FLAML version 2.3.3 model-based assessment of gross primary productivity at forest, grassland, and cropland ecosystem sites

Jie Lai [a, b], Yuan Zhang [a], Anzhi Wang [a], Wenli Fei [a], Yiwei Diao [c], Rongping Li [d], Jiabin Wu [a, *]

[a] CAS Key Laboratory of Forest Ecology and Silviculture, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China

[b] University of Chinese Academy of Sciences, Beijing 101408, China

[c] Key Laboratory of Ecosystem Carbon Source and Sink, China Meteorological Administration (ECSS-CMA), Wuxi University, Wuxi, 214105, China

[d] Institute of Atmospheric Environment, China Meteorological Administration, Shenyang 110016, China

*** Correspondence: wujb@iae.ac.cn**

Abstract

Accurately estimating Gross Primary Productivity (GPP) in terrestrial ecosystems is essential for understanding the global carbon cycle. Satellite-based Light Use Efficiency (LUE) models are commonly employed for simulating GPP. However, the variables and algorithms related to environmental limiting factors differ significantly across various LUE models~~.~~, leading to high uncertainty in GPP estimation. In this work, we developed a series of FLAML-LUE models ~~tailored for~~with different ~~ecosystems~~variable combinations. These models utilize the Fast Lightweight Automated Machine Learning (FLAML) framework, using variables of LUE models, to investigate the potential of estimating site-scale GPP. Incorporating meteorological data, eddy covariance measurements, and remote sensing indices, we employed FLAML-LUE models to assess the impact of various variable combinations on GPP across different temporal scales, including daily, 8-day, 16-day, and monthly intervals. Cross-validation analyses indicated that the ~~effectiveness of FLAML-LUE models for~~

~~forest ecosystems varied significantly across different sites, with R^2 values ranging from 0.56 to~~ FLAML-LUE model performs excellently in GPP prediction, accurately simulating both its temporal variations and magnitude, particularly in mixed forests and coniferous forests, with average R^2 values for daily-scale simulations reaching 0.92 and 0.91, respectively. However, the model performed less effectively in alpine shrubland and typical grassland ecosystems, though it still outperformed both MODIS GPP and PML GPP in terms of performance. Furthermore, the model's adaptability under extreme climate conditions was evaluated, and the results showed that high temperatures and high VPD lead to a slight decrease in model accuracy, though R^2 remains around 0.8. Under drought conditions, the model's performance improved slightly in croplands and evergreen broadleaf forests, although it declined at some sites.~~0.94. For grassland ecosystems, R^2 values ranged from 0.62 to 0.87, and for cropland ecosystems, R^2 values ranged from 0.78 to 0.88. Extending the time scale of input data could significantly enhance the accuracy of model simulations. Specifically, the average R^2 increased from 0.82 to 0.92 for forest ecosystems, 0.79 to 0.83 for grassland ecosystems, and 0.84 to 0.87 for farmland ecosystems. Additionally, the importance ranking method indicated that vegetation index and temperature were the most important variables for GPP estimation in forest, grassland, and farmland ecosystems, while the importance of the moisture index was relatively low.~~ This study offers an approach to estimate GPP fluxes and evaluate the impact of variables on GPP estimation. It has the potential to be applied in predicting GPP for different vegetation types at a regional scale.

Keywords: Light Use Efficiency; Gross Primary Productivity; Automated Machine Learning; Fast Lightweight Automated Machine Learning

1. Introduction

The global carbon budget mainly addresses the carbon reserves in the atmosphere, oceans, and terrestrial (Barbour, 2021), with terrestrial ecosystems being vital for regulating the global carbon cycle (Gherardi and Sala, 2020; Landry and Matthews, 2016). Terrestrial ecosystems primarily absorb atmospheric carbon dioxide through the process of plant photosynthesis, which is crucial for regulating climate and mitigating global warming (Sellers et al., 2018; Beer et al., 2010; Cox et al., 2000). Gross primary productivity (GPP) is a critical measure of carbon exchange between terrestrial ecosystems and the atmosphere~~.~~ (Menefee et al., 2023). Accurate quantification of GPP is essential for evaluating carbon balance and comprehending the response of terrestrial ecosystems to climate change (Sellers et al., 2018).

The primary method currently used for measuring CO_2 exchange between ecosystems and the atmosphere is the eddy covariance technique (Chen et al., 2020; Yu et al., 2016). This technique precisely measures Net Ecosystem Exchange (NEE), which is the difference between the carbon released by ecosystem respiration (ER) and the carbon taken up by photosynthesis (Bhattacharyya et al., 2013). While flux observation sites based on the eddy covariance (EC) technique can dynamically monitor site-scale carbon fluxes, expanding their findings to larger regional scales remains challenging, mainly due to the sparse and spatially non-uniform distribution of flux sites (Xie et al., 2023; Jung et al., 2020). Remote sensing data is widely used in ecosystem carbon cycle research as it can provide information on the spatial dynamics of vegetation and climate at a larger scale (Xiao et al., 2019). By extrapolating spatially using models that incorporate remote sensing and climate data, it is possible to estimate global GPP based on observations of GPP at the site level. Therefore, remote sensing has become a crucial data resource for estimating GPP (Cai et al., 2021; Xiao et al., 2019; Wang et al., 2011).

Light Use Efficiency (LUE) models based on satellite observations are commonly employed to simulate GPP~~.~~ (Zhang et al., 2023; Zhang et al., 2015; Jiang et al., 2014). Such models include Physiological Principles Predicting Growth using Satellite data (3-PGS, Coops and Waring, 2001), the Carnegie-Ames- Stanford Approach Model (CASA, Potter et al., 1993), the Eddy Covariance–Light Use Efficiency Model (EC-LUE, Yuan et al., 2010, 2007), the MODIS Global Terrestrial Gross and Net Primary Production (MOD17, Running et al., 2004), the Vegetation Photosynthesis Model (VPM, Xiao et al., 2003), and the Vegetation Photosynthesis and Respiration Model (VPRM, Mahadevan et al., 2008). Among all the forecasting methods (Coops and Waring, 2001; Potter et al., 1993), the LUE model is widely utilized for simulating the spatio-temporal dynamics of GPP due to its simplicity and strong theoretical foundation. Over the past few decades, numerous GPP models utilizing LUE have been developed (Pei et al., 2022).

Despite significant advances in LUE theory for GPP estimation, uncertainties persist in GPP models utilizing LUE. Firstly, differences in environmental limiting factors among various LUE models contribute significantly to the uncertainty in GPP estimation. For example, Cai et al. (2014) found a strong positive correlation between water effectiveness and GPP estimate factors, while other studies found that the LUE model estimates of GPP were strongly correlated with the vegetation index, which affects the photosynthetic capacity of vegetation through leaf nitrogen content (Peltoniemi et al., 2012; Ercoli, 1993).

Recently, with the massive accumulation of satellite data and ground-based observations, more and more studies have applied machine learning (ML) methods to model ecosystem processes (Zhao et al., 2019; Alemohammad et al., 2017; Chaney et al., 2016). ML is a modeling solution that ~~is fundamentally different~~differs from simple regression models and complex simulation models in its approach. It is very effective in handling large-scale multivariate data with complex relationships between predictors (Reichstein et al., 2019; Tramontana et al., 2016). These data-driven ~~ML~~ models are ~~well~~ particularly suited for ~~addressing~~capturing nonlinear ecosystem dynamics but often require large training datasets and ~~complex issues across different ecosystems. They provide innovative~~may lack explicit links to real-world processes. However, their ability to uncover spatial patterns without process-based constraints makes them valuable for spatial predictions. Consequently, ML-based approaches ~~for simulating GPP by solving the nonlinear relationship. These models are less reliant on theoretical assumptions. Therefore, many researchers prefer this method~~have gained popularity in recent years. For example, Kong et al. (2023) developed a hybrid model that combines ML and LUE model to estimate GPP. This hybrid model improves the LUE model by integrating a machine learning approach (MLP, multi-layer perceptron), and estimates GPP using the MLP-based LUE framework along with additional required inputs.

Chang et al. ~~(2023)~~(2023) constructed RFR-LUE models that utilize the ~~RFR~~Random Forest Regression (RFR) algorithm with variables of LUE models to assess the potential of site-scale GPP estimation.

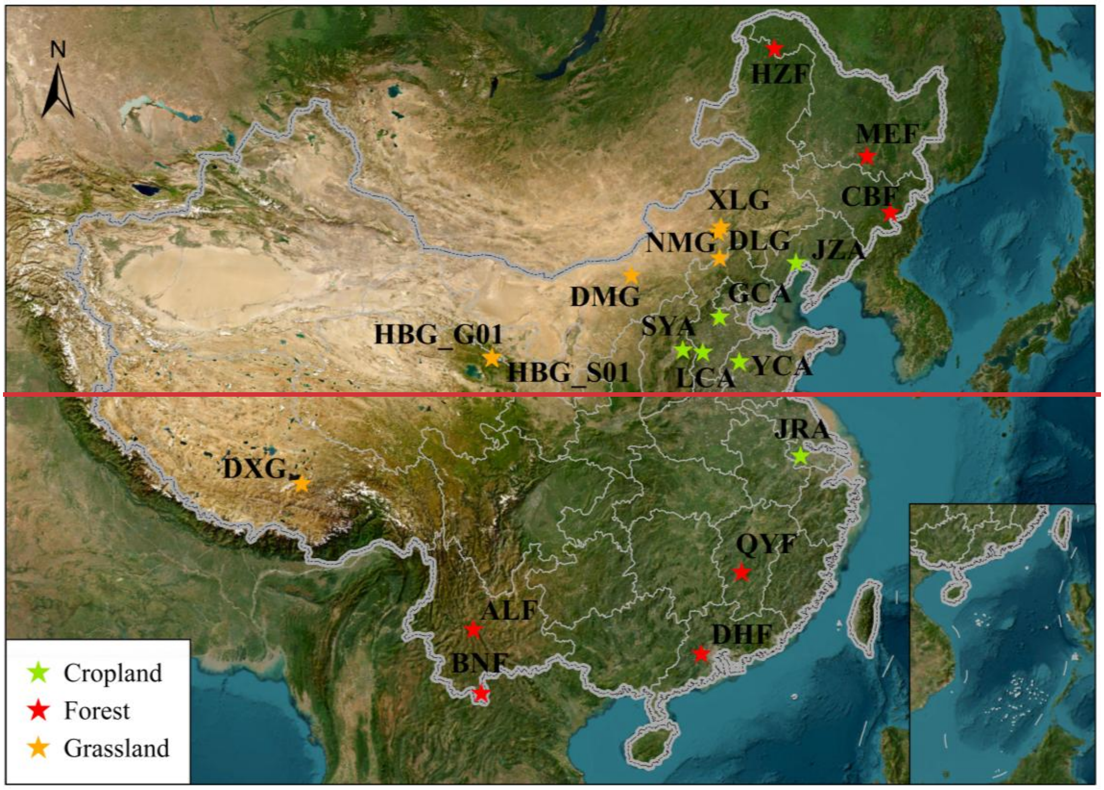
Lately, Automated Machine Learning (AutoML) has demonstrated significant potential in constructing data-driven models automatically (C. Zhang et al., 2023; Zheng et al., 2023). Numerous sophisticated open-source AutoML frameworks have been suggested by computer scientists, including ~~AutoWeka~~Automated WEKA ~~(Thornton et al., 2013), H2O~~(Auto-WEKA, Thornton et al., 2013), H2O AutoML ~~(LeDell and Poirier, 2020), TPOT~~(H2O, LeDell and Poirier, 2020), Tree-based Pipeline Optimization Tool ~~(Melanie, 2023), AutoGluo~~(TPOT, Melanie, 2023), Automated Machine Learning with Gluon ~~(Erickson et al., 2020), FLAML~~(AutoGluon, Erickson et al., 2020), Fast Lightweight Automated Machine Learning ~~(C. Wang et al., 2021), and AutoKera~~(FLAML, C. Wang et al., 2021), and AutoKeras (Rosebrock, 2019). These frameworks are extensively used in finance, manufacturing, healthcare, and mobile communications, among other fields (Adams et al., 2020), with FLAML being particularly favored for its efficiency in rapid prototyping and deployment in research and production settings. ~~FLAML (Fast Lightweight Automated Machine Learning)~~FLAML is a powerful framework for AutoML, known for its speed in identifying top-performing models and optimal hyperparameters through parallel optimization and smart search algorithms. FLAML integrates several effective search strategies, outperforming other leading AutoML libraries on large benchmarks even with constrained budgets (C. Wang et al., 2021).

In this research, a new model called FLAML-LUE was created by combining FLAML model with LUE-based models, the latter provides the key variables of vegetation growth for modeling. Such knowledge-and-data-driven models aim to reduce the large uncertainty in estimating GPP. ~~Considering the variations of the optimal moisture factor and vegetation index factor for different ecosystems (Wang et al., 2023; Wu et al., 2010), this study thus develops different models specifically for forest, grassland, and cropland ecosystems. The main goals of this study were (1) to compare the overall performance of the models simulating GPP with different input variables (moisture factor and vegetation index) and at four temporal scales; (2) to analyze monthly differences between observed and simulated values in different cover types; (3) to analyze the importance of the various input indicators for GPP modeling under different ecosystems~~The specific objectives of this study are: (1) to evaluate the overall performance of models using different input variables, including the fraction of photosynthetically active radiation absorbed by vegetation (fPAR) and various water stress indicators, across multiple sites and vegetation types based on eddy covariance observations; (2) to assess model performance under extreme climatic conditions, such as high temperature, elevated vapor pressure deficit (VPD), and drought.

2. ~~Material~~Materials and methods

2.1 Site description

~~Fig.~~**Figure 1** displays the geographical locations of the 20 flux sites selected for the study. These sites are situated in various climatic zones and ecosystem types including forest, grassland, and cropland. The observation data for these sites comes from the Science Data Bank (SDB, https://www.scidb.cn/en/). Detailed information about the sites is provided in **Table 1**.



~~Fig.~~

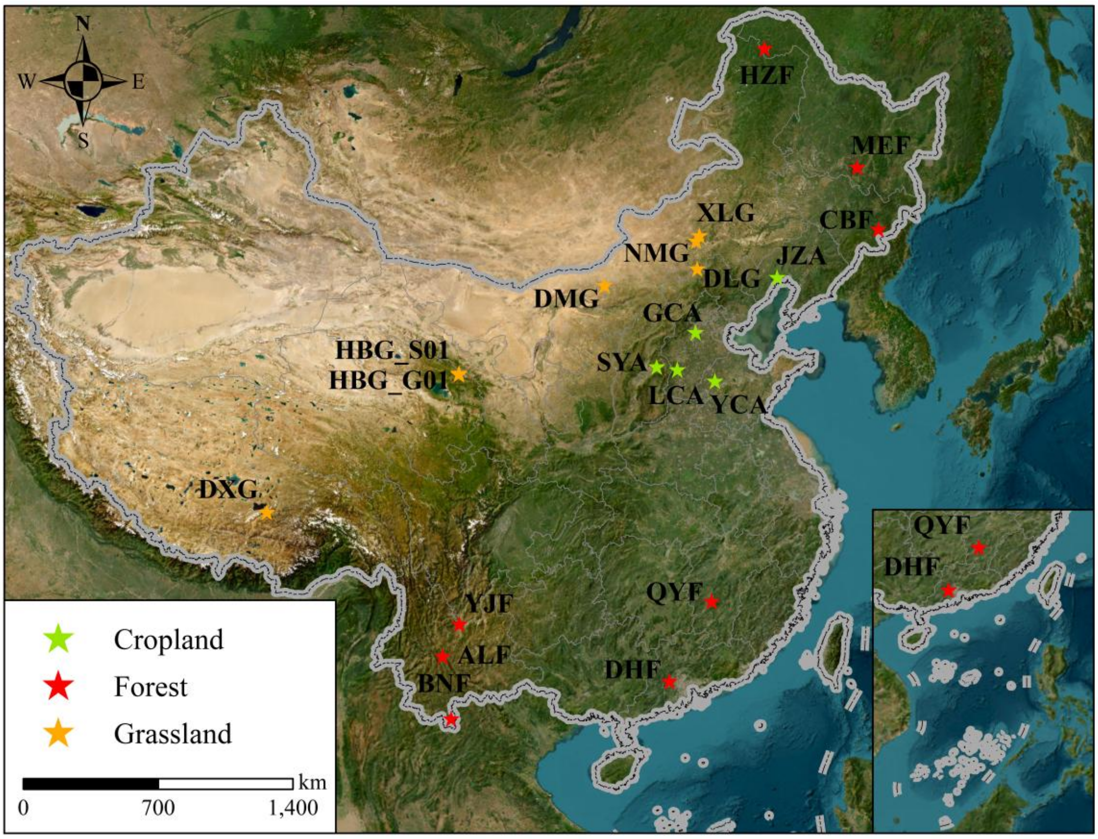


Figure 1. The location map of the flux site is based on the map approved by the National Surveying and Mapping Bureau of China (Approval No. GS (2019)1822). The topographic map is derived from data provided by Esri, Maxar, Earthstar Geographics, and the GIS User Community (Service Layer Credits).

~~**Table1**~~

Table 1

Basic information on the 20 flux stations.

Site	Longitude (°E)	Latitude (°N)	Ecosystem type	Time Range	Classified
HZF	123.018	51.781	Forest	2014-2018	~~Needle-leaved~~NF
MEF	127.668	45.417	Forest	2016-2018	~~Deciduous Broadleaved~~DBF
CBF	128.096	42.403	Forest	2003-2010	~~Mixed~~MF
QYF	115.058	26.741	Forest	2003-2010	~~Needle-leaved~~NF
~~DHF~~ALF	101.028	24.541	Forest	2009-2013~~Mixed coniferous and broad-leaved forests~~	EBF~~2003-2010~~

~~ALF~~DHF	112.534	23.173	Forest	~~Evergreen Broadleaved forests~~2003-2010	MF~~2009-2013~~
BNF	101.577	21.614	Forest	2003-2015	~~Evergreen Broadleaved~~EBF
~~XLG~~YJF	~~Grassland~~101.827	~~Mowing grasslands~~26.080	Forest~~2006-2014~~	~~Grassland~~2013-2015	SAV
~~NMG~~XLG	~~Grassland~~116.671	~~Temperate steppe~~43.554	Grassland	2006-2014	GRA
~~DLG~~NMG	116.404	43.326	Grassland	2003-2010~~Typical grasslands~~	Grassland
~~DMG~~DLG	116.284	42.047	Grassland	2006-2015~~-2018~~	Grassland
~~HBG_G01~~DMG	110.328	41.644	Grassland	2015-~~2020~~2018	~~Alpine Meadow~~Grassland
HBG_~~S01~~G01	101.313	37.613	Grassland	~~Alpine shrub-meadow~~2015-2020	~~2003-2013~~MEA
~~DXG~~HBG_S01	101.331	37.665	Grassland	2003-~~2010~~2013	~~Alpine Meadow~~SHR
~~JZA~~DXG	~~Cropland~~91.066	~~Spring corn~~30.497	Grassland~~2005-2014~~	~~Single Cropping~~2003-2010	MEA
~~GCA~~JZA	121.202	41.148	Cropland	2005-2014~~Winter wheat-Summer corn~~	SC~~2020-2022~~
~~SYA~~GCA	115.735	39.149	Cropland	2020-2022~~Spring corn~~	DC~~2012-2014~~
~~LCA~~SYA	113.200	37.750	Cropland	2012-2014~~Winter wheat-Summer corn~~	SC~~2013-2017~~
~~YCA~~LCA	114.413	37.531	Cropland	2013-2017~~Winter wheat-Summer corn~~	DC~~2003-2010~~

~~JRA~~YCA	116.570	36.829	Cropland	2003-2010~~Winter wheat - Summer rice~~	~~2015-2020~~DC

Note: Vegetation types in the table are classified based on the land cover characteristics of each flux site and are used in subsequent model simulations. NF: Needle-leaved Forest; DBF: Deciduous Broadleaved Forest; MF: Mixed Forest; EBF: Evergreen Broadleaved Forest; SAV: Savannas; GRA: Typical Grassland; MEA: Alpine Meadow; SHR: Shrubs; SC: Single Cropping; DC: Double Cropping.

2.2 Data

2.2.1 Eddy covariance data

~~Eddy covariance (~~EC~~)~~ data were collected at 20 sites, including ~~7~~8 forests sites, 7 grasslands sites, and ~~6~~5 cropland sites (**Table 1**). ~~Back third of long time series data from ALF, CBF, and QYF Stations data were used for forest model validation, and in the same way, a third of DLG, DXG, and HBG Stations data were used for grassland models validation, a third of JZA and YCA Stations data were used for cropland models validation. None of the validation data were involved in the model training process.~~

Flux and meteorological data were collected every half hour from the mentioned sites. The flux and meteorological data underwent standardized quality control and corrections, ensuring high reliability and making them suitable for validating various GPP models and remote sensing observations. However, ER data were missing at some sites ~~have no ER data, so~~(DLG, LCA, XLG). To address this~~ study is based on the nocturnal breathing extrapolation method: ~~, the Lloyd & Taylor equation (Reichstein et al., 2005; Lloyd and Taylor, 1994)~~. The shortwave radiation Rg values (10W/m^2) determined the separation of daytime and nighttime data. In this study, the response function established by the temperature of nocturnal ER data was extended to the daytime to obtain the daytime ER.~~

~~$R_{eco} = R_{eco.ref}\exp\left(E_0\left(\frac{1}{T_{ref}-T_0}-\frac{1}{T_{air}-T_0}\right)\right)$ (1)~~

~~In the above~~ was applied to estimate ER based on nocturnal respiration data. Daytime and nighttime periods were distinguished using shortwave radiation (Rg), with a threshold of 10 W/m². The temperature–response relationship derived from nighttime ER was extrapolated to estimate daytime ER. This is a commonly used method for processing flux data at flux tower sites.

$$R_{eco} = R_{eco.ref}\exp\left(E_0\left(\frac{1}{T_{ref}-T_0}-\frac{1}{T_{air}-T_0}\right)\right) \tag{1}$$

$$GPP = ER - NEE \tag{2}$$

In equation~~,~~ (1), R_{eco} is the nocturnal ecosystem respiration value, $R_{eco.ref}$ is the ER value at the reference temperature, T_{ref} is the reference temperature (298.16K), E_0 is constant (308.56K), T_0 is the minimum temperature at which respiration stops, set at 227.13K, and T_{air} is the air temperature or soil temperature (K). Daytime GPP was then estimated by subtracting NEE from the total daytime ER.

~~We can then estimate the total ecosystem productivity of the ecosystem during the day by subtracting the net ecosystem exchange from the total ER during the day.~~

~~$GPP = ER - NEE$ (2)~~

~~In the above equation, GPP represents the carbon uptake by plants during photosynthesis. ER denotes CO_2 released through ecosystem respiration from aboveground plant parts, roots, and soil, occurring both day and night. NEE reflects the net carbon gain or loss within the ecosystem.~~

~~The pre-processed flux data are first aggregated into daily, 8-day, 16-day, and monthly intervals. Then, daily values are further aggregated to 8-day, 16-day, and~~

~~monthly resolutions applying suitable methods. A detailed flow illustrating the processing of each variable is shown in Fig. 2.~~

2.2.2 ~~Remote sensing~~MODIS data

In this study, remote sensing data were primarily ~~came~~obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS~~ and~~). ~~ERA5-LAND~~~~.~~ MODIS data offer a spatial resolution of 500 meters and an 8-day temporal resolution~~, while ERA5-LAND data have a spatial resolution of 0.1° and a daily temporal resolution~~. These datasets were sourced from the Google Earth Engine (GEE) platform (Gorelick et al., 2017). To align with the spatial and temporal scales of flux tower observations and reduce the impact of missing data (Schmid, 2002), we applied the Savitzky-Golay smoothing filter with a window size of 10 to process the vegetation indices. Vegetation and water indices derived from MODIS data ~~from GEE were used to derive~~included the enhanced vegetation ~~and water indices, including~~ index (EVI~~,~~), normalized difference vegetation index (NDVI~~, LAI,~~), and land surface water index (LSWI~~,~~), which were calculated using the formulas presented in **Table 2**. ~~Temperature and PDSI index data were obtained from the ERA5-LAND product. The Maximum Value Composite (MVC) method was used to aggregate multi-temporal vegetation indices (VIs), ensuring alignment with the model simulation time steps.~~

2.2.3 ERA5-LAND

ERA5-Land (Hersbach et al., 2020) is a global high-resolution reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) under the Copernicus Climate Change Service (C3S). It provides hourly land surface variables at a spatial resolution of 0.1°, generated using a dedicated land surface model driven by the ERA5 climate reanalysis. The dataset integrates advanced land surface modeling and data assimilation techniques, offering a wide range of variables such as air temperature, soil moisture, precipitation, and snow depth. In this study, site-specific variables including air temperature (T), soil water content (SW), precipitation (Pre), and leaf area index (LAI) were extracted from ERA5-Land. In addition, photosynthetically active radiation (PAR), evapotranspiration fraction (EF), VPD and relative humidity (RH) were calculated and derived from available ERA5-Land variables using GEE.

2.2.4 SPEI Database, Version 2.10

The SPEI Database, Version 2.10 (Vicente-Serrano et al., 2010) provides global data of the Standardized Precipitation-Evapotranspiration Index (SPEI) across temporal scales from 1 to 48 months. Developed by the Climatic Research Unit (CRU), this dataset combines precipitation and potential evapotranspiration (PET) to assess drought conditions. Negative SPEI values indicate drought, while positive values signify wet periods. In this study, SPEI values less than -1.5 were used to identify drought months at each flux station, highlighting significant moisture deficits that affect vegetation growth and ecosystem productivity (Qian et al., 2024).

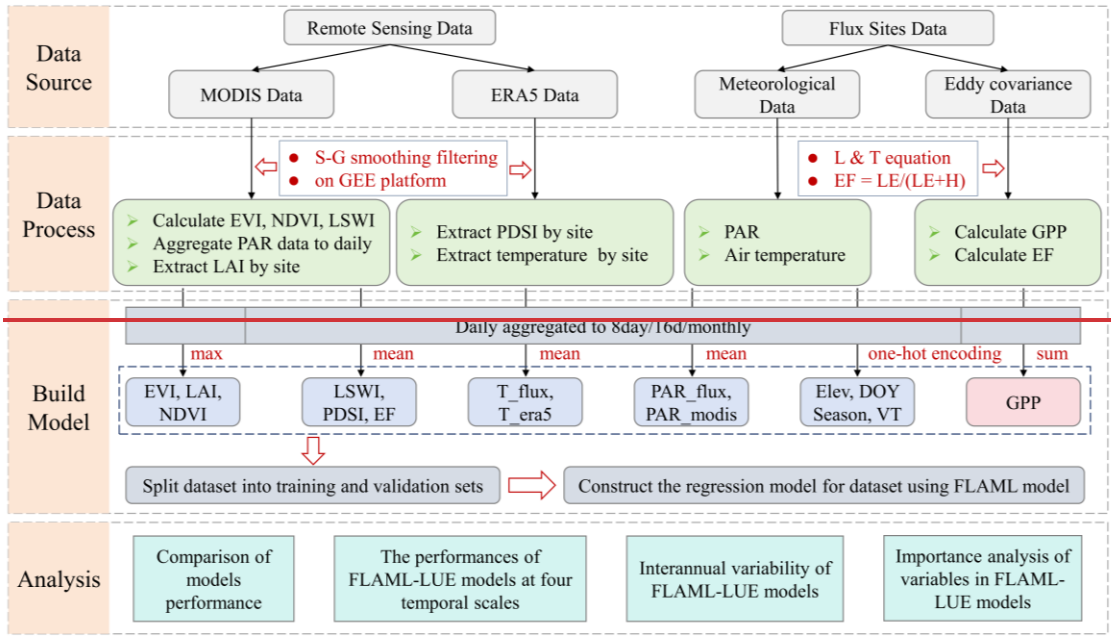
2.3 Model Construction

~~Most LUE models usually have four groups of variables: PAR, VIs, temperature, and water. In past studies, NDVI, EVI, or LAI were used as indicators of the proportion of PAR absorbed. In addition, different moisture indices were added to the LUE model~~

~~to account for water stress, including LSWI, Palmer drought severity index (PDSI), and evapotranspiration fraction (EF) indicators. In this study, all above-mentioned variables were used to build the LUE model.~~

Most LUE models typically incorporate four main groups of variables: PAR, fPAR, temperature, and water-related stress indicators. In previous studies, vegetation indices such as EVI, NDVI, or LAI have been widely used as proxies for fPAR, representing the fraction of PAR absorbed by the plant canopy (Chang et al., 2023; Qian et al., 2024). In this study, we selected six water-related indicators based on their ecological relevance: plant-based indicators (LSWI and EF), soil-based indicators (SW), and atmospheric indicators (VPD, precipitation, and relative humidity). Previous research has shown that plant-based indicators like LSWI and EF effectively capture canopy-level drought stress (Anderson et al., 2007; Xiao et al., 2004). Soil moisture regulates water availability at the root level, which strongly influences photosynthetic activity, particularly under water-limited conditions (Vicca et al., 2014; Reichstein et al., 2007). Meanwhile, atmospheric indicators such as VPD, precipitation, and RH influence stomatal conductance and transpiration by altering the vapor pressure gradient between the leaf surface and the surrounding air (Wang et al., 2018; Novick et al., 2016). To assess the relative importance of these different types of water stress indicators in estimating GPP, we developed machine learning models using each group individually. This allowed us to identify the most effective type of water-related variable for simulating GPP across diverse ecosystems within the LUE modeling framework.

