https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

Geoscientific Model Development
Discussions

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. In this work, we developed a series of FLAML-LUE models tailored for different ecosystems. 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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.
Geoscientific Model Development
Discussions

from 0.56 to 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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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 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 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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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 from simple regression models and complex simulation models. 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-suited for addressing nonlinear and complex issues across different ecosystems. They provide innovative approaches for simulating GPP by solving the nonlinear relationship. These models are less reliant on theoretical assumptions. Therefore, many researchers prefer this method in recent years. 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) constructed RFR-LUE models that utilize the 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 (Thornton et al., 2013), H2O (LeDell and Poirier, 2020), TPOT (Melanie, 2023), AutoGluo (Erickson et al., 2020), FLAML (C. Wang et al., 2021), and AutoKera (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) 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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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.

2. Material and methods

2.1 Site description

Fig. 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.

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.
Geoscientific Model Development
Open Access
Discussions

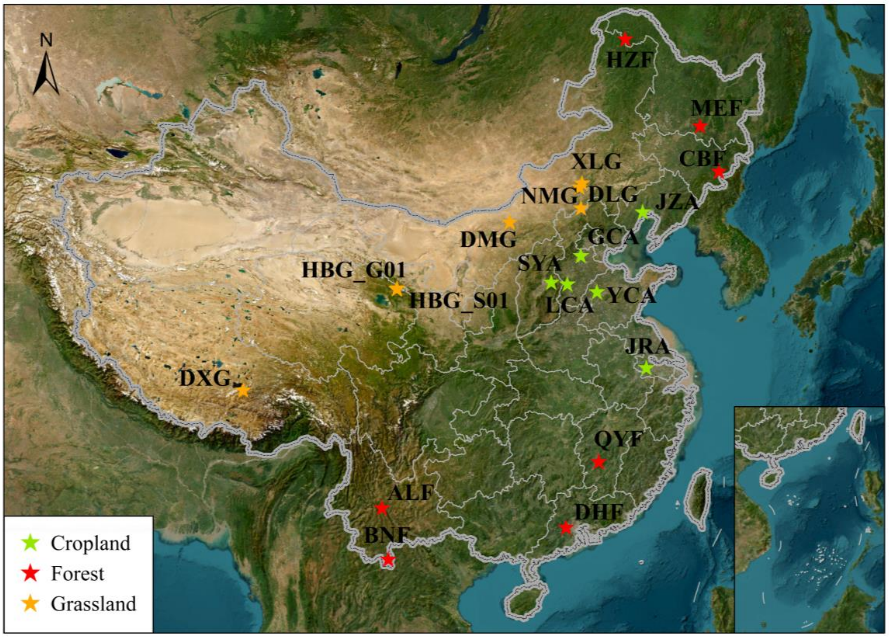

Fig. 1. The location map of the flux site is based on the map approved by the National Surveying and Mapping Bureau of China (Approval No. GS (2019)1822). The topographic map is derived from data provided by Esri, Maxar, Earthstar Geographics, and the GIS User Community (Service Layer Credits).

Table1

Basic information on the 20 flux stations.

Site	Ecosystem type	Surface cover type	Time Range	Classified
HZF	Forest	Coniferous forest	2014-2018	Needle-leaved
MEF	Forest	Deciduous broad-leaved forests	2016-2018	Deciduous Broadleaved
CBF	Forest	Broad-leaved Korean pine forests	2003-2010	Mixed
QYF	Forest	Artificial coniferous forests	2003-2010	Needle-leaved
DHF	Forest	Mixed coniferous and broad-leaved forests	2003-2010	Mixed
ALF	Forest	Evergreen Broadleaved forests	2009-2013	Evergreen Broadleaved
BNF	Forest	Tropical rainforest	2003-2015	Evergreen Broadleaved
XLG	Grassland	Mowing grasslands	2006-2014	Grassland
NMG	Grassland	Temperate steppe	2003-2010	Grassland
DLG	Grassland	Typical grasslands	2006-2015	Grassland
DMG	Grassland	Temperate desert steppe	2015-2018	Grassland
HBG_G01	Grassland	Alpine meadow	2015-2020	Alpine Meadow
HBG_S01	Grassland	Alpine shrub-meadow	2003-2013	Shrub
DXG	Grassland	Alpine meadow	2003-2010	Alpine Meadow
JZA	Cropland	Spring corn	2005-2014	Single Cropping
GCA	Cropland	Winter wheat - Summer corn	2020-2022	Double Cropping

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

SYA	Cropland	Spring corn	2012-2014	Single Cropping
LCA	Cropland	Winter wheat - Summer corn	2013-2017	Double Cropping
YCA	Cropland	Winter wheat - Summer corn	2003-2010	Double Cropping
JRA	Cropland	Winter wheat - Summer rice	2015-2020	Double Cropping

2.2 Data

2.2.1 Eddy covariance data

Eddy covariance (EC) data were collected at 20 sites, including 7 forests, 7 grasslands, and 6 cropland (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, some sites have no ER data, so this study is based on the nocturnal breathing extrapolation method: 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.

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

In the above equation, 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).

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.

$$GPP = ER - NEE \quad (2)$$

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 data

In this study, remote sensing data primarily came from 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. MODIS data from GEE were used to derive vegetation and water indices,

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.
Geoscientific Model Development
Discussions

including EVI, NDVI, LAI, and LSWI, which were calculated using the formulas 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.

2.3 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.

The flowchart of this study is shown in Fig. 2.

