

# 2001-2022 global gross primary productivity dataset using an ensemble model based on random forest

Xin Chen<sup>1</sup>, Tiexi Chen<sup>1,2,3\*</sup>, Xiaodong Li<sup>4</sup>, Yuanfang Chai<sup>5</sup>, Shengjie Zhou<sup>1</sup>, Renjie Guo<sup>6</sup>, Jie Dai<sup>1</sup>

<sup>1</sup>School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China.

<sup>2</sup>Qinghai Provincial Key Laboratory of Plateau Climate Change and Corresponding Ecological and Environmental Effects, Qinghai University of Science and Technology, Xining 810016, China

<sup>3</sup>School of Geographical Sciences, Qinghai Normal University, Xining 810008, Qinghai, China.

<sup>4</sup>Qinghai Institute of Meteorological Science, Xining 810008, Qinghai, China.

<sup>5</sup>Department of Earth Sciences, Vrije Universiteit Amsterdam, Boelelaan 1085, 1081 HV, Amsterdam, the Netherlands

<sup>6</sup>Faculty of Geographical Science, Beijing Normal University, Beijing, China.

*Correspondence to:* Tiexi Chen (txchen@nuist.edu.cn)

**Abstract.** Advancements in remote sensing technology have significantly contributed to the improvement of models for estimating terrestrial gross primary productivity (GPP). However, discrepancies in the spatial distribution and interannual variability within GPP datasets pose challenges to a comprehensive understanding of the terrestrial carbon cycle. In contrast to previous models that rely on remote sensing and environmental variables, we developed an ensemble model based on the random forest (ERF model). This model used GPP outputs from established models (EC-LUE, GPP-kNDVI, GPP-NIRv, Revised-EC-LUE, VPM, MODIS) as inputs to estimate GPP. The ERF model demonstrated superior performance, explaining 85.1% of the monthly GPP variations at 170 sites, surpassing the performance of selected GPP models (67.7%-77.5%) and an independent random forest model using remote sensing and environmental variables (81.5%). Additionally, the ERF model improved accuracy across each month and various subvalues, mitigating the issue of "high value underestimation and low value overestimation" in GPP estimates. Over the period from 2001 to 2022, the global GPP estimated by the ERF model was 132.7 PgC yr<sup>-1</sup>, with an increasing trend of 0.42 PgC yr<sup>-2</sup>, which is comparable to or slightly better than the accuracy of other mainstream GPP datasets in term of validation results of GPP observations independent of FLUXNET (ChinaFlux). Importantly, for the growing number of GPP datasets, our study provides a way to integrate these GPP datasets, which may lead to a more reliable estimate of global GPP.

## 31 **1 Introduction**

32 Gross primary productivity (GPP) is the largest carbon flux in the global carbon cycle, and serves as the primary input of  
33 carbon into the terrestrial carbon cycle. Uncertainties in GPP estimates can propagate to other carbon flux estimates, making  
34 it crucial to clarify the spatio-temporal patterns of GPP (Xiao et al., 2019; Ruehr et al., 2023). However, global GPP is variously  
35 estimated from 90 PgC yr<sup>-1</sup> to 160 PgC yr<sup>-1</sup> across different studies, with these variations becoming more pronounced when  
36 scaled down to regional scales or specific ecosystem types. This variability underscores the necessity for innovative methods  
37 to reduce uncertainty in GPP estimates (Jung et al., 2019; Ryu et al., 2019; Anav et al., 2015).

38 The light use efficiency (LUE) model is one of the most widely adopted methods for estimating GPP. It assumes that GPP is  
39 proportional to the photosynthetically active radiation absorbed by vegetation, and optimizes the spatio-temporal pattern of  
40 GPP through meteorological constraints such as temperature and moisture (Pei et al., 2022). However, variations in these  
41 constraints varies significantly, leading to differences of over 10% in model explanatory power. (Yuan et al., 2014). Recent  
42 studies have proposed some novel vegetation indices that have been shown to be effective proxies for GPP through theoretical  
43 derivation and observed validation (Badgley et al., 2017; Camps-Valls et al., 2021). However, these vegetation indices often  
44 use only remote sensing data as an input for estimating long-term GPP without considering meteorological factors, which has  
45 led to some controversy (Chen et al., 2024; Dechant et al., 2020; Dechant et al., 2022). Both LUE and vegetation index models  
46 use linear mathematical formulas to estimate GPP, but ecosystems are inherently complex, and the biases introduced by these  
47 numerical models increase the uncertainty of GPP estimates. Machine learning models have shown great potential for  
48 improving GPP estimates in previous studies (Jung et al., 2020; Guo et al., 2023). These model are trained by non-physical  
49 means directly using GPP observations and selected environmental and vegetation variables, and the performance of the  
50 models depends on the number and quality of observed data and the representativeness of input data. Nevertheless, direct  
51 validation from flux towers of FLUXNET reveals that these models typically explain only about 70% of monthly GPP  
52 variations, with similar performance to other GPP models (Wang et al., 2021; Badgley et al., 2019; Zheng et al., 2020; Jung et  
53 al., 2020). Due to deviations in the model structure, a common limitation across these models is the poor estimate of monthly  
54 extreme GPP, leading to the phenomenon of "high value overestimation and low value overestimation" (Zheng et al., 2020).  
55 Especially for extremely high values, which usually occur during the growing season and largely determine the annual totals  
56 and interannual fluctuations of GPP, this underestimation may hinder our understanding of the global carbon cycle.

57 It is challenging for a single model to provide accurate estimates for all global regions. Ensemble models have outperformed  
58 individual models in previous studies, potentially addressing some inherent issues in model estimate (Chen et al., 2020; Yao  
59 et al., 2014). Traditional multi-model ensemble methods usually use a simple multi-model average or a weighted Bayesian  
60 average. However, these methods typically assign fixed weights to each model and are essentially linear combinations. Recent  
61 studies have incorporated machine learning techniques to multi-model ensembles to establish nonlinear relationships between  
62 multiple simulated target variables and real target variable, improving simulation performance (Bai et al., 2021; Yao et al.,

63 2017; Tian et al., 2023). Whether this method can improve some common problems with individual GPP models, such as high  
64 value underestimation and low value overestimation, is not clear and needs to further investigation.  
65 In this study, we attempt to use an ensemble model based on the random forest (ERF model) to improve global GPP estimate.  
66 Specifically, the work of this study includes the following: (1) Recalibrating parameters for each model, and comparing the  
67 performance of six GPP models and the ERF model; (2) Focusing on the phenomenon of “high value underestimation and low  
68 value overestimation” in each model, and evaluating the performance of each model across different months, vegetation types  
69 and subvalues (high value, median value, low value); (3) Developing a global GPP dataset using the ERF model and validating  
70 its generalization using GPP observations from ChinaFlux.

## 71 2 Method

### 72 2.1 Data at the global scale

73 In this study, we selected remote sensing data from the Moderate Resolution Imaging Spectroradiometer (MODIS) and  
74 meteorological data from ERA5 to estimate global GPP (Hersbach et al., 2020). For the remote sensing data, surface reflectance  
75 (red band, near infrared band, blue band and shortwave infrared band), leaf area index (LAI) and fraction of photosynthetically  
76 active radiation (FPAR) were used. For meteorological data, we selected average air temperature, dew point temperature,  
77 minimum air temperature, total solar radiation and direct solar radiation. Dew point temperature and air temperature were used  
78 to calculate saturated vapor pressure difference (VPD) (Yuan et al., 2019), and diffuse solar radiation (DifSR) was derived as  
79 the difference between total solar radiation and direct solar radiation. Minimum air temperature was obtained from the hourly  
80 air temperature. CO<sub>2</sub> data were obtained from the monthly average carbon dioxide levels measured by the Mauna Loa  
81 Observatory in Hawaii. Table 1 provides an overview of the datasets used in this study.

82

83 **Table 1.** Overview of the datasets used in this study.

Variable	Dataset	Spatial resolution	Temporal resolution	Temporal coverage
Surface reflectance (red band and near infrared band)	MCD43C4	0.05 °	daily	2001-2022
Surface reflectance (red band, near infrared band, blue band and shortwave infrared band)	MOD09CMG	0.05 °	daily	2001-2022
LAI	MOD15A2H	500m	8d	2001-2022
FPAR	MOD15A2H	500m	8d	2001-2022
Average air temperature (AT)	ERA5-land	0.1 °	Monthly	2001-2022
Dew point temperature (DPT)	ERA5-land	0.1 °	Monthly	2001-2022

Minimum air temperature (MINT)	ERA5-land	0.1 °	Monthly	2001-2022
Total solar radiation (TSR)	ERA5 monthly data on single levels	0.25 °	Monthly	2001-2022
Direct solar radiation (DirSR)	ERA5 monthly data on single levels	0.25 °	Monthly	2001-2022
CO <sub>2</sub>	NOAA's Earth System Research Laboratory Harvested Area and Yield for 175 Crops	/	Monthly	2001-2022
Distribution map of C4 crops	MCD12C1	1/12 °	Annual	2000
Land use		0.05 °	Annual	2010

84

85 Previous studies have shown that the photosynthetic capacity of C4 crops is much higher than that of C3 crops (Chen et al.,  
86 2014; Chen et al., 2011), so it is necessary to divide the cropland into C3 crops and C4 crops. To estimate the global GPP, we  
87 used the "175 Crop harvested Area and yield" dataset, which describes the global harvested area and yield of 175 crops in  
88 2000 (Monfreda et al., 2008). We extracted the sum of the area ratios of all C4 crops (corn, corn feed, sorghum, sorghum feed,  
89 sugarcane, millet) at each grid as the coverage of C4 crops (Figure S1). Consequently, the estimated value of cropland GPP  
90 can be expressed as: coverage of C3 crops × simulated GPP value of C3 crops + coverage of C4 crops × simulated GPP value  
91 of C4 crops, which has been used in a previous study (Guo et al., 2023).