The flowchart of this study is shown in ~~Fig.~~**Figure 2**.



~~Fig.~~

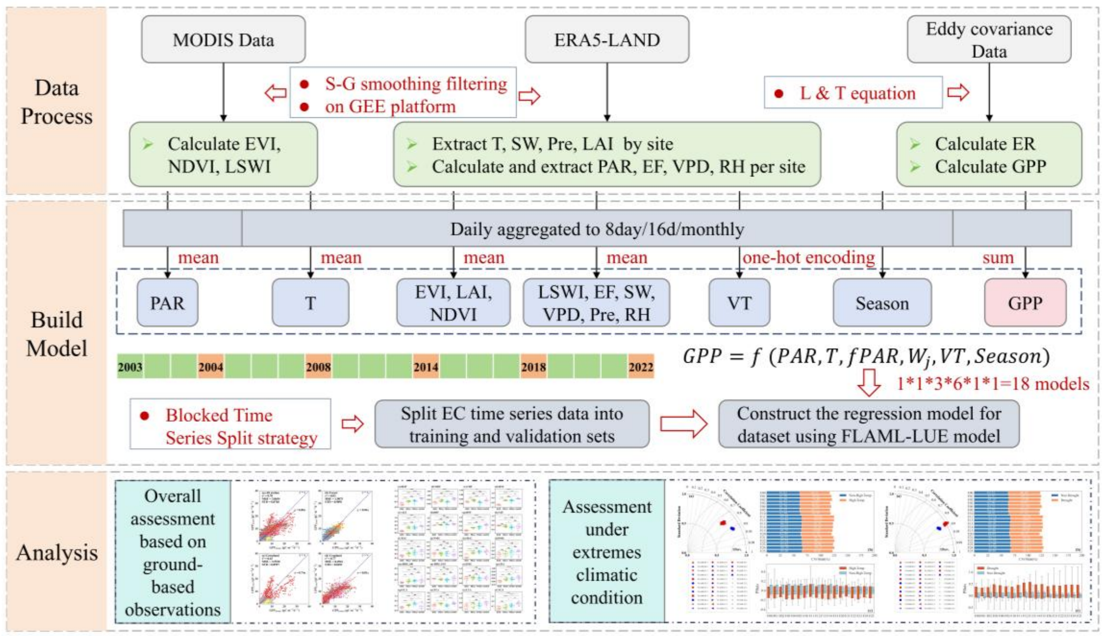


Figure 2. Flowchart of this study. S-G smoothing filtering: Savitzky-Golay smoothing filtering method, L & T equation: Lloyd & Taylor equation.

2.3.1 Data pre-processing and splitting strategy

The primary datasets for estimating GPP with FLAML-LUE models include multi-year continuous EC flux data, satellite-based observations, and ERA5-Land climate reanalysis data. Prior research (Jung et al., 2011) has demonstrated notable seasonal fluctuations in GPP, we divided the time series data into four distinct seasons. ~~Additionally, we incorporate the day of year (DOY) indicator into the model. Research has demonstrated that topography significantly influences GPP modeling (Xie and Li, 2020). Therefore, we include elevation as an additional variable. Moreover, the vegetation cover type, which varies across different ecosystems, greatly impacts the accuracy of GPP simulation (Chang et al., 2023).~~Moreover, the vegetation cover type, which varies across different ecosystems, greatly impacts the accuracy of GPP simulation (Chang et al., 2023). Hence, we integrate vegetation type as a factor in our model.

The pre-processed dataset was divided into training and testing sets using the Blocked Time Series Split strategy. Given the temporal dependency of the data, standard cross-validation is not suitable for time series analysis (Reichstein et al., 2019). Instead, a block-based and non-continuous split is applied to preserve the temporal structure. In this approach, the time series is partitioned into several non-overlapping continuous training blocks (e.g., 2003-2005, 2007-2009, 2011-2013, 2015-2017, 2019-2021), with independent years reserved as the validation set following each training block (e.g., 2006, 2010, 2014, 2018, 2022). This strategy ensures that the temporal order is maintained, preventing future data from leaking into the training process and thus avoiding invalid predictions. Additionally, the method incorporates validation over multiple periods, enabling the assessment of model generalization across different climate conditions, which is crucial for evaluating the model's robustness under varying environmental scenarios.

Table 2

Predictor variables for driving the FLAML models and their specifications.

	Variable	Acquired method (formula)	Original Spatial Resolution	Data Source
	EVI	$2.5 \times (R_{nir} - R_{red}) / (R_{nir} + 6.0 \times R_{red} - 7.5 \times R_{blue} + 1)$	500m	MOD09GA
~~Vegetation indices~~fPAR	NDVI	$(R_{nir} - R_{red}) / (R_{nir} + R_{red})$		
	LAI	-	∼10km~~500m~~	~~MCD15A3H~~ERA5-Land
	LSWI	$(R_{nir} - R_{swir}) / (R_{nir} + R_{swir})$	500m	MOD09GA
	~~PDSI~~	~~-~~	~~∼10km~~	~~ERA5~~
	EF ~~(%)~~ (%)	EF = LE~~/(~~/(LE~~+~~+H)	∼~~1km~~10km	~~SDB~~ERA5-Land
	SW (m^3/m^3)	-	∼10km	ERA5-Land
Water	VPD	$VPD = e_s - e$ $e=6.112\times\exp((17.67\times T_d)+(243.5+T_d))$ $e_s=6.112\times\exp((17.67\times T)+(243.5+T))$	∼10km	ERA5-Land
	Pre (mm)	-	∼10km	ERA5-Land
	RH (%)	$RH = (e/e_s) \times 100$	∼10km	ERA5-Land
Radiation	PAR(μmol m^{-2} s^{-1})	-	∼~~1km~~10km	~~SDB~~ERA5-Land
	~~PAR(μmol m-2 s-1)~~	~~-~~	~~500m~~	~~MCD18C2~~
Temperature	T~~_flux~~ (℃)	-	∼~~1km~~10km	~~SDB~~ERA5-Land
	~~T_era5 (℃)~~	~~-~~	~~∼10km~~	~~ERA5~~
~~Vegetation Types~~VT	EBF, DBF, ~~CF~~NF, MF, GRA, MEA, SHR, ~~Grassland Croplands~~SC, DC	One-hot encoding	invariant	-
Season	Spring, Summer, Autumn, Winter	One-hot encoding	invariant	-
~~DOY~~	~~Days of year~~	~~-~~	~~invariant~~	~~-~~
~~Terrain~~	~~Elevation~~	~~-~~	~~90m~~	~~SRTM90~~

Note: EVI: Enhanced Vegetation Index, NDVI: Normalized Difference Vegetation Index, LAI: Leaf Area Index, LSWI: Land Surface Water Index, EF: Evaporative Fraction, SW: Surface Soil Moisture, VPD: Vapor Pressure Deficit, Pre: Precipitation, RH: Relative Humidity, PAR: Photosynthetically Active Radiation, and T: Air Temperature. NF: Needle-leaved Forest; DBF: Deciduous Broadleaved Forest; MF: Mixed Forest; EBF: Evergreen Broadleaved Forest; SAV: Savannas; GRA: Typical Grassland; MEA: Alpine Meadow; SHR: Shrubs; SC: Single Cropping; DC: Double Cropping. In the formulas for EVI and NDVI, R_{nir}, R_{red}, R_{blue}, R_{swir} represent the surface reflectance in the near-infrared (NIR), red, and blue spectral bands, respectively. In the EF calculation formula, LE refers to latent heat flux, while H represents sensible heat flux. In the RH formula, e is the actual vapor pressure, e_s is the saturation vapor pressure, T_d is the dew point temperature, and T is the air temperature.

2.3.2 Automated Machine Learning (AutoML)

Instead of applying a specific ML method like RF for building regression models, we utilize the lightweight Python library ~~“~~"FLAML~~”~~" version 2.3.3 ~~(C. Wang et al., 2021)~~(Wang et al., 2021) for the AutoML task. This library refines the search process by balancing computational cost and model error, and it iteratively selects the learner, hyperparameters, sample size, and resampling strategy ~~(C. Wang et al., 2021)~~(Wang et al., 2021). For our modeling approach, we set up the AutoML for regression tasks using the "auto" option for the estimator list, focused on optimizing the R^2 metric, and used a time step of 120 seconds (2 minutes) for each AutoML run. The "auto" option includes a range of tree-based methods, such as LightGBM (Ke et al., 2017), XGBoost (Chen and Guestrin, 2016), CatBoost (Prokhorenkova et al., 2018), RF (Breiman, 2001), and Extra-Trees (Geurts et al., 2006).

2.3.3 Model development

Eighteen FLAML-LUE model variations were constructed for ~~each site and time scale~~all sites, using multiple permutations of ~~eight~~six input factor groups, as described in Eq. (3). **Table 3** displays the model number based on different variable configurations.

~~$GPP = f\ (PAR, T, VI_i, W_j, VT, Season, DOY, Elevation)$ (3)~~

~~Here~~

$$GPP = f\ (PAR, T, fPAR, W_j, VT, Season) \qquad (3)$$

where, the ~~VI_i~~$fPAR$ include EVI, NDVI, and LAI; W_j denotes moisture factors including LSWI, EF, ~~and~~SW, PDSI~~; VT_i~~, Pre, RH; VT represents vegetation types, in which forest ecosystems include: ~~Needle-leaved, Deciduous Broadleaved, Mixed, and Evergreen Broadleaved; Grassland~~EBF, DBF, NF, MF, and SAV; grassland ecosystems include ~~grasslands, meadows~~GRA, MEA, and ~~shrub~~SHR, and farmland ecosystems include ~~single cropping and double cropping. Season~~SC and DC; *Season* represents the season in which the original data were acquired. ~~DOY represents the days of the year.~~

~~Each ecosystem has 18 indicator combinations, which are divided into two groups based on different data sources, the FLAML00-FLAML08 combination uses the ground-based observations as the input data, and the FLAML10-FLAML18 combination uses remote sensing data as the main input data.~~

Table 3

~~Input data for different models~~

Input variable combinations of fPAR and water stress indicators.

Group ~~(Flux)~~	Input variables	Group ~~(RS)~~	Input variables	Group	Input variables
FLAML00	~~PAR, T_flux, EVI, LSWI, Season, DOY, Elevation, Vegetation Types~~NDVI, LSWI	FLAML10	~~PAR_modis, T_era5, EVI, LSWI, Season, DOY, Elevation, Vegetation Types~~EVI, LSWI	FLAML20	LAI, LSWI
FLAML01	~~PAR, T_flux, EVI, PDSI, Season, DOY, Elevation, Vegetation Types~~NDVI, EF	FLAML11	EVI, EF~~PAR_modis, T_era5, EVI, PDSI, Season, DOY, Elevation, Vegetation Types~~	FLAML21	LAI, EF
FLAML02	~~PAR, T_flux,~~NDVI, SW~~, Season, DOY, Elevation, Vegetation Types~~	FLAML12	~~PAR_modis, T_era5, EVI, EF, Season, DOY, Elevation, Vegetation Types~~EVI, SW	FLAML22	LAI, SW
FLAML03	~~PAR, T_flux, NDVI, LSWI, Season, DOY, Elevation, Vegetation Types~~NDVI, VPD	FLAML13	~~PAR_modis, T_era5, NDVI, LSWI, Season, DOY, Elevation, Vegetation~~	FLAML23	LAI, VPD

			~~Types~~EVI, VPD		
FLAML04	~~PAR, T_flux, NDVI, PDSI, Season, DOY, Elevation, Vegetation Types~~NDVI, Pre	FLAML14	~~PAR_modis, T_era5, NDVI, PDSI, Season, DOY, Elevation, Vegetation Types~~EVI, Pre	FLAML24	LAI, Pre
FLAML05	~~PAR, T_flux, NDVI, EF, Season, DOY, Elevation, Vegetation Types~~NDVI, RH	FLAML15	~~PAR_modis, T_era5, NDVI, EF, Season, DOY, Elevation, Vegetation Types~~EVI, RH	FLAML25	LAI, RH
~~FLAML06~~	~~PAR, T_flux, LAI, LSWI, Season, DOY, Elevation, Vegetation Types~~	~~FLAML16~~	~~PAR_modis, T_era5, , Season, DOY, Elevation, Vegetation Types~~		
~~FLAML07~~	~~PAR, T_flux, LAI, PDSI, Season, DOY, Elevation, Vegetation Types~~	~~FLAML17~~	~~PAR_modis, T_era5, LAI, PDSI, Season, DOY, Elevation, Vegetation Types~~		
~~FLAML08~~	~~PAR, T_flux, LAI, EF, Season, DOY, Elevation, Vegetation Types~~	~~FLAML18~~	~~PAR_modis, T_era5, LAI, EF, Season, DOY, Elevation, Vegetation Types~~		

Note: EVI: Enhanced Vegetation Index, NDVI: Normalized Difference Vegetation Index, LAI: Leaf Area Index, LSWI: Land Surface Water Index, EF: Evaporative Fraction, SW: Surface Soil Moisture, VPD: Vapor Pressure Deficit, Pre: Precipitation, RH: Relative Humidity.

2.3.4 Model performance evaluation methods

~~Model performance in this study was assessed in two main ways. We assessed the ability of the FLAML-LUE model to capture changes in GPP at different sites and time scales (daily, 8-day, 16-day, monthly), as well as its representativeness of interannual changes in GPP. We compared model-derived annual average GPP to EC-GPP measurements at each site and scale and analyzed standard deviations to measure the model's ability to capture the magnitude of change. Performance metrics included coefficient of determination (~~R^2~~), root mean square error (RMSE), mean bias, and regression slope between simulated and observed values. Paired t-tests were used to determine whether the differences in performance between different temporal~~

~~resolutions were statistically significant, with a significance level of 0.05. Statistical analyses were performed in Python 3.9 using the following libraries: numpy, pandas, scipy, matplotlib, sklearn, and flaml. Additionally, R was used with the following libraries: ggplot2, ggpubr, and readxl.~~

$$R = \frac{\frac{1}{T}\sum_{t=1}^{T}(f_t - \bar{f})(o_t - \bar{o})}{\sigma_f \sigma_o} \quad (4)$$

$$nuRMSE = \frac{uRMSE}{\sigma_o} = \frac{1}{\sigma_o}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[\left(f_t - \bar{f}\right) - (o_t - \bar{o})\right]^2} \quad (5)$$

$$\sigma_f = \frac{\sigma_f}{\sigma_o} = \frac{1}{\sigma_o}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\left(f_t - \bar{f}\right)\right)^2} \quad (6)$$

$$\sigma_o = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left((o_t - \bar{o})\right)^2} \quad (7)$$

~~The Taylor diagram (Taylor, 2001) visually represents the alignment between model simulations and observations by displaying the correlation coefficient (R), normalized unbiased root mean square error (nuRMSE), and normalized standard deviation (SD). The Taylor Skill Score (TSS) quantifies how closely a model's simulation aligns with observations in this diagram. It is defined as follows:~~

$$TSS = \frac{4(1+R)}{\left(\hat{\sigma}_f + \frac{1}{\hat{\sigma}_f}\right)^2 (1+R_0)} \quad (8)$$

$$\hat{\sigma}_f = \frac{\sigma_f}{\sigma_o} \quad (9)$$

~~Where σ_f and σ_o represent the standard deviations of the model simulation and observations, respectively~~To evaluate the simulation accuracy of the FLAML-LUE model in estimating GPP, we employed a suite of widely used statistical metrics to quantify the agreement between modeled and observed values (Qian et al., 2024; Chang et al., 2023; Tramontana et al., 2016). Specifically, we calculated the coefficient of determination (R^2), Pearson correlation coefficient (R), normalized unbiased root mean square error (nuRMSE), and normalized standard deviation (NSD, $\hat{\sigma}_f$), based on GPP observations from flux towers and model simulations. The Taylor diagram (Taylor, 2001) was utilized to provide a visual summary of the model's performance, incorporating R, nuRMSE, and NSD.

$$R^2 = \frac{\left[\sum_{t=1}^{T}\left(f_t - \bar{f}\right)(o_t - \bar{o})\right]^2}{\sum_{t=1}^{T}\left(f_t - \bar{f}\right)^2 \sum_{t=1}^{T}(o_t - \bar{o})^2} \tag{4}$$

$$R = \frac{\frac{1}{T}\sum_{t=1}^{T}\left(f_t - \bar{f}\right)(o_t - \bar{o})}{\sigma_f \sigma_o} \tag{5}$$

$$nuRMSE = \frac{uRMSE}{\sigma_o} = \frac{1}{\sigma_o}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[\left(f_t - \bar{f}\right) - (o_t - \bar{o})\right]^2} \tag{6}$$

$$\hat{\sigma}_f = \frac{\sigma_f}{\sigma_o} = \frac{1}{\sigma_o}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\left(f_t - \bar{f}\right)\right)^2} \tag{7}$$

$$\sigma_o = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left((o_t - \bar{o})\right)^2} \tag{8}$$

where, o_t represents the observed GPP from the flux tower, f_t denotes the simulated GPP from FLAML-LUE model, $\bar{o}$ represents the average of observed GPP from the flux tower, $\bar{f}$ represents the average of estimated GPP from the GPP product, t represents the corresponding ID for the GPP data, and n represents the total count of GPP data for the site. σ_o represent the standard deviations of the observed GPP. A higher R^2 value indicates better consistency between the estimated GPP and the flux GPP.

In addition, the Taylor Skill Score (TSS) was computed to quantitatively assess the overall agreement between simulations and observations, with higher values indicating better performance.

$$TSS = \frac{4(1+R)}{\left(\hat{\sigma}_f + \frac{1}{\hat{\sigma}_f}\right)^2 (1+R_0)} \tag{9}$$

where σ_f represent the standard deviations of the model simulation, and R_0 denotes the maximum possible correlation coefficient (in this study, R_0 = 1). The TSS ranges from 0 to 1, with a higher TSS indicating better overall model performance relative to the observations.

~~2.3.5 Feature Importance Analysis~~

~~In Data Science, “feature importance" scores indicate how useful a feature is in predicting the target variable. These scores differ depending on the learning algorithm, resulting in varying magnitudes. For instance, Extra-Trees assesses feature importance by the reduction in mean squared error, LightGBM by the frequency a feature is used in tree splits, and XGBoost by the average information gain from splits. However, model interpretability remains a complex challenge, and there is no consensus on the best technique for determining the significance of features. Shapley Additive exPlanations (SHAP, Lundberg and Lee, 2017) provide a unified approach for model interpretation. However, their assumption that ML predictions can be broken down into individual feature contributions may not apply to highly nonlinear models (Gosiewska and Biecek, 2019). Thus, we use the default feature importance metrics from the AutoML-selected algorithm, as they are widely accepted by researchers in the field.~~

~~Then, we introduce a “ranking score” metric to standardize feature importance comparisons across various algorithms. For each estimator, features are ranked from~~

~~least to most important and assigned a score accordingly: the least important feature gets a score of 1, the next gets 2, and so on. This approach normalizes feature importance across different models, providing a unified scale for comparison, ranging from 1 (least important) to the total number of features (most important).~~

~~3. Results~~

~~3.1 Overall FLAML models performances on forest ecosystem~~

~~3.1.1 Performance Evaluation of Models~~

~~To examine the performance of each model in forest ecosystems and at the site level, the accuracy of the 18 FLAML-LUE models was evaluated using the site data from ALF, CBF, and QYF stations as the forest ecosystems model test set. The algorithms adopted by each FLAML-LUE model under the forest ecosystems are shown in Table S1. Table 4 shows the R^2, RMSE, and SD of the 18 FLAML-LUE models in the forest station test set. Cross-validation analysis shows that there are few differences between FLAML-LUE models under different combinations of input data.~~

~~**Table 4**~~

~~R^2, SD, RMSE for the forest ecosystems model test set.~~

To further investigate model bias across sites, the percent bias (PBias) was introduced (Qian et al., 2024). Positive PBias values indicate overestimation by the model, while negative values suggest underestimation. The closer the PBias is to zero, the more accurate the model's estimations. The calculation formula is as follows:

$$\cancel{\mathrm{FLAML}}PBias = \frac{\sum_{t=1}^{T}(f_t - o_t)}{\sum_{t=1}^{T} o_t} \times 100\% \qquad \cancel{R^2}(10)$$

~~FLAML00~~	~~0.90~~	~~0.864~~	~~0.974~~	~~0.311~~	~~0.9552~~

~~FLAML01~~	~~0.88~~	~~0.832~~	~~1.056~~	~~0.338~~	~~0.9412~~
~~FLAML02~~	~~0.88~~	~~0.838~~	~~1.047~~	~~0.335~~	~~0.9431~~
~~FLAML03~~	~~0.89~~	~~0.882~~	~~1.033~~	~~0.330~~	~~0.9522~~
~~FLAML04~~	~~0.89~~	~~0.888~~	~~1.027~~	~~0.330~~	~~0.9558~~
~~FLAML05~~	~~0.88~~	~~0.875~~	~~1.049~~	~~0.335~~	~~0.9521~~
~~FLAML06~~	~~0.89~~	~~0.878~~	~~1.000~~	~~0.320~~	~~0.9550~~
~~FLAML07~~	~~0.89~~	~~0.881~~	~~1.019~~	~~0.326~~	~~0.9553~~
~~FLAML08~~	~~0.89~~	~~0.875~~	~~1.022~~	~~0.327~~	~~0.9544~~
~~FLAML10~~	~~0.89~~	~~0.896~~	~~0.997~~	~~0.319~~	~~0.9606~~
~~FLAML11~~	~~0.88~~	~~0.861~~	~~1.070~~	~~0.343~~	~~0.9491~~
~~FLAML12~~	~~0.87~~	~~0.871~~	~~1.096~~	~~0.351~~	~~0.9483~~
~~FLAML13~~	~~0.88~~	~~0.876~~	~~1.053~~	~~0.337~~	~~0.9532~~
~~FLAML14~~	~~0.88~~	~~0.885~~	~~1.093~~	~~0.351~~	~~0.9528~~
~~FLAML15~~	~~0.87~~	~~0.880~~	~~1.130~~	~~0.362~~	~~0.9476~~
~~FLAML16~~	~~0.88~~	~~0.880~~	~~1.049~~	~~0.335~~	~~0.9531~~
~~FLAML17~~	~~0.89~~	~~0.964~~	~~1.015~~	~~0.325~~	~~0.9710~~
~~FLAML18~~	~~0.87~~	~~0.898~~	~~1.099~~	~~0.352~~	~~0.9551~~
~~Flux(average)~~	~~0.89~~	~~0.868~~	~~1.025~~	~~0.328~~	
~~ERA5(average)~~	~~0.88~~	~~0.890~~	~~1.067~~	~~0.342~~	
~~Forest(average)~~	~~0.88~~	~~0.879~~	~~1.046~~	~~0.335~~	

To evaluate the model's ability to capture GPP dynamics under extreme climate conditions, we identified heatwaves and high VPD events using the 95th percentile of historical meteorological records (Stefanon et al., 2012; Anderson and Bell, 2010). Drought events were defined as months with SPEI less than -1.5 (Ayantobo et al., 2019; Gumus, 2023) . These definitions enabled us to evaluate model performance under extreme environmental stresses (Qian et al., 2024, 2023).

$$CV_{Rmse} = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(f_t - o_t)^2}}{\bar{o}} \times 100\% \qquad (11)$$

To determine whether model performance differed significantly across temporal resolutions (daily, 8-day, 16-day, and monthly), we conducted paired t-tests at a 0.05

significance level. All statistical analyses were performed in Python 3.9 using libraries including numpy, pandas, scipy, matplotlib, sklearn, and flaml. Complementary visualizations were produced in R using ggplot2, ggpubr, and readxl.

3. Results

3.1 Overall Model Evaluation Based on Ground-Based Observations

To evaluate the model performance at the site level, the accuracy of the 18 FLAML-LUE models was assessed using test datasets from individual flux tower sites. The algorithms adopted by each FLAML-LUE model are shown in **Table S1**. **Figure 3** presents the R, nuRMSE, and NSD values for the 18 models. As shown in **Figure 3u**, the model performance shows relatively small differences across different combinations of input indicators. Specifically (**Table 4**), the overall R^2 of the different FLAML-LUE models ranged from 0.78 to 0.82, while nuRMSE values ranged from 0.4240 to 0.4670.

Among the fPAR-related indices, the model driven by EVI performed slightly better ($R^2 = 0.82$, nuRMSE = 0.4265) than those driven by NDVI ($R^2 = 0.80$, nuRMSE = 0.4524) and LAI ($R^2 = 0.79$, nuRMSE = 0.4561). Regarding moisture stress indicators, the model using LSWI as input achieved the best performance ($R^2 = 0.82$, nuRMSE = 0.4298), followed by those using VPD ($R^2 = 0.80$, nuRMSE = 0.4455) and RH ($R^2 = 0.80$, nuRMSE = 0.4450). Models driven by EF ($R^2 = 0.80$, nuRMSE = 0.4487), SW ($R^2 = 0.80$, nuRMSE = 0.4505), and Pre ($R^2 = 0.80$, nuRMSE = 0.4503) performed slightly worse, though the differences were minimal.

As shown in **Table 5**, the performance of the FLAML-LUE model varies considerably across different sites, with the average R^2 ranging from 0.17 at DXG to 0.92 at CBF and HBG_G01. Notably, this variation was primarily attributed to site-level differences rather than the combinations of input indicators (**Figure 3**), highlighting the influence of land cover type and climatic conditions on model performance.

The best model performance was observed at the HZF, MEF, CBF and HBG_G01 sites ($R^2 > 0.85$, TSS > 0.9), followed by QYF, DLG, JZA, and SYA ($R^2 > 0.75$, TSS > 0.88). Within forest ecosystems, the model performed better in MF, NF, and DBF than in EBF (ALF, BNF) and savannas (YJF). MF, which include both evergreen conifers and deciduous broadleaf species, exhibit distinct seasonal variations that can be effectively captured by satellite imagery. In contrast, EBF show minimal seasonal greenness variation, leading to larger modeling bias in GPP estimation.

In grassland ecosystems, the model performed better for shrublands and typical steppe than for alpine meadows (**Tables S4 and S5**). Alpine meadows, characterized by short growing seasons and harsh high-altitude climates, often experience strong environmental disturbances and large GPP fluctuations, making them more difficult to model accurately. In contrast, typical steppe and alpine shrublands display clearer phenological rhythms and stronger photosynthetic activity, making their GPP dynamics easier to capture.

In cropland ecosystems, all sites demonstrated relatively strong model performance ($R^2 > 0.6$, TSS > 0.80). Compared to natural grasslands or alpine meadows, croplands are usually monocultures with stable phenology and simpler canopy structures, which aid in more accurate GPP modeling.

Notably, at the DXG site, the model achieved a high TSS (0.8326) but a relatively low R^2 (0.17), primarily due to the large performance variation among different index combinations. As shown in **Table S4**, all six NDVI-driven models (FLAML10-FLMAL15) have negative R^2 values, significantly reducing the overall model accuracy at this site.

Table 4

Summary of evaluation metrics for FLAML-LUE model performance across all validation sites.