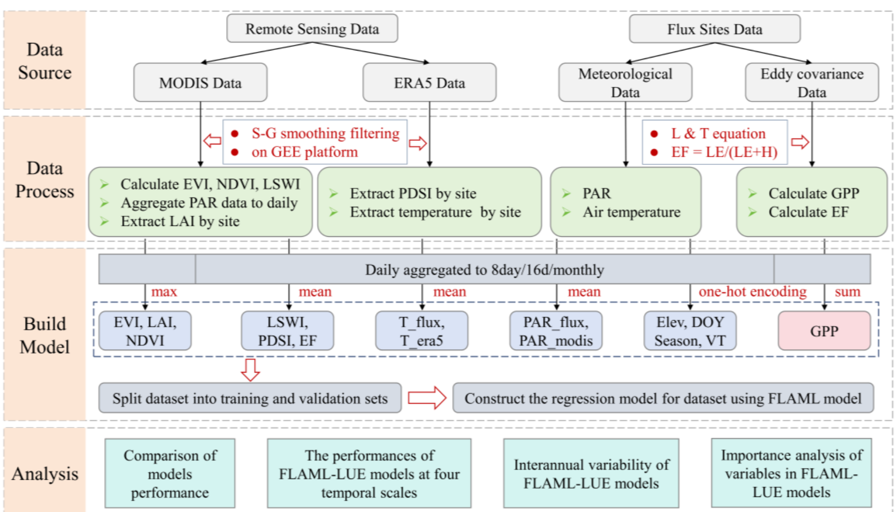

Fig. 2. Flowchart of this study. S-G smoothing filtering: Savitzky-Golay smoothing filtering method, L & T equation: Lloyd & Taylor equation.

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

2.3.1 Data pre-processing

The primary datasets for estimating GPP with FLAML-LUE models include multi-year continuous EC flux data, satellite-based observations, and climate 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). Hence, we integrate vegetation type as a factor in our model.

Table 2
Predictor variables for driving the FLAML models and their specifications.

	Variable	Acquired method (formula)	Original Spatial Resolution	Data Source
Vegetation indices	EVI	$2.5 \times (R_{nir} - R_{red}) / (R_{nir} + 6.0 \times R_{red} - 7.5 \times R_{blue} + 1)$	500m	MOD09GA
	NDVI	$(R_{nir} - R_{red}) / (R_{nir} + R_{red})$		
	LAI	-	500m	MCD15A3H
Water	LSWI	$(R_{nir} - R_{swir}) / (R_{nir} + R_{swir})$	500m	MOD09GA
	PDSI	-	~10km	ERA5
	EF（%）	LE/ (LE + H)	~1km	SDB
Radiation	PAR(μmol m^{-2} s^{-1})	-	~1km	SDB
	PAR(μmol m^{-2} s^{-1})	-	500m	MCD18C2
Temperature	T_flux（℃）	-	~1km	SDB
	T_era5（℃）	-	~10km	ERA5
Vegetation Types	EBF, DBF, CF, MF Grassland Croplands	One-hot encoding	invariant	-
Season	Spring, Summer, Autumn, Winter	One-hot encoding	invariant	-
DOY	Days of year	-	invariant	-
Terrain	Elevation	-	90m	SRTM90

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) 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). 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, using multiple permutations of eight input factor groups, as described in Eq. (3). Table 3 displays the model number based on different variable configurations.

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

Here, the VIi include EVI, NDVI, and LAI; W_j denotes moisture factors including LSWI, EF, and PDSI; VT_i represents vegetation types, in which forest ecosystems include: Needle-leaved, Deciduous Broadleaved, Mixed, and Evergreen Broadleaved; Grassland ecosystems include grasslands, meadows and shrub, and farmland ecosystems include single cropping and double cropping. 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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.
Geoscientific Model Development Discussions

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

Group (Flux)	Input variables	Group (RS)	Input variables
FLAML00	PAR, T_flux, EVI, LSWI, Season, DOY, Elevation, Vegetation Types	FLAML10	PAR_modis, T_era5, EVI, LSWI, Season, DOY, Elevation, Vegetation Types
FLAML01	PAR, T_flux, EVI, PDSI, Season, DOY, Elevation, Vegetation Types	FLAML11	PAR_modis, T_era5, EVI, PDSI, Season, DOY, Elevation, Vegetation Types
FLAML02	PAR, T_flux, EVI, EF, Season, DOY, Elevation, Vegetation Types	FLAML12	PAR_modis, T_era5, EVI, EF, Season, DOY, Elevation, Vegetation Types
FLAML03	PAR, T_flux, NDVI, LSWI, Season, DOY, Elevation, Vegetation Types	FLAML13	PAR_modis, T_era5, NDVI, LSWI, Season, DOY, Elevation, Vegetation Types
FLAML04	PAR, T_flux, NDVI, PDSI, Season, DOY, Elevation, Vegetation Types	FLAML14	PAR_modis, T_era5, NDVI, PDSI, Season, DOY, Elevation, Vegetation Types
FLAML05	PAR, T_flux, NDVI, EF, Season, DOY, Elevation, Vegetation Types	FLAML15	PAR_modis, T_era5, NDVI, EF, Season, DOY, Elevation, Vegetation Types
FLAML06	PAR, T_flux, LAI, LSWI, Season, DOY, Elevation, Vegetation Types	FLAML16	PAR_modis, T_era5, LAI, LSWI, 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

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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

Geoscientific Model Development
Discussions
Open Access
EGU

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.

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

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

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

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

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, 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,

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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.

FLAML	R^2	SD	RMSE	nuRMSE	TSS
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	

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 $gC \cdot m^{-2}d^{-1}$) and the models driven with ERA5

© Author(s) 2025. CC BY 4.0 License.

(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·m^{-2}d^{-1}, and 0.719 gC·m^{-2}d^{-1}, respectively. The model performed slightly better on the coniferous forest at QYF station than on the broad-leaved Korean

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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.

© Author(s) 2025. CC BY 4.0 License.

