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

Xin Chen[1], Tiexi Chen[1,2,3*], Xiaodong Li[4], Yuanfang Chai[5], Shengjie Zhou[1], Renjie Guo[6], Jie Dai[1]

[1]School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China.
[2]Qinghai Provincial Key Laboratory of Plateau Climate Change and Corresponding Ecological and Environmental Effects, Qinghai University of Science and Technology, Xining 810016, China
[3]School of Geographical Sciences, Qinghai Normal University, Xining 810008, Qinghai, China.
[4]Qinghai Institute of Meteorological Science, Xining 810008, Qinghai, China.
[5]Department of Earth Sciences, Vrije Universiteit Amsterdam, Boelelaan 1085, 1081 HV, Amsterdam, the Netherlands
[6]Faculty of Geographical Science, Beijing Normal University, Beijing, China.

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

Abstract. Advancements~~The continuous developmentadvancement~~ ~~of~~in remote sensing technology ~~has~~have significantly~~been~~ contributed to the improvement of~~instrumental in improving~~ models for estimating terrestrial gross primary productivity (GPP). However, discrepancies in spatial distribution and interannual variability within GPP datasets pose challenges to a comprehensive understanding of the terrestrial carbon cycle.~~However, challenges arise from inconsistent spatial distributions and interannual variations in GPP datasets, which hinderimpeding our comprehensive understanding of the entire terrestrial carbon cycle.~~ In contrast to previous models that rely~~relying~~ on remote sensing and environmental variables, we developed ~~a~~ an ensemble model based on random forest~~, named GPP$_{ERF}$~~ (ERF model). This model used~~utilized~~ the GPP ~~GPP~~ outputs ~~from~~of established ~~remote sensing-based~~ models (EC-LUE, GPP-kNDVI, GPP-NIRv, Revised-EC-LUE, VPM, MODIS) as input~~s~~ to estimate GPP~~for GPP estimatesestimatesestimations~~. The ERF model~~GPP$_{ERF}$~~ demonstrated superior~~demonstratedshoweddemonstrated significant effectiveness by,~~ explaining ~~83.7~~85.1% of the monthly GPP variation~~s~~ ~~in GPP across~~at 17~~1~~0 sites~~. This performance~~ and surpassing the performance of ~~outperformedsurpassed~~ both ~~the~~ selected ~~remote sensing~~GPP models (~~72.4~~67.7%-77.~~1~~5%) and an independent random forest model using remote sensing and environmental variables (~~77~~81.~~7~~5%). Additionally, the ERF model ~~GPP$_{ERF}$ also exhibitedshowed the higher~~ improved the accuracy across~~in~~ each month and various~~different~~ subvalues, mitigating the issue ~~improvingwhich improved the phenomenon~~ of "high value underestimation and low value overestimation" in GPP estimates. Over the period from 2001 to 2022, the global ~~estimated~~ GPP estimated ~~value using~~by the ERF model~~the ensemble model based on random forest~~ was 13~~1~~2.~~2~~7 PgC yr^{-1}, with an increasing trend of ~~corresponding toexhibiting a trend of~~ 0.4~~5~~2 PgC yr^{-2}, which is comparable to or slightly better than the accuracy of other mainstream GPP datasets in term of validation results from ChinaFlux ~~. In addition~~~~Furthermore, the evaluation results using theutilizing flux sites fromof ChinaFlux indicated showedindicated that the dateset exhibited good generalization.~~ In summary, the ERF model offers a reliable alternative for reducing uncertainties in

GPP estimate, providing a more dependable global GPP estimate.~~the random forestmachine learning-based ensemble methodmodel helps to reduce the uncertainty in the estimation of a single remote sensingGPP model and provides a more reliable estimateestimation of global GPP.~~

1 Introduction

Gross primary productivity (GPP) is the largest carbon flux in the global carbon cycle, and serves as the primary input of ~~it is also the input of~~ carbon ~~to~~into the terrestrial carbon cycle~~during the carbon cycle~~. Uncertainties in GPP estimates~~estimation~~~~the estimation of GPP~~ ~~will be~~can ~~further~~ propagated to other carbon flux estimates, making it crucial~~so it is important~~ to clarify the spatio-temporal patterns of GPP (Xiao et al., 2019; Ruehr et al., 2023). However, global GPP is variously estimated from~~various~~~~different studies estimate global GPP to be between~~at 90 PgC yr^{-1} ~~and~~to 160 PgC yr^{-1} across different studies, with these variations becoming~~and these uncertainties become~~~~this uncertainty may~~can ~~be even~~ more pronounced when scaled down to~~extended to~~ regional scales or specific ecosystem types~~,~~. This variability underscores the necessity for innovative methods~~so it is necessary to develop some new methods~~ to reduce the uncertainty ~~of~~in GPP estimates (Jung et al., 2019; Ryu et al., 2019; Anav et al., 2015).

The light use efficiency (LUE) model is one of the most widely adopted methods ~~used models~~ for estimating GPP~~, which~~. It assumes that GPP is proportional to the photosynthetically active radiation absorbed by vegetation, and optimizes the spatio-temporal pattern of GPP through meteorological constraints such as temperature and water (Pei et al., 2022). However, variations~~the forms of~~ in these ~~meteorological~~ constraints varies significantly~~greatly~~, ~~and this difference alone can lead~~leading to~~result in a~~ differences of over~~more than~~ 10% in model explanatory power.~~the explanatory power of the models~~ (Yuan et al., 2014). Recent studies have proposed some novel~~new~~ vegetation indices~~, which~~ that have been shown to be effective proxies for GPP through theoretical derivation and observed~~observational~~ validation~~ by observations~~ (Badgley et al., 2017; Camps-Valls et al., 2021). However, these vegetation indices often use only remote sensing data as an input for estimating long-term GPP without considering~~taking~~ meteorological factors~~ into account~~, which has led to~~caused~~ some controversy (Chen et al., 2024; Dechant et al., 2020; Dechant et al., 2022). Both LUE and vegetation index models~~Both the LUE model and the vegetation index model~~ use a combination of linear mathematical formulas to estimate GPP. However, ecosystems are inherently~~highly~~ complex~~,~~, and the biases introduced ~~into a process~~ by these numerical models~~this numerical model~~ ~~will~~ increase the uncertainty in the estimates of the final product (GPP)~~ estimates~~. ~~The m~~Machine~~Machine~~ learning models has ~~been~~ shown great potential for improving GPP estimates in previous studies~~in previous studies to havethat it has great potential for improvingto improve GPP estimates~~ (Jung et al., 2020; Guo et al., 2023). Th~~is~~ese model ~~areis~~are trained by non-physical means directly using GPP observations and selected environmental and vegetation variables, and the performance of the model depends on~~model performance is related to~~ the number and quality of ~~the~~ observed data and the representativeness of ~~the~~ input data. ~~Machine learning has also been widely used in recent years due to its advantages such as the fact that nono need for parameter calibration is required and the reliable model accuracy.~~ Nevertheless, direct validation from flux towers of FLUXNET reveal~~show~~s that these models typically explain~~s~~ only ~~explains~~ about 70% of ~~the~~ monthly GPP variation~~s in GPP~~, with similar performance to other GPP models (Wang et al., 2021; Badgley et al., 2019; Zheng et al., 2020; Jung et al., 2020). Due to ~~the~~ deviations ~~of~~in the model structure, ~~there is~~ a common limitation~~issueproblem~~ across these models is poor estimate~~estimatione~~ of monthly extreme GPP~~in these models, that is, the estimation of the monthly extreme GPP is poor~~, ~~and~~leading to the phenomenon of "high value overestimation~~underestimate,~~ and low value overestimate~~ion~~"~~ occurs~~ (Zheng et al., 2020). ~~Currently, there are several remote sensing data-driven methods to estimate GPP, including light use efficiency (LUE) models, vegetation index models, machine learning models, and process models . Direct validation of flux towers from FLUXNET shows that these models usually only explain about 70% of the monthly variation in GPP . One possible reason is that remote sensing models cannot fully characterize all the processes of photosynthesis. This is understandable, most of the existing models use linear or nonlinear mathematical formulas to express a certain process of photosynthesis. However, the ecosystem is highly complex, the bias introduced by such a numerical model in a process will increase the uncertainty in the final product (GPP) estimates. For example, in the LUE model, the difference in the meteorological constraints alone can lead to a difference of more than 10% in the explanatory power of the model . As an important factor affecting photosynthesis, some models consider the effect of~~ ~~CO_2~~ ~~fertilization. However, a study revealed that the effect of~~ ~~CO_2~~ ~~fertilization showed a significant negative trend in the past 40 years, and this process may be missing in the model . Limited by the imperfection of the model mechanism, adjusting the model parameters is the most effective way to improve the simulation accuracy. The usual practice of the modeler is to divide the directly observed GPP data according to different vegetation types, and randomly select the testset through the cross-validation method to calibrate and validate the model parameters. However, this method is based on the assumption that the model parameters of the same vegetation type in different regions are roughly the same. In fact, the photosynthetic characteristics of the same vegetation type are also quite different in different regions. A typical example is the difference between C3 and C4 crops in the cropland, the GPP of C4 crops during the growing season may reach 600-800 gC~~ ~~m^{-2} $month^{-1}$~~~~, accounting for more than 60% of the annual GPP, in contrast, the GPP of C3 crops in the growing season is only 200-300~~ ~~m^{-2} $month^{-1}$~~~~, or even lower . Some other studies have also found that the maximum carboxylation rate (Vcmax) that determines photosynthesis at the leaf scale not only varies with vegetation types, but also depends on environmental factors . The same vegetation type also has a difference of 40umol~~ ~~m^{-2} s^{-1}~~ ~~in different geographical areas , all of which may lead to uncertainties in GPP estimate. A widespread problem is that the deviation of model structure and model parameters may lead to poor estimation of GPP in the monthly extreme value, and the phenomenon of "high value underestimation and low value overestimate" occurs.~~ Especially for extremely high values, which usually occur during the growing season and largely determine the annual value and inter-annual fluctuations~~variation~~ of GPP, this underestimation may hinder our understanding of the global~~entire~~ carbon cycle ~~process~~.

It is challenging~~difficult~~ for a single model to provide accurate estimates for all global regions~~havegive a good estimate forestimation in all regions of the worldglobe~~. Ensemble models have been shown to outperform single models in previous studies, potentially addressing some inherent issues~~Previous studies have shown that an ensemble model maycan perform better than a single model, which may improve some potential problems in model estimation~~ in model estimate~~estimationone~~ (Chen et al., 2020; Yao et al., 2014). Traditional multi-model ensemble methods usually use ~~a~~ simple multi-model ~~simple~~ average or a ~~bayesian~~ weighted bayesian average. However, these methods typically assign fixed weights to each model and are essentially linear combinations.~~usually only provide fixed weights for each model, and are essentially linear combinations between multiple models.~~ Recent~~Some recent~~ studies have applied~~apply~~ machine learning methods to multi-model ensembles to establish nonlinear relationships between multiple simulated target variables and real target variable~~s~~, ~~enhanc~~improving simulation performance~~, improving the to improve simulation performance~~ (Bai et al., 2021; Yao et al., 2017; Tian et al., 2023). Whether this method can improve some common problems with a single GPP model, such as high value underestimation and low value overestimation, is not clear and needs to further investigation~~ be further explored.However, few studies have applied this method to the global GPP estimation, which providesis a novelnew idea for improvingto improve some common problems of a single remote sensing model (such as high value underestimation and ground low value overestimation).~~

In this study, we attempt to use an ensemble model based on the random forest (ERF model)~~an ensemble model based on machine learning methods~~ to improve global GPP estimat~~ione~~~~the estimation of global GPP~~. Specifically, the work of this study includes the following~~ points~~: (1) ~~After r~~Re~~-~~calibrating the parameters ~~of~~for each model, and comparing the performance of ~~fiveseven~~six ~~remote sensing~~GPP models and the ERF~~ensemble~~ model~~s was~~~~ere compared~~; (2) Focusing on the phenomenon of "high value underestimation and low value overestimation" in each model,~~,~~ and evaluating~~comparingcompared~~ the performance of each model in different~~each~~ month~~s~~s, ~~each~~ vegetation type~~s~~s and ~~different~~ sub~~-~~values (high value, median value, low value); (3) Developing a global GPP dataset using ~~an ensemble~~the ERF mode~~l~~l ~~based on machine learning methods, and using GPP observations from ChinaFlux as a complementary validation set to test the generalization of this dataset, i.e. the extent to which the dataset captures changes in GPP in regions where fewer sites are includedused in the modeling process.~~ and validate its generalization using GPP observations from ChinaFlux.

2 Method

2.1 Data at the global scale

In this study, we selected remote sensing data from the Moderate Resolution Imaging Spectroradiometer (MODIS) and meteorological data from EAR5~~ Land~~ to estimate ~~the~~ global GPP (Hersbach et al., 2020). For the remote sensing data, surface reflectance (red band, near infrared band, blue band and shortwave infrared band), leaf area ~~I~~index (LAI) and ~~F~~fraction of ~~P~~photosynthetically ~~A~~active ~~R~~radiation (FPAR) were used~~ in this study~~. For meteorological data, we selected average air temperature, dew point temperature, minimum air temperature, total solar radiation~~,~~ and direct solar radiation.

~~The d~~Dew point temperature and air temperature were used to calculate ~~the~~ saturated vapor pressure difference (VPD) (Yuan et al., 2019), and ~~the~~ diffuse solar radiation was derived~~calculated~~ as the difference between ~~the~~ total solar radiation and ~~the~~ direct solar radiation. ~~m~~Minimum air temperature was obtained from the hourly air temperature. ~~The~~ CO_2 data were obtained from~~comes from~~ the monthly average carbon dioxide levels measured by the Mauna Loa Observatory in Hawaii. Table 1 provides an overview of the datasets used in this study.~~Table 1 shows the details of these data.~~

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 ~~A~~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~~hourly~~	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_2	NOAA's Earth System Research Laboratory	/	Monthly	2001-2022
Distribution map of C4 crops	Harvested Area and Yield for 175 Crops	1/12 °	Annual	2000
Land use	MCD12C1	0.05 °	Annual	2010

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

The land use map ~~comes~~was derived from the IGBP classification of MCD12Q1, and 2010 was ~~selected~~chosen as the reference year (that is, land use data is unchanged in the simulation of global GPP). In order to meet the requirements~~need~~ of subsequent research, ~~the~~ land cover types were ~~combined~~grouped into 9 categories: ~~d~~Deciduous Broadleaf Forest (DBF), ~~e~~Evergreen Needleleaved~~coniferous~~ ~~f~~Forest (ENF), Evergreen Broadleaf Forest (EBF), Mixed Forest (MF), Grassland (GRA), Cropland (including CRO-C3 and CRO-C4), Savannah (SAV), Shrub (SHR), Wetland (WET).

Finally~~Ultimately~~, for higher resolution data, we gridded the dataset to 0.05 ° by averaging all pixels whose center fell within each 0.05 ° grid cell for upscaling. For lower resolution data, we used the nearest neighbor resampling to 0.05 °. ~~all data were resampled to a spatial resolution of 0.05 °, while~~ In addition, MODIS data ~~from MODIS~~ were aggregated to a monthly scale to ensure spatio-temporal~~meet spatiotemporal~~ consistency.