92 The land use map was derived from the IGBP classification of MCD12C1, and 2010 was chosen as the reference year (that is,  
93 land use data is unchanged in the simulation of global GPP). In order to meet the requirements of subsequent research, land  
94 cover types were grouped into 9 categories: Deciduous Broadleaf Forest (DBF), Evergreen Needleleaved Forest (ENF),  
95 Evergreen Broadleaf Forest (EBF), Mixed Forest (MF), Grassland (GRA), Cropland (including CRO-C3 and CRO-C4),  
96 Savannah (SAV), Shrub (SHR), Wetland (WET).

97 Finally, for higher resolution data, we gridded the dataset to 0.05 ° by averaging all pixels whose center fell within each 0.05 °  
98 grid cell for upscaling. For lower resolution data, we used the nearest neighbor resampling method to 0.05 °. In addition,  
99 MODIS data were aggregated to a monthly scale to ensure spatio-temporal consistency.

## 100 2.2 Observation data at the site scale

101 GPP observations were sourced from the FLUXNET 2015 dataset, which includes carbon fluxes and meteorological variables  
102 from more than 200 flux sites around the world (Pastorello et al., 2020). GPP cannot be obtained directly from flux sites and

103 usually needs to be obtained by dismantling the Net Ecosystem Exchange. We chose a monthly level GPP based on the  
 104 nighttime partitioning method and retained only high quality data ( $NEE\_VUT\_REF\_QC > 0.8$ ) for every year, ultimately  
 105 selecting 170 sites with 10932 monthly values for this study. In addition, we selected monthly average air temperature, total  
 106 solar radiation and VPD. The site observations do not provide direct solar radiation, so we extracted data from ERA5 covering  
 107 the flux tower. Monthly minimum air temperature was derived from hourly air temperature. Since some required model data  
 108 are not directly available at flux sites, LAI and FPAR were extracted from MOD15A2H (500 m), and surface reflectance data  
 109 (red band, near infrared band, blue band and shortwave infrared band) were derived from MCD43A4 (500 m) and MOD09A1  
 110 (500 m). These data are roughly similar to the footprint of the flux site and can represent the land surface of the site (Chu et  
 111 al., 2021).

### 112 2.3 GPP estimate model

113 We selected six independent models to estimate GPP in this study. These models are widely used with few model parameters  
 114 and have demonstrated reliable accuracy in previous studies (Zheng et al., 2020; Zhang et al., 2017; Badgley et al., 2017). The  
 115 six models are EC-LUE, Revised-EC-LUE, NIRv-based linear model, kNDVI-based linear model, VPM, MODIS. The VPM,  
 116 MODIS and EC-LUE are LUE models based on remote sensing data and meteorological data (Yuan et al., 2007; Running et  
 117 al., 2004; Xiao et al., 2004). Zheng et al., (2020) proposed the Revised-EC-LUE model, which divides the canopy into sunlit  
 118 and shaded leaves, improving the estimate of global GPP (Zheng et al., 2020). The NIRv and kNDVI are novel vegetation  
 119 indices calculated from the red and near-infrared bands of the reflectance spectrum (Badgley et al., 2017; Camps-Valls et al.,  
 120 2021). Similar to solar induced chlorophyll fluorescence, they exhibit a linear relationship with GPP and are considered  
 121 effective proxies for GPP. Detailed descriptions of all models can be found in Text S1.

122 To reduce uncertainty in GPP estimates from a single model, we used the ERF model, the basic idea of which is to restructure  
 123 the simulated values of multiple models. In this study, we directly used the ERF model to establish the relationship between  
 124 the GPP simulated by the above six models and GPP observations. In addition, for comparison with the ERF model, we also  
 125 used the random forest (RF) method for modeling. In this study, we used average air temperature, minimum air temperature,  
 126 VPD, direct solar radiation, diffuse solar radiation, FPAR and LAI to estimate GPP. Both models used the random forest  
 127 method, which has been widely used in previous studies of GPP estimate (Jung et al., 2020; Guo et al., 2023). Random forest  
 128 is an ensemble learning algorithm that combines the outputs of multiple decision trees to produce a single result, and is  
 129 commonly used for classification and regression problems (Belgiu and Drăguț, 2016). In the regression problem, the output  
 130 result of each decision tree is a continuous value, and the average of all decision tree outputs is taken as the final result. An  
 131 overview of all models used can be found in Table 2.

132 **Table 2.** Overview of the models used in this study.

ID	Model	Input data	Output
1	EC-LUE	FPAR, VPD, AT, SRAD, CO <sub>2</sub>	GPP <sub>EC</sub>

2	Revised-EC-LUE	LAI, VPD, AT, DifSR, DirSR, CO <sub>2</sub>	GPP <sub>REC</sub>
3	kNDVI-GPP	Red band and near infrared band (MCD43)	GPP <sub>kNDVI</sub>
4	NIRv-GPP	Red band and near infrared band (MCD43)	GPP <sub>NIRv</sub>
5	VPM	Red band, near infrared band, blue band, shortwave infrared band (MOD09), AT, SRAD	GPP <sub>VPM</sub>
6	MODIS	FPAR, SRAD, MINT, VPD	GPP <sub>MODIS</sub>
7	Random forest model (RF)	LAI, FPAR, AT, MINT, VPD, DifSR, DirSR	GPP <sub>RF</sub>
8	Ensemble model based on random forest (ERF)	GPP <sub>EC</sub> , GPP <sub>REC</sub> , GPP <sub>kNDVI</sub> , GPP <sub>NIRv</sub> , GPP <sub>MODIS</sub> , GPP <sub>VPM</sub>	GPP <sub>ERF</sub>

133

#### 134 **2.4 Model parameter calibration and validation**

135 FLUXNET only provides GPP observations and meteorological data, lacking direct measurements for LAI, FPAR, and surface  
136 reflectance, so remote sensing data is needed. Considering the variety of remote sensing data sources, such as MODIS and  
137 AVHRR, it is evident that calibrating the same GPP model with different remote sensing data can yield varied parameters. In  
138 addition, the number of sites used to calibrate model parameters is also an important influencing factor for model parameters.  
139 The original parameters of these models were calibrated with only a limited number of sites (e.g., 95 sites for Revised EC-  
140 LUE and 104 for NIRv) (Wang et al., 2021; Zheng et al., 2020). Therefore, to reduce the impact of the uncertainty of model  
141 parameters on simulation results, we did not use original parameters and conducted parameter calibration for GPP models  
142 across different vegetation types. For EC-LUE, Revised EC-LUE, VPM and MODIS, the Markov chain Monte Carlo method  
143 was used to calibrate model parameters. Traditionally, the mean of the posterior distribution of parameters is taken as the  
144 optimal value. However, previous studies have indicated that some model parameters are not well constrained when calibrating  
145 multiple model parameters (Xu et al., 2006; Wang et al., 2017), so we selected the parameter with the smallest root-mean-  
146 square error (RMSE) as the optimal parameter in each iteration. For each vegetation type, we randomly selected 70% of the  
147 sites for parameter calibration, and repeated the process 200 times. In order to avoid overfitting, we adopted the mean of the  
148 200 calibrated parameters as the final model parameters. Similarly, for the two vegetation index models, we randomly selected  
149 70% of the sites in each vegetation type for parameter calibration, repeating the process 200 times and using the mean of the  
150 200 calibrated parameters as the final model parameters.

151 After obtaining GPP estimates from the six GPP models, we evaluated the simulation performance of the RF model and the  
152 ERF model respectively. For both models, we evaluated the model performance using 5-fold cross-validation, where the  
153 process was repeated 200 times, and the mean of the 200 GPP estimates was considered the final GPP estimate. In addition,  
154 we used a second validation method where 70% of the data was selected for modeling and only the remaining 30% was  
155 validated, a process that was repeated 200 times. We utilized the determination coefficient ( $R^2$ ) and RMSE as metrics to

156 evaluate the simulation performance of all models. Additionally, we used the ratio of GPP simulations to GPP observations  
157 (Sim/Obs) to measure whether the model overestimates or underestimates.

## 158 **2.5 Global GPP estimate based on ERF model and its uncertainty.**

159 Based on the ERF model, we estimated global GPP for 2001-2022 (ERF\_GPP). It is important to note that in this process, we  
160 used all the site data to build the model. The uncertainties of ERF\_GPP can be attributed to two primary factors: the influence  
161 of the number of GPP observations and the influence of the number of features (that is, the simulated GPP). For the first type  
162 of uncertainty, we randomly selected 80% of the data to build a model and simulate the multi-year average of global GPP. The  
163 process was repeated 100 times, yielding 100 sets of multi-year averages of ERF\_GPP. Their standard deviations were  
164 considered as the uncertainty of ERF\_GPP caused by the number of GPP observations. For the second type of uncertainty, we  
165 selected different number of features to build a model and simulate the multi-year average of global GPP. A total of 56 sets of  
166 multi-year averages of ERF\_GPP were obtained. The standard deviation of different combinations was considered to be the  
167 uncertainty of ERF\_GPP caused by the number of features.

## 168 **2.6 Evaluation of the generalization of different GPP datasets**

169 The majority of flux sites in FLUXNET are concentrated in Europe and North America, it is unclear whether the different GPP  
170 estimate methods are suitable for regions with sparse flux sites. Recently, ChinaFlux has published GPP observations from  
171 several sites, offering an opportunity to evaluate the generalization of different GPP datasets. However, the spatial resolution  
172 of most GPP datasets is  $0.05^\circ$ , and a direct comparison with GPP observations at flux sites is challenging. Therefore, we  
173 extracted  $0.05^\circ$  MODIS land use covering the flux sites. If the vegetation type of the flux site matched the MODIS land use,  
174 the site was used for the analysis. Finally, a total of 12 flux sites were selected (Figure S2), and Table S1 shows the information  
175 of these sites. The same procedure was applied to FLUXNET, resulting in the selection of 52 sites (Figure S2). It should be  
176 noted that due to the absence of meteorological data from some sites in Chinaflux, we did not validate all GPP models at the  
177 site scale (500 m).