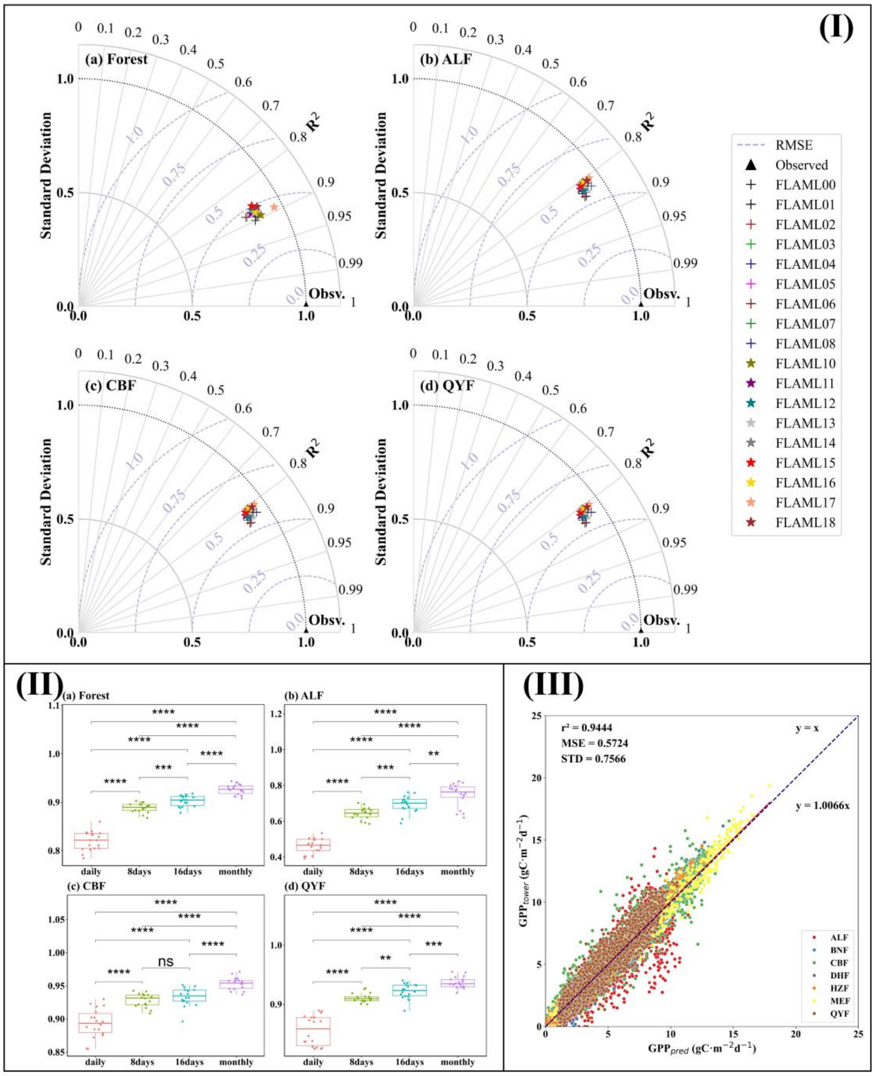

Fig. 3. (I) Normalized Taylor diagrams showing the overall performance of all FLAML-LUE models in (a) forest ecosystem, (b) ALF, (c) CBF, and (d) QYF. (II) Boxplots of 18 model performances (R^2) at different temporal scales in forest ecosystems, ALF, CBF, and QYF. Asterisks indicate significant differences 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.

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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.
Geoscientific Model Development
Discussions
Open Access

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.

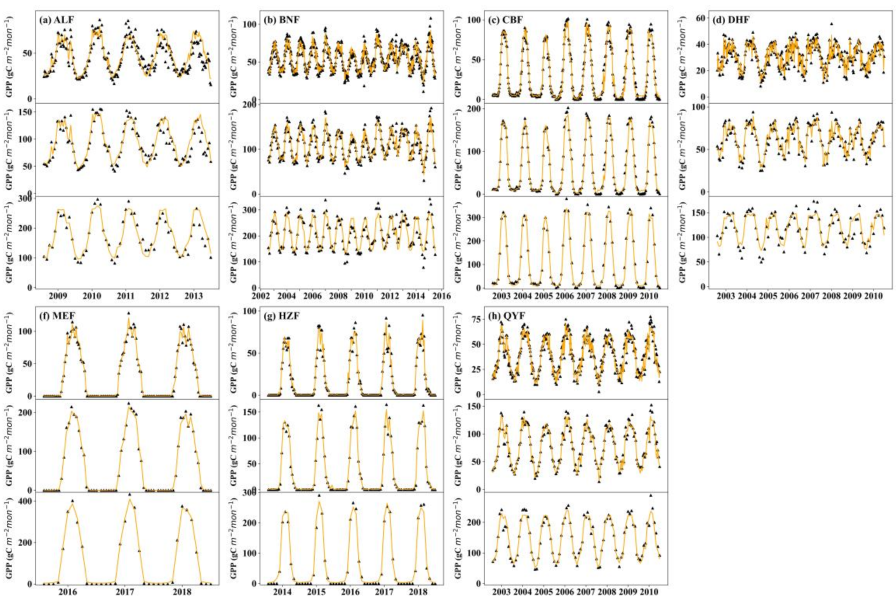

Fig. 4. Plot of simulated GPP time-series variation at ALF, BNF, CBF, DHF, HZF, MEF, and QYF sites, with black triangles representing tower-based observations and orange solid lines representing

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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.

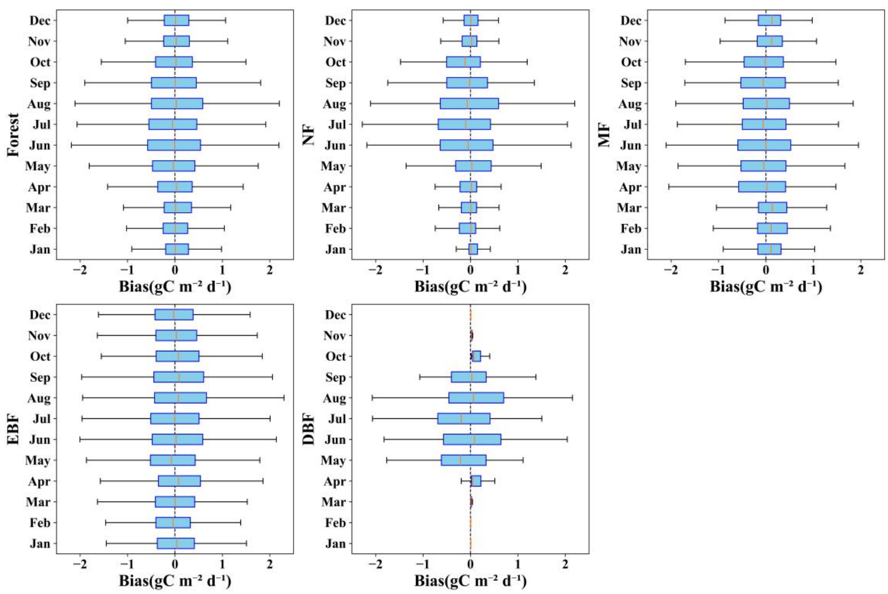

Fig. 5. The monthly bias of FLAML-LUE models among vegetation types. NF: needle-leaf forest,

