Wu et al. (2012) looked at the IAV of net ecosystem fluxes by fitting the parameters of a semi-empirical model at shorter timescales to capture the seasonality, but also annual variability of model parameters. The approach allows testing the role of changes in ecosystem functioning in the IAV of carbon fluxes (Richardson et al., 2007). They concluded that climate and parametric variability control IAV of ecosystem fluxes at shorter and longer timescales, respectively. Simultaneously, Fatichi and Ivanov (2014) highlighted the role of climate when using 200 years of hourly synthetic meteorological data to force an ecohydrological model to find that the random occurrence of favourable weather conditions at certain hours of the day can be a major predictor of IAV of net primary production (NPP). This statistical relationship was corroborated by Zscheischler et al. (2016) using actual flux data from EC sites from forested areas in North America, where the 91^{st} percentile values of hourly GPP flux, i.e., peak GPP values, substantially contributed to the IAV of GPP flux. These studies highlight the correlation between the distribution tails and the IAV in EC fluxes. However, there is no robust pattern across sites nor do they challenge there is no variability in ecosystem function.

More recently, a model selection study compared an ensemble of 5600 possible semi-empirical LUE model structures to find a global best model structure (Bao et al., 2022). The best LUE model is calibrated at a daily timescale per site and explains the variability of GPP fluxes across the FLUXNET EC network (Pastorello et al., 2020), considering the effect of various environmental conditions on maximum LUE through partial sensitivity functions. Though the best global model performed similarly to the best model selected for each site at the daily resolution, it failed to represent the variability of annually aggregated GPP fluxes for 74% of sites, i.e., the Nash-Sutcliffe efficiency (NSE) of model performance (Nash & Sutcliffe, 1970) was below or equal to 0.5. This finding may be attributed to (1) the use of daily data in the study, as the model had no information on the favourable conditions that occurred in a diurnal cycle and failed to simulate the diurnal GPP peaks which had a major influence on IAV (Fatichi & Ivanov, 2014; Zscheischler et al., 2016; Bao et al., 2022), (2) the assumption of invariance in ecosystem function, i.e., values of model parameters remain constant for all site–years in a site, and (3) the need to explicitly consider different timescales in the cost function (Desai, 2010).

In contrast, Mengoli et al. (2022) proposed an optimality-based framework (Wang et al., 2017; Stocker et al., 2020), i.e., process-based P-model which simulates GPP mechanistically and differentiates between instantaneous and acclimated photosynthetic responses. This model demonstrated its capability in simulating half-hourly GPP dynamics at ten EC sites, covering four vegetation classes for limited time periods. Whereas, the performance of this modelling framework across sites representing diverse climate–vegetation features and various temporal resolutions were not evaluated. Though this modelling framework considers the effect of temperature, vapour pressure deficit (VPD), atmospheric CO_2 concentration, solar radiation, and the fraction of absorbed photosynthetically active radiation (fAPAR), it does not explicitly consider the effect of drought stress on GPP variability at sub-daily scale. Recently, Mengoli et al. (2023) proposed an improved version of this model by incorporating climatic aridity and calculating a scaling factor for GPP. However, in the improved model, the scaling factor could only be applied to improve the simulation of daily GPP.

The challenge to correctly reproduce IAV is also apparent on a global scale. Anav et al. (2015) further drew attention to the disagreement in annual GPP, modelled by various global GPP modelling frameworks, such as a data–driven model–tree–ensemble (Jung et al., 2011), process-based models in the TRENDY project (Sitch et al., 2015), and the CARBONES dataset (Kuppel et al., 2013) which was derived using a hybrid approach. These discrepancies highlight that site level limitations in simulating IAV propagate to larger scales where additional mechanisms play a role in the IAV of ecosystem fluxes, such as natural or anthropogenic disturbances and land–use landcover change (McGuire et al., 2001; Bultan et al., 2022).

As such, here we explore ecosystem-level estimations of GPP flux to systematically investigate how various factors can be linked to describing the IAV of GPP flux, such as peak values of diurnal GPP, climatic conditions, and variables represented by model parameters, which are usually hard to measure directly and can be difficult to interpret even when various modelling approaches are adopted. We tested the impact of the constant or time–varying parameterizations and evaluated their performances in capturing GPP variability at various temporal aggregation scales, especially at the annual scale. We also tested the hypothesis that observational constraints complement and enhance theoretically-grounded process formulations and that improving the model simulations at the sub-daily scale improves the prediction of IAV of GPP. Additional analysis on parameter inversion approaches and cost functions, as well as on parametric variability are treated in a companion paper [companion paper citation, in prep]. In this study, we aim to answer

1. How well does a mechanistic model perform compared to a semi-empirical model across various temporal scales with different model parameterization approaches?
2. Can the performance of a mechanistic model be improved if drought stress is included?
3. What factors influence the variability of model performance at different temporal scales?
4. How much are the differences in model performance between a mechanistic and a semi-empirical model as well as across plant–functional types (PFT) and climate–vegetation types?
5. Does improved simulations of peak diurnal GPP lead to improved simulations of IAV of GPP?

2 Methods and data

In this study, we focused on parameterization of both a semi-empirical model, at daily and sub-daily scales, and a mechanistic model at a sub-daily scale using various parameterization strategies consisting of different subsets of data and cost functions (Fig. 1). Thereafter, we performed forward runs of models with calibrated parameters at the temporal resolution of model parameterization data and evaluated model performances at different temporal aggregations (Fig. 1). The following sections describe each methodological step in a detailed manner.

2.1 Models

2.1.1 Mechanistic model: P-model of Mengoli

Stocker et al. (2020) proposed the first version of the P-model based on theories formulated by Wang et al. (2017), which unified the classic Farquhar–von Caemmerer–Berry (FvCB) model (Farquhar et al., 1980) with the simplified formation of big leaf LUE models (Monteith, 1972). The probable reasons behind using the 'P' in the P-model are (1) 'P' stands for photosynthesis, (2) classically, GPP used to be denoted by 'P' (Monteith, 1972), and (3) the initial of the lead author (Prentice et al., 2014) who formulated the theories behind the model starts with 'P' (B. D. Stocker, personal communication, May 06, 2024). The underlying equations of the P-model were formulated based on the optimality principle (Prentice et al., 2014) and the coordination principle (J.-L. Chen et al., 1993; Maire et al., 2012). According to the optimality principle, plants aim to optimize the cost of transpiring water to assimilate CO_2 through the stomata. In the P-model, the ratio of leaf internal and ambient CO_2 concentration ($\chi = C_i/C_a$) is calculated for which the above-described cost is minimal, and the sensitivity (ξ) of χ to VPD is predicted. The coordination principle describes the achievement of equilibrium between the maximum rate of carboxylation ($V_{c_{max}}$) and electron transport (J_{max}) by the plants.

Mengoli et al. (2022) adapted the first version of the P-model to simulate half-hourly GPP dynamics. Here, we applied this same model at an hourly scale and called it the $\mathrm{P_{hr}}$ model. The major improvement in this version was defining an explicit differentiation between instantaneous (such as RuBisCo and light-limited carbon assimilation), and photosynthetic responses ($V_{c_{max}}$, J_{max}, and ξ) which acclimate over time in response to environmental conditions. One of the important aspects of this $\mathrm{P_{hr}}$ model is that the parameters associated with cellular biochemistry acclimate to favourable conditions during the day over a period of time or acclimation time (A_t). In this study, we considered the favourable condition as the average of three hourly input data points in the middle of the day from 11:00 (hh:mm) LT, 12:00 (hh:mm) LT, and 13:00 (hh:mm) LT. A rolling mean of the average condition from mid-day was taken over the A_t, which was used to calculate optimal values of the model parameters, as described in Mengoli et al. (2022). The value of A_t was calibrated as a parameter in our study (Table 1). We chose the mid-

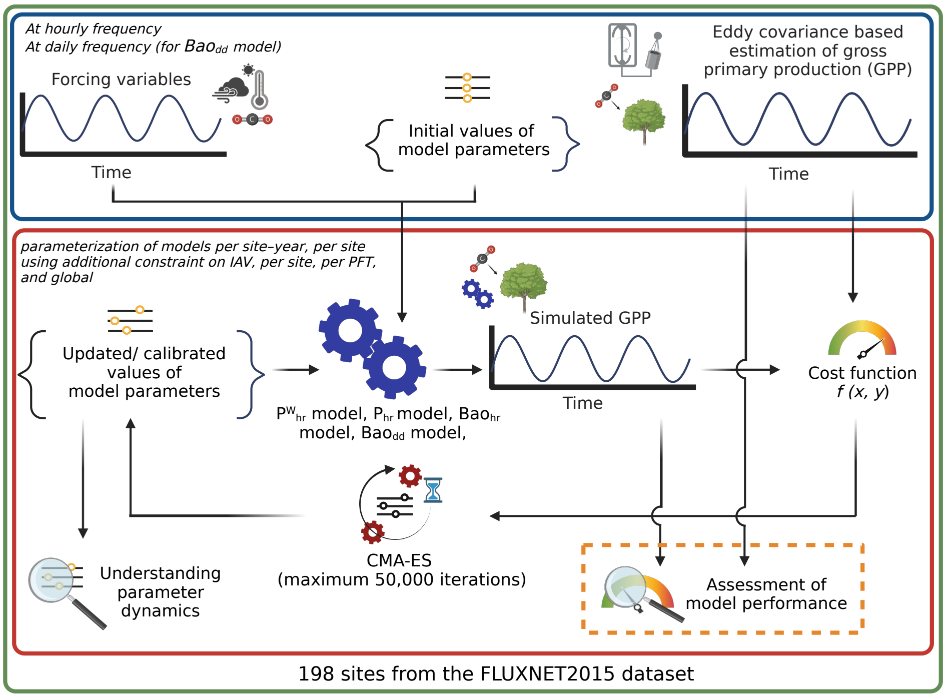

Figure 1. Graphical representation of the model–data–integration workflow adopted in this study. The blue box indicates the preparation of forcing and observation data at hourly and daily scales for each site, as well as defines the initial value of parameters and their range by surveying literature. Then five different model parameterization tasks were performed for the light use efficiency (LUE) model from Bao et al. (2022) at hourly scale ($\mathrm{Bao_{hr}}$ model) and at daily scale ($\mathrm{Bao_{dd}}$ model), P-model from Mengoli et al. (2022) at hourly scale ($\mathrm{P_{hr}}$ model), and $\mathrm{P_{hr}}$ model with an explicit drought stress function ($\mathrm{P^{W}_{hr}}$ model) using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) (Hansen & Kern, 2004), which is indicated by the red box. The cost function (f) is a function of observed (y) and simulated (x) gross primary production. The green box denotes that the whole workflow was applied for the 198 sites from the FLUXNET2015 dataset (Pastorello et al., 2020). The dotted orange box highlights the focus of this study. The parameter dynamics is explored in detail in a companion paper [companion paper citation, in prep]. The figure was created in BioRender. De, R. (2024).

day and rolling mean approach from Mengoli et al. (2022) as it produced the best results in their evaluations of the $\mathrm{P_{hr}}$ model at the half-hourly scale.

One of the known limitations of the $\mathrm{P_{hr}}$ model is its tendency to overestimate GPP fluxes in water–limited ecosystems, as no explicit representation of soil moisture conditions was included (Mengoli et al., 2022, 2023). In order to relax such drawbacks here we used the water availability index (WAI) as a proxy of soil moisture (Tramontana et al., 2016; Boese et al., 2019; Bao et al., 2022). The WAI represents the spatial and temporal dynamics in plant available water based on a simple hydrological model where storage is controlled by precipitation and evapotranspiration. We further introduced a drought stress function that additionally scaled the GPP estimates of the $\mathrm{P_{hr}}$ model, and we denoted this new version as the $\mathrm{P^{W}_{hr}}$ model. We calibrated ten parameters in the $\mathrm{P^{W}_{hr}}$ model in which nine parameters were in the hydrological model and the drought stress function (Table 1). Further details on the implementation of the $\mathrm{P^{W}_{hr}}$ model, along with the drought stress function can be found in Sect. S1.1 and S1.2.

Table 1. Description, range, initial values, and units of calibrated model parameters

fX/ model name	Symbol	Definition	Initial value	Lower bound	Upper bound	Unit	Reference
P^{W}_{hr}, P_{hr} models	A_t	Length of acclimation time	18	1	100	days	after Mengoli et al. (2022)
Bao_{hr}, Bao_{dd} models	ε_{max}	Maximum light use efficiency	0.04	0	0.4	$\mu molCO_2 \cdot \mu mol\ photons^{-1}$	Bao et al. (2022)
fT (Bao_{hr}, Bao_{dd} models)	T_{opt}	Optimal temperature	10	5	35	°C	Bao et al. (2022)
	k_T	Sensitivity to temperature changes	2	1	20	$°C^{-1}$	Bao et al. (2022)
	α_{fT}	Lag parameter for temperature effect	0.29	0	0.9	-	Bao et al. (2022)
fVPD (Bao_{hr}, Bao_{dd} models)	κ	Sensitivity to VPD changes	-5 × 10^{-5}	-0.01	-1 × 10^{-5}	Pa^{-1}	Bao et al. (2022)
	C_κ	Sensitivity to atmospheric CO_2 concentration changes	0.4	0	10	-	Bao et al. (2022)
	C_{a0}	Minimum optimal atmospheric CO_2 concentration	380	340	390	ppm	Bao et al. (2022)
	C_m	CO_2 fertilization intensity indicator	2000	100	4000	ppm	Bao et al. (2022)
fL (Bao_{hr}, Bao_{dd} models)	γ	Light saturation curvature indicator	2 × 10^{-3}	0	0.05	$\mu mol\ photons^{-1}\ m^2 \cdot s$	Bao et al. (2022)
fCI (Bao_{hr}, Bao_{dd} models)	μ	Sensitivity to cloudiness index changes	0.5	10^{-3}	1	-	Bao et al. (2022)
fW (P^{W}_{hr}, P_{hr}, Bao_{hr}, Bao_{dd} models)	W_I	Optimal soil moisture	0.26	0.01	0.99	$mm \cdot mm^{-1}$	Bao et al. (2022)
	k_W	Sensitivity to soil moisture changes	-11	-5	-30	-	Bao et al. (2022)
	α	Lag parameter for soil moisture effect	0.98	0	1	-	Bao et al. (2022)
WAI (P^{W}_{hr}, P_{hr}, Bao_{hr}, Bao_{dd} models)	AWC	Available water capacity	100	1	1000	mm	Bao et al. (2022)
	θ	Rate of evapotranspiration	0.05	10^{-4}	0.1	$mm \cdot h^{-1}$	Bao et al. (2022)
	PET_{scalar}	Multiplicative scalar for potential evapotranspiration	1.2	0	5	-	Trautmann et al. (2018)
	MR_{tair}	Snow melt rate for temperature	0.125	0	0.5	$mm \cdot° C^{-1} \cdot h^{-1}$	Trautmann et al. (2018)
	MR_{netrad}	Snow melt rate for net radiation	0.0375	0	0.125	$mm \cdot MJ^{-1} \cdot h^{-1}$	Trautmann et al. (2018)
	sn_a	Sublimation resistance	0.44	0	3	-	Bao et al. (2022)

2.1.2 Semi-empirical model: Bao model

Vegetation stores energy from absorbed solar radiation in the form of biochemical energy through photosynthesis. The efficiency of the photosynthetic apparatus in performing this energy conversion is termed as light use efficiency (ε). In a LUE model, GPP is calculated as the product of instantaneous ε, photosynthetic photon flux density ($PPFD$), and the fraction of incident photosynthetically active radiation that is absorbed by vegetation ($fAPAR$). Instantaneous ε reaches its maximum, i.e., ε_{max}, when all environmental factors are optimal for photosynthesis. Instantaneous ε is determined as the product between ε_{max} and the partial sensitivity functions (fX) for the different environmental factors controlling GPP, such as air temperature (T), VPD, available soil water supply (W), absorbed photosynthetic photon flux ($L = PPFD \times fAPAR$), the cloudiness index (CI, Table A1), and atmospheric CO_2 concentration (Mäkelä et al., 2008; Horn & Schulz, 2011; Bao et al., 2022).

$$GPP_{sim} = (\boldsymbol{\varepsilon_{max}} \cdot fT \cdot fVPD \cdot fL \cdot fCI \cdot fW) \cdot PPFD \cdot fAPAR \tag{1}$$

$$fT = \frac{2 \cdot exp\left(-\frac{T_f - \boldsymbol{T_{opt}}}{\boldsymbol{k_T}}\right)}{1 + \left(exp\left(-\frac{T_f - \boldsymbol{T_{opt}}}{\boldsymbol{k_T}}\right)\right)^2} \tag{2}$$

$$T_f(t) = (1 - \boldsymbol{\alpha_{fT}}) \cdot T(t) + \boldsymbol{\alpha_{fT}} \cdot T_f(t-1) \tag{3}$$

$$fVPD = exp\left(\boldsymbol{\kappa} \cdot \left(\frac{\boldsymbol{C_{a0}}}{CO_2}\right)^{\boldsymbol{C_\kappa}} \cdot VPD\right) \cdot \left(1 + \frac{CO_2 - \boldsymbol{C_{a0}}}{CO_2 - \boldsymbol{C_{a0}} + \boldsymbol{C_m}}\right) \tag{4}$$

$$fL = \frac{1}{\boldsymbol{\gamma}(PPFD \cdot fAPAR) + 1} \tag{5}$$

$$fCI = CI^{\boldsymbol{\mu}} \tag{6}$$

$$fW = \frac{1}{1 + exp(\boldsymbol{k_W}(W_{f_t} - \boldsymbol{W_I}))} \tag{7}$$

$$W_{f_t} = (1 - \boldsymbol{\alpha}) \cdot W_t + \boldsymbol{\alpha} \cdot W_{f_{t-1}} \tag{8}$$