2.2 Observation data at the site scale

~~The modeling used~~ GPP observations ~~-~~were sourced from the FLUXNET 2015 dataset, which includes carbon fluxes and meteorological variables from more than 200 flux sites around the world (Pastorello et al., 2020). GPP cannot be obtained directly from ~~the~~ flux sites and usually needs to be obtained by dismantling the Net Ecosystem Exchange. We chose a ~~month-scale~~monthly level GPP based on the nighttime partitioning method and retained only high quality data (NEE_VUT_REF_QC > 0.8) for every year, ultimately~~and finally~~ selecting~~selected~~ 17~~1~~0 sites with 10~~824~~932 monthly values for this study. In addition, we selected monthly average air temperature, total solar radiation and VPD ~~on the monthly scale were selected~~. The site observations do not provide direct solar radiation, so we extracted data from the ERA5 covering the flux tower. ~~The m~~Monthly minimum air temperature ~~is~~was derived~~obtained~~ from ~~the~~ hourly air temperature. Since some required model data are~~part of the data required by~~for ~~the model is~~ not directly available at ~~the~~ flux sites, ~~surface reflectance,~~ LAI and FPAR were extracted~~on at a scale of~~from MOD15A2H (500 m) ~~were extracted,~~, and surface reflectance data (red band, near infrared band, blue band and shortwave infrared band) were~~are~~ derived from MCD43A4 (500 m) and MOD09A1 (500 m). ~~which~~These data are roughly similar to the footprint of the flux site and can represent the land surface of the site ~~situation~~ (Chu et al., 2021).

2.3 GPP estimation model~~Remote sensing models and ensemble models for estimating GPP~~

We selected six independent models to estimate GPP in this study.~~In this study, fivesevensix independent remote sensing models were selected to estimate GPP.~~ These models are widely used with few model parameters and have demonstrated reliable accuracy~~have shown reliable model accuracy~~ in previous studies (Zheng et al., 2020; Zhang et al., 2017; Badgley et al., 2017). The ~~sevensix~~~~five~~ models are EC-LUE, Revised-EC-LUE, NIRv-based linear model, kNDVI-based linear model, VPM, MODIS~~ and traditional random forest model~~~~ using remote sensing and environmental variables~~. The VPM, MODIS and EC-LUE ~~is a~~are LUE models based on~~driven by~~ remote sensing data and meteorological data~~. These models assumes that GPP is proportional to the photosynthetically active radiation absorbed by the canopy, and the seasonal variation of GPP is corrected by meteorological constraints~~ (Yuan et al., 2007; Running et al., 2004; Xiao et al., 2004)~~;~~. Recently, Zheng et al. ~~revised the EC-LUE model and~~(2020) proposed the Revised-EC-LUE model, which divides the canopy into sunlit and shaded leaves~~, and considers long-term changes in CO_2 to~~, ~~improve~~improving~~d~~ the estimation of global GPP (Zheng et al., 2020). The NIRv and kNDVI are newly proposed vegetation indices~~, which are~~ calculated from the red and near-infrared bands of the reflectance spectrum (Badgley et al., 2017; Camps-Valls et al., 2021). Similar to ~~the S~~solar induced chlorophyll fluorescence~~ (SIF)~~, they exhibit~~exhibithave~~ a linear relationship with~~withto the~~ GPP and are considered ~~to be~~ effective proxies for ~~the~~ GPP. Detailed descriptions of all models can be found~~are presented~~ in Text S1. ~~The randomRandom forest (RF) method has beenis widely used in GPP estimation, which usuallyand typically uses meteorological variables and the vegetation index for modeling . In this study, we used average air temperature, minimum air temperature, VPD, direct solar radiation, diffuse solar radiationradiation, FPAR and LAI to estimate GPP, similar to the variables selected in some previous studies..~~

To reduce ~~the~~ uncertainty in ~~estimating~~ GPP estimat~~tione~~ from a single model, we ~~also~~ used the ~~an ensemble model based on the random forest (ERF)a multi-modalmodel ensemble method~~ERF model, the basic idea of which is to restructure~~re-model~~ the simulated values of multiple models. ~~Random forest is an ensemble learning algorithm that combines the outputs of multiple decision trees to produce a single result, and is commonly used for classification and regression problems. In the regression problem, the output result of each decision tree is a continuous value, and the average of the output results of all decision trees is taken as the final result.~~ In this study~~, an ensemble model based on the random forest (ERF) method was used,. In contrast to theUnlike traditional machine learningRF methods, that is,~~ we directly used the ~~random forest method~~ERF model~~s~~ to establish the relationship between the GPP simulated by the above ~~four~~six models and ~~the~~ GPP observations.~~-~~ In addition, for comparison with the ERF model, we also used the random forest (RF) method for modeling. In this study, we used average air temperature, minimum air temperature, VPD, direct solar radiation, diffuse solar radiation, FPAR and LAI to estimate GPP. Both models used the random forest method, which has been widely used in previous studies of GPP estimati~~one, similar to the variables selected in some previous studies~~ (Jung et al., 2020; Guo et al., 2023). Random forest is an ensemble learning algorithm that combines the outputs of multiple decision trees to produce a single result, and is commonly used for classification and regression problems (Belgiu and Drăguț, 2016). In the regression problem, the output result of each decision tree is a continuous value, and the average of the output results of all decision trees is taken as the final result. An overview~~A summary~~ of all models used can be found~~is shown~~ in Table 2.

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

ID	Model	Input data	Output
1	EC-LUE	FPAR, VPD, AT, SRAD, CO_2	GPP_{EC}
2	Revised-EC-LUE	LAI, VPD, AT, DifSR, DirSR, CO_2	GPP_{REC}
3	kNDVI-GPP	Red band and near infrared band	GPP_{kNDVI}
4	NIRv-GPP	Red band and near infrared band	GPP_{NIRv}
5	VPM	Red band, near infrared band, blue band, shortwave infrared band, AT, SRAD	GPP_{VPM}
6	MODIS	FPAR, SRAD, MINT, VPD	GPP_{MODIS}
~~5~~7	~~Traditional r~~Random forest model (RF)	LAI, FPAR, AT, MINT, VPD, DifSR, DirSR~~SRAD, LAI~~	GPP_{RF}
~~6~~8	Ensemble model based on random forest (ERF)	GPP_{EC}, GPP_{REC}, GPP_{kNDVI}, GPP_{NIRv}, GPP_{MODIS}, GPP_{VPM}	GPP_{ERF}

2.4 Model parameter calibration and ~~V~~validation

FLUXNET only provides GPP observations and meteorological data, lacking direct measurements for LAI, FPAR, and surface reflectance,~~while LAI, FPAR and surface reflectanceother data are not provided,~~ so only remote sensing data can be used. Considering the variety of remote sensing data sources, such as MODIS and AVHRR, it is evident that calibrating the same GPP model with different remote sensing data can yield varied parameters.~~However, there are many sources of remote sensing data available, such as MODIS, AVHRR, etc., so using different remote sensing data to calibrate the same GPP model may produce different model parameters.~~ In addition, the number of sites used to calibrate model parameters is also an important influencing factor for model parameters. The original parameters of these models were calibrated with only a limited~~small~~ number of sites (e.g., 95 sites ~~were used~~ for Revised EC-LUE and 104 for NIRv) (Wang et al., 2021; Zheng et al., 2020)~~.~~. Therefore~~Due to the difference between meteorological data and vegetation data~~, to reduce the impact of the uncertainty of ~~the~~ model parameters- on simulation results, we did not use original~~default~~ parameters~~ in the model, but~~ and conducted~~carried out~~ parameter calibration ~~and model validation~~ for ~~all remote sensing~~GPP models across~~according to~~ different vegetation types. For EC-LUE~~ and~~, Revised EC-LUE, VPM and MODIS, the Markov chain Monte Carlo method ~~(MCMC)~~ was used to calibrate ~~the~~ model parameters. Traditionally,~~The~~~~In the traditional MCMC method,~~ ~~usually takes~~ the mean ~~value~~ of the posterior distribution of ~~the~~ parameters is ~~usually~~ taken as the optimal value~~,~~. However,~~while~~ previous studies have indicated~~shown~~ that some model parameters are~~cannot be~~ not well constrained when calibrating multiple model parameters (Xu et al., 2006; Wang et al., 2017), so we selected~~use~~ the parameter with the smallest root-mean-square error (RMSE) as the optimal parameter in each iteration. For each vegetation type, we randomly selected 70% of the sites for parameter calibration, and repeated the process~~ was repeated~~ 200 times. In order to avoid overfitting, we adopted~~tookused~~ the mean of the 200 calibrated parameters as the final model parameters. Similarly, for the two vegetation index models, we randomly selected 70% of the sites in each vegetation type for parameter calibration~~re~~, peating the process 200 times~~. The process was repeated 200 times,~~ and using the mean of the 200 calibrated parameters ~~was used~~ as the final model parameters. After obtaining GPP estimates from the six ~~four remote sensing~~GPP models, we evaluated~~tested~~ the simulation performance of the ~~traditional random forest~~RF model and the ~~random forest-based ensemble~~ERF model respectively. For both models, we evaluate~~tested~~d the model performance using 5-fold cross-validation, where the process was repeated 200 times, and the mean of the 200 GPP estimates was considered the final GPP estimate~~mean of the GPP estimated 200 times as~~~~was~~ ~~the final GPP estimate~~. We utilized the determination coefficient~~Goodness of Ffit~~ (R^2) and RMSE as metrics to evaluate~~were used to measure~~ the simulation performance of all models. Additionally, ~~In addition, weW~~we used the ratio of GPP simulations to GPP observations (Sim/Obs) to measure whether the model ~~was~~ overestimate~~d~~s or underestimate~~d~~s.

2.5 Global GPP estimation based on ERF model and its uncertainty.

Based on the ERF model~~site-scale model~~, we estimated global GPP for 2001-2022 (ERF_GPP). The uncertainties~~uncertainty~~ of ERF_GPP can be attributed to two primary factors~~mainly comes from two aspects~~, one is the influence of the number of GPP observations, and the other is the influence of the number of features (that is, the simulated GPP)~~ used in the modeling process~~. For the first type of uncertainty~~For the first uncertainty~~, we randomly selected 80% of the data to build a model and simulate the multi-year average of global GPP. The process was repeated 100 times, yielding 100 sets~~and 100 groups~~ of multi-year averages of ERF_GPP~~ were obtained~~. Their standard deviations were considered as~~to be~~ the uncertainty of ERF_GPP caused by the number of GPP observations. For the second type of uncertainty, we selected ~~choose~~ different number of features to build a model~~s~~ and simulate the multi-year average of global GPP. A total of 56 sets~~ groups~~ of multi-year averages ~~of~~for ERF_GPP were~~are~~ obtained. The standard deviation of different combinations ~~i~~was considered to be the uncertainty of ERF_GPP caused by the number of features.

2.~~5~~6 Evaluation of the generalization of different GPP datasets

The majority of flux sites~~Most of the flux sites~~ in ~~Fluxnet2015~~FLUXNET are concentrated~~locatedconcentrated~~ in Europe and North America, it is unclear~~not clear~~ whether the different GPP estimation methods are suitable for ~~some~~ regions with sparse flux sites. Recently, ChinaFlux has published GPP observations from several~~multiple~~ sites, offering an opportunity to~~which which providesproviding an opportunity to~~ ~~test~~evaluate the generalization of ~~the~~ different GPP datasets. However, the spatial resolution of most GPP datasets is 0.05°, and a direct comparison with GPP observations at flux sites is challenging. Therefore, we extracted 0.05° MODIS land use covering the flux sites. ~~tower, and wheni~~If the vegetation type ~~of vegetation observed by~~ of the flux site~~tower~~ matched~~was consistent with~~ the MODIS land use, the site was used for the analysis. Finally, a total of 12 flux sites were selected (Figure S2), and Table S1 shows the information of these sites. The same procedure was applied to~~done for~~ FLUXNET, resulting in the selection of 52 sites~~and a total of 52 sites were selected~~ (Figure S2). It should be noted that due to the absence of meteorological data from some sites in Chinaflux, we did not validate all GPP models at the site scale (500 m).

~~Based on site-scale models, we estimated the global GPP for 2001-2022 using an ensemble model based on random forest~~~~ERF model~~ ~~(ERF_GPP).~~ We ~~test~~evaluated the generalization of ERF_GPP ~~on~~at 12 ChinaFlux sites and 52 FLUXNET sites. In addition, we selected a number of widely used GPP datasets for comparison, including BESS (Li et al., 2023), GOSIF (Li and Xiao, 2019), ~~ECGC~~FLUXCOM (Jung et al., 2020), NIRv (Wang et al., 2021), Revise-EC-LUE (Zheng et al., 2020), MODIS (Running et al., 2004), VPM (Zhang et al., 2017), which ~~are~~were generated using different GPP estimation methods. These GPP ~~product~~datasets all have a spatial resolution of ~~0.05 °~~500 m-0.5 °, similar to the resampling process in section 2.1, we have unified them to 0.05 °~~, avoiding the uncertainty of GPP validation introduced due to resolution differences.~~ The common time range for these ~~product~~datasets spanned from 2001 to 2018~~is 2001-2018~~, and the temporal~~time~~ resolution ~~has been~~was unified~~standardized~~was ~~unified~~ to monthly to ~~to~~ match the~~be consistent with~~ GPP observations.

