Agricultural and Forest Meteorology 372 (2025) 110701

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Global distribution pattern in characteristics of gross primary productivity response to soil water availability

Shanning Bao [a,*], Nuno Carvalhais [b,c,d], Jian Xu [a], Jing M. Chen [e,f], Yang Lei [a], Gegen Tana [a], Changgui Lin [a], Jiancheng Shi [a]

[a] National Space Science Center, Chinese Academy of Sciences, 100190, Beijing, PR China
[b] Department for Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, 07745, Jena, Germany
[c] Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[d] ELLIS Unit Jena, 07745, Jena, Germany
[e] School of Geographical Science, Key Laboratory for Humid Subtropical Eco-Geographical Processes of the Ministry of Education, Fujian Normal University, Fuzhou 350008, PR China
[f] Department of Geography and Planning, University of Toronto, 100St. George St., Toronto, Ontario M5S3G3, Canada

ARTICLE INFO

Keywords:
Gross primary productivity
Soil water availability
Response curve feature
Spatial distribution
Lag effect
Hybrid model

ABSTRACT

Understanding how carbon assimilation rates respond to water availability is crucial for diagnosing global carbon and water cycles. This study aims to investigate characteristics and drivers of gross primary productivity (GPP) responses to soil water availability using three parameters from a light-use-efficiency (LUE) model: $\boldsymbol{W_L}$, $\boldsymbol{k_W}$ and $\boldsymbol{\alpha_W}$, representing the inflection point, slope and lag effect of GPP response to soil water availability changes, respectively. We followed a hybrid modeling approach coupling an artificial neural network with the LUE model to derive model parameters and examine intricate relationships between these parameters and features characterizing climate, vegetation, nutrient deposition, soil properties and elevation across 196 eddy covariance sites. Relationships between the LUE model parameters and observed ecosystem properties were analyzed using partial dependence plots and Shapley additive explanation dependence plots. Our results revealed significant statistical differences in parameters across plant functional types. Specifically, forests exhibited lower inflection points, responding more steeply and immediately to water availability changes, contrasting with smoother and lagged responses from open shrubs. Vegetation seasonality, represented by variability of enhanced vegetation index (EVI) and seasonal EVI, was the most influential noncategorical factor, followed by soil properties. Notably, the relationships were predominantly nonlinear. Additionally, older forest ecosystems generally showed lower vulnerability while responding more steeply to relative soil water availability changes than younger forests. While aridity was less influential on parameter variability than anticipated, aridity seasonality was a primary driver for the inflection point. High temperatures and substantial diurnal and annual temperature ranges were linked to pronounced lag effects. Despite these findings, challenges remain regarding model accuracy on annual scales, parameter uncertainties and interactions between features. Overall, this study underscores the spatial heterogeneity of GPP responses to soil water availability and highlights the importance of considering variability in model parameters and GPP sensitivities across space and time.

1. Introduction

Soil water availability is a primary driver of temporal variability in carbon assimilation rates (Jung et al., 2017; Humphrey et al., 2021). Water limitation is reported to have an increasing impact on ecosystem processes, especially on gross primary productivity (GPP) (Gampe et al., 2021). Many modeling studies have explored the spatial patterns of GPP to study its sensitivity to water supply (the degree of GPP response to water supply), using various photosynthesis models (Jung et al., 2017; Humphrey et al., 2021; Li et al., 2023; Stocker et al., 2018). However, the factors underlying these patterns remain unclear. Additionally, critical aspects related to the shape of the GPP sensitivity to changes in water availability, such as the inflection point, response slope and lag effect, are not well understood. Investigating the dominant factors

* Corresponding author.
E-mail address: baoshanning@nssc.ac.cn (S. Bao).

https://doi.org/10.1016/j.agrformet.2025.110701
Received 6 March 2024; Received in revised form 6 June 2025; Accepted 16 June 2025
Available online 27 June 2025

influencing the GPP responses to water availability is essential for more accurate quantification of climate controls on GPP and future carbon uptake under climate change (Greve et al., 2014; Monteith, 1972).

In photosynthesis models, the sensitivity of GPP to environmental conditions is quantified through model parameters that describe how GPP responds to changing meteorological drivers. For example, the MODIS GPP product, based on a light use efficiency (LUE) model, includes four parameters related to temperature (T) and vapor pressure deficit (VPD) that define the response slopes and thresholds of GPP to T and VPD variations (Running et al., 2004). Another example is the optimal soil water or soil moisture threshold, a key parameter in photosynthesis models, which represents the inflection point of GPP response to soil water availability (Yuan et al., 2014; Stocker et al., 2020; Lawrence et al., 2019). These model parameters regulate the simulated responses of GPP to changes in environmental conditions and are key for analyzing and comparing the level of GPP sensitivity, i.e., whether GPP is significantly affected (i.e., high sensitivity) or remains stable (i.e., low sensitivity). By examining the relationships between the different model parameters and observed ecosystem characteristics (such as vegetation, climate and soil properties), we can better understand how GPP-climate sensitivities are shaped by local conditions.

Model parameters related to GPP sensitivities to soil water and other factors were usually determined based on plant functional types (PFTs) or calibrated locally using carbon flux data from eddy covariance (EC) sites such as those from FLUXNET. For example, the MODIS GPP model relies on a PFT-specific lookup table. Similarly, many photosynthesis models based on the Farquhar et al. (1980) framework also adopt PFT-based parameterization. A representative case was the photosynthesis module in the community land model (CLM) series. It applied the PFT-based parameterization approach for the specific leaf area and leaf nitrogen for carboxylation, while fixing water-stress-related parameters globally. Other models, including dynamic global vegetation models (DGVMs) in Trends in Net Land-Atmosphere Carbon Exchanges (TRENDY) and global climate models in the Coupled Model Intercomparison Project Phase 6 (CMIP6) ensemble, also used PFT-based or globally uniform parameterizations (Sitch et al., 2008; Eyring et al., 2016). Such approaches ignore the intra-PFT parameter variability, which can reduce simulation accuracy and higher uncertainties (Groenendijk et al., 2011). Although the calibration approach can improve model performance by capturing local parameter variability, its global application is limited by the sparse coverage of observational data. As a result, many studies calibrate parameters at selected sites and extrapolate them based on PFT classification (Yuan et al., 2019), site-similarity (Carvalhais et al., 2010) or through globally unified parameters (Wang et al., 2017). However, these extrapolation approaches assume that photosynthesis model parameters vary with PFT only, neglecting other possible dominating factors, e.g., climate and soil properties.

Previous studies have explored the relationships between photosynthesis model parameters and ecosystem properties, revealing strong associations with vegetation traits and climatic conditions (Madani et al., 2014; Wang et al., 2010; Croft and al., 2017). These findings suggested that the parameter variability reflects the response of vegetation to local edaphoclimatic conditions. However, the specific drivers and spatial patterns of parameters governing GPP responses to soil water remain underexplored. While prior research has emphasized the role of root-zone water storage capacity in modulating GPP (Stocker et al., 2023) and noted the complexity of vegetation sensitivity to soil moisture (Li et al., 2023), it has not directly addressed the factors shaping these model parameters. Some studies, such as those by Horn and Schulz (2011) and Peaucelle et al. (2019), have attempted to predict photosynthesis model parameters using plant traits and climate features to improve model extrapolation ability. However, these studies based on calibrated parameters were hampered by parameter equifinality and limited in detecting the relationship between parameters and ecosystem features.

Hybrid models coupling traditional process models with machine learning techniques enable the exploration of parameter variability. Bao et al. (2023) proposed a simultaneous model parameter inversion and extrapolation (SPIE) approach which can parametrize an LUE model based on ecosystem properties. The approach embeds the traditional LUE model into a neural network structure, which can output GPP and model parameters simultaneously. Within this approach, relying on a cross-validation strategy to minimise equifinality and over-fitting issues, we here explore the variability of the parameters determining GPP responses to soil water and its relationship with ecosystem properties.

In this study, we focus on analysing three parameters characterising the response curve of GPP to soil water availability in a LUE model ($\boldsymbol{W_I}$, $\boldsymbol{k_W}$ and $\boldsymbol{\alpha_W}$, see Section 2.1) with the aim of exploring the ecosystem features explaining their spatial variability. First, we demonstrate that the LUE model effectively captures the spatiotemporal variability of GPP, within which the GPP response to soil moisture is shaped by these three parameters: the inflection point of the response curve ($\boldsymbol{W_I}$), the response slope ($\boldsymbol{k_W}$) and lag coefficient of soil water effect ($\boldsymbol{\alpha_W}$). The hypothesis is that the emerging sensitivity of GPP to soil water availability can be explained by plant functional type, bioclimatic conditions and soil properties associated with soil water holding capacity. To test the hypothesis, we applied the SPIE hybrid modeling approach (see Section 2.4) to quantify relationships between these parameters and various features representing vegetation, climate, atmospheric nutrient deposition, terrain and soil properties (introduced in Section 2.3 and see the flowchart in Section 2.6). Two machine learning interpretation approaches (see Section 2.5) were used to address the parameter-feature relationship and identify the dominant features. Furthermore, these parameters at the global scale were produced and their spatial patterns were analyzed. Ultimately, this study steps forward in methods to enhance the understanding of how ecosystem features influence the sensitivity of carbon assimilation rates to soil water availability.

2. Data and methods

2.1. Light use efficiency model

We selected the LUE model from a large ensemble of models with various combinations of environmental drivers and functions representing GPP responses to these drivers (Bao et al., 2022). It was evaluated against observations across different site groups and plant-climate types at daily, weekly, monthly and annual scales, and the results showed that the model significantly outperforms other models. The model has a typical LUE structure (as shown in Eq. (1)) which defines GPP as the product of the maximum light use efficiency ($\boldsymbol{\varepsilon_{max}}$), absorbed photosynthetically active radiation (APAR) and environmental sensitivity functions (refer to fT, fVPD, fCa, fW, fL and fCI in Eq. (1)).

$$\mathrm{GPP} = \boldsymbol{\varepsilon_{max}} \cdot \mathrm{APAR} \cdot f\mathrm{T} \cdot f\mathrm{VPD} \cdot f\mathrm{C_a} \cdot f\mathrm{W} \cdot f\mathrm{L} \cdot f\mathrm{CI} \quad (1)$$

Above (Eq. (1)), APAR represents the product of the incoming photosynthetically active radiation (PAR) and the fraction of the absorbed photosynthetically active radiation (FAPAR).

The environmental sensitivity functions quantify how key environmental factors influence the photosynthesis process. Specifically: fT represents the sensitivity of GPP to T; fVPD corresponds to the sensitivity of GPP to VPD; $f\mathrm{C_a}$ denotes the sensitivity to atmospheric CO_2 concentration ($\mathrm{C_a}$); fW refers to the sensitivity to the relative soil water availability index (W); and fL and fCI represent the sensitivity to light intensity (L) and cloudiness index (CI), respectively. Together, these functions describe the modulation of GPP by temperature, VPD, CO_2 level, soil water availability, light saturation and diffuse radiation, as formulated in Eqs. (2)-(9). The corresponding response curves for these functions are illustrated in Figure S1.

$$fT = \frac{2e^{-\left(T_f - \boldsymbol{T_{opt}}\right)/\boldsymbol{k_T}}}{1 + e^{\left(-\left(T_f - \boldsymbol{T_{opt}}\right)/\boldsymbol{k_T}\right)^2}} \quad (2)$$

$$fVPD = e^{\boldsymbol{\kappa}\left(\frac{\boldsymbol{C_{a0}}}{C_a}\right)^{\boldsymbol{C_\kappa}} VPD} \quad (3)$$

$$C_a = 1 + \frac{C_a - \boldsymbol{C_{a0}}}{C_a - \boldsymbol{C_{a0}} + \boldsymbol{C_m}} \quad (4)$$

$$fW = \frac{1}{1 + e^{\boldsymbol{k_W}\left(W_f - \boldsymbol{W_I}\right)}} \quad (5)$$

$$fL = \frac{1}{1 + \boldsymbol{\gamma}\cdot PAR\cdot FAPAR} \quad (6)$$

$$fCI = CI^{\boldsymbol{\mu}} \quad (7)$$

$$T_f(t) = (1 - \boldsymbol{\alpha_T})\cdot T(t) + \boldsymbol{\alpha_T}\cdot T_f(t-1) \quad (8)$$

$$W_f(t) = (1 - \boldsymbol{\alpha_W})\cdot W(t) + \boldsymbol{\alpha_W}\cdot W_f(t-1) \quad (9)$$

All sensitivity functions range from zero to one except fC_a, which represents the fertilization effect of CO_2 on GPP (i.e., $fC_a \geq 1$). The other sensitivity functions represent the environmental stress on GPP (i.e., $0 \leq fT \leq 1$, $0 \leq fVPD \leq 1$, $0 \leq fW \leq 1$, $0 \leq fL \leq 1$, $0 \leq fCI \leq 1$).