FLAML	R^2	R	NSD	nuRMSE	TSS
FLAML00	**0.82**	0.91	0.8806	0.4240	**0.9378**
FLAML01	0.82	0.90	0.8717	0.4301	0.9340
FLAML02	0.82	0.90	0.8810	0.4299	0.9365
FLAML03	0.82	0.91	0.8748	0.4250	0.9360
FLAML04	0.82	0.91	0.8763	0.4254	0.9363
FLAML05	0.82	0.91	0.8691	0.4244	0.9346
FLAML10	0.82	0.90	0.8638	0.4277	0.9323
FLAML11	0.79	0.89	0.8641	0.4620	0.9237
FLAML12	0.79	0.89	0.8686	0.4597	0.9256
FLAML13	0.79	0.89	0.8592	0.4539	0.9244
FLAML14	0.79	0.89	0.8629	0.4585	0.9243
FLAML15	0.80	0.89	0.8671	0.4525	0.9271
FLAML20	0.81	0.90	0.8610	0.4376	0.9291
FLAML21	0.79	0.89	0.8551	0.4542	0.9230
FLAML22	0.79	0.89	0.8597	0.4618	0.9225
FLAML23	0.79	0.89	0.8562	0.4577	0.9225
FLAML24	0.78	0.88	0.8543	**0.4670**	0.9194
FLAML25	0.79	0.89	0.8590	0.4582	0.9232
Statistics					
EVI	**0.82**	0.90	0.8756	0.4265	**0.9359**
NDVI	0.80	0.89	0.8643	0.4524	0.9262

LAI	0.79	0.89	0.8576	0.4561	0.9233
LSWI	**0.82**	0.90	0.8685	0.4298	**0.9330**
EF	0.80	0.89	0.8636	0.4487	0.9269
SW	0.80	0.89	0.8698	0.4505	0.9282
VPD	0.80	0.90	0.8634	0.4455	0.9276
Pre	0.80	0.89	0.8645	0.4503	0.9267
RH	0.80	0.90	0.8650	0.4450	0.9283

Note: The statistics represent the mean values of R^2, R, NSD, nuRMSE, and TSS across all combinations in which the respective variable was involved. Bold numbers indicate the highest values, while underlined numbers represent the lowest values.

Table 5

Mean evaluation metrics for different combinations of fPAR and water stress indicators at each site.

Station Name	R^2	R	NSD	nuRMSE	TSS
HZF	0.85	0.93	0.9839	0.3685	0.9650
MEF	0.91	0.96	0.8989	0.2918	0.9679
CBF	**0.92**	0.97	0.8687	0.2716	0.9644
QYF	0.75	0.89	0.8171	0.4677	0.9057
ALF	0.64	0.83	0.6250	0.5950	0.7387
DHF	0.55	0.75	0.7831	0.6671	0.8224
BNF	0.37	0.67	0.8119	0.7540	0.8003
YJF	0.43	0.68	0.6702	0.7348	0.7130
XLG	0.49	0.75	0.9877	0.6980	0.8736
NMG	0.40	0.64	0.6334	0.7685	0.6673
DLG	0.78	0.89	0.9509	0.4543	0.9425
DMG	0.59	0.78	0.6941	0.6204	0.7742
HBG_G01	**0.92**	0.96	0.9040	0.2750	**0.9715**
HBG_S01	0.53	0.82	1.1390	0.6556	0.8945
DXG	0.17	0.83	1.3421	0.7631	0.8326
JZA	0.80	0.91	0.7697	0.4373	0.8916
GCA	0.62	0.82	0.9519	0.5950	0.9014
SYA	0.81	0.92	0.7606	0.4294	0.8854
LCA	0.64	0.80	0.7830	0.5898	0.8488
YCA	0.64	0.80	0.7117	0.5991	0.8043
All	0.80	0.90	0.8658	0.4450	0.9285

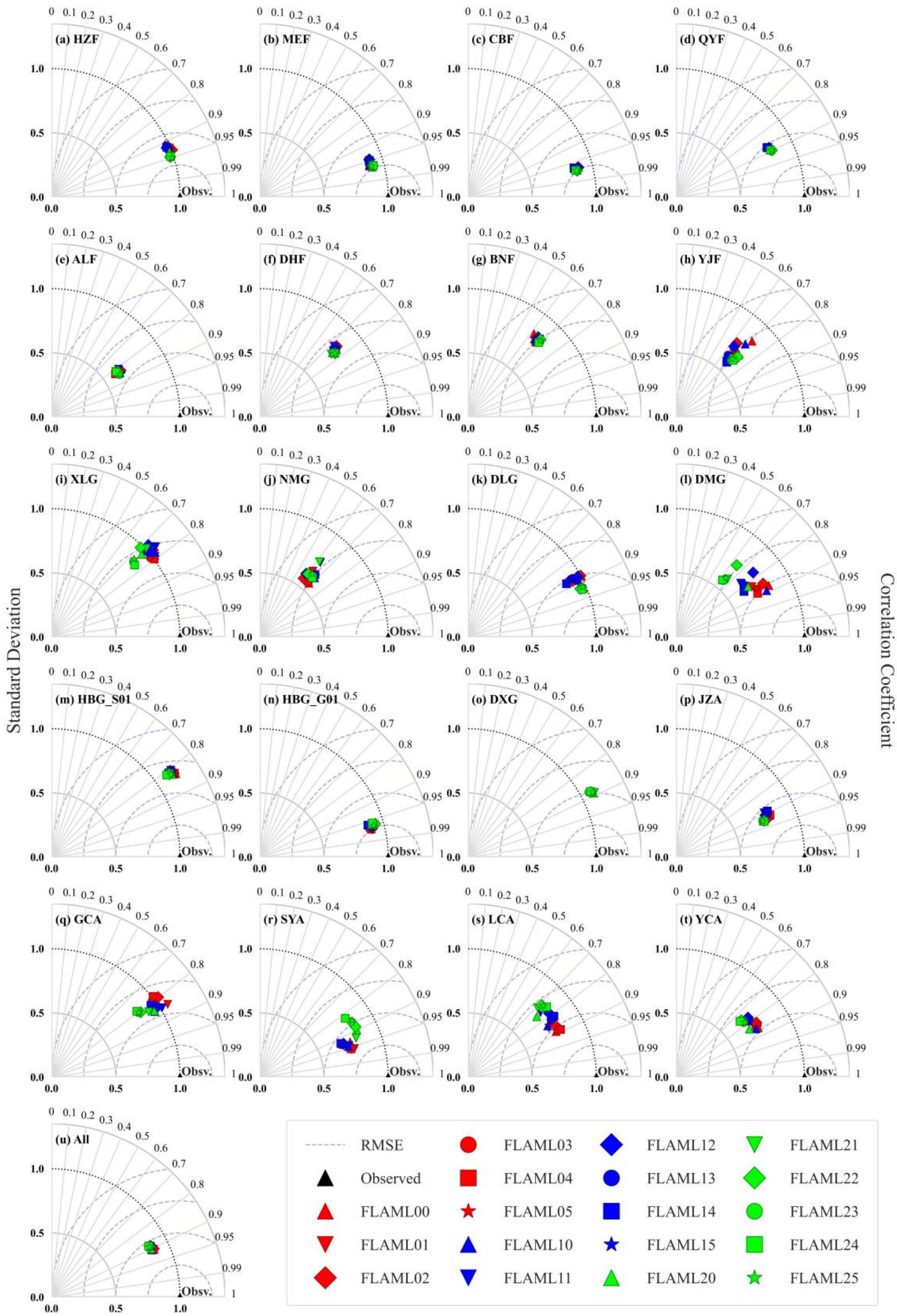


Figure 3. Normalized Taylor diagrams showing the performance of the FLAML-LUE model at various sites based on observed GPP data. Each point represents a specific combination of fPAR and water stress factor used in the model simulation. Different colors denote different fPAR products:

red for EVI, blue for NDVI, and green for LAI. Marker shapes indicate the type of water stress factor: "+" for LSWI, "×" for EF, diamond for SW, circle for VPD, square for Pre, and star for RH. Points closer to the reference point (R = 1, NSD = 1) indicate better agreement between simulated and observed GPP. Panels (a)–(h) correspond to eight forest sites, (i)–(o) to seven grassland sites, (p)–(t) to five cropland sites, and (u) presents an overall model evaluation on the validation dataset across all sites.

From an ecosystem perspective, **Table 7** indicate that the FLAML-LUE model achieves the highest fitting accuracy in forest ecosystems (R^2 = 0.83, nuRMSE = 0.4162), followed by cropland ecosystems (R^2 = 0.72, nuRMSE = 0.5258), and the lowest in grassland ecosystems (R^2 = 0.71, nuRMSE = 0.5407). The slope of the fitted line in **Figure 7** is less than 1 for all ecosystem types, indicating that the FLAML-LUE model tends to underestimate GPP, particularly in croplands and grasslands.

Tables S2, S3, and **Table 6** collectively demonstrate that the model's performance varies across ecosystem types depending on the choice of fPAR-related variables. In forest ecosystems, the model is relatively insensitive to different fPAR and water-related inputs, with the LAI-driven model achieving the best performance. This can be attributed to LAI's ability to capture forest canopy structure, thereby improving fPAR estimates. In contrast, the model's performance is more sensitive to the choice of input variables in cropland and grassland ecosystems. In croplands, the EVI-driven model performs best, followed by LAI and then NDVI, although the performance differences are moderate. In grasslands, however, the NDVI-driven model performs worst, especially at the DXG site, likely due to NDVI's sensitivity to soil background and saturation in sparse and heterogeneous vegetation. EVI, with reduced saturation and higher sensitivity to biomass, shows better performance in structured cropland areas. Overall, the EVI and LSWI driven model (FLAML00) exhibits the best performance across all ecosystem types.

To further investigate model accuracy across different land cover types, **Figure 5** presents the R^2 values of five forest types, three grassland types, and two cropland types under different models. In general, model performance varies little within the same land cover type but differs substantially across types. Specifically, DBF, NF, MF, and SC exhibit higher simulation accuracy, followed by GRA, SHR, and DC, while EBF, SAV, and MEA perform the worst. These results are consistent with the Taylor diagram in **Figure 3**.

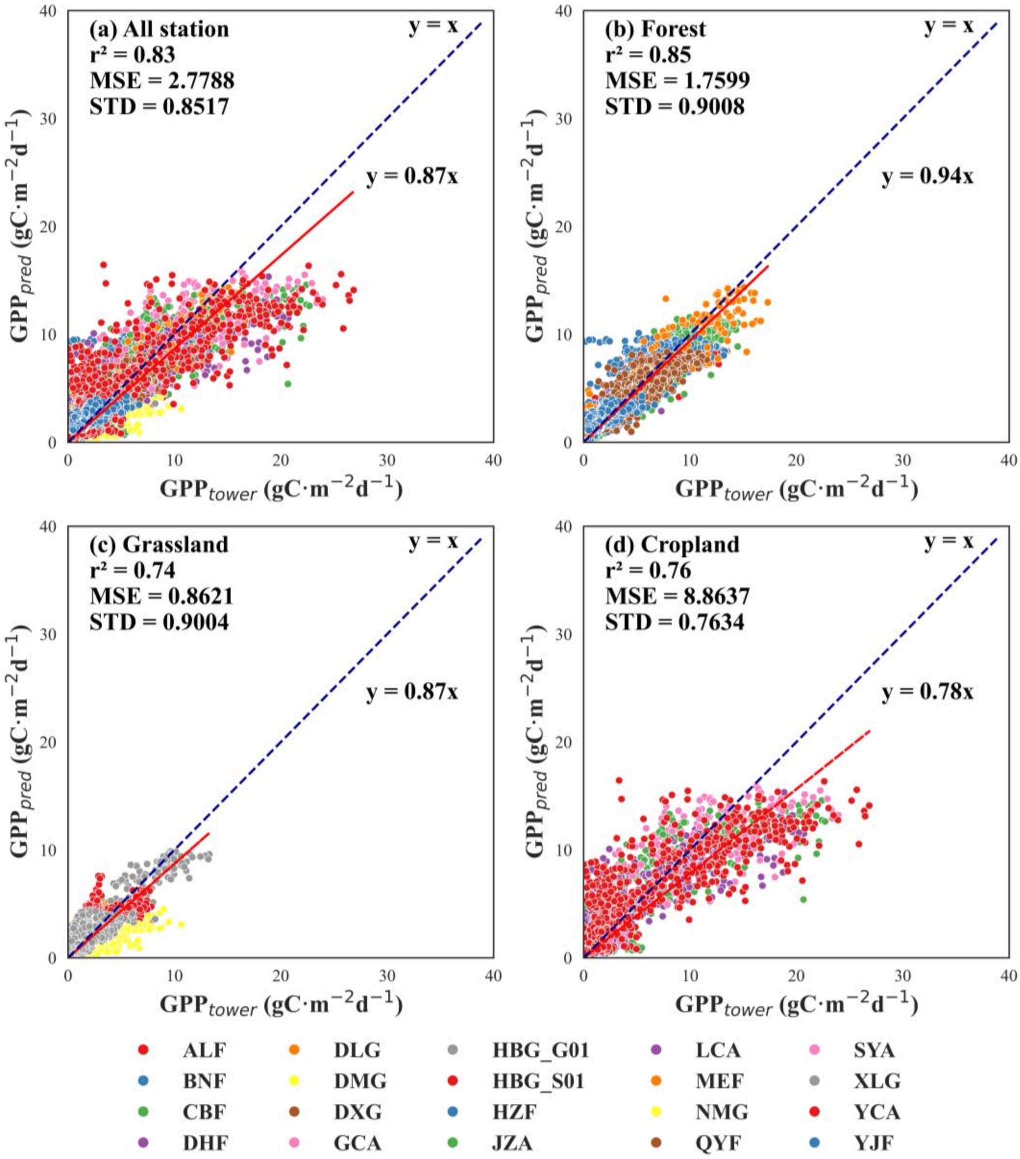


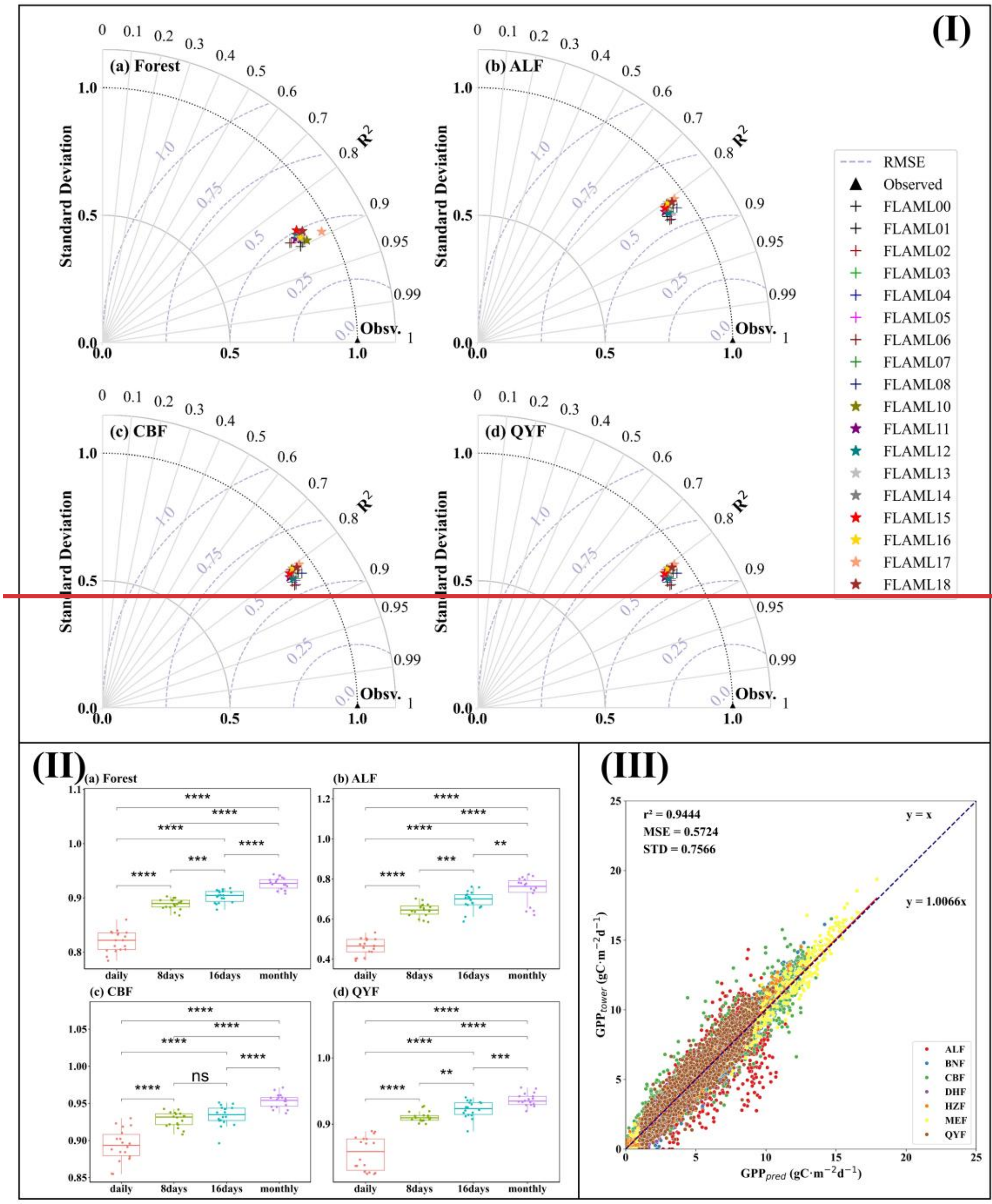
Figure 4. Scatterplot of observed GPP vs. ~~As shown in Table 4, the cross-validation analysis showed that the average R^2 for the four temporal scales under forest ecosystems was 0.82-0.93. There was little difference in performance between the models driven with flux data (FLAML00 - FLAML08, $R^2 = 0.89$, RMSE = 1.025 gC·m^{-2}d^{-1}) and the models driven with ERA5 (FLAML10 - FLAML18, $R^2 = 0.88$, RMSE = 1.067 gC·m^{-2}d^{-1}). However, the models driven using EVI ($R^2 = 0.89$, RMSE = 1.040 gC·m^{-2}d^{-1}) performed slightly better than NDVI ($R^2 = 0.88$, RMSE = 1.064 gC·m^{-2}d^{-1}) and LAI ($R^2 = 0.89$, RMSE = 1.034 gC·m^{-2}d^{-1}). The model driven with LSWI ($R^2 = 0.89$, RMSE = 1.018 gC·m^{-2}d^{-1}) performed slightly better than PDSI ($R^2 = 0.89$, RMSE = 1.047 gC·m^{-2}d^{-1}) and EF ($R^2 = 0.88$, RMSE = 1.074 gC·m^{-2}d^{-1}).~~

~~Fig. 3 shows the Taylor diagrams of the performance of all FLAML-LUE models in three forest sites: ALF, CBF, and QYF. The R^2, nuRMSE, and SD of different combinations of variables under forest ecosystems were slightly different, and the TSS values ranged from 0.9412 – 0.9710. The best performance was the FLAML17 combination with the largest TSS of 0.9710.~~

~~It is worth noting that the differences in model performance are mainly between forest types rather than different combinations of input variables. For the CBS mixed forests and QYF needle-leaf, models with various input combinations show high R^2 and low RMSE (Table S2, Table S3, Table S4). The average R^2 of the four temporal scales of CBF broadleaf Korean pine forest was 0.92-0.94, and the average R^2 of FLAML00-FLAML08 and FLAML10-FLAML18 were both 0.93 and the average RMSE was 1.153 gC·m^{-2}d^{-1}, 1.137 gC·m^{-2}d^{-1}, respectively. The average R^2 of the four temporal~~

~~scales of the coniferous forests in QYF ranged from 0.89 to 0.93, and the average R^2 of FLAML00-FLAML08 and FLAML10-FLAML18 were 0.92 and 0.90, with an average RMSE of 0.657 $gC\cdot m^{-2}d^{-1}$, and 0.719 $gC\cdot m^{-2}d^{-1}$, respectively. The model performed slightly better on the coniferous forest at QYF station than on the broad-leaved Korean pine forest at CBF station. A significant discrepancy was observed at the ALF station, which had an average R^2 for the four temporal scales ranging from 0.56 to 0.70. The average R^2 of FLAML00-FLAML08 and FLAML10-FLAML18 were 0.66 and 0.61, with average RMSE values of 1.173 $gC\cdot m^{-2}d^{-1}$ and 1.261 $gC\cdot m^{-2}d^{-1}$, respectively. In forest ecosystems, mixed forests (CBF) and evergreen needle-leaf forests (QYF) generally show better model performance than evergreen broad-leaf forests (ALF). Mixed forests, consisting of both evergreen needle-leaved and deciduous broadleaf species, display significant seasonal variations that can be effectively captured by satellite imagery. In contrast, evergreen broad-leaf forests have minimal seasonal changes in greenness, leading to higher modeling biases in GPP estimation.~~

~~A best-fit line between GPP_{tower} and GPP_{pred} was determined for all sites as one dataset using linear regression (Fig. 3 (III)). The R^2 for all sites differed less from the results for the sites analyzed individually. As shown in Fig. 3 (III), the slope of the fitted line was close to but slightly greater than 1, indicating that the FLAML-LUE model underestimated the GPP of forest ecosystems.~~



~~**Fig. 3.** (I) Normalized Taylor diagrams showing the overall performance of all FLAML-LUE models in (a) forest ecosystem, (b) ALF, (c) CBF, and (d) QYF. (II) Boxplots of 18 model performances (R^2) at different temporal scales in forest ecosystems, ALF, CBF, and QYF.~~simulated GPP. Different colored dots represent different sites. Note: The simulated GPP values represent the mean of FLAML00 to FLAML25.

Table 6

Summary of evaluation metrics for FLAML-LUE model performance across all validation sites.

FLAML	R^2	TSS

	Forest	Grass	Crop	Forest	Grass	Crop
FLAML00	0.83	0.73	0.77	0.9476	0.9241	0.9004
FLAML01	0.84	0.71	0.75	0.9472	0.9187	0.8946
FLAML02	0.84	0.70	0.75	0.9516	0.9167	0.8966
FLAML03	0.84	0.71	0.76	0.9485	0.9169	0.8971
FLAML04	0.84	0.72	0.76	0.9475	0.9157	0.8991
FLAML05	0.84	0.72	0.76	0.9487	0.9171	0.8927
FLAML10	0.83	0.72	0.76	0.9463	0.9213	0.8861
FLAML11	0.83	0.68	0.70	0.9464	0.9124	0.8696
FLAML12	0.84	0.67	0.70	0.9487	0.9091	0.8717
FLAML13	0.83	0.69	0.71	0.9459	0.9083	0.8696
FLAML14	0.83	0.69	0.70	0.9450	0.9060	0.8713
FLAML15	0.84	0.69	0.71	0.9486	0.9096	0.8746
FLAML20	0.85	0.73	0.73	0.9525	0.9219	0.8718
FLAML21	0.85	0.71	0.70	0.9531	0.9186	0.8575
FLAML22	0.86	0.70	0.68	0.9549	0.9150	0.8545
FLAML23	0.86	0.71	0.69	0.9539	0.9153	0.8535
FLAML24	0.85	0.72	0.67	0.9532	0.9145	0.8465
FLAML25	0.86	0.71	0.68	0.9542	0.9163	0.8561
Statistics						
EVI	0.84	**0.72**	**0.76**	0.9485	0.9182	0.8968
NDVI	0.83	0.69	0.72	0.9468	0.9111	0.8738
LAI	**0.85**	0.71	0.69	0.9536	0.9169	0.8566
LSWI	0.84	**0.73**	**0.75**	0.9488	0.9224	0.8861
EF	0.84	0.70	0.72	0.9489	0.9166	0.8739
SW	0.84	0.69	0.71	0.9517	0.9136	0.8743
VPD	0.84	0.70	0.72	0.9495	0.9135	0.8734
Pre	0.84	0.71	0.71	0.9486	0.9121	0.8723
RH	0.84	0.70	0.72	0.9505	0.9143	0.8745

Table 7

Mean evaluation metrics for different combinations of fPAR and water stress indicators across various ecosystems.

Ecosystem	R^2	R	$\hat{\sigma}_f$	nuRMSE	TSS

ALL	0.80	0.90	0.8658	0.4450	0.9285
Forest	**0.83**	0.91	0.8958	0.4162	**0.9431**
Grassland	0.71	0.84	0.9187	0.5407	0.9154
Croplands	0.72	0.85	0.7893	0.5258	0.8757

Note: The evaluation metrics for all sites and different ecosystem types were calculated based on the average of 18 simulation results.

Regarding CV_{RMSE}, SHR shows the largest error, followed by MEA, GRA, SC, and DC, while the five forest types show the smallest errors. This may be attributed to the greater GPP variability in grassland and cropland ecosystems, which are more strongly influenced by climatic variability and anthropogenic activities, leading to higher model uncertainty. In contrast, forest ecosystems have more stable structures and continuous carbon exchange processes, resulting in more robust model performance. Although alpine meadow is classified as grassland ecosystems, their extreme climatic conditions, short growing season, and high sensitivity to temperature and precipitation further increase the uncertainty of GPP simulation, leading to higher errors.

In terms of PBias, SHR consistently shows a pronounced overestimation across all models. Similarly, SAV and MEA are also generally overestimated in all models, though to a lesser extent than SHR. EBF exhibits a slight overestimation as well. Other vegetation types display only minor underestimation or overestimation. Overall, the models perform best for DBF, NF, and MF, followed by EBF, MEA, SC, and DC, while the simulation accuracy is relatively poor for SAV, SC, and especially SHR.

Biases also differ among grassland ecosystems, especially for typical grasslands, alpine meadows, and shrublands. Typical grasslands tend to be underestimated, while alpine meadows and shrublands are often overestimated. These biases may result from the model's limited ability to capture seasonal changes in water availability and its interaction with temperature. Typical grasslands usually show high productivity when water is sufficient, especially in spring and summer. If the model fails to reflect these seasonal patterns, it can lead to underestimation. In contrast, productivity in alpine meadows is mainly limited by low temperatures and a short growing season. If the model does not fully consider these constraints, it may overestimate photosynthesis and thus GPP. For shrublands, overestimation may be due to high spatial heterogeneity, including a mix of shrubs, grasses, and bare soil. This complexity is difficult to capture in remote sensing data (e.g., fPAR) and model inputs, leading to possible overestimation of productivity.

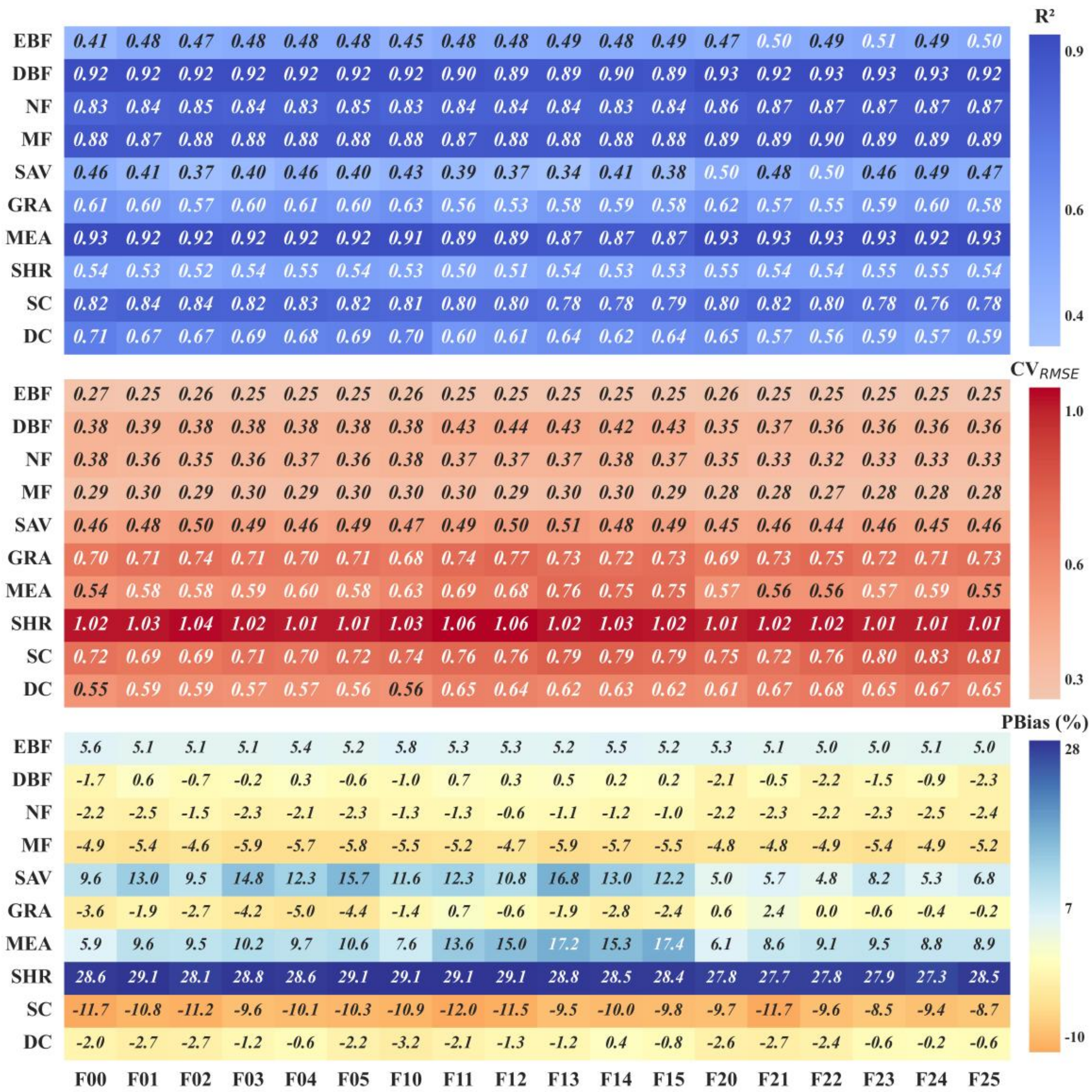


Figure 5. Comparison of R^2, CV_{RMSE}, and PBias of GPP estimates from different FLAML-LUE models across various land cover types. Note: F00 represents FLAML00, and so on.