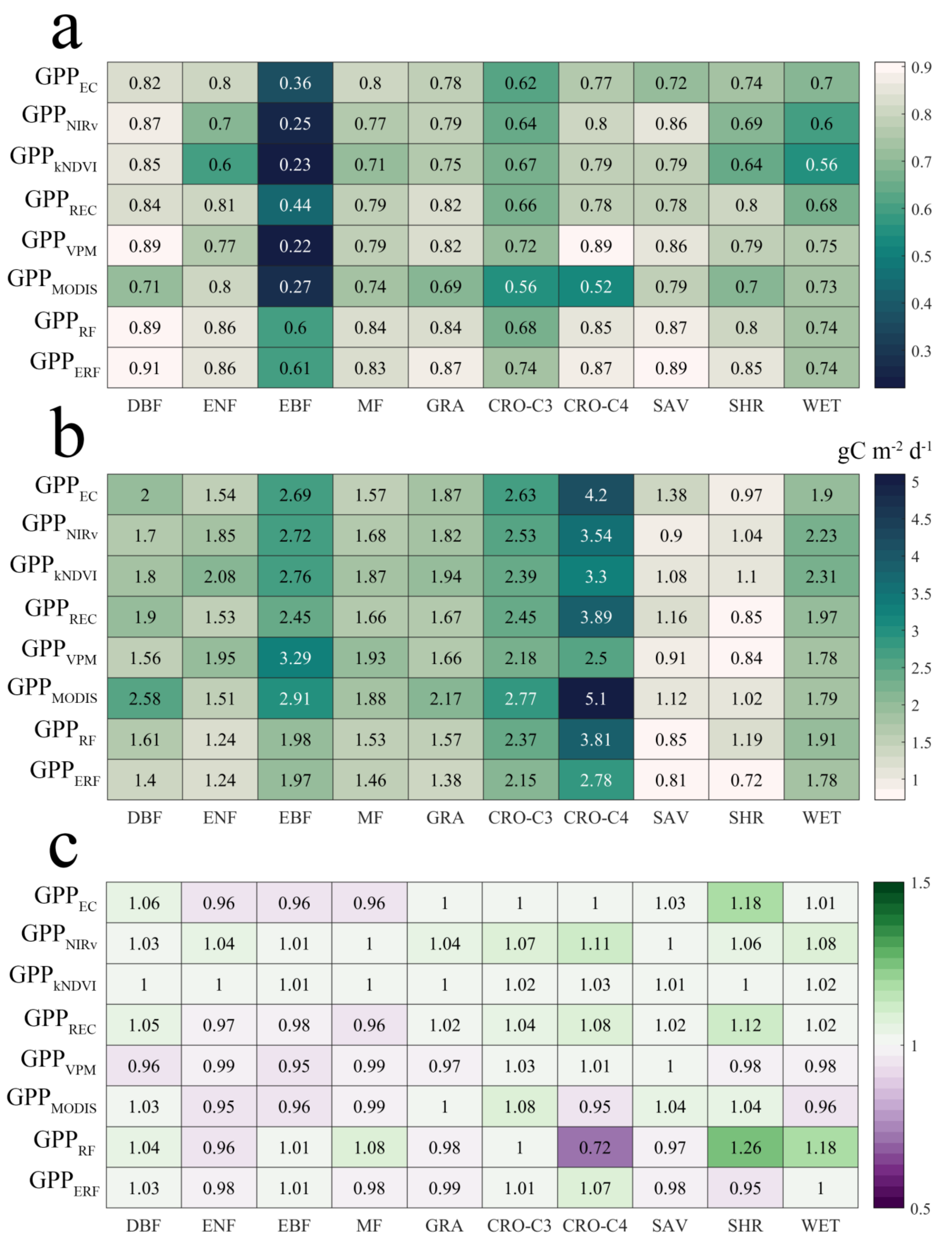
3 Result

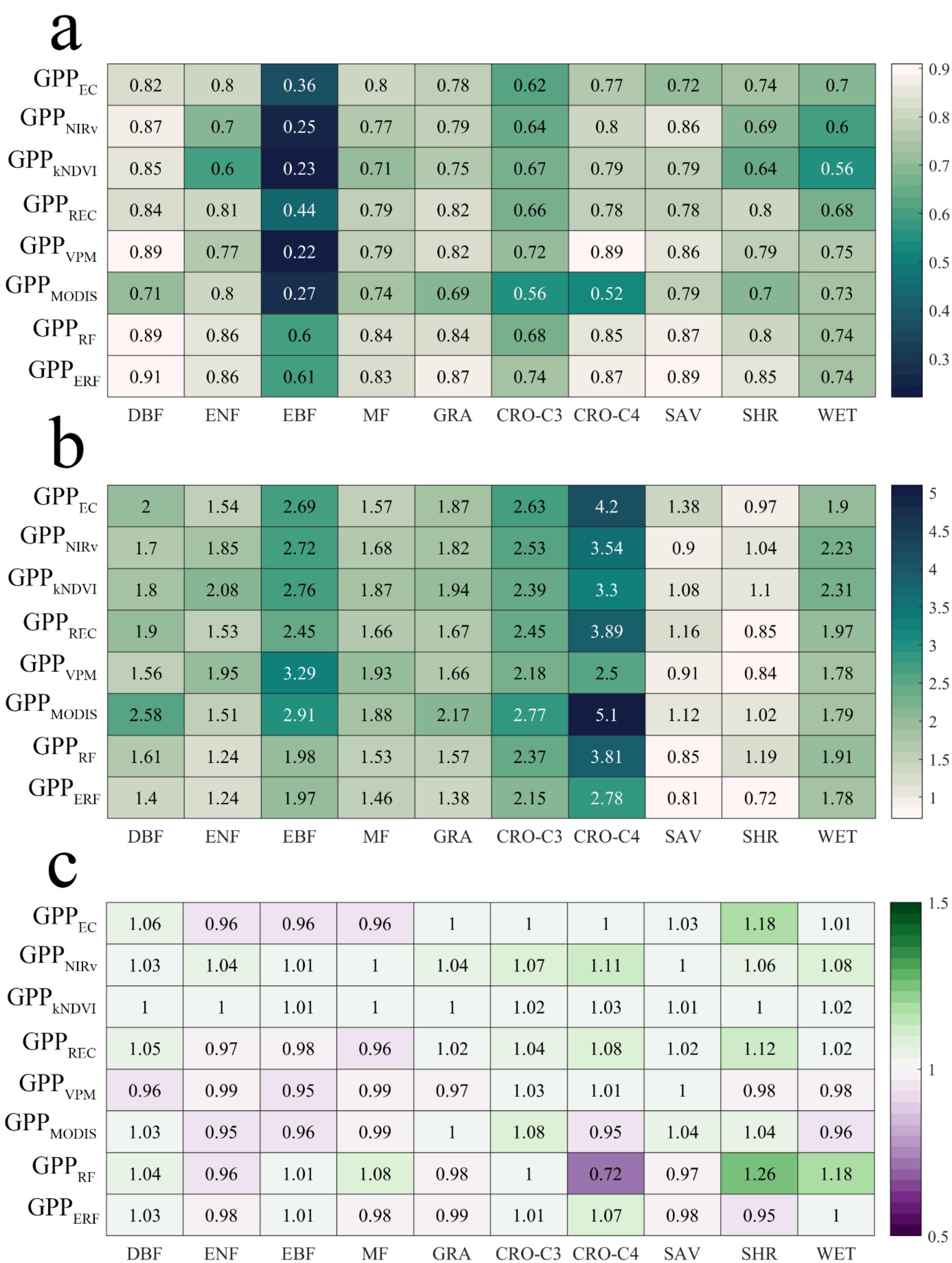
3.1 Performance of ~~six~~GPP models at site scale

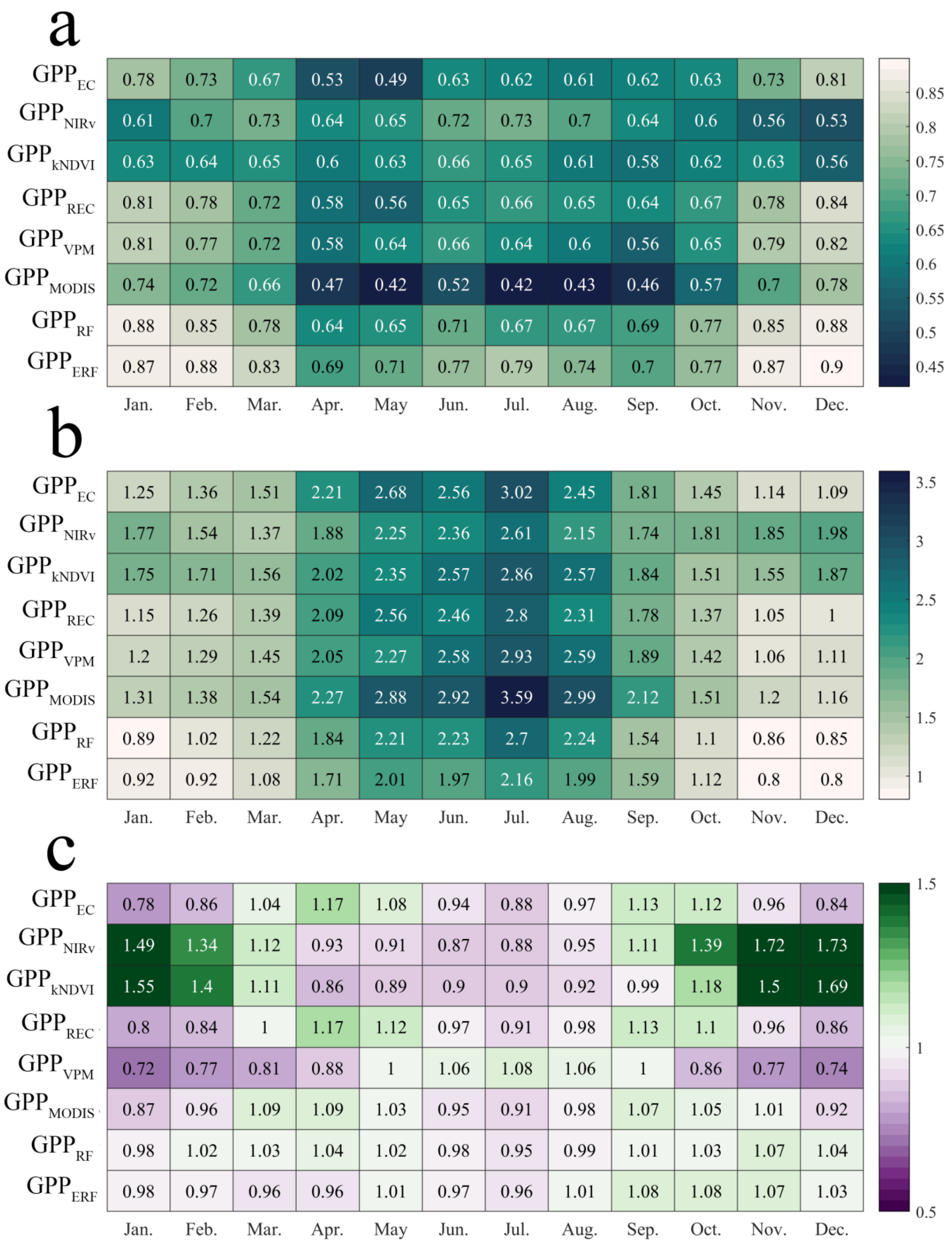
Table S2-S~~5~~7 show~~s~~ the optimization results of ~~four~~~~six~~ six GPP model parameters~~parameters of the remote sensing~~~~GPP models~~ ~~parameters~~. Consistent with~~Similar to~~ the previous study, in ~~the EC-LUE model, VPDM and the~~ Revised-EC-LUE model, the light use efficiency parameter of shade leaves was significantly higher than that of ~~the~~ sunlit leaves (Zheng et al., 2020). It is necessary to divide the cropland into C3 crops and C4 crops. In all models, the light use efficiency parameters of C4 crops were significantly higher than those of C3 crops, which was particularly~~especially~~ reflected in the two vegetation index models of GPP_{kNDVI} and GPP_{NIRv}, the slope of the linear regression directly reflected~~was a direct reflection of~~ the difference in the photosynthetic capacity of the different crops.

Figure 1 shows the performance of all models across different~~on the~~ vegetation types. Overall, the performance of the ~~ensemble~~ERF model was better than that of the ~~remote sensing~~GPP models. GPP_{ERF} ~~always~~ had the high~~est~~er accuracy among all models, with R^2 between 0.61-0.91 and RMSE between 0.~~8~~72-~~3~~ 2.78 gC m^{-2} d^{-1}. In contrast, ~~in EBF,~~ the LUE and vegetation index models performed relatively poorly in EBF, with R^2 below 0.5.~~the performance of the two vegetation index models was relatively poor, especially for evergreen forests, the~~ ~~R^2 of GPP_{kNDVI} and GPP_{NIRv} was much significantly lower than other models.~~ It is worth noting that compared to other vegetation types, the RMSE was highest for cropland~~of cropland was the higher~~, with ~~5~~ 6 out of ~~6~~ 8 models ~~in~~for C4 ~~C~~crop exceeding 3 gC m^{-2} d^{-1}, ~~which suggested~~suggesting that these existing GPP models may not properly capture the~~track~~ seasonal changes in cropland GPP. ~~No significant estimation bias in vegetation type was found in four remote sensing model with calibration parameters and the ensemble model.~~Four remote

~~sensing~~Six models with calibration parameters and ~~the ensemble~~ERF model were found to have no significant deviation across~~in~~ vegetation types. However, GPP_{RF} was significantly underestimated ~~in~~for C4 crops and ~~significantly~~ overestimated ~~in~~for SHR~~ and WET~~.


a
GPP_EC 0.82 0.8 0.36 0.8 0.78 0.62 0.77 0.72 0.74 0.7
GPP_NIRv 0.87 0.7 0.25 0.77 0.79 0.64 0.8 0.86 0.69 0.6
GPP_kNDVI 0.85 0.6 0.23 0.71 0.75 0.67 0.79 0.79 0.64 0.56
GPP_REC 0.84 0.81 0.44 0.79 0.82 0.66 0.78 0.78 0.8 0.68
GPP_VPM 0.89 0.77 0.22 0.79 0.82 0.72 0.89 0.86 0.79 0.75
GPP_MODIS 0.71 0.8 0.27 0.74 0.69 0.56 0.52 0.79 0.7 0.73
GPP_RF 0.89 0.86 0.6 0.84 0.84 0.68 0.85 0.87 0.8 0.74
GPP_ERF 0.91 0.86 0.61 0.83 0.87 0.74 0.87 0.89 0.85 0.74
DBF ENF EBF MF GRA CRO-C3 CRO-C4 SAV SHR WET
0.9 0.8 0.7 0.6 0.5 0.4 0.3
b
gC m-2 d-1
GPP_EC 2 1.54 2.69 1.57 1.87 2.63 4.2 1.38 0.97 1.9
GPP_NIRv 1.7 1.85 2.72 1.68 1.82 2.53 3.54 0.9 1.04 2.23
GPP_kNDVI 1.8 2.08 2.76 1.87 1.94 2.39 3.3 1.08 1.1 2.31
GPP_REC 1.9 1.53 2.45 1.66 1.67 2.45 3.89 1.16 0.85 1.97
GPP_VPM 1.56 1.95 3.29 1.93 1.66 2.18 2.5 0.91 0.84 1.78
GPP_MODIS 2.58 1.51 2.91 1.88 2.17 2.77 5.1 1.12 1.02 1.79
GPP_RF 1.61 1.24 1.98 1.53 1.57 2.37 3.81 0.85 1.19 1.91
GPP_ERF 1.4 1.24 1.97 1.46 1.38 2.15 2.78 0.81 0.72 1.78
DBF ENF EBF MF GRA CRO-C3 CRO-C4 SAV SHR WET
5 4.5 4 3.5 3 2.5 2 1.5 1
c
GPP_EC 1.06 0.96 0.96 0.96 1 1 1 1.03 1.18 1.01
GPP_NIRv 1.03 1.04 1.01 1 1.04 1.07 1.11 1 1.06 1.08
GPP_kNDVI 1 1 1.01 1 1 1.02 1.03 1.01 1 1.02
GPP_REC 1.05 0.97 0.98 0.96 1.02 1.04 1.08 1.02 1.12 1.02
GPP_VPM 0.96 0.99 0.95 0.99 0.97 1.03 1.01 1 0.98 0.98
GPP_MODIS 1.03 0.95 0.96 0.99 1 1.08 0.95 1.04 1.04 0.96
GPP_RF 1.04 0.96 1.01 1.08 0.98 1 0.72 0.97 1.26 1.18
GPP_ERF 1.03 0.98 1.01 0.98 0.99 1.01 1.07 0.98 0.95 1
DBF ENF EBF MF GRA CRO-C3 CRO-C4 SAV SHR WET
1.5 1 0.5