The LUE model parameters (in **bold** and *italic*) represent the characteristics of GPP response curves (Figure S1) to the forcing variables. *t* refers to the time step. Eqs. (8)-(9) are the lag functions for GPP responses to temperature in the boreal climate and to soil moisture in the arid climate, respectively. The units and ranges of these parameters are listed in Table S1, Table 1.

The study focused on three parameters of the LUE model that characterize the soil water response of GPP (hereafter referred to as water-related parameters): $\boldsymbol{W_I}$, $\boldsymbol{k_W}$ and $\boldsymbol{\alpha_W}$. These three parameters respectively represent the inflection point, the slope of water response function and the lagged effect in GPP responses to W changes (see Fig. 1 for the variation of *f*W with these three parameters). Specifically, $\boldsymbol{W_I}$ is the water limitation threshold at which GPP is reduced by half. A lower $\boldsymbol{W_I}$ indicates that GPP remains near its potential until W becomes very low, reflecting lower vulnerability to W, while a higher $\boldsymbol{W_I}$ represents greater vulnerability. Meanwhile, $\boldsymbol{k_W}$ quantifies the steepness of GPP response to W changes. A lower $\boldsymbol{k_W}$, or a higher absolute value of $\boldsymbol{k_W}$, signifies a sharper GPP response under water stress, whereas a higher $\boldsymbol{k_W}$ corresponds to a smoother response. Similarly, $\boldsymbol{\alpha_W}$ indicates the degree of lag in the GPP response to W changes, with higher values ($\geq$0.8) representing a more pronounced lag and a slower adjustment of the photosynthetic rate to soil moisture fluctuations. In contrast, lower $\boldsymbol{\alpha_W}$ or $\boldsymbol{\alpha_W}$=0 means an immediate response. Together, these three parameters were used to diagnose the vulnerability of GPP to W, the sharpness of its response and the degree of lag in its reaction to changes in W.

Table 1
List of the LUE model parameters.

Parameters	Meanings	Range	Units	References
$\boldsymbol{\varepsilon_{max}}$	Maximum light use efficiency	0 – 10	$gC\bullet MJ^{-1}$	(Bao et al., 2022)
$\boldsymbol{T_{opt}}$	Optimal air temperature	5 – 35	°C	(Horn and Schulz, 2011)
$\boldsymbol{k_T}$	Response slope to temperature changes	1 – 20	–	
$\boldsymbol{\alpha_T}$	Lag coefficient for temperature effect	0.0 – 0.9	–	
$\boldsymbol{\kappa}$	Response slope to vapor pressure deficit changes	-10^{-1} – -10^{-4}	kPa^{-1}	(Bao et al., 2022)
$\boldsymbol{C_\kappa}$	Response slope to atmospheric CO_2 concentration changes	0 – 10	–	
$\boldsymbol{C_{a0}}$	Minimum optimal atmospheric CO_2 concentration	340 – 390	ppm	
$\boldsymbol{C_m}$	CO_2 fertilization intensity indicator	100 – 4000	ppm	
$\boldsymbol{W_I}$	Inflection point of response curve to soil water	0.01 – 0.99	$m^3\bullet m^{-3}$	(Horn and Schulz, 2011)
$\boldsymbol{k_W}$	Response slope to soil water changes	−30 – −5	–	
$\boldsymbol{\alpha_W}$	Lag coefficient for soil water effect	0.0 – 0.9	–	
$\boldsymbol{\gamma}$	Light saturation curvature indicator	0 – 1	$MJ^{-1}\bullet m^2\bullet d$	(Bao et al., 2022)
$\boldsymbol{\mu}$	Response slope to cloudiness index changes	10^{-3} – 1	–	
AWC	Available soil water content	1–1000	mm	
$\boldsymbol{\theta}$	Soil water content decay coefficient	0.001–0.2	–	

2.2. Forcing data

The forcing data (listed in Table S1) for the LUE model were collected from 196 EC sites (Figure S2) provided by FLUXNET (https://fluxnet.org/). These sites encompass eleven plant functional types (Table S2), including evergreen needleleaf forests (ENF), grasslands (GRA), deciduous broadleaf forests (DBF), croplands (CRO), wetlands (WET), evergreen broadleaf forests (EBF), open shrublands (OSH), woody savannas (WSA), mixed forests (MF), savannas (SAV) and closed shrublands (CSH), organized by site number.

Among the forcing data shown in Eqs. (1)-(9), T (°C) and VPD (hPa) were directly obtained from the site observations. PAR was derived from the observed daily global radiation (R_g, $MJ\bullet m^{-2}\bullet d^{-1}$) multiplied by 0.45 (representing the fraction of visible light bands). FAPAR was represented by the normalized difference vegetation index (NDVI) for their linear relationship. The NDVI was calculated using the red and near-infrared bands of MODIS directional hemispherical reflectance product at the resolution of 500 m (MCD43A3), averaged from the nearest four pixels. The gaps in temporal NDVI at each site were filled with the *FluxnetEO* dataset (Walther et al., 2022). CI was equal to one minus the global radiation by the potential radiation, which is linearly correlated to the fraction of diffuse radiation to the global radiation. C_a (ppm) was estimated based on the linear interpolation approach using the monthly mean C_a for the period of the observational meteorological data at each EC site. The sources of these forcing data along with the processing steps are detailed in Table S2.

W was calculated based on potential evapotranspiration (ET_p, mm) and precipitation (P, mm) using two parameters, the maximum available soil water content (***AWC***) and ET-decay coefficient ($\boldsymbol{\theta}$). These calculations are detailed in Eqs. (10)–(13), adapted from previous studies (Tramontana et al., 2016; Trautmann et al., 2022).

$$P_{min} = \min(P(t), \boldsymbol{AWC} - W_a(t-1)) \quad (10)$$

$$ET_{sim}(t) = \min\left(ET_p(t), \boldsymbol{\theta}\cdot(W_a(t-1) + P_{min})\right) \quad (11)$$

$$W_a(t) = W_a(t-1) - ET_{sim}(t) + P_{min} \quad (12)$$

$$W = W_a/\boldsymbol{AWC} \quad (13)$$

This calculation assumes a 'bucket model' describing the dynamics of water available for plants, where water is lost through evapotranspiration and gained from precipitation. At time step *t*, P_{min}, the amount of daily precipitation made available to vegetation, is the smaller value between precipitation (P, mm) and the difference between the water availability indicator (W_a, mm) at the previous time step (*t*-1) and the parameter ***AWC*** (mm, a constant per pixel/site). Simulated evapotranspiration (ET_{sim}) at time *t* is equal to the smaller value between ET_p at time *t* and the product of $\boldsymbol{\theta}$ and the sum of W_a at time *t*-1 and P_{min}. W_a

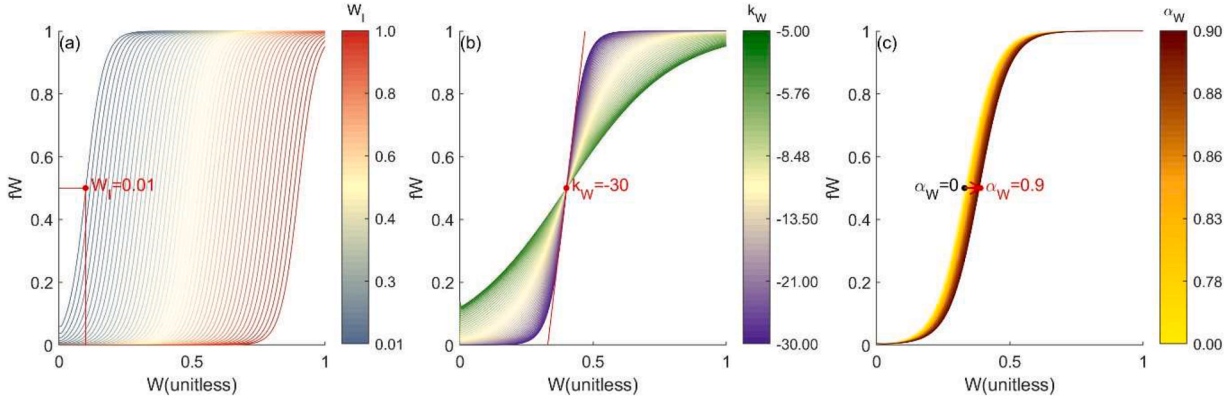

Fig. 1. The sensitivity function of GPP to W (*f*W) changes with three water-related parameters ($\boldsymbol{W_I}$, $\boldsymbol{k_W}$ and $\boldsymbol{\alpha_W}$). (a) $\boldsymbol{W_I}$ represents the inflection point of *f*W, which is equal to the W resulting in *f*W=0.5, exemplified as the red point shown in (a); (b) $\boldsymbol{k_W}$ is the slope of *f*W, representing the response gradient to W changes; (c) $\boldsymbol{\alpha_W}$ represents the degree of the lag effect. The *f*W response to W is instantaneous when $\boldsymbol{\alpha_W}$ is close to zero, as shown by the blue curve on the left. When α_W is ≥0.8, the *f*W response to W is lagged by more than one day.

represents the available water content for the plants and is derived from ET_{sim} and P_{min}. The two parameters, ***AWC*** and $\boldsymbol{\theta}$, were predicted along with the LUE model parameters. Snow sublimation was accounted for using the approach described by Trautmann et al. (2022). Before estimation, all state variables were initialized with a spin-up of five years. W is equal to normalized WAI, ranging from zero to one. In this study, we used W as a metric to represent the soil water availability for carbon assimilation.

To evaluate model accuracy and to invert model parameters (introduced in Section 2.4), we collected GPP (hereafter referred to as GPP_{obs}, $gC•m^{-2}•d^{-1}$) and evapotranspiration (hereafter referred to as ET_{obs}, mm). GPP_{obs} was estimated using the observed net ecosystem exchange (NEE, $MJ•m^{-2}•d^{-1}$) based on the nighttime partitioning approach (Reichstein et al., 2005). ET_{obs} was estimated from latent heat flux (LE, $MJ•m^{-2}•d^{-1}$) and T. The uncertainties of NEE and LE were also collected.