In this study, we used the LUE model of Bao et al. (2022, 2023) since it emerged as a robust representation from the systematic comparison across the large diversity of LUE formulations in the literature. The model selection followed a Bayesian approach that leveraged on the evaluation of modelling performance across FLUXNET EC sites (Pastorello et al., 2020) when forced and calibrated with daily data for each site. We denoted this model as the $\mathrm{Bao_{hr}}$ model when we parameterized at hourly scale, and as the $\mathrm{Bao_{dd}}$ model when we parameterized at daily scale. The model is described in Eqs. (1) to (8), where fT, $fVPD$, fW, fL, and fCI are partial sensitivity functions for T, VPD, W, L, and CI, respectively. In this case, W and fW were calculated similar to the implementation in $\mathrm{P_{hr}^{W}}$ model, i.e., with a simple hydrological model (Sect. S1.1) and drought stress function (Eq. 7 and 8 are same as Eq. S1 and S2), respectively. Bold terms in the Eq. (1) to (8) are model parameters, and their initial values, units, and ranges are described in Table 1. The physical ranges for most of the parameters were based on Bao et al. (2022, 2023) and Trautmann et al. (2018). The $fVPD$ term, viz. Eq. (4), also accounts for atmospheric CO_2 concentration. The partial sensitivity functions range from zero to one (except the 2^{nd} part of Eq. 4 which can be greater than one), where a value of zero completely diminishes, and of one completely favours GPP. In this study, we changed the denominator of Eq. (2) in comparison to the original exponential function $exp\left(-\frac{T_f - T_{opt}}{k_T}\right)^2$ of Bao et al. (2022, 2023), as the revised version produced a more realistic range of fT (Fig. S3). Sensitivity functions fT and fW also consider a lag effect of T and W. The lag effect of temperature was considered for Temperate, Boreal, and Polar regions where the first letter of the Köppen–Geiger (KG) climate class is ‘C’, ‘D’, ‘E’, and that of soil water supply was considered for arid regions where the first letter of the KG climate class is 'B' (Rubel et al., 2017; Beck et al., 2018).

2.2 Data used

We selected 198 eddy covariance sites, for which the required forcing and observation or derived data for model parameterization were available from the FLUXNET2015 dataset (Pastorello et al., 2020; FLUXNET.org, 2024a). A list of these sites can be found in Table S2 (Sect. S3) and their spatial distributions is plotted in Fig. S4. The variables which were used to force, and parameterize models as well as data processing steps such as gap–filling, and quality control are described in detail in Table A1, Appendix B and Appendix C. We prepared these data in both hourly and daily resolutions.

In our study, a total of 13 different PFTs (as defined in FLUXNET.org, 2024b) were represented: croplands (CRO; 19 sites), deciduous broadleaf forests (DBF; 25 sites), deciduous needle leaf forest (DNF; one site), evergreen broadleaf forests (EBF; 13 sites), evergreen needle leaf forests (ENF; 47 sites), grasslands (GRA; 35 sites), mixed forests (MF; nine sites), closed shrublands (CSH; three sites), open shrublands (OSH; 13 sites), savannas (SAV; six sites), permanent wetlands (WET; 20 sites), woody savannas (WSA; six sites), and land cover under snow for most of the year (SNO; one site). The major KG climate classes (Rubel et al., 2017; Beck et al., 2018; FLUXNET.org, 2024c) are represented by 12 tropical sites, 18 arid sites, 87 temperate sites, 71 boreal sites, and 10 polar sites. We also classified sites into 9 climate–vegetation types, similar to Bao et al. (2022), in which seven sites are tropical forests (TropicalF), five sites are tropical grassland (TropicalG), six sites are arid forest (AridF), 12 sites are arid grassland (AridG), 51 sites are temperate forest (TemperateF), 36 sites are temperate grassland (TemperateG), 52 sites are boreal forest (BorealF), 19 sites are boreal grassland (BorealG), and 10 sites have polar vegetation.

2.3 Model parameterization

We primarily defined four different parameterization strategies consisting of various subsets of data to calibrate the model parameters controlling hourly GPP dynamics. These parameterization strategies were used to determine a vector of calibrated parameter values (1) for each site–year, (2) for each site, (3) for each PFT, and (4) for all sites at once (global parameterization). We also performed another parameterization per site using a modified cost function which used an additional constraint on the IAV of GPP ($Cost^{IAV}$). We parameterized and forced the $\mathrm{Bao_{hr}}$ model, and the $\mathrm{Bao_{dd}}$ model using hourly and daily data, respectively to perform a comparative analysis (Table 2). The $\mathrm{P_{hr}^{W}}$ model and the $\mathrm{P_{hr}}$ model were only parameterized and forced using hourly data (Table 2).

We used Python (Python Core Team, 2021) implementation (pycma v3.3.0.1) of the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) (Hansen & Kern, 2004; Hansen et al., 2019) as our global search algorithm to find the values of model parameters for which cost function reached its minimum. This is a derivative-free, evolutionary algorithm, which is designed to find global minima in a rugged parameter space.

$$Cost_{i_P} = (1 - GPP_{NNSE_i}) + (1 - ET_{NNSE_i}) \tag{9}$$

$$Cost_{i_{Bao}} = (1 - GPP_{NNSE_i}) + (1 - ET_{NNSE_i}) + Cost_{ideal} + Cost_{non\_ideal} \tag{10}$$

A robust cost function is a necessity for the numerical optimizer to find the global minimum. The cost functions for $\mathrm{P_{hr}^{W}}$, $\mathrm{P_{hr}}$ models ($Cost_{i_P}$) and the $\mathrm{Bao_{hr}}$, $\mathrm{Bao_{dd}}$ models ($Cost_{i_{Bao}}$) were calculated as Eq. (9) and (10), respectively, in case of per site–year

Table 2. Description of models and tasks accomplished with each specific model. The tasks are described in the footnote of the table.

Models	**Description**	**Parameterization strategies**				
		per site–year	**per site using Cost$^{\mathbf{IAV}}$**	**per site**	**per PFT**	**global**
P_{hr}	P-model of Mengoli et al. (2022) parameterized using hourly data	a, d	a, d	a, d	a, d	a, d
P_{hr}^{W}	P-model of Mengoli et al. (2022) with an additional constraint on drought stress and parameterized using hourly data	a, b, d, e	a, b, d, e	a, b, d, e	a, b, d, e	a, b, d, e
Bao_{hr}	LUE model of Bao et al. (2022) parameterized using hourly data	a, c, d, e	a, c, d, e	a, c, d, e	a, c, d, e	a, c, d, e
Bao_{dd}	LUE model of Bao et al. (2022) parameterized using daily data	a, c, d	a, c, d	a, c, d	a, c, d	a, c, d

a: evaluation of model performance across timescale with different model types, parameterization strategies, and cost functions.
b: evaluation of a mechanistic model with an explicit drought stress function.
c: evaluation of a semi-empirical model with different temporal resolutions of data used for model parameterization.
d: factors behind variability of model performance across timescales.
e: variability of annual model performance with model performance in simulating diurnal gross primary productivity (GPP) peaks.

and per–site parameterization. Here, i is either a site or site–year based on parameterization type. For PFT-specific model parameterization, the cost functions were $\sum_{i=1}^{N_{PFT}} Cost_{i_P}$ and $\sum_{i=1}^{N_{PFT}} Cost_{i_{Bao}}$ for P_{hr}^{W}, P_{hr} models and Bao_{hr}, Bao_{dd} models, respectively. i denotes a site and N_{PFT} denotes the total number of sites in a specific PFT. In the case of global model parameterization, the cost functions were $\sum_{i=1}^{N} Cost_{i_P}$ and $\sum_{i=1}^{N} Cost_{i_{Bao}}$ for the P_{hr}^{W} model and Bao_{hr}, Bao_{dd} models, respectively. i denotes a site and N denotes the total number of sites used in this study.

$$NNSE_i = \frac{1}{2 - NSE_i} \tag{11}$$

$$NSE_i = 1 - \frac{\sum_{t=1}^{N_{t,i}} \left(\sigma_{weight_{t,i}} \cdot (EC_{t,i} - sim_{t,i})\right)^2}{\sum_{t=1}^{N_{t,i}} \left(\sigma_{weight_{t,i}} \cdot (EC_{t,i} - \overline{EC_{t,i}})\right)^2} \tag{12}$$

$$\sigma_{weight_{t,i}} = 1 - \frac{\sigma_{t,i} - min(\sigma_i)}{max(\sigma_i) - min(\sigma_i)} \tag{13}$$

GPP_{NNSE_i} and ET_{NNSE_i} were calculated (Eq. 11) as a weighted normalized NSE, viz. NNSE (Hundecha & Merz, 2012) between the time series of good quality data points (see Appendix B for the selection criteria) of EC derived and simulated GPP and ET, respectively. The GPP and ET derived from EC measurements are denoted as GPP_{EC} and ET_{LE}, respectively. The simulated GPP and ET are denoted as GPP_{sim} and ET_{sim} (see Fig. S1 for calculation of ET_{sim}), respectively. We considered ET as well in our cost function to better constrain the parameters of the simple hydrological model used in this study. The NNSE values (Nash & Sutcliffe, 1970) are between zero and one, where one is the best, and zero is the worst agreement between observed and simulated data. Here, we used these normalized values so that minimizing $(1-NNSE)$ always results in better model performance in comparison to using $(1-NSE)$, where NSE can have values between $-\infty$ (worst agreement) and one (best agreement). In Eq. (12), $N_{t,i}$ is the total number of good quality data points from each timestep t for a site–year or site i. σ in Eq. (13) is random uncertainty (which is the standard deviation of fluxes in a sliding window of ±5 days and ±1hour of the time-of-day of the current timestamp) of NEE or ET (Table A1).

$$\begin{aligned} Cost_{ideal} = &((1 - max(fT_r)) + (1 - max(fVPD_r)) + (1 - max(fW_r)) \\ &+ (1 - max(fL_r))) \cdot 10^3 \end{aligned} \tag{14}$$

$$\begin{aligned} Cost_{non\_ideal} = &\sum_r ((fT_r - \theta_{fT})(T < 0°C \;\&\; fT_r > \theta_{fT})) \\ &+ \sum_r ((fVPD_r - \theta_{fVPD})(VPD > 2000Pa \;\&\; fVPD_r > \theta_{fVPD})) \\ &+ \sum_r ((fW_r - \theta_{fW})(W < 0.01 \;\&\; fW_r > \theta_{fW}) \end{aligned} \tag{15}$$

The $Cost_{ideal}$ and $Cost_{non\_ideal}$ were introduced as a regularizers in $Cost_{i_{Bao}}$ to avoid over-fitting of the sensitivity functions (Bao et al., 2022, 2023). These cost function components ensure that values of partial sensitivity functions were not penalized and favoured under ideal and non-ideal conditions, respectively. The ideal and non-ideal conditions were determined by certain constant thresholds for all sites. Equation (14) ensured that the partial sensitivity functions, fT (Eq. 2), only left part of the $fVPD$ (Eq. 4), fW (Eq. 7) and fL (Eq. 5) approaches one, when certain ideal environmental conditions ($PPFD$ $\in$ [0 to 600 μmol photons $\cdot\, m^{-2} \cdot s^{-1}$], $fAPAR \in$ [0 to 1], $T \in$ [-5 to 40 °C], $VPD \in$ [0 to 4500 Pa], $W \in$ [0 to 1]) occur (these ranges are denoted by subscript r), so that the ε_{max} in Eq. (1) reaches its maximum potential. The factor 10^3 in Eq. (14) was included to match the ranges of all other components in the cost function for the Bao_{hr}, Bao_{dd} models ($Cost_{i_{Bao}}$) so that all the components had equal weight. Equation (15) penalized the cases when the values of fT (Eq. 2), only left part of $fVPD$ (Eq. 4), and fW (Eq. 7), were greater than a certain threshold ($\theta_{fT} = 0.2$, $\theta_{fVPD} = 0.9$, $\theta_{fW} = 0.2$) under non-ideal conditions ($T < 0$ °C, $VPD > 2000$ Pa, $W < 0.01$) for photosynthesis.

$$Cost_{i_P}^{IAV} = (1 - GPP_{NNSE_i}) + (1 - GPP_{NNSE_i}^{y}) + (1 - ET_{NNSE_i}) \tag{16}$$

$$Cost_{i_{Bao}}^{IAV} = (1 - GPP_{NNSE_i}) + (1 - GPP_{NNSE_i}^{y}) + (1 - ET_{NNSE_i}) + Cost_{ideal} + Cost_{non\_ideal} \tag{17}$$

$$GPP_{NNSE_i}^{y} = \frac{1}{2 - GPP_{NSE_i}^{y}} \tag{18}$$

$$GPP_{NSE_i}^{y} = 1 - \frac{\sum_{t=1}^{N_{t,i}} \left(\sigma_{weight_{t,i}}^{y} \cdot (EC_{t,i}^{y} - sim_{t,i}^{y})\right)^2}{\sum_{t=1}^{N_{t,i}} \left(\sigma_{weight_{t,i}}^{y} \cdot (EC_{t,i}^{y} - \overline{EC_{t,i}^{y}})\right)^2} \tag{19}$$

$$\sigma_{weight_{t,i}}^{y} = 1 - \frac{\sigma_{t,i}^{y} - min(\sigma_i^{y})}{max(\sigma_i^{y}) - min(\sigma_i^{y})} \tag{20}$$

$$EC_{t,i}^{y} = \sum_{t=1}^{t} EC_{t,y,i};\ sim_{t,i}^{y} = \sum_{t=1}^{t} sim_{t,y,i};\ \sigma_{t,i}^{y} = \sum_{t=1}^{t} \sigma_{t,y,i} \tag{21}$$

In the case of per–site–year parameterization using cost functions in Eq. (9) and (10), we fitted the model so that the annual average of GPP can be captured well for each site–year. Whereas, in the case of per–site parameterization using cost functions in Eq. (9) and (10), the model was parameterized for each site. We performed another experiment as a balance between these two experiments using the $Cost^{IAV}$, which is similar to Desai (2010) to put an additional constraint on IAV, and parameterized $\mathrm{P_{hr}^{W}}$, $\mathrm{P_{hr}}$, $\mathrm{Bao_{hr}}$, and $\mathrm{Bao_{dd}}$ models for each of the EC sites. The cost functions, $Cost_{i_P}^{IAV}$ for $\mathrm{P_{hr}^{W}}$, $\mathrm{P_{hr}}$ models (Eq. 16) and $Cost_{i_{Bao}}^{IAV}$ for $\mathrm{Bao_{hr}}$, and $\mathrm{Bao_{dd}}$ models (Eq. 17) now include an additional term $(1 - GPP_{NNSE_i}^{y})$ to constrain the annual cumulative sum of GPP flux from each site i. $EC_{t,i}^{y}$, $sim_{t,i}^{y}$, and $\sigma_{t,i}^{y}$ (Eq. 21) are cumulative sums of GPP_{EC}, GPP_{sim}, and σ_{NEE} from start of each year y to timestep t for each site i, respectively.

2.4 Simulating and evaluating GPP estimates

2.4.1 Forward runs

In the case of the site–year parameterization, we performed a forward run for each site–year using the respective set of calibrated parameter values and forcing data for that year. Afterwards, we concatenated GPP_{sim} from all the years for a given site to assess model performance. For per–site parameterization using $Cost^{IAV}$, and per–site parameterization, we used site-specific values of calibrated parameters to perform site-level model evaluation. We also applied calibrated model parameters for a certain PFT to simulate GPP at all the sites which belong to a certain PFT. Similarly, for the global parameterization, a single set of calibrated parameter values was used to simulate GPP for each site.