a
GPP EC: 0.82 0.8 0.36 0.8 0.78 0.62 0.77 0.72 0.74 0.7
GPP NIRv: 0.87 0.7 0.25 0.77 0.79 0.64 0.8 0.86 0.69 0.6
GPP kNDVI: 0.85 0.6 0.23 0.71 0.75 0.67 0.79 0.79 0.64 0.56
GPP REC: 0.84 0.81 0.44 0.79 0.82 0.66 0.78 0.78 0.8 0.68
GPP VPM: 0.89 0.77 0.22 0.79 0.82 0.72 0.89 0.86 0.79 0.75
GPP MODIS: 0.71 0.8 0.27 0.74 0.69 0.56 0.52 0.79 0.7 0.73
GPP RF: 0.89 0.86 0.6 0.84 0.84 0.68 0.85 0.87 0.8 0.74
GPP ERF: 0.91 0.86 0.61 0.83 0.87 0.74 0.87 0.89 0.85 0.74
DBF ENF EBF MF GRA CRO-C3 CRO-C4 SAV SHR WET
0.9 0.8 0.7 0.6 0.5 0.4 0.3
b
GPP EC: 2 1.54 2.69 1.57 1.87 2.63 4.2 1.38 0.97 1.9
GPP NIRv: 1.7 1.85 2.72 1.68 1.82 2.53 3.54 0.9 1.04 2.23
GPP kNDVI: 1.8 2.08 2.76 1.87 1.94 2.39 3.3 1.08 1.1 2.31
GPP REC: 1.9 1.53 2.45 1.66 1.67 2.45 3.89 1.16 0.85 1.97
GPP VPM: 1.56 1.95 3.29 1.93 1.66 2.18 2.5 0.91 0.84 1.78
GPP MODIS: 2.58 1.51 2.91 1.88 2.17 2.77 5.1 1.12 1.02 1.79
GPP RF: 1.61 1.24 1.98 1.53 1.57 2.37 3.81 0.85 1.19 1.91
GPP ERF: 1.4 1.24 1.97 1.46 1.38 2.15 2.78 0.81 0.72 1.78
DBF ENF EBF MF GRA CRO-C3 CRO-C4 SAV SHR WET
5 4.5 4 3.5 3 2.5 2 1.5 1
c
GPP EC: 1.06 0.96 0.96 0.96 1 1 1 1.03 1.18 1.01
GPP NIRv: 1.03 1.04 1.01 1 1.04 1.07 1.11 1 1.06 1.08
GPP kNDVI: 1 1 1.01 1 1 1.02 1.03 1.01 1 1.02
GPP REC: 1.05 0.97 0.98 0.96 1.02 1.04 1.08 1.02 1.12 1.02
GPP VPM: 0.96 0.99 0.95 0.99 0.97 1.03 1.01 1 0.98 0.98
GPP MODIS: 1.03 0.95 0.96 0.99 1 1.08 0.95 1.04 1.04 0.96
GPP RF: 1.04 0.96 1.01 1.08 0.98 1 0.72 0.97 1.26 1.18
GPP ERF: 1.03 0.98 1.01 0.98 0.99 1.01 1.07 0.98 0.95 1
DBF ENF EBF MF GRA CRO-C3 CRO-C4 SAV SHR WET
1.5 1 0.5


a
GPP_EC 0.78 0.73 0.67 0.53 0.49 0.63 0.62 0.61 0.62 0.63 0.73 0.81
GPP_NIRv 0.61 0.7 0.73 0.64 0.65 0.72 0.73 0.7 0.64 0.6 0.56 0.53
GPP_kNDVI 0.63 0.64 0.65 0.6 0.63 0.66 0.65 0.61 0.58 0.62 0.63 0.56
GPP_REC 0.81 0.78 0.72 0.58 0.56 0.65 0.66 0.65 0.64 0.67 0.78 0.84
GPP_VPM 0.81 0.77 0.72 0.58 0.64 0.66 0.64 0.6 0.56 0.65 0.79 0.82
GPP_MODIS 0.74 0.72 0.66 0.47 0.42 0.52 0.42 0.43 0.46 0.57 0.7 0.78
GPP_RF 0.88 0.85 0.78 0.64 0.65 0.71 0.67 0.67 0.69 0.77 0.85 0.88
GPP_ERF 0.87 0.88 0.83 0.69 0.71 0.77 0.79 0.74 0.7 0.77 0.87 0.9
Jan. Feb. Mar. Apr. May Jun. Jul. Aug. Sep. Oct. Nov. Dec.
0.85 0.8 0.75 0.7 0.65 0.6 0.55 0.5 0.45
b
GPP_EC 1.25 1.36 1.51 2.21 2.68 2.56 3.02 2.45 1.81 1.45 1.14 1.09
GPP_NIRv 1.77 1.54 1.37 1.88 2.25 2.36 2.61 2.15 1.74 1.81 1.85 1.98
GPP_kNDVI 1.75 1.71 1.56 2.02 2.35 2.57 2.86 2.57 1.84 1.51 1.55 1.87
GPP_REC 1.15 1.26 1.39 2.09 2.56 2.46 2.8 2.31 1.78 1.37 1.05 1
GPP_VPM 1.2 1.29 1.45 2.05 2.27 2.58 2.93 2.59 1.89 1.42 1.06 1.11
GPP_MODIS 1.31 1.38 1.54 2.27 2.88 2.92 3.59 2.99 2.12 1.51 1.2 1.16
GPP_RF 0.89 1.02 1.22 1.84 2.21 2.23 2.7 2.24 1.54 1.1 0.86 0.85
GPP_ERF 0.92 0.92 1.08 1.71 2.01 1.97 2.16 1.99 1.59 1.12 0.8 0.8
Jan. Feb. Mar. Apr. May Jun. Jul. Aug. Sep. Oct. Nov. Dec.
3.5 3 2.5 2 1.5 1
c
GPP_EC 0.78 0.86 1.04 1.17 1.08 0.94 0.88 0.97 1.13 1.12 0.96 0.84
GPP_NIRv 1.49 1.34 1.12 0.93 0.91 0.87 0.88 0.95 1.11 1.39 1.72 1.73
GPP_kNDVI 1.55 1.4 1.11 0.86 0.89 0.9 0.9 0.92 0.99 1.18 1.5 1.69
GPP_REC 0.8 0.84 1 1.17 1.12 0.97 0.91 0.98 1.13 1.1 0.96 0.86
GPP_VPM 0.72 0.77 0.81 0.88 1 1.06 1.08 1.06 1 0.86 0.77 0.74
GPP_MODIS 0.87 0.96 1.09 1.09 1.03 0.95 0.91 0.98 1.07 1.05 1.01 0.92
GPP_RF 0.98 1.02 1.03 1.04 1.02 0.98 0.95 0.99 1.01 1.03 1.07 1.04
GPP_ERF 0.98 0.97 0.96 0.96 1.01 0.97 0.96 1.01 1.08 1.08 1.07 1.03
Jan. Feb. Mar. Apr. May Jun. Jul. Aug. Sep. Oct. Nov. Dec.
1.5 1 0.5

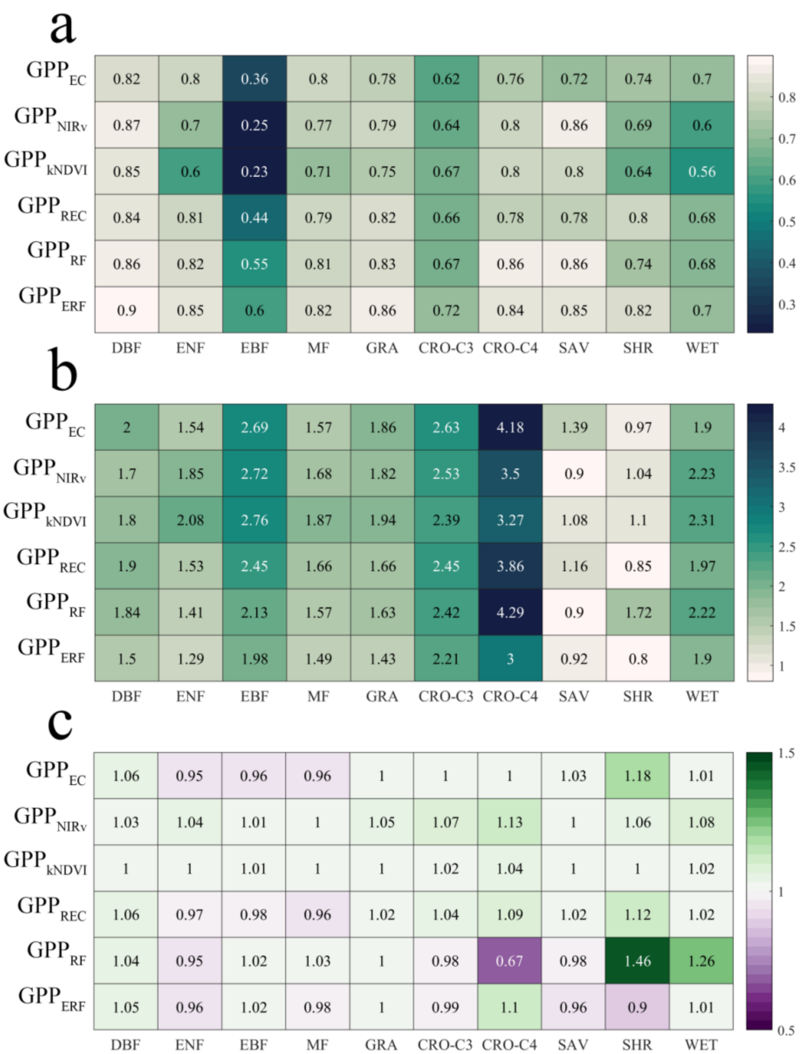
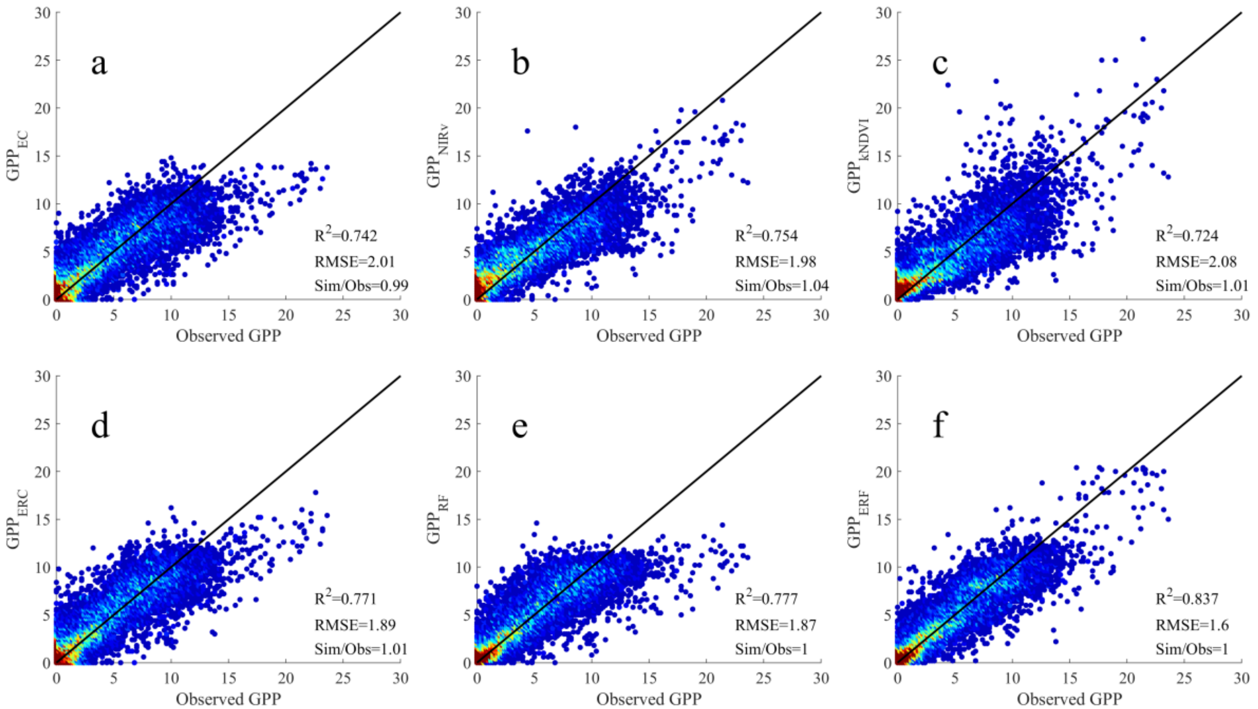


Figure 1. The performance of the ~~six~~eight models on different vegetation types. a, b and c represent R^2, RMSE, and Sim/Obs respectively.

~~We furtheralso counted the simulation performance of the different models at each site. As shown in Figure S3, we averaged the evaluation indicators of all sites and found that the accuracy of GPP_{ERF} was the highest, R^2 was 0.75, RMSE was 1.53 gC m^{-2} d^{-1}, Sim/Obs was also the closest to 1, which was 1.04.~~ Combining the results of all flux sites, GPP_{ERF} ~~could~~ explained ~~83.7~~85.1% of the monthly GPP variations, while the ~~five~~seven ~~remote sensing~~GPP models only explained ~~72.4~~67.7%-~~77.7~~81.5% of the monthly GPP variations~~variation in GPP~~ (Figure 2). In order to further prove the robustness of the ERF model, we also used GPP models with original parameters for modeling and validation. As shown in Figure S3, the performance of these GPP models decreased significantly, with R^2 ranging from 0.570 to 0.719 and RMSE ranging from 2.29 to 3.81 gC m^{-2} d^{-1}. The phenomenon of "high underestimation" and "low overestimation" was also pronounced~~serious~~. However, the ERF model maintained~~showed~~ a consistent advantage, with R^2 significantly higher than other GPP models

(0.856). In addition, we tested the effect of the amount of GPP on the accuracy of the ERF model. As shown in Table S8, as the number of GPP in the ERF model increase~~s~~d, the performance gain of the model gradually decrease~~s~~d.