Across the four temporal scales, the performance of the 18 FLAML-LUE models improves as the temporal resolution becomes coarser. The average R^2 across 20 sites increases from 0.64 at the daily scale to 0.74 at the monthly scale (**Table S8**), while the average nuRMSE decreases from 0.5518 to 0.4088. Paired t-tests show that, except for YJF, NMG, DMG, DXG, and YCA, the FLAML-LUE model exhibits significantly lower R^2 at the daily scale than at longer temporal scales ($p < 0.05$, **Figure 6**). For these five sites, model performance remains relatively stable across different temporal scales.

Furthermore, compared to the daily scale, the nuRMSE decreases by 12.97%, 16.52%, and 25.92% at the 8-day, 16-day, and monthly scales, respectively, indicating that the uncertainty of the FLAML-LUE model is significantly reduced at coarser temporal resolutions.

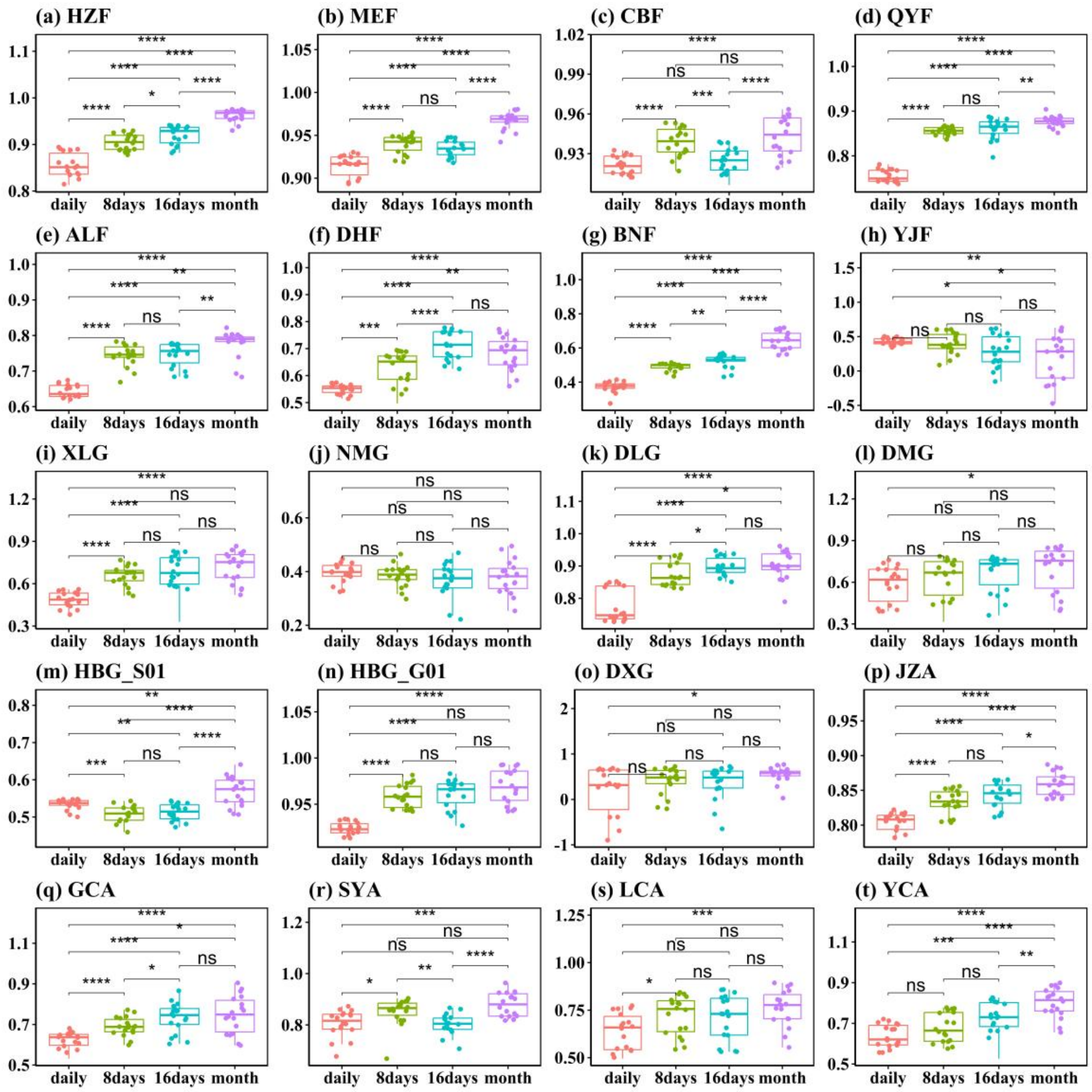


Figure 6. Asterisks indicate significant differences between the R^2 at the four temporal resolutions (Kruskal-Wallis test), ****p values < 0.0001, ***p values < 0.001, **p values < 0.01, *p values ⩽ 0.05, and ns indicates no significance (p > 0.05). ~~(III) Scatterplot of observed GPP vs. simulated GPP in forest ecosystems. Different colored dots represent different site values.~~

Overall, the accuracy of FLAML-LUE models constructed using different combinations of fPAR and water stress indicators showed limited variation, with the FLAML00 model (fPAR = EVI, water = LSWI) demonstrating the best performance. However, the model exhibited considerable differences in performance across ecosystem types, with the highest accuracy observed in forest ecosystems, followed by croplands and then grasslands. Further analysis by specific vegetation cover types revealed that the model performed best for DBF, NF, and MF, followed by GRA, MEA SC, and DC, while its performance was relatively poor for EBF, SAV, and particularly SHR (PBias > 27%, CVrmse > 1, $R^2 < 0.6$). In addition, evaluation across different temporal scales indicated that model uncertainty decreased with increasing time intervals, suggesting that the FLAML-LUE model exhibits greater robustness and reliability at coarser temporal resolutions.

3.2 Model Evaluation Under Extreme Climatic Conditions

Numerous studies have demonstrated that climate extremes such as heatwaves, droughts, and high atmospheric VPD can substantially alter ecosystem dynamics and reduce carbon uptake capacity (Frank et al., 2015; Reichstein et al., 2013). These extreme events can suppress photosynthesis, increase respiration, and disrupt the balance of carbon exchange between vegetation and the atmosphere. In order to evaluate the robustness and reliability of the FLAML-LUE models under such stress conditions, this study further investigates model performance in simulating GPP under three types of climate extremes: high temperature, high VPD, and drought. By analyzing the response of model accuracy and bias under these scenarios, we aim to assess its applicability and limitations in extreme environmental conditions.

3.2.1 Performance Under High Temperature Events

Figure 7 shows the performance of 18 FLAML-LUE models under high-temperature and non-high-temperature conditions. The results indicate a significant decline in model accuracy under high-temperature conditions. As shown in **Figure 7a**, the models perform well under non-high-temperature conditions, with the R values of all 18 FLAML-LUE models exceeding 0.9. However, under high-temperature conditions, the Taylor diagram reveals a significant decrease in model performance, with correlation coefficients dropping and a substantial increase in nuRMSE, indicating a reduced ability to capture GPP dynamics.

Interestingly, as shown in **Figure 7b**, the CV_{RMSE} values under non-high-temperature conditions are generally higher than under high-temperature conditions. This may be due to higher observed GPP values under high temperatures, resulting in a larger denominator for CV_{RMSE}, which can reduce the CV_{RMSE} despite larger prediction errors. Overall, the difference in prediction bias between high-temperature and non-high-temperature conditions is minimal.

Figure 7c shows that, under high-temperature conditions, the PBias fluctuates more significantly, with more stations showing severe overestimation or underestimation. Specifically, some models (e.g., FLAML00, FLAML01, FLAML11, FLAML15, FLAML21) overestimate GPP at certain sites under high-temperature conditions, while all models show more severe underestimation at other sites. Models driven by LAI (FLAML20 – FLAML25) exhibit smaller bias variations under non-high-temperature conditions, with PBias mainly ranging from -0.3 to 0.3.

In conclusion, high-temperature conditions increase model uncertainty, with all models exhibiting varying degrees of overestimation or underestimation across sites. Models incorporating VPD, precipitation, and relative humidity as water stress factors perform better overall, indicating greater robustness under high-temperature stress.

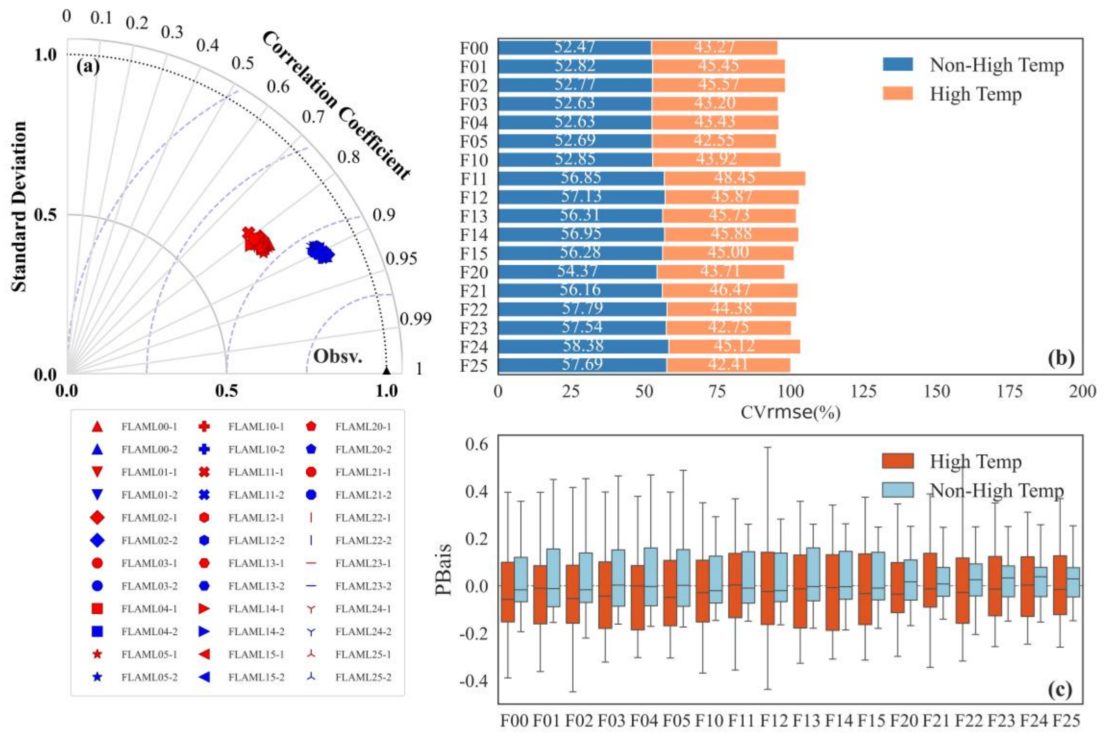


Figure 7. The comparison of GPP products performance under high temperature and non-high temperature (In the Taylor diagram, 1 represents high temperature, 2 represents non-high temperature).

Differences in model performance under high-temperature and non-high-temperature conditions are pronounced across various land cover types. **Figure 8** compares the estimation accuracy of different land cover types under both conditions. Overall, model accuracy in simulating GPP is significantly lower under high-temperature conditions, with R^2 values showing a notable decline. Specifically, for the NF type, the R^2 under high temperatures approaches a negative value, indicating very low explanatory power, whereas under non-high-temperature conditions, R^2 ranges from 0.83 to 0.87. Notably, the FLAML13 model for Savannas shows a drastic decrease in R^2 from 0.38 under non-high-temperature conditions to -1.46 under high-temperature conditions, performing even worse than the mean of the data during high temperatures.

Corresponding to **Figure 7**, CV_{RMSE} is generally lower under high-temperature conditions than under non-high-temperature conditions. The SHR type exhibits a higher coefficient of variation, while PBias shows more pronounced fluctuations. For SHR and EBF, the models tend to overestimate GPP under both temperature conditions, with overestimation more pronounced under high temperatures. In contrast, MEA shows underestimation under high-temperature conditions but overestimation under non-high-temperature conditions. Overall, most land cover types exhibit a greater degree of underestimation under high-temperature conditions. Nevertheless, the MF type maintains relatively high simulation accuracy. In contrast, the NBF, NF, and SC types are more strongly affected by high temperatures, with NF showing negative simulation accuracy under high-temperature conditions and SC exhibiting marked variations in PBias.

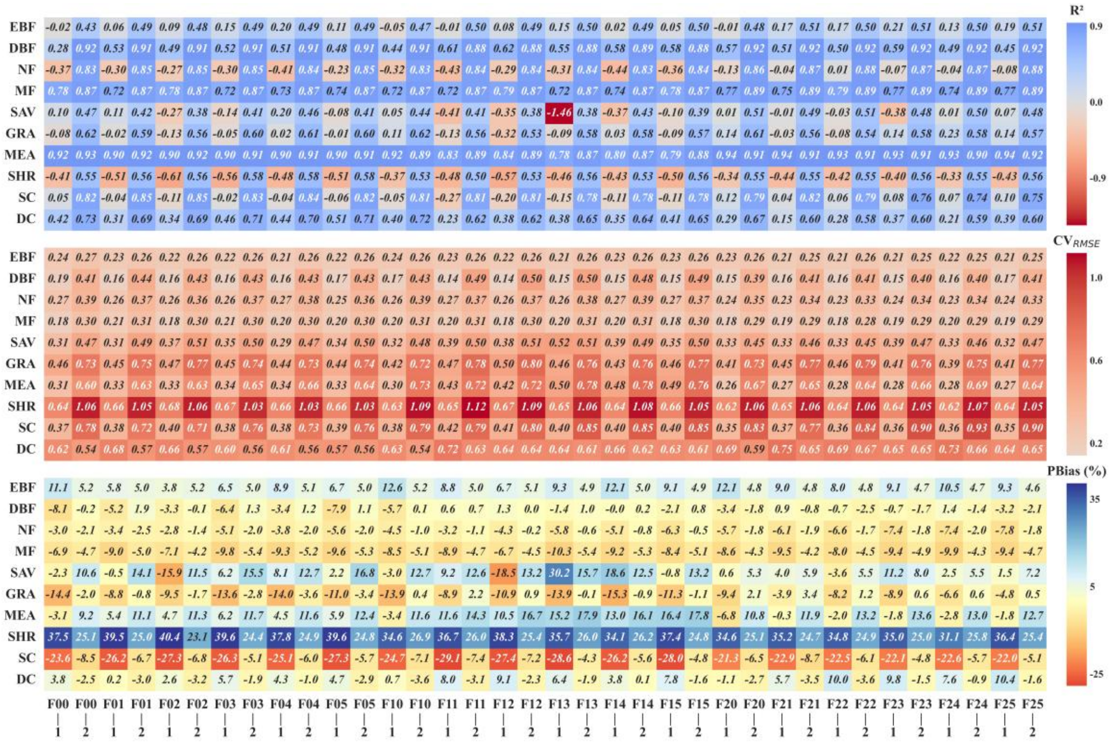


Figure 8. Comparison of statistical indicators (R^2, CV_{RMSE}, PBias) of FLAML-LUE model under high temperature conditions and non-high temperature conditions for different land cover types (1 represents high temperature, 2 represents non-high temperature).

3.2.2 Performance Under High VPD

Figure 9 shows the performance of the 18 FLAML-LUE models under high and non-high VPD conditions. Unlike the high-temperature scenario, the statistical metrics of all models exhibit only a slight decline under high VPD, indicating a less pronounced impact on model performance. As shown in **Figure 9a**, the variability in model performance increases under high VPD conditions. However, **Figure 9b** reveals that CVRMSE values are generally higher under non-high VPD conditions, a trend consistent with the results observed under high-temperature conditions.

Under high VPD, PBias exhibits significant fluctuations compared to non-high VPD conditions (**Figure 9c**). Specifically, the average PBias across sites is higher under high VPD, whereas it is lower under non-high VPD. In high VPD conditions, models driven by EVI show smaller differences in PBias across sites, with values primarily ranging from -0.4 to 0.5. In contrast, FLAML05 shows larger differences in PBias between sites under non-high VPD, with overestimations at some sites. Overall, model performance under high VPD shows greater uncertainty, with both overestimations and underestimations occurring across different sites. In general, EVI-driven models perform more consistently under both high and non-high VPD conditions.

Model performance also differs across land cover types under high and non-high VPD conditions. **Figure 10** compares the estimation accuracy for various land cover types under both conditions. Overall, GPP simulation accuracy for certain cover types (e.g., DBF, MF, MEA, SC, DC) shows little difference between high and non-high VPD conditions. Although R^2 values for some land cover types are significantly lower under high VPD than under non-high VPD, the impact of high VPD on model performance is smaller compared to high temperature. The most notable example is the FLAML13 model for Savannas, where R^2 drops significantly from -1.46 under non-high VPD to -0.39 under high VPD, performing worse than the mean data value under high VPD.

Similar to high-temperature conditions, CV_{RMSE} under high VPD is generally lower than under non-high VPD. MEA shows a larger coefficient of variation, and PBias exhibits more noticeable fluctuations. For the EBF and SHR type, models tend to overestimate GPP in both high and non-high VPD conditions, with the overestimation being more pronounced under high VPD. SC and GRA models show significant underestimation under high VPD. DBF, NF, and MF perform relatively well under high

VPD, while SC underestimates GPP under both conditions, and DC overestimates GPP under high VPD but underestimates it under non-high VPD. Overall, compared to high-temperature conditions, the effect of high VPD on estimation errors is smaller across different land cover types.

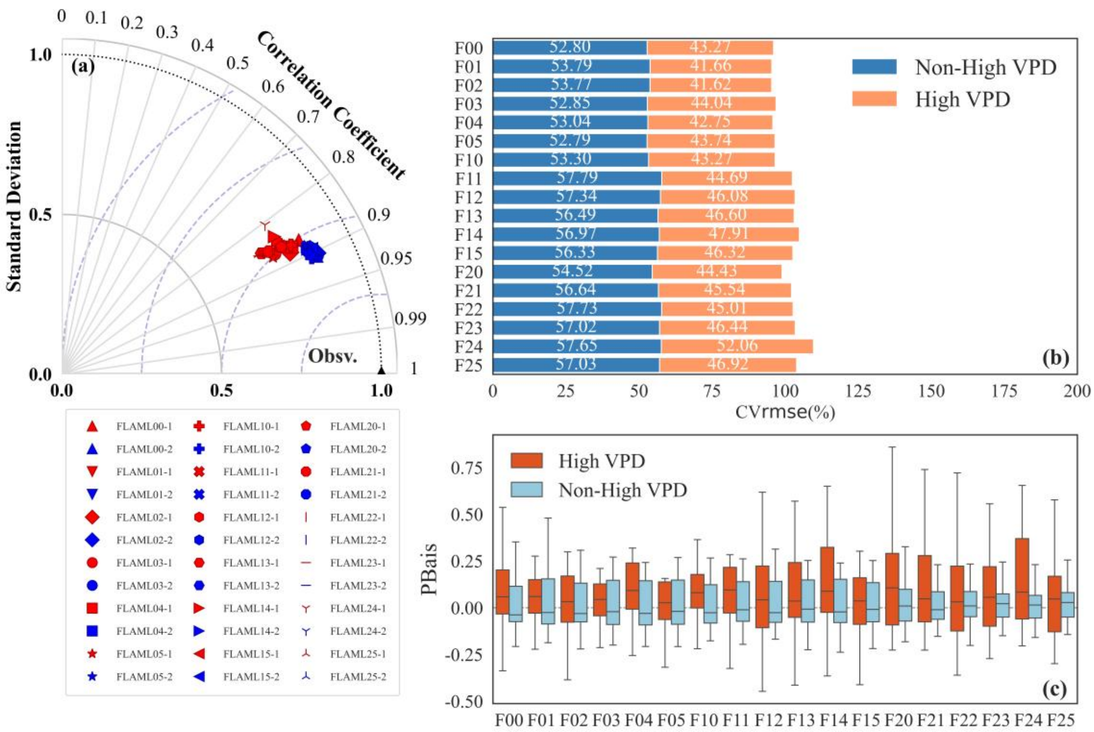


Figure 9. The comparison of GPP products performance under high VPD and non-high VPD (In the Taylor diagram, 1 represents high VPD, 2 represents non-high VPD).

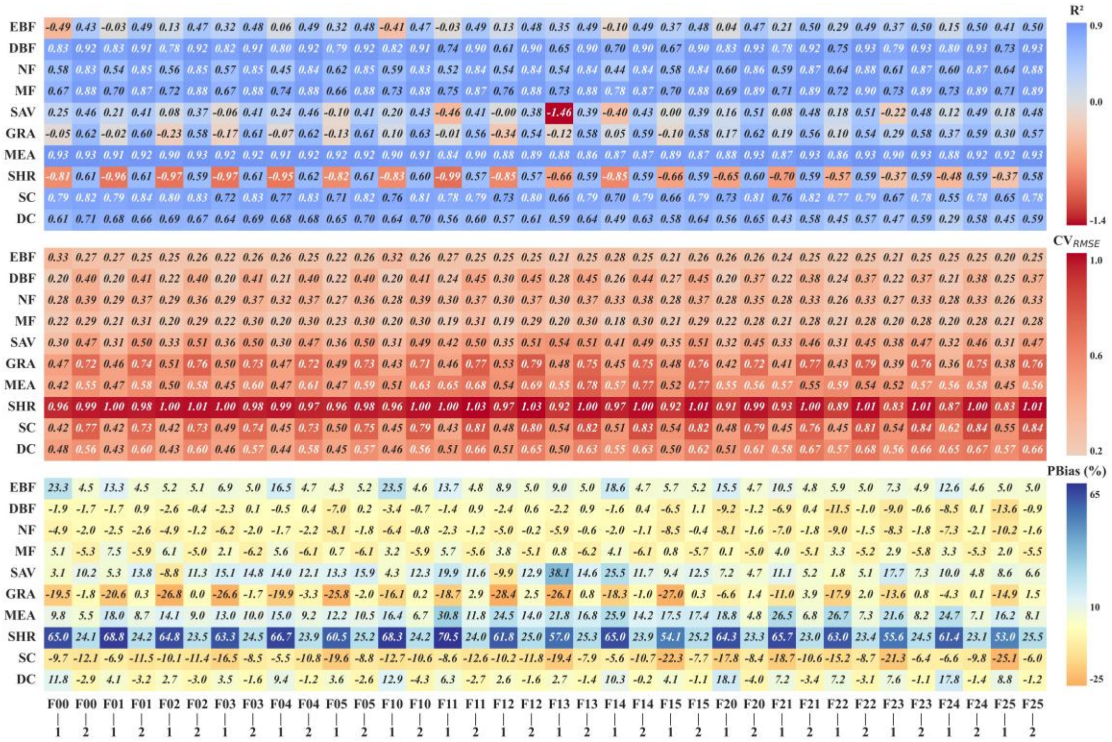


Figure 10. Comparison of statistical indicators (R^2, CV_{RMSE}, PBias) of FLAML-LUE model under high VPD conditions and non-high VPD conditions for different land cover types (1 represents high VPD, 2 represents non-high VPD).

3.2.3 Performance Under Drought Conditions

Figure 11 presents the simulation performance of the 18 FLAML-LUE models under drought and non-drought conditions. Unlike the decline in performance under high temperature and high VPD conditions, the model shows similar or even slightly better accuracy under drought compared to non-drought conditions. This may be attributed to an overall reduction in GPP and its variability during drought periods, which potentially makes it easier for the models to capture the general trend and thereby improves simulation accuracy.

Compared to the boxplots under non-drought conditions, drought notably increases the variability in PBias across sites for all models, particularly due to substantial overestimation at certain sites. In contrast, the degree of underestimation remains similar to that under non-drought conditions. Among the models, those driven by EVI exhibit the best overall performance, followed by those using LAI as the vegetation indicator.

Figure 12 shows that drought substantially affects GPP estimation accuracy across most land cover types. For certain types, such as savannas and deciduous broadleaf forests, no data were available during drought months, making performance evaluation under drought impossible. For other land cover types, the impact of drought varies significantly. Specifically, EBF, MEA, and DC show higher R^2 values under drought, while NF, MF, GRA, SHR, and SC perform better under non-drought conditions. Among them, MF and SHR have the lowest simulation accuracy under drought but perform relatively well during non-drought periods.

Regarding CV_{RMSE}, all land cover types except MEA and NF exhibit lower values under drought conditions, consistent with the results in **Figure 11a**. MEA shows the largest coefficient of variation, indicating greater variability in model performance under drought. In terms of PBias, NF, MEA, and SHR exhibit the highest errors. On average, model errors increase under drought across most land cover types. Except for EBF and GRA, most types show severe overestimation or underestimation during drought periods.

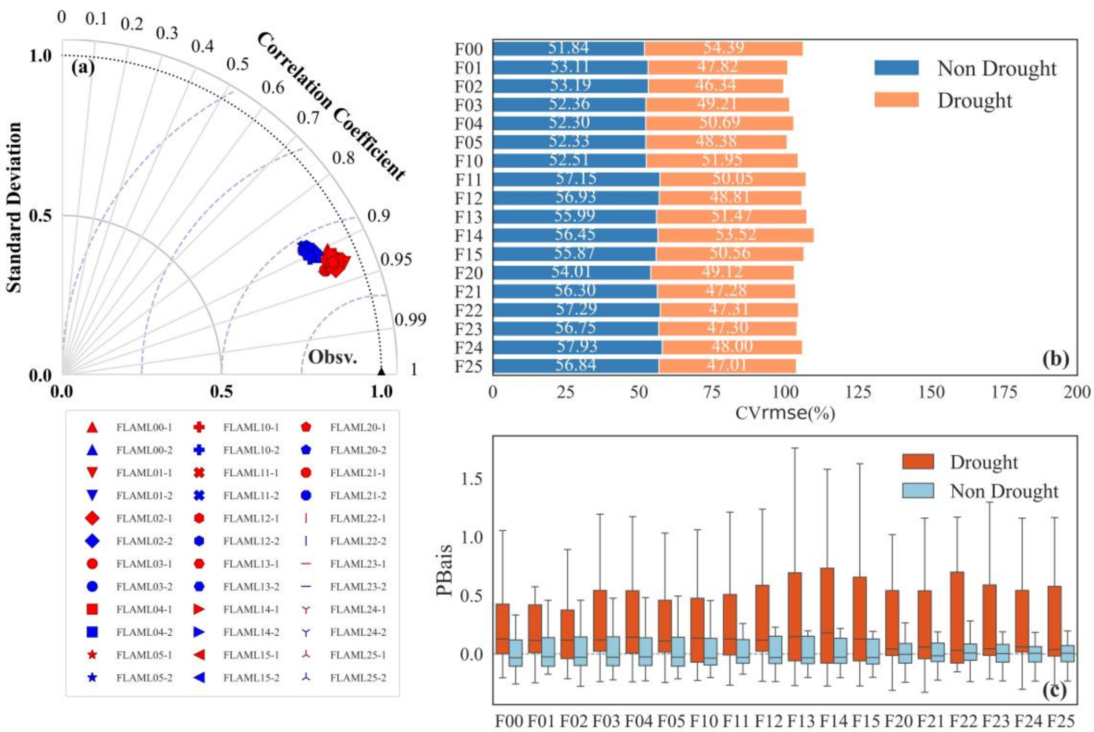


Figure 11. The comparison of GPP products performance under drought and non-drought (In the Taylor diagram, 1 represents drought, 2 represents non drought).