2.3. Collected features

We collected both site-level and global features (listed in Table 2) to predict parameters based on SPIE (introduced in Section 2.4) at local and global scales. These features included PFT, main climate types, bioclimatic variables (BIO1–19), aridity index (AI1–2), vegetation index features (VIF1–8), forest age, atmospheric nutrient deposition, soil properties and elevation (see definitions and units in Table 2). At site-level, PFT and elevation were collected from the site information on the FLUXNET website. The climate types were extracted from Koeppen Geiger climate classification types. The 19 bioclimatic variables and aridity index were calculated based on monthly temperature, precipitation and evapotranspiration products (Fick and Hijmans, 2017; Zomer et al., 2022) from 1970 to 2000 with a 1 km resolution provided by WorldClim. The vegetation index features were calculated based on the enhanced vegetation index (EVI, multiplied by 100), which was computed using the red, near-infrared and blue bands of Landsat 5, 7 and 8 from 1986–2015 (downloaded via Google Earth Engine) around each site with a window size of 1 km. The forest age was extracted from the global forest age product at 1 km resolution which was estimated based on inventories, biomass and climate data (Besnard and al., 2021). The ages of cropland and grassland were set to be zero since most of them are annual plants and their sensitivities do not change with ages. The atmospheric nitrogen and phosphorus deposition datasets were extracted from the nearest pixel of the modelled global nutrient deposition product from 1986–2015 (Rodhe et al., 2002). Moreover, we collected 16 soil property variables for the surface soil layer at each site from the mean of the nearest pixels of SoilGrids (Poggio et al., 2021) within a window of 1 km.

At the global scale, the input features were collected from the same sources as the local scale except for PFT and elevation. We extracted PFT from MODIS 1 km land cover type product (MCD12Q1) in 2001 and applied the elevation product derived from the SRTM data at 0.0083 degrees. All global input features were resampled to the spatial resolution of 1 km based on the linear interpolation approach.

All input features were normalized and then standardized by subtracting the mean and dividing by the standard deviation. These steps ensure efficient learning, faster convergence, and improved performance of the neural network by maintaining numerical stability and avoiding issues caused by varying input scales. Categorical variables were processed using the one-hot encoding approach (Brownlee, 2020), a standard approach for converting categorical data into a numerical format compatible with neural network processing.

2.4. Simultaneous parameter inversion and extrapolation (SPIE) approach

We applied the SPIE approach (Bao et al., 2023) to determine the water-related parameters and other LUE model parameters. With this approach, the model parameters can be inverted based on observational data and the collected features, allowing us to investigate how ecosystem features shape parameter variation. These relationships can then be used to extrapolate model parameters based on the features of the target area. Here, the detailed workflow involved coupling the LUE model with an artificial neural network (3 hidden layers and 64 neurons per layer). All LUE model parameters and WAI parameters were predicted using the neural network with the standardized input features introduced in Section 2.3. The predicted parameters were then used to parameterize the LUE model and WAI to estimate GPP and ET with the temporal forcing data from EC sites (as explained in Section 2.2). We optimized and constrained the neural network by the cost function including GPP errors (sum of the squared difference between GPP_{obs} and GPP_{sim} divided by the standard deviation of net ecosystem exchange measurements), ET errors (sum of the squared difference between ET_{obs} and ET_{sim} divided by the uncertainties of LE measurements) and some constraints on sensitivity functions (see equations S1-S5), as explained in section S1. The outputs of the whole framework include the predicted parameters (only spatially changed) and the estimated temporal GPP and ET (see the flowchart in Bao et al., 2023). Since the neural network was trained against the observations and effectively captured the relationship between the ecosystem features and model parameters, it was applied to analyze the spatial variability and extrapolate the parameters.

To mitigate the effect of different training datasets on the derived relationship, we trained the neural network ten times based on different groups of datasets. Each training dataset included all PFTs and climate types of the sites comprehensively to reduce extrapolation bias. GPP_{sim}

Table 2
Definition of input features.

Name	Short names	Definitions	Class
Categorical variables	CRO	Croplands	Vegetation feature
	CSH	Closed shrublands	
	DBF	Deciduous broadleaf forests	
	EBF	Evergreen broadleaf forests	
	ENF	Evergreen needleleaf forests	
	GRA	Grasslands	
	MF	Mixed forests	
	OSH	Open shrublands	
	SAV	Savannas	
	WET	Wetlands	
	WSA	Woody savannas	
	Boreal	Boreal climate in Koeppen-Geiger climate classification	Climate feature
	Arid	Arid climate in Koeppen-Geiger climate classification	
Bioclimatic variables	BIO1	Annual mean temperature in °C	
	BIO2	Mean diurnal range in °C (mean of monthly maximum temperature minus minimum temperature)	
	BIO3	Isothermality (BIO2 divided by BIO7 and 100)	
	BIO4	temperature seasonality in 10^2 °C (standard deviation of temperature multiplied by 100)	
	BIO5	Max temperature of warmest month in °C	
	BIO6	Min temperature of coldest month in °C	
	BIO7	Temperature annual range in °C (BIO5 minus BIO6)	
	BIO8	Mean temperature of wettest quarter in °C	
	BIO9	Mean temperature of driest quarter in °C	
	BIO10	Mean temperature of warmest quarter in °C	
	BIO11	Mean temperature of coldest quarter in °C	
	BIO12	Annual precipitation in mm	
	BIO13	Precipitation of wettest month in mm	
	BIO14	Precipitation of driest month in mm	
	BIO15	Precipitation seasonality (coefficient of variation)	
	BIO16	Precipitation of wettest quarter in mm	
	BIO17	Precipitation of driest quarter in mm	
	BIO18	Precipitation of warmest quarter in mm	
	BIO19	Precipitation of coldest quarter in mm	
Aridity features	AI1	Mean annual aridity index (ratio between mean annual precipitation and potential evapotranspiration)	
	AI2	seasonality of aridity index (standard deviation of mean monthly aridity index)	
Vegetation index features	VIF1	Annual mean EVI	Vegetation feature
	VIF2	EVI seasonality (standard deviation of monthly EVI relative to the mean EVI)	
	VIF3	Max EVI of warmest month	
	VIF4	Min EVI of coldest month	
	VIF5	Mean EVI of wettest quarter	
	VIF6	Mean EVI of driest quarter	
	VIF7	Mean EVI of warmest quarter	
	VIF8	Mean EVI of coldest quarter	
	Age	Forest age (in years)	
Atmospheric nutrient deposition	$Ndep_{NHX}$	Average atmospheric nitrogen deposition (NH_3 and NH_4)	Deposited nutrient feature
	$Ndep_{NOY}$	Average atmospheric nitrogen deposition (NO and NO_2)	
	Pdep	Average atmospheric phosphorus deposition	
Soil properties	BDRLOG	Probability of occurrence (0–100 %) of R horizon	Soil Property
	BDTICM	Absolute depth to bedrock (in cm)	
	BLDFIE	Bulk density (fine earth) in kg/m^3	
	CECSOL	Cation exchange capacity of soil in cmol/kg	
	CLYPPT	Clay content (0–2 mm) mass fraction in %	
	CRFVOL	Coarse fragments volumetric in %	
	ORCDRC	Soil organic carbon content (fine earth fraction) in g/kg	
	PHIHOX	Soil pH*10 in H_2O	
	PHIKCL	Soil pH*10 in KCl	
	SLTPPT	Silt content (2–50 mm) mass fraction in %	
	SNDPPT	Sand content (50–2000 mm) mass fraction in %	
	AWCh1	Derived available soil water capacity (volumetric fraction) with field capacity (FC) = pF 2.0	
	AWCh2	Derived available soil water capacity (volumetric fraction) with FC = pF 2.3	
	AWCh3	Derived available soil water capacity (volumetric fraction) with FC = pF 2.5	
	WWP	Derived available soil water capacity (volumetric fraction) until wilting point	
	AWCtS	Saturated water content (volumetric fraction) teta-S	
Elevation	elev	Elevation (in m)	Terrain feature

and ET_{sim} were validated according to the ten-fold cross-validation strategy. The following results and analyses were all based on the cross-validated parameters and GPP.

To assess the epistemic uncertainty of the neural network (defined as the unexplained variations arising from neural network structures and parameters), we applied drop-out training as a Bayesian approximation (Gal and Ghahramani, 2016). The drop-out procedure was repeated ten times, and the resulting standard deviations of parameter values were used to assess the uncertainty in predicted parameters, following the approach introduced by Hollinger and Richardson (Hollinger and Richardson, 2005). The standard deviation of Nash-Sutcliffe model efficiency (NSE, as Eq. (14) shows) was used to represent the uncertainty in simulated GPP accuracy for its sensitivity and suitability in evaluating ecological models (McCuen et al., 2006).

$$NSE = 1 - \frac{\sum_{t=1}^{N}(GPP_{obs}(t) - GPP_{sim}(t))^2}{\sum_{t=1}^{T}(GPP_{obs}(t) - \overline{GPP_{obs}})^2} \quad (14)$$

2.5. Interpretation approaches for spatial variability of water-related parameters

We used the following two approaches, partial dependence plot (PDP) and Shapley additive explanation (SHAP) dependence plot, to interpret the dependence of the three water-related parameters on SPIE inputs (introduced in Section 2.3) and identify the primary drivers of their spatial variability. For each parameter, we defined the categorical

variables with significant differences and the top ten most important non-categorical variables identified by both approaches as its primary drivers. Here we applied Kolmogorov-Smirnov test to assess the significance of the statistical differences between the PDP and SHAP values for specific types and other sites.

2.5.1. Partial dependence plot

The PDP is a typical approach to analyze the responses of predictions to the input features of interest for interpreting a machine learning model (Friedman, 2001). We computed the partial dependence of the predicted model parameters on each input feature based on Eq. (15).

$$D_i(x_i) = \frac{1}{N}\left(\sum_{j=1}^{N} f\left(x_i, x_{-i}^{j}\right)\right) \quad (15)$$

In the equation, $D_i(x_i)$ represents the partial dependence of feature i at point x_i. N is the number of observations (referring to the number of sites). $f(x_i, x_{-i}^{j})$ denotes the predicted parameter, which depends on the feature x_i and the remaining features x_{-i}. First, we generated a sequence of 100 evenly spaced numbers within the range of the input feature x_i across sites while keeping all the other features (x_{-i}) fixed. The combinations of these generated values and the fixed features were then input into the trained neural network to predict parameters of the test sites (datasets excluded from the training process). The mean of predicted parameters, based on x_i and x_{-i} (i.e., $f(x_i, x_{-i}^{j})$), across all sites was used to create the PDP. These plots display the response of the predicted parameter to changes in the relevant feature. For categorical features (PFT and climate types), partial dependence is calculated based on two values of the target feature (0 and 1) combined with the other features.

2.5.2. Shapley additive explanation dependence plot

The SHAP dependence plot is an alternative to PDP for the global interpretation of machine learning models. The SHAP value of each feature was calculated based on the deviation of the predicted model parameter at a certain input from the average prediction (Lundberg and Lee, 2017) (Eq. (16)).

$$\varphi_i(v_x) = \frac{1}{M}\sum_{S\subseteq\mathfrak{M}} \frac{v_x(S\cup i) - v_x(S)}{(M-1)!/|S|!(M-|S|-1)!} \quad (16)$$

The $\varphi_i(v_x)$ represents the SHAP value of feature i at point x_i with value function v (i.e., the predicted parameter value at x_i). M is the number of all features and $\mathfrak{M}$ is the set of all features. S refers to a subset of features or a feature in $\mathfrak{M}$. $|S|$ is the cardinality of the set S. The $v_x(S)$ is the parameter value relative to the features in set S for the point x. The SHAP value represents the individual contribution of a feature to the predicted parameter. We created the SHAP dependence plots based on the average SHAP value across sites to analyze the relationship between the predicted parameter and each feature.

2.6. Flowchart of this study

As the flowchart of this study presented (Fig. 2), we utilized the SPIE approach (marked by the dashed frame) to predict the global water-related parameters ($\boldsymbol{W_I}$, $\boldsymbol{k_W}$ and $\boldsymbol{\alpha_W}$) based on model inputs. And we analyzed the relationship between these parameters and the model inputs to explore how the water-related of GPP is influenced by the ecosystem properties.