~~**Figure 2.** Comparison between the GPP simulations of the six models and the GPP observations. a-f represents GPP_{EC}, GPP_{NIRv}, GPP_{kNDVI}, GPP_{REC}, GPP_{RF}, GPP_{ERF}, respectively.~~

In summary~~Overall~~, GPP_{ERF} showed~~exhibited~~ high accuracy in terms of~~ site scale,~~ vegetation type~~,~~ and the ~~the~~ ability to interpret monthly variations in GPP, which also illustrate~~d~~s the potential of ~~machine learning-based ensemble~~ERF model~~s~~ to improve ~~in improving~~ GPP estimation. However, it was observed~~we also found~~ that most ~~of the~~ GPP simulations ~~have~~exhibited the phenomenon of “high value underestimation and low value overestimation~~overestimate~~”. For example, GPP_{EC}, GPP_{REC}~~,~~, ~~and~~ GPP_{MODIS}~~,~~ and GPP_{RF} showed obvious underestimation in the months when the monthly GPP value surpassed~~was above~~~~greater than~~ ~~10~~ 15 gC m^{-2} d^{-1} (Figure 2)~~,~~. Therefore, it is ~~therefore~~ necessary to evaluate the performance of different models in each month and ~~in~~ different subvalues.

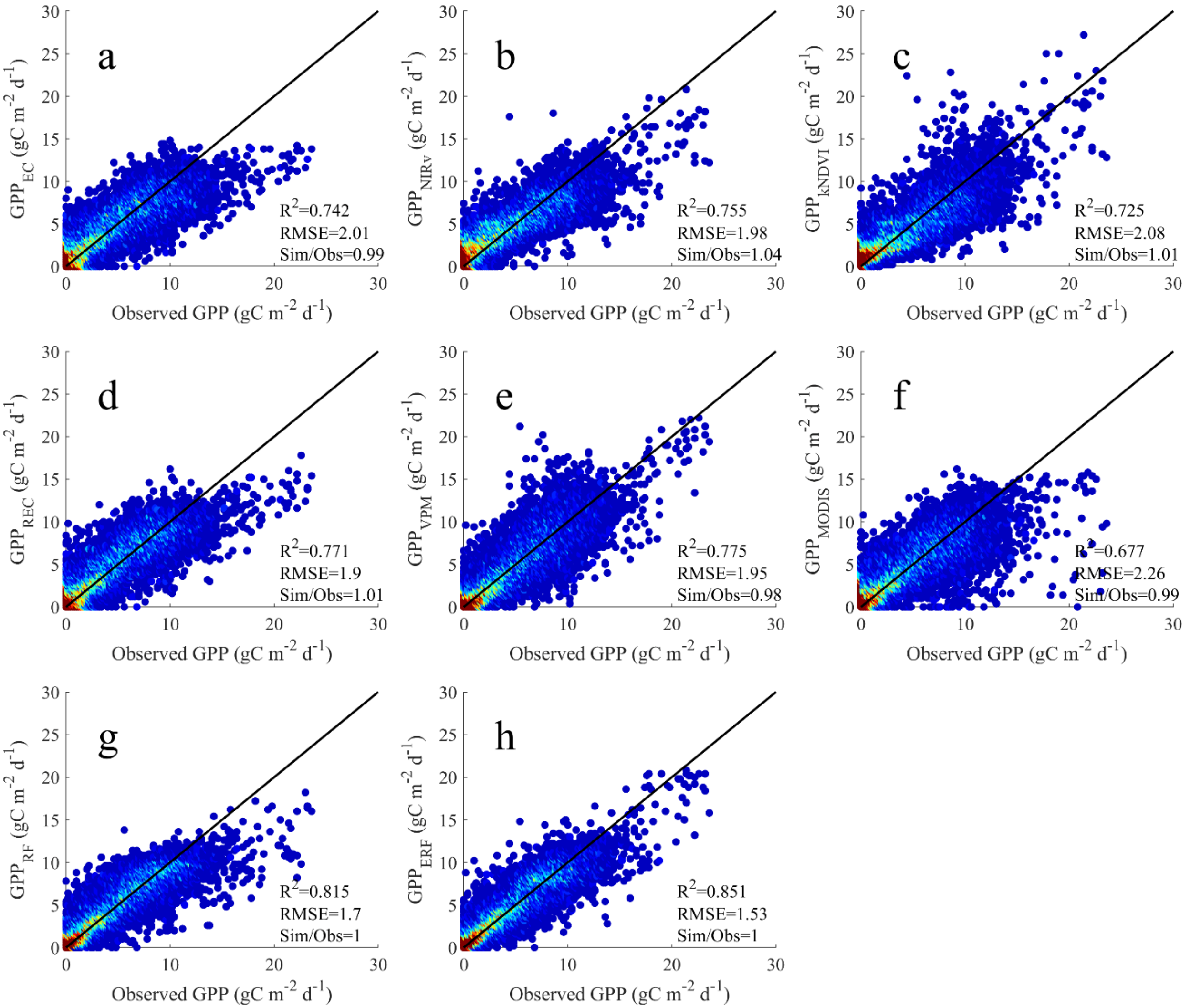
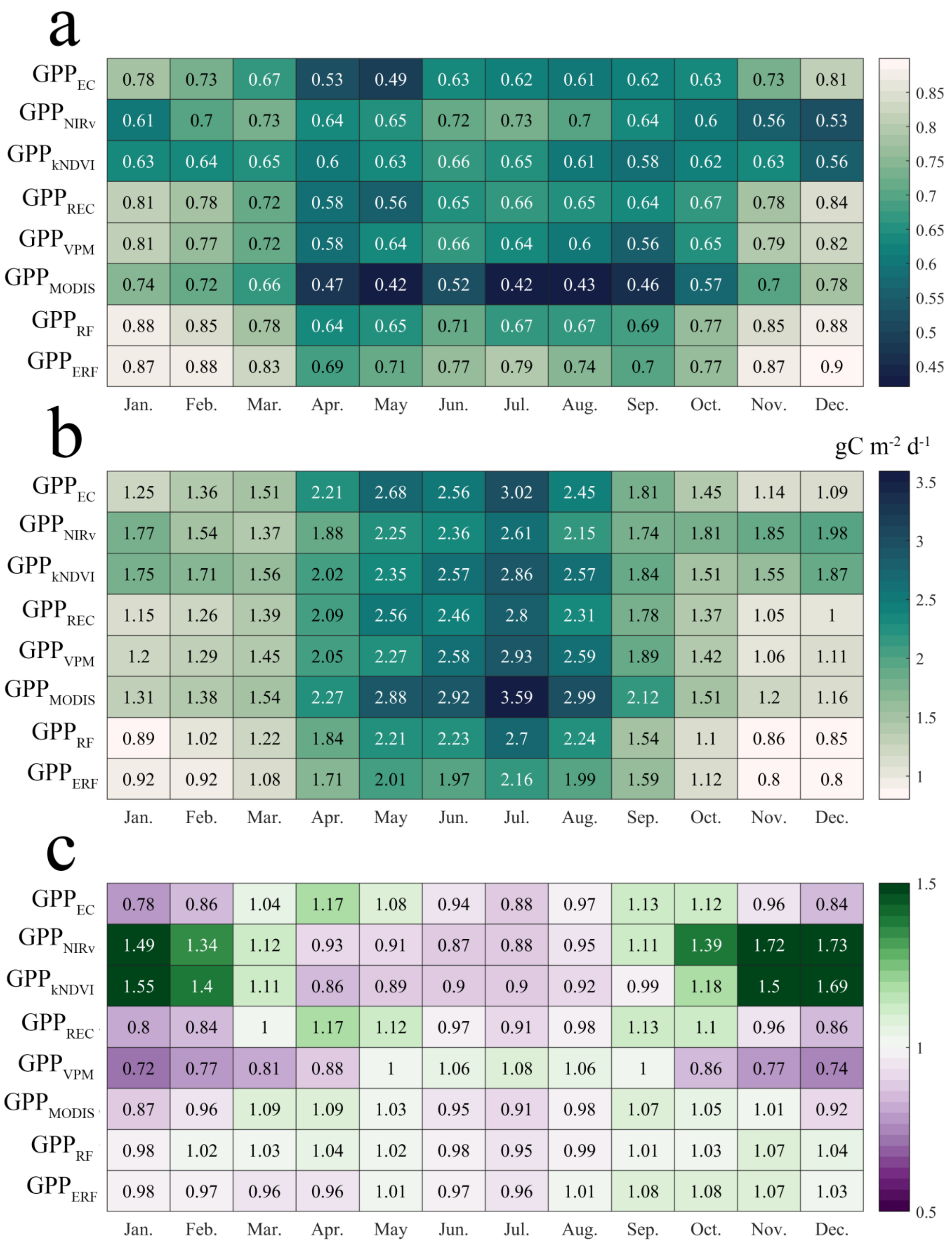


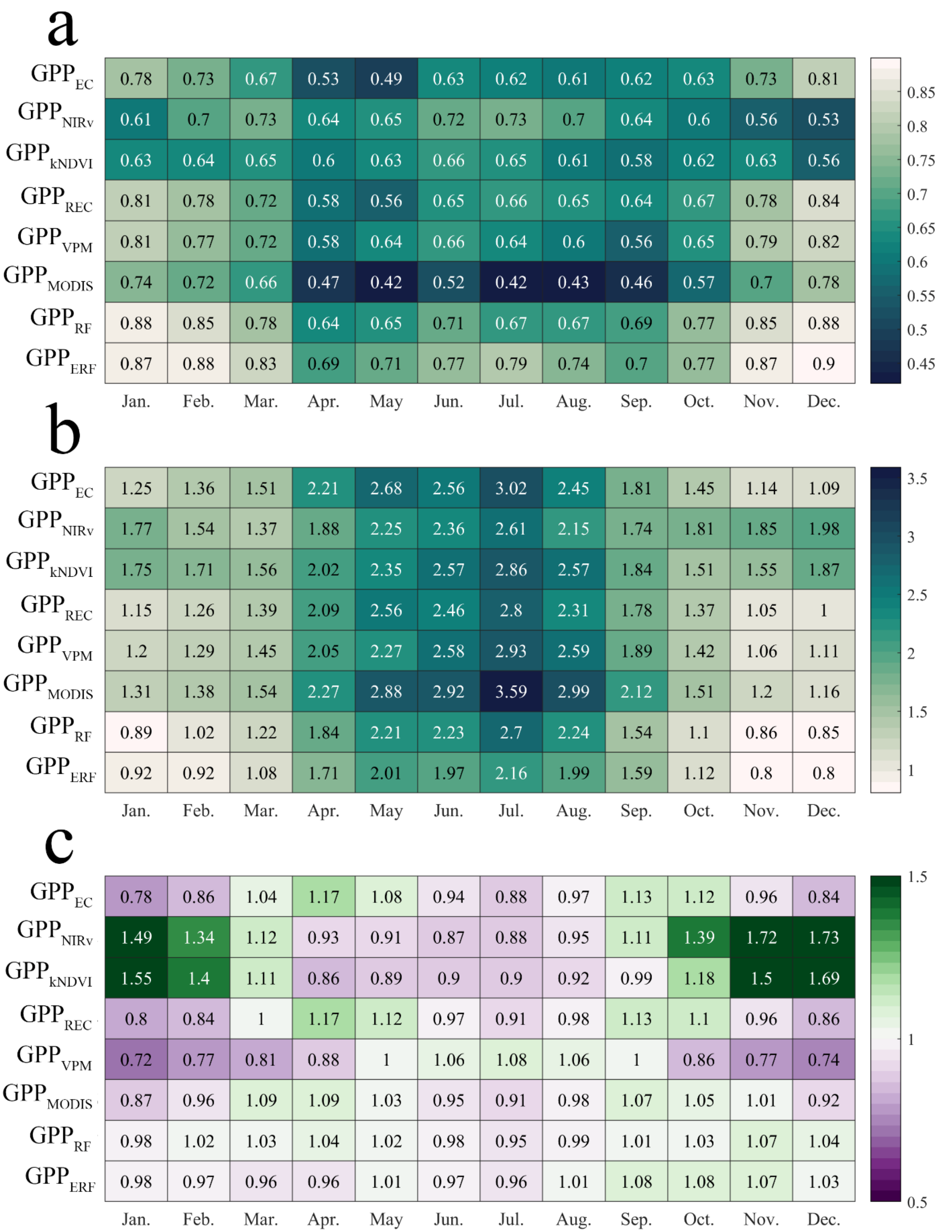
Figure 2. Comparison between the GPP simulations of the ~~six~~eight models and the GPP observations. a-~~f~~ h represents GPP_{EC}, GPP_{NIRv}, GPP_{kNDVI}, GPP_{REC}, GPP_{VPM}, GPP_{MODIS}, GPP_{RF}, GPP_{ERF}, respectively.

3.2 Performance of ~~six~~GPP models in each month and different subvalues

Figure 3 shows the simulation accuracy of the eight ~~six~~ models in each month. The ERF model maintained a higher accuracy than other GPP models, with GPP_{ERF} consistently achieving higher R^2 and lower RMSE in most months, and no evident instances of "high value underestimation and low value overestimation". In contrast, the accuracy of other~~the~~ ~~remote sensing~~GPP models was less satisfactory accuracy~~not satisfactory~~, especially during~~for~~ winter (most flux sites are concentrated in the Northern Hemisphere~~northern hemisphere~~), the LUE models tended to underestimate~~d~~ ~~the~~ GPP~~ per month~~, and the Sim/Obs remained at 0.~~78~~72-~~0.96~~1.01, although~~but~~ R^2 values were~~was~~ above 0.7~~,~~. Meanwhile,~~while~~ the vegetation index models overestimated GPP, Sim/Obs remained at 1.34-1.73, and R^2 values were~~was~~ relatively low, mostly around 0.6.