2.4.2 Model performance metrics

We performed forward runs at an hourly scale and averaged the hourly simulations to daily, weekly, monthly, and annual temporal frequencies to calculate model performance measures at different temporal aggregations. Model performance was only evaluated for temporal aggregations from daily to annual for the $\mathrm{Bao_{dd}}$ model. We applied a data screening procedure (Appendix C) before calculating model performance measures. We evaluated how well a model can simulate the IAV of GPP based on how well a model simulated the annual average GPP for a site. In this study, we performed most of our analysis using NSE (Nash & Sutcliffe, 1970) and normalized NSE, viz. NNSE (which is $\frac{1}{2-\mathrm{NSE}}$) as NSE indicates the degree to which scatter between observed and simulated data fits to the 1:1 line. In addition, we calculated the square of the Pearson correlation coefficient (R^2) (PCC, 2008) which explains whether the dispersion of observed and simulated data matches and in the case of an unbiased model, values of NSE will be closer to values of R^2. Whereas, if a model is systematically biased, it will result in higher R^2 values, but bad NSE values (Krause et al., 2005). We also calculated Root Mean Squared Error (RMSE) (Chai & Draxler, 2014) to quantify how closely the mean of simulated data matches with the mean of the observed data.

$$NSE = 2 \cdot \alpha_{NSE} \cdot r - \alpha_{NSE}^2 - \beta_n^2 \tag{22}$$

$$\alpha_{NSE} = \frac{\sigma_{sim}}{\sigma_{EC}} \tag{23}$$

$$\beta_n = \frac{\mu_{sim} - \mu_{EC}}{\sigma_{EC}} \tag{24}$$

Moreover, using Eq. (22) to (24), we decomposed NSE values to linear correlation (r), relative variability (α_{NSE}), and bias (β_n) in some cases to investigate which of these were improved or diminished between different model parameterization strategies (Gupta et al., 2009). In Eq. (23), and (24), σ_{sim} and σ_{EC} are standard deviations of GPP_{sim} and GPP_{EC}, respectively, μ_{sim} and μ_{EC} are mean GPP_{sim} and GPP_{EC}, respectively. We calculated these metrics using the Python (Python Core Team, 2021) package Permetrics v1.5.0 (Van Thieu, 2023; Van Thieu & Mirjalili, 2023), and the definition of each of the model performance metrics can be found in the package documentation.

2.4.3 Factors associated with simulating GPP flux

We selected potential factors that can affect model performance at different temporal resolutions. These factors can be of two types. There were factors which we determined based on our experiment design, which included model types (P_{hr}^{W} model, P_{hr} model, Bao_{hr} model, and Bao_{dd} model), parameterization strategies (per site–year, per site using $Cost^{IAV}$, per site, per PFT, and global parameterization), number of years with good quality data (Appendix C) in a site. Whereas, other factors represent site-specific characteristics, including PFT, KG climate class, and climate–vegetation types.

First, we conducted Levene's test (Levene, 1960) to find out if the assumption of homoscedasticity is fulfilled across groups in the controlling factors. Then, we performed an N-way Analysis of Variance (ANOVA) (Kaufmann & Schering, 2014) with the potential controlling factors to determine which of them played a major role in determining model performance at hourly and annual temporal scales. For analysis at an hourly scale, the Bao_{dd} model was not included as this model produced simulations at a daily scale. We performed two N-way ANOVA analyses once including the performance of the P_{hr} model, and then excluding the performance of the P_{hr} model. The Levene's test and N-way ANOVA analyses were implemented using SciPy v1.11.3 (Virtanen et al., 2020) and statsmodels v0.14.0 (Seabold & Perktold, 2010), respectively.

2.4.4 Evaluating GPP estimates in water–limited ecosystems

We investigated to determine whether explicit accounting of the drought stress function in the P_{hr}^{W} model had improved its performance at arid sites. For this purpose, we chose the aridity index (AI) to determine which sites were arid or semi-arid, as this index provided a numerical representation of moisture availability (Zomer et al., 2022) at a location. The AI values were calculated by dividing the average precipitation (P) per hour by the average potential evapotranspiration (PET) per hour for the whole observation period at a site.

We drew examples from a few site-specific results to highlight different aspects of the behaviour of P_{hr}^{W} and P_{hr} models for ecosystems with contrasting soil moisture controls on GPP and with a larger availability of good–quality measurements. For this purpose, we chose a water–limited semi-arid site (annual average precipitation of 318 mm) in central Australia (Alice Springs, AU-ASM). This site also features a complex mixture of Mulga woodland and savanna (Cleverly et al., 2013; Pastorello et al., 2020). In contrast, we also highlighted the behaviours of $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ and P_{hr} models in an irrigated cropland (Mead - irrigated continuous maize site, US-Ne1) in the mid-western U.S.A (Amos et al., 2005; Pastorello et al., 2020).

2.4.5 Effect of temporal resolution of data used for model parameterization on model performance

We parameterized the LUE model of Bao et al. (2022) with hourly and daily data for $\mathrm{Bao}_{\mathrm{hr}}$ model and $\mathrm{Bao}_{\mathrm{dd}}$ model, respectively. We performed a comparison between these two versions of the model to highlight whether the resolution of data used for model parameterization can substantially affect the prediction of the annual average or IAV of GPP fluxes. Here, we also drew a site-specific example from an energy–limited deciduous forest in central Germany (Hainich, DE-Hai) as this site had a very long observation period (Knohl et al., 2003).

2.4.6 Evaluation between modelling experiments of various complexities

We formulated our experiments using models and parameterization strategies consisting of varying numbers of model parameters to be calibrated. The number of parameters calibrated for a detailed parameterization strategy, such as per site–year parameterization was substantially higher than a generic parameterization strategy, such as global parameterization. We used Akaike's Information Criterion (AIC) to investigate whether a complex modelling experiment with a higher number of parameters can better simulate GPP (Burnham & Anderson, 2004).

$$AIC = n\ log\left(\frac{\sum(EC_i - sim_i)^2}{n}\right) + 2K \tag{25}$$

$$AIC_c = n\ log\left(\frac{\sum(EC_i - sim_i)^2}{n}\right) + 2K + \frac{2K(K+1)}{n-K-1} \tag{26}$$

Following recommendations of Burnham and Anderson (2004), we used Eq. (25) to calculate AIC when $n/K > 40$, where n is the total number of observations and K is the total number of parameters. Otherwise, we used a corrected version of AIC ($\mathrm{AIC_c}$, Eq. 26). Though the values of AIC or $\mathrm{AIC_c}$ can be in any range, the lowest value of AIC or $\mathrm{AIC_c}$ determines the preferred modelling experiments. EC_i and sim_i are i^{th} observations of EC-derived GPP and simulated GPP, respectively in Eq. (25) and (26). We considered GPP_{sim} from all the four variations of models, i.e., $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model, P_{hr} model, $\mathrm{Bao}_{\mathrm{hr}}$ model, and $\mathrm{Bao}_{\mathrm{dd}}$ model for calculation of AIC or AIC_c. We calculated AIC at hourly and daily aggregations by concatenating good quality (Appendix C) hourly or daily data, and daily averages GPP_{EC} and GPP_{sim} from all the days from all sites. Similarly, we used monthly and annual aggregations for calculating AIC_c at monthly and annual scales, respectively. AIC_c was calculated at monthly and annual aggregation, as n was usually smaller than K in these cases. The value of K was the total number of model parameters calibrated for all the site–years, for all the sites, for all the PFT, and for a specific model in case of per site–year parameterization, per site parameterization using $Cost^{IAV}$, per site parameterization, per PFT parameterization, and global parameterization, respectively.

2.4.7 Simulating GPP peaks

We assessed model performance in predicting peak GPP_{EC}. We defined peak GPP_{EC} and peak GPP_{sim} as the 90$^{\text{th}}$ percentiles of hourly GPP_{EC} ($P90_{GPP_{EC}}$) and GPP_{sim} ($P90_{GPP_{sim}}$), respectively, following the concept of good hours by Zscheischler et al. (2016) and Fatichi and Ivanov (2014). We calculated $P90_{GPP_{EC}}$ and $P90_{GPP_{sim}}$ for each site–year considering only good quality hourly data (Appendix B). We compared the ratios of peak GPP_{sim} from $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model and $\mathrm{Bao}_{\mathrm{hr}}$ model to GPP_{EC} for each parameterization strategy in order to identify possible biases.

$$\Delta NNSE_{P90} = NNSE_{P90}^{j1} - NNSE_{P90}^{j2} \tag{27}$$

$$\Delta NNSE_{y} = NNSE_{y}^{j1} - NNSE_{y}^{j2} \tag{28}$$

We furthermore investigated whether improving the simulation of peaks of GPP_{EC} improved the simulation of IAV of GPP. We calculated NNSE between $P90_{GPP_{EC}}$ and $P90_{GPP_{sim}}$ ($NNSE_{P90}^{j}$) from all the site–years in a site considering only good site–years and only for sites with more than 3 years of good quality data (Appendix C) for a parameterization strategy j. Similarly, we calculated NNSE between the annual average of GPP_{EC} and GPP_{sim} ($NNSE_{y}^{j}$) for sites with more than 3 years of good quality data (Appendix C) for a parameterization strategy j. Then, differences between $NNSE_{P90}^{j1}$ and $NNSE_{y}^{j2}$ were calculated for a pair of parameterization strategies where $j1$ and $j2$ are two different parameterization experiments, for both $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model and $\mathrm{Bao}_{\mathrm{hr}}$ model (Eq. 27 and 28). Correlation between $\Delta NNSE_{P90}$ and $\Delta NNSE_{y}$ were then investigated to study whether a certain parameterization strategy for a given model better captured the GPP_{EC} peaks, and thus contributed to higher annual model performance.

3 Results

3.1 Overall model performance

All four models, i.e., $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$, P_{hr}, $\mathrm{Bao}_{\mathrm{hr}}$, and $\mathrm{Bao}_{\mathrm{dd}}$ models performed significantly better at the hourly scale than the annual scale (Fig. 2). The use of an additional constraint on IAV, i.e., $Cost^{IAV}$ did not contribute to better model performance across sites at an annual scale and performed closer to parameterization per site and poorer than site–year parameterization (Fig. 2). The median model performance was highest for the model parameterization per site–year among all model parameterization strategies (per site–year, per site using $Cost^{IAV}$, per site, per PFT, and global parameterization) for all four models (Table D1). Model parameterization per site–year also produced the best model performance at all temporal aggregation levels including annual aggregation (Fig. 2 and Sect. S2.1). $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model performed substantially better for the majority of the sites compared to P_{hr} model at all temporal aggregation levels as it explicitly considered site-specific water availability (Fig. 2 and Sect. S2.1). Comparison of model performances at different temporal aggregations also revealed that $\mathrm{Bao}_{\mathrm{hr}}$ and $\mathrm{Bao}_{\mathrm{dd}}$ models performed slightly better than the $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model across all timescales (hourly, daily, weekly, monthly, and annual), as the $\mathrm{Bao}_{\mathrm{hr}}$ and $\mathrm{Bao}_{\mathrm{dd}}$ models were more flexible than the $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model and captured ecosystem response with a broad range of parameters (Fig. 2, Table D1 and Sect. S2.1). For example, the median NNSE(s) at the hourly resolution were 0.827 and 0.853 for the $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model and the $\mathrm{Bao}_{\mathrm{hr}}$ model, respectively. Conversely, at the annual resolution, the median NNSE(s) were 0.543 and 0.661 for the $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model and $\mathrm{Bao}_{\mathrm{hr}}$ model, respectively.

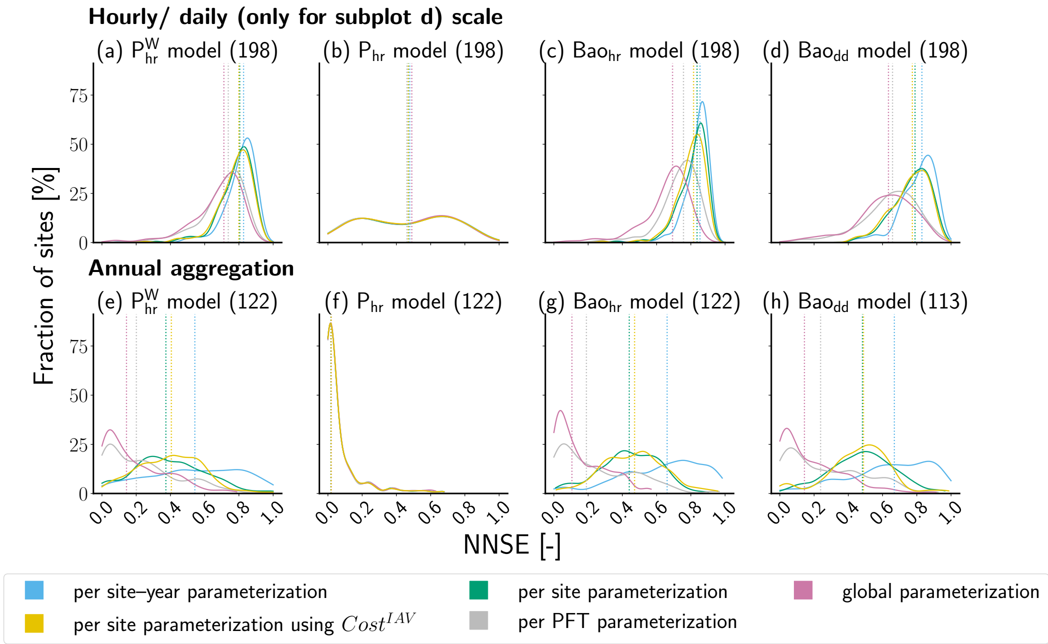

Figure 2. Distributions of model performance measure (normalized Nash-Sutcliffe efficiency, viz. NNSE) at hourly/daily scale (first row) and at annual timescale (second row) from P-model of Mengoli et al. (2022) with drought stress, parameterized at hourly scale (P^{W}_{hr}), P-model of Mengoli et al. (2022) without drought stress, parameterized at hourly scale (P_{hr}), global best model of Bao et al. (2022) parameterized at hourly scale (Bao_{hr}), and global best model of Bao et al. (2022) parameterized at daily scale (Bao_{dd}). For the Bao_{dd} model, subplot (d) shows model performance at daily scale as this model was parameterized at daily scale. $Cost^{IAV}$ denotes the usage of an additional constraint on annual gross primary production flux during per–site parameterization. The dotted vertical lines represent the median model performances, which are summarized in Table D1. The numbers in parentheses beside the model name on top of each of the sub-figures represent the total number of sites. The model performance at an annual scale was calculated for fewer sites as some sites have a very low measurement period (Appendix C).

3.2 Factors behind variability in model performance

We summarized the percentage contributions of factors which influenced model performance at hourly and annual scales and found that most of the variability in model performance came from how we designed our modelling experiments (Fig. 3). Model types was a crucial factor when the P_{hr} model was included in N-way ANOVA analysis, as this model had comparatively poor performance at both hourly and annual scales and resulted in greater variability in the NNSE values (69% and 60% contribution to the sum of squares of the regression, viz. SSR in hourly and annual scale, respectively). We then excluded the P_{hr} model from further analysis to uncover the other factors behind the model performance and found that for the hourly scale, the model performance varied the most across the groups of KG classes (31.4% contribution to the SSR), followed by parameterization type (31.0% to the SSR) and climate–vegetation type (23.2% contribution to the SSR) (Fig. 3). However, at an annual scale, the parameterization strategy strongly affected (71.7% contribution to the SSR) the model performance, as per–site–year parameterization usually better simulated the annual GPP_{obs} compared to other parameterization strategies. The number of good years (Appendix C) used for calculating annual NNSE also exerted a small influence (3.8% contribution to the SSR) on the annual model performance. In general, there were only slight performance differences between models when the P_{hr} model was not considered, and model parameterization played a bigger role in the variability of model performance.

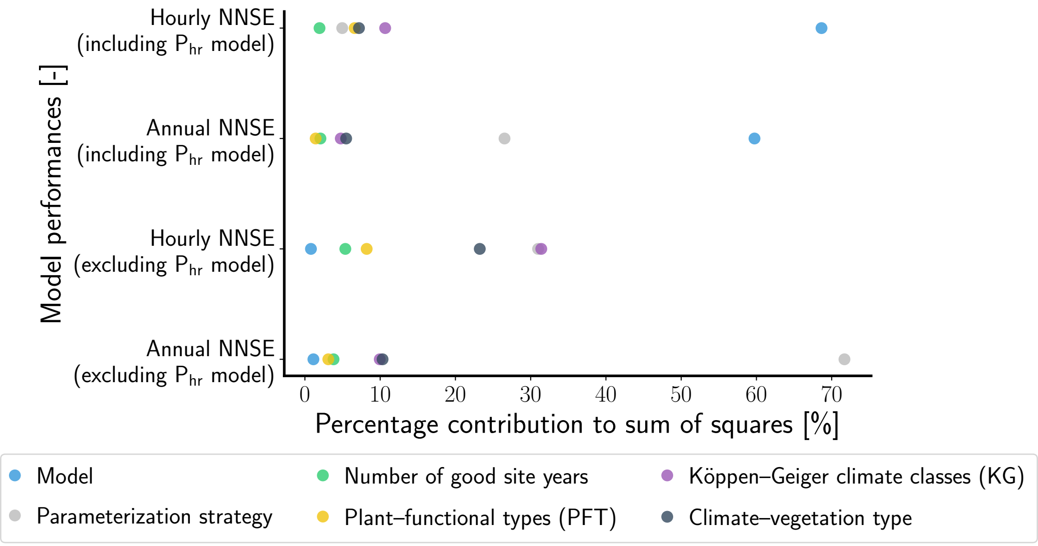

Figure 3. Percentage contributions of factors influencing variability in model performance (normalized Nash-Sutcliffe efficiency, viz. NNSE) in the sum of squares in N-way Analysis of Variance (ANOVA). The percentage contributions show the influence of various factors on hourly and annual model performance when the P-model (Mengoli et al., 2022) without any explicit drought stress function, parameterized at hourly scale (P_{hr} model) was considered in the analysis, as well as on hourly and annual model performance excluding the P_{hr} model. The sum of squares of residual was removed before plotting the percentage contributions of the factors and only the explained variance is shown.