3. Results

3.1. Performance of SPIE

The SPIE simulation of GPP showed robust spatiotemporal performance at different time scales (Fig. 3). NSE and determination coefficient (R^2) of the mean GPP per day-of-year, week and month across sites reached 0.78, 0.79 and 0.80, respectively (Fig. 3a-c). At the annual scale, the NSE and R^2 of GPP_{sim} were not as good as at shorter time scales, but the normalized root mean squared error was lower (NRMSE=0.31). At the site level, SPIE can perform well ($R^2 \geq 0.4$) at 86 %, 85 %, 83 % and 59 % of the sites at daily, weekly, monthly and annual scales, respectively, while cannot capture the GPP variations at the remaining sites ($R^2 < 0.4$). As another output of the LUE model, the simulation accuracy of ET was also good (NSE=0.55,0.57,0.59 and 0.35). Although the GPP_{sim} based on SPIE (in cross-validation) underperformed site calibrations (trained on all sites), SPIE overperformed using optimized parameters per PFT (OPT-PFT, shown in Figure S3). It demonstrates that the seasonal variability of GPP, as predicted using the SPIE-predicted parameters, is accurately captured with a low mean annual error.

At the site level, low uncertainties in GPP prediction accuracy and predicted parameters associated with neural network structures were

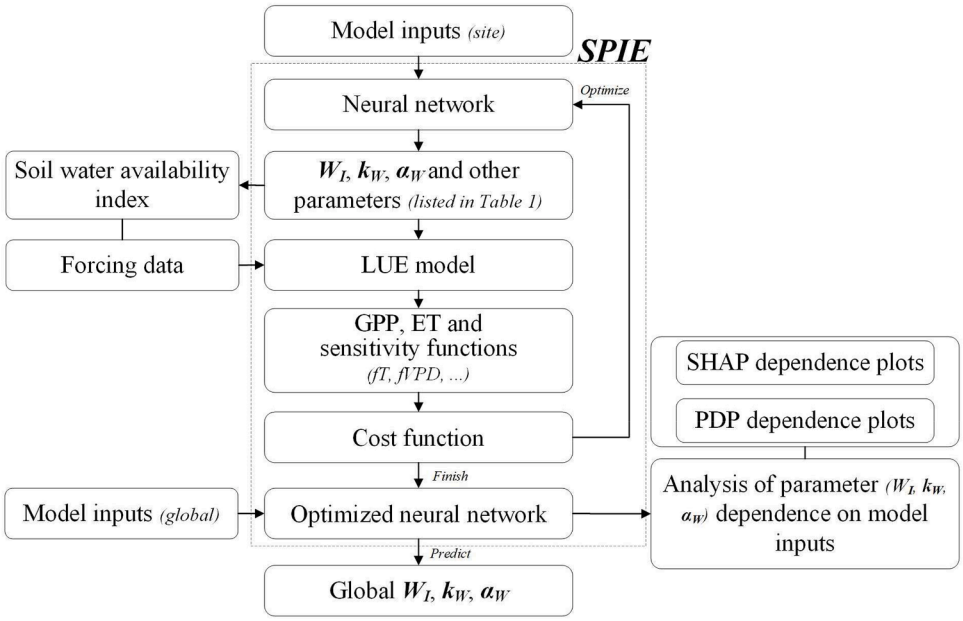

Fig. 2. Flowchart of this study.

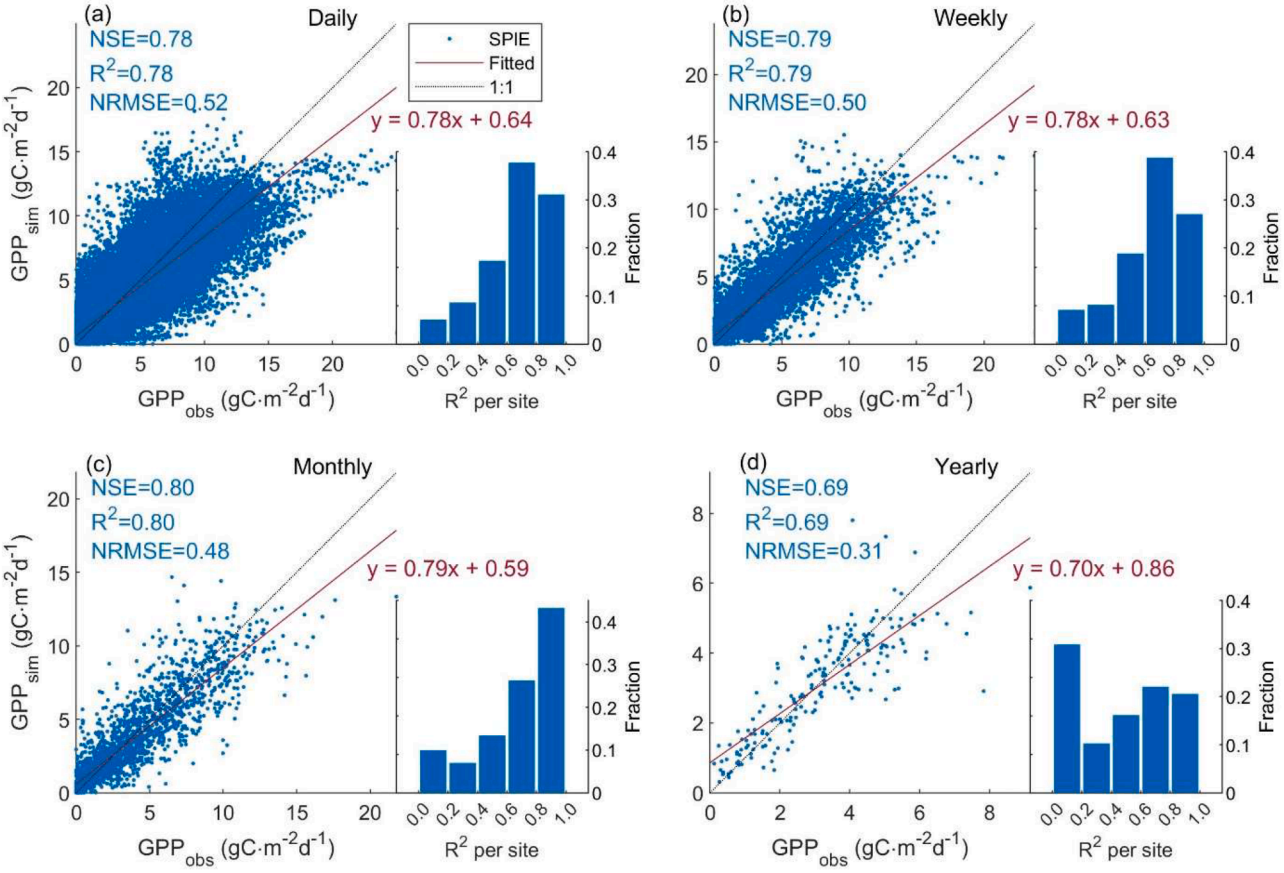

Fig. 3. Comparison between mean seasonal cycle GPP_{obs} and GPP_{sim} based on SPIE-predicted parameters (in blue). Each point in the scatter graphs represents the mean GPP per day, week, month or per year at a site, respectively. The Nash-Sutcliffe model efficiency (NSE), determination coefficient (R^2), and normalized root mean squared error (NRMSE) are displayed in the upper left corner of each plot. The histogram in the right part of each plot is the distribution of the site-level R^2.

observed (Fig. 4). The accuracy of mean seasonal cycle GPP showed robustness at daily, weekly and monthly scales with the uncertainty value of NSE below 0.01. However, at the annual scale, the uncertainty more than doubled. Among the three parameters, the uncertainty of predicted $\boldsymbol{k_W}$ was higher than the other two, though overall uncertainty remained low.

3.2. *Response of water-related parameters to model inputs*

The PDP and SHAP dependence plots (Figs. 5 and 6) showed that the spatial variability of $\boldsymbol{W_I}$ was influenced by vegetation characteristics, soil properties and climatic features. Although the PDP and SHAP have different units (see section S2), the curves exhibited similar trends in the relationship between $\boldsymbol{W_I}$ and input features (Figs. 5 and 6), especially after unit transformation (Figure S5 and Figure S6). MF, DBF and SAV had significantly lower $\boldsymbol{W_I}$ values than other types (first row of Figs. 5

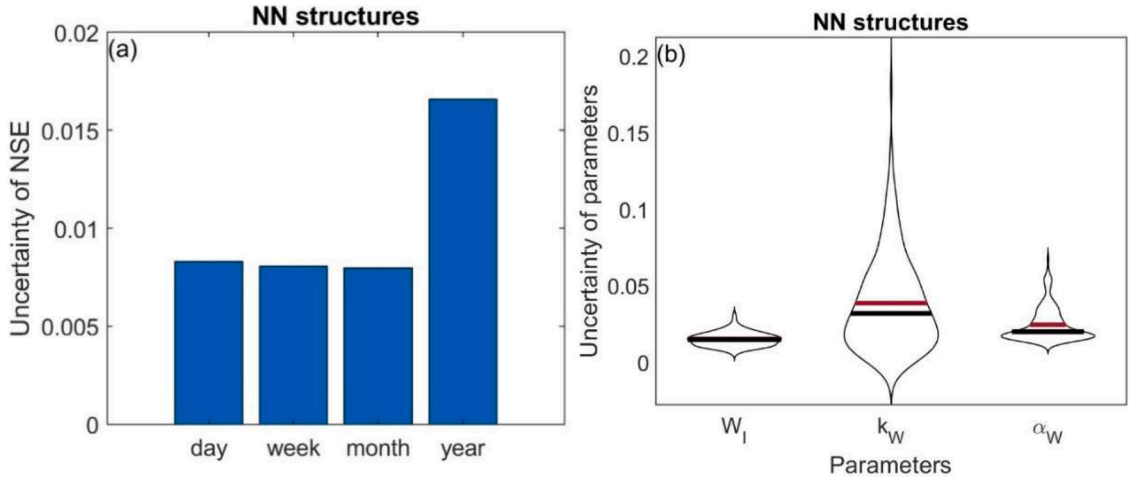

Fig. 4. Uncertainty in Nash-Sutcliffe efficiency (NSE) of mean seasonal cycle GPP (a) and predicted parameter scalars (b) related to neural network structures.

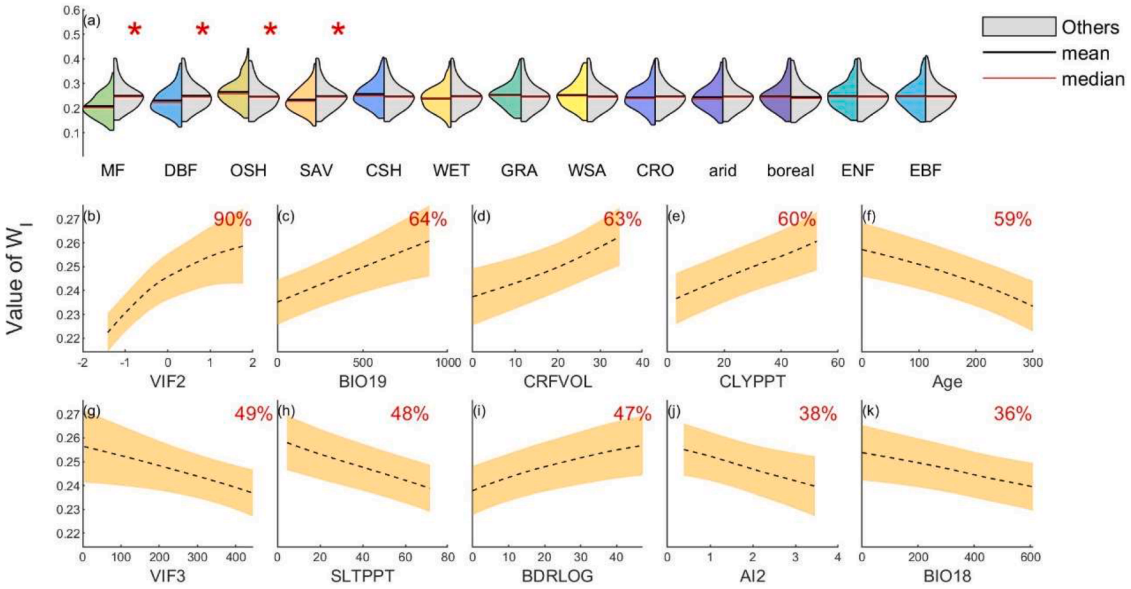

Fig. 5. Partial dependence plots showing the relationship between W_I and input features in cross-validation. Categorical features (the first row, (1)) are arranged based on differences in the mean predicted W_I, with red asterisks denoting statistically significant differences ($p < 0.05$) between predicted W_I of a specific category and those of other categories (grey-color). The width of the half violin plots represents the probability density function. Non-categorical features are sorted by their relative variance (displayed in the upper-right corner of (2)- (11)). The brown shading in (2)-(11) indicates the standard deviation of the predicted W_I across training groups.