178 We evaluated the generalization of ERF\_GPP at 12 ChinaFlux sites and 52 FLUXNET sites. In addition, we selected a number  
179 of widely used GPP datasets for comparison, including BESS (Li et al., 2023), GOSIF (Li and Xiao, 2019), FLUXCOM:  
180 random forest-based version (FLUXCOM-RF) and ensemble version (FLUXCOM-ENS) (Jung et al., 2020), NIRv (Wang et  
181 al., 2021), Revise-EC-LUE (Zheng et al., 2020), MODIS (Running et al., 2004), VPM (Zhang et al., 2017), which were  
182 generated using different GPP estimate methods. These GPP datasets all have a spatial resolution of  $500\text{ m}-0.5^\circ$ , similar to the  
183 resampling process in section 2.1, we have unified them to  $0.05^\circ$ . The common time range for these datasets spanned from  
184 2001 to 2018, and the temporal resolution was unified to monthly to match the GPP observations.

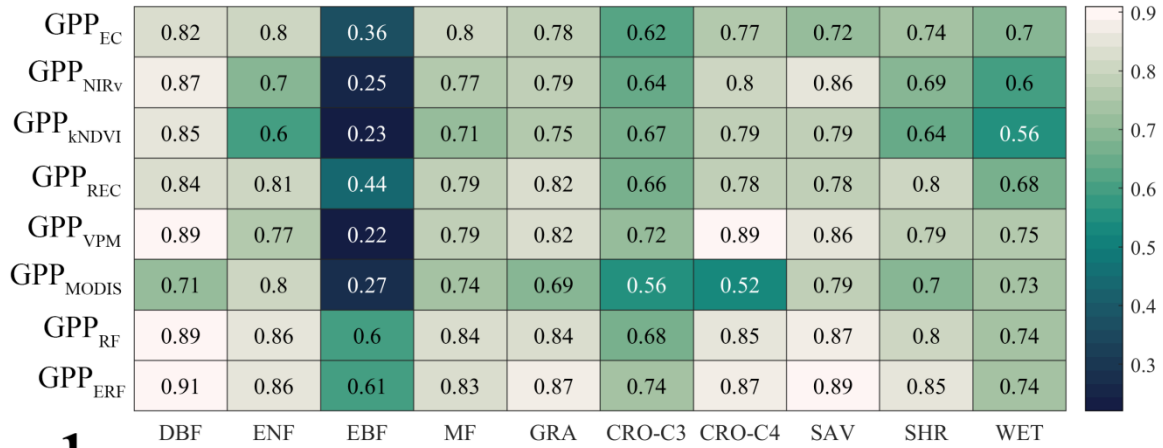
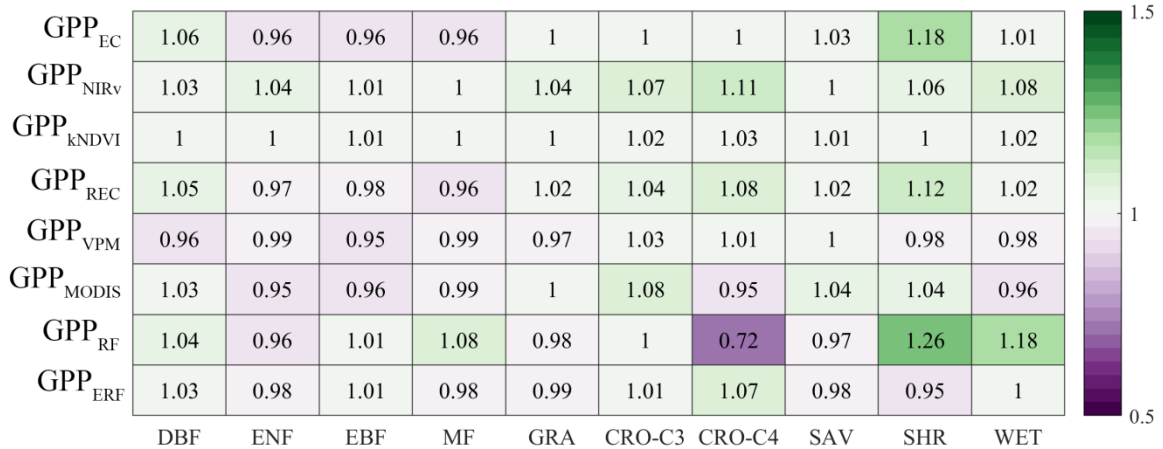
## 185 3 Result

### 186 3.1 Performance of GPP models at site scale

187 Table S2-S7 show the optimization results of the six GPP model parameters. Consistent with previous study, in the Revised  
188 EC-LUE model, the light use efficiency parameter of shade leaves was significantly higher than that of sunlit leaves (Zheng  
189 et al., 2020). It is necessary to divide cropland into C3 crops and C4 crops. In all models, the light use efficiency parameters  
190 of C4 crops were significantly higher than those of C3 crops, which was particularly reflected in the two vegetation index  
191 models of  $GPP_{kNDVI}$  and  $GPP_{NIRv}$ , the slope of the linear regression directly reflected the difference in photosynthetic capacity  
192 of the different crops.

193 Figure 1 shows the performance of all models across different vegetation types. Overall, the performance of the ERF model  
194 was better than that of the other GPP models.  $GPP_{ERF}$  had the higher accuracy among all models, with  $R^2$  between 0.61-0.91  
195 and RMSE between 0.72-2.78  $gC\ m^{-2}\ d^{-1}$ . In contrast, the LUE and vegetation index models performed slightly weaker,  
196 especially in EBF, where  $R^2$  was both below 0.5. It is worth noting that compared to other vegetation types, the RMSE was  
197 highest for cropland, with 6 out of 8 models for C4 crop exceeding 3  $gC\ m^{-2}\ d^{-1}$ , suggesting that these existing GPP models  
198 may not properly capture the seasonal changes in cropland GPP. The six models with calibration parameters and ERF model  
199 were found to have no significant deviation across vegetation types. However,  $GPP_{RF}$  was significantly underestimated for C4  
200 crops and overestimated for SHR.



**a****b****c**

202 **Figure 1.** The performance of the eight models on different vegetation types. a, b and c represent  $R^2$ , RMSE, and Sim/Obs respectively.

203 Combining the results of all flux sites,  $GPP_{ERF}$  explained 85.1% of the monthly GPP variations, while the seven GPP models

204 only explained 67.7%-81.5% of the monthly GPP variations (Figure 2). Another validation method also showed similar results

205 (Figure S3). In order to further prove the robustness of the ERF model, we also used GPP models with original parameters for

206 modeling and validation. As shown in Figure S4, the performance of these GPP models decreased significantly, with  $R^2$  ranging

207 from 0.570 to 0.719 and RMSE ranging from 2.29 to 3.81  $gC\ m^{-2}\ d^{-1}$ . The phenomenon of "high value underestimation and

208 low value overestimation" was also pronounced. However, the ERF model maintained a consistent advantage, with  $R^2$

209 significantly higher than other GPP models (0.856). In addition, we tested the effect of the number of GPP models on the

210 accuracy of the ERF model. As shown in Table S8, as the number of GPP in the ERF model increased, the performance gain

211 of the model gradually decreased.

212 In summary,  $GPP_{ERF}$  showed high accuracy in terms of vegetation type and the ability to interpret monthly variations in GPP,

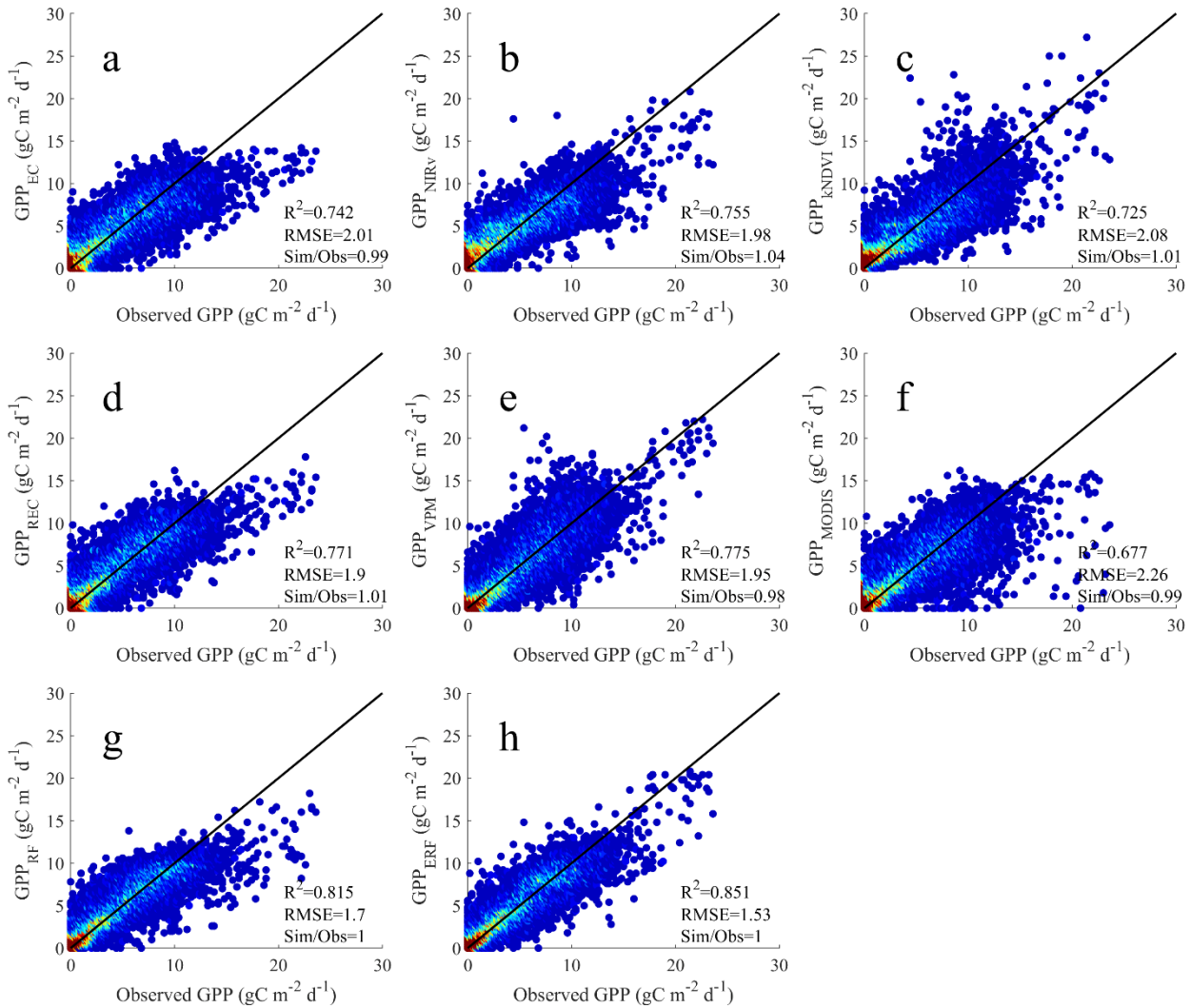
213 which also illustrates the potential of the ERF model to improve GPP estimate. However, it was observed that most GPP