3.3 Effect of drought stress on model performance

The performance of the P_{hr}^{W} model to predict the annual average of GPP_{EC} from each of the site–years substantially improved in comparison to P_{hr} model (from an NSE of -0.94 to 0.93) after explicit consideration of soil water supply in the model (Fig. 4). Most of this improvement came from the better prediction of the annual average of GPP_{EC} at the arid and semi-arid sites (with AI values lower than 0.5). For the semi-arid site (AU-ASM) the predicting performance of the P_{hr}^{W} model for all the parameterization strategies largely benefited from the explicit inclusion of soil water supply constraints (Fig. 5). The systematic bias in model simulations was also improved after the inclusion of a drought stress constraint, as well as the modelling bias also improved from a generalized to a detailed parameterization strategy. Although the coupling of a simple hydrological model which calculated water–availability based on precipitation and evapotranspiration and inclusion of drought stress function generally improved the P_{hr}^{W} model for most of the site–years at an arid site, the model failed to capture the GPP_{EC} (Sect. S2.2, Fig. S6) at an irrigated cropland site (US-Ne1), as the simple hydrological model which we used to calculate water–availability lacked representation of human management.

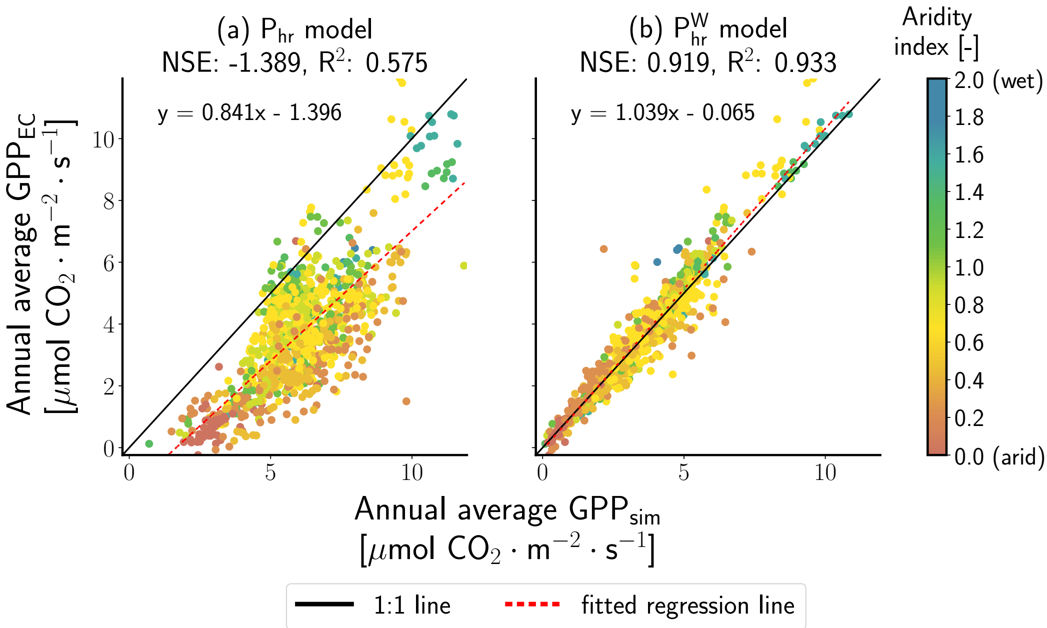

Figure 4. Scatter plot of the annual average (from good quality site–years, see Appendix C) of eddy covariance measurements derived gross primary production (GPP_{EC}) versus simulated gross primary production (GPP_{sim}) from P-model of Mengoli et al. (2022) parameterized at hourly scale **(a)** without drought stress (P_{hr} model) and **(b)** with drought stress (P_{hr}^W model). The results in this plot are from parameterization for each site–year. We only used good–quality site–years in this figure (Appendix C). The dots in the scatter represent a site–year and are coloured by the aridity index (AI) of the site. The model performance metrics (Nash-Sutcliffe efficiency, viz. NSE) are shown at the top of each subplot. The equations of fitted regression lines are shown in respective subplots.

3.4 Effect of temporal resolution of the data used for model parameterization on model performance

The use of hourly data to constrain Bao_{hr} model parameters and aggregating hourly values of GPP_{sim} to annual scale did not have a significant effect on the Bao_{hr} model performance in comparison to parameterization of the same model with daily data, i.e., Bao_{dd} model, in simulating the annual average of GPP_{EC} (Fig. 6) for each site–year. The value of NSE slightly decreased from 0.964 to 0.956 for the Bao_{hr} model compared to the Bao_{dd} model, and both models performed almost equally well. Here, for the Bao_{hr} model, we also focus on a site-specific example at a site (DE-Hai) in central Germany with a deciduous broadleaf forest where the Bao_{hr} model proved to be capable of simulating annual average of GPP_{EC} flux relatively well when the model was parameterized for each site–year and each site (Fig. 7). However, GPP_{EC} was underestimated in cases of PFT-specific and global parameterization. For this specific site, the Bao_{hr} model performed relatively better in comparison to Bao_{dd} model (Fig. 7 and S7).

3.5 Role of parameterization strategies on model performance

Model performances at an annual scale generally increased with a more detailed parameterization strategy (Fig. 8). For the P_{hr}^W model, the median differences in annual model performance between the most detailed parameterization strategy, i.e., site–year

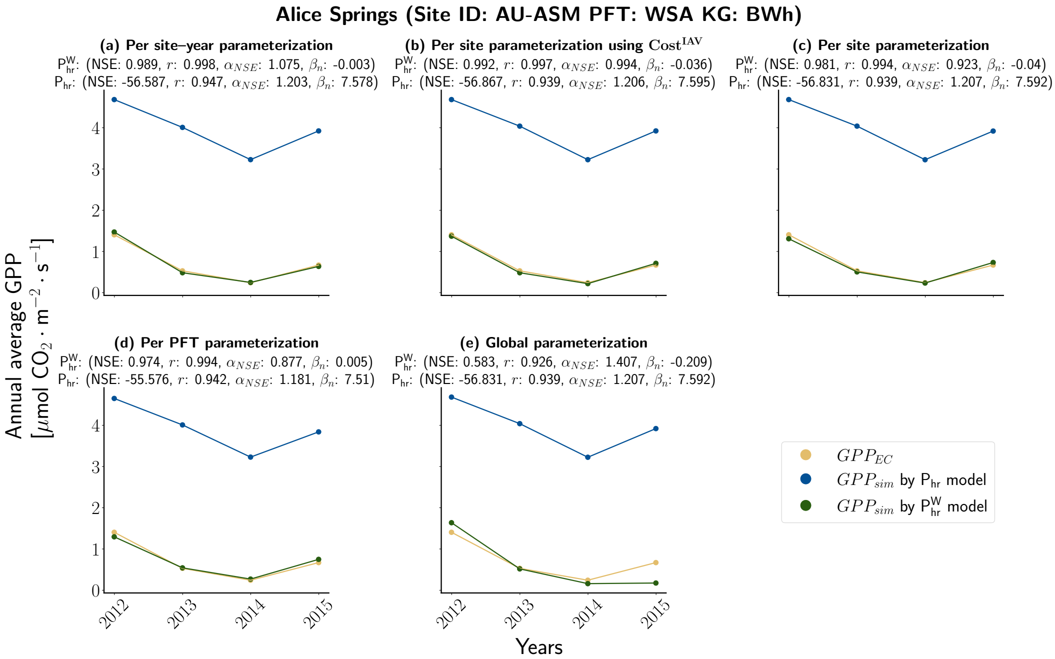

Figure 5. Comparison of annual average of gross primary production (GPP) derived by eddy covariance measurements (GPP_{EC}), and GPP simulated (GPP_{sim}) by the P-model of Mengoli et al. (2022) parameterized at hourly scale without drought stress (P_{hr} model) and with drought stress (P_{hr}^{W} model). The five subplots show simulated GPP from **(a)** site–year specific parameterization, **(b)** site-specific parameterization using an additional constraint on inter–annual variability in the cost function ($Cost^{IAV}$), **(c)** site-specific parameterization, **(d)** plant–functional types (PFT) specific parameterization, and **(e)** global parameterization. The values of model performance measures (Nash-Sutcliffe efficiency, viz. NSE, correlation coefficient, viz. r, relative variability, viz. α_{NSE}, and bias, viz. β_n) are shown on top of respective subplots. This site is dominated by Mulga (*Acacia aneura*), and had an annual average temperature of ≈22 °C, and an annual average precipitation of ≈318 mm during the observation period (Cleverly et al., 2013; Pastorello et al., 2020). The site ID, PFT, and Köppen–Geiger climate class (KG) of the site are provided on top of the figure in bold.

parameterization, and other detailed parameterization strategies, i.e., per site parameterization using $Cost^{IAV}$ and per site parameterization were small, which were 0.12, and 0.11, respectively. In contrast, the median differences in annual model performance between the most detailed parameterization strategy, i.e., site–year parameterization, and other generalized parameterization strategies, i.e., PFT-specific parameterization, and global parameterization were quite large, which were 0.28, and 0.37, respectively. Similarly, for the Bao_{hr} model, the median differences in annual model performance between the most detailed parameterization strategy, i.e., site–year parameterization, and other detailed parameterization strategies, i.e., per site parameterization using $Cost^{IAV}$ and per site parameterization were 0.20, and 0.21, respectively. In contrast, the median differences in annual model performance between the most detailed parameterization strategy, i.e., site–year parameterization, and other generalized parameterization strategies, i.e., PFT-specific parameterization, and global parameterization were 0.36, and 0.50, respectively. The positive values of median annual model performance confirm the high-

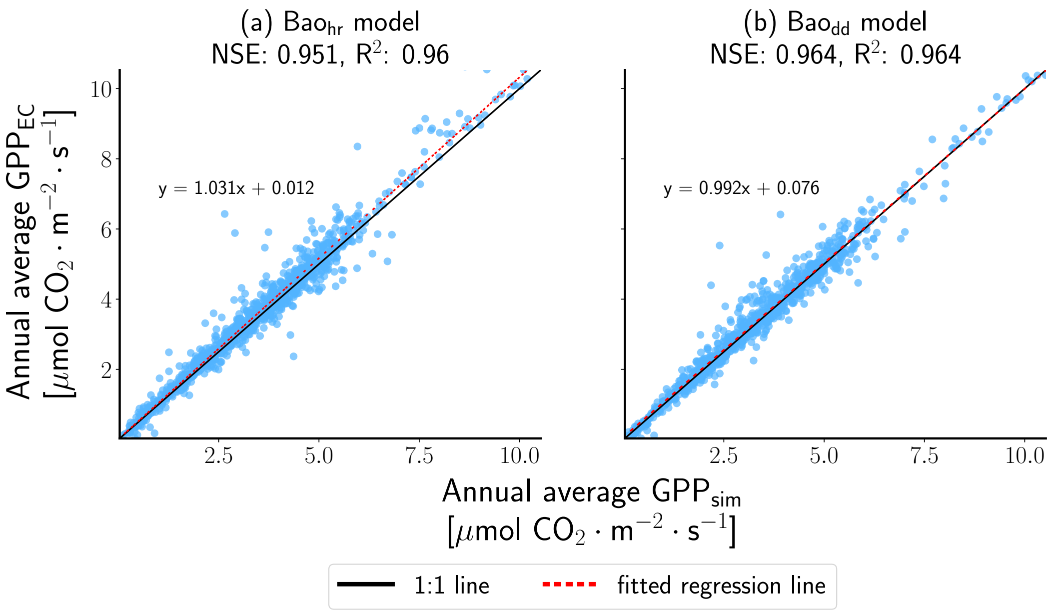

Figure 6. Scatter plot of annual average (from good quality site–years, see Appendix C) eddy covariance derived gross primary production (GPP_{EC}) versus simulated gross primary production (GPP_{sim}) by the light use efficiency model of Bao et al. (2022) parameterized at hourly (Bao$_{\mathrm{hr}}$ model) and daily scale (Bao$_{\mathrm{dd}}$ model) for each site–year. The plots show the performance of the **(a)** Bao$_{\mathrm{hr}}$ model, and **(b)** Bao$_{\mathrm{dd}}$ model. The dots in the scatter represent a site–year. The model performance metrics (Nash-Sutcliffe efficiency, viz. NSE) are shown on the top of each subplot. The equations of fitted regression lines are shown in respective subplots.

est median performance of site–year parameterization compared to the other four parameterization strategies.

At an hourly scale, differences in model performance between a pair of similar parameterization strategies, such as a pair of detailed parameterization (i.e., between site–year-specific and site-specific) or a pair of generalized parameterization (i.e., between per PFT and global) approaches for both models were small (Fig. S8). However, this difference can be higher between a detailed and a generalized model parameterization strategy. The median differences in hourly NNSE between site–year-specific and site-specific model parameterization were 0.02 and 0.01 for the $\mathrm{P}^{\mathrm{W}}_{\mathrm{hr}}$ model and Bao$_{\mathrm{hr}}$ model, respectively. In contrast, the median differences in hourly NNSE between site–year-specific and global model parameterization were 0.11 and 0.16 for the $\mathrm{P}^{\mathrm{W}}_{\mathrm{hr}}$ model and Bao$_{\mathrm{hr}}$ model, respectively.

The median annual model performance between per–site parameterization using $Cost^{IAV}$ and per–site parameterization were relatively small, which were 0.01 and 0.02 for $\mathrm{P}^{\mathrm{W}}_{\mathrm{hr}}$ model and Bao$_{\mathrm{hr}}$ model, respectively, and it shows the additional constraint on IAV of GPP flux in the cost function did not substantially improve annual model performance. At hourly scale, the median differences in model performance between per–site parameterization using $Cost^{IAV}$ and per–site parameterization were also negligible, which were 0.00 and -0.01 for $\mathrm{P}^{\mathrm{W}}_{\mathrm{hr}}$ model and Bao$_{\mathrm{hr}}$ model, respectively. Though the per–site parameterization using $Cost^{IAV}$ did not improve the annual model performance, it also did not degrade the hourly model performance.

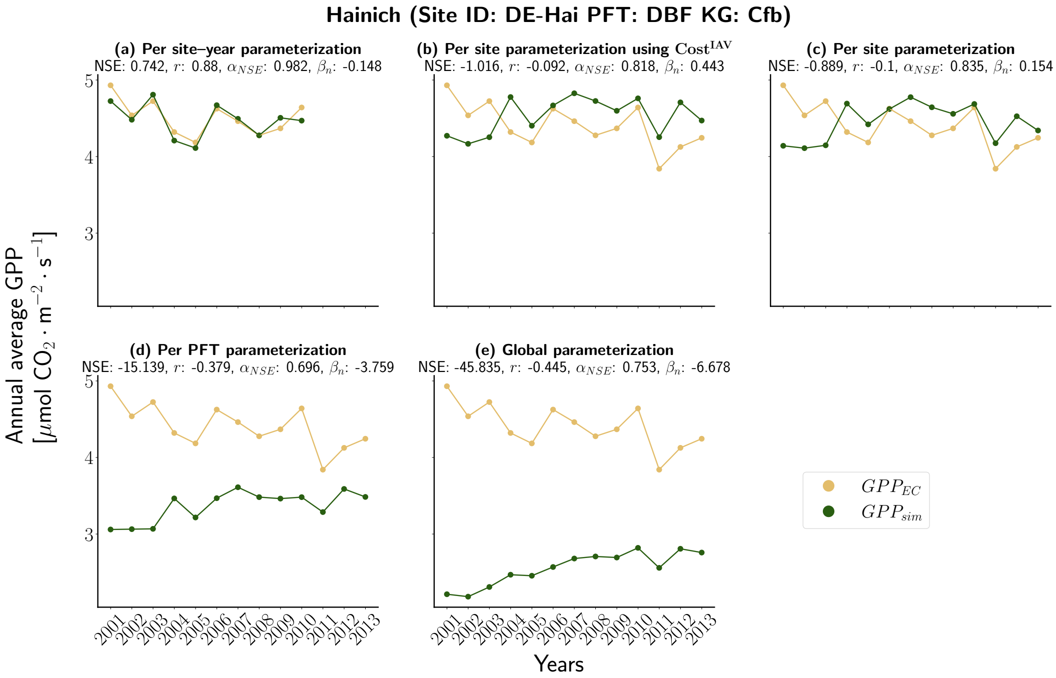

Figure 7. Comparison of annual average of gross primary production (GPP) derived by eddy covariance measurements (GPP_{EC}), and GPP simulated (GPP_{sim}) by the light use efficiency model of Bao et al. (2022), which was parameterized with hourly data (Bao$_{\mathrm{hr}}$ model). The five subplots show simulated GPP from **(a)** site–year specific parameterization, **(b)** site-specific parameterization using an additional constraint on inter–annual variability in the cost function ($Cost^{IAV}$), **(c)** site-specific parameterization, **(d)** plant–functional types (PFT) specific parameterization, and **(e)** global parameterization. The years 2010 to 2012 could not be parameterized in the case of site–year parameterization, as there were no good quality evapotranspiration estimates from latent heat flux measurements for those years. The values of model performance measures (Nash-Sutcliffe efficiency, viz. NSE, correlation coefficient, viz. r, relative variability, viz. α_{NSE}, and bias, viz. β_n) are shown on top of respective subplots. This site represents an average 140-year-old deciduous forest (Tamrakar et al., 2018) with a distinct seasonal cycle and an annual average temperature of ≈8.3 °C, and an annual average precipitation of 750–800 mm during the observation period (Knohl et al., 2003; Pastorello et al., 2020). The site ID, PFT, and Köppen-Geiger climate class (KG) of the site are provided on top of the figure in bold.