a
GPP_EC 0.78 0.73 0.67 0.53 0.49 0.63 0.62 0.61 0.62 0.63 0.73 0.81
GPP_NIRv 0.61 0.7 0.73 0.64 0.65 0.72 0.73 0.7 0.64 0.6 0.56 0.53
GPP_kNDVI 0.63 0.64 0.65 0.6 0.63 0.66 0.65 0.61 0.58 0.62 0.63 0.56
GPP_REC 0.81 0.78 0.72 0.58 0.56 0.65 0.66 0.65 0.64 0.67 0.78 0.84
GPP_VPM 0.81 0.77 0.72 0.58 0.64 0.66 0.64 0.6 0.56 0.65 0.79 0.82
GPP_MODIS 0.74 0.72 0.66 0.47 0.42 0.52 0.42 0.43 0.46 0.57 0.7 0.78
GPP_RF 0.88 0.85 0.78 0.64 0.65 0.71 0.67 0.67 0.69 0.77 0.85 0.88
GPP_ERF 0.87 0.88 0.83 0.69 0.71 0.77 0.79 0.74 0.7 0.77 0.87 0.9
Jan. Feb. Mar. Apr. May Jun. Jul. Aug. Sep. Oct. Nov. Dec.
0.85 0.8 0.75 0.7 0.65 0.6 0.55 0.5 0.45
b
gC m-2 d-1
GPP_EC 1.25 1.36 1.51 2.21 2.68 2.56 3.02 2.45 1.81 1.45 1.14 1.09
GPP_NIRv 1.77 1.54 1.37 1.88 2.25 2.36 2.61 2.15 1.74 1.81 1.85 1.98
GPP_kNDVI 1.75 1.71 1.56 2.02 2.35 2.57 2.86 2.57 1.84 1.51 1.55 1.87
GPP_REC 1.15 1.26 1.39 2.09 2.56 2.46 2.8 2.31 1.78 1.37 1.05 1
GPP_VPM 1.2 1.29 1.45 2.05 2.27 2.58 2.93 2.59 1.89 1.42 1.06 1.11
GPP_MODIS 1.31 1.38 1.54 2.27 2.88 2.92 3.59 2.99 2.12 1.51 1.2 1.16
GPP_RF 0.89 1.02 1.22 1.84 2.21 2.23 2.7 2.24 1.54 1.1 0.86 0.85
GPP_ERF 0.92 0.92 1.08 1.71 2.01 1.97 2.16 1.99 1.59 1.12 0.8 0.8
Jan. Feb. Mar. Apr. May Jun. Jul. Aug. Sep. Oct. Nov. Dec.
3.5 3 2.5 2 1.5 1
c
GPP_EC 0.78 0.86 1.04 1.17 1.08 0.94 0.88 0.97 1.13 1.12 0.96 0.84
GPP_NIRv 1.49 1.34 1.12 0.93 0.91 0.87 0.88 0.95 1.11 1.39 1.72 1.73
GPP_kNDVI 1.55 1.4 1.11 0.86 0.89 0.9 0.9 0.92 0.99 1.18 1.5 1.69
GPP_REC 0.8 0.84 1 1.17 1.12 0.97 0.91 0.98 1.13 1.1 0.96 0.86
GPP_VPM 0.72 0.77 0.81 0.88 1 1.06 1.08 1.06 1 0.86 0.77 0.74
GPP_MODIS 0.87 0.96 1.09 1.09 1.03 0.95 0.91 0.98 1.07 1.05 1.01 0.92
GPP_RF 0.98 1.02 1.03 1.04 1.02 0.98 0.95 0.99 1.01 1.03 1.07 1.04
GPP_ERF 0.98 0.97 0.96 0.96 1.01 0.97 0.96 1.01 1.08 1.08 1.07 1.03
Jan. Feb. Mar. Apr. May Jun. Jul. Aug. Sep. Oct. Nov. Dec.
1.5 1 0.5


a
GPP EC: 0.78 0.73 0.67 0.53 0.49 0.63 0.62 0.61 0.62 0.63 0.73 0.81
GPP NIRv: 0.61 0.7 0.73 0.64 0.65 0.72 0.73 0.7 0.64 0.6 0.56 0.53
GPP kNDVI: 0.63 0.64 0.65 0.6 0.63 0.66 0.65 0.61 0.58 0.62 0.63 0.56
GPP REC: 0.81 0.78 0.72 0.58 0.56 0.65 0.66 0.65 0.64 0.67 0.78 0.84
GPP VPM: 0.81 0.77 0.72 0.58 0.64 0.66 0.64 0.6 0.56 0.65 0.79 0.82
GPP MODIS: 0.74 0.72 0.66 0.47 0.42 0.52 0.42 0.43 0.46 0.57 0.7 0.78
GPP RF: 0.88 0.85 0.78 0.64 0.65 0.71 0.67 0.67 0.69 0.77 0.85 0.88
GPP ERF: 0.87 0.88 0.83 0.69 0.71 0.77 0.79 0.74 0.7 0.77 0.87 0.9
Jan. Feb. Mar. Apr. May Jun. Jul. Aug. Sep. Oct. Nov. Dec.
0.85 0.8 0.75 0.7 0.65 0.6 0.55 0.5 0.45
b
GPP EC: 1.25 1.36 1.51 2.21 2.68 2.56 3.02 2.45 1.81 1.45 1.14 1.09
GPP NIRv: 1.77 1.54 1.37 1.88 2.25 2.36 2.61 2.15 1.74 1.81 1.85 1.98
GPP kNDVI: 1.75 1.71 1.56 2.02 2.35 2.57 2.86 2.57 1.84 1.51 1.55 1.87
GPP REC: 1.15 1.26 1.39 2.09 2.56 2.46 2.8 2.31 1.78 1.37 1.05 1
GPP VPM: 1.2 1.29 1.45 2.05 2.27 2.58 2.93 2.59 1.89 1.42 1.06 1.11
GPP MODIS: 1.31 1.38 1.54 2.27 2.88 2.92 3.59 2.99 2.12 1.51 1.2 1.16
GPP RF: 0.89 1.02 1.22 1.84 2.21 2.23 2.7 2.24 1.54 1.1 0.86 0.85
GPP ERF: 0.92 0.92 1.08 1.71 2.01 1.97 2.16 1.99 1.59 1.12 0.8 0.8
Jan. Feb. Mar. Apr. May Jun. Jul. Aug. Sep. Oct. Nov. Dec.
3.5 3 2.5 2 1.5 1
c
GPP EC: 0.78 0.86 1.04 1.17 1.08 0.94 0.88 0.97 1.13 1.12 0.96 0.84
GPP NIRv: 1.49 1.34 1.12 0.93 0.91 0.87 0.88 0.95 1.11 1.39 1.72 1.73
GPP kNDVI: 1.55 1.4 1.11 0.86 0.89 0.9 0.9 0.92 0.99 1.18 1.5 1.69
GPP REC: 0.8 0.84 1 1.17 1.12 0.97 0.91 0.98 1.13 1.1 0.96 0.86
GPP VPM: 0.72 0.77 0.81 0.88 1 1.06 1.08 1.06 1 0.86 0.77 0.74
GPP MODIS: 0.87 0.96 1.09 1.09 1.03 0.95 0.91 0.98 1.07 1.05 1.01 0.92
GPP RF: 0.98 1.02 1.03 1.04 1.02 0.98 0.95 0.99 1.01 1.03 1.07 1.04
GPP ERF: 0.98 0.97 0.96 0.96 1.01 0.97 0.96 1.01 1.08 1.08 1.07 1.03
Jan. Feb. Mar. Apr. May Jun. Jul. Aug. Sep. Oct. Nov. Dec.
1.5 1 0.5

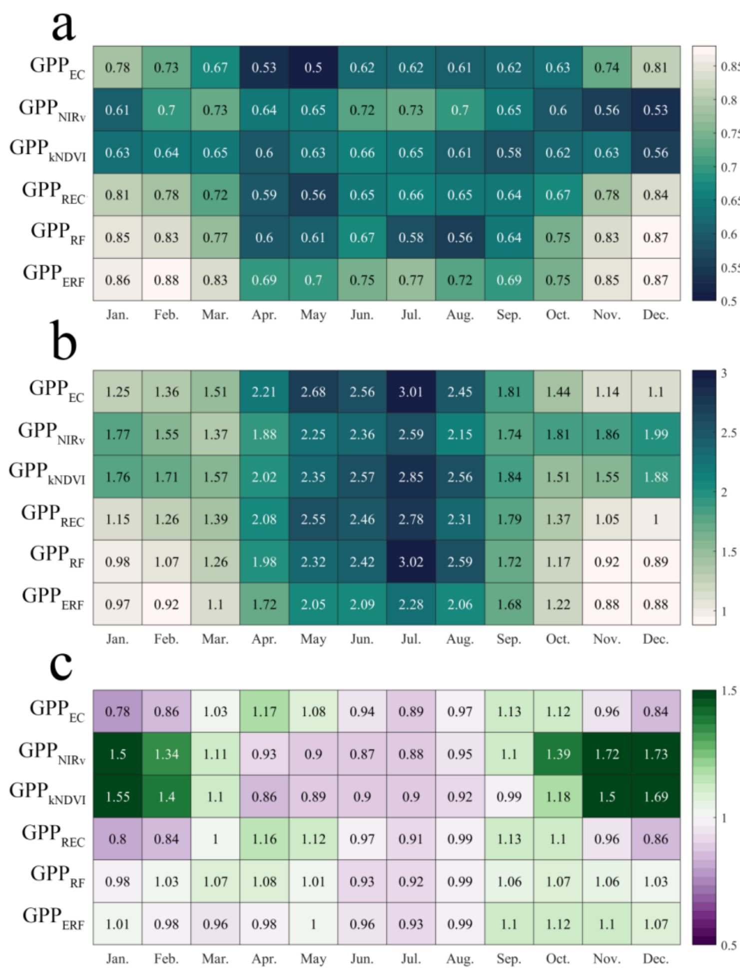
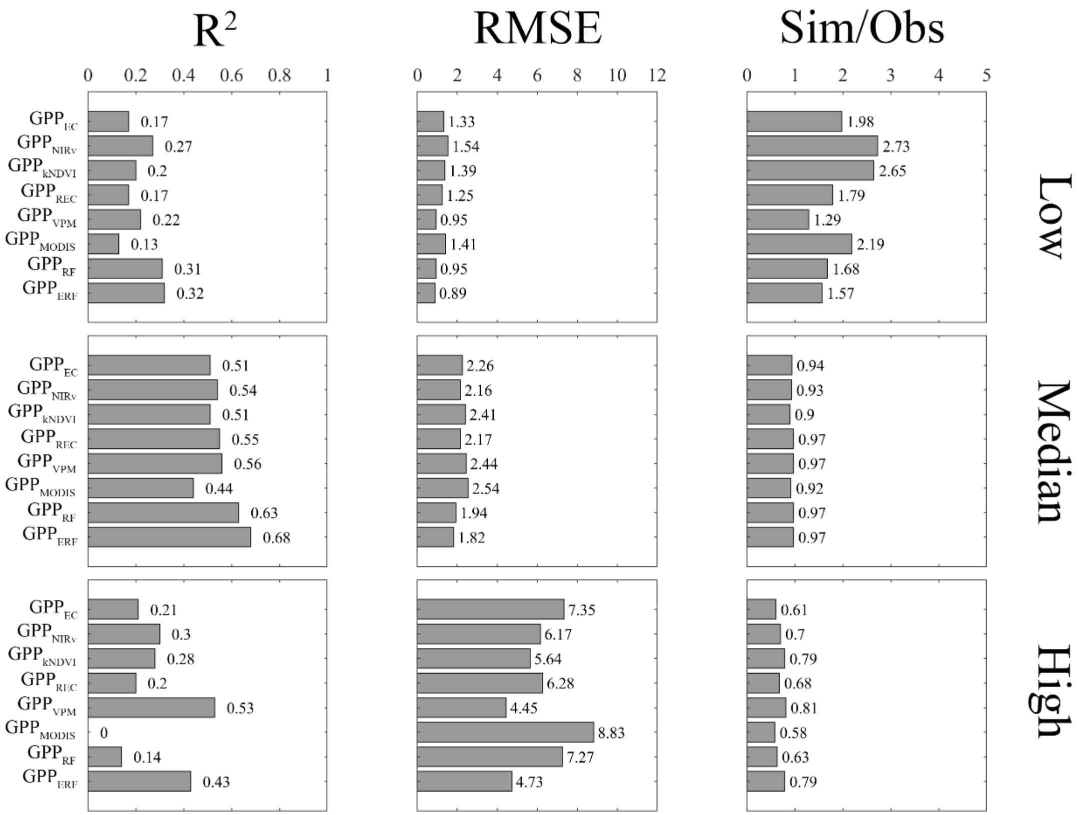


Figure 3. Performance of the ~~six~~ eight models in each month. a, b and c represent R^2, RMSE, and Sim/Obs respectively.

We compared the performance of all models in different subvalues, including high value (GPP > ~~10~~ 15 gC m^{-2} d^{-1}), median value (~~10~~ 15 gC m^{-2} d^{-1} > GPP > 2 gC m^{-2} d^{-1}), low value (GPP < 2 gC m^{-2} d^{-1}). For~~In the~~ extreme values, ~~all~~ most models performed poorly (Figure 4), ~~the~~with R^2 ~~of the~~for ~~remote sensing~~GPP models falling~~was~~ ~~all~~ below 0.3, and only GPP_{VPM} showing better performance in the high-value range.~~while in the high value was relatively good.~~ GPP_{ERF} demonstrated~~showed~~ some improvement in both low and high values, with R^2 ~~was~~ 0.32 and 0.43, RMSE ~~was~~ of 0.89 and 4.73 gC m^{-2} d^{-1}, and Sim/Obs ~~was~~ closer to 1, respectively.~~GPP_{ERF} showed a more significantobvious improvement in the high value, R^2 increased to 0.3843, the RMSE decreased to 3.03 gC m^{-2} d^{-1}, Sim/Obs also increased to 0.82, and only a slight improvement in the low value.~~ In the median value range, all models performed well, with no~~without~~ significant~~serious~~ bias in the GPP estimateion~~GPP estimation biases~~. The R^2 of ~~the remote sensing~~GPP models ranged from 0.44 to 0.68~~was between 0.434 and 0.68~~, and the RMSE remained between 1.~~71~~82 and 2.~~1~~54 gC m^{-2} d^{-1}. Further analysis~~presentations~~ were~~as~~ made at two typical sites, it ~~i~~was obvious that GPP_{EC}, GPP_{REC} and GPP_{MODIS} on CN-Qia exhibited~~showed~~ obvious underestimation during the growing season (Figure S4). On CH_Lae, GPP_{kNDVI} and GPP_{VPM} ~~are~~were significantly overestimated (Figure S5). In contrast, at both sites, GPP_{ERF} ~~i~~was more consistent with observations, indicating~~meaning~~ that the superior~~good~~ performance of GPP_{ERF} ~~i~~was due to the correction on the time series (although not perfectly corrected~~it iwas not well corrected~~ at all sites).~~It could be seen that there was a large deviation in the estimation of the existing remote sensing~~GPP ~~model~~s deviated greatly ~~in the GPP extreme value, and the estimation in the median value was relatively good, while the ensemble model based on the machine learning method~~ERF model ~~could improve the simulation accuracy of~~ extreme value~~high value, which was of great significance for accurately estimating the annual values and inter-annual variation of GPP in terrestrial ecosystems.~~


R2
RMSE
Sim/Obs
0 0.2 0.4 0.6 0.8 1
0 2 4 6 8 10 12
0 1 2 3 4 5
GPPEC
GPPNIRv
GPPkNDVI
GPPREC
GPPVPM
GPPMODIS
GPPRF
GPPERF
0.17 0.27 0.2 0.17 0.22 0.13 0.31 0.32
1.33 1.54 1.39 1.25 0.95 1.41 0.95 0.89
1.98 2.73 2.65 1.79 1.29 2.19 1.68 1.57
Low
0.51 0.54 0.51 0.55 0.56 0.44 0.63 0.68
2.26 2.16 2.41 2.17 2.44 2.54 1.94 1.82
0.94 0.93 0.9 0.97 0.97 0.92 0.97 0.97
Median
0.21 0.3 0.28 0.2 0.53 0 0.14 0.43
7.35 6.17 5.64 6.28 4.45 8.83 7.27 4.73
0.61 0.7 0.79 0.68 0.81 0.58 0.63 0.79
High