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.
Geoscientific Model Development Discussions

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

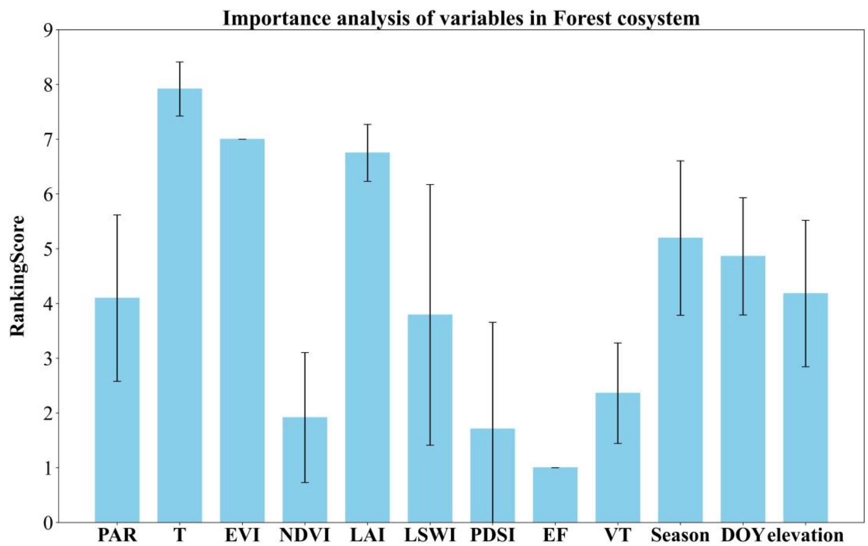

Fig. 6. Average variables importance of forest ecosystem in FALML-LUE models. LSWI: land surface water index, PDSI: Palmer Drought Severity Index, EF: evaporative fraction, EVI: enhanced vegetation index, NDVI: normalized difference vegetation index, LAI: leaf area index, T: temperature, PAR: photosynthetically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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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.

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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 $gC \cdot m^{-2}d^{-1}$, 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 $gC \cdot m^{-2}d^{-1}$. The model driven with EF (R^2 = 0.83, RMSE = 0.833 $gC \cdot m^{-2}d^{-1}$) performed slightly better than LSWI (R2 = 0.82, RMSE = 0.873 $gC \cdot m^{-2}d^{-1}$) and PDSI (R^2 = 0.81, RMSE = 0.877 $gC \cdot m^{-2}d^{-1}$).

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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.

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

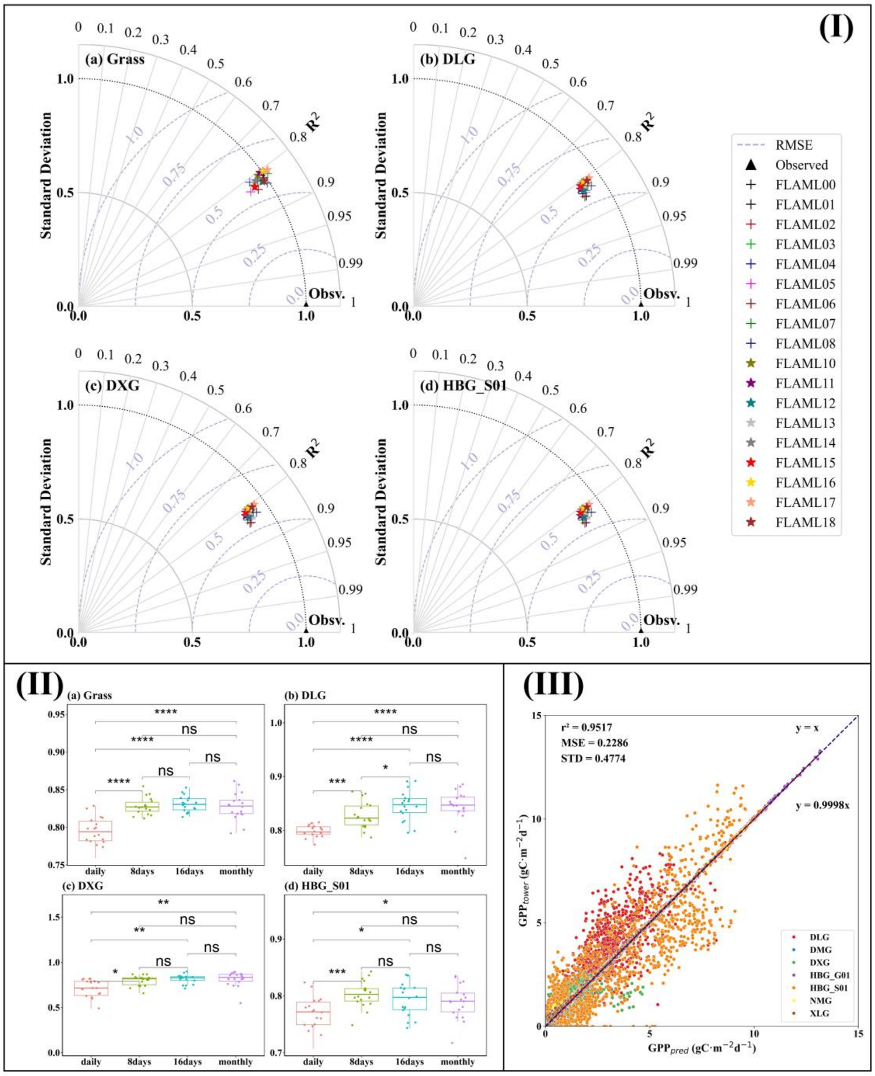

Fig. 7. (I) Normalized Taylor diagrams showing the overall performance of all FLAML-LUE models in (a) grass ecosystem, (b) DXG, (c) DLG, and (d) HBG_S01. (II) Boxplots of 18 model performances (R2) at different temporal scales in grassland ecosystems, DXG, DLG, and HBG_S01. (III) Scatterplot of observed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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.
Geoscientific Model Development Discussions
Open Access