214 simulations exhibited the phenomenon of "high value underestimation and low value overestimation". For example,  $GPP_{EC}$ ,

215  $GPP_{REC}$ ,  $GPP_{MODIS}$  and  $GPP_{RF}$  showed obvious underestimation in the months when the monthly GPP value surpassed 15  $gC$

216  $m^{-2}\ d^{-1}$  (Figure 2). Therefore, it is necessary to evaluate the performance of different models in each month and different

217 subvalues.



218

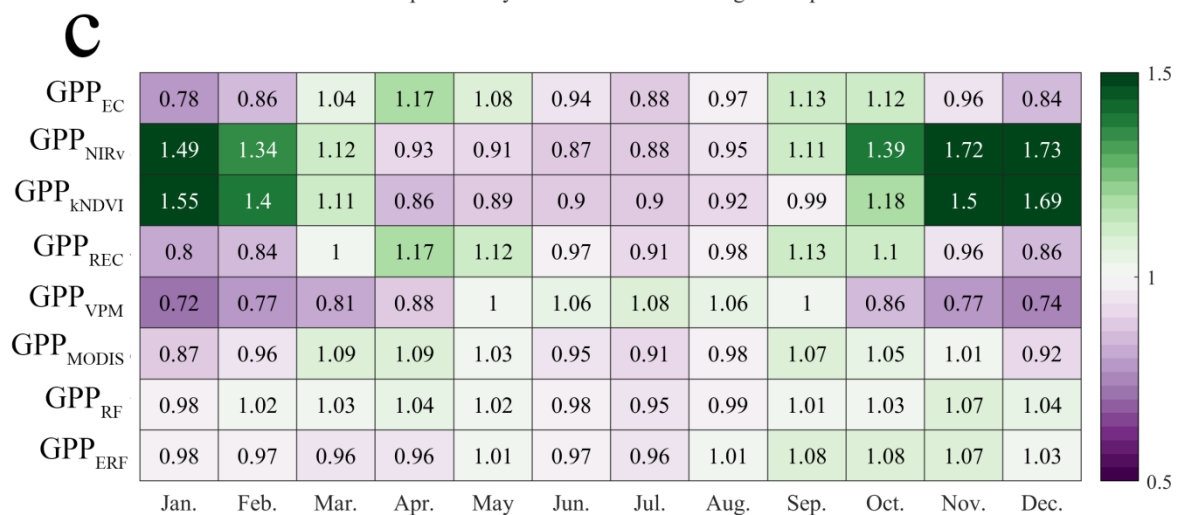
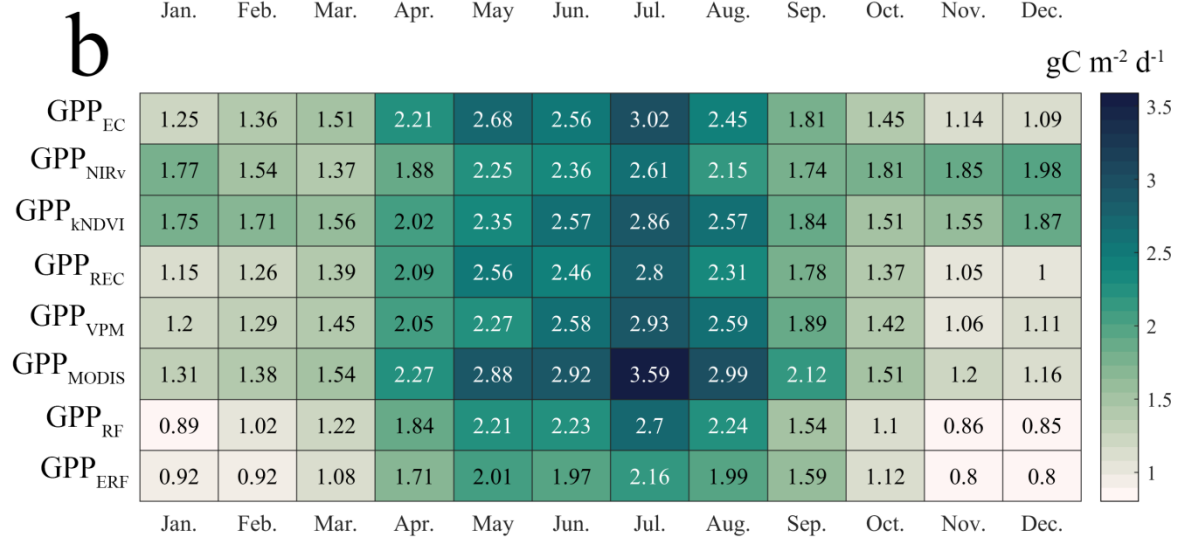
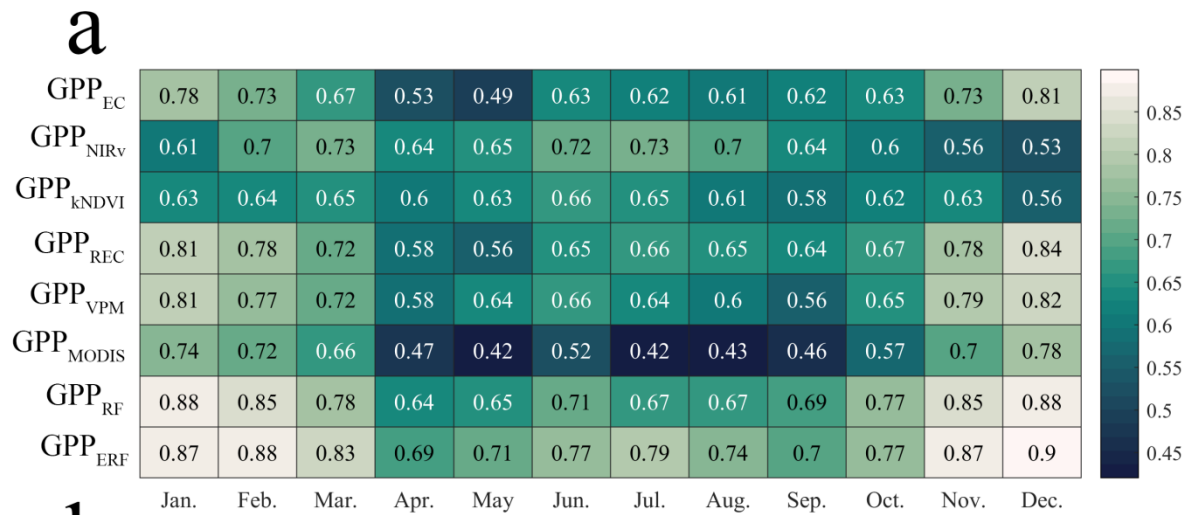
219 **Figure 2.** Comparison between the GPP simulations of the eight models and the GPP observations. a-h represents GPP<sub>EC</sub>, GPP<sub>NIRv</sub>, GPP<sub>KNDVI</sub>,  
 220 GPP<sub>REC</sub>, GPP<sub>VPM</sub>, GPP<sub>MODIS</sub>, GPP<sub>RF</sub>, GPP<sub>ERF</sub>, respectively.