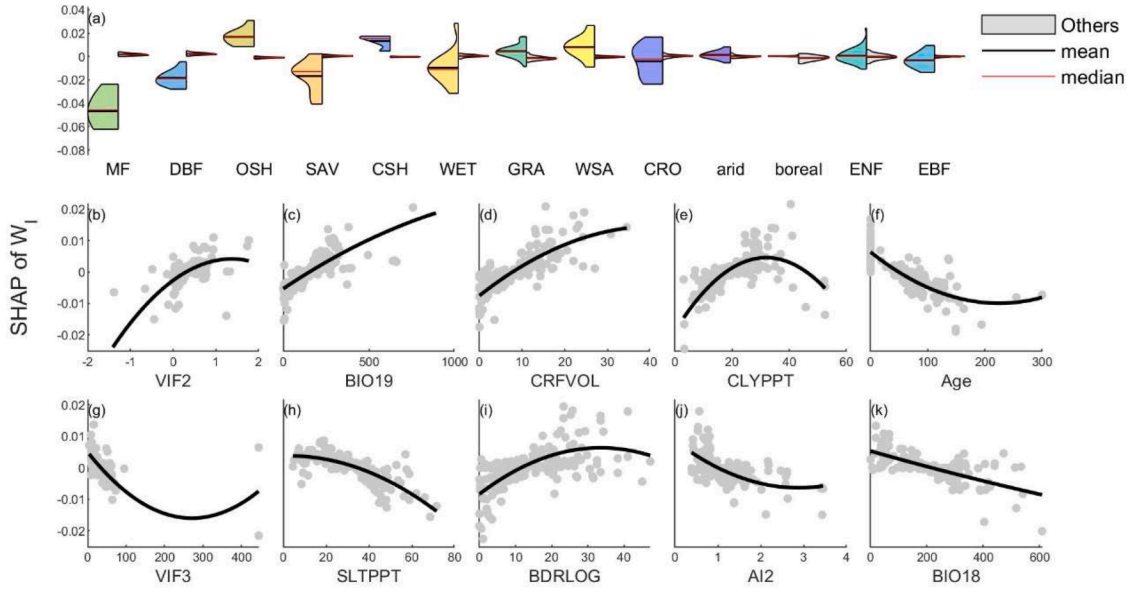

Fig. 6. Shapley additive explanation (SHAP) dependence plots showing the relationship between W_I and input features in cross-validation. All features are arranged in the same order as the partial dependence plots. The grey points in (2)-(11) represent the SHAP values for individual sites and the black lines are the fitted quadratic polynomial curve.

and 6), indicating generally lower vulnerability to W. In contrast, OSH had higher $\boldsymbol{W_I}$, suggesting a higher vulnerability. The differences among other PFTs were not statistically significant, while the mean and median $\boldsymbol{W_I}$ were higher for CSH, GRA and WSA sites. Within the set of continuous variables, the EVI variability (VIF2, representing vegetation seasonality) emerged as the most important variable to explain the variability of $\boldsymbol{W_I}$, followed by precipitation in the coldest quarter (BIO19), soil texture attributes (CRFVOL, CLYPPT and SLTPPT, representing coarse fragments, soil clay and silt contents, respectively) and forest age (Age). Furthermore, the maximum EVI in the warmest month (VIF3), R horizon probability, aridity seasonality (AI2) and the precipitation during the warmest season (BIO18) also contributed notably to the variability of $\boldsymbol{W_I}$. The contribution of other features to $\boldsymbol{W_I}$ was relatively small (shown in Figure S4). These response curves in both PDP and SHAP plots revealed that higher $\boldsymbol{W_I}$ was associated with strong vegetation seasonality, higher precipitation in the coldest season, increased coarse fragment, silt contents, younger forest age, lower EVI in the warmest season, reduced aridity seasonality, and lower precipitation in the warmest season. However, SHAP dependence plots showed nonlinear and nonmonotonic relationship between $\boldsymbol{W_I}$ and soil properties, especially clay content. The results demonstrated that the primary factors driving $\boldsymbol{W_I}$ include PFTs, vegetation seasonality, precipitation

features, soil properties, forest age, warm-season EVI and aridity seasonality.

The spatial variability of k_W was influenced by factors similar to those influencing W_I. Although PDP and SHAP plots displayed opposite trends due to their different units in Figs. 7 and 8, they represent the same response patterns (refer to the PDP plots converted to the same units as SHAP plots, Figure S8–9). Among PFTs, MF, DBF, and WET had statistically lower k_W, indicating a sharper response of GPP to W changes. Conversely, shrublands, including OSH and CSH, WSA, and GRA exhibited higher mean and median k_W, reflecting smoother responses to W. Similar to W_I, VIF2 had the largest contribution to the spatial variability of k_W. Other important factors for both parameters included forest age (Age), silt content (SLTPPT) and bulk density (BLDFIE). Moreover, several soil properties related to water holding capacity, soil quality and root penetration depth, such as absolute soil depth (BDTICM), wilting point (WWP), cation exchange capacity (CECSOL), soil pH (PHIHOX), and R horizon probability (BDRLOG), affected significantly the spatial variability of k_W. Other notable contributors included mean EVI of the warmest quarter (VIF7) and mean temperature of the driest quarter (BIO9). The response curves in both PDP and SHAP plots showed that k_W increased with vegetation seasonality, soil depth, EVI in the warmest season, soil pH, and temperature of the driest season, whereas decreased with forest age, wilting point, cation exchange capacity, and silt content. Notably, the relationship between k_W and the top ten non-categorical variables except CECSOL and BIO9 were all non-linear. These findings highlighted that k_W is dominated by PFTs, vegetation seasonality, forest age, soil properties and dry-season temperature.

The parameter representing the lagged response of GPP to W change, α_W, had different influential factors from W_I and k_W in PDP and SHAP plots (Figs. 9 and 10). OSH had statistically higher α_W values compared to other types, suggesting a more pronounced lag effect. While the differences were not statistically significant, arid vegetation had higher mean and median α_W values than vegetation in other climate types. In contrast, α_W for DBF was significantly lower, indicating that DBF tends to respond promptly to W changes. Other PFTs showed no statistically significant differences. Three vegetation features, EVI variability (VIF2), mean EVI of the wettest quarter (VIF5) and the maximum EVI of the warmest month (VIF3) emerged as the most important contributors. Thermal features, including the maximum temperature of the warmest month (BIO5), mean temperature of the warmest quarter (BIO10), mean diurnal and annual temperature range (BIO2 and BIO7), also played significant roles. Additionally, soil properties such as soil depth (BDTICM and BDRLOG) and cation exchange capacity (CECSOL), indicative of root penetration depth and soil fertility, further influenced α_W. The response curves in both PDP and SHAP plots showed that vegetation with strong seasonality, growing in higher temperatures in the warmest month and season, larger diurnal and annual temperature ranges and deeper soil depth tend to have a lagged response to W changes. Notably, relationships between α_W and bioclimatic variables were almost linear whereas the relationships with soil properties and vegetation features were primarily nonlinear. Overall, α_W was primarily driven by vegetation characteristics, followed by thermal features and soil properties related to root depth and nutrient availability.

To summarize, these water-related parameters, W_I, k_W and α_W, varied across PFTs and strongly responded to vegetation seasonality, including EVI variability and seasonal EVI values, and soil properties. DBF and MF exhibited lower values for all three parameters, while OSH showed the opposite trend. W_I and k_W were both influenced by forest age and soil texture whereas k_W and α_W were jointly affected by thermal characteristics.

3.3. Distribution maps of water-related parameters

The global distribution maps of W_I and k_W revealed similar patterns, while α_W displayed a distinct pattern (Fig. 11a and b). W_I and k_W were higher in subtropical and temperate dry regions (red and yellow color in Fig. 11a and green color in Fig. 11b) and lower in the wet tropics and high northern latitudes (blue color in Fig. 11a and yellow and purple color in Fig. 11b). These two parameters were strongly correlated ($r >$ 0.7) in 55 % of the areas. α_W was close to one (median=0.83) in most areas of arid climates (as classified by Koeppen-Geiger climate classification, Fig. 11c), where W_I and k_W were higher as well. However, α_W was significantly lower in the remaining arid regions and approached zero in other climate types. The global distribution of k_W showed a slightly different density distribution pattern from the local scale (refers to the histogram in Fig. 11b), whereas W_I and α_W exhibited similar distribution patterns. Overall, these results indicated that most vegetation in arid climates tends to exhibit delayed and slower responses to variations in W and faces a higher risk of water limitations.

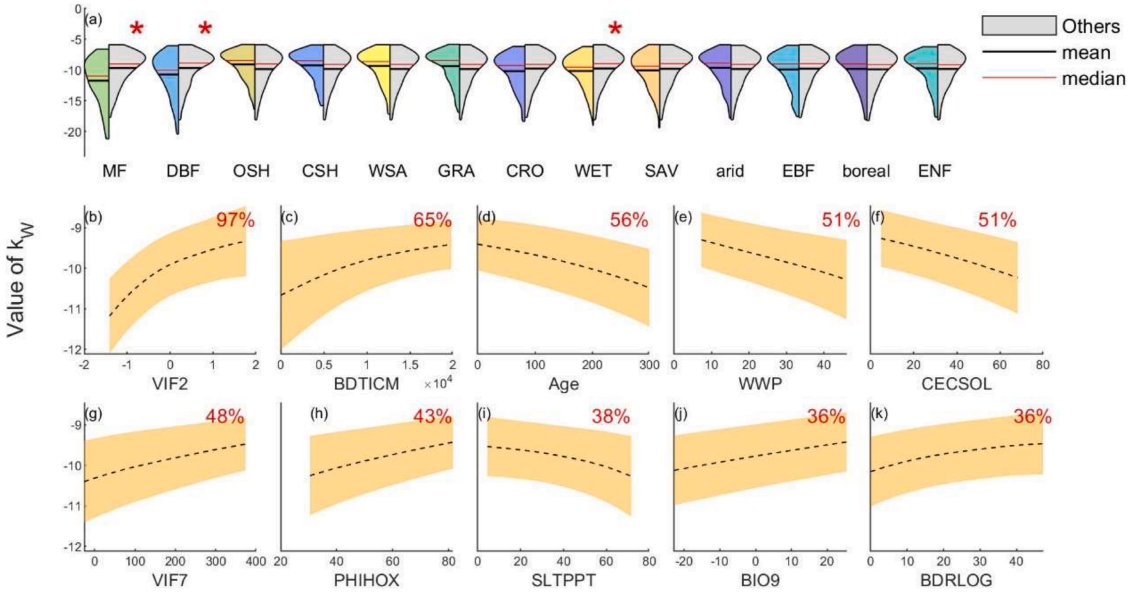

Fig. 7. Partial dependence plots showing relationships between k_W and input features.