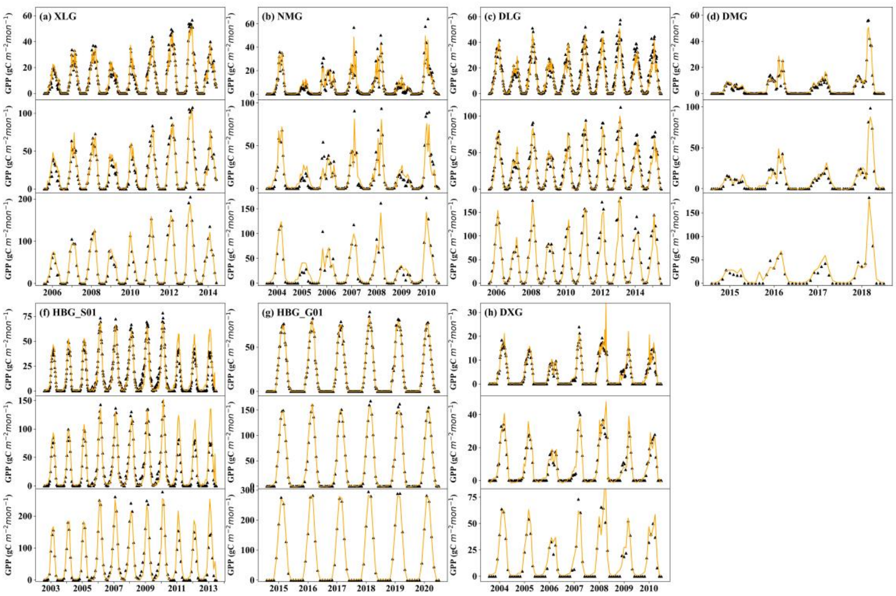

Fig. 8. Plot of simulated GPP time-series variation at DLG, DMG, DXG, HBG_G01, HBG_S01, NMG, and XLG sites.

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,

© Author(s) 2025. CC BY 4.0 License.

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.

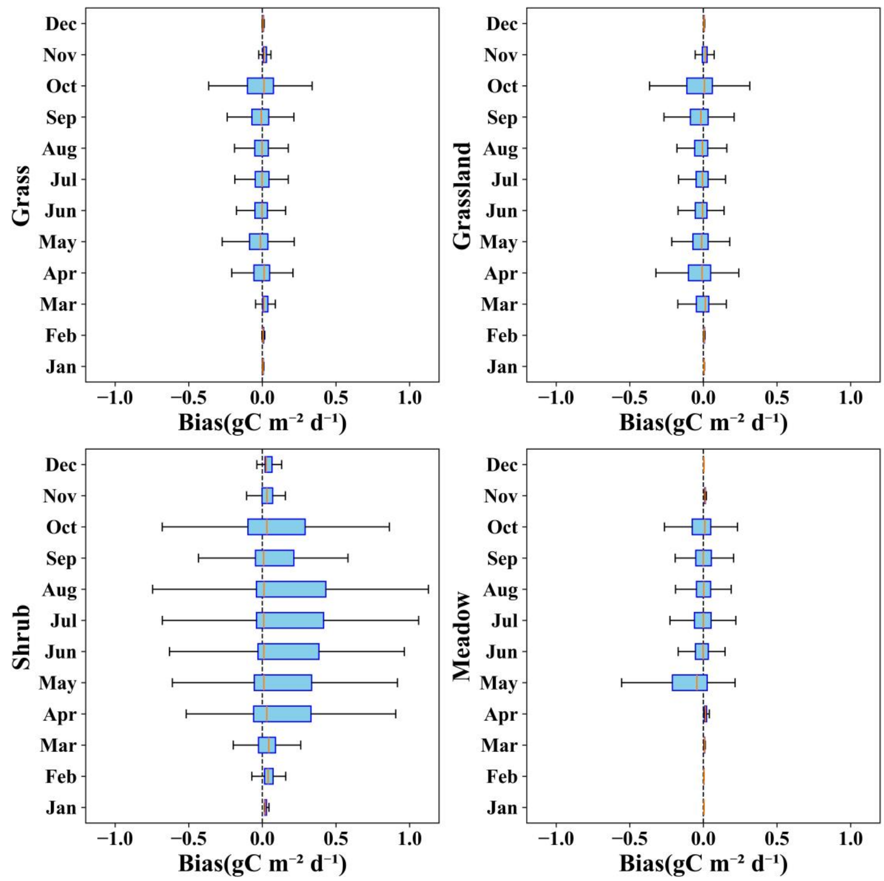

Fig. 9. The monthly bias of FLAML-LUE models among grass types. Grassland: typical grassland, Shrub: alpine shrub, Meadow: alpine meadow.

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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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.

© Author(s) 2025. CC BY 4.0 License.