221

### 222 3.2 Performance of GPP models in each month and different subvalues

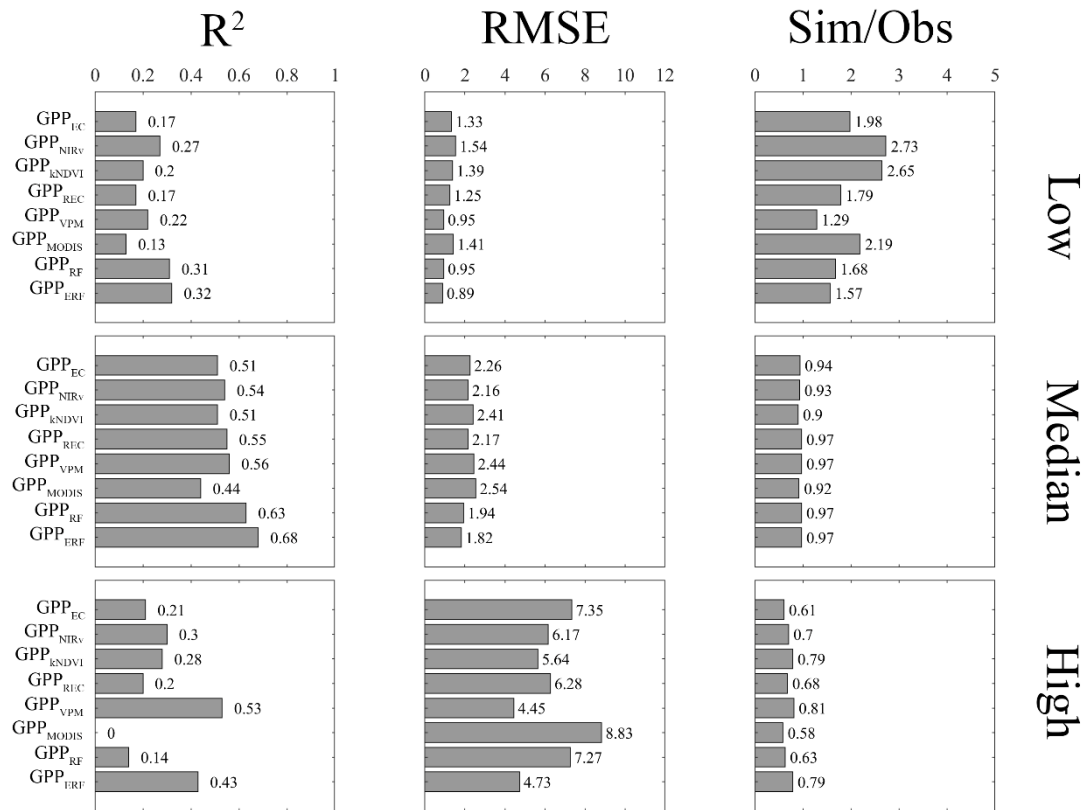
223 Figure 3 shows the simulation accuracy of the eight models in each month. The ERF model maintained a higher accuracy than  
 224 other GPP models, with GPP<sub>ERF</sub> consistently achieving higher  $R^2$  and lower RMSE in most months, and no evident  
 225 phenomena of "high value underestimation and low value overestimation". In contrast, the accuracy of other GPP models  
 226 was less satisfactory accuracy, especially during winter (most flux sites are concentrated in the Northern Hemisphere), the

227 LUE models tended to underestimate GPP, and the Sim/Obs remained at 0.72-1.01, although  $R^2$  were above 0.7. Meanwhile,  
228 the vegetation index models overestimated GPP, Sim/Obs remained at 1.34-1.73, and  $R^2$  were relatively low, mostly around  
229 0.6.



231 **Figure 3.** Performance of the eight models in each month. a, b and c represent  $R^2$ , RMSE, and Sim/Obs respectively.

232 We compared the performance of all models in different subvalues, including high value ( $GPP > 15 \text{ gC m}^{-2} \text{ d}^{-1}$ ), median value  
 233 ( $15 \text{ gC m}^{-2} \text{ d}^{-1} > GPP > 2 \text{ gC m}^{-2} \text{ d}^{-1}$ ), low value ( $GPP < 2 \text{ gC m}^{-2} \text{ d}^{-1}$ ). For extreme values, most models performed poorly  
 234 (Figure 4), with  $R^2$  for GPP models falling below 0.3, and only  $GPP_{VPM}$  showing better performance in the high value.  $GPP_{ERF}$   
 235 demonstrated some improvement in both low and high values, with  $R^2$  0.32 and 0.43, RMSE of 0.89 and 4.73  $\text{gC m}^{-2} \text{ d}^{-1}$ , and  
 236 Sim/Obs closer to 1, respectively. In the median value, all models performed better, with no significant bias in the GPP estimate.  
 237 The  $R^2$  of GPP models ranged from 0.44 to 0.68, and the RMSE remained between 1.82 and 2.54  $\text{gC m}^{-2} \text{ d}^{-1}$ . Further analysis  
 238 was made at two typical sites, it was obvious that  $GPP_{EC}$ ,  $GPP_{REC}$  and  $GPP_{MODIS}$  on CN-Qia exhibited obvious underestimation  
 239 during the growing season (Figure S5). On CH\_Lae,  $GPP_{kNDVI}$  and  $GPP_{VPM}$  were significantly overestimated (Figure S6). In  
 240 contrast, at both sites,  $GPP_{ERF}$  was more consistent with observations, indicating that the superior performance of  $GPP_{ERF}$  was  
 241 due to the corrections on the time series.



242

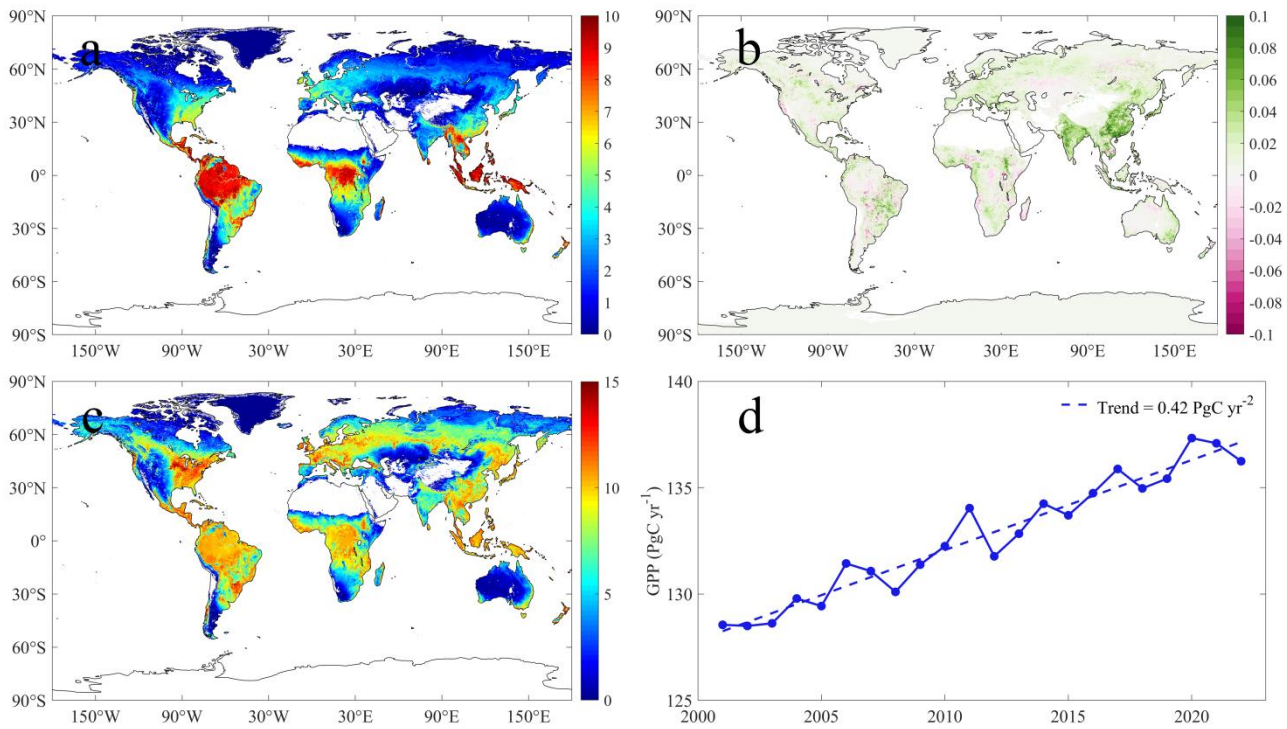
243 **Figure 4.** Performance of eight models in different subvalues.

### 244 3.3 Temporal and spatial characteristics of ERF\_GPP and its generalization evaluation

245 Figure 5a shows the spatial distribution of the multi-year average of ERF\_GPP. The high values of GPP were mainly  
246 concentrated in tropical areas, exceeding  $10 \text{ gC m}^{-2} \text{ d}^{-1}$ , and relatively high in southeastern North America, Europe and southern  
247 China, about  $4\text{-}6 \text{ gC m}^{-2} \text{ d}^{-1}$ . From 2001-2022, China and India showed the fastest increase in GPP, mostly at  $0.1 \text{ gC m}^{-2} \text{ d}^{-1}$   
248 (Figure 5b), similar to a previous study that reported that China and India led the global greening (Chen et al., 2019). We  
249 further investigated the annual maximum GPP, as shown in Figure 5c, and the North American corn belt was the global leader  
250 in GPP at more than  $15 \text{ gC m}^{-2} \text{ d}^{-1}$ , compared to only  $10 \text{ gC m}^{-2} \text{ d}^{-1}$  in most tropical forests. In 2001-2022, the global GPP was  
251  $132.7 \pm 2.8 \text{ PgC yr}^{-1}$ , with an increasing trend of  $0.42 \text{ PgC yr}^{-2}$  (Figure 5d). The lowest value was  $128.6 \text{ PgC yr}^{-1}$  in 2001, and  
252 the highest value was  $136.2 \text{ PgC yr}^{-1}$  in 2020.

253 The results of the two uncertainty analyses consistently indicated that ERF\_GPP exhibited higher uncertainty in tropical  
254 regions (Figures S7 and S8), and the uncertainty of ERF\_GPP caused by the number of GPP observations was relatively small,  
255 the standard deviation of 100 simulations was about  $0.3 \text{ gC m}^{-2} \text{ d}^{-1}$  in the tropics and lower in other regions, below  $0.1 \text{ gC m}^{-2}$   
256  $\text{d}^{-1}$ . In contrast, the uncertainty of ERF\_GPP caused by the number of features was more pronounced, especially when fewer  
257 features were included in the models. It is worth noting that when the number of features was five, the uncertainty was already  
258 substantially less, and the standard deviation was generally lower than  $0.5 \text{ gC m}^{-2} \text{ d}^{-1}$ .

259



260

261 **Figure 5.** Spatial distribution and interannual changes of ERF\_GPP during 2001-2022. a represents the multi-year average, b represents the  
262 trend, c represents the annual maximum, and d represents the interannual change of GPP.