3.6 Differences between modelling experiments involving a mechanistic and a semi-empirical model structure

The lowest AIC values were obtained for per site–year parameterization for all the models at hourly and daily scales or aggregations, suggesting the sum of squares errors (SSE) was substantially reduced even when a comparatively complex parameterization strategy with a large number of model parameters was chosen (Table 3). The AIC values gradually increased from per site–year, per site, per PFT to global parameterization at hourly and daily scales or aggregations for all three models which are $\mathrm{P}^{\mathrm{W}}_{\mathrm{hr}}$ model, Bao$_{\mathrm{hr}}$ model, and Bao$_{\mathrm{dd}}$ model (Table 3). Semi-empirical models, i.e., Bao$_{\mathrm{hr}}$ model and Bao$_{\mathrm{dd}}$ model also had mostly lower values of AIC compared to mechanistic $\mathrm{P}^{\mathrm{W}}_{\mathrm{hr}}$ model even though more parameters were parameterized for these models (Table 3). At the daily scale, the Bao$_{\mathrm{dd}}$ model had the lowest AIC for all the parameterization experiments due to the pa-

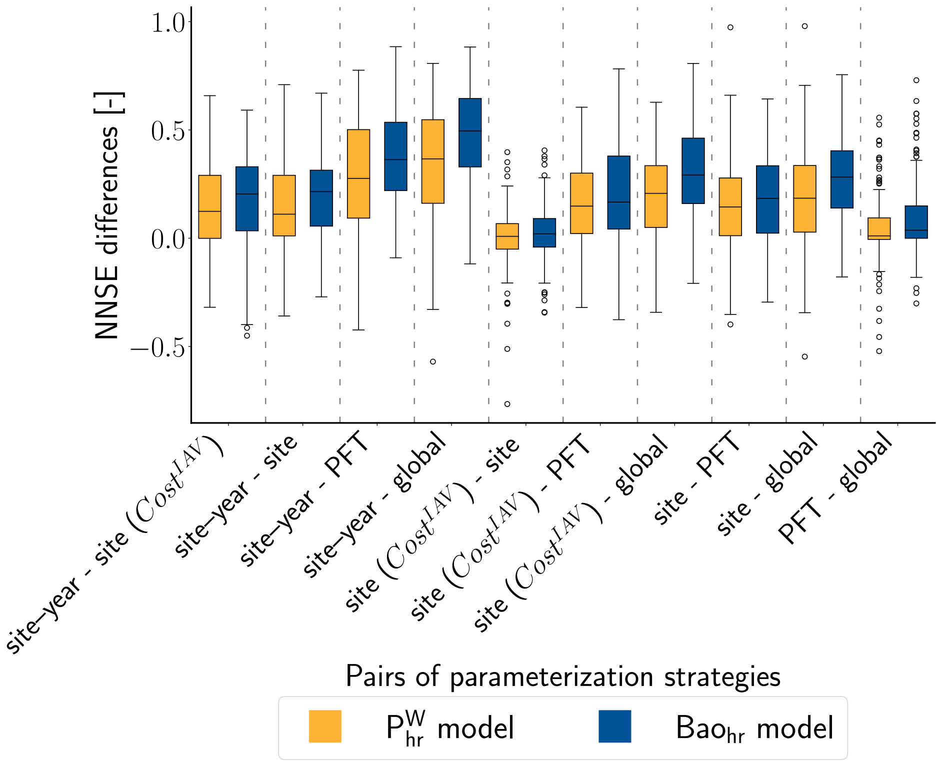

Figure 8. Distributions of the differences between model performance measures (normalized Nash-Sutcliffe efficiency, viz. NNSE) calculated at annual scale, from various pairs of model parameterization experiments conducted for the P-model of Mengoli et al. (2022) with drought stress, parameterized with hourly data (P^{W}_{hr} model) and the light use efficiency model of Bao et al. (2022), parameterized with hourly data (Bao_{hr} model). $Cost^{IAV}$ in parentheses denotes the usage of an additional constraint on annual gross primary production flux during per–site parameterization. The boxes are spanned between the first and third quartiles of the differences, and the line in the middle represents the median. The whiskers show the farthest data point from the box within 1.5× of the interquartile range. The circles represent the outliers that go beyond the limits of the whiskers. The vertical dotted grey line separates each pair of model parameterization strategies.

rameterization at daily scale. Whereas, for the other two models, parameterization and forward runs were performed at an hourly scale and then simulations were aggregated to daily resolution. The P_{hr} model was not included in AIC or AIC_c analysis as previous results proved this model significantly underperformed compared to the other models, and this will always result in higher AIC or AIC_c values.

At monthly and annual scales, we show the differences in AIC_c values between P^{W}_{hr}, Bao_{hr}, and Bao_{dd} models for the same parameterization strategy, and not between parameterization strategies in a same model. The reason behind this is AIC_c values largely depend on the relationship between sample size, i.e., n, and the total number of parameters which were parameterized, i.e., K. The values of AIC_c became very large even when a significantly smaller SSE was obtained, and they became unreliable when the value of n was closer to K. For example, at monthly aggregation, per site–year parameterization of the P^{W}_{hr} model had a very high AIC_c value of 1.30×10^{6} even when it had the lowest SSE among all the five parameterization strategies (Table S3, S4, and S5). The

Bao_{dd} model proved to be better able to capture the seasonal cycle, i.e., monthly GPP estimates compared to the other two models for most of the parameterization experiments considering the number of parameters parameterized (Table 3). However, the P^{W}_{hr} model had the lowest AIC_c value in the case of per site parameterization using $Cost^{IAV}$ and per site parameterization at monthly aggregation. In contrast, at an annual scale, the P^{W}_{hr} model had mostly the lowest AIC_c values, and some of the experiments also suffered from the above-described unreliable AIC_c estimates, where n and K had similar values (Table 3, S3, S4, and S5).

Table 3. Akaike's Information Criterion (AIC) or corrected AIC (AIC_c) values for modelling experiments of various complexities

Temporal scale/ aggregation	**Models**	**parameterization strategies**				
		per site–year	**per site using $\mathbf{Cost^{IAV}}$**	**per site**	**per PFT**	**global**
Hourly scale (AIC)	P^{W}_{hr}	1.72×10^{7}	1.84×10^{7}	1.86×10^{7}	2.16×10^{7}	2.25×10^{7}
	Bao_{hr}	1.57×10^{7}	1.71×10^{7}	1.78×10^{7}	2.11×10^{7}	2.41×10^{7}
Daily scale/ aggregation (AIC)	P^{W}_{hr}	4.58×10^{5}	5.05×10^{5}	5.11×10^{5}	6.88×10^{5}	7.42×10^{5}
	Bao_{hr}	3.86×10^{5}	4.35×10^{5}	4.50×10^{5}	6.74×10^{5}	8.00×10^{5}
	Bao_{dd}	2.78×10^{5}	3.25×10^{5}	3.44×10^{5}	5.10×10^{5}	5.65×10^{5}
Monthly aggregation ($\mathbf{AIC_c}$)	P^{W}_{hr}	1.30×10^{6}	1.63×10^{4}	1.63×10^{4}	2.04×10^{4}	2.25×10^{4}
	Bao_{hr}	-4.71×10^{4}	1.81×10^{4}	1.77×10^{4}	2.01×10^{4}	2.42×10^{4}
	Bao_{dd}	-3.39×10^{4}	1.70×10^{4}	1.73×10^{4}	1.55×10^{4}	1.79×10^{4}
Annual aggregation ($\mathbf{AIC_c}$)	P^{W}_{hr}	-3.56×10^{3}	-5.22×10^{3}	-5.56×10^{3}	9.09×10^{2}	9.28×10^{2}
	Bao_{hr}	-3.93×10^{3}	-3.87×10^{3}	-4.07×10^{3}	1.45×10^{3}	1.21×10^{3}
	Bao_{dd}	-3.70×10^{3}	-3.51×10^{3}	-3.53×10^{3}	1.10×10^{3}	8.31×10^{2}

3.7 Model performances across different plant–functional types

Generally better model performances were achieved with both the P^{W}_{hr} model and the Bao_{hr} model when parameterized detailed model parameterization strategies were used (Fig. 9). In this analysis, we removed the per–site parameterization experiment using $Cost^{IAV}$ as it performed very similar to per–site parameterization, and also we did not consider the P_{hr} model as it produced poor performance across all the PFTs. The highest median NNSEs were obtained with per–site–year parameterization for all the PFTs. For the P^{W}_{hr} model parameterization experiments, the highest median value of NNSE was found for CSH for per–site–year parameterization (median NNSE: 0.88), DBF for per–site parameterization (median NNSE: 0.85), CSH and DBF for per-PFT parameterization (median NNSE: 0.81), and CSH for global parameterization (median NNSE: 0.80). However, CSH had only three sites and highest median model performance for CSH should be interpreted with caution. For the Bao_{hr} model parameterization experiments, the highest median value of NNSE was found for DBF for per–site–year parameterization (median NNSE: 0.89), DBF for per–site parameterization (median NNSE: 0.88), MF for per-PFT parameterization (median NNSE: 0.84), and DBF and MF for global parameterization (median NNSE: 0.77). We found similar results also for climate–vegetation types

(Sect. 2.2), where a more detailed parameterization strategy achieved higher model performance than a generalized parameterization strategy (Sect. 2.6 and Fig. S9).

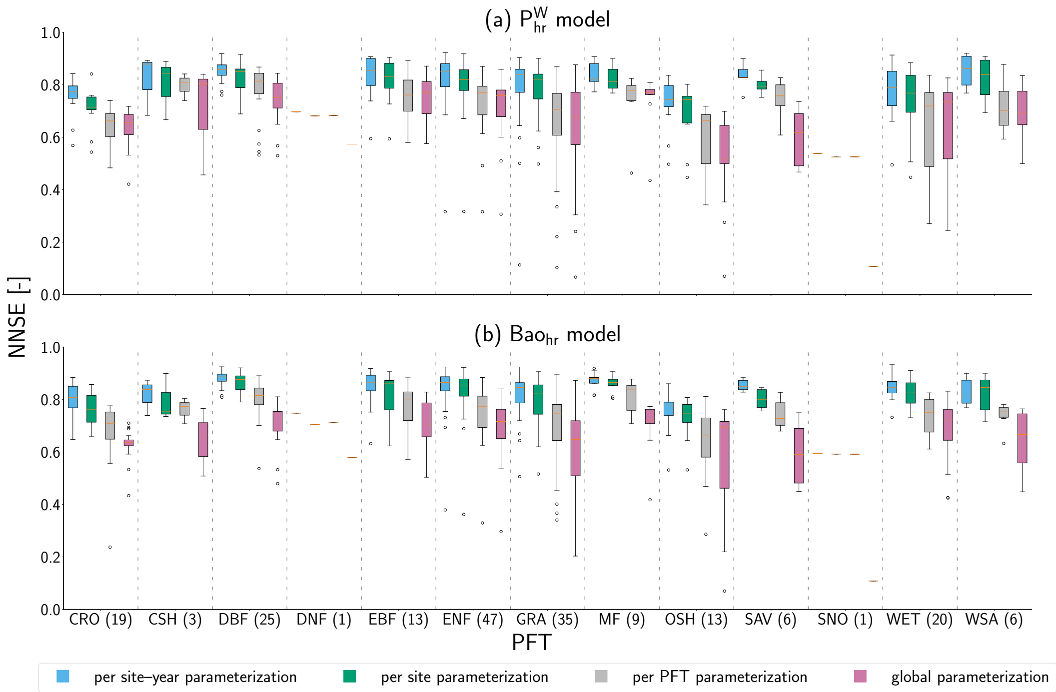

Figure 9. Box-plots showing the range of the hourly model performance metric (normalized Nash-Sutcliffe efficiency, viz. NNSE), for the sites in different plant–functional types (PFT), and different parameterization experiments. The subplots show the model performance for **(a)** P-model of Mengoli et al. (2022) with drought stress function, parameterized with hourly data (P^{W}_{hr} model), and **(b)** the light use efficiency model of Bao et al. (2022) parameterized with hourly data (Bao_{hr} model). The numbers in parentheses beside the name of each PFT on the x-axis are the number of sites present in a specific PFT. The boxes are spanned between the first and third quartiles of NNSE values, and the line in the middle represents the median. The whiskers show the farthest data point from the box within 1.5× of the interquartile range. The circles represent the outliers that go beyond the limits of the whiskers. For, deciduous needle-leaf forests (DNF), and areas covered by snow (SNO) only the median value could be shown as these PFTs have only one site. The vertical dotted grey line separates each PFT.

3.8 Correlation between annual model performance and model performance in simulating diurnal GPP peaks

One of the crucial reasons behind poor annual model performance (Fig. 2 and Table D1) can be the inability of both the P^{W}_{hr} model and the Bao_{hr} model to capture the peaks of GPP_{EC} (Fig. S10). Specifically, $P90_{GPP_{EC}}$ was highly underestimated in the case of global parameterization. The median of the ratio of $P90_{GPP_{sim}}$ to $P90_{GPP_{EC}}$ were 0.77 and 0.53 for the P^{W}_{hr} model and the Bao_{hr} model, respectively, during global parameterization. The underestimation generally decreased with more detailed parameterization strategies with little differences between both the models. The median values of the ratio of $P90_{GPP_{sim}}$ to $P90_{GPP_{EC}}$ were 0.95 and 0.93 for the site–year parameterization of the P^{W}_{hr} model and the Bao_{hr} model, respectively. Moreover, the lower values of the interquartile range (IQR) of these ratios signify the importance of site–year parameterization compared to per PFT or global parameterization to reliably capture the peak GPP_{EC} in diurnal cycles for most of the sites and to attain better model performance at the sub-daily scale. The values of IQR were 0.1 for both the P_{hr}^{W} model and the Bao_{hr} model in the case of site–year parameterization, 0.44 and 0.38 for the P_{hr}^{W} model and the Bao_{hr} model, respectively in the case of PFT-specific parameterization, and 0.44 and 0.42 for the P_{hr}^{W} model and the Bao_{hr} model, respectively in the case of global parameterization.

We further found that if a certain parameterization strategy better simulated the $P90_{GPP_{EC}}$ for each site–year, it corresponded to a comparatively better annual model performance for a site which is demonstrated by the positive values of Pearson correlation coefficients (Fig. 10). Here also, a detailed parameterization strategy, such as site–year parameterization resulted in a better simulation of $P90_{GPP_{EC}}$, and thus better annual model performance for most of the sites compared to a generalized parameterization strategy, such as global parameterization. In this case when $j1$ was site–year parameterization and $j2$ was global parameterization, 91% and 93% sites had higher $NNSE_{P90}^{j1}$ than $NNSE_{P90}^{j2}$ and corresponding $NNSE_{y}^{j1}$ than $NNSE_{y}^{j2}$ for the P_{hr}^{W} model and the Bao_{hr} model, respectively. When $j1$ was parameterization per site using $Cost^{IAV}$ and $j2$ was parameterization per site, respectively, only 34.43% and 41.8% had positive values of $\Delta NNSE_{P90}$ (i.e., $NNSE_{P90}^{j1} > NNSE_{P90}^{j2}$) and corresponding $\Delta NNSE_{y}$ (i.e., $NNSE_{y}^{j1} > NNSE_{y}^{j2}$) for the P_{hr}^{W} model and the Bao_{hr} model, respectively. This signified that using an additional constraint related to the IAV of GPP in the cost function during model parameterization did not improve the prediction of peak GPP values for most of the sites.