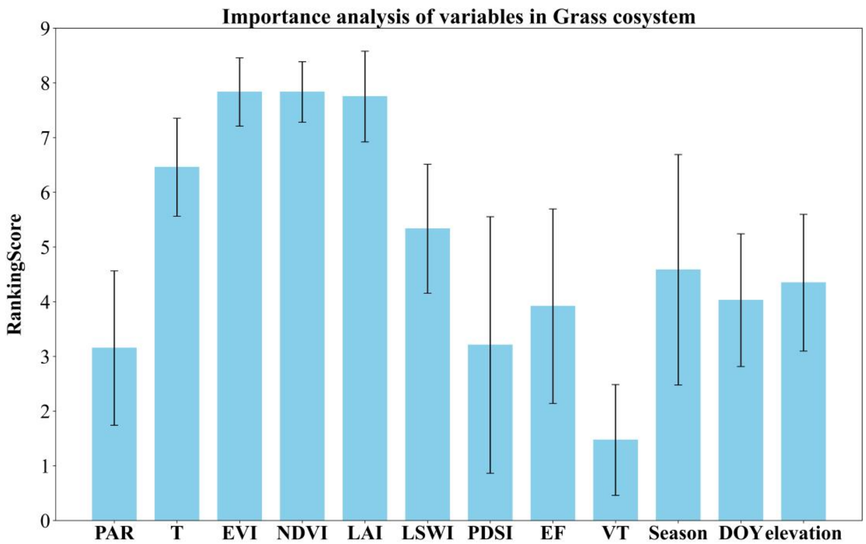

Fig. 10. Average variables importance of grassland ecosystems in FALML-LUE models.

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 $gC \cdot m^{-2}d^{-1}$)

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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

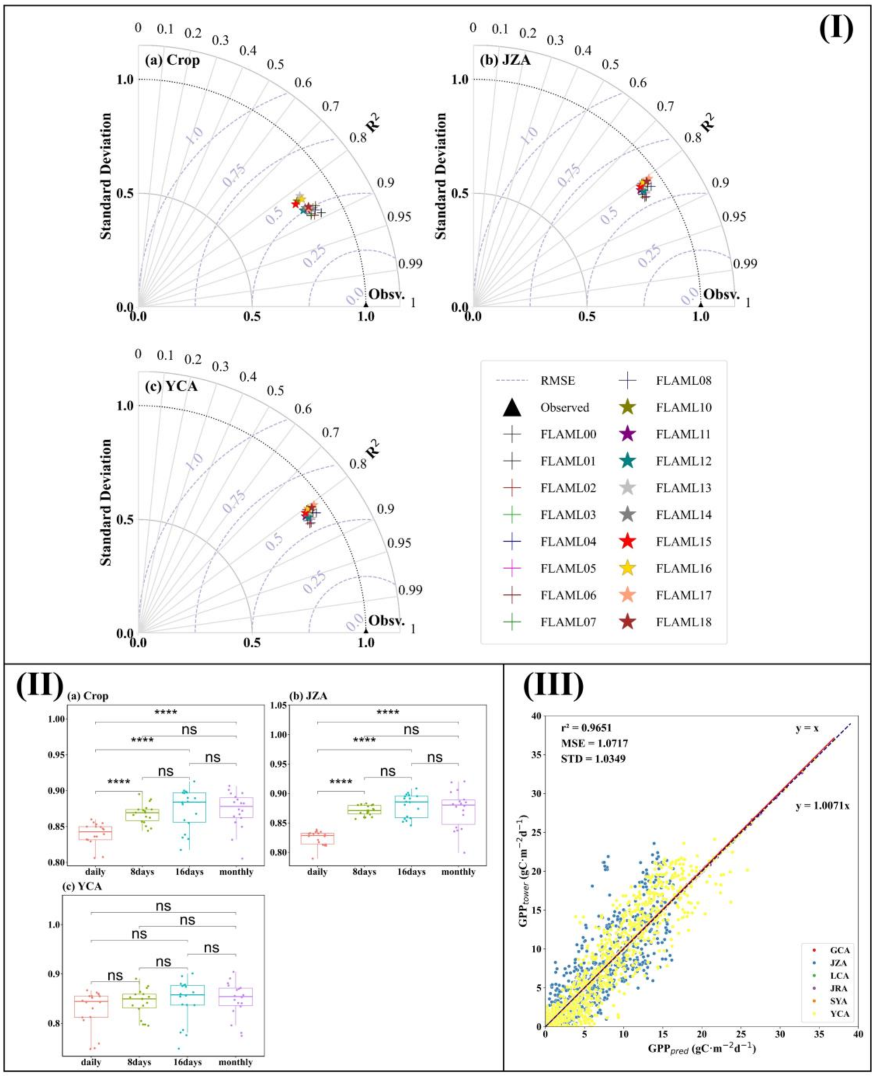

Fig. 11. (I) Normalized Taylor diagrams showing the overall performance of all FLAML-LUE models in (a) cropland ecosystem, (b) JZA, and (c) YCA. (II) Boxplots of 18 model performances (R2) at different temporal scales in crop ecosystem, JZA, and YCA. (III) Scatterplot of observed GPP vs. simulated GPP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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

The average R^2 was 0.86 and the average RMSE was 1.724 $gC \cdot m^{-2}d^{-1}$ for the single cropping farmland station (JZA), while the average R^2 was 0.84 and the average RMSE was 2.400 $gC \cdot m^{-2}d^{-1}$ 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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