263

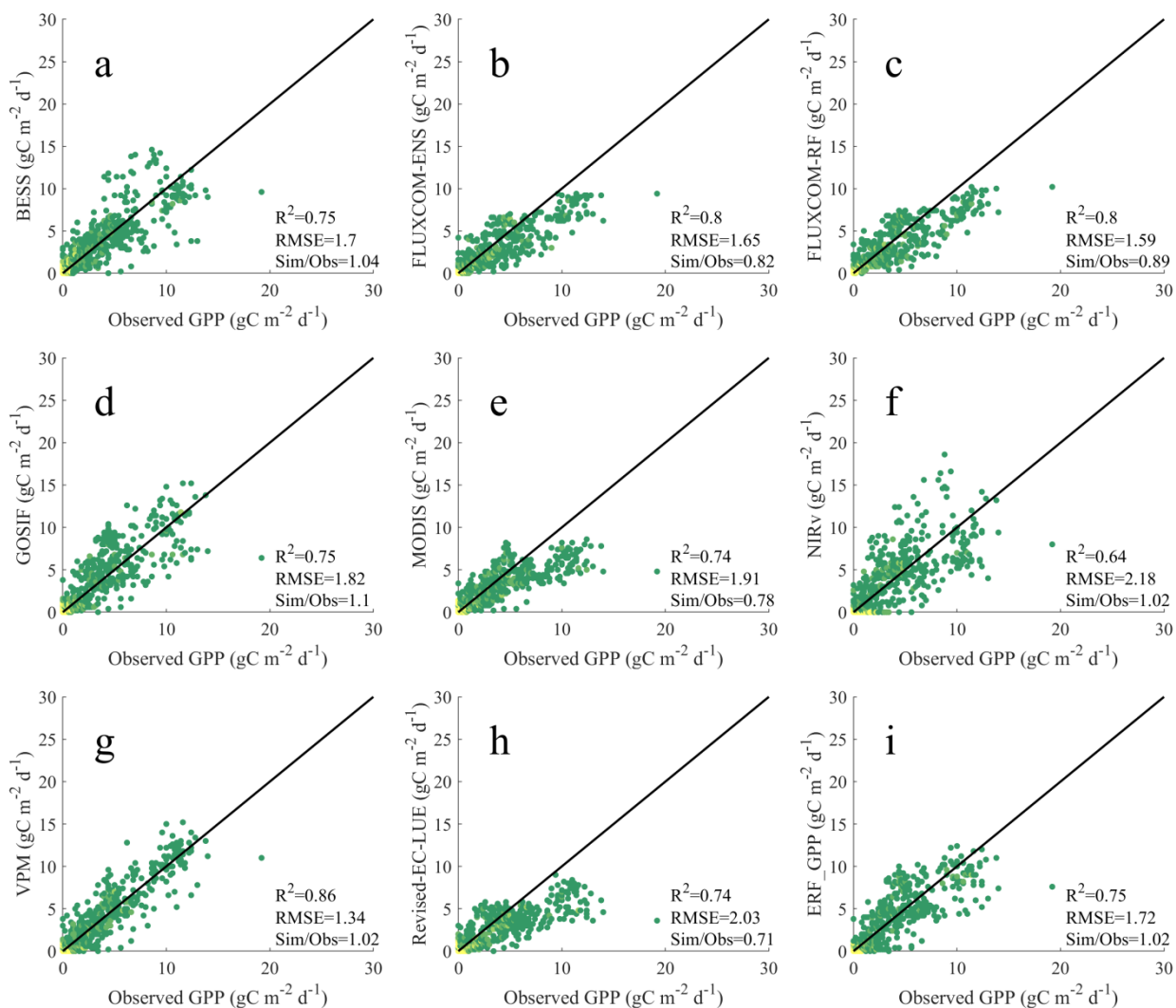
264 As shown in Figure 6, ERF\_GPP and other GPP datasets were validated using GPP observations from ChinaFlux. Among all  
265 models,  $GPP_{VPM}$  demonstrated the best performance, with  $R^2$  of 0.86 and RMSE of  $1.34 \text{ gC m}^{-2} \text{ d}^{-1}$ . ERF\_GPP also exhibited  
266 high generalization, with  $R^2$  of 0.75, RMSE of  $1.72 \text{ gC m}^{-2} \text{ d}^{-1}$ , there was no “high value underestimation and low value  
267 overestimation”, which was comparable to the accuracy of BESS and GOSIF. However, the simulation accuracy of the other  
268 GPP datasets in Chinaflux was relatively poor, with the  $R^2$  of NIRv being only 0.64, while FLUXCOM-ENS, FLUXCOM-  
269 RF, MODIS and Revised EC-LUE were significantly underestimated, with the Sim/Obs being only 0.71-0.89. In the validation  
270 of FLUXNET, the  $R^2$  of FLUXCOM-ENS, MODIS, and Revised EC-LUE ranged from 0.57 to 0.67, and the RMSE ranged  
271 from 2.67 to  $3.3 \text{ gC m}^{-2} \text{ d}^{-1}$ , and exhibited different degrees of underestimation (Figure S9). Other GPP datasets demonstrated  
272 similar performance, with ERF\_GPP being the best ( $R^2 = 0.74$ ,  $RMSE = 2.26 \text{ gC m}^{-2} \text{ d}^{-1}$ ).

273

274

275





276

277 **Figure 6.** Comparison between the GPP datasets and the GPP observations from ChinaFlux. a-i represents BESS, FLUXCOM-ENS,  
 278 FLUXCOM-RF, GOSIF, MODIS, NIRv, VPM, Revise-EC-LUE, ERF\_GPP, respectively.

## 279 4 Discussion

### 280 4.1 Performance analysis of different models

281 After parameter calibration, both LUE and vegetation index models obtained reliable model accuracy. However, noticeable  
 282 errors persist in different months and subvalues, indicating the prevalent phenomenon of "high value underestimation and low  
 283 value overestimation" (Figures 1-4). In addition to MODIS, the GPP simulated by the other three LUE models is generally  
 284 underestimated in winter (Figure 3), which may be caused by biases in the parameters used in meteorological constraints. In

285 the expression form of the temperature constraint adopted by LUE models, the maximum temperature, minimum temperature  
286 and optimum temperature for limiting photosynthesis are all constants, however these values may not be fixed (Huang et al.,  
287 2019; Grossiord et al., 2020). A previous study has demonstrated that the GPP estimate could be effectively improved by using  
288 dynamic temperature parameters (Chang et al., 2021). Moreover, the form of meteorological constraint is also an important  
289 influencing factor. Compared with other LUE models, VPM does not use VPD constraints, but incorporates land surface water  
290 index from satellite observations as constraints (Xiao et al., 2004), which may be the reason why the model performs better  
291 than other models at high value (Figure 4). Conversely, the two vegetation index models overestimated GPP in winter, and  
292 even overestimated by 70% in December. The vegetation index model does not consider meteorological constraints that believe  
293 that all environmental impacts on vegetation have been included in the vegetation index (kNDVI, NIRv). However, it is a fact  
294 that under high temperature or low radiation, the vegetation index may still maintain the appearance of high photosynthesis  
295 (greening), while in fact the GPP is low (Doughty et al., 2021; Yang et al., 2018; Chen et al., 2024). Furthermore, the  
296 relationship between these vegetation indices and GPP is not robust, and the vegetation indices based on reflectance may have  
297 hysteresis (Wang et al., 2022).

298 Compared to other GPP models, the ERF model demonstrated better performance ( $R^2 = 851$ ). Since there are no physical  
299 constraints, the machine learning model needs to find the relationship between explanatory variables and target variable from  
300 a large amount of training data (such as  $GPP=f(LAI,T,P, \text{etc.})$ ). Therefore, the reliability of the model usually depends on the  
301 representativeness of the training data. For example, LAI can explain GPP to a large extent, while complex modeling  
302 relationships are still needed from LAI to GPP. The difference between the ERF model and the RF model lies in the explanatory  
303 variables. The ERF model uses multiple GPP simulations that are more representative and aligned with the target variable,  
304 thus making the GPP simulations more accurate. In other words, the ERF model does not need to take into account the  
305 uncertainties of the model structure (such as meteorological constraints) and model parameters (such as maximum light use  
306 efficiency), but rather focuses on the uncertainties inherent in the simulated GPP. To further clarify the impact of explanatory  
307 variables on the ERF model, we conducted a feature importance analysis (Figure S10). From an average of 200 times, the  
308 results of the ERF model did not depend on a single GPP simulation. Even  $GPP_{MODIS}$ , with the highest relative importance,  
309 accounted for no more than 25%, suggesting that the ERF model behaves more like a weighted average of multiple GPP  
310 simulations. In addition, it is important to emphasize that the accuracy of the ERF model is still robust even for GPP simulations  
311 of original parameters (Figure S4), which means that we can try to use this method to integrate the currently published GPP  
312 data sets to obtain a more accurate global GPP estimate.

313 It is worth noting that in the study of Tian et al. (2023), the ERF model was also used to improve the GPP estimate. Our study  
314 extends this work in several ways. Firstly, parameter calibration was carried out in our study so that the final validation results  
315 are comparable, that is, differences in model performance are mainly due to the uncertainty of the model structure. Secondly,  
316 our study focused on the phenomenon of "high value underestimation and low value overestimation" of GPP model, with  
317 results indicating that the ERF model performed well across various vegetation types, months, and subvalues. Finally, we

318 generated the ERF\_GPP dataset and validated it on different observational datasets, further confirming the robustness of the  
319 ERF model in GPP estimate.

## 320 **4.2 Robustness of ERF\_GPP**

321 Since the current GPP datasets are generated based on remote sensing and FLUXNET GPP observations, there is a strong  
322 similarity in spatial distribution among all GPP datasets. Therefore, the validation of GPP observations independent of  
323 FLUXNET is crucial. Validation results from GPP observations of ChinaFlux indicated that ERF\_GPP exhibited good  
324 generalization in China ( $R^2=0.75$ ), which was slightly lower than the accuracy of 5-fold-cross-validation during modeling,  
325 possibly due to the mismatch between the  $0.05^\circ$  GPP estimate and the footprint of the flux tower (Chu et al., 2021). In addition,  
326 the validation of FLUXNET further confirms the reliability of ERF\_GPP. Overall, this is comparable to or slightly better than  
327 the simulation accuracy of current mainstream GPP datasets. We also observed a clear improvement in the spatial maximum  
328 value of ERF\_GPP in some corn growing regions, such as the North American Corn Belt (Figure 5c), which is supported by  
329 previous studies showing that C4 crops have much higher GPP peaks than other vegetation types (Yuan et al., 2015; Chen et  
330 al., 2011).