4 Discussion

4.1 Uncertainties in modelling experiments

Any model–data–integration study is prone to uncertainties related to both data and the model. The EC dataset used in our study has a couple of well-known uncertainties. For example, the NEE measurements from eddy covariance can have uncertainties due to the accumulation of atmospheric CO_2 under the canopy at night (storage) and a sudden turbulent mixing during the morning when the stable nighttime boundary layer breaks up, or because of advection of atmospheric CO_2 out of the control volume sampled by the eddy covariance system (D. Baldocchi et al., 2000; Aubinet, 2008; Jocher et al., 2018). The GPP fluxes that we used were derived from NEE measurements by extrapolating the nighttime respiration of ecosystem (Reichstein et al., 2005) to daytime. Moreover, GPP can be estimated based on another well-known algorithm, the daytime partitioning method (Lasslop et al., 2010). We preferred nighttime partitioning as only respiration is modelled in this method. In daytime partitioning, both GPP and respiration are modelled, resulting in higher prediction errors. We believe the uncertainties in our modelling results due to the choice of partitioning algorithm should be small as quantified in a previous study by Desai et al. (2008). We also used ET in the cost function which is equivalent to latent heat flux. The mismatch between the summation of latent, sensible, and ground heat fluxes with net radiation calculated using incoming and outgoing radiation, the so-called lack of energy–balance closure, remains a long-standing challenge with EC measurements (Foken, 2008; Mauder et al., 2020; Zhang et al., 2024). Quality control of millions of data points at an hourly scale was also challenging, especially when we merged data from various sources, such as in-situ measurements, modelled re-analysis data, and remote sensing-based estimates. Another major uncertainty arises from the mismatch between the footprint of EC towers and the grid of remote sensing data which were used to calculate vegetation indices (Chu et al., 2021). The PFT classification of sites based on a simple PFT classification method may not accurately represent the vegetation of some of the sites. For example, a site at Alice Springs (AU-ASM) in central Australia was classified as a savanna in FLUXNET2015 (Pastorello et

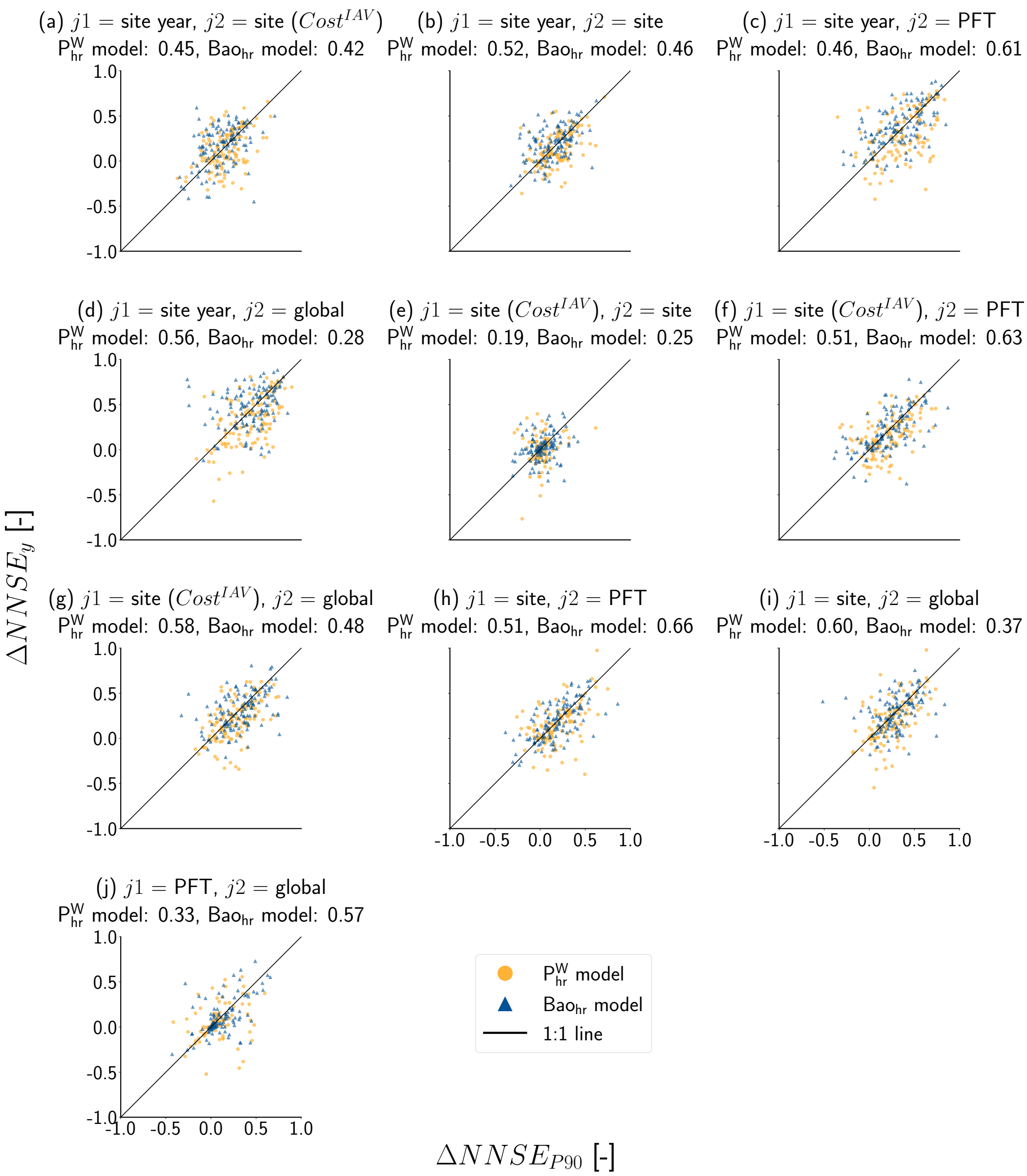

Figure 10. Scatter between differences in model performance in simulating peak gross primary production, viz. GPP ($\Delta NNSE_{P90}$) and annual average of GPP ($\Delta NNSE_y$). $j1$ and $j2$ are a pair of parameterization strategies for which differences are calculated in each subplot from (a) to (j). Each dot in the scatter represents a site. The P_{hr}^{W} model and Bao_{hr} model are P-model of Mengoli et al. (2022) with drought stress function and the light use efficiency model of Bao et al. (2022), both were parameterized with hourly data. The values on top of each subplot indicate the Pearson correlation coefficient (PCC, 2008) for respective models.

al., 2020). In fact, this site is dominated by a discontinuous canopy of Mulga (*Acacia aneura*) that has needle leaves and a seasonal understory grassy layer (Cleverly et al., 2013). This site can be classified as a woody savanna as well. An arctic site in Bayelva (SJ-Blv) has a combination of snow, wet grounds, and specific tundra vegetation (Boike et al., 2018) which were not well represented by the snow classification of FLUXNET2015

(Pastorello et al., 2020). Another limitation of the dataset is that sites are mostly clustered in European and North American countries and, hence do not necessarily represent global ecosystem functioning particularly well due to sampling bias (Papale et al., 2015). Similarly, some PFTs are represented by very few sites, which makes PFT-specific parameterization challenging.

4.2 General performance of models in simulating GPP

In this study, we evaluated the performance of the optimality-based $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ and P_{hr} models across a wide range of sites, representing various vegetation and climate types. We uncovered poor model performance of the P_{hr} model at many sites, especially at arid sites. Calculating *WAI* using a simple hydrological model and inducing a moisture stress function, i.e., the introduction of the $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model substantially improved model performance to simulate the annual average of GPP fluxes across many sites, including both water–limited and energy–limited sites. The inclusion of the moisture stress function to the P_{hr} model not only improved the annual model performance but also improved the model performance across all the temporal scales or aggregation levels. This highlights the importance of the representation of soil moisture conditions for modelling approaches, which are aimed at accurately representing ecosystem functioning and vegetation response. However, the coupling of the hydrological model raised the need to calibrate nine more parameters, which counters the vision of developing a parameter–sparse approach using theories that demand a lower site or site–year specific fine–tuning of model parameters (Prentice et al., 2015). Further experimentation is needed to find a fine balance between the number of key model parameters, which require calibration, and an accurate representation of ecosystem functioning.

Coming to the differences in model structure, we found that semi-empirical models ($\mathrm{Bao}_{\mathrm{hr}}$ and $\mathrm{Bao}_{\mathrm{dd}}$ models) performed statistically better, i.e, had a lower value of AIC compared to mechanistic model ($\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model) at hourly and daily scale or aggregations for most of the parameterization experiments even though the semi-empirical modelling experiments needed more parameters to be parameterized. At the monthly aggregation level, the seasonal cycles were also significantly better captured by the parameter–heavy semi-empirical model parameterized with daily data ($\mathrm{Bao}_{\mathrm{dd}}$ model) for most of the parameterization experiments. However, at the annual aggregation level, the mechanistic model, i.e., the $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model was comparatively better in most cases and a more flexible semi-empirical model with a higher number of parameters did not have a substantial improvement in annual model performance.

Though the partial sensitivity functions of environmental variables used in the $\mathrm{Bao}_{\mathrm{hr}}$ model and the $\mathrm{Bao}_{\mathrm{dd}}$ model were found to be applicable for most of the sites, they can be of many different types and may vary across site conditions (Bao et al., 2022). The EC sites were also affected by human management, such as irrigation, harvesting, and mowing as well as natural disturbances, such as fire, and pest attacks. These factors can affect the IAV of GPP flux which was estimated from EC measurements. Models used in this study may not be able to account for all of these factors due to structural limitations. For example, in the hydrological model, we only used precipitation and ET to calculate the mass balance of water. However, human management (such as irrigation and drainage) can play an important role, and the *WAI* estimates in managed sites, such as at an irrigated maize site (US-Ne1) may not be accurate.

4.3 The importance of the parameterization approach to estimate IAV of GPP

We also emphasize the importance of determining the parameterization approach by the inter–comparison of parameterization strategies. parameterization of model parameters largely determines model performance and parameterized parameters capture the individual characteristics of sites or climatic events of site–years (Wu et al., 2012). Detailed model parameterization strategies, such as parameterization specific to site–years also comparatively better predicted the annual average of GPP fluxes and year-specific parameters explained some parts of the IAV of GPP flux. As the fast rate of change in climatic characteristics has become more frequent in recent years, developing a generalized model structure to simulate carbon fluxes between years and/or between sites of similar vegetation types has become even more challenging (Knauer et al., 2023). Moreover, the generalized model parameterization strategy, i.e., global parameterization was also dominated by PFTs, such as ENF and GRA which were represented by many sites, and certain PFT, such as DNF was represented by only one site in the FLUXNET2015 dataset (Pastorello et al., 2020). This may imply that global parameterization or parameter up-scaling experiments using the FLUXNET2015 dataset (Pastorello et al., 2020) may result in biased parameter sets that cannot be generalized to the global scale or a weighted site representation may be necessary in this case. Besides model parameterization, a recent study by (Zou et al., 2024) also found that the importance of each independent driver, and their relative contributions varies over time and ecosystem type. The relative importance of forcing variables may also be another factor besides model parameterization.

Though we have demonstrated the capability of the $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model that included drought stress and the $\mathrm{Bao}_{\mathrm{hr}}$ model to simulate the hourly fluxes of GPP, accurate estimation of IAV of GPP fluxes at the site level with these models requires further developments. Particularly, both models failed to capture the peak GPP in diurnal cycles at many sites even after model parameterization at a sub-daily scale and using an additional constraint on the IAV of GPP in the cost function. These underestimations at an hourly scale may have accumulated to a larger error when the fluxes were aggregated at an annual scale to study the IAV. We also showed that comparatively better model performances were achieved when the GPP peaks per site–year were better simulated. These results are similar to another study by Lin et al. (2023), directed at evaluating terrestrial ecosystem models' capability in explaining the IAV of GPP which also found an underestimation of GPP. However, it is also true that some peak values of diurnal GPP can also be an outlier produced by data processing algorithms, such as the gap-filling algorithm, and it is hard to distinguish these outliers from true GPP values. Moreover, it was found both the $\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model and the $\mathrm{Bao}_{\mathrm{hr}}$ model showed the highest model performance at the sub-daily scale mostly for forest sites compared to savannas or grasslands, this in turn led to the poor simulation of IAV at many sites which are not forests.

The poor representation of IAV of GPP can be attributed to either limitations of models or model parameterization strategies. It is important to discover which seasonal phases of the GPP dynamics for particular vegetation types or climatic zones are not well represented in models simulating the IAV of GPP. It is particularly important to focus on the meteorological sensitivity of GPP during periods of high productivity where improvements in the prediction of high fluxes would tend to improve the description of IAV. Another aspect could also be to decompose the metric (Gupta et al., 2009) used in the cost function or develop a more detailed model evaluation to understand which other parts of the time series were not well constrained during model parameterization.

5 Conclusions

We have demonstrated the capability of an improved version of an optimality-based mechanistic model ($\mathrm{P}_{\mathrm{hr}}^{\mathrm{W}}$ model) and a semi-empirical LUE model ($\mathrm{Bao}_{\mathrm{hr}}$ and $\mathrm{Bao}_{\mathrm{dd}}$ model) to simulate sub-daily GPP fluxes across 198 EC sites, representing 13 different vegetation types including forests, grasslands, savannas, croplands, and tundra. We conclude that explicit accounting of drought stress in the optimality-based ecosystem model is a necessity as it proved to be an important factor in controlling GPP fluxes at all temporal scales including at annual aggregation. We found that the semi-empirical model mostly produced better results at hourly, daily, and monthly scales compared to the mechanistic model. However, at an annual scale, the improvement in the performance of the semi-empirical model was not significant even though more parameters were parameterized to flexibly capture the ecosystem dynamics. While these models generally performed well in simulating hourly GPP dynamics, the small errors at the sub-daily scale, particularly related to the estimation of GPP peaks, accumulated to bigger errors at the annual scale and led to poor performance of models in explaining the IAV of GPP. We found that comparatively better annual model performance could be achieved when the peaks of GPP were better simulated. Moreover, both models performed better mostly at forest sites compared to grasslands or savannas which may also lead to poor estimation of IAV of GPP at many sites globally. Our results further suggest the need to focus on sub-daily GPP dynamics during the various seasonal phases, especially highly productive ones, towards an improved constraint on GPP sensitivities. Hence, better annual model performance with a detailed parameterization strategy, such as site–year parameterization, signifies that temporally varying model parameters are necessary to better capture the variations of annual average GPP and indicate that ecosystem functioning is not stable between years. We believe these new understandings can guide us towards developing models and parameterization strategies for simulating the inter–annual variations in ecosystem GPP more successfully, and improve our understanding of the global carbon cycle response to changing climatic conditions.

Appendix A Data description

Table A1: Description of forcing and model parameterization data

Abbreviation	**Definition**	**Unit**	**Variable name in dataset/ remarks**	**Reference**
GPP_{EC}[a]	GPP derived from EC based net ecosystem exchange (NEE) using night-time partitioning* method	μmol$CO_2 \cdot m^{-2} \cdot s^{-1}$	`GPP_NT_VUT_USTAR50`	Pastorello et al. (2020); Reichstein et al. (2005)
σ_{NEE}	Random uncertainty for NEE	μmol$CO_2 \cdot m^{-2} \cdot s^{-1}$	`NEE_VUT_USTAR50_RANDUNC`	Pastorello et al. (2020)
LE	Latent heat flux	$W \cdot m^{-2}$	`LE_F_MDS`	Pastorello et al. (2020)
σ_{LE}	Random uncertainty for latent heat flux	$W \cdot m^{-2}$	`LE_RANDUNC`	Pastorello et al. (2020)
SW_IN[b]	Incoming shortwave radiation	$W \cdot m^{-2}$	`SW_IN_F`	Pastorello et al. (2020)
$NETRAD$[b, c]	Net radiation	$W \cdot m^{-2}$	`NETRAD`	Pastorello et al. (2020)
SW_IN_POT	Potential incoming shortwave radiation	$W \cdot m^{-2}$	`SW_IN_POT`	Pastorello et al. (2020)
$PPFD\_IN$[a]	Incoming photosynthetic photon flux density	μmol photons$\cdot m^{-2} \cdot s^{-1}$	$PPFD\_IN$ gap-filled with $2.04 \times SW\_IN$	Pastorello et al. (2020); see Sect. 3.4.2 of Stocker et al. (2020) for the gap-filling equation

Continued on next page

Table A1 – *Continued from previous page*

Abbreviation	**Definition**	**Unit**	**Variable name in dataset/ remarks**	**Reference**
T^{b}	Air temperature	°C	`TA_F_MDS`	Pastorello et al. (2020)
VPD^{b}	Vapor pressure deficit	Pa	`VPD_F_MDS`	Pastorello et al. (2020)
$P^{b,d}$	Precipitation	$mm \cdot h^{-1}$ or $mm \cdot d^{-1}$	`P`	Pastorello et al. (2020)
CO_2	Atmospheric CO_2 concentration dry air mole fractions from quasi-continuous measurements at Mauna Loa	ppm	`co2_mlo_surface-insitu_1_ccgg_DailyData` (interpolated linearly to hourly scale). The measurements from Mauna Loa were used for all sites as the CO_2 concentration measurements at EC sites are often noisy and discontinuous.	Thoning et al. (2021)
elev	Site elevation	m a.s.l.	Collected from literature	Bao et al. (2022)
$ET_{LE}{}^{a}$	Evapotranspiration derived from *LE* flux	$mm \cdot h^{-1}$ or $mm \cdot d^{-1}$	Calculated from *LE* with a dependency on *T*	Henderson-Sellers (1984)
σ_{ET}	Random uncertainty for *ET_LE*	$mm \cdot h^{-1}$ or $mm \cdot d^{-1}$	Calculated from *LE_RANDUNC* with a dependency on *T*	Henderson-Sellers (1984)
PET	Potential evapotranspiration	$mm \cdot h^{-1}$ or $mm \cdot d^{-1}$	Calculated from *T*, *NETRAD* and *elev* using the method of Priestley and Taylor	Priestley and Taylor (1972)
CI	Cloudiness index	-	Calculated as $1 - \left(\frac{SW\_IN}{SW\_IN\_POT}\right)$	Bao et al. (2022); Fu and Rich (1999); Turner et al. (2006)
WAI	Water availability indicator	mm	Described in Sect. S1 of the supplement	Bao et al. (2022); Tramontana et al. (2016); Trautmann et al. (2018)
W	Soil water supply	$mm \cdot mm^{-1}$	calculated as $\frac{WAI}{AWC}$ (*AWC* is defined in Table 1)	Bao et al. (2022)
NDVI	Normalized difference vegetation index	-	Daily *NDVI* from FluxnetEO v2 (MODIS) was linearly interpolated to hourly	Walther et al. (2022, 2023)