Geoscientific Model Development
Discussions
Open Access

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

3.3.2 Analysis of interannual GPP variability

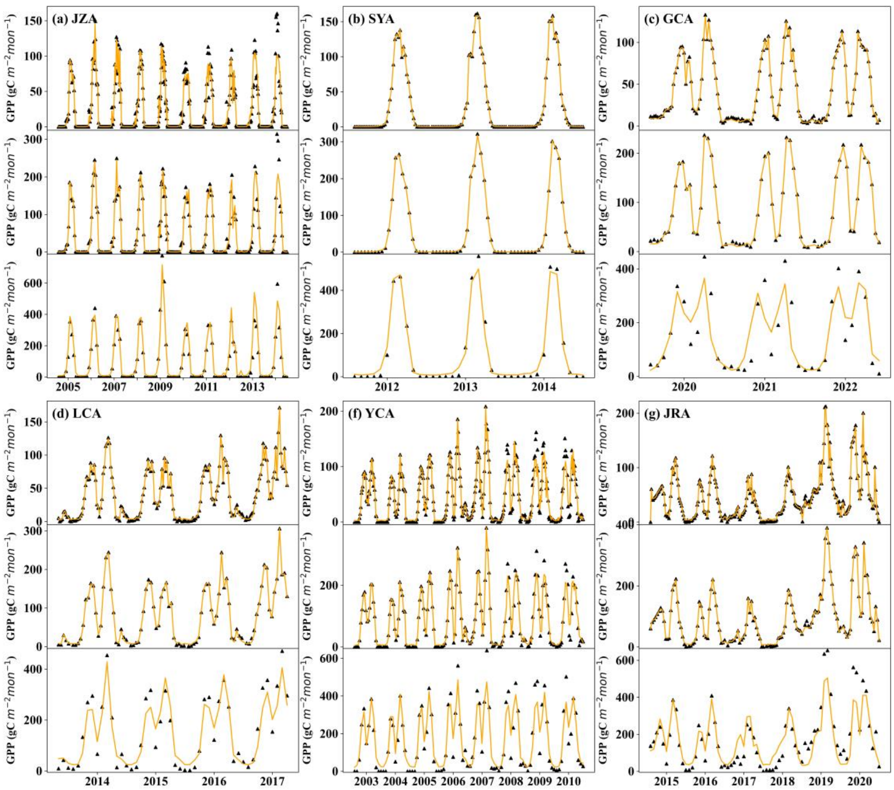

Fig. 12. Plot of simulated GPP time-series variation at GCA, JRA, JZA, LCA, SYA, and YCA sites.

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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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.

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

Geoscientific Model Development
Discussions

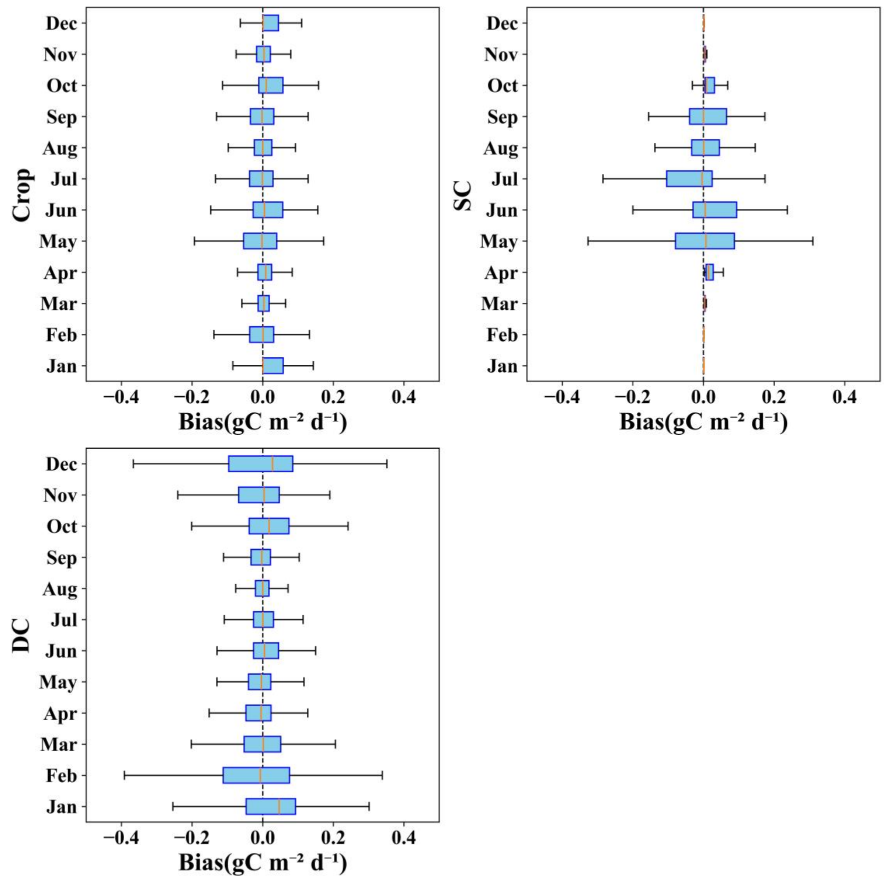

Fig. 13. The monthly bias of FLAML-LUE models among cropland types. SC: single cropping, DC: double cropping.

3.3.3 Analysis of the importance of variables

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

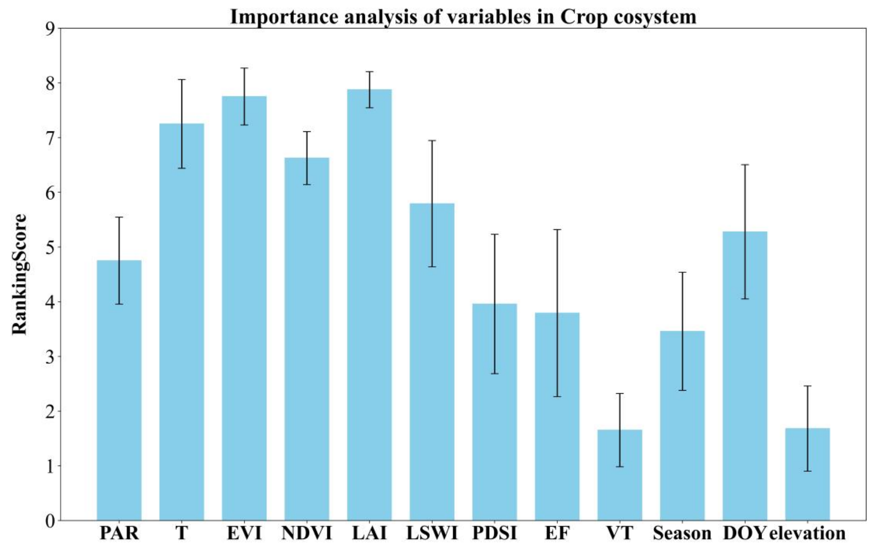

Fig. 14. Average variables importance of farm ecosystem in FALML-LUE models.