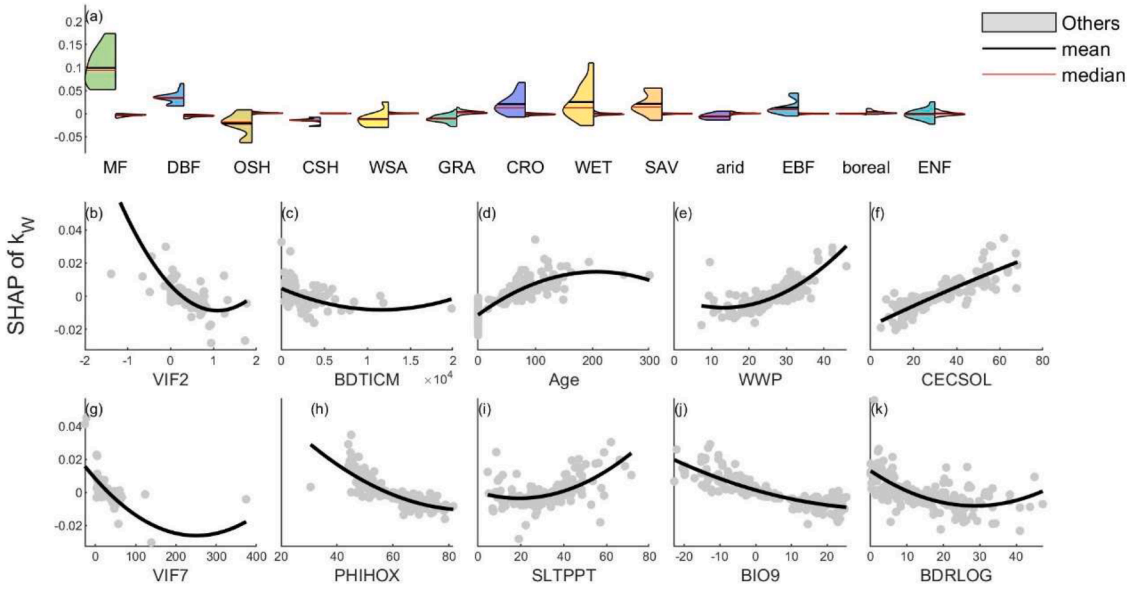

Fig. 8. Shapley additive explanation (SHAP) dependence plots showing relationships between k_W and input features.

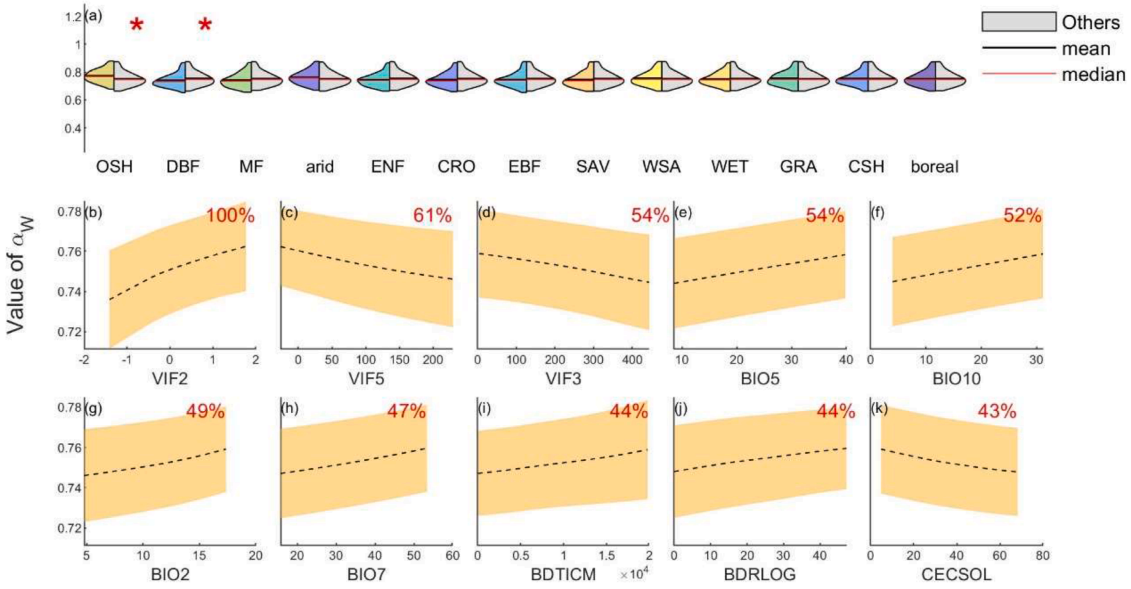

Fig. 9. Partial dependence plots of α_W to input features.

4. Discussion

The results support the hypothesis that the sensitivity of GPP to W, as represented by the three parameters, is influenced by vegetation types as well as edaphoclimatic properties translated into bioclimatic and soil properties. However, climate aridity and most of the soil properties were not drivers for every parameter. Given that the LUE model effectively captured the spatiotemporal variability of GPP across eddy covariance sites, the physical meaning of these parameters and the real-world datasets used to predict the parameters, our findings offer insights into the different factors influencing functional controls of GPP sensitivity to W. In this section, we discuss the main factors behind parameter variability and how the spatial patterns of these three parameters relate to findings from other studies. Furthermore, we highlight the challenges when interpreting model parameters with machine learning approaches.

4.1. Drivers for GPP sensitivity to soil water availability

The PDP and SHAP results suggest that the response of GPP to W variations, represented by the three water-related parameters, varies across PFTs, aligning with Horn and Schulz (2011). Forests, particularly MF and DBF, exhibited lower values for the inflection point, response slope and lag coefficient. This indicates that forests are less susceptible to water limitations, and respond promptly with a steep slope to W changes. In contrast, open shrubs are more vulnerable to low W, and exhibit a smoother and lagged response to W changes. This finding aligns with the observations of Whitecross et al. (2017), highlighting a slower response to W changes can lead to more moderate water usage, potentially regulated by stomatal control. During dry periods, this regulation helps conserve water, extending the period of water availability. Conversely, during periods of increased water supply, such a strategy also regulates water consumption and growth, which can be advantageous in regions with irregular precipitation patterns. However,

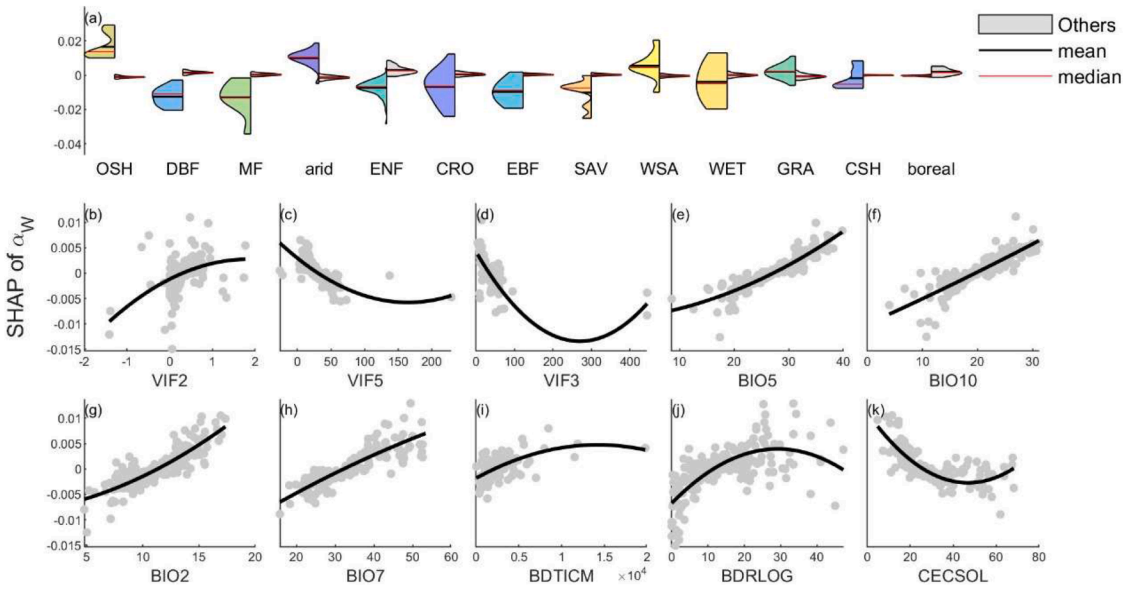

Fig. 10. Shapley additive explanation (SHAP) dependence plots of α_W to input features.

the association of the plant type with water limitation does not imply a unidirectional causation. While PFTs differ in their physiological responses to water stress, their spatial distributions are also shaped by long-term environmental selection. Forests typically emerge in regions with less frequent or less intense water stress, whereas grasses and shrubs are often better adapted to arid climates. This highlights the importance of considering bi-directional interactions between vegetation and environmental constraints.

The seasonality of vegetation emerges as the most important non-categorical variable to characterize the GPP sensitivity to W. This includes the EVI variability (VIF2) and the seasonal EVI values such as VIF3, 5 and 7. The importance of vegetation seasonality can, in part, be attributed to the strong correlation between the dynamics of vegetation indices and carbon assimilation rates (Wang et al., 2010; Horn and Schulz, 2011). However, vegetation seasonality is both a cause and a consequence of water availability. On one hand, plants adapt their growth cycles and phenological strategies in response to seasonal water availability (Yu et al., 2023). On the other hand, these seasonal traits, such as prolonged growing seasons or sharp peaks in EVI or GPP, can influence how efficiently vegetation utilizes available water. Our findings suggest that regions with lower seasonal variability and higher seasonal EVI are generally associated with fewer water constraints and sharper responses of GPP to changes in W. Conversely, ecosystems with higher seasonal EVI variability and lower seasonal EVI tend to show more constrained and delayed responses. Thus, vegetation seasonality reflects a complex interplay between environmental effects and adaptive ecological traits, reinforcing the need to interpret its role within a reciprocal, not merely causal, framework.

Forest age is another important feature influencing both the inflection point and the slope of soil water response function. The results suggest that older trees tend to be less vulnerable under the same dry conditions and exhibit a sharper response to W changes compared to younger trees. Our findings agree with the on-site study of trees older than 25 years by Jia et al. (2017), showing higher average soil moisture content for older trees. Additionally, the wet and warm environments commonly associated with older trees (e.g., >50 years), such as rainforests, along with the potentially deeper root systems, may further enhance their sensitivity and responsiveness to water availability.

Soil properties have been demonstrated to play a pivotal role in modulating vegetation responses to water stress (Fernandez-Illescas et al., 2001), though their influence is intricate, as revealed by our results. The relationships between soil properties and SHAP values of the parameters are nonlinear (Figs. 6, 8 and 10), with the exception of the relationship between response slope and cation exchange capacity. We recognize the R horizon probability as a crucial soil property for all three parameters. An intermediate probability of reaching R horizon reduces the inflection point and lag effect, while improving the response slope of GPP to W. Similarly, a moderate cation exchange capacity diminishes the lag effect. However, medium soil clay content increases the inflection point and hence the risk of soil water stress. The non-linear relationship between these parameters with soil properties might be the result of combined and interactive effects of various soil properties across different soil layers (Manrique et al., 1991).