Continued on next page

Table A1 – *Continued from previous page*

Abbreviation	**Definition**	**Unit**	**Variable name in dataset/ remarks**	**Reference**
$fAPAR$	Fraction of incident photosynthetic photon flux that is absorbed by vegetation	-	Linear relationship between $NDVI$ and $fAPAR$ was assumed. $\begin{cases} NDVI, & \text{if } NDVI > 0 \\ 0, & \text{if } NDVI \leq 0 \end{cases}$	Bao et al. (2022); Myneni et al. (1997)
QC[a]	Data quality flags	-	1.0 (good quality), 0.5 (medium quality), and 0.0 (bad quality) in the case of hourly data, which is the fraction of good quality measured or gap-filled data from two half-hours. In the case of daily, QC can have any values between 0.0 and 1.0, which is a fraction representing the percentage of good quality measured or gap-filled data in a day. The daily data with $QC > 0.8$ was considered good.	Pastorello et al. (2020); Nelson et al. (2024)

a: For GPP_{EC}, the QC flags of NEE, and for ET_{LE} the QC flags of LE were used, as they were derived from the respective variables. QC flags of SW_IN were used to determine bad and medium quality data of $PPFD\_IN$, which were replaced with a gap-filling procedure.

b: Bad, medium quality (value of QC is 0 and 0.5) data and gaps were filled with downscaled (Besnard et al., 2019) ERA5 (Hersbach et al., 2023) or ERA-Interim v2.0 data (Berrisford et al., 2011).

c: We have collected good quality SW_IN and $NETRAD$ values from all the sites and fitted a linear regression model using the RANdom SAmple Consensus (RANSAC) algorithm (Fischler & Bolles, 1981) to determine the relation between them. The fitted equation ($NETRAD = 0.7066 \times SW\_IN - 0.1345$) was used to fill gaps in $NETRAD$ using SW_IN. The gap-filling with regression was only applied for a few sites at hourly scale.

d: At hourly scale, the data gaps or bad quality data in P were filled by distributing the daily downscaled P (Besnard et al., 2019) from ERA-Interim v2.0 (Berrisford et al., 2011) for a certain day to the hourly timesteps, based on hourly P from gridded ERA5 data (Hersbach et al., 2023).

*We preferred the night-time partitioning (Reichstein et al., 2005) over daytime partitioning (Lasslop et al., 2010) as only respiration is modelled in this case and GPP is derived as the difference between measured NEE and respiration. Whereas, in the daytime partitioning method, GPP is modelled as well and can have prediction errors due to uncertain model parameters.

Appendix B Data screening for model parameterization

We used only good–quality data to calibrate model parameters. At hourly scale, we selected GPP_{EC} and ET_{LE} as good quality data when the values of their respective QC flag were 1 (Table A1). At the daily scale, we considered a GPP_{EC} and ET_{LE} data point as good when the value of the QC flag was greater than 0.8. We also removed any data gaps from observed and simulated data, σ_{NEE}, and σ_{LE} (Table A1). There were certain negative values in our GPP_{EC} data, as it was calculated using night-time based partitioning method (Reichstein et al., 2005). In this case, if a negative GPP_{EC} value occurred, when the SW_IN (Table A1) is zero i.e., during night hours, we replaced those data points with 0 and used them in the cost function. If the negative GPP_{EC} occurred during day hours, we excluded them.

Appendix C Data screening for evaluation of model performance

The good quality data at an hourly scale were selected using the same criteria described in Appendix B. The data screening at a daily scale was also similar to Appendix B, when the LUE model was parameterized using daily data. For all other cases, we assigned a flag (0 = not considered, 1 = considered) to identify which data points were considered during model parameterization. We aggregated this flag to daily, weekly, monthly, and annual scales by taking averages. Then this flag indicated the fraction of good quality data used to calculate a data point in a certain temporal resolution. We only used data points at certain temporal resolutions which were calculated using more than 50% (flag value > 0.5) good quality data points from hourly/daily resolution. We calculated monthly, and annual model performance metrics for a certain site if at least three data points were present. We couldn't calculate annual metrics for 76 and 85 sites due to low numbers of good quality site–years when the annual data was aggregated from hourly, and daily data, respectively. The monthly metrics were not calculated for the three sites due to the same reason when they were aggregated from daily data.

Appendix D Median values of model performance

The median values of the model performance metric, i.e., NNSE which are plotted in Fig. 2 are summarized in Table D1.

Table D1. Median NNSE obtained at each modelling experiment at hourly/daily scale and annual aggregations

Temporal scale/ aggregation	**Models**	**parameterization strategies**				
		per site–year	**per site using $\mathbf{Cost^{IAV}}$**	**per site**	**per PFT**	**global**
Hourly/ daily scale	P^{W}_{hr}	0.827	0.799	0.805	0.738	0.712
	P_{hr}	0.478	0.470	0.469	0.461	0.490
	Bao_{hr}	0.853	0.816	0.836	0.757	0.693
	Bao_{dd}	0.830	0.774	0.790	0.658	0.634
annual aggregation	P^{W}_{hr}	0.543	0.405	0.373	0.201	0.143
	P_{hr}	0.018	0.019	0.018	0.018	0.019
	Bao_{hr}	0.661	0.471	0.440	0.190	0.105
	Bao_{dd}	0.669	0.489	0.482	0.238	0.143

Open Research Section

The codes that were used to perform all the necessary analyses and plot all the figures in this study are available at De (2024). The data from eddy covariance sites are available from FLUXNET (Pastorello et al., 2020; FLUXNET.org, 2024a). The FluxnetEO MODIS version 2 dataset is available at Walther et al. (2022, 2023). ERA5 dataset and ERA-Interim v2.0 data are available from Hersbach et al. (2023) and Berrisford et al. (2011), respectively. Atmospheric CO_2 measurements at Mauna Loa observatory are available at Thoning et al. (2021).

Acknowledgments

We would like to express our gratitude to the FLUXNET community, including regional networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, FluxnetCanada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOSSiberia, and USCCC for their contributions in developing the eddy covariance based FLUXNET2015 dataset, which was invaluable to our study. We also thank all the site investigators for generously sharing their data with the respective networks. The FLUXNET eddy covariance data processing and harmonization was carried out by the ICOS Ecosystem Thematic Center, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux and AsiaFlux offices. We acknowledge suggestions received from Jacob Nelson regarding the FLUXNET dataset, suggestions received from Sophia Walther regarding the FluxnetEO dataset, assistance received from Ulrich Weber in some of the data processing steps, assistance received from Giulia Mengoli with P-model codes, assistance received from Nikolaus Hansen regarding the usage of pycma, and members of the Model–Data Integration group at Max Planck Institute for Biogeochemistry for sharing constructive feedback. We thank Olaf Kolle, J. Rose Cleverly, and Shirley Papuga for providing clarifications and further information on site characteristics. We thank Donatella Zona, J. Rose Cleverly, Pasi Kolari, Julia Boike, Iris Feigenwinter, Donatella Spano, Javier Houspanossian, and Wang Yi for providing feedback on an earlier version of the manuscript. RD acknowledges the funding from the International Max Planck Research School for Global Biogeochemical Cycles (IMPRS-gBGC) for carrying out his doctoral research. TT considers this a contribution to the Swedish National Space Agency projects (SNSA Dnr 2021-00144; 2021-00111) and also acknowledges funding from FORMAS (Dnr 2021-00644; 2023-02436). LS acknowledges support from the Ministry of Education, Youth and Sports of the Czech Republic within the CzeCOS program (grant number LM2023048) and the AdAgriF project (CZ.02.01.01/00/22_008/0004635). LH acknowledges funding from the SNF funded projects ICOS-CH Phase 1, 2, and 3 (20FI21_148992, 20FI20_173691, and 20F120_198227). M Roland and BG acknowledge the support of the Research Foundation Flanders (FWO) for the support to ICOS Research Infrastructure. LM acknowledges the funding from the Forest Services, Autonomous Province of Bozen. Large language models (LLM), i.e., ChatGPT (version GPT-4-1106-vision-preview API) and GitHub Copilot (v1.172.0) were used to generate certain code snippets to perform some of the analyses in this study, and to generate code documentation, followed by careful evaluations by authors. However, no texts in this research paper were written with the aid of LLM.

References

Amos, B., Arkebauer, T. J., & Doran, J. W. (2005, March). Soil Surface Fluxes of Greenhouse Gases in an Irrigated Maize-Based Agroecosystem. *Soil Science Society of America Journal*, *69*(2), 387–395. Retrieved from `http://dx.doi.org/10.2136/sssaj2005.0387` doi: 10.2136/sssaj2005.0387

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., … Zhao, M. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, *53*(3), 785-818. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015RG000483` doi: 10.1002/2015RG000483

Aubinet, M. (2008). Eddy covariance CO_2 flux measurements in nocturnal conditions: an analysis of the problem. *Ecological Applications*, *18*(6), 1368–1378. Retrieved 2024-02-15, from http://www.jstor.org/stable/40062261

Baldocchi, D., Chu, H., & Reichstein, M. (2018). Inter-annual variability of net and gross ecosystem carbon fluxes: A review. *Agricultural and Forest Meteorology*, *249*, 520-533. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168192317301806 doi: 10.1016/j.agrformet.2017.05.015

Baldocchi, D., Finnigan, J., Wilson, K., Paw U, K., & Falge, E. (2000). On measuring net ecosystem carbon exchange over tall vegetation on complex terrain. *Boundary-Layer Meteorology*, *96*, 257–291. doi: 10.1023/A:1002497616547

Baldocchi, D. D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology*, *9*(4), 479-492. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2486.2003.00629.x doi: 10.1046/j.1365-2486.2003.00629.x

Bao, S., Alonso, L., Wang, S., Gensheimer, J., De, R., & Carvalhais, N. (2023). Toward robust parameterizations in ecosystem-level photosynthesis models. *Journal of Advances in Modeling Earth Systems*, *15*(8), e2022MS003464. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003464 doi: 10.1029/2022MS003464

Bao, S., Wutzler, T., Koirala, S., Cuntz, M., Ibrom, A., Besnard, S., ... Carvalhais, N. (2022). Environment-sensitivity functions for gross primary productivity in light use efficiency models. *Agricultural and Forest Meteorology*, *312*, 108708. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168192321003944 doi: 10.1016/j.agrformet.2021.108708

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018, October). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. Retrieved 2024-02-08, from https://www.nature.com/articles/sdata2018214 doi: 10.1038/sdata.2018.214

Berrisford, P., Dee, D., Poli, P., Brugge, R., Fielding, M., Fuentes, M., ... Simmons, A. (2011, 11/2011). The ERA-Interim archive version 2.0 (Computer software manual No. 1). Shinfield Park, Reading: ECMWF.

Besnard, S., Carvalhais, N., Arain, M. A., Black, A., Brede, B., Buchmann, N., ... Reichstein, M. (2019, February). Memory effects of climate and vegetation affecting net ecosystem CO_2 fluxes in global forests. *PLOS ONE*, *14*(2), e0211510. Retrieved from http://dx.doi.org/10.1371/journal.pone.0211510 doi: 10.1371/journal.pone.0211510

BioRender. De, R. (2024). *https://BioRender.com/i01x768.* (last accessed: 06-Sep-2024)

Boese, S., Jung, M., Carvalhais, N., Teuling, A. J., & Reichstein, M. (2019). Carbon–water flux coupling under progressive drought. *Biogeosciences*, *16*(13), 2557–2572. Retrieved from https://bg.copernicus.org/articles/16/2557/2019/ doi: 10.5194/bg-16-2557-2019

Boike, J., Juszak, I., Lange, S., Chadburn, S., Burke, E., Overduin, P. P., ... Westermann, S. (2018). A 20-year record (1998–2017) of permafrost, active layer and meteorological conditions at a high arctic permafrost research site (bayelva, spitsbergen). *Earth System Science Data*, *10*(1), 355–390. Retrieved from https://essd.copernicus.org/articles/10/355/2018/ doi: 10.5194/essd-10-355-2018

Bultan, S., Nabel, J. E., Hartung, K., Ganzenmüller, R., Xu, L., Saatchi, S., & Pongratz, J. (2022). Tracking 21st century anthropogenic and natural carbon fluxes through model-data integration. *Nature Communications*, *13*(1), 5516. doi: 10.1038/s41467-022-32456-0

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261–

304. doi: 10.1177/0049124104268644

Burton, C. A., Renzullo, L. J., Rifai, S. W., & Van Dijk, A. I. J. M. (2023). Empirical upscaling of OzFlux eddy covariance for high-resolution monitoring of terrestrial carbon uptake in australia. *Biogeosciences*, *20*(19), 4109–4134. Retrieved from `https://bg.copernicus.org/articles/20/4109/2023/` doi: 10.5194/bg-20-4109-2023

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. Retrieved from `https://gmd.copernicus.org/articles/7/1247/2014/` doi: 10.5194/gmd-7-1247-2014

Chen, J. (2021). *Biophysical Models and Applications in Ecosystem Analysis*. East Lansing: Michigan State University Press. Retrieved from `https://muse.jhu.edu/book/82816`

Chen, J.-L., Reynolds, J. F., Harley, P. C., & Tenhunen, J. D. (1993). Coordination theory of leaf nitrogen distribution in a canopy. *Oecologia*, *93*(1), 63–69. doi: 10.1007/BF00321192

Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., … Zona, D. (2021). Representativeness of eddy-covariance flux footprints for areas surrounding ameriflux sites. *Agricultural and Forest Meteorology*, *301-302*, 108350. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0168192321000332` doi: 10.1016/j.agrformet.2021.108350

Cleverly, J., Boulain, N., Villalobos-Vega, R., Grant, N., Faux, R., Wood, C., … Eamus, D. (2013). Dynamics of component carbon fluxes in a semi-arid acacia woodland, central australia. *Journal of Geophysical Research: Biogeosciences*, *118*(3), 1168-1185. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrg.20101` doi: 10.1002/jgrg.20101

Dannenberg, M. P., Barnes, M. L., Smith, W. K., Johnston, M. R., Meerdink, S. K., Wang, X., … Biederman, J. A. (2023). Upscaling dryland carbon and water fluxes with artificial neural networks of optical, thermal, and microwave satellite remote sensing. *Biogeosciences*, *20*(2), 383–404. Retrieved from `https://bg.copernicus.org/articles/20/383/2023/` doi: 10.5194/bg-20-383-2023

De, R. (2024). *Scripts for analyses presented in "Addressing challenges in simulating inter–annual variability of gross primary production"*. Zenodo. Retrieved from `https://github.com/de-ranit/iav_gpp_p_bao` doi: 10.5281/zenodo.13729515

Desai, A. R. (2010). Climatic and phenological controls on coherent regional interannual variability of carbon dioxide flux in a heterogeneous landscape. *Journal of Geophysical Research: Biogeosciences*, *115*(G3). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JG001423` doi: 10.1029/2010JG001423

Desai, A. R., Richardson, A. D., Moffat, A. M., Kattge, J., Hollinger, D. Y., Barr, A., … Stauch, V. J. (2008). Cross-site evaluation of eddy covariance GPP and RE decomposition techniques. *Agricultural and Forest Meteorology*, *148*(6), 821-838. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0168192307003000` doi: 10.1016/j.agrformet.2007.11.012

Farquhar, G. D., von Caemmerer, S. v., & Berry, J. A. (1980). A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *planta*, *149*(1), 78–90. doi: 10.1007/BF00386231

Fatichi, S., & Ivanov, V. Y. (2014). Interannual variability of evapotranspiration and vegetation productivity. *Water Resources Research*, *50*(4), 3275-3294. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013WR015044` doi: 10.1002/2013WR015044

Fischler, M. A., & Bolles, R. C. (1981, jun). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, *24*(6), 381–395. Retrieved from https://doi.org/10.1145/358669.358692 doi: 10.1145/358669.358692

FLUXNET.org. (2024a). *FLUXNET2015 Dataset.* (https://fluxnet.org/data/fluxnet2015-dataset/, last accessed: 08-Feb-2024)

FLUXNET.org. (2024b). *IGBP classification.* (https://fluxnet.org/data/badm-data-templates/igbp-classification/, last accessed: 08-Feb-2024)

FLUXNET.org. (2024c). *Köppen Climate Classification.* (https://fluxnet.org/data/badm-data-templates/koppen-climate-classification/, last accessed: 08-Feb-2024)

Foken, T. (2008). The energy balance closure problem: An overview. *Ecological Applications*, *18*(6), 1351-1367. Retrieved from https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/06-0922.1 doi: 10.1890/06-0922.1

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* (Vol. 1, pp. 1–31).