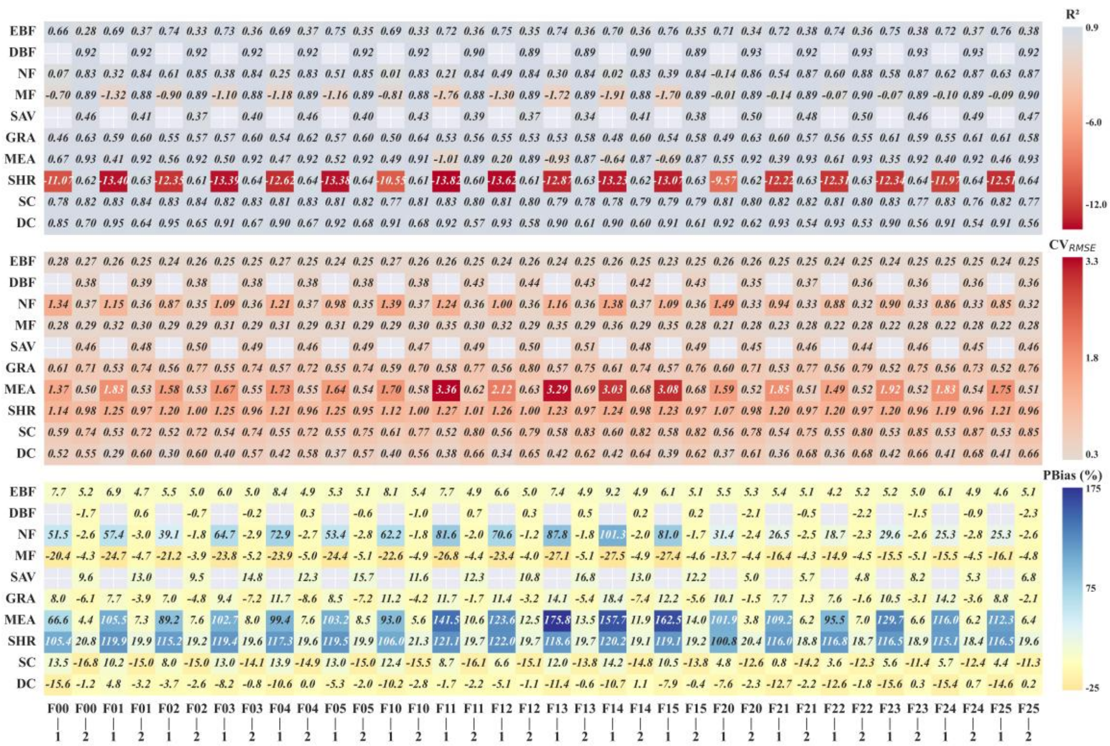


Figure 12. Comparison of statistical indicators (R^2, CV_{RMSE}, PBias) of FLAML-LUE model under drought conditions and non-drought conditions for different land cover types (1 represents drought, 2 represents non-drought).

4. Discussion

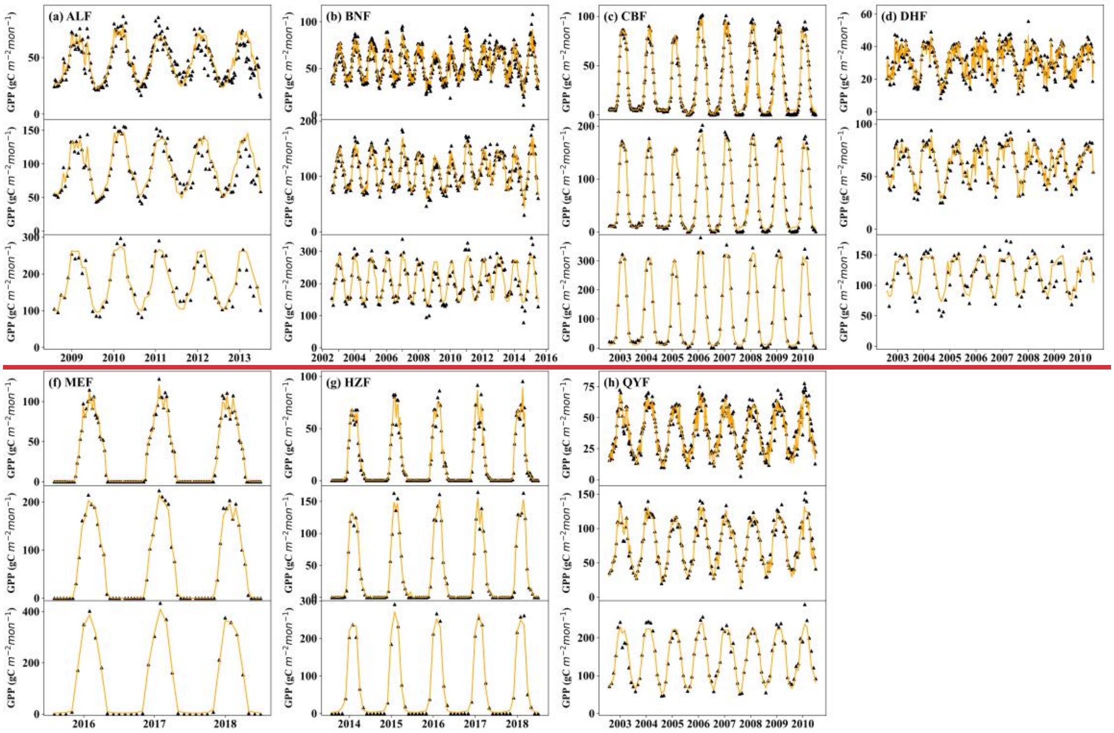
Model performance is highly influenced by the algorithms used, the underlying processes, and how GPP responds to varying environmental conditions (Chang et al., 2023). A detailed comparison of the FLAML-LUE models across different ecosystems showed that performance varied depending on the input variables, vegetation types, and time scales (Chang et al., 2023; Harris et al., 2021).

4.1 Performance comparison of FLAML-LUE models for different ecosystems

~~Under forest ecosystems, for all four temporal scales, the 18 FLAML-LUE models showed better accuracy as time aggregates to larger intervals. as shown by the increased R^2 from 0.82 to 0.93. Paired t-tests revealed that the daily performance (R^2) of the FLAML-LUE model was significantly lower than that of the other temporal scales across all sites ($p < 0.01$, Fig. 3(II)). In addition, the RMSE of the 8-day, 16-day, and monthly GPP (FLAML-LUE) also decreased significantly by 26.88%, 33.18%, and 41.34%, respectively, when compared to the daily-scale results, suggesting that the uncertainty in these models becomes smaller at the longer temporal scale. The slopes of the linear regression relationships between the simulated and observed GPP approach 1 with improving temporal resolution at ALF, CBF, and QYF sites.~~

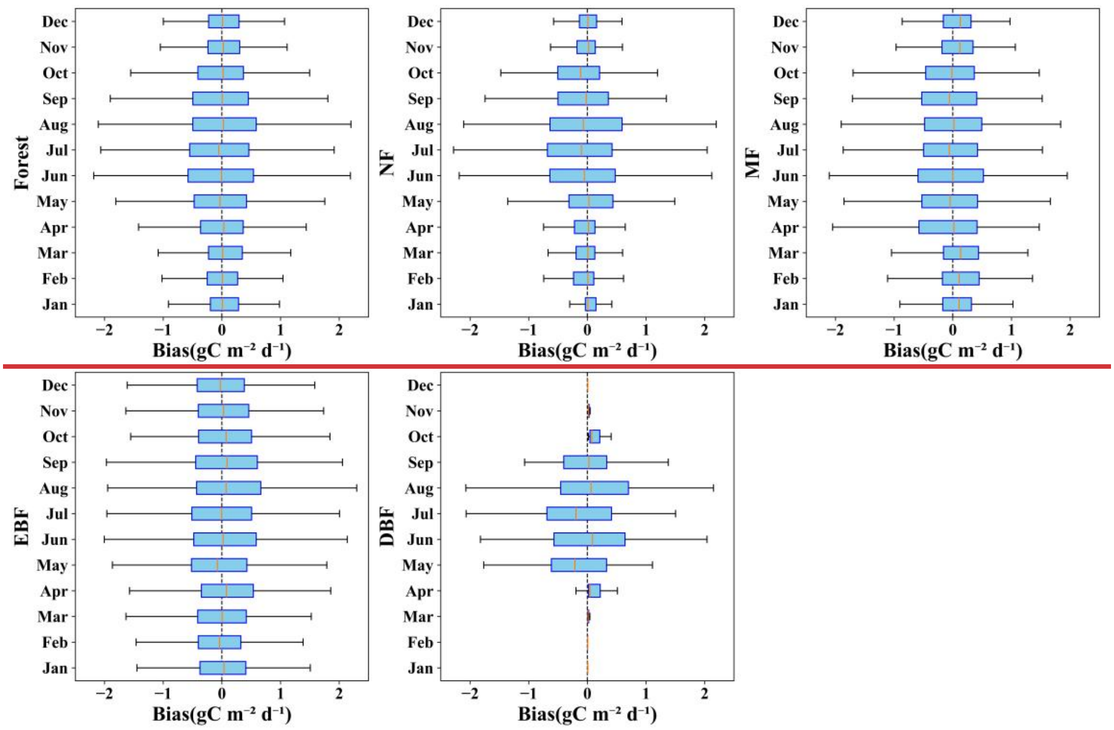
~~3.1.2 Analysis of interannual GPP variability~~

~~Based on the Taylor diagram TSS skill scores, it can be seen that the forest ecosystems have the highest GPP simulation accuracy under the combination of FLAML17 indicators, as shown in Table S2.~~



~~**Fig. 4.** Plot of simulated GPP time-series variation at ALF, BNF, CBF, DHF, HZF, MEF, and QYF sites, with black triangles representing tower-based observations and orange solid lines representing model-simulated data.~~

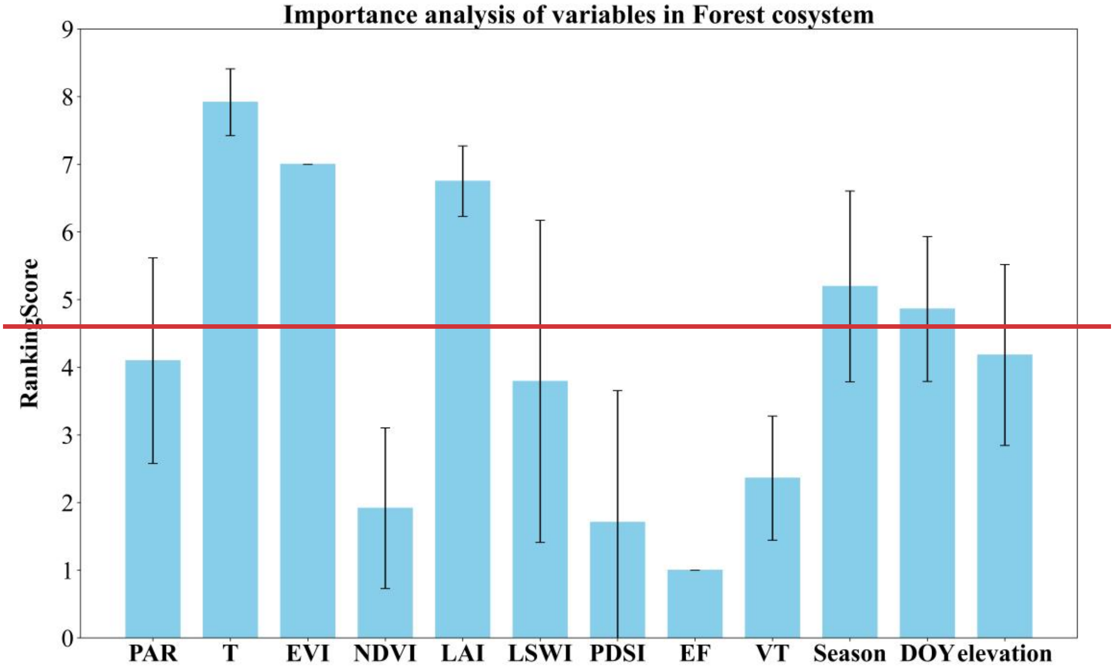
~~Fig. 4 shows that the simulated GPP closely aligns with the observed GPP values in terms of seasonal patterns at the 8-day, 16-day, and monthly scales. The simulated and observed GPP in forest ecosystems exhibit strong seasonality, with the lowest values in spring, peaking in summer, and declining through fall and winter. Forest ecosystems showed a peak of growth in the summer. In addition, the average annual GPP varied greatly among sites (Table S5). Among the forest ecosystems, tropical rainforest sites (BNF), subtropical evergreen broadleaf forests (ALF) had the highest annual GPP, followed by subtropical planted coniferous forests (QYF), deciduous broadleaf forests (MEF) and mixed coniferous and broadleaf forests (CBF, DHF), and the lowest annual average GPP was found in the cold-temperate coniferous forests (HZF). In summary, the FLAML-LUE model accurately modeled this inter-site variation in GPP and showed seasonal variations in GPP.~~



~~**Fig. 5.** The monthly bias of FLAML-LUE models among vegetation types. NF: needle-leaf forest, MF: mixed forest, EBF: evergreen broad-leaf forest, DBF: deciduous broad-leaf forest.~~

~~We examined the monthly discrepancies between observed and simulated values across various forest types in forest ecosystems. Fig. 5 shows that the forest ecosystems model underestimated GPP in spring and summer on average, and although the forest ecosystems GPP simulation was biased in all months, it generally showed a larger bias in summer. There were significant differences in bias between forest types, with the model performing better in capturing the seasonal dynamics of coniferous and deciduous broadleaf forests.~~

~~3.1.3 Analysis of the importance of variables~~



~~**Fig. 6.** Average variables importance of forest ecosystem in FALML-LUE models. LSWI: land surface water index, PDSI: Palmer Drought Severity Index, EF: evaporative fraction, EVI: enhanced vegetation index, NDVI: normalized difference vegetation index, LAI: leaf area index, T: temperature, PAR: photosynthetically active radiation, VT: vegetation type.~~

~~Fig. 6 shows the importance of each variable in the FLAML-LUE model for the forest ecosystem. The FLAML-LUE model utilizes AutoML algorithms based on different combinations of metrics to find the optimal algorithm and appropriate hyperparameters. Since different ML algorithms were selected for different temporal scales and different combinations of indicators, and different methods were used to calculate the importance of each indicator, the ranking assignment method was used to assign the importance of each indicator. Based on the average importance of 4 temporal scales and 18 combinations of indicators, it can be seen that in forest ecosystems, the importance of temperature is greater than other variables in the model. The importance of EVI and LAI is much higher than that of NDVI among the three vegetation indices, which is also consistent with the results in section 3.1.1, that is, the simulation performance of the model consisting of the combination of indicators EVI and LAI is~~

~~better than that of the combination of NDVI indicators. The importance of LSWI is higher than PDSI and EF among the water stress factors. Forest ecosystem GPP exhibits clear seasonal variation, with temperature and VI emerging as the most critical variables in the ML model for GPP estimation. These factors significantly impact the accuracy of predictions.~~

~~3.2 Overall FLAML models performances on grassland ecosystem~~

~~3.2.1 Performance Evaluation of Models~~

~~To examine the performance of each model in grassland ecosystems and at the site level, the accuracy of the 18 FLAML-LUE models was evaluated using the site data from DXG, DLG and HBG_S01 Stations as the grassland ecosystem model test set. Table 5 shows the~~ ~~R^2~~, ~~RMSE and SD of the 18 FLAML-LUE models with the grass station test set. Table S6 shows the algorithms adopted by each FLAML-LUE model under the grassland ecosystems.~~

~~**Table 5**~~

~~R^2, SD, RMSE for the grassland ecosystems model test set.~~

~~FLAML~~	~~R^2~~	~~SD~~	~~RMSE~~	~~nuRMSE~~	~~TSS~~
~~FLAML00~~	~~0.82~~	~~0.961~~	~~0.863~~	~~0.424~~	~~0.9525~~
~~FLAML01~~	~~0.82~~	~~0.987~~	~~0.857~~	~~0.421~~	~~0.9543~~
~~FLAML02~~	~~0.84~~	~~0.942~~	~~0.816~~	~~0.401~~	~~0.9558~~
~~FLAML03~~	~~0.82~~	~~0.935~~	~~0.858~~	~~0.422~~	~~0.9508~~
~~FLAML04~~	~~0.81~~	~~0.928~~	~~0.886~~	~~0.436~~	~~0.9466~~
~~FLAML05~~	~~0.83~~	~~0.909~~	~~0.832~~	~~0.409~~	~~0.9502~~
~~FLAML06~~	~~0.82~~	~~0.992~~	~~0.859~~	~~0.422~~	~~0.9544~~
~~FLAML07~~	~~0.82~~	~~1.015~~	~~0.865~~	~~0.425~~	~~0.9548~~
~~FLAML08~~	~~0.84~~	~~0.991~~	~~0.819~~	~~0.402~~	~~0.9585~~
~~FLAML10~~	~~0.81~~	~~0.976~~	~~0.890~~	~~0.437~~	~~0.9509~~
~~FLAML11~~	~~0.80~~	~~0.990~~	~~0.897~~	~~0.441~~	~~0.9512~~
~~FLAML12~~	~~0.83~~	~~0.976~~	~~0.845~~	~~0.415~~	~~0.9555~~
~~FLAML13~~	~~0.82~~	~~0.951~~	~~0.874~~	~~0.430~~	~~0.9508~~

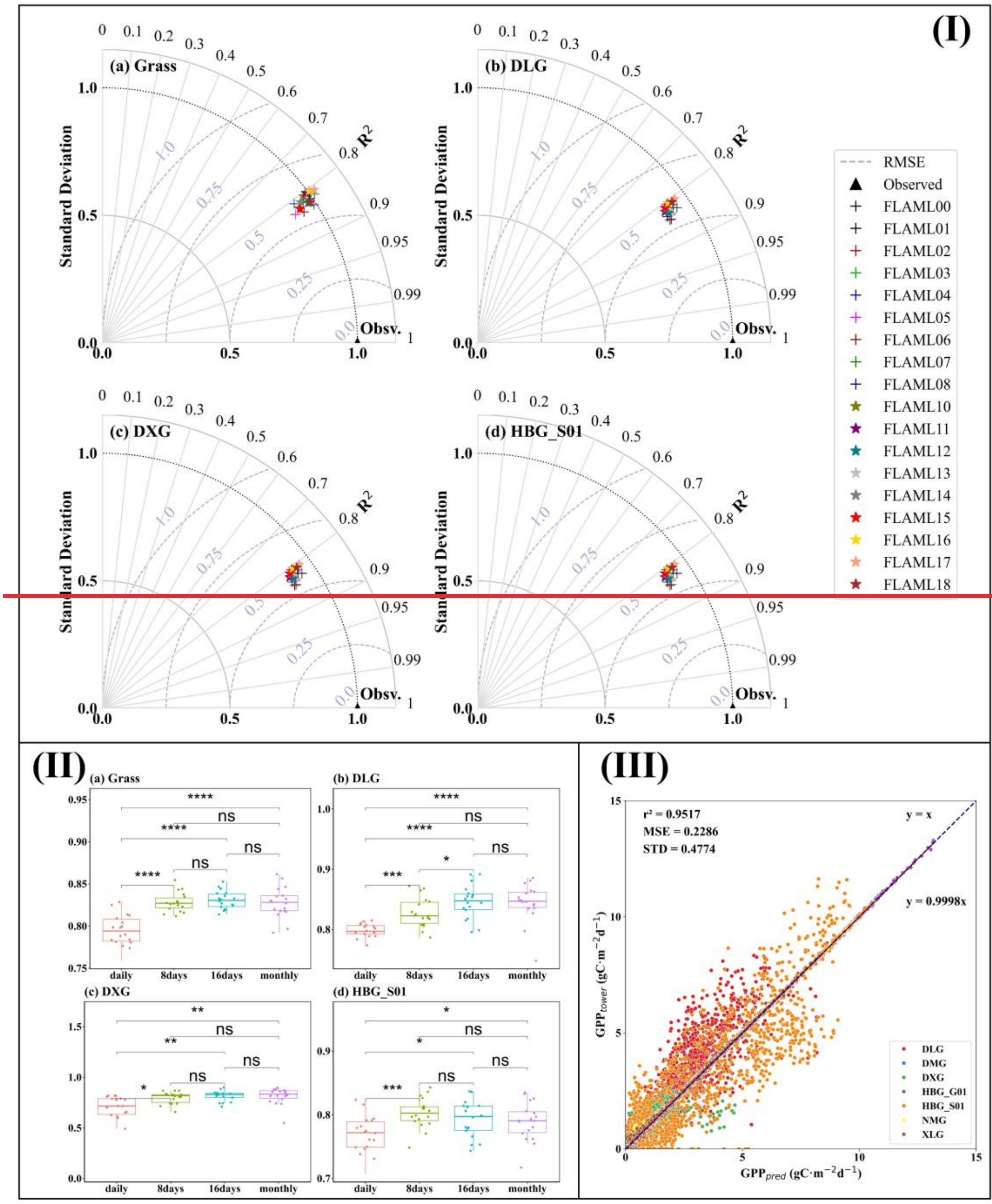
~~FLAML14~~	~~0.82~~	~~0.955~~	~~0.871~~	~~0.428~~	~~0.9517~~
~~FLAML15~~	~~0.83~~	~~0.936~~	~~0.843~~	~~0.414~~	~~0.9527~~
~~FLAML16~~	~~0.81~~	~~1.004~~	~~0.895~~	~~0.440~~	~~0.9515~~
~~FLAML17~~	~~0.81~~	~~1.024~~	~~0.885~~	~~0.435~~	~~0.9528~~
~~FLAML18~~	~~0.83~~	~~0.984~~	~~0.841~~	~~0.413~~	~~0.9563~~
~~Flux(average)~~	~~0.82~~	~~0.962~~	~~0.851~~	~~0.418~~	
~~ERA5(average)~~	~~0.81~~	~~0.977~~	~~0.871~~	~~0.428~~	
~~Forest(average)~~	~~0.82~~	~~0.970~~	~~0.861~~	~~0.423~~	

~~As shown in Table 5, the cross-validation analysis showed that the average R^2 for the four temporal scales under grassland ecosystems was 0.80-0.84. The models driven by the flux data performed slightly better than the one driven by the ERA5 data, with average R^2 of 0.82, 0.81, and RMSE of 0.851, 0.871 gC·m^{-2}d^{-1}, respectively. In grassland ecosystems, models driven by different vegetation indices had equal mean R^2 values of 0.82 and RMSE values of 0.861 gC·m^{-2}d^{-1}. The model driven with EF (R^2 = 0.83, RMSE = 0.833 gC·m^{-2}d^{-1}) performed slightly better than LSWI (R2 = 0.82, RMSE = 0.873 gC·m^{-2}d^{-1}) and PDSI (R^2 = 0.81, RMSE = 0.877 gC·m^{-2}d^{-1}).~~

~~Fig. 7 shows the Taylor diagrams of the performance of all FLAML-LUE models in grassland ecosystems, DXG, DL, and HBG_S01. The R^2, nuRMSE, and SD of different combinations of variables under grassland ecosystems were slightly different, and the TSS values ranged from 0.9466 - 0.9585, among which the best performance was the FLAML08 combination with the largest TSS of 0.9585.~~

~~Similar to forest ecosystems, the main differences in the prediction accuracy of the FLAML-LUE model for grassland ecosystems were between grass types rather than between different combinations of indicators. It is clear that the simulation accuracy of GPP for grassland ecosystems is lower than that for forest ecosystems, and there are~~

~~also significant differences between grass types. For typical grassland, the FLAML-LUE model performed best with an average R^2 of 0.83 and an RMSE of 0.779 $gC \cdot m^{-2}d^{-1}$, followed by alpine scrub with an average R^2 of 0.79 and an RMSE of 0.459 $gC \cdot m^{-2}d^{-1}$, and stations with alpine meadows the worst performance, with an average R^2 of 0.78 and an RMSE of 0.461 $gC \cdot m^{-2}d^{-1}$ (Table S7, S8, S9). It is worth noting that the model simulated the alpine meadows with the lowest RMSE for GPP, which is since the average daily GPP of alpine meadows is smaller than that of typical grassland and alpine scrub.~~



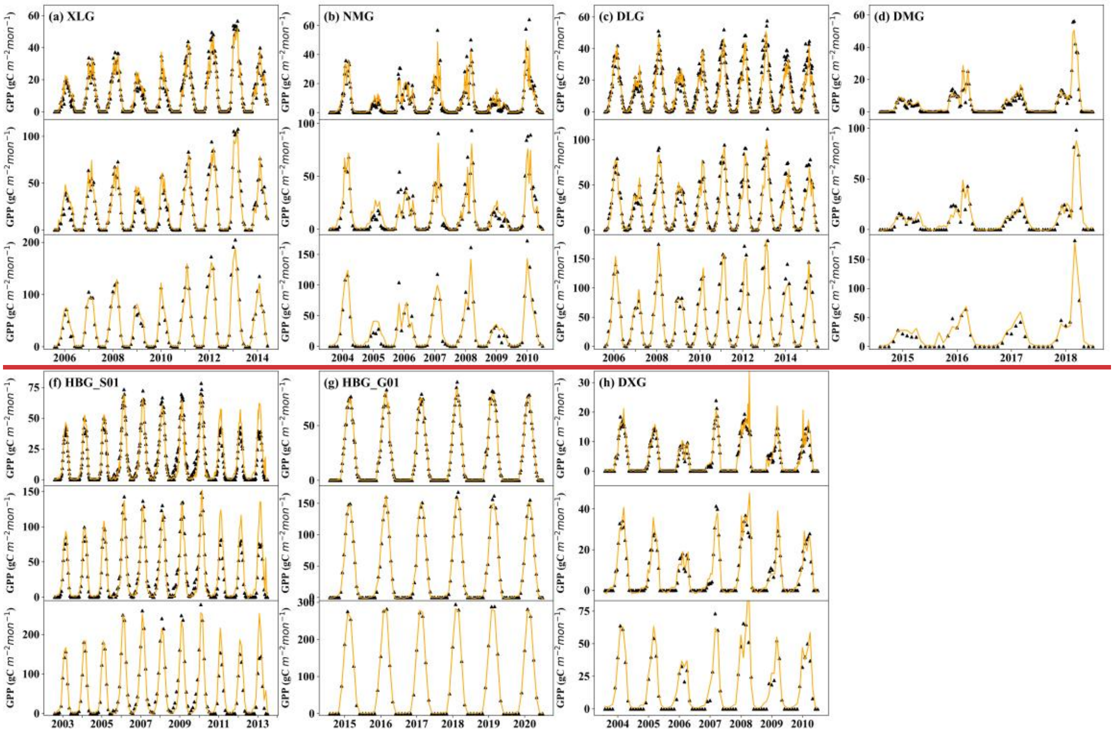
~~**Fig. 7.** (I) Normalized Taylor diagrams showing the overall performance of all FLAML-LUE models in (a) grass ecosystem, (b) DXG, (c) DLG, and (d) HBG_S01. (II) Boxplots of 18 model performances (R2) at different temporal scales in grassland ecosystems, DXG, DLG, and HBG_S01. (III) Scatterplot of observed GPP vs. simulated GPP in grassland ecosystems.~~

~~A best-fit line between tower-based GPP and predicted GPP was determined for all grass ecosystem sites as one dataset using linear regression (Fig. 7 (III)). The R^2 for all sites differed less from the results for the sites analyzed individually. As shown in Fig. 10, the slope of the fitted line was close to, but slightly less than 1, indicating that the~~

~~FLAML-LUE model overestimated the GPP of grassland ecosystems.~~

~~In grassland ecosystems, for all four temporal scales, the 18 FLAML-LUE models showed higher accuracy as temporal aggregation increased from daily to monthly. The FLAML-LUE model shows a marked improvement in validation accuracy at extended time scales, with the average R^2 rising from 0.80 to 0.83. Paired t-tests revealed that for grassland ecosystems and at the DXG and DLG stations, the FLAML-LUE model's performance (R^2) was significantly lower at the daily scale compared to other time scales ($p < 0.01$, Fig. 7 (II)). However, at station HBG_S01, model performance at the daily scale was only lower than the 8-day time scale, and not significantly different from other time scales. In addition, the RMSE of the 8-day, 16-day, and monthly GPP (FLAML-LUE) were also significantly lower by 12.10%, 13.36%, and 12.62%, respectively, compared to the daily-scale results, indicating that the uncertainty associated with these models diminishes at extended time scales.~~