The relationship between bioclimatic variables and these parameters were all close to linear. The results highlighted a small role of aridity for all the three parameters, while the aridity seasonality is one of the primary drivers for the inflection point of GPP response to W. Among the climate features, seasonal rainfall is another important driver of the inflection point, indicating that seasonal precipitation influences the vulnerability to W (Hajek and Knapp, 2022). The thermal factors primarily affected the degree of lag effect, including the temperature during the warmest season, as well as the diurnal and annual temperature ranges. This finding contrasts with the results of Horn and Schulz (2011), which suggest that the lag effect can be partially explained by the correlation between temperature and evaporative fraction. This may be caused by differences in ecosystem features used to feed machine learning models between these two studies. Our results reveal that high temperatures and temperature ranges lead to a more pronounced lag effect.

Overall, alongside plant functional types, vegetation, soil and climatic factors, such as vegetation seasonality, forest age, soil properties, aridity seasonality, seasonal rainfall, seasonal temperature mean and ranges, emerge as the primary factors associated to the spatial variability in GPP responses to W. While these findings are based on statistical associations, these results suggest that the shape and intensity of GPP responses to W arise from complex plant-environmental interactions, reflecting long-term ecological adaptation and feedbacks. Under extreme climate events, ecosystems with sharper and earlier GPP declines in responses to W deficits may be more vulnerable to rapid-onset droughts, while ecosystems characterized by lagged or buffered responses may exhibit lower vulnerability and slower recovery. Identifying these traits and environmental contexts related to these response patterns can inform predictions of ecosystem carbon dynamics under intensifying climate extremes and guide adaptive management

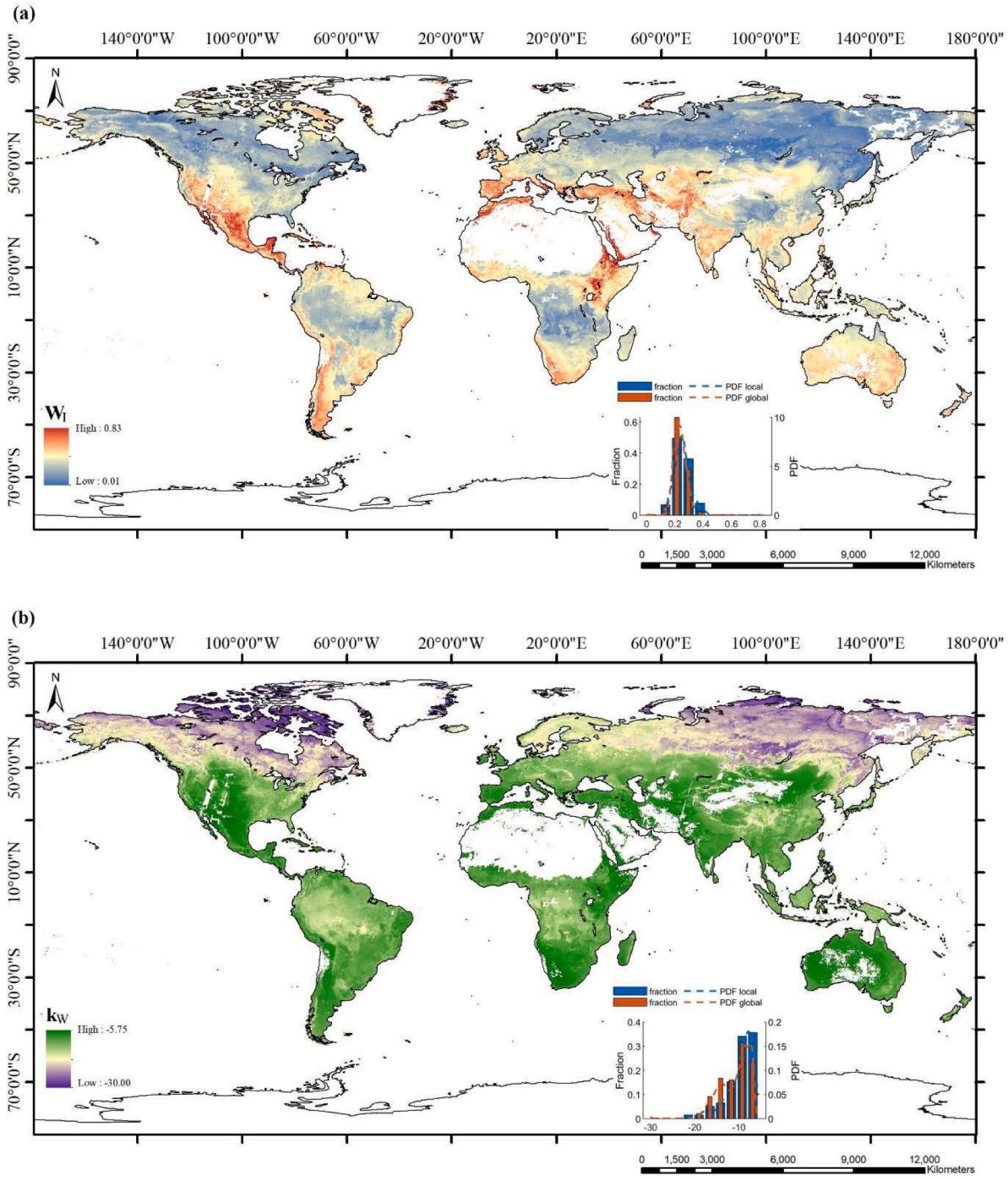

Fig. 11. Global distribution maps of (a)W_I, (b)k_W and (c)α_W in 0.0083° (smoothed using a 25×25 moving median window). The histogram shows the possibility distribution function (PDF, dotted lines) and fraction (bars) of the predicted parameters at sites (in blue) and at the global scale (in red).

strategies.

4.2. Global patterns of water-related parameters

The global distribution of $\boldsymbol{W_I}$ and $\boldsymbol{k_W}$ generally have similar patterns with the climate aridity (Ullah et al., 2022; Asadi Zarch et al., 2015). It is also similar to the pattern of precipitation sensitivity to leaf area index changes [Fig. 2a in (Cui et al., 2022)], suggesting the spatial variability of vegetation sensitivities to water supply. However, the pattern is different from the root-zone water storage estimated by the maximum cumulative water deficit during dry seasons (Stocker et al., 2023), particularly in the subtropical dry regions (e.g., southern Spain, India and South Asia). In most of the arid areas, the response of GPP to water availability changes has a temporal lag (brown area in Fig. 11c). This lag effect might be associated with the vegetation's strategy to mitigate water stress resulting from insufficient rainfall or alterations in its seasonality (Souza et al., 2016). The strategy is a result of a plant's system to absorb and recover from disturbance or stress (Kinzig et al., 2006; Siqueira et al., 2008), which has been reported at local scales (Horn and Schulz, 2011; Siqueira et al., 2008; Zhang et al., 2011). Our results reveal the spatial relationship between the response of GPP to W and water supply, which can support conservation efforts in regions where

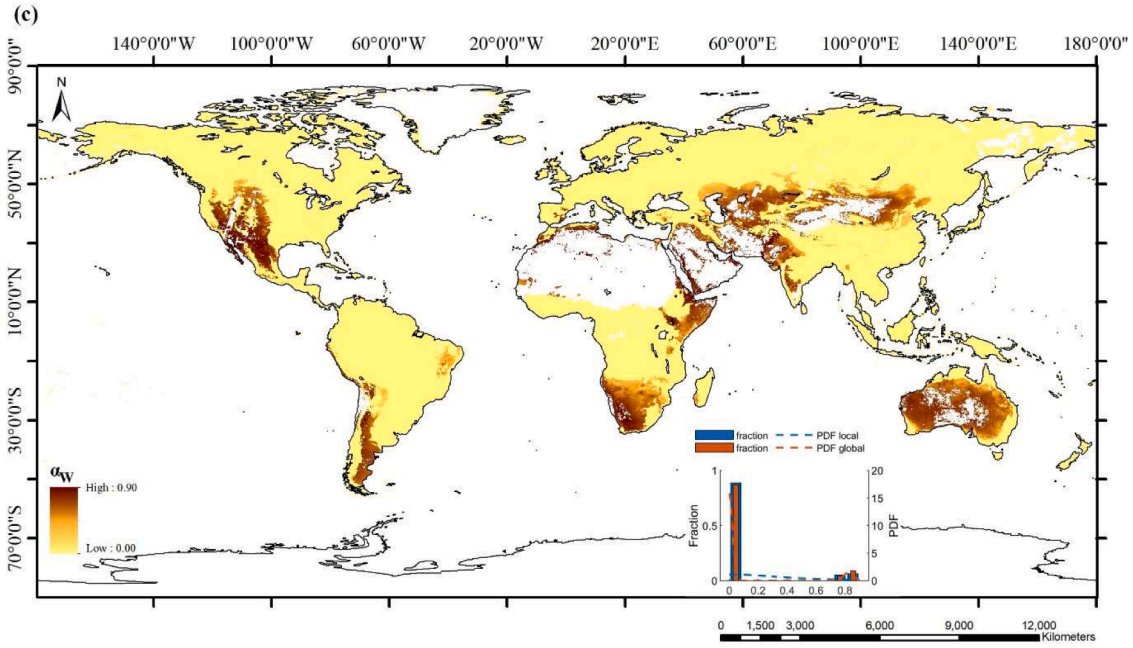

Fig. 11. (*continued*).

soil moisture plays a critical role in sustaining vegetation productivity.

All the parameter maps resulting from this study, which can be directly used to parameterize the LUE model by multiplying with a scalar, are made available to the community. These could support further modeling and data analysis exercises, including exploring local to global changes in GPP given varying phenology and climate scenarios.

4.3. Challenges in parameter interpretation

Although the relative importance of the input features (i.e., the magnitude of the partial dependence and SHAP dependence) differ across approaches and datasets, our results show consistent trends in dependence plots between PDP and SHAP. Nevertheless, several challenges require further investigation in future studies.

The performance of simulated GPP by SPIE is good across sites (NSE=0.78, R^2=0.78 and NRMSE=0.52), while it is low ($R^2 \leq 0.4$) at 16 % of the sites, particularly for evergreen broadleaf forest, tropical and polar sites where the uncertainty is relatively higher. This lower performance may hinder the analysis of parameter variability at these sites, contributing to increased error and uncertainty in the wet tropics and Arctic region, as indicated by the large parametric uncertainties in tropics and humid areas (Figure S13a-c). A probable contributing factor is the low data availability in these regions (Figure S2), which can impede regional interpretation of the links between local ecosystem features and parameter variability, while also inflating uncertainties in the posterior estimation of GPP fields. Consequently, further inclusion of tropical datasets is crucial for improving GPP modelling accuracy and gaining insights into spatial changes in GPP climate sensitivity.

The high GPP values at the daily scale were systematically underestimated at several sites, a common issue observed in many GPP models (Bao et al., 2022; De et al., 2024). Although the underlying causes remain unclear, (De et al., 2024) demonstrated that addressing this bias could substantially improve the modeled interannual variability of GPP. Thus, future studies should aim to uncover the mechanisms behind this underestimation and develop methods to enhance peak GPP modeling at daily and hourly scales.

The interactions among these input features for the neural network might affect the model interpretation. Understanding how these features influence each other is crucial for improving the model accuracy and interpretability.

Beyond the spatial controls in parameter variability there is relevance in understanding and quantifying sources of temporal variability in GPP climate sensitivities. In addition to instantaneous responses to environmental changes, plants undergo physiological and structural adaptions over weekly to monthly time scales to optimize their growth potential during their lifetime, i.e., acclimation (Stillman, 2003; Walters, 2005). Mengoli et al. (2022) demonstrated the importance of incorporating temperature acclimation by dynamically adjusting parameters representing the maximum carboxylation and electron transport rates, leading to improvement in model performance. Despite existing ambiguities in the formulation and effectiveness of acclimation principles (De et al., 2024), such results highlight the necessity for challenging temporal invariance in model parameters in future studies to tackle, e.g., interannual variability in ecosystem fluxes.