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. 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

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 combinations of indicators, while the differences were not significant, and the main differences in prediction accuracy were ascribed to site differences rather than model differences. FLAML-LUE had the best simulation performance for mixed forests in CBF, and planted coniferous forests in QYF, with R^2 of 0.93 and 0.90, respectively; followed by single cropping farmland in the Jinzhou site,

© Author(s) 2025. CC BY 4.0 License.

double cropping farmland in the Yucheng site and typical grassland DLG. Over the alpine meadow and alpine ecosystem, the model performance was poorer, with an R^2 of 0.79; and even worse at the MEF site, with an average R^2 of 0.63. 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 16-day and monthly scales was better than that at 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.

In addition, our results indicate that forest and agricultural fields have greater carbon sequestration capacity and higher annual fluxes than grasslands (Table S5, S10,

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

S14), 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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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.

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

4.3 Comparison with other studies

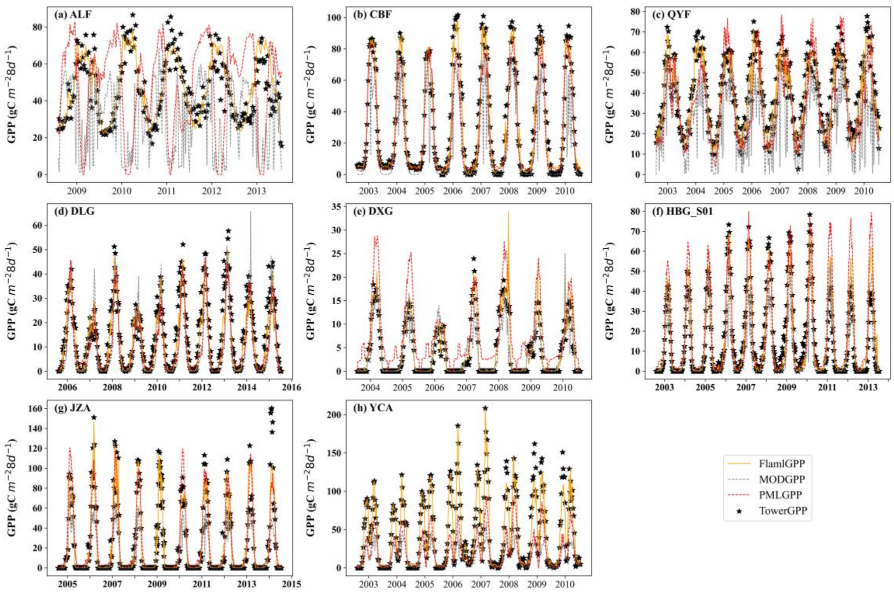

Fig. 15. Comparing 8-day GPP from FLAML-LUE, PML, MOD17 models, and EC observations.

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% of the GPP variation.

Compared to two GPP products (MODIS GPP, 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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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_R^2	MOD_R^2	PML_R^2
	ALF	0.79	0.24	0.33
Forest	CBF	0.98	0.78	0.93
	QYF	0.96	0.54	0.74
	DLG	0.93	0.76	0.77
Grass	DXG	0.89	0.74	0.82
	HBG_S01	0.92	0.81	0.83
Crop	JZA	0.94	0.84	0.85
	YCA	0.96	0.71	0.78

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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.
Geoscientific Model Development Discussions
Open Access

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 can be obtained from ERA5 reanalysis data. Thus, the model can be extended from the site scale to the regional and even global scale.

5. Conclusion

In this study, the FLAML-LUE model was applied to estimate GPP at four different time scales across 20 sites in China. Our findings indicate that the FLAML-LUE model excels at predicting GPP, capturing both its temporal variations and magnitude. It performs particularly well in mixed and evergreen coniferous forests, with mean R^2 values of 0.93 and 0.91, respectively. In addition, extending the time scale of input data can further enhance model accuracy. Specifically, the mean R^2 of forest ecosystems increased from 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 using the importance ranking method showed that vegetation index 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 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 sites but generally underestimated the GPP peaks. These results suggest that the FLAML-LUE model is highly capable of predicting GPP and has significant potential for scaling up GPP from flux footprints to larger areas, enhancing our 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 to improve GPP modeling in these ecosystems in the future.

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 we can find detailed documentation about FLAML at https://microsoft.github.io/FLAML/. The observation data for these sites can be

© Author(s) 2025. CC BY 4.0 License.

downloaded from https://www.scidb.cn/en/ (Science Data Bank). 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

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.

© Author(s) 2025. CC BY 4.0 License.

https://doi.org/10.5194/bg-14-4101-2017

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, 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

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

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.

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

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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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.

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

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

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.

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., 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

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

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

© Author(s) 2025. CC BY 4.0 License.

from AVHRR and MODIS sensors over South Florida. Remote Sens. Environ. 93, 77–86. https://doi.org/10.1016/j.rse.2004.06.020

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., 2023. Assessment of Six Machine Learning Methods for Predicting Gross Primary Productivity in Grassland. Remote Sens. 15, 3475. https://doi.org/10.3390/rs15143475

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, J., Sun, G., Chu, H.-S., Liu, J., McNulty, S.G., Noormets, A., John, R., Ouyang, Z., Zha, T., Li, H., Guan, W., Chen, J., 2014. Long-term variability in the water budget and its controls in an oak-dominated temperate forest. Hydrol. Process. 28, 6054–6066. https://doi.org/10.1002/hyp.10079

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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

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

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

https://doi.org/10.5194/gmd-2024-169
Preprint. Discussion started: 17 January 2025
© Author(s) 2025. CC BY 4.0 License.

Estimates Over India: A Modeling Testbed. J. Adv. Model. Earth Syst. 15, e2022MS003099. https://doi.org/10.1029/2022MS003099