331 Due to the increasing trend of drought, the constraining effect of water on vegetation is gradually increasing, and some studies  
332 have reported the decoupling phenomenon of LAI and GPP under some specific conditions (Jiao et al., 2021; Hu et al., 2022).  
333 However, in China and India with significant greening, GPP continues to increase in most datasets, and ERF\_GPP supports this  
334 view. This phenomenon may be attributed to the low drought pressure on croplands in China and India due to irrigation, which  
335 poses less constraint on GPP (Ambika and Mishra, 2020; Ai et al., 2020). The global estimate of ERF\_GPP is  $132.7 \pm 2.8$  PgC  
336  $\text{yr}^{-1}$ , which is close to estimates from most previous studies (Wang et al., 2021; Badgley et al., 2019). A study have suggested  
337 that global GPP may reach 150-175 PgC  $\text{yr}^{-1}$  (Welp et al., 2011), however, there is no further evidence to support this view.  
338 ERF\_GPP exhibited higher uncertainty in tropical regions, similar reports have been made in previously published GPP  
339 datasets (Badgley et al., 2019; Guo et al., 2023). The scarcity of flux observations in these regions (Pastorello et al., 2020),  
340 coupled with the well-known issue of cloud pollution and saturation in remote sensing data in the tropics (Badgley et al., 2019),  
341 exacerbates the uncertainty in GPP estimates for these regions. Therefore, in future studies, on the one hand, more flux  
342 observations in tropical regions are needed, and on the other hand, attempts can be made to combine optical and microwave  
343 data to improve GPP estimate.

## 344 **4.3 Limitations and uncertainties**

345 In this study, we improved GPP estimate based on the ERF model. Nonetheless, there are still some limitations and  
346 uncertainties due to the availability of data and methods. First, C4 crop distribution maps were used in our study to improve  
347 estimates of cropland GPP. However, it is important to note that this dataset only represents the spatial distribution of crops  
348 around the year 2000, which introduce uncertainty into GPP simulations of cropland in a few C3 and C4 alternating areas.  
349 Secondly, the ERF model considers six GPP simulations, and it is not clear whether adding more GPP simulations to the model

350 can further improve the GPP estimate. Finally, our model did not consider the effect of soil moisture on GPP, and some  
351 previous studies have highlighted the importance of incorporating soil moisture in GPP estimates, especially for dry years  
352 (Stocker et al., 2019; Stocker et al., 2018).

## 353 **5 Conclusion**

354 In this study, we compared the performance of the ERF model with other GPP models at the site scale, especially for the  
355 phenomenon of "high value underestimation and low value overestimation", and further developed the ERF\_GPP dataset.  
356 Overall,  $GPP_{ERF}$  had higher model accuracy, explaining 85.1% of the monthly GPP variations, and demonstrated reliable  
357 accuracy in different months, vegetation types and subvalues. Over the period from 2001 to 2022, the global estimate of  
358 ERF\_GPP was  $132.7 \pm 2.8$  PgC yr<sup>-1</sup>, corresponding to an increasing trend of 0.42 PgC yr<sup>-2</sup>. Validation results from ChinaFlux  
359 indicated that ERF\_GPP had good generalization. For the current emerging GPP estimate models, the ERF model provides an  
360 alternative method that lead to better model accuracy.

## 361 **Data and code availability**

362 The ERF\_GPP for 2001-2022 is available at <https://doi.org/10.6084/m9.figshare.24417649> (Chen et al., 2023). The spatial  
363 resolution of ERF\_GPP is 0.05° and the temporal resolution is monthly. Code is available from the author upon reasonable  
364 request.

## 365 **Author contributions**

366 X.C. and T.X.C. conceived the scientific ideas and designed this research framework. X.C. compiled the data, conducted  
367 analysis, prepared figures. X.C., T.X.C. and Y.F.C. wrote the manuscript. D.X.L., R.J.G., J.D., and S.J.Z. gave constructive  
368 suggestions for improving the manuscript.

## 369 **Acknowledgments**

370 This study was supported by the National Natural Science Foundation of China (No. 42130506, 42161144003 and 31570464)  
371 and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX23\_1322).

## 372 **Declaration of interests**

373 The authors have not disclosed any competing interests.

- 375 Ai, Z., Wang, Q., Yang, Y., Manevski, K., Yi, S., and Zhao, X.: Variation of gross primary production, evapotranspiration and  
376 water use efficiency for global croplands, *Agricultural and Forest Meteorology*, 287, 10.1016/j.agrformet.2020.107935, 2020.
- 377 Ambika, A. K. and Mishra, V.: Substantial decline in atmospheric aridity due to irrigation in India, *Environmental Research*  
378 *Letters*, 15, 10.1088/1748-9326/abc8bc, 2020.
- 379 Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N. C., Peylin,  
380 P., Piao, S., Sitch, S., Viovy, N., Wiltshire, A., and Zhao, M.: Spatiotemporal patterns of terrestrial gross primary production:  
381 A review, *Reviews of Geophysics*, 53, 785-818, 10.1002/2015rg000483, 2015.
- 382 Badgley, G., Field, C. B., and Berry, J. A.: Canopy near-infrared reflectance and terrestrial photosynthesis, *Science advances*,  
383 3, e1602244, 2017.
- 384 Badgley, G., Anderegg, L. D., Berry, J. A., and Field, C. B.: Terrestrial gross primary production: Using NIRV to scale from  
385 site to globe, *Global change biology*, 25, 3731-3740, 2019.
- 386 Bai, Y., Zhang, S., Bhattarai, N., Mallick, K., Liu, Q., Tang, L., Im, J., Guo, L., and Zhang, J.: On the use of machine learning  
387 based ensemble approaches to improve evapotranspiration estimates from croplands across a wide environmental gradient,  
388 *Agricultural and Forest Meteorology*, 298, 108308, 2021.
- 389 Belgiu, M. and Drăguț, L.: Random forest in remote sensing: A review of applications and future directions, *ISPRS journal of*  
390 *photogrammetry and remote sensing*, 114, 24-31, 2016.
- 391 Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D.,  
392 Muñoz-Marí J., García-Haro, F. J., and Guanter, L.: A unified vegetation index for quantifying the terrestrial biosphere,  
393 *Science Advances*, 7, eabc7447, 2021.
- 394 Chang, Q., Xiao, X. M., Doughty, R., Wu, X. C., Jiao, W. Z., and Qin, Y. W.: Assessing variability of optimum air temperature  
395 for photosynthesis across site-years, sites and biomes and their effects on photosynthesis estimation, *Agricultural and Forest*  
396 *Meteorology*, 298, 10.1016/j.agrformet.2020.108277, 2021.
- 397 Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R. K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R., Tommervik,  
398 H., Bala, G., Zhu, Z., Nemani, R. R., and Myneni, R. B.: China and India lead in greening of the world through land-use  
399 management, *Nature Sustainability*, 2, 122-129, 10.1038/s41893-019-0220-7, 2019.
- 400 Chen, X., Chen, T., Li, X., Chai, Y., Zhou, S., Guo, R., Dai, J.: 2001-2022 global gross primary productivity dataset using an  
401 ensemble model based on random forest. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.24417649.v2>, 2023
- 402 Chen, T., van der Werf, G. R., Dolman, A. J., and Groenendijk, M.: Evaluation of cropland maximum light use efficiency  
403 using eddy flux measurements in North America and Europe, *Geophysical Research Letters*, 38, 10.1029/2011gl047533, 2011.
- 404 Chen, T., Van Der Werf, G., Gobron, N., Moors, E., and Dolman, A.: Global cropland monthly gross primary production in  
405 the year 2000, *Biogeosciences*, 11, 3871-3880, 2014.
- 406 Chen, X., Chen, T., Liu, S., Chai, Y., Guo, R., Dai, J., Wang, S., Zhang, L., and Wei, X.: Vegetation Index-Based Models  
407 Without Meteorological Constraints Underestimate the Impact of Drought on Gross Primary Productivity, *Journal of*  
408 *Geophysical Research: Biogeosciences*, 129, e2023JG007499, 2024.
- 409 Chen, Y., Yuan, H., Yang, Y., and Sun, R.: Sub-daily soil moisture estimate using dynamic Bayesian model averaging, *Journal*  
410 *of Hydrology*, 590, 125445, 2020.
- 411 Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., Metzger, S., Kumar, J., and Arain, M. A.:  
412 Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites, *Agricultural and Forest*  
413 *Meteorology*, 301, 108350, 2021.
- 414 Dechant, B., Ryu, Y., Badgley, G., Köhler, P., Rascher, U., Migliavacca, M., Zhang, Y., Tagliabue, G., Guan, K., and Rossini,  
415 M.: NIRVP: A robust structural proxy for sun-induced chlorophyll fluorescence and photosynthesis across scales, *Remote*  
416 *Sensing of Environment*, 268, 112763, 2022.
- 417 Dechant, B., Ryu, Y., Badgley, G., Zeng, Y., Berry, J. A., Zhang, Y., Goulas, Y., Li, Z., Zhang, Q., and Kang, M.: Canopy  
418 structure explains the relationship between photosynthesis and sun-induced chlorophyll fluorescence in crops, *Remote Sensing*  
419 *of Environment*, 241, 111733, 2020.
- 420 Doughty, R., Xiao, X. M., Qin, Y. W., Wu, X. C., Zhang, Y., and Moore, B.: Small anomalies in dry-season greenness and  
421 chlorophyll fluorescence for Amazon moist tropical forests during El Niño and La Niña, *Remote Sensing of Environment*, 253,  
422 10.1016/j.rse.2020.112196, 2021.