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1), 80-91. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022169409004843 doi: 10.1016/j.jhydrol.2009.08.003

Hansen, N., Akimoto, Y., & Baudis, P. (2019, February). *CMA-ES/pycma on Github.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.2559634 doi: 10.5281/zenodo.2559634

Hansen, N., & Kern, S. (2004). Evaluating the CMA Evolution Strategy on multimodal test functions. In X. Yao et al. (Eds.), *Parallel problem solving from nature - ppsn viii* (pp. 282–291). Berlin, Heidelberg: Springer.

Henderson-Sellers, B. (1984). A new formula for latent heat of vaporization of water as a function of temperature. *Quarterly Journal of the Royal Meteorological Society*, *110*(466), 1186-1190. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49711046626 doi: https://doi.org/10.1002/qj.49711046626

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., … Thépaut, J.-N. (2023). *ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS).* (last accessed: 23-Mar-2024) doi: 10.24381/cds.adbb2d47

Horn, J. E., & Schulz, K. (2011). Identification of a general light use efficiency model for gross primary production. *Biogeosciences*, *8*(4), 999–1021. Retrieved from https://bg.copernicus.org/articles/8/999/2011/ doi: 10.5194/bg-8-999-2011

Hundecha, Y., & Merz, B. (2012). Exploring the relationship between changes in climate and floods using a model-based analysis. *Water Resources Research*, *48*(4). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR010527 doi: 10.1029/2011WR010527

Jez, J. M., Topp, C. N., Siebers, M. H., Gomez-Casanovas, N., Fu, P., Meacham-Hensold, K., … Bernacchi, C. J. (2021, 02). Emerging approaches to measure photosynthesis from the leaf to the ecosystem. *Emerging Topics in Life Sciences*, *5*(2), 261-274. Retrieved from 10.1042/ETLS20200292 doi: 10.1042/ETLS20200292

Jocher, G., Marshall, J., Nilsson, M. B., Linder, S., De Simon, G., Hörnlund, T., … Peichl, M. (2018). Impact of canopy decoupling and subcanopy advection on the annual carbon balance of a boreal scots pine forest as derived from eddy covariance. *Journal of Geophysical Research: Biogeosciences*, *123*(2), 303-325. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JG003988 doi: 10.1002/2017JG003988

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., … 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. *Journal of Geophysical Research: Biogeosciences*, *116*(G3). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JG001566 doi: 10.1029/2010JG001566

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., … Reichstein, M. (2020). Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the fluxcom approach. *Biogeosciences*, *17*(5), 1343–1365. Retrieved from https://bg.copernicus.org/articles/17/1343/2020/ doi: 10.5194/bg-17-1343-2020

Kaufmann, J., & Schering, A. (2014). Analysis of variance ANOVA. In *Wiley statsref: Statistics reference online.* John Wiley & Sons, Ltd. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat06938 doi: 10.1002/9781118445112.stat06938

Knauer, J., Cuntz, M., Smith, B., Canadell, J. G., Medlyn, B. E., Bennett, A. C., … Haverd, V. (2023). Higher global gross primary productivity under future climate with more advanced representations of photosynthesis. *Science Advances*, *9*(46), eadh9444. Retrieved from https://www.science.org/doi/abs/10.1126/sciadv.adh9444 doi: 10.1126/sciadv.adh9444

Knohl, A., Schulze, E.-D., Kolle, O., & Buchmann, N. (2003). Large carbon uptake by an unmanaged 250-year-old deciduous forest in central germany. *Agricultural and Forest Meteorology*, *118*(3), 151-167. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168192303001151 doi: 10.1016/S0168-1923(03)00115-1

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89–97. Retrieved from https://adgeo.copernicus.org/articles/5/89/2005/ doi: 10.5194/adgeo-5-89-2005

Kuppel, S., Chevallier, F., & Peylin, P. (2013). Quantifying the model structural error in carbon cycle data assimilation systems. *Geoscientific Model Development*, *6*(1), 45–55. Retrieved from https://gmd.copernicus.org/articles/6/45/2013/ doi: 10.5194/gmd-6-45-2013

Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., … Wohlfahrt, G. (2010). Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Global Change Biology*, *16*(1), 187-208. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2486.2009.02041.x doi: 10.1111/j.1365-2486.2009.02041.x

Levene, H. (1960). Contributions to probability and statistics: essays in honor of harold hotelling. In I. Olkin (Ed.), (pp. 278–292). Stanford University Press, Palo Alto.

Lin, S., Hu, Z., Wang, Y., Chen, X., He, B., Song, Z., … Yuan, W. (2023). Underestimated interannual variability of terrestrial vegetation production by terrestrial ecosystem models. *Global Biogeochemical Cycles*, *37*(4), e2023GB007696. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023GB007696 (e2023GB007696 2023GB007696) doi: 10.1029/2023GB007696

Maire, V., Martre, P., Kattge, J., Gastal, F., Esser, G., Fontaine, S., & Soussana, J.-F. (2012, 06). The coordination of leaf photosynthesis links C and N fluxes in C3 plant species. *PLOS ONE*, *7*(6), 1-15. Retrieved from 10.1371/journal.pone.0038345 doi: 10.1371/journal.pone.0038345

Mauder, M., Foken, T., & Cuxart, J. (2020). Surface-energy-balance closure over land: A review. *Boundary-Layer Meteorology*, *177*, 395–426. doi: 10.1007/

s10546-020-00529-6

McGuire, A. D., Sitch, S., Clein, J. S., Dargaville, R., Esser, G., Foley, J., … Wittenberg, U. (2001). Carbon balance of the terrestrial biosphere in the twentieth century: Analyses of CO_2, climate and land use effects with four process-based ecosystem models. *Global Biogeochemical Cycles*, *15*(1), 183-206. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000GB001298 doi: 10.1029/2000GB001298

Mengoli, G., Agustí-Panareda, A., Boussetta, S., Harrison, S. P., Trotta, C., & Prentice, I. C. (2022). Ecosystem photosynthesis in land-surface models: A first-principles approach incorporating acclimation. *Journal of Advances in Modeling Earth Systems*, *14*(1). doi: 10.1029/2021MS002767

Mengoli, G., Harrison, S. P., & Prentice, I. C. (2023). A global function of climatic aridity accounts for soil moisture stress on carbon assimilation. *EGUsphere*, *2023*, 1–19. Retrieved from https://egusphere.copernicus.org/preprints/2023/egusphere-2023-1261/ doi: 10.5194/egusphere-2023-1261

Monteith, J. (1972). Solar radiation and productivity in tropical ecosystems. *Journal of applied ecology*, *9*(3), 747–766.

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 Transactions on Geoscience and remote sensing*, *35*(6), 1380–1393.

Mäkelä, A., Pulkkinen, M., Kolari, P., Lagergren, F., Berbigier, P., Lindroth, A., … 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 Biology*, *14*(1), 92-108. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2486.2007.01463.x doi: 10.1111/j.1365-2486.2007.01463.x

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part i — a discussion of principles. *Journal of Hydrology*, *10*(3), 282-290. Retrieved from https://www.sciencedirect.com/science/article/pii/0022169470902556 doi: 10.1016/0022-1694(70)90255-6

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., … Jung, M. (2024). X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X. *EGUsphere*, *2024*, 1–51. Retrieved from https://egusphere.copernicus.org/preprints/2024/egusphere-2024-165/ doi: 10.5194/egusphere-2024-165

Papale, D., Black, T. A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., … Ráduly, B. (2015). Effect of spatial sampling from european flux towers for estimating carbon and water fluxes with artificial neural networks. *Journal of Geophysical Research: Biogeosciences*, *120*(10), 1941-1957. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JG002997 doi: 10.1002/2015JG002997

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., … others (2020, July). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, *7*(1), 225. Retrieved 2023-11-15, from https://www.nature.com/articles/s41597-020-0534-3 doi: 10.1038/s41597-020-0534-3

PCC. (2008). Pearson's correlation coefficient. In *Encyclopedia of public health [kirch w. (ed.)]* (pp. 1090–1091). Dordrecht: Springer Netherlands. Retrieved from https://doi.org/10.1007/978-1-4020-5614-7_2569 doi: 10.1007/978-1-4020-5614-7_2569

Plummer, S. (2006). On validation of the MODIS gross primary production product. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(7), 1936-1938. doi: 10.1109/TGRS.2006.872521

Prentice, I. C., Dong, N., Gleason, S. M., Maire, V., & Wright, I. J. (2014). Balancing the costs of carbon gain and water transport: testing a new theoretical framework for plant functional ecology. *Ecology Letters*, *17*(1), 82-91. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.12211 doi: 10.1111/ele.12211

Prentice, I. C., Liang, X., Medlyn, B. E., & Wang, Y.-P. (2015). Reliable, robust and realistic: the three r's of next-generation land-surface modelling. *Atmospheric Chemistry and Physics*, *15*(10), 5987–6005. Retrieved from https://acp.copernicus.org/articles/15/5987/2015/ doi: 10.5194/acp-15-5987-2015

Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly weather review*, *100*(2), 81–92.

Python Core Team. (2021). Python: A dynamic, open source programming language [Computer software manual]. Retrieved from https://www.python.org/ (Python version 3.10)

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., . . . Valentini, R. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*, *11*(9), 1424-1439. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2486.2005.001002.x doi: 10.1111/j.1365-2486.2005.001002.x

Richardson, A. D., Hollinger, D. Y., Aber, J. D., Ollinger, S. V., & Braswell, B. H. (2007). Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange. *Global Change Biology*, *13*(4), 788-803. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2486.2007.01330.x doi: 10.1111/j.1365-2486.2007.01330.x

Rubel, F., Brugger, K., Haslinger, K., & Auer, I. (2017, 04). The climate of the european alps: Shift of very high resolution Köppen-Geiger climate zones 1800 – 2100. *Meteorologische Zeitschrift*, *26*(2), 115-125. Retrieved from http://dx.doi.org/10.1127/metz/2016/0816 doi: 10.1127/metz/2016/0816

Ruehr, S., Keenan, T. F., Williams, C., Zhou, Y., Lu, X., Bastos, A., . . . Terrer, C. (2023). Evidence and attribution of the enhanced land carbon sink. *Nature Reviews Earth & Environment*, *4*(8), 518–534. doi: 10.1038/s43017-023-00456-3

Running, S. W., Thornton, P. E., Nemani, R., & Glassy, J. M. (2000). Global terrestrial gross and net primary productivity from the earth observing system. In O. E. Sala, R. B. Jackson, H. A. Mooney, & R. W. Howarth (Eds.), *Methods in ecosystem science* (pp. 44–57). New York, NY: Springer New York. Retrieved from 10.1007/978-1-4612-1224-9_4 doi: 10.1007/978-1-4612-1224-9_4

Schimel, D. (2001). Preface. In E.-D. Schulze et al. (Eds.), *Global biogeochemical cycles in the climate system* (p. xvii-xxi). San Diego: Academic Press. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780126312607500029 doi: 10.1016/B978-012631260-7/50002-9

Schimel, D., Stephens, B. B., & Fisher, J. B. (2015). Effect of increasing CO_2 on the terrestrial carbon cycle. *Proceedings of the National Academy of Sciences*, *112*(2), 436-441. Retrieved from https://www.pnas.org/doi/abs/10.1073/pnas.1407302112 doi: 10.1073/pnas.1407302112

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th python in science conference.*

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., . . . Myneni, R. (2015). Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, *12*(3), 653–679. Retrieved from https://bg.copernicus.org/articles/12/653/2015/ doi: 10.5194/bg-12-653-2015

Stocker, B. D., Wang, H., Smith, N. G., Harrison, S. P., Keenan, T. F., Sandoval, D., . . . Prentice, I. C. (2020). P-model v1.0: an optimality-based light use efficiency model for simulating ecosystem gross primary production. *Geoscientific Model Development*, *13*(3), 1545–1581. Retrieved from https://gmd.copernicus.org/articles/13/1545/2020/ doi: 10.5194/gmd-13-1545-2020

Tamrakar, R., Rayment, M. B., Moyano, F., Mund, M., & Knohl, A. (2018). Implications of structural diversity for seasonal and annual carbon dioxide fluxes in two temperate deciduous forests. *Agricultural and Forest Meteorology*, *263*, 465-476. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168192318302934 doi: 10.1016/j.agrformet.2018.08.027

Thoning, K., Crotwell, A., & Mund, J. (2021). Atmospheric carbon dioxide dry air mole fractions from continuous measurements at mauna loa, hawaii, barrow, alaska, american samoa and south pole. 1973-2020 version 2021-08-09. *National Oceanic and Atmospheric Administration (NOAA), Global Monitoring Laboratory (GML), Boulder, Colorado, USA*. doi: 10.15138/yaf1-bk21

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., ... Papale, D. (2016). Predicting carbon dioxide and energy fluxes across global fluxnet sites with regression algorithms. *Biogeosciences*, *13*(14), 4291–4313. Retrieved from https://bg.copernicus.org/articles/13/4291/2016/ doi: 10.5194/bg-13-4291-2016

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. *Hydrology and Earth System Sciences*, *22*(7), 4061–4082. Retrieved from https://hess.copernicus.org/articles/22/4061/2018/ doi: 10.5194/hess-22-4061-2018

Turner, D. P., Ritts, W. D., Styles, J. M., Yang, Z., Cohen, W. B., Law, B. E., & Thornton, P. E. (2006). A diagnostic carbon flux model to monitor the effects of disturbance and interannual variation in climate on regional NEP. *Tellus B: Chemical and Physical Meteorology*, *58*(5), 476-490. Retrieved from 10.1111/j.1600-0889.2006.00221.x doi: 10.1111/j.1600-0889.2006.00221.x

Van Thieu, N. (2023, August). *Permetrics: A framework of performance metrics for machine learning models.* Zenodo. Retrieved from https://github.com/thieu1995/permetrics doi: 10.5281/zenodo.3951205

Van Thieu, N., & Mirjalili, S. (2023). Mealpy: An open-source library for latest meta-heuristic algorithms in python. *Journal of Systems Architecture*. doi: 10.1016/j.sysarc.2023.102871

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. doi: 10.1038/s41592-019-0686-2

Walther, S., Besnard, S., Nelson, J. A., El-Madany, T. S., Migliavacca, M., Weber, U., ... Jung, M. (2022). Technical note: A view from space on global flux towers by MODIS and landsat: the FluxnetEO data set. *Biogeosciences*, *19*(11), 2805–2840. Retrieved from https://bg.copernicus.org/articles/19/2805/2022/ doi: 10.5194/bg-19-2805-2022

Walther, S., Nelson, J. A., Besnard, S., & Weber, U. (2023). *The FluxnetEO dataset (MODIS).* ICOS ERIC - Carbon Portal. Retrieved from https://meta.icos-cp.eu/collections/mdtEHjyujUDsC9vgMv5eeH8B doi: 10.18160/0KWD-3RRW

Wang, H., Prentice, I. C., Keenan, T. F., Davis, T. W., Wright, I. J., Cornwell, W. K., ... Peng, C. (2017). Towards a universal model for carbon dioxide uptake by plants. *Nature plants*, *3*(9), 734–741. doi: 10.1038/s41477-017-0006-8

Wohlfahrt, G., & Gu, L. (2015). The many meanings of gross photosynthesis and their implication for photosynthesis research from leaf to globe. *Plant, Cell & Environment*, *38*(12), 2500-2507. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/pce.12569 doi:

10.1111/pce.12569
Wu, J., van der 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*(1), 13–28. Retrieved from https://bg.copernicus.org/articles/9/13/2012/ doi: 10.5194/bg-9-13-2012
Xiao, X., Jin, C., & Dong, J. (2014). Gross primary production of terrestrial vegetation. In J. M. Hanes (Ed.), *Biophysical applications of satellite remote sensing* (pp. 127–148). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from 10.1007/978-3-642-25047-7_5 doi: 10.1007/978-3-642-25047-7_5
Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L. L., Baldocchi, D., ... Wofsy, S. C. (2007). Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. *Agricultural and Forest Meteorology*, *143*(3), 189-207. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168192306003881 doi: 10.1016/j.agrformet.2006.12.001
Zhang, W., Nelson, J. A., Miralles, D. G., Mauder, M., Migliavacca, M., Poyatos, R., ... Jung, M. (2024). A new post-hoc method to reduce the energy imbalance in eddy covariance measurements. *Geophysical Research Letters*, *51*(2), e2023GL107084. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023GL107084 (e2023GL107084 2023GL107084) doi: 10.1029/2023GL107084
Zomer, R. J., Xu, J., & Trabucco, A. (2022). Version 3 of the global aridity index and potential evapotranspiration database. *Scientific Data*, *9*(1), 409. doi: 10.1038/s41597-022-01493-1
Zou, H., Chen, J., Li, X., Abraha, M., Zhao, X., & Tang, J. (2024). Modeling net ecosystem exchange of CO_2 with gated recurrent unit neural networks. *Agricultural and Forest Meteorology*, *350*, 109985. Retrieved from https://www.sciencedirect.com/science/article/pii/S016819232400100X doi: 10.1016/j.agrformet.2024.109985
Zscheischler, J., Fatichi, S., Wolf, S., Blanken, P. D., Bohrer, G., Clark, K., ... Seneviratne, S. I. (2016). Short-term favorable weather conditions are an important control of interannual variability in carbon and water fluxes. *Journal of Geophysical Research: Biogeosciences*, *121*(8), 2186-2198. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JG003503 doi: 10.1002/2016JG003503