5. Conclusion

This study identifies the primary factors explaining the spatial variability of the three parameters controlling the GPP response to soil water availability in a LUE model. These three parameters represent the vulnerability to relative soil moisture, response steepness and the magnitude of lag effect. Our results demonstrate a robust representation of GPP observations, and highlight that the shape of the GPP sensitivity to soil water availability varies between plant functional types and is largely affected by vegetation seasonality, including EVI variability and seasonal EVI values. The impact of soil properties is found to be nonlinear and intrinsic. Notably, the probability of R horizon occurrence has a significant impact on all three parameters. Furthermore, forest age and soil texture influence both the inflection point and slope of the response curve, while seasonal temperatures and temperature ranges play a crucial role in determining the lag effect. The spatial distribution of the inflection point and response slope mirrors climate aridity, though not all vegetation in arid climates exhibits a lagged response to soil water availability changes.

These results extend beyond the hypothesis that GPP sensitivity to soil water availability varies across plant functional types and argue for accommodating variability within PFT by quantifying the role of local

edaphoclimatic conditions and phenological responses in driving parametric variability. Our findings further indicate that climate aridity alone, as well as most of the soil properties, are not primary drivers of this sensitivity. This research provides insights into the intrinsic relationship between ecosystem features and photosynthesis sensitivity, supporting the idea that the vegetation's functional response to climate emerges from long-term plant-environment interactions.

We acknowledge the remaining outstanding challenges related to the underrepresentation of tropical and polar ecosystems in the observational network, parameter uncertainty, model representativeness, and the complexity of interactions between ecosystem features. These suggest that future studies, in addition to improving LUE modeling performance, should focus on dimensionality reduction to enhance generalization capabilities. Since various aspects of GPP response to soil water availability vary with vegetation, climate and soil properties, our study emphasizes the importance of accounting for both spatial and, eventually, temporal variability in vegetation photosynthesis sensitivity when analyzing and forecasting global carbon assimilation.

CRediT authorship contribution statement

Shanning Bao: Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Nuno Carvalhais:** Writing – review & editing, Supervision, Methodology. **Jian Xu:** Writing – review & editing. **Jing M. Chen:** Writing – review & editing. **Yang Lei:** Writing – review & editing, Funding acquisition. **Gegen Tana:** Writing – review & editing. **Changgui Lin:** Writing – review & editing. **Jiancheng Shi:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

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

Acknowledgement

The study is supported by the National Key Research and Development Program of China (No. 2022YFB3903300), the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) "Dynamic monitoring and simulation of water cycle in Asian water tower area" (No. 2019QZKK0206), and the Youth Innovation Funding of the National Space Science Center, Chinese Academy of Sciences. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The FLUXNET eddy covariance data processing and harmonization was carried out by the ICOS Ecosystem Thematic Center, AmeriFlux Management Project and Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux and AsiaFlux offices.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2025.110701.

Data availability

Data will be made available on request.

References

Asadi Zarch, M.A., Sivakumar, B., Sharma, A., 2015. Assessment of global aridity change. J Hydrol 520, 300–313.
Bao, S., et al., 2022. Environment-sensitivity functions for gross primary productivity in light use efficiency models. Agric For Meteorol 312, 108708.
Bao, S., et al., 2023a. Toward robust parameterizations in ecosystem-level photosynthesis models. Journal of Advances in Modeling Earth Systems 15 (8), e2022MS003464.
Bao, S., et al., 2023b. Towards robust parameterizations in ecosystem-level photosynthesis models. Journal of Advances in Modeling Earth Systems 15 (8), e2022MS003464.
Besnard, S., et al., 2021. Mapping global forest age from forest inventories, biomass and climate data. Earth System Science Data 13 (10), 4881–4896.
Brownlee, J., 2020. Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python. Machine Learning Mastery.
Carvalhais, N., et al., 2010. Deciphering the components of regional net ecosystem fluxes following a bottom-up approach for the Iberian Peninsula. Biogeosciences 7 (11), 3707–3729.
Croft, H., et al., 2017. Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. Glob Chang Biol 23 (9), 3513–3524.
Cui, J., et al., 2022. Global water availability boosted by vegetation-driven changes in atmospheric moisture transport. Nat Geosci 15 (12), 982–988.
De, R., et al., Addressing challenges in simulating inter-annual variability of gross primary production. 2024.
Eyring, V., et al., 2016. Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development 9 (5), 1937–1958.
Farquhar, G.D., von Caemmerer, S.v., Berry, J.A., 1980. A biochemical model of photosynthetic CO 2 assimilation in leaves of C 3 species. planta 149, 78–90.
Fernandez-Illescas, C.P., et al., 2001. The ecohydrological role of soil texture in a water-limited ecosystem. Water Resour Res 37 (12), 2863–2872.
Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International journal of climatology 37 (12), 4302–4315.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann Stat 1189–1232.
Gal, Y., Ghahramani, Z., 2016. Dropout as a bayesian approximation: representing model uncertainty in deep learning. In: international conference on machine learning. PMLR.
Gampe, D., et al., 2021. Increasing impact of warm droughts on northern ecosystem productivity over recent decades. Nat Clim Chang 11 (9), 772–779.
Greve, P., et al., 2014. Global assessment of trends in wetting and drying over land. Nat Geosci 7 (10), 716–721.
Groenendijk, M., et al., 2011. Assessing parameter variability in a photosynthesis model within and between plant functional types using global Fluxnet eddy covariance data. Agric For Meteorol 151 (1), 22–38.
Hajek, O.L., Knapp, A.K., 2022. Shifting seasonal patterns of water availability: ecosystem responses to an unappreciated dimension of climate change. New Phytologist 233 (1), 119–125.
Hollinger, D.Y., Richardson, A.D., 2005. Uncertainty in eddy covariance measurements and its application to physiological models. Tree Physiol. 25 (7), 873–885.
Horn, J., Schulz, K., 2011. Spatial extrapolation of light use efficiency model parameters to predict gross primary production. Journal of Advances in Modeling Earth Systems 3 (4).
Humphrey, V., et al., 2021. Soil moisture–atmosphere feedback dominates land carbon uptake variability. Nature 592 (7852), 65–69.
Jia, X., et al., 2017. Soil moisture decline due to afforestation across the Loess Plateau, China. J Hydrol 546, 113–122.
Jung, M., et al., 2017. Compensatory water effects link yearly global land CO2 sink changes to temperature. Nature 541 (7638), 516–520.
Kinzig, A.P., et al., 2006. Resilience and regime shifts: assessing cascading effects. Ecology and society 11 (1).
Lawrence, D.M., et al., 2019. The community land model version 5: description of new features, benchmarking, and impact of forcing uncertainty. Journal of Advances in Modeling Earth Systems, 11 (12), 4245–4287.
Li, W., et al., 2023. Widespread and complex drought effects on vegetation physiology inferred from space. Nat Commun 14 (1), 4640.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc., Long Beach, California, USA, pp. 4768–4777.
Madani, N., et al., 2014. Improving ecosystem productivity modeling through spatially explicit estimation of optimal light use efficiency. Journal of Geophysical Research: Biogeosciences 119 (9), 1755–1769.
Manrique, L.A., Jones, C.A., Dyke, P.T., 1991. Predicting soil water retention characteristics from soil physical and chemical properties. Commun Soil Sci Plant Anal 22 (17–18), 1847–1860.
McCuen, R.H., Z. Knight, and A.G.J.J.o.h.e. Cutter, Evaluation of the Nash–Sutcliffe efficiency index. 2006. 11(6): p. 597–602.
Mengoli, G., et al., 2022. Ecosystem photosynthesis in land-surface models: a first-principles approach incorporating acclimation. Journal of Advances in Modeling Earth Systems 14 (1), e2021MS002767.
Monteith, J.L., 1972. Solar radiation and productivity in tropical ecosystems. Journal of applied ecology 9 (3), 747–766.

Peaucelle, M., et al., 2019. Covariations between plant functional traits emerge from constraining parameterization of a terrestrial biosphere model. Global Ecology and Biogeography 28 (9), 1351–1365.
Poggio, L., et al., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7 (1), 217–240.
Reichstein, M., et al., On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. 2005. 11(9): p. 1424–1439.
Rodhe, H., Dentener, F., Schulz, M., 2002. The global distribution of acidifying wet deposition. Environ. Sci. Technol. 36 (20), 4382–4388.
Running, S.W., et al., 2004. A continuous satellite-derived measure of global terrestrial primary production. Bioscience 54 (6), 547–560.
Siqueira, M., Katul, G., Porporato, A., 2008. Onset of water stress, hysteresis in plant conductance, and hydraulic lift: scaling soil water dynamics from millimeters to meters. Water Resour Res 44 (1).
Sitch, S., et al., 2008. Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five dynamic Global Vegetation models (DGVMs). Glob Chang Biol 14 (9), 2015–2039.
Souza, R., et al., 2016. Vegetation response to rainfall seasonality and interannual variability in tropical dry forests. Hydrol Process 30 (20), 3583–3595.
Stillman, J.H., 2003. Acclimation capacity underlies susceptibility to climate change. Science 301 (5629), 65. -65.
Stocker, B.D., et al., 2018. Quantifying soil moisture impacts on light use efficiency across biomes. New Phytologist 218 (4), 1430–1449.
Stocker, B.D., et al., 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.
Stocker, B.D., et al., 2023. Global patterns of water storage in the rooting zones of vegetation. Nat Geosci 16 (3), 250–256.
Tramontana, G., et al., 2016. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. Biogeosciences 13 (14), 4291–4313.
Trautmann, T., et al., 2022. The importance of vegetation in understanding terrestrial water storage variations. Hydrol. Earth Syst. Sci. 26 (4), 1089–1109.
Ullah, S., et al., 2022. Spatiotemporal changes in global aridity in terms of multiple aridity indices: an assessment based on the CRU data. Atmospheric Research, 268, 105998.
Walters, R.G., 2005. Towards an understanding of photosynthetic acclimation. J. Exp. Bot. 56 (411), 435–447.
Walther, S., et al., 2022. A view from space on global flux towers by MODIS and Landsat: the FluxnetEO dataset (Landsat). Biogeosciences 2805–2840.
Wang, H., et al., 2010. Deriving maximal light use efficiency from coordinated flux measurements and satellite data for regional gross primary production modeling. Remote Sens Environ 114 (10), 2248–2258.
Wang, H., et al., 2017. Towards a universal model for carbon dioxide uptake by plants. Nat Plants 3 (9), 734–741.
Whitecross, M., Witkowski, E., Archibald, S., 2017. Savanna tree-grass interactions: a phenological investigation of green-up in relation to water availability over three seasons. South African Journal of Botany 108, 29–40.
Yu, T., et al., 2023. Interannual and seasonal relationships between photosynthesis and summer soil moisture in the Ili River basin, Xinjiang, 2000–2018. Science of The Total Environment 856, 159191.
Yuan, W., et al., 2014. Global comparison of light use efficiency models for simulating terrestrial vegetation gross primary production based on the LaThuile database. Agric For Meteorol 192, 108–120.
Yuan, W., et al., 2019. Increased atmospheric vapor pressure deficit reduces global vegetation growth. Sci Adv 5 (8), eaax1396.
Zhang, G., et al., 2011. Responses of grassland vegetation to climatic variations on different temporal scales in Hulun Buir Grassland in the past 30 years. Journal of Geographical Sciences 21, 634–650.
Zomer, R.J., Xu, J., Trabucco, A., 2022. Version 3 of the global aridity index and potential evapotranspiration database. Sci Data 9 (1), 409.