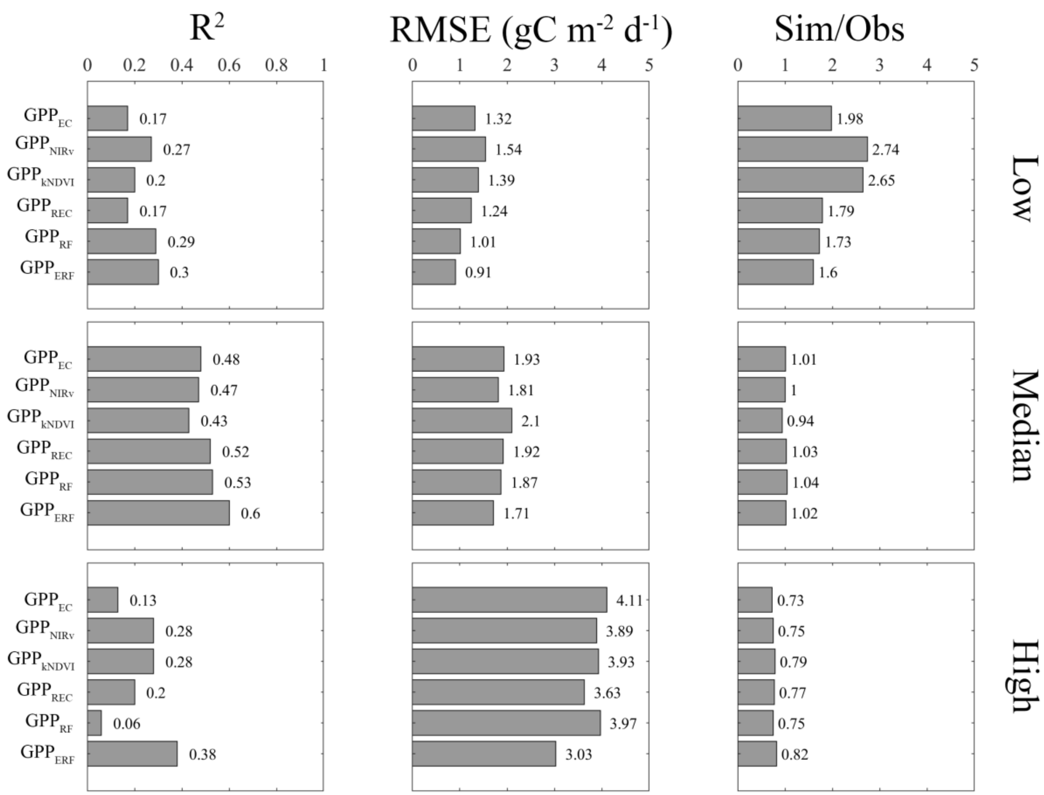


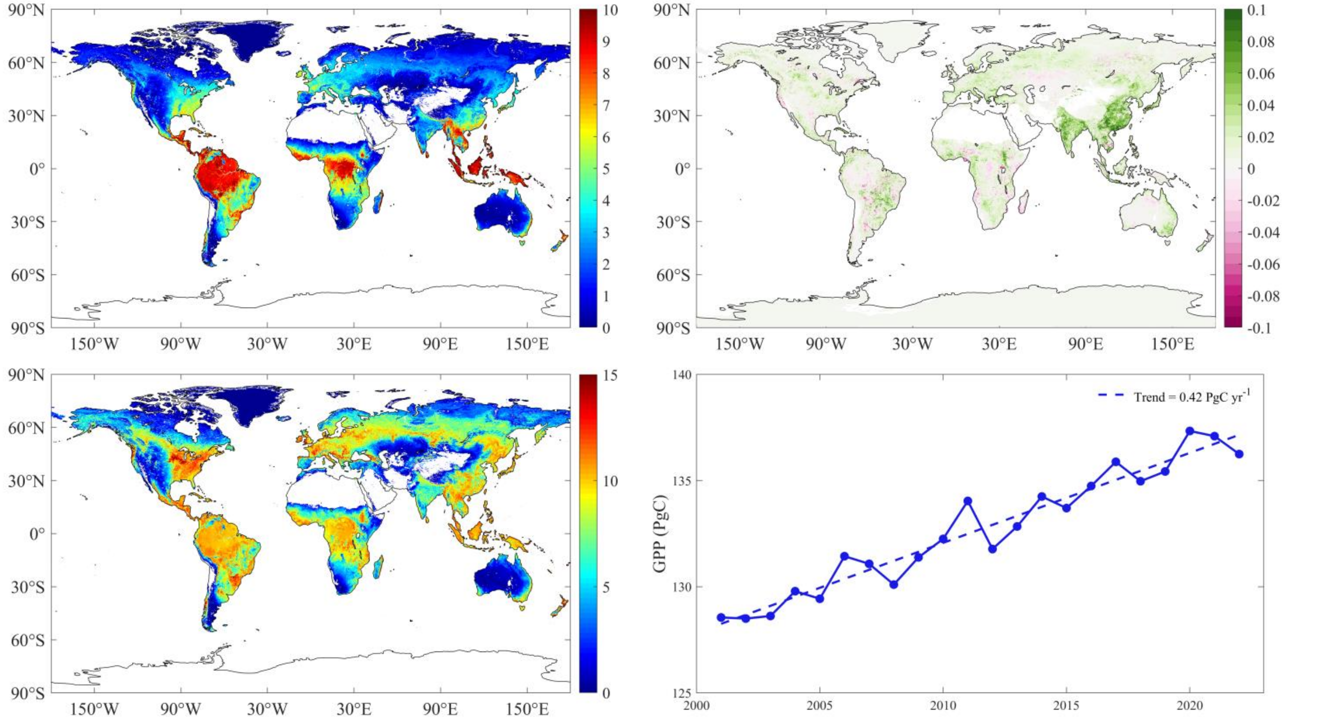
Figure 4. Performance of ~~six~~eight models in different subvalues.

3.3 Temporal and spatial characteristics of ERF GPP and its generalization evaluation~~Global GPP estimation based on ensemble model and its generalization evaluation~~

~~Based on remote sensing data and meteorological data, we estimated the global GPP from 2001 to 2022 using the ensemble model based on random forest~~ERF model~~.~~ Figure 5a shows the spatial distribution of the multi-year average of ERF_GPP. The high value of GPP was mainly concentrated in tropical areas, exceeding 10 gC m^{-2} d^{-1}, and relatively high in southeastern North America, Europe and southern China, about 4-6 gC m^{-2} d^{-1}. From 2001-2022, China and India showed the fastest increase in GPP, mostly at 0.1 gC m^{-2} d^{-1} (Figure 5b), similar to a previous study that reported that China and India led the global greening (Chen et al., 2019). We further investigate~~estimated~~d the annual maximum GPP, as shown in Figure 5c, and the North American corn belt was ~~by far~~ the global leader in GPP at more than 15 gC m^{-2} d^{-1}, compared to only 10 gC m^{-2} d^{-1}~~gC m-2 d-1~~ in most tropical forests. In 2001-2022, the global GPP was 13~~1~~2.~~2~~7 ± ~~3.1~~2.8 PgC yr^{-1}, ~~the~~with a trend ~~was~~of 0.4~~5~~2 PgC yr^{-2}~~;~~. ~~t~~The lowest value was 12~~6.4~~8.6 PgC yr^{-1} in 2001, and the highest value was 13~~5.9~~6.2 PgC yr^{-1} in 2020 (Figure 5d).

The results of the two uncertainty analyses consistently indicate~~show~~d that ERF_GPP exhibit~~presents~~ed a high uncertainty in ~~the~~ tropical regions (Figures S6 and S7), and the uncertainty of ERF_GPP~~ERF~~ caused by the number of GPP observations

~~i~~was relatively small, the standard deviation of 100 simulations ~~i~~was about 0.3 gC m^{-2} d^{-1} in the tropics and lower in other regions, below 0.1 gC m^{-2} d^{-1}. In contrast, the ERF_GPP~~ERF~~ caused by the number of features ~~i~~was much more uncertain, especially when~~if~~ the number of features ~~i~~was small. It is worth noting that when the number of features ~~i~~was 5, the uncertainty ~~i~~was already substantially less, and the standard deviation ~~i~~was generally lower than 0.5 gC m^{-2} d^{-1}.


90°N
60°N
30°N
0°
30°S
60°S
90°S
150°W
90°W
30°W
30°E
90°E
150°E
10
9
8
7
6
5
4
3
2
1
0
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
15
10
5
0
Trend = 0.42 PgC yr-1
GPP (PgC)
140
135
130
125
2000
2005
2010
2015
2020

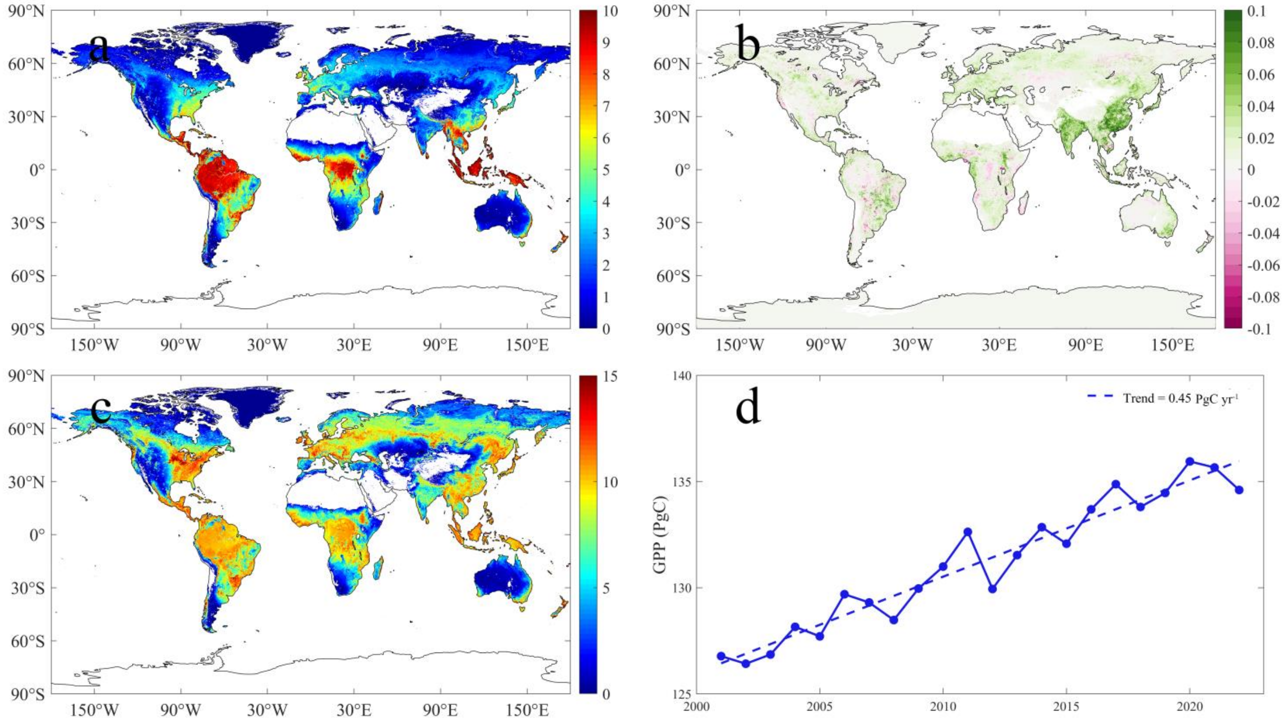
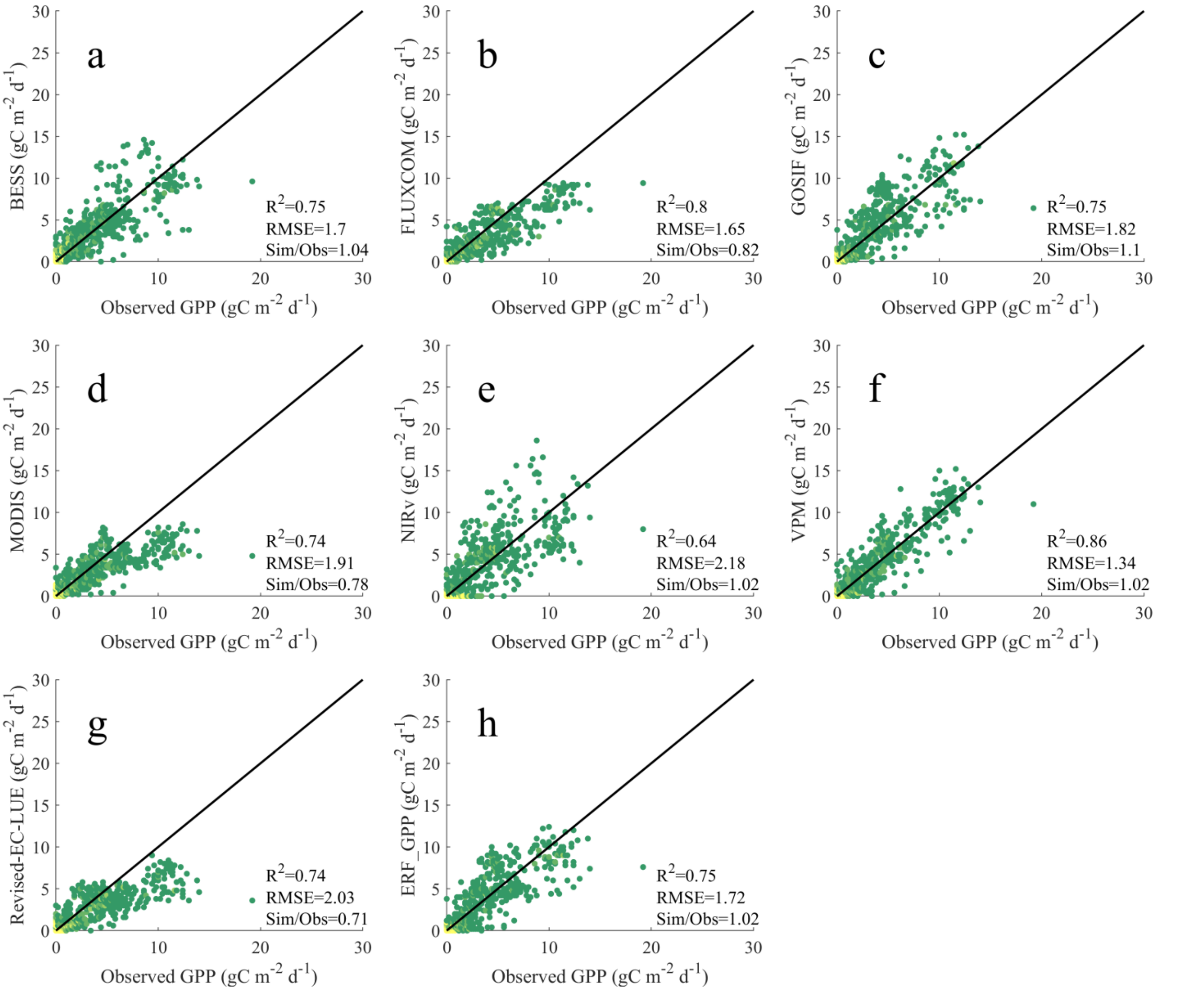


Figure 5. Spatial distribution and interannual change of ERF_GPP during 2001-2022. a represents the multi-year average, b represents the trend, c represents the annual maximum, and d represents the interannual change of GPP.