~~3.2.2 Analysis of interannual GPP variability~~

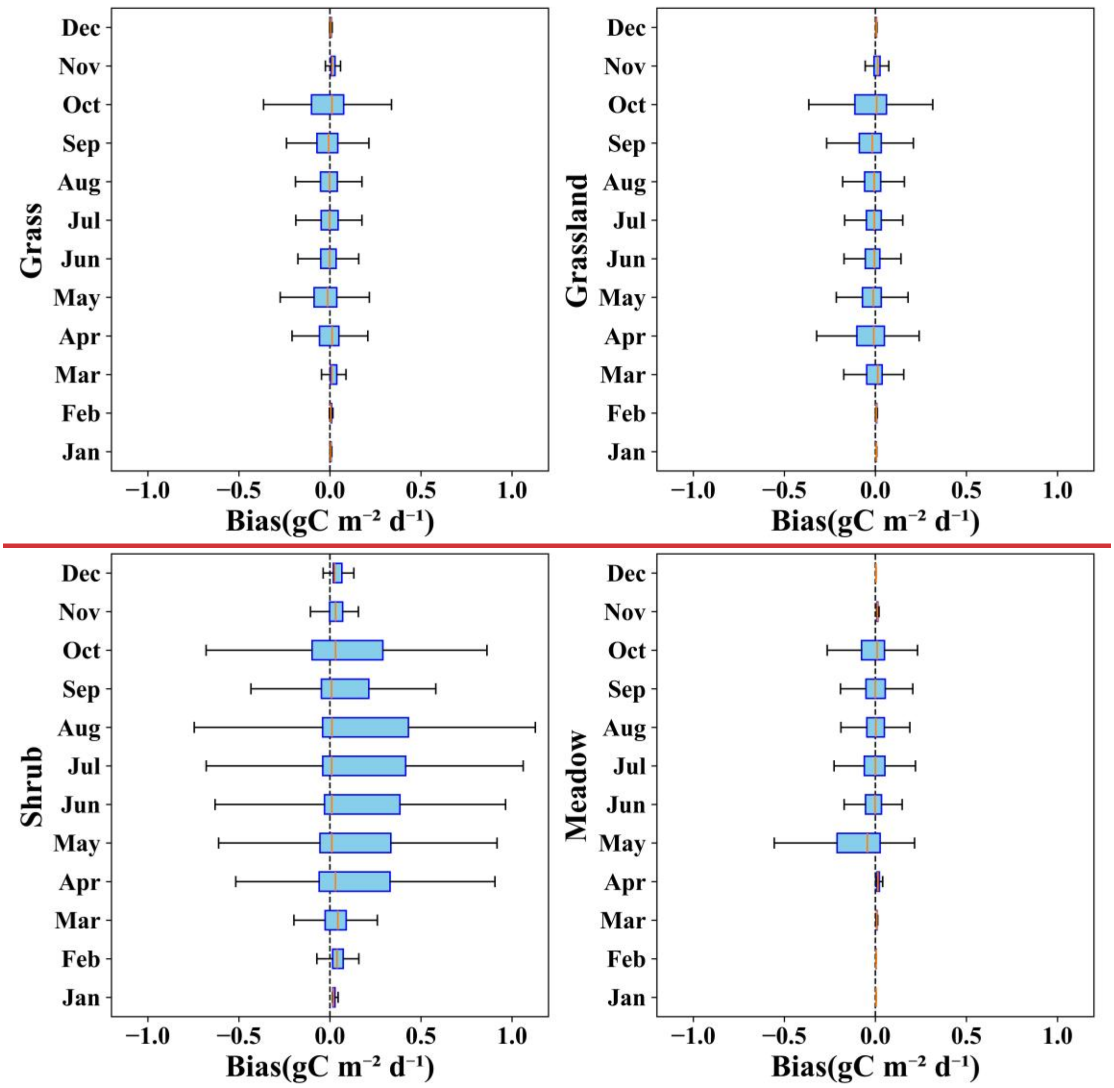


~~**Fig. 8.** Plot of simulated GPP time-series variation at DLG, DMG, DXG, HBG_G01, HBG_S01, NMG, and XLG sites.~~

~~Based on the Taylor diagram TSS skill scores, it can be seen that the grassland ecosystems have the highest GPP simulation accuracy under the combination of FLAML08 indicators, as shown in Table S6.~~

~~Fig. 8 shows that the FLAML-LUE model can simulate seasonal dynamics similar to the observed GPP, as can be seen from their long-term evolution courses at the seven grass ecosystem sites (DLG, DMG, DXG, HBG_G01, HBG_S01, NMG, XLG). Although the overall trend was simulated correctly, it is clear that the FLAML-LUE model does not capture the GPP peaks in grassland ecosystems well. For the simulation of typical grassland sites (DLG, DMG, NMG, XLG), the model performance was generally poor for NMG site, and the GPP values were poorly simulated during the peak growing seasons. In addition, it is more difficult to simulate GPP at the meadow sites (DXG and HBG_G01), especially for the summer peak simulation at DXG site,~~

~~which was too high compared to measured GPP. This is possibly due to the special geographic location and survival environment of alpine. In conclusion, the simulation of summer peaks of different grass types for GPP in grass ecosystems did not perform well.~~



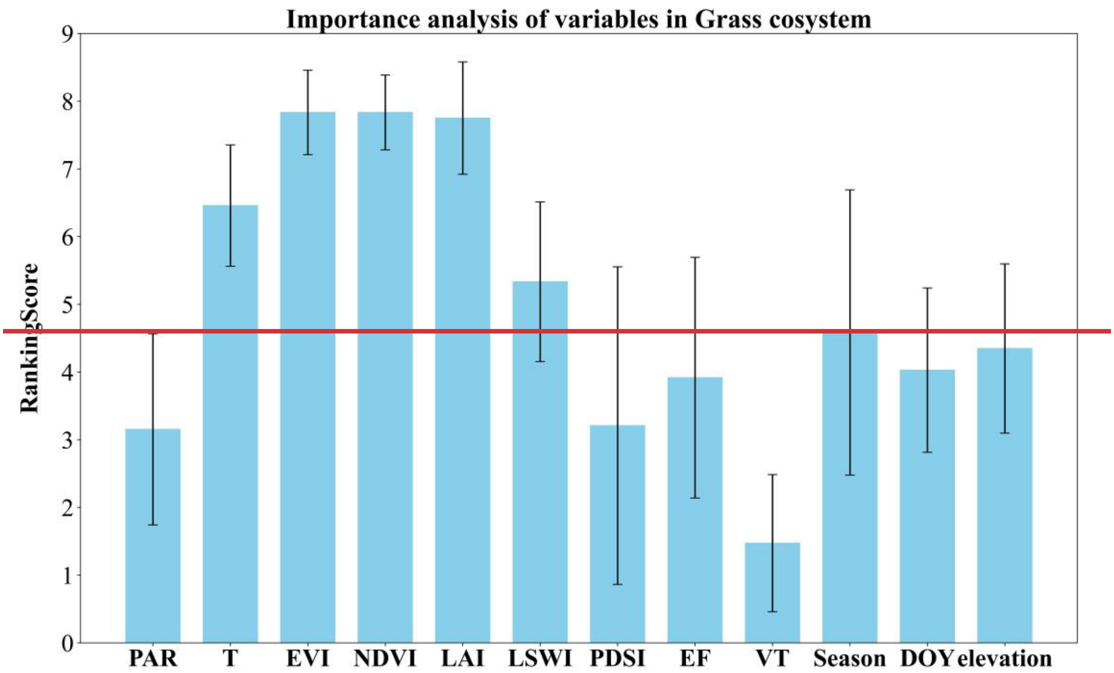
~~**Fig. 9.** The monthly bias of FLAML-LUE models among grass types. Grassland: typical grassland, Shrub: alpine shrub, Meadow: alpine meadow.~~

~~We examined the monthly discrepancies between observed and modeled values for different farm types in the grassland ecosystem. Fig. 9 shows that the simulated values of GPP from the grass ecosystem model for typical grassland and alpine scrub have biases in all months, and the biases were generally larger in summer and were all~~

~~overestimated. The gross primary productivity in spring and winter was smaller, and the corresponding deviations were smaller. Similarly, the GPP simulations for alpine meadows were underestimated and had smaller deviations, as seen in Fig. 9.~~

~~3.2.3 Analysis of the importance of variables~~

~~Fig. 10 shows the importance of the variables in the FLAML-LUE model for grassland ecosystems. It can be seen that the importance of NDVI is the highest among all the variables in the grass ecosystem model. The importance of LAI was the lowest among the three vegetation indices, while it is still higher than that of the other variables, indicating that vegetation indices are very important for modeling the GPP of grassland ecosystems. The importance score of temperature ranked just below the three vegetation indices, proving that temperature is also one of the most important indicators for estimating GPP in grassland ecosystems. In grassland ecosystems, the moisture index LSWI had a higher importance in modeling the GPP compared to PDSI and EF, and overall, the grass ecosystem showed a higher importance score for the moisture index than the forest ecosystem. Generally, forest vegetation has stronger water storage capacity and a higher ability to utilize deep soil water when compared to grasses, thus making forests more resistant to drought during meteorological droughts. Therefore, grass ecosystem simulated GPP were more sensitive to the moisture index.~~



~~**Fig. 10.** Average variables importance of grassland ecosystems in FALML-LUE models.~~

~~3.3 Overall FLAML models performances on cropland ecosystem~~

~~3.3.1 Performance Evaluation of Models~~

~~To examine the performance of each model in the cropland ecosystem and at the site level, the accuracy of the 18 FLAML-LUE models was evaluated using the site data from JZA and YCA stations as the cropland ecosystems model test set. Table 6 shows the R^2, RMSE, and SD of the 18 FLAML-LUE models with the cropland station test set. The algorithms adopted by each FLAML-LUE model under the cropland ecosystems are shown in Table S11.~~

~~As shown in Table 6, the cross-validation analysis showed that the average R^2 for the four temporal scales under cropland ecosystems was 0.82-0.89. The models driven by the flux data performed slightly better than the one driven by the ERA5 data, with their average R^2 of 0.88, 0.85, and RMSE of 1.908, 2.108 $gC{\cdot}m^{-2}d^{-1}$, respectively. However, the models driven using EVI (R^2 = 0.87, RMSE = 1.955 $gC{\cdot}m^{-2}d^{-1}$)~~

~~performed slightly better than NDVI (R^2 = 0.85, RMSE = 2.069 gC·m^{-2}d^{-1}) and LAI (R^2 = 0.86, RMSE = 1.999 gC·m^{-2}d^{-1}). The model driven with PDSI (R^2 = 0.87, RMSE = 1.952 gC·m^{-2}d^{-1}) performed slightly better than EF (R^2 = 0.87, RMSE =1.991 gC·m^{-2}d^{-1}) and LSWI (R^2 = 0.85, RMSE = 2.080 gC·m^{-2}d^{-1}).~~

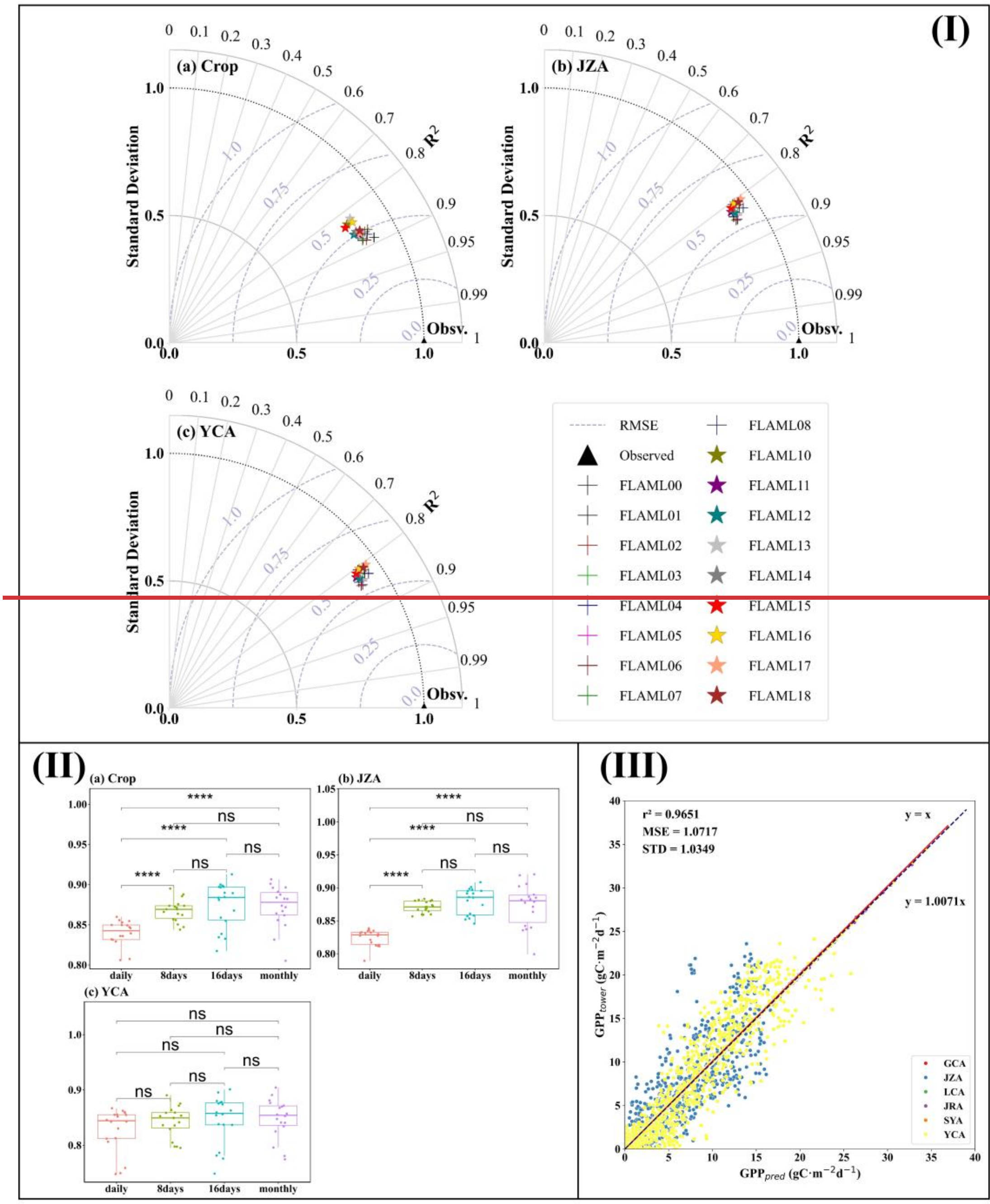
~~Fig. 11 shows the Taylor diagrams of the performance of all FLAML-LUE models in cropland ecosystems, JZA station, and YCA station. The R^2, nuRMSE, and SD of different combinations of variables under cropland ecosystems were slightly different, and the TSS values ranged from 0.9253 – 0.9622, among which the best performance was the FLAML00 combination with the largest TSS of 0.9622.~~

~~**Table 6**~~

~~R^2, SD, RMSE for the cropland ecosystems model test set.~~

FLAML	R^2	SD	RMSE	nuRMSE	TSS
FLAML00	0.89	0.904	1.812	0.626	0.9622
FLAML01	0.88	0.859	1.858	0.611	0.9490
FLAML02	0.89	0.873	1.832	0.594	0.9542
FLAML03	0.87	0.883	1.966	0.647	0.9524
FLAML04	0.87	0.851	1.963	0.640	0.9425
FLAML05	0.87	0.872	1.967	0.592	0.9475
FLAML06	0.87	0.897	1.981	0.616	0.9532
FLAML07	0.88	0.864	1.882	0.672	0.9483
FLAML08	0.88	0.886	1.912	0.596	0.9535
FLAML10	0.83	0.838	2.230	0.621	0.9282
FLAML11	0.87	0.858	1.983	0.598	0.9430
FLAML12	0.86	0.840	2.015	0.635	0.9365
FLAML13	0.82	0.861	2.281	0.633	0.9319
FLAML14	0.86	0.853	2.042	0.644	0.9384
FLAML15	0.84	0.825	2.195	0.603	0.9253
FLAML16	0.83	0.861	2.212	0.668	0.9348
FLAML17	0.87	0.864	1.985	0.631	0.9418
FLAML18	0.86	0.868	2.025	0.629	0.9454
Flux(average)	0.88	0.877	1.908	0.622	

~~ERA5(average)~~	~~0.85~~	~~0.852~~	~~2.108~~	~~0.629~~
~~Forest(average)~~	~~0.86~~	~~0.864~~	~~2.008~~	~~0.625~~



~~**Fig. 11.** (I) Normalized Taylor diagrams showing the overall performance of all FLAML-LUE models in (a) cropland ecosystem, (b) JZA, and (c) YCA. (II) Boxplots of 18 model performances (R2) at different temporal scales in crop ecosystem, JZA, and YCA. (III) Scatterplot of observed GPP vs. simulated GPP in crop ecosystem.~~

~~Unlike forest and grassland ecosystems, the performance of the FLAML-LUE model did not differ significantly among different farm types in cropland ecosystems.~~

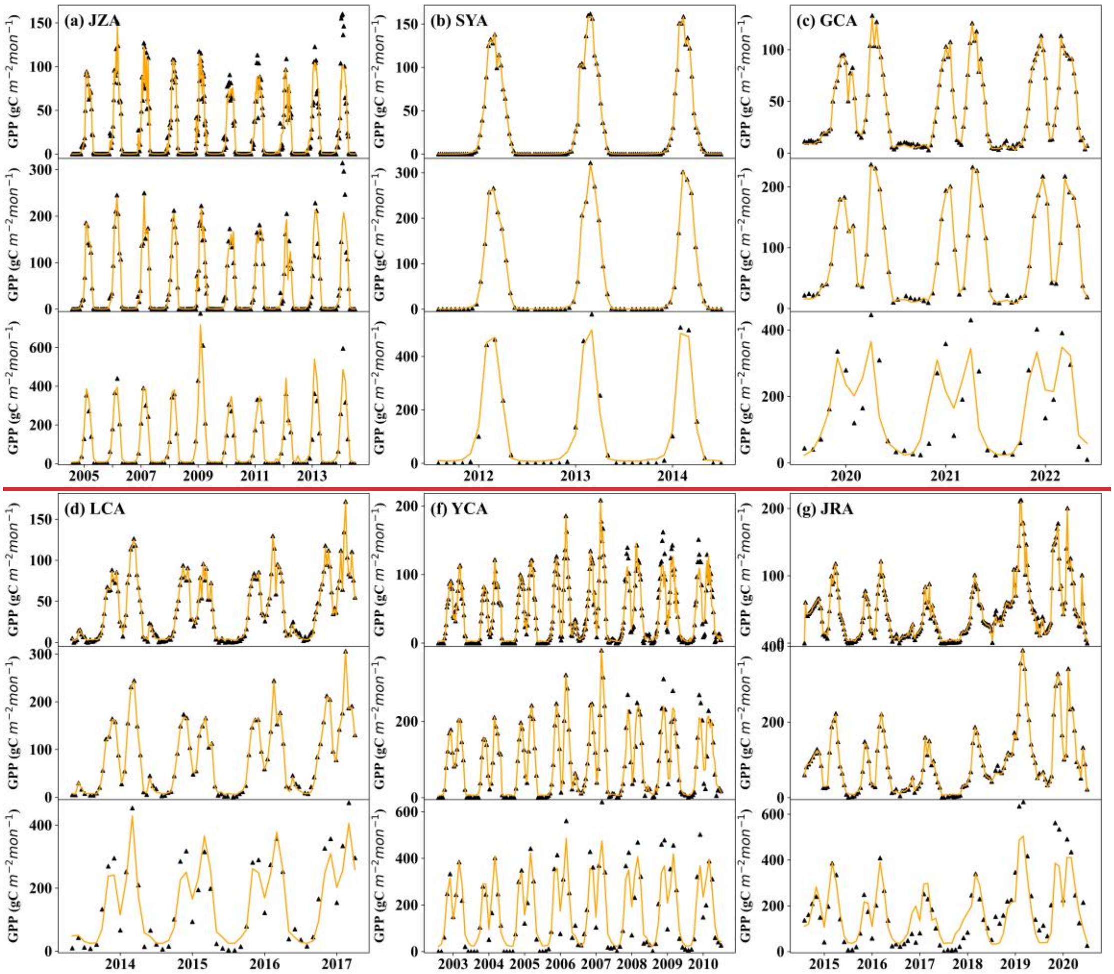
~~The average R^2 was 0.86 and the average RMSE was 1.724 gC·m^{-2}d^{-1} for the single cropping farmland station (JZA), while the average R^2 was 0.84 and the average RMSE was 2.400 gC·m^{-2}d^{-1} for the double cropping farmland (YCA). The simulation of the single cropping farmland was slightly better than the double cropping farmland (Table S12, S13).~~

~~A best-fit line between GPP_{tower} and GPP_{pred} was determined for all cropland ecosystem sites as one dataset using linear regression (Fig. 11 (III)). The R^2 for all sites differed less from the results for the sites analyzed individually. As shown in Fig. 11 (III), the slope of the fitted line was close to but slightly less than 1, indicating that the FLAML-LUE model overestimated the GPP of cropland ecosystems.~~

~~In the cropland ecosystems, the average R^2 increased from 0.84 at the daily scale to 0.87 at the 16-day scale as can be seen, and the 18 FLAML-LUE models showed higher accuracy as the temporal aggregation increased from the daily to the monthly. The model generally showed significantly lower performance (R^2) at the daily scale than at other time scales ($p < 0.001$, Fig. 11(II)(a)), while there was no remarkable difference in the model performance at all four time scales for the YCA ($p > 0.05$, Fig. 11(II) (c)). The difference in model performance between the 16-day and monthly scales was not significant at all stations ($p > 0.05$, Fig. 11(II)). In addition, RMSE of the GPP (FLAML-LUE) was also significantly reduced by 14.70%, 18.61%, and 19.79% for the 8-day, 16-day, and monthly GPP, respectively, when compared to the daily-scale results, suggesting that the uncertainty in these models becomes smaller at the longer temporal scale. At JZA and YCA, the slopes of the linear regression relationship~~

~~between the modeled GPP and the observed GPP converge to 1 as the time scale improves.~~

~~3.3.2 Analysis of interannual GPP variability~~

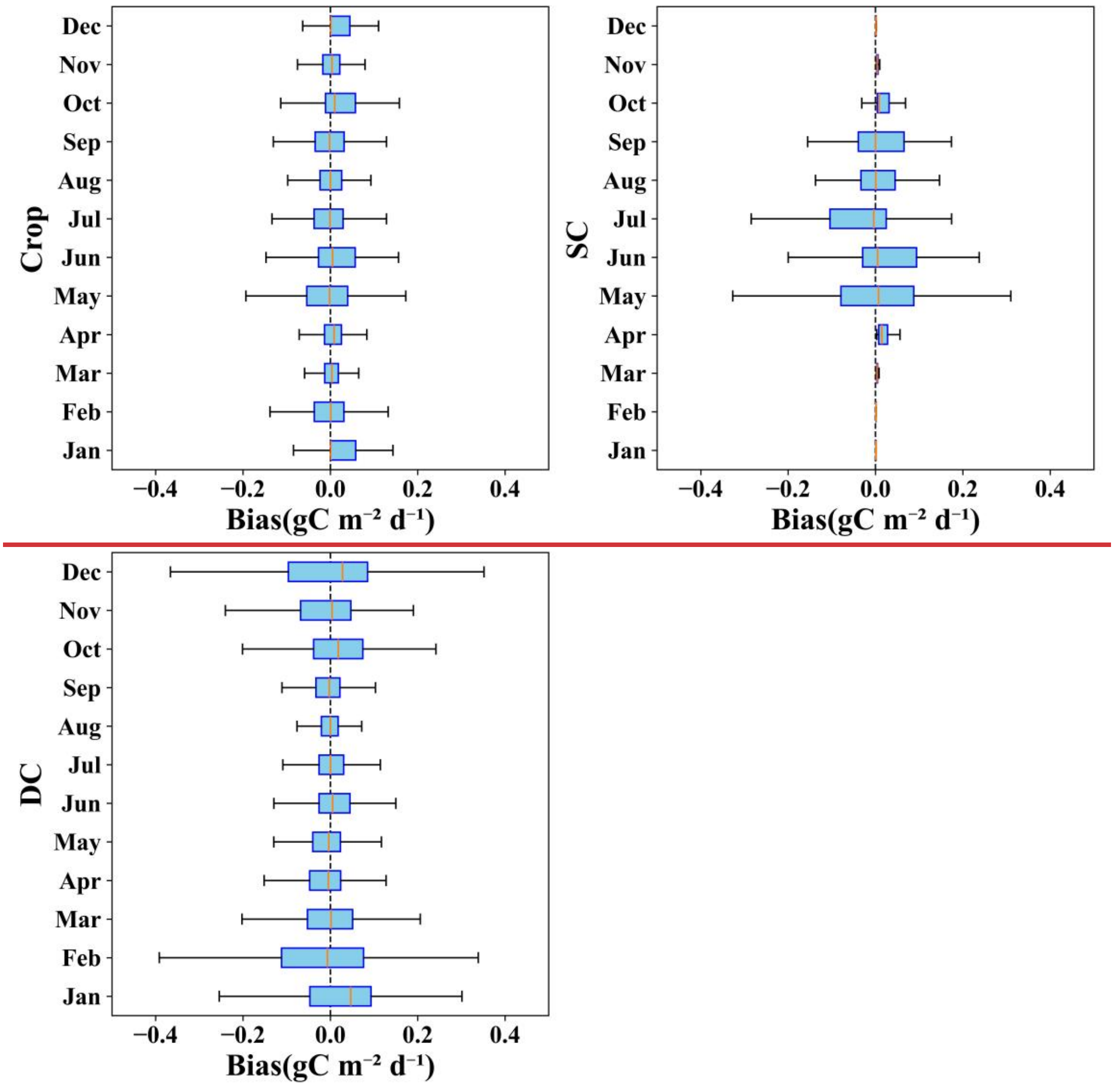


~~**Fig. 12.** Plot of simulated GPP time-series variation at GCA, JRA, JZA, LCA, SYA, and YCA sites.~~

~~Fig. 12 shows that simulated GPP aligns closely with the observed GPP values, showing comparable seasonal patterns at the 8-day, 16-day, and monthly scales. In farmland ecosystems, simulated GPP values from different farm types show different seasonal dynamics. Farmland with spring maize (JZA, SYA), a single-crop system, shows a single GPP peak during the harvest season. In comparison, double-cropping systems, with cycles of winter wheat and summer corn, display GPP peaks in both May and August. In addition, the average annual GPP of farmlands with different crop~~

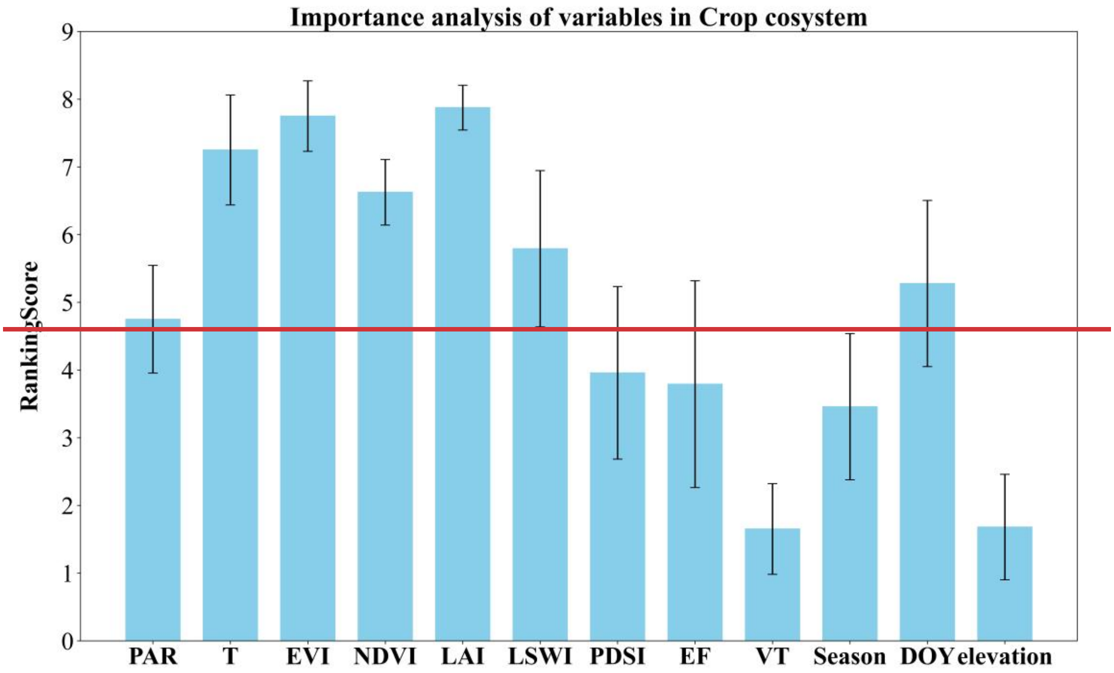
~~rotation schemes varied greatly (Table S14). The annual mean GPP of double-cropping farmland was higher than that of single-cropping farmland. In conclusion, the FLAML-LUE model accurately modeled the differences in GPP among farmland types and showed seasonal variations in GPP among farmland types.~~

~~We examined the monthly discrepancies between observed and modeled values for different farm types in the agroecosystem. Fig. 13 shows that the agroecosystem model averagely overestimated GPP values in spring and fall (positive bias), while slightly underestimated it in summer. Although the agroecosystem GPP simulations were biased in all months, the biases were generally larger in spring and fall. There were significant differences in bias between farmland types. The model over double cropping farmland showed small biases in simulated GPP for all months of the year, while it overestimated GPP in the spring and fall, and underestimated GPP in the summer over the single cropping farmland. This suggests that the model performance for the single cropping farmland still need to be improved.~~



~~**Fig. 13.** The monthly bias of FLAML-LUE models among cropland types. SC: single cropping, DC: double cropping.~~

~~3.3.3 Analysis of the importance of variables~~



~~**Fig. 14.** Average variables importance of farm ecosystem in FALML-LUE models.~~

~~Fig. 14 shows the importance of the variables in the cropland ecosystem FLAML-LUE model. It can be seen that the importance of LAI is the highest among all the variables in the farm ecosystem model. The importance of NDVI was the lowest among the three vegetation indices, while it is still higher than that of the other variables, indicating that vegetation indices are very important for modeling the GPP of cropland ecosystems. The importance score of temperature was second only to LAI and EVI, and similar to forest and grassland ecosystems, temperature is also one of the important indexes for modeling GPP in farm ecosystem. In addition, the moisture stress factor was also important, and unlike forest and grass ecosystems, the most important moisture factor for constructing the GPP simulation model in cropland ecosystems was LSWI, followed by PDSI, and EF was the lowest.~~

~~4.1. Discussion~~

~~Model performance is highly influenced by the algorithms used, the underlying~~

~~processes, and how GPP responds to varying environmental conditions (Chang et al., 2023). A detailed comparison of the FLAML-LUE models across different ecosystems showed that performance varied depending on the input variables, vegetation types, and time scales (Chang et al., 2023; Harris et al., 2021).~~

~~4.11.1 Performance comparison of FLAML-LUE models for different ecosystems~~

In this study, FLAML-LUE models were constructed for ~~different ecosystems,~~ different combinations of variables and different time scales based on AutoML algorithms. On the whole, the modeled GPP values agree well with the GPP estimated based on the EC tower, and the FLAML-LUE models performed better in capturing the magnitude and seasonal dynamics of the GPP, which indicated that it was feasible to estimate the GPP using AutoML algorithms. Further, all three ecosystems showed good model performance driven by observational data. Comparisons across various ecosystems indicate that the model exhibited superior performance over forest ecosystems compared to grassland and agricultural ecosystems, as evidenced by the average R^2 values.