423 Grossiord, C., Buckley, T. N., Cernusak, L. A., Novick, K. A., Poulter, B., Siegwolf, R. T., Sperry, J. S., and McDowell, N.  
424 G.: Plant responses to rising vapor pressure deficit, *New Phytologist*, 226, 1550-1566, 2020.

425 Guo, R., Chen, T., Chen, X., Yuan, W., Liu, S., He, B., Li, L., Wang, S., Hu, T., Yan, Q., Wei, X., and Dai, J.: Estimating  
426 Global GPP From the Plant Functional Type Perspective Using a Machine Learning Approach, *Journal of Geophysical*  
427 *Research-Biogeosciences*, 128, 10.1029/2022jg007100, 2023.

428 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., and  
429 Schepers, D.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999-2049, 2020.

430 Hu, Z., Piao, S., Knapp, A. K., Wang, X., Peng, S., Yuan, W., Running, S., Mao, J., Shi, X., and Ciais, P.: Decoupling of  
431 greenness and gross primary productivity as aridity decreases, *Remote Sensing of Environment*, 279, 113120, 2022.

432 Huang, M., Piao, S., Ciais, P., Peñuelas, J., Wang, X., Keenan, T. F., Peng, S., Berry, J. A., Wang, K., and Mao, J.: Air  
433 temperature optima of vegetation productivity across global biomes, *Nature ecology & evolution*, 3, 772-779, 2019.

434 Jiao, W., Wang, L., Smith, W. K., Chang, Q., Wang, H., and D'Odorico, P.: Observed increasing water constraint on vegetation  
435 growth over the last three decades, *Nature Communications*, 12, 10.1038/s41467-021-24016-9, 2021.

436 Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein,  
437 M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, *Scientific data*, 6, 1-14, 2019.

438 Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P.,  
439 and Carvalhais, N.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM  
440 approach, *Biogeosciences*, 17, 1343-1365, 2020.

441 Li, B., Ryu, Y., Jiang, C., Dechant, B., Liu, J., Yan, Y., and Li, X.: BESSv2.0: A satellite-based and coupled-process model  
442 for quantifying long-term global land-atmosphere fluxes, *Remote Sensing of Environment*, 295, 10.1016/j.rse.2023.113696,  
443 2023.

444 Li, X. and Xiao, J.: A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS,  
445 and Reanalysis Data, *Remote Sensing*, 11, 10.3390/rs11050517, 2019.

446 Monfreda, C., Ramankutty, N., and Foley, J. A.: Farming the planet: 2. Geographic distribution of crop areas, yields,  
447 physiological types, and net primary production in the year 2000, *Global Biogeochemical Cycles*, 22, 10.1029/2007gb002947,  
448 2008.

449 Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., and  
450 Humphrey, M.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, *Scientific data*,  
451 7, 1-27, 2020.

452 Pei, Y., Dong, J., Zhang, Y., Yuan, W., Doughty, R., Yang, J., Zhou, D., Zhang, L., and Xiao, X.: Evolution of light use  
453 efficiency models: Improvement, uncertainties, and implications, *Agricultural and Forest Meteorology*, 317, 108905, 2022.

454 Ruehr, S., Keenan, T. F., Williams, C., Zhou, Y., Lu, X., Bastos, A., Canadell, J. G., Prentice, I. C., Sitch, S., and Terrer, C.:  
455 Evidence and attribution of the enhanced land carbon sink, *Nature Reviews Earth & Environment*, 4, 518-534,  
456 10.1038/s43017-023-00456-3, 2023.

457 Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H.: A continuous satellite-derived  
458 measure of global terrestrial primary production, *Bioscience*, 54, 547-560, 2004.

459 Ryu, Y., Berry, J. A., and Baldocchi, D. D.: What is global photosynthesis? History, uncertainties and opportunities, *Remote*  
460 *sensing of environment*, 223, 95-114, 2019.

461 Stocker, B. D., Zscheischler, J., Keenan, T. F., Prentice, I. C., Penuelas, J., and Seneviratne, S. I.: Quantifying soil moisture  
462 impacts on light use efficiency across biomes, *New Phytologist*, 218, 1430-1449, 10.1111/nph.15123, 2018.

463 Stocker, B. D., Zscheischler, J., Keenan, T. F., Prentice, I. C., Seneviratne, S. I., and Penuelas, J.: Drought impacts on terrestrial  
464 primary production underestimated by satellite monitoring, *Nature Geoscience*, 12, 264-+, 10.1038/s41561-019-0318-6, 2019.

465 Tian, Z., Yi, C., Fu, Y., Kutter, E., Krakauer, N. Y., Fang, W., Zhang, Q., and Luo, H.: Fusion of Multiple Models for  
466 Improving Gross Primary Production Estimation With Eddy Covariance Data Based on Machine Learning, *Journal of*  
467 *Geophysical Research: Biogeosciences*, 128, e2022JG007122, <https://doi.org/10.1029/2022JG007122>, 2023.

468 Wang, J., Dong, J., Yi, Y., Lu, G., Oyler, J., Smith, W., Zhao, M., Liu, J., and Running, S.: Decreasing net primary production  
469 due to drought and slight decreases in solar radiation in China from 2000 to 2012, *Journal of Geophysical Research:*  
470 *Biogeosciences*, 122, 261-278, 2017.

471 Wang, S., Zhang, Y., Ju, W., Qiu, B., and Zhang, Z.: Tracking the seasonal and inter-annual variations of global gross primary  
472 production during last four decades using satellite near-infrared reflectance data, *Science of the Total Environment*, 755,  
473 142569, 2021.

474 Wang, X., Biederman, J. A., Knowles, J. F., Scott, R. L., Turner, A. J., Dannenberg, M. P., Köhler, P., Frankenberg, C., Litvak,  
475 M. E., and Flerchinger, G. N.: Satellite solar-induced chlorophyll fluorescence and near-infrared reflectance capture  
476 complementary aspects of dryland vegetation productivity dynamics, *Remote Sensing of Environment*, 270, 112858, 2022.

477 Welp, L. R., Keeling, R. F., Meijer, H. A. J., Bollenbacher, A. F., Piper, S. C., Yoshimura, K., Francey, R. J., Allison, C. E.,  
478 and Wahlen, M.: Interannual variability in the oxygen isotopes of atmospheric CO<sub>2</sub> driven by El Niño, *Nature*,  
479 477, 579-582, 10.1038/nature10421, 2011.

480 Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J. A., Huete, A. R., Ichii, K., Ni, W., Pang, Y., and Rahman, A. F.:  
481 Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years, *Remote Sensing of Environment*, 233,  
482 111383, 2019.

483 Xiao, X., Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore III, B., and Ojima, D.: Modeling gross primary  
484 production of temperate deciduous broadleaf forest using satellite images and climate data, *Remote sensing of environment*,  
485 91, 256-270, 2004.

486 Xu, T., White, L., Hui, D., and Luo, Y.: Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in  
487 parameter estimation and model prediction, *Global Biogeochemical Cycles*, 20, 2006.

488 Yang, J., Tian, H. Q., Pan, S. F., Chen, G. S., Zhang, B. W., and Dangal, S.: Amazon drought and forest response: Largely  
489 reduced forest photosynthesis but slightly increased canopy greenness during the extreme drought of 2015/2016, *Global  
490 Change Biology*, 24, 1919-1934, 10.1111/gcb.14056, 2018.

491 Yao, Y., Liang, S., Li, X., Chen, J., Liu, S., Jia, K., Zhang, X., Xiao, Z., Fisher, J. B., and Mu, Q.: Improving global terrestrial  
492 evapotranspiration estimation using support vector machine by integrating three process-based algorithms, *Agricultural and  
493 Forest Meteorology*, 242, 55-74, 2017.

494 Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J. B., Zhang, N., Chen, J., Cheng, J., Zhao, S., and Zhang, X.: Bayesian multimodel  
495 estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations, *Journal of  
496 Geophysical Research: Atmospheres*, 119, 4521-4545, 2014.

497 Yuan, W., Cai, W., Nguy-Robertson, A. L., Fang, H., Suyker, A. E., Chen, Y., Dong, W., Liu, S., and Zhang, H.: Uncertainty  
498 in simulating gross primary production of cropland ecosystem from satellite-based models, *Agricultural and Forest  
499 Meteorology*, 207, 48-57, 10.1016/j.agrformet.2015.03.016, 2015.

500 Yuan, W., Cai, W., Xia, J., Chen, J., Liu, S., Dong, W., Merbold, L., Law, B., Arain, A., and Beringer, J.: Global comparison  
501 of light use efficiency models for simulating terrestrial vegetation gross primary production based on the LaThuile database,  
502 *Agricultural and Forest Meteorology*, 192, 108-120, 2014.

503 Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L. L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A. H., and Goulden,  
504 M. L.: Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production  
505 across biomes, *Agricultural and Forest Meteorology*, 143, 189-207, 2007.

506 Yuan, W., Zheng, Y., Piao, S., Ciais, P., Lombardozzi, D., Wang, Y., Ryu, Y., Chen, G., Dong, W., and Hu, Z.: Increased  
507 atmospheric vapor pressure deficit reduces global vegetation growth, *Science advances*, 5, eaax1396, 2019.

508 Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., and Dong, J.: A global moderate resolution dataset of gross primary  
509 production of vegetation for 2000–2016, *Scientific data*, 4, 1-13, 2017.

510 Zheng, Y., Shen, R., Wang, Y., Li, X., Liu, S., Liang, S., Chen, J. M., Ju, W., Zhang, L., and Yuan, W.: Improved estimate of  
511 global gross primary production for reproducing its long-term variation, 1982–2017, *Earth System Science Data*, 12, 2725-  
512 2746, 2020.

513

514