~~Additionally, the models constructed for each ecosystem showed different performances under different~~Although model performance differences across indicator combinations were minimal, EVI-driven FLAML-LUE models slightly outperformed those driven by NDVI. This highlights the key role of EVI in GPP estimation, as it offers more comprehensive atmospheric correction and is less susceptible to saturation from green reflectance compared to NDVI. Additionally, model performance varied significantly across sites.

Based on the evaluation metrics, the optimal model selected was FLAML00 (EVI + LSWI). Under this combination of indicators, ~~while the differences were not significant, and the main differences in prediction accuracy~~ the FLAML-LUE model demonstrated the best performance in mixed forests at CBF, deciduous broadleaf forests at MEF, and alpine meadows at HBG_G01, with R^2 values of 0.92, 0.92, and 0.93, respectively. The next best performances were ~~ascribed to site differences rather than model differences. FLAML-LUE had the best simulation performance for mixed forests in CBF, and planted~~ observed in coniferous forests ~~in~~at QYF~~, with R^2 of 0.93 and 0.90, respectively; followed by~~ and HZF, single~~-~~-cropping farmland ~~in the Jinzhou site~~at JZA and SYA stations, double~~-~~-cropping farmland ~~in the Yucheng site~~at YCA, and typical ~~grassland~~grasslands at DLG~~. Over the alpine meadow and~~ and DMG sites. In contrast, the model performed poorly in alpine ~~ecosystem, the model performance was poorer~~shrub and alpine ecosystems, with an ~~R^2~~R^2 of 0.~~79;~~54, and ~~even worse~~the worst performance was observed at the ~~MEF~~BNF site, with an average ~~R^2 of 0.63.~~ R^2 of only 0.28. Mixed forests exhibit distinct seasonal variations that satellite imagery can effectively capture, while evergreen broadleaf forests (ALF and BNF) show minimal seasonal changes in vegetation cover or greenness, making accurate predictions challenging. Alpine shrublands have more complex vegetation structures and less distinct seasonal variations in vegetation cover, which makes it harder for the model to capture the dynamics accurately. In contrast, alpine meadows exhibit more pronounced seasonal variations in vegetation cover, which makes the model more effective in capturing GPP dynamics. For non-forest ecosystems, the highest R^2 values were observed in agricultural fields and typical grasslands, followed by alpine meadows and alpine shrublands.

Mixed forests display clear seasonal variations that satellite imagery can effectively capture. However, evergreen broadleaf forests (ALF) have slight seasonal variations in vegetation cover or greenness, making it difficult for the model to predict. For non-forest ecosystems, the highest R^2 was found in agricultural fields and typical grasslands, followed by alpine meadows and alpine scrub. In addition, the differences in model performance were also reflected in different temporal scales. In general, the model simulation performance at the 16-day and monthly scales was better than that at the daily scale, and the performances of different temporal scales for forest, grassland, and cropland ecosystems were consistent with previous studies.

~~Discrepancies in the comparison between observed and simulated values varied across ecosystems, with models for grassland and forest ecosystems generally underestimating GPP (exhibiting a negative bias) in spring and summer, while displaying satisfactory performance in other seasons. The GPP during spring and winter remains relatively low, and hence correspondingly smaller deviations of modeling values. Overall, the FLAML-LUE model performed well in capturing interannual variability in GPP, while it encounters challenges in accurately capturing the dynamic fluctuations of GPP throughout the growing season.~~

This study did not distinguish between rainfed and irrigated agricultural systems, considering only the crop rotation types. Specifically, JZA and SYA represent rainfed systems, whereas GCA, LCA, and YCA are irrigated. Future research could incorporate this distinction to improve the accuracy of carbon flux estimates in cropland ecosystems. This distinction is important for interpreting model results under water-limited conditions.

In addition, our results indicate that forest and agricultural fields have greater carbon sequestration capacity and higher annual fluxes than grasslands (Table ~~S5~~S9, S10, ~~S14~~S11), aligning with previous research outcomes (Y. Wang et al., 2021; Zhang et al., 2007). However, due to the annual harvest of crops, approximately 76% of the on-farm biomass is removed, resulting in limited long-term carbon storage capacity (Zhang et al., 2007). With the exception of tropical rainforests (i.e., BNF), the annual carbon production of planted forests (i.e., QYF) is higher than that of natural forests (i.e., CBF, DHF), which implies that planted forests possess significant potential for carbon assimilation, functioning as robust carbon sinks.

~~4.2 Impact of variables on GPP estimation~~

~~The estimated GPP in this research closely matched the GPP measured by the EC tower. However, the important characterizing factors affecting the models varied across different ecosystems. For forest ecosystem, temperature was the most primary variable for model construction, while the vegetation index was the most important factor for characterizing grass ecosystem and agroecosystem GPP. Our study is consistent with previous studies, indicating that, in addition to temperature data, vegetation index are the crucial drivers for accurately predicting GPP. High variability in greenness existed in grassland and scrub over the phenological cycle, as well as in agricultural land under~~

~~anthropogenic management patterns, while models were less effective in predicting evergreen broadleaf forests, with lower variability in greenness. A common problem is the high uncertainty in predicting evergreen forest GPP with many satellite-driven GPP models. This study found that the FLAML-LUE model using EVI slightly outperformed the one using NDVI, highlighting EVI's superior role in GPP estimation. EVI offers better atmospheric correction and is less affected by green radiation saturation compared to NDVI. Recent research indicates that satellite observations of solar-induced chlorophyll fluorescence (SIF) provide a more accurate picture of the dynamics of plant photosynthesis. It is a more effective indicator for modeling subtropical evergreen vegetation. Future studies should consider incorporating SIF into models to assess its potential for improving performance in evergreen forests.~~

~~Compared to temperature and radiation, moisture plays a more crucial role in regulating GPP. Recent research suggests that water stress is the primary source of uncertainty in GPP estimations (Zhang and Ye, 2022). At the same site, the FLAML-LUE model's performance driven by the three moisture indices was highly consistent across the six sites (QYZ, CBS, DLG, HBG_S01, JZA, YCA). However, for the DXG and ALF stations, the performance of the model varied with the moisture index. When PDSI was used for DXG and ALF, the R^2 values of these models were low at 0.75 and 0.60, respectively. Our results showed low importance for all moisture indices at all sites. However, moisture indices were more important in non-forest than in forest, suggesting that forests are less sensitive to water stress. This finding is consistent with the results of previous studies (Zhang et al., 2015; Sims et al., 2014; Xie et al., 2014),~~

~~which may be due to that forest vegetation has strong water storage capacity and the ability to utilize the deep soil water (Bi et al., 2015). Water variables were more crucial for grasslands compared to other ecosystems, indicating that grasslands with shallow root vegetation are less drought tolerant. In this context, future grassland management should prioritize scientific conservation planning and improved water management strategies.~~

4.2 Model Performance Variations Under Extreme Conditions

In the context of global warming and the increasing frequency of extreme climate events, the adaptability and stability of GPP estimation models in extreme environments have become crucial. This study systematically evaluated the performance of the FLAML-LUE model under high-temperature, high-VPD, and drought scenarios by grouping the validation set. The results showed a general decline in the model's accuracy across all three extreme climate conditions, with varying performance depending on the scenario, highlighting the complexity of vegetation carbon absorption responses to climate stress.

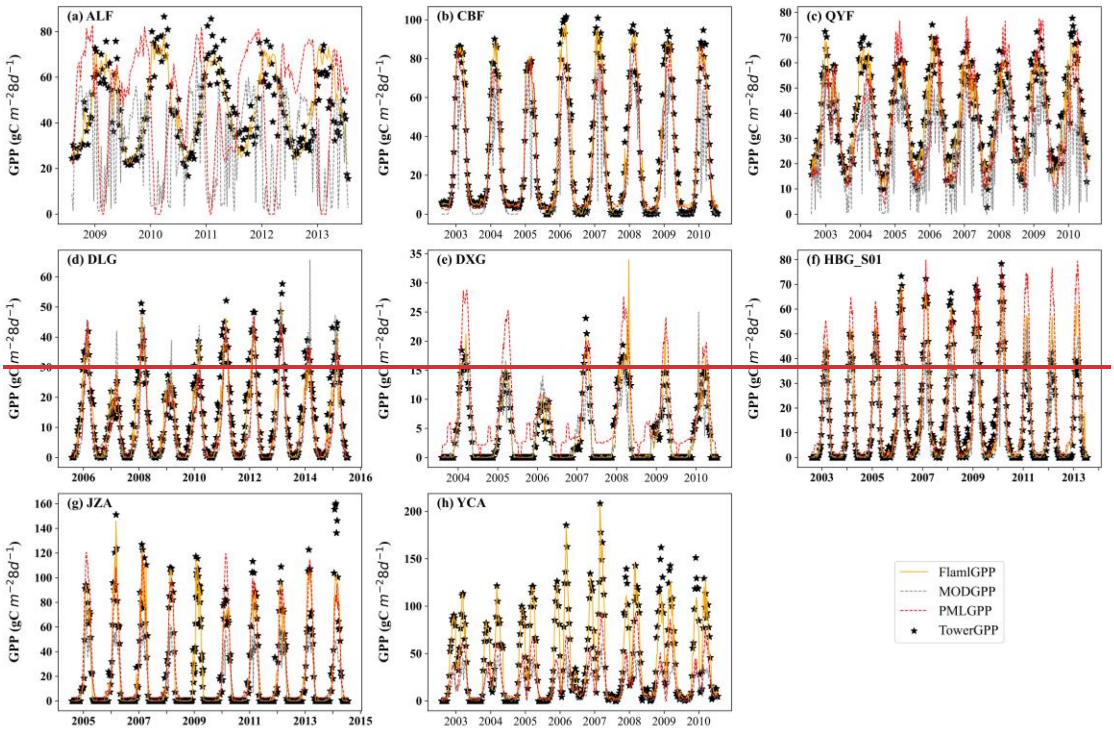
In high-temperature conditions, the model generally underestimated GPP. This could be due to the suppression of photosynthesis caused by high temperatures. High temperatures increase transpiration stress, causing stomatal closure to reduce water loss, which limits CO_2 input and lowers photosynthetic rates (Qu et al., 2020; Reichstein et al., 2013). Additionally, high temperatures can cause leaf damage and senescence, reducing LAI and overall photosynthetic potential (A. Chen et al., 2021; Y. Chen et al., 2021). Although the FLAML-LUE model accounts for fPAR and water stress factors, it may not fully capture rapid responses such as leaf damage or sudden declines in LAI, which likely contribute to the reduced accuracy under high-temperature conditions. Moreover, the model does not explicitly account for the lag effect of leaf senescence, which may further worsen estimation bias (Frank et al., 2015).

Under high VPD conditions, the model showed significant uncertainty, with some areas overestimating GPP and others underestimating it. This inconsistency likely arises from the diverse water stress mechanisms induced by high VPD. Guo et al. (2015) noted that high VPD does not always reflect the true level of water stress in plants, leading to the potential overestimation of GPP. Conversely, in extreme VPD scenarios, where stomata close to reduce carbon absorption, the model may underestimate GPP if it fails to recognize this regulatory behavior (Li et al., 2016). Additionally, the FLAML-LUE model does not explicitly consider leaf energy load or light inhibition, which may contribute to the model's higher errors under high VPD conditions (Rigden et al., 2020).

Although the model's performance decreased at some sites under drought conditions, its overall accuracy improved under these scenarios. This improvement may be due to the stronger limiting effect of drought on vegetation growth, allowing the model to more accurately capture the suppressive impact of water stress on photosynthesis. In drought conditions, water scarcity limits carbon absorption, leading to a substantial reduction in GPP (McDowell et al., 2008). As a result, the model's estimates are more likely to align with the actual limitation of carbon absorption. Thus, under drought conditions, the model may underestimate GPP, which can be more accurate, while in wetter environments, where water stress is less pronounced, the model may overestimate GPP, reducing its accuracy. Additionally, under drought, the model is likely better at capturing the direct effects of water shortage on plant physiology, reducing interference from other environmental variables and improving prediction accuracy (Zhou et al., 2019).

Although the FLAML-LUE model demonstrates strong predictive capabilities under normal climate conditions, there is still room for improvement under extreme scenarios. One potential limitation is the insufficient representation of rapid plant response mechanisms (e.g., leaf damage and sudden declines in LAI) in the current input features (Frank et al., 2015; Reichstein et al., 2013). Future research could incorporate high-temporal-resolution vegetation indices, such as solar-induced chlorophyll fluorescence (SIF), to better capture dynamic changes in plant metabolic activity and stress responses under extreme conditions (Yi et al., 2024; Pagán et al., 2019). Including lag variables or cumulative stress indices could also enhance the model's ability to handle delayed physiological responses after stress events (Frank et al., 2015). Furthermore, future studies should expand the scope to include a broader range of climate events that affect GPP, such as floods and low temperatures, in addition to high temperature, high VPD, and drought (Wang et al., 2023). Vegetation in different regions responds differently to these events, with low temperatures and frost being especially important for high-latitude ecosystems.

4.3 Comparison with other ~~studies~~products



~~Fig.~~

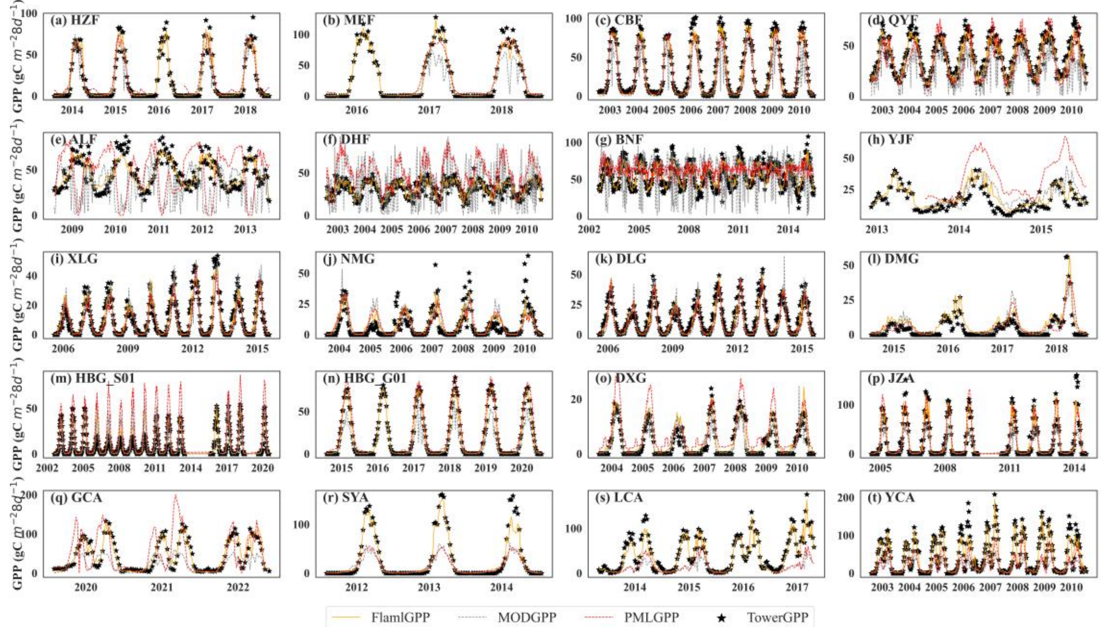


Figure 15. Comparing 8-day GPP from FLAML-LUE, PML, MOD17 models, and EC observations.

This study attempted to predict the GPP of different sites using the FLAML model based on the LUE model variables. The results showed that the AutoML algorithm is a promising GPP estimation method, which explains on average ~~63%-93~~75%-98% of the

GPP variation.

Compared to two GPP products (MODIS GPP, and PML GPP), the GPP from this study showed the highest precision (**Table 7**) and better consistency with flux tower-based GPP under different ecosystems. Overall, the FLAML-LUE model used in this study had the best simulation performance. These findings highlight the potential of the FLAML algorithm for accurately estimating GPP. The FLAML-LUE model is a data-driven ML approach that builds relationships based on dependent and explanatory variables. This enables it to effectively simulate the complex nonlinear interactions across diverse ecosystems (Tramontana et al., 2016). This advantage is even more prominent at the global scale, considering that more flux tower data are available for model construction.

Table 7
R^2 of 8-day GPP simulated by FLAML-LUE, PML, and MOD17 at different ecosystems validation sites.

Ecosystem	Station	~~FLAML_~~R2			~~MOD_R²~~ ~~PML_R²~~TSS		
		FLAML	MOD	PML	FLAML	MOD	PML
~~Forest~~ALL	~~ALF~~ALL	0.~~79~~93	0.~~24~~71	0.~~33~~78	0.9657	0.2677	0.5675
Forest	~~CBF~~HZF	0.~~98~~95	**0.~~78~~88**	0.~~93~~91	0.9843	**0.9672**	0.9569
	~~QYF~~MEF	**0.~~96~~98**	0.~~54~~78	**0.~~74~~95**	0.9868	0.7664	**0.9571**
~~Grass~~	~~DLG~~CBF	0.~~93~~98	0.~~76~~78	0.~~77~~93	**0.9903**	0.8860	0.9567
	~~DXG~~QYF	0.~~89~~95	0.54	0.74	0.~~829~~833	0.8634	0.9231
	ALF	0.87	0.24	0.34	0.9054	0.2455	0.1812
	DHF	0.83	0.27	0.45	0.9527	0.3030	0.5851
	BNF	0.81	0.05	0.02	0.9025	0.3370	0.3337
	YJF	0.75	0.31	0.42	0.9334	0.7759	0.5820
	XLG	0.92	0.76	0.79	0.9651	0.9343	0.9008
Grass	NMG	0.67	0.48	0.41	0.8288	0.8340	0.7436
	DLG	0.92	0.76	0.77	0.9787	**0.9349**	0.9320

	DMG	0.82	0.68	0.57	0.9537	0.9080	0.8611
	HBG_S01	0.~~92~~89	0.78	0.81	0.~~83~~9718	0.9284	0.7175
	HBG_G01	0.99	0.91	0.97	**0.9947**	0.7546	**0.9911**
	DXG	0.90	0.75	0.82	0.9737	0.9134	0.9105
Crop	JZA	0.~~94~~95	0.84	0.85	**0.9786**	**0.6009**	**0.9582**
	GCA	0.89	0.33	0.19	0.9708	0.4889	0.6748
	SYA	**0.96**	**0.92**	**0.92**	0.9666	0.3708	0.3948
	LCA	0.94	0.57	0.48	0.9731	0.2433	0.3959
	YCA	0.~~96~~93	0.71	0.78	0.9657	0.2677	0.5675

Note: Bold numbers indicate the highest values, while underlined numbers represent the lowest values.

However, further work is needed to evaluate the FLAML-LUE model's suitability and accuracy, considering its limitations. In particular, it tends to underestimate high GPP and overestimate low GPP. In addition, the model performance in GPP estimation is highly dependent on ecosystem type. Our findings indicated that mixed forests, deciduous broadleaf forests, and agricultural lands had higher prediction accuracies. While grass sites such as alpine scrub and alpine meadows were predicted with large uncertainties, consistent with results from other studies (Y. Wang et al., 2021; Yuan et al., 2014). This is still a big challenge in accurately estimating GPP.

In general, satellite imagery accurately captures the seasonal leaf phenology of DBF and MF canopies (e.g., spring leaf unfolding and fall senescence). Additionally, the key environmental factors influencing vegetation production during different phenological phases are well-defined (Yuan et al., 2014), making them well-suited for FLAML-LUE modeling. In contrast, the ambiguous seasonal leaf area changes in EBF and the low variability of GPP in NMG ecosystems result in poorer model performance, and empirical methods struggle to estimate GPP variability in these areas (Tramontana et al., 2016).

Model performance is heavily influenced by the quality of the driver data and the typicality of the flux towers. In this study, meteorological indices are obtained directly from spatially explicit reanalysis products. Remotely sensed variables (e.g., NDVI and EVI~~, LAI~~, LSWI) serve as proxies for vegetation growth and seasonal changes and are crucial for scaling simulations from site to regional levels. These gridded indices are directly derived from satellite reflectance bands. Large-area EFs can be obtained using LE and Hs calculations from ERA5 reanalysis data or can be derived using NDVI temperature triangulation (Venturini et al., 2004). ~~PDSI~~LAI, VPD, Pre, and RH can be obtained from ERA5 reanalysis data. Thus, the model can be extended from the site scale to the regional and even global scale. Building on this foundation, we will develop a long-term gridded GPP dataset for China using the FLAML-LUE framework to analyze its spatiotemporal variations over multiple years. This dataset will allow us to investigate long-term GPP trends across different climate zones and vegetation types, as well as their responses to key environmental drivers. By comparing GPP estimates across regions and years, we will also assess model uncertainties and identify potential areas for improvement.

5. Conclusion

In this study, the FLAML-LUE model was ~~applied to estimate GPP at four different time scales~~ developed based on data from 20 flux observation sites across ~~20 sites in~~ China~~. Our findings indicate~~, integrating the FLAML algorithm with key variables from the LUE model. The results demonstrate that the FLAML-LUE model ~~excels at predicting GPP, capturing~~performs excellently in GPP prediction, accurately simulating both its temporal variations and magnitude~~. It performs ~~, particularly ~~well ~~in mixed forests and ~~evergreen ~~coniferous forests~~, with mean R^2 values of~~. The average R^2 for daily-scale simulations reached 0.~~93~~92 and 0.91, respectively. ~~In addition, ~~Further analysis showed that extending the ~~time~~temporal scale of input data ~~can further enhance~~significantly improves model accuracy. ~~Specifically, the mean R^2 of forest ecosystems increased from ~~ In a comparison of models with different variable combinations, it was found ~~0.89 to 0.93, that of grassland ecosystems from 0.79 to 0.83, and ~~that ~~of farmland ecosystems from 0.84 to 0.87. Analysis of the importance of the variables by ~~the model driven by EVI outperformed those driven by NDVI and LAI. The model using LSWI as the ~~importance ranking method showed that vegetation index~~driving variable performed better than those with EF, SW, CPD, Pre, and ~~temperature were the most important variables for GPP estimation in forest, grassland and farmland ecosystems, while the importance of moisture index was relatively low. Of which, temperature were the ~~RH as primary variables~~ in the construction of FLAML-LUE models for forest, grassland and farmland ecosystems. The GPP time-series plots for each site indicated that the FLAML model was able to simulate seasonal dynamics more accurately at most of the ~~, with the EVI+LSWI combination yielding the best performance. Additionally, the model's prediction accuracy decreased under high temperature and high VPD conditions. However, under drought conditions, the overall prediction accuracy increased, although it decreased at some sites~~ but generally underestimated the GPP peaks. These results suggest that~~.

In summary, the FLAML-LUE model ~~is highly capable of predicting~~

~~GPP~~demonstrates strong applicability and ~~has significant~~ potential for wider application in GPP estimation. It holds promise for scaling ~~up GPP from flux footprints to larger areas, enhancing our~~from site-level to regional or even global levels, contributing to a deeper understanding of carbon ~~dynamics. However, it is important to note that the FLAML-LUE model demonstrates limited performance in alpine meadows, highlighting the need for further research~~cycling processes. However, the model's applicability in unique ecosystems, such as alpine shrublands, remains limited, and its ability to adapt to extreme climate events requires further enhancement. Future work should focus on optimizing the model structure and parameter settings to improve ~~GPP modeling in these ecosystems in the future~~its robustness and generalization across diverse ecological environments.

CRediT authorship contribution statement

J.L., Y.Z. and J.W. conceived the study. J.L. collected and processed the data. J.L. and Y.Z. drafted the manuscript. A.W., Y.Z., R.L and W.D. funded the study~~,~~. J.L., Y.Z., A.W, W.F. and J.W. checked the negatives and touched up. All authors have read and agreed to the embellished manuscript.

Data availability

A Fast Library for Automated Machine Learning & Tuning (FLAML) is a Python library, and detailed documentation about FLAML can be found on GitHub. We have uploaded the related source code and documentation to Zenodo (https://doi.org/10.5281/zenodo.14874754, Laijie, 2025). The flux observation data and the Python source code of the FLAML-LUE used in this paper are also archived on

Zenodo (https://doi.org/10.5281/zenodo.14542880, Laijie, 2024).

Declaration of competing ~~interest~~interests

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

Acknowledgments

This study was financially supported by the National Key Research and Development Program of China (Grant Number: 2022YFF1300501), the Natural Science Foundation of Liaoning Province (Grant Number: 2024-BSBA-62), the Open Research Fund Project of Key Laboratory of Ecosystem Carbon Source and Sink, China Meteorological Administration (Grant Number: ECSS-CMA202305), the Fundamental Research Funds of the Chinese Academy of Meteorological Sciences (Grant Number: 2024Z001). This work utilized eddy covariance data obtained from ChinaFlux. We appreciate all the staff at ChinaFlux for providing high-quality measurement data to the scientific community.

References

Adams, M.D., Massey, F., Chastko, K., Cupini, C., 2020. Spatial modelling of particulate matter air pollution sensor measurements collected by community scientists while cycling, land use regression with spatial cross-validation, and applications of machine learning for data correction. Atmos. Environ. 230, 117479. https://doi.org/10.1016/j.atmosenv.2020.117479

Alemohammad, S.H., Fang, B., Konings, A.G., Aires, F., Green, J.K., Kolassa, J., Miralles, D., Prigent, C., Gentine, P., 2017. Water, Energy, and Carbon with Artificial Neural Networks (WECANN): a statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence. Biogeosciences 14, 4101–4124. https://doi.org/10.5194/bg-14-4101-2017

Anderson, G.B., Bell, M.L., 2010. Heat Waves in the United States: Mortality Risk during Heat Waves and Effect Modification by Heat Wave Characteristics in 43 U.S. Communities. Environ. Health Perspect. 119, 210. https://doi.org/10.1289/ehp.1002313

Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A., Kustas, W.P., 2007. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 2. Surface moisture climatology. J. Geophys. Res. Atmospheres 112. https://doi.org/10.1029/2006JD007507

Ayantobo, O.O., Li, Y., Song, S., 2019. Multivariate Drought Frequency Analysis using Four-Variate Symmetric and Asymmetric Archimedean Copula Functions. Water Resour. Manag. 33, 103–127. https://doi.org/10.1007/s11269-018-2090-6

Barbour, M.T., 2021. Estimating Organic Carbon Burial in Freshwater Impoundments with a Rapid-Assessment Model and Geospatial Analysis (M.S.).

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M.A., Baldocchi, D., Bonan, G.B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K.W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F.I., Papale, D., 2010. Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate. Science 329, 834–838. https://doi.org/10.1126/science.1184984

Bhattacharyya, P., Neogi, S., Singha Roy, K., Rao, K.S., 2013. Gross primary production, ecosystem respiration and net ecosystem exchange in Asian rice paddy: An eddy covariance-based approach. Curr. Sci. 104, 67–75.

~~Bi, J., Knyazikhin, Y., Choi, S., Park, T., Barichivich, J., Ciais, P., Fu, R., Ganguly, S., Hall, F., Hilker, T., Huete, A., Jones, M., Kimball, J., Lyapustin, A.I., Mõttus, M., Nemani, R.R., Piao, S., Poulter, B., Saleska, S.R., Saatchi, S.S., Xu, L., Zhou, L., Myneni, R.B., 2015. Sunlight mediated seasonality in canopy structure and photosynthetic activity of Amazonian rainforests. Environ. Res. Lett. 10, 064014. https://doi.org/10.1088/1748-9326/10/6/064014~~

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Cai, W., Ullah, S., Yan, L., Lin, Y., 2021. Remote Sensing of Ecosystem Water Use Efficiency: A Review of Direct and Indirect Estimation Methods. Remote Sens. 13, 2393. https://doi.org/10.3390/rs13122393

Cai, W., Yuan, W., Liang, S., Liu, S., Dong, W., Chen, Y., Liu, D., Zhang, H., 2014. Large Differences in Terrestrial Vegetation Production Derived from Satellite-Based Light Use Efficiency Models. Remote Sens. 6, 8945–8965. https://doi.org/10.3390/rs6098945

Chaney, N.W., Herman, J.D., Ek, M.B., Wood, E.F., 2016. Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning. J. Geophys. Res. Atmospheres 121, 13,218-13,235. https://doi.org/10.1002/2016JD024821

Chang, X., Xing, Y., Gong, W., Yang, C., Guo, Z., Wang, D., Wang, J., Yang, H., Xue, G., Yang, S., 2023. Evaluating gross primary productivity over 9 ChinaFlux sites based on random forest regression models, remote sensing, and eddy covariance data. Sci. Total Environ. 875, 162601. https://doi.org/10.1016/j.scitotenv.2023.162601

Chen, A., Huang, L., Liu, Q., Piao, S., 2021. Optimal temperature of vegetation productivity and its linkage with climate and elevation on the Tibetan Plateau. Glob. Change Biol. 27, 1942–1951. https://doi.org/10.1111/gcb.15542

Chen, S.-P., Cui-Hai, Y.O.U., Zhong-Min, H.U., Zhi, C., Lei-Ming, Z., Qiu-Feng, W., 2020. Eddy covariance technique and its applications in flux observations of terrestrial ecosystems.

Chin. J. Plant Ecol. 44, 291. https://doi.org/10.17521/cjpe.2019.0351

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Chen, Y., Feng, X., Fu, B., Wu, X., Gao, Z., 2021. Improved Global Maps of the Optimum Growth Temperature, Maximum Light Use Efficiency, and Gross Primary Production for Vegetation. J. Geophys. Res. Biogeosciences 126, e2020JG005651. https://doi.org/10.1029/2020JG005651

Coops, N.C., Waring, R.H., 2001. The use of multiscale remote sensing imagery to derive regional estimates of forest growth capacity using 3-PGS. Remote Sens. Environ. 75, 324–334. https://doi.org/10.1016/S0034-4257(00)00176-0

Cox, P.M., Betts, R.A., Jones, C.D., Spall, S.A., Totterdell, I.J., 2000. Erratum: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408, 750–750. https://doi.org/10.1038/35047138

Ercoli, L., 1993. Relationship between nitrogen and chlorophyll content and spectral properties in maize leaves. Eur. J. Agron.

Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., Smola, A., 2020. AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data. https://doi.org/10.48550/arXiv.2003.06505

Frank, Dorothea, Reichstein, M., Bahn, M., Thonicke, K., Frank, David, Mahecha, M.D., Smith, P., van der Velde, M., Vicca, S., Babst, F., Beer, C., Buchmann, N., Canadell, J.G., Ciais, P., Cramer, W., Ibrom, A., Miglietta, F., Poulter, B., Rammig, A., Seneviratne, S.I., Walz, A., Wattenbach, M., Zavala, M.A., Zscheischler, J., 2015. Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts. Glob. Change Biol. 21, 2861–2880. https://doi.org/10.1111/gcb.12916

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63, 3–42. https://doi.org/10.1007/s10994-006-6226-1

Gherardi, L.A., Sala, O.E., 2020. Global patterns and climatic controls of belowground net carbon fixation. Proc. Natl. Acad. Sci. 117, 20038–20043. https://doi.org/10.1073/pnas.2006715117

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ., Big Remotely Sensed Data: tools, applications and experiences 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

~~Gosiewska, A., Biecek, P., 2019. auditor: an R Package for Model-Agnostic Visual Validation and Diagnostics. R J. 11, 85–98.~~

Gumus, V., 2023. Evaluating the effect of the SPI and SPEI methods on drought monitoring over Turkey. J. Hydrol. 626, 130386. https://doi.org/10.1016/j.jhydrol.2023.130386

Guo, Q., Hu, Z., Li, S., Yu, G., Sun, X., Zhang, L., Mu, S., Zhu, X., Wang, Y., Li, Y., Zhao, W., 2015. Contrasting responses of gross primary productivity to precipitation events in a water-limited and a temperature-limited grassland ecosystem. Agric. For. Meteorol. 214–215, 169–177. https://doi.org/10.1016/j.agrformet.2015.08.251

Harris, N.L., Gibbs, D.A., Baccini, A., Birdsey, R.A., de Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M.C., Herold, M., Houghton, R.A., Potapov, P.V., Suarez, D.R., Roman-Cuesta, R.M., Saatchi, S.S., Slay, C.M., Turubanova, S.A., Tyukavina, A., 2021. Global maps of twenty-first century forest carbon fluxes. Nat. Clim. Change 11, 234–240. https://doi.org/10.1038/s41558-020-00976-6

~~He, H., Zhang, Liyun, Gao, Y., Ren, X., Zhang, Li, Yu, G., Wang, S., 2015. Regional representativeness assessment and improvement of eddy flux observations in China. Sci. Total Environ. 502, 688–698. https://doi.org/10.1016/j.scitotenv.2014.09.073~~

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803

Jiang, G., Sun, R., Zhang, L., Liu, S., Xu, Z., Qiao, C., 2014. Analysis of light use efficiency and gross primary productivity based on remote sensing data over a phragmites-dominated wetland in Zhangye, China, in: Land Surface Remote Sensing II. Presented at the Land Surface Remote Sensing II, SPIE, pp. 571–578. https://doi.org/10.1117/12.2068840

Jung, M., Reichstein, M., Margolis, H.A., Cescatti, A., Richardson, A.D., Arain, M.A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B.E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E.J., Papale, D., Sottocornola, M., Vaccari, F., Williams, C., 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J. Geophys. Res. Biogeosciences 116. https://doi.org/10.1029/2010JG001566

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D.S., Haverd, V., Köhler, P., Ichii, K., Jain, A.K., Liu, J., Lombardozzi, D., Nabel, J.E.M.S., Nelson, J.A., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., Reichstein, M., 2020. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. Biogeosciences 17, 1343–1365. https://doi.org/10.5194/bg-17-1343-2020

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30 (NIPS 2017), Advances in Neural Information Processing Systems. Presented at the 31st Annual Conference on Neural Information Processing Systems (NIPS), Neural Information Processing Systems (nips), La Jolla.

Kong, D., Yuan, D., Li, H., Zhang, J., Yang, S., Li, Y., Bai, Y., Zhang, S., 2023. Improving the Estimation of Gross Primary Productivity across Global Biomes by Modeling Light Use Efficiency through Machine Learning. Remote Sens. 15, 2086. https://doi.org/10.3390/rs15082086

Landry, J.-S., Matthews, H.D., 2016. Non-deforestation fire vs. fossil fuel combustion: The source of CO2 emissions affects the global carbon cycle and climate responses. Biogeosciences 13, 2137–2149. https://doi.org/10.5194/bg-13-2137-2016

LeDell, E., Poirier, S., 2020. H2O AutoML: Scalable Automatic Machine Learning.

Li, H., Zhang, F., Li, Y., Wang, J., Zhang, L., Zhao, L., Cao, G., Zhao, X., Du, M., 2016. Seasonal and inter-annual variations in CO2 fluxes over 10 years in an alpine shrubland on the Qinghai-Tibetan Plateau, China. Agric. For. Meteorol. 228–229, 95–103. https://doi.org/10.1016/j.agrformet.2016.06.020

Lloyd, J., Taylor, J.A., 1994. On the Temperature Dependence of Soil Respiration. Funct. Ecol. 8, 315–323. https://doi.org/10.2307/2389824

~~Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30 (NIPS 2017), Advances in Neural Information Processing Systems. Presented at the 31st Annual Conference on Neural Information Processing Systems (NIPS), Neural Information Processing Systems (nips), La Jolla.~~

Mahadevan, P., Wofsy, S.C., Matross, D.M., Xiao, X., Dunn, A.L., Lin, J.C., Gerbig, C., Munger, J.W., Chow, V.Y., Gottlieb, E.W., 2008. A satellite-based biosphere parameterization for net ecosystem CO2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM). Glob. Biogeochem. Cycles 22. https://doi.org/10.1029/2006GB002735

McDowell, N., Pockman, W.T., Allen, C.D., Breshears, D.D., Cobb, N., Kolb, T., Plaut, J., Sperry, J., West, A., Williams, D.G., Yepez, E.A., 2008. Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? New Phytol. 178, 719–739. https://doi.org/10.1111/j.1469-8137.2008.02436.x

Melanie, 2023. TPOT: All about this Machine Learning Python library. Data Sci. Courses DataScientest. URL https://datascientest.com/en/tpot-all-about-this-machine-learning-python-library (accessed 6.2.24).

Menefee, D., Lee, T.O., Flynn, K.C., Chen, J., Abraha, M., Baker, J., Suyker, A., 2023. Machine learning algorithms improve MODIS GPP estimates in United States croplands. Front. Remote Sens. 4. https://doi.org/10.3389/frsen.2023.1240895

Novick, K.A., Ficklin, D.L., Stoy, P.C., Williams, C.A., Bohrer, G., Oishi, A.C., Papuga, S.A., Blanken, P.D., Noormets, A., Sulman, B.N., Scott, R.L., Wang, L., Phillips, R.P., 2016. The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. Nat. Clim. Change 6, 1023–1027. https://doi.org/10.1038/nclimate3114

Pagán, B.R., Maes, W.H., Gentine, P., Martens, B., Miralles, D.G., 2019. Exploring the Potential of Satellite Solar-Induced Fluorescence to Constrain Global Transpiration Estimates. Remote Sens. 11, 413. https://doi.org/10.3390/rs11040413

Pei, Y., Dong, J., Zhang, Y., Yuan, W., Doughty, R., Yang, J., Zhou, D., Zhang, L., Xiao, X., 2022. Evolution of light use efficiency models: Improvement, uncertainties, and implications. Agric. For. Meteorol. 317, 108905. https://doi.org/10.1016/j.agrformet.2022.108905

Peltoniemi, M., Pulkkinen, M., Kolari, P., Duursma, R.A., Montagnani, L., Wharton, S., Lagergren, F., Takagi, K., Verbeeck, H., Christensen, T., Vesala, T., Falk, M., Loustau, D., Mäkelä, A., 2012. Does canopy mean nitrogen concentration explain variation in canopy light use efficiency across 14 contrasting forest sites? Tree Physiol. 32, 200–218.

https://doi.org/10.1093/treephys/tpr140

Potter, C.S., Randerson, J.T., Field, C.B., Matson, P.A., Vitousek, P.M., Mooney, H.A., Klooster, S.A., 1993. Terrestrial ecosystem production: A process model based on global satellite and surface data. Glob. Biogeochem. Cycles 7, 811–841. https://doi.org/10.1029/93GB02725

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features, in: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., CesaBianchi, N., Garnett, R. (Eds.), ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 31 (NIPS 2018), Advances in Neural Information Processing Systems. Presented at the 32nd Conference on Neural Information Processing Systems (NIPS), Neural Information Processing Systems (nips), La Jolla.

Qian, L., Yu, X., Zhang, Z., Wu, L., Fan, J., Xiang, Y., Chen, J., Liu, X., 2024. Assessing and improving the high uncertainty of global gross primary productivity products based on deep learning under extreme climatic conditions. Sci. Total Environ. 957, 177344. https://doi.org/10.1016/j.scitotenv.2024.177344

Qian, L., Zhang, Z., Wu, L., Fan, S., Yu, X., Liu, X., Ba, Y., Ma, H., Wang, Y., 2023. High uncertainty of evapotranspiration products under extreme climatic conditions. J. Hydrol. 626, 130332. https://doi.org/10.1016/j.jhydrol.2023.130332

Qu, L., De Boeck, H.J., Fan, H., Dong, G., Chen, J., Xu, W., Ge, Z., Huang, Z., Shao, C., Hu, Y., 2020. Diverging Responses of Two Subtropical Tree Species (Schima superba and Cunninghamia lanceolata) to Heat Waves. Forests 11, 513. https://doi.org/10.3390/f11050513

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M.D., Seneviratne, S.I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D.C., Papale, D., Rammig, A., Smith, P., Thonicke, K., van der Velde, M., Vicca, S., Walz, A., Wattenbach, M., 2013. Climate extremes and the carbon cycle. Nature 500, 287–295. https://doi.org/10.1038/nature12350

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1

Reichstein, M., Ciais, P., Papale, D., Valentini, R., Running, S., Viovy, N., Cramer, W., Granier, A., Ogee, J., Allard, V., Aubinet, M., Bernhofer, C., Buchmann, N., Carrara, A., Grunwald, T., Heimann, M., Heinesch, B., Knohl, A., Kutsch, W., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J., Pilegaard, K., Pumpanen, J., Rambal, S., Schaphoff, S., Seufert, G., Soussana, J.-F., Sanz, M.-J., Vesala, T., Zhao, M., 2007. Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. Glob. Change Biol. 13, 634–651. https://doi.org/10.1111/j.1365-2486.2006.01224.x

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Glob. Change Biol. 11, 1424–1439. https://doi.org/10.1111/j.1365-2486.2005.001002.x

Rigden, A.J., Mueller, N.D., Holbrook, N.M., Pillai, N., Huybers, P., 2020. Combined influence of soil moisture and atmospheric evaporative demand is important for accurately predicting US maize yields. Nat. Food 1, 127–133. https://doi.org/10.1038/s43016-020-0028-7

Rosebrock, A., 2019. Auto-Keras and AutoML: A Getting Started Guide. PyImageSearch. URL https://pyimagesearch.com/2019/01/07/auto-keras-and-automl-a-getting-started-guide/ (accessed 6.2.24).

Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., Hashimoto, H., 2004. A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. BioScience 54, 547–560. https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2

Schmid, H.P., 2002. Footprint modeling for vegetation atmosphere exchange studies: a review and perspective. Agric. For. Meteorol., FLUXNET 2000 Synthesis 113, 159–183. https://doi.org/10.1016/S0168-1923(02)00107-7

Sellers, P.J., Schimel, D.S., Moore, B., Liu, J., Eldering, A., 2018. Observing carbon cycle–climate feedbacks from space. Proc. Natl. Acad. Sci. 115, 7860–7868. https://doi.org/10.1073/pnas.1716613115

~~Sims, D.A., Brzostek, E.R., Rahman, A.F., Dragoni, D., Phillips, R.P., 2014. An improved approach for remotely sensing water stress impacts on forest C uptake. Glob. Change Biol. 20, 2856–2866. https://doi.org/10.1111/gcb.12537~~

Stefanon, M., D'Andrea, F., Drobinski, P., 2012. Heatwave classification over Europe and the Mediterranean region. Environ. Res. Lett. 7, 014023. https://doi.org/10.1088/1748-9326/7/1/014023

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmospheres 106, 7183–7192. https://doi.org/10.1029/2000JD900719

Thornton, C., Hutter, F., Hoos, H.H., Leyton-Brown, K., 2013. Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms. https://doi.org/10.48550/arXiv.1208.3719

Tramontana, G., Jung, M., Schwalm, C.R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M.A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., Papale, D., 2016. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. Biogeosciences 13, 4291–4313. https://doi.org/10.5194/bg-13-4291-2016

Venturini, V., Bisht, G., Islam, S., Jiang, L., 2004. Comparison of evaporative fractions estimated from AVHRR and MODIS sensors over South Florida. Remote Sens. Environ. 93, 77–86. https://doi.org/10.1016/j.rse.2004.06.020

Vicca, S., Bahn, M., Estiarte, M., van Loon, E.E., Vargas, R., Alberti, G., Ambus, P., Arain, M.A., Beier, C., Bentley, L.P., Borken, W., Buchmann, N., Collins, S.L., de Dato, G., Dukes, J.S., Escolar, C., Fay, P., Guidolotti, G., Hanson, P.J., Kahmen, A., Kröel-Dulay, G., Ladreiter-Knauss, T., Larsen, K.S., Lellei-Kovacs, E., Lebrija-Trejos, E., Maestre, F.T., Marhan, S., Marshall, M., Meir, P., Miao, Y., Muhr, J., Niklaus, P.A., Ogaya, R., Peñuelas, J., Poll, C., Rustad, L.E., Savage, K., Schindlbacher, A., Schmidt, I.K., Smith, A.R., Sotta, E.D., Suseela, V., Tietema, A., van Gestel, N., van Straaten, O., Wan, S., Weber, U., Janssens, I.A., 2014. Can current moisture responses predict soil CO_2 efflux under altered precipitation regimes? A synthesis of manipulation experiments. Biogeosciences 11, 2991–

3013. https://doi.org/10.5194/bg-11-2991-2014

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. https://doi.org/10.1175/2009JCLI2909.1

Wang, C., Wu, Q., Weimer, M., Zhu, E., 2021. FLAML: A Fast and Lightweight AutoML Library. https://doi.org/10.48550/arXiv.1911.04706

Wang, H., ~~Shao, W., Hu, Y., Cao, W., Zhang, Y~~Guan, H., Liu, B., Chen, X., 2023. ~~Assessment~~Impacts of ~~Six Machine Learning Methods for Predicting Gross Primary Productivity~~climate extremes on vegetation dynamics in ~~Grassland. Remote Sens. 15, 3475.~~a transect along the Hu Line of China. Ecol. Indic. 155, 111043. https://doi.org/10.~~3390/rs15143475~~1016/j.ecolind.2023.111043

Wang, H., He, B., Zhang, Y., Huang, L., Chen, Z., Liu, J., 2018. Response of ecosystem productivity to dry/wet conditions indicated by different drought indices. Sci. Total Environ. 612, 347–357. https://doi.org/10.1016/j.scitotenv.2017.08.212

Wang, J., Liu, J., Cao, M., Liu, Y., Yu, G., Li, G., Qi, S., Li, K., 2011. Modelling carbon fluxes of different forests by coupling a remote-sensing model with an ecosystem process model. Int. J. Remote Sens. 32, 6539–6567. https://doi.org/10.1080/01431161.2010.512933

Wang, Y., Li, R., Hu, J., Fu, Y., Duan, J., Cheng, Y., 2021. Daily estimation of gross primary production under all sky using a light use efficiency model coupled with satellite passive microwave measurements. Remote Sens. Environ. 267, 112721. https://doi.org/10.1016/j.rse.2021.112721

~~Wei, S., Yi, C., Fang, W., Hendrey, G., 2017. A global study of GPP focusing on light-use efficiency in a random forest regression model. Ecosphere 8, e01724. https://doi.org/10.1002/ecs2.1724~~

~~Wu, C., Munger, J.W., Niu, Z., Kuang, D., 2010. Comparison of multiple models for estimating gross primary production using MODIS and eddy covariance data in Harvard Forest. Remote Sens. Environ. 114, 2925–2939. https://doi.org/10.1016/j.rse.2010.07.012~~

Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J.A., Huete, A.R., Ichii, K., Ni, W., Pang, Y., Rahman, A.F., Sun, G., Yuan, W., Zhang, L., Zhang, X., 2019. Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. Remote Sens. Environ. 233, 111383. https://doi.org/10.1016/j.rse.2019.111383

Xiao, X., Braswell, B., Zhang, Q., Boles, S., Frolking, S., Moore, B., 2003. Sensitivity of vegetation indices to atmospheric aerosols: continental-scale observations in Northern Asia. Remote Sens. Environ. 84, 385–392. https://doi.org/10.1016/S0034-4257(02)00129-3

~~Xie~~Xiao, X., Hollinger, D., Aber, J., ~~Sun, G., Chu, H.-S., Liu, J., McNulty, S.G., Noormets,~~ Goltz, M., Davidson, E.A., ~~John, R., Ouyang, Z., Zha, T., Li, H., Guan, W., Chen, J., 2014. Long-term variability in the water budget and its controls~~Zhang, Q., Moore, B., 2004. Satellite-based modeling of gross primary production in an ~~oak-dominated temperate~~evergreen needleleaf forest. ~~Hydrol. Process. 28, 6054–6066.~~Remote Sens. Environ. 89, 519–534. https://doi.org/10.~~1002/hyp.10079~~1016/j.rse.2003.11.008

Xie, M., Ma, X., Wang, Y., Li, C., Shi, H., Yuan, X., Hellwich, O., Chen, C., Zhang, W., Zhang, C., Ling, Q., Gao, R., Zhang, Y., Ochege, F.U., Frankl, A., De Maeyer, P., Buchmann, N., Feigenwinter, I., Olesen, J.E., Juszczak, R., Jacotot, A., Korrensalo, A., Pitacco, A., Varlagin, A., Shekhar, A., Lohila, A., Carrara, A., Brut, A., Kruijt, B., Loubet, B., Heinesch, B., Chojnicki, B., Helfter, C., Vincke, C., Shao, C., Bernhofer, C., Brümmer, C., Wille, C., Tuittila, E.-S., Nemitz, E., Meggio, F., Dong, G., Lanigan, G., Niedrist, G., Wohlfahrt, G., Zhou, G., Goded, I., Gruenwald, T., Olejnik, J., Jansen, J., Neirynck, J., Tuovinen, J.-P., Zhang, J., Klumpp, K., Pilegaard, K., Šigut, L., Klemedtsson, L., Tezza, L., Hörtnagl, L., Urbaniak, M., Roland, M., Schmidt, M., Sutton, M.A., Hehn, M., Saunders, M., Mauder, M., Aurela, M., Korkiakoski, M., Du, M., Vendrame, N., Kowalska, N., Leahy, P.G., Alekseychik, P., Shi, P., Weslien, P., Chen, S., Fares, S., Friborg, T., Tallec, T., Kato, T., Sachs, T., Maximov, T., di Cella, U.M., Moderow, U., Li, Y., He, Y., Kosugi, Y., Luo, G., 2023. Monitoring of carbon-water fluxes at Eurasian meteorological stations using random forest and remote sensing. Sci. Data 10, 587. https://doi.org/10.1038/s41597-023-02473-9

~~Xie, X., Li, A., 2020. An Adjusted Two-Leaf Light Use Efficiency Model for Improving GPP Simulations Over Mountainous Areas. J. Geophys. Res. Atmospheres 125, e2019JD031702. https://doi.org/10.1029/2019JD031702~~

Yi, K., Li, R., Scanlon, T.M., Lerdau, M.T., Berry, J.A., Yang, X., 2024. Impact of atmospheric dryness on solar-induced chlorophyll fluorescence: Tower-based observations at a temperate forest. Remote Sens. Environ. 306, 114106. https://doi.org/10.1016/j.rse.2024.114106

Yu, G., Ren, W., Chen, Z., Zhang, Leiming, Wang, Q., Wen, X., He, N., Zhang, Li, Fang, H., Zhu, X., Gao, Y., Sun, X., 2016. Construction and progress of Chinese terrestrial ecosystem carbon, nitrogen and water fluxes coordinated observation. J. Geogr. Sci. 26, 803–826. https://doi.org/10.1007/s11442-016-1300-5

Yuan, W., Cai, W., Xia, J., Chen, J., Liu, S., Dong, W., Merbold, L., Law, B., Arain, A., Beringer, J., Bernhofer, C., Black, A., Blanken, P.D., Cescatti, A., Chen, Y., Francois, L., Gianelle, D., Janssens, I.A., Jung, M., Kato, T., Kiely, G., Liu, D., Marcolla, B., Montagnani, L., Raschi, A., Roupsard, O., Varlagin, A., Wohlfahrt, G., 2014. Global comparison of light use efficiency models for simulating terrestrial vegetation gross primary production based on the LaThuile database. Agric. For. Meteorol. 192–193, 108–120. https://doi.org/10.1016/j.agrformet.2014.03.007

Yuan, W., Liu, S., Yu, G., Bonnefond, J.-M., Chen, J., Davis, K., Desai, A.R., Goldstein, A.H., Gianelle, D., Rossi, F., Suyker, A.E., Verma, S.B., 2010. Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. Remote Sens. Environ. 114, 1416–1431. https://doi.org/10.1016/j.rse.2010.01.022

Yuan, W., Liu, S., Zhou, Guangsheng, Zhou, Guoyi, Tieszen, L.L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A.H., Goulden, M.L., Hollinger, D.Y., Hu, Y., Law, B.E., Stoy, P.C., Vesala, T., Wofsy, S.C., 2007. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. Agric. For. Meteorol. 143, 189–207. https://doi.org/10.1016/j.agrformet.2006.12.001

Zhang, C., Tian, X., Zhao, Y., Lu, J., 2023. Automated machine learning-based building energy load prediction method. J. Build. Eng. 80, 108071. https://doi.org/10.1016/j.jobe.2023.108071

Zhang, W.L., Chen, S.P., Chen, J., Wei, L., Han, X.G., Lin, G.H., 2007. Biophysical regulations of carbon fluxes of a steppe and a cultivated cropland in semiarid Inner Mongolia. Agric. For. Meteorol. 146, 216–229. https://doi.org/10.1016/j.agrformet.2007.06.002

Zhang, Y., Song, C., Sun, G., Band, L.E., Noormets, A., Zhang, Q., 2015. Understanding moisture stress on light use efficiency across terrestrial ecosystems based on global flux and remote-sensing data. J. Geophys. Res. Biogeosciences 120, 2053–2066. https://doi.org/10.1002/2015JG003023

Zhang, ~~Y., Ye, A., 2022. Uncertainty analysis of multiple terrestrial gross primary productivity products. Glob. Ecol. Biogeogr. 31, 2204–2218. https://doi.org/10.1111/geb.13578~~
~~Zhang, ~~Z., Guo, J., Jin, S., Han, S., 2023. Improving the ability of PRI in light use efficiency estimation by distinguishing sunlit and shaded leaves in rice canopy. Int. J. Remote Sens. 44, 5755–5767. https://doi.org/10.1080/01431161.2023.2252165

Zhao, W.L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., Qiu, G.Y., 2019. Physics-Constrained Machine Learning of Evapotranspiration. Geophys. Res. Lett. 46, 14496–14507. https://doi.org/10.1029/2019GL085291

Zheng, Z., Fiore, A.M., Westervelt, D.M., Milly, G.P., Goldsmith, J., Karambelas, A., Curci, G., Randles, C.A., Paiva, A.R., Wang, C., Wu, Q., Dey, S., 2023. Automated Machine Learning to Evaluate the Information Content of Tropospheric Trace Gas Columns for Fine Particle Estimates Over India: A Modeling Testbed. J. Adv. Model. Earth Syst. 15, e2022MS003099. https://doi.org/10.1029/2022MS003099

Zhou, S.-X., Prentice, I.C., Medlyn, B.E., 2019. Bridging Drought Experiment and Modeling: Representing the Differential Sensitivities of Leaf Gas Exchange to Drought. Front. Plant Sci. 9. https://doi.org/10.3389/fpls.2018.01965