As shown in Figure 6, ~~the generalizations of~~ ERF_GPP and other GPP datasets were validated using GPP observations from ChinaFlux. Among~~Of~~ all the models, GPP_{VPM} demonstrated~~has~~ the best performance, with R^2 of 0.86 and RMSE of 1.34 gC m^{-2} d^{-1}.~~Overall, in China,~~ ERF_GPP also exhibited ~~has~~d ~~a~~ high generalization, with R^2 of 0.75, RMSE of 1.7~~5~~2 gC m^{-2} d^{-1}, there was no "high value underestimation and low value overestimation", which was comparable to the ~~simulation~~ accuracy of BESS~~, ECGC~~ and GOSIF. However, the simulation accuracy of the other ~~two~~ GPP datasets in Chinaflux was relatively poor, with the R^2 of NIRv being only 0.64, while FLUXCOM, MODIS and ~~the~~ Revised EC-LUE w~~as~~ere significantly underestimated, with the Sim/Obs being only 0.71-0.82~~0.71~~. In the validation of FLUXNET, the R^2 of FLUXCOM, MODIS, and Revised EC-LUE ranged from 0.57 to 0.67, and the RMSE ranged from 2.67 to 3.3 gC m^{-2} d^{-1}, and exhibit~~showed~~ different degrees of underestimation (Figure S~~3~~8). Other GPP datasets demonstrated~~showed~~ similar performance, with ERF_GPP being the best (R^2 = 0.74, RMSE = 2.26 gC m^{-2} d^{-1}).~~ Notably, in the high values, all models exhibited significant underestimation, which may be caused by the 0.05 °resolution being inconsistent with the flux tower footprint.~~

~~We further examined the different GPP datasets at each site, similar to the results at all sites, the ERF_GPP was relatively robust, with~~ ~~R^2~~ ~~and RMSE of 0.77 and 1.49 gC~~ ~~m^{-2} d^{-1}~~~~, respectively (Figure S4). The~~ ~~R^2~~ ~~of NIRv and Revised EC-LUE was~~

~~0.68 and 0.69, and Revised EC-LUE also showed a significant underestimate (Sim/Obs at 0.66). It should be noted that from the perspective of the average simulation accuracy of each site, BESS seemed to overestimate the GPP (Sim/Obs at 1.2).~~


a
BESS (gC m-2 d-1)
R2=0.75
RMSE=1.7
Sim/Obs=1.04
b
FLUXCOM (gC m-2 d-1)
R2=0.8
RMSE=1.65
Sim/Obs=0.82
c
GOSIF (gC m-2 d-1)
R2=0.75
RMSE=1.82
Sim/Obs=1.1
d
MODIS (gC m-2 d-1)
R2=0.74
RMSE=1.91
Sim/Obs=0.78
e
NIRv (gC m-2 d-1)
R2=0.64
RMSE=2.18
Sim/Obs=1.02
f
VPM (gC m-2 d-1)
R2=0.86
RMSE=1.34
Sim/Obs=1.02
g
Revised-EC-LUE (gC m-2 d-1)
R2=0.74
RMSE=2.03
Sim/Obs=0.71
h
ERF_GPP (gC m-2 d-1)
R2=0.75
RMSE=1.72
Sim/Obs=1.02
Observed GPP (gC m-2 d-1)

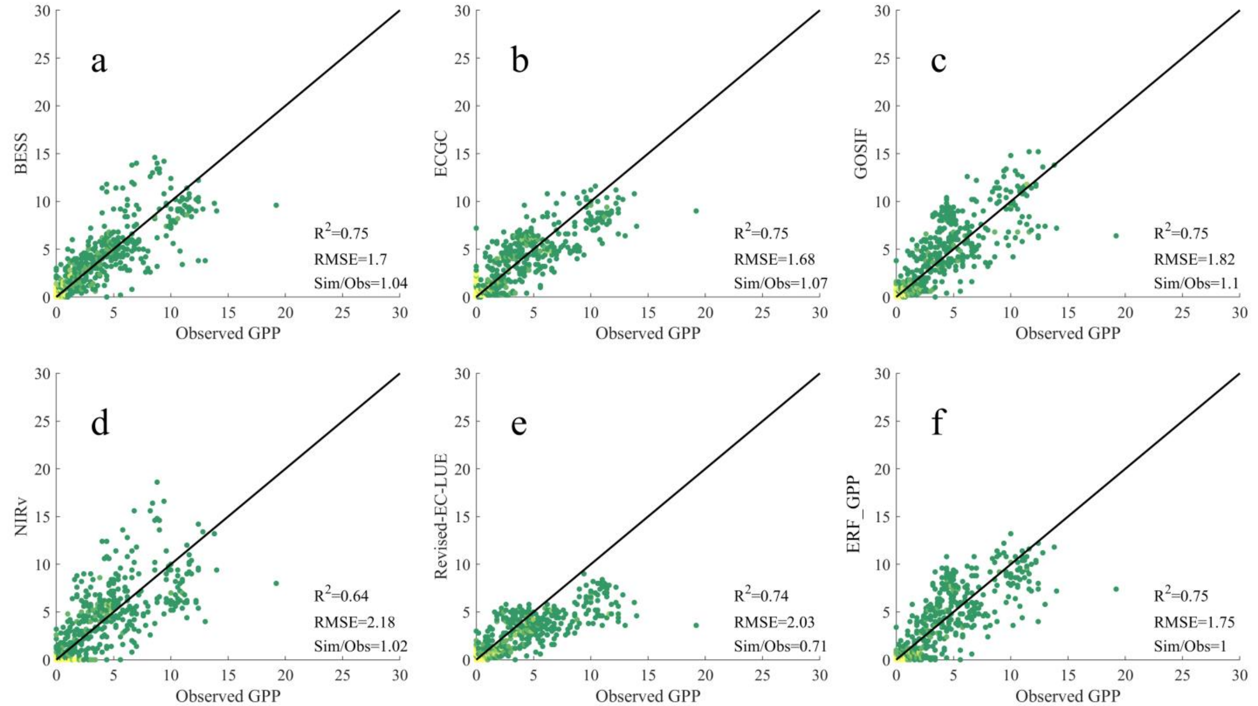


Figure 6. Comparison between the GPP datasets and the GPP observations from ChinaFlux. a-h represents BESS, FLUXCOM, GOSIF, MODIS, NIRv, VPM, Revise-EC-LUE, ERF_GPP, respectively.

4 Discussion

4.1 Performance analysis of different models

After parameter calibration, both LUE and vegetation index models obtained reliable model accuracy~~,~~. However, noticeable errors persist in different months and subvalues, indicating the prevalent phenomenon of "high value underestimation and low value overestimation".~~but there are still obvious errors in different months and different sub-values, that is, the phenomenon of "high value underestimation and low value overestimation" generally exists~~ (Figures 1-4). ~~With the continuous development of remote sensing technology and carbon cycle models, the existing models for estimating GPP are gradually increasing, including LUE models, process models, machine learning models and the newly developed vegetation index models (such as SIF, NIRV, KNDVI), these "big class" models also include many "small classes". For example, the differences in the meteorological constraintsenvironmental restriction function in the LUE model are extended to CASA, VPM, EC-LUE and other models . A recent study collected the response functions of GPP to different environmental variables, and under the LUE theory, 5600 LUE models could be generated ., Tthese different model structures greatly increase the uncertainty of global GPP estimation, which make people still confused about the annual value and inter-annual~~

~~trend of global GPP. All models can obtain a reliable model accuracy after calibrating the parameters, but there are still obvious errors in different vegetation types, different months and different sub-values (Figures 1-4).however none of the model accuracy is particularly outstanding, so it is urgent to provide a new method to further improve the accuracy of GPP estimation.~~

~~Multi-model ensemble may be a proven approach, and previous studies have shown that even simple multi-model average can reduce model estimation uncertainty. In this study, we used an ensemble modelsERF model to improve the estimation of GPP. Compared with the remote sensingother GPP models, the ensembleERF model could indeed show higher accuracy, the R^2 reached 0.83751, which is significantly higher than the accuracy of the machine learningRF model based on meteorological variables and remote sensing variables (R^2=0.777815). Since there are no physical constraints, machine learning models need to find the relationship between explanatory variables and target variables from a large amount of training data (such as GPP=f (LAI,T,P, etc.)) , so the reliability of the model usually depends on the representativeness of training data, such as LAI can explain GPP to a large extent., however, due to the complexity of the surrounding environment of flux sites, it is difficult to guarantee consistent modeling relationships even for the same vegetation type. The difference between ensemble models based on machine learningERF model lies in the differences in explanatory variables. These explanatory variables are the results of multiple model simulations, and these results are usually more representative and more consistent with the relationship between the target variables, which makes the GPP simulations more accurate.~~

~~The simulation results of different models in each months and different subvalues showed that the existing GPP estimation model widely existed the phenomenon of "high value underestimation and low value overestimate".~~ In addition to MODIS, the GPP simulated by the other three LUE models is generally underestimated in winter (Figure 3), which may be caused by biases in the parameters used in the meteorological constraints.~~For the LUE model, this phenomenon is most obvious in winter (Figure 3), and the GPP was underestimated by about 20%, which may be due to the deviation in the form of environmental factor.~~ In the expression form of the temperature constraint adopted by the LUE models, the maximum temperature, minimum temperature and optimum temperature for limiting photosynthesis are all constants, however these values may not be fixed (Huang et al., 2019; Grossiord et al., 2020). A ~~P~~previous study has demonstrated~~shown~~ that the GPP estimation~~estimation of GPP~~ could~~can~~ be effectively improved by using dynamic temperature parameters (Chang et al., 2021). Moreover~~In addition~~, the form of meteorological constraint is also an important influencing factor. Compared with other LUE models, VPM does not use VPD constraints, but incorporates~~uses~~ land surface water index from satellite observations as constraints (Xiao et al., 2004), ~~-~~which may be the reason why the model performs better than other models at high values.~~which makes the model perform better than other models at high values.~~ Conversely, ~~T~~the two vegetation index models overestimated GPP~~were overestimated~~ in winter, and even overestimated by 70% in December. The vegetation index model does not consider ~~the~~ meteorological constraints~~constraints of environmental factors~~. that ~~They~~ believe that all environmental impacts on vegetation have been included in the vegetation index (kNDVI, NIRv~~),~~). However, it is a fact that under high temperatures or low radiation, the vegetation index may still maintain the appearance of high photosynthesis (greening), while in fact the GPP is low (Doughty et al., 2021; Yang et al., 2018; Chen et al., 2024). Furthermore~~however,~~

~~this aspect is still controversial , and~~In addition, the relationship between these vegetation indices and GPP is not robust, and the vegetation indices based on reflectance may have hysteresis (Wang et al., 2022)~~, and our results also showed that only using vegetation indices modeling GPP should be carefully considered. In the low value and high value, the effects of all remote sensing models are not ideal, which may be caused by the model structure itself. Simple mathematical expressions cannot characterize the entire photosynthesis process, and these models are usually only empirical or semi-empirical. Although the ensemble model based on machine learning can improve this phenomenon to a certain extent, it still depends on the reliability of the remote sensing model in the extreme value. Therefore, we believe that in the future model development, it is necessary to focus on the simulation performance of GPP in the extreme value.~~.

Compared to other GPP models, the ERF model demonstrated~~showed~~ ~~good~~better performance (R^2 = 851). Since there are no physical constraints, the machine learning model needs to find the relationship between explanatory variables and target variable~~s~~ from a large amount of training data (such as GPP=f (LAI,T,P, etc.)). Therefore, the reliability of the model usually depends on the representativeness of the training data. For example, LAI can explain GPP to a large extent, while complex modeling relationships are still needed from LAI to GPP. The difference between the ERF model and the RF model lies in the~~is the difference in~~ explanatory variables. The ERF model leverages multiple GPP simulations that are more representative and aligned with the target variable~~These explanatory variables are the result of multiple model simulations that are generally more representative and more consistent with the relationship between the target variables~~, thus~~which~~ ~~makes~~making the GPP simulations more accurate. In other words, the ERF model does not need to take into account the uncertainties of the model structure (such as meteorological constraints) and model parameters (such as maximum light use efficiency), but rather focuses on the uncertainties inherent in the simulated GPP~~only the uncertainties of the simulated GPP~~.

To further clarify the impact of explanatory variables on the ERF model, we conducted a feature importance analysis (Figure S~~6~~9). From an average of 200 times, the results of the ERF model did not depend on a single GPP simulation. Even GPP_{MODIS}, with the highest relative importance, accounted for no more than 25%,~~which had the highest relative importance, was no more than 25%,~~ suggesting that the ERF model behaves more like a weighted average of multiple GPP simulations.~~so it looks more like a weighted average of multiple GPP simulations.~~

It is worth noting that in the study of Tian et al. (2023), the ERF model was also used to improve the GPP estimation. Our research extends this work in several ways.~~On this basis, our research is further extended.~~ Firstly, parameter calibration was carried out in our study so that the final validation results are~~were~~ comparable, that is, ~~the~~ differences in model performance ~~was~~are mainly due to the uncertainty of the model structure. Secondly, our ~~research~~study focuses on the phenomenon of "~~low value overestimation~~ high value underestimation and ~~high value underestimation~~low value overestimation" of ~~the~~ GPP model, with results indicating that the ERF model performed well across various vegetation types, months, and subvalues.~~and the research results show that the ensembleERF model hasd a good performance in different vegetation types, different months, and different subvalues.~~ Finally, we generated the ERF_GPP dataset ~~the ERF model was used to estimate the global GPP~~ and validate~~d on~~ it on different observational data-sets, ~~which~~ further confirming~~proves~~ the robustness of the ERF model in GPP estimation.

4.2 Robustness of ~~global GPP estimation based on ensemble model~~ERF_GPP

In this study, based on site-scale validation, we demonstrate the reliability of the ~~random forest-based ensemble~~ERF model in GPP estimation. However, further discussion is needed regarding the robustness of~~what needs to be discussed further is whether~~ the spatial distribution, spatial trends and global totals of ERF_GPP~~ are reliable~~. Since the current GPP datasets are generated based on remote sensing observation, there is a strong similarity in spatial distribution among all GPP datasets.~~all GPP datasets are very similar in spatial distribution.~~ Therefore, the validation of GPP observations independent of FLUXNET~~2015~~ is~~are very~~ crucial~~important~~. Validation results from GPP observations ~~from~~of ChinaFlux indicated~~show~~ that GPP_{ERF} exhibited~~showed~~ good generalization in China (R^2=0.75), which was slightly lower than the accuracy of~~ the~~ 5~~-~~-fold~~-~~-cross~~-~~-validation during modeling, possibly due to the mismatch between the 0.05 ° GPP and the footprint of the flux tower (Chu et al., 2021). In addition, the validation of FLUXNET further confirms~~shows~~ the reliability of ERF_GPP. Overall, ~~however, ~~this is comparable to or slightly better than the simulation accuracy of current mainstream GPP datasets. ~~In addition, w~~We also observed~~found~~ a clear improvement in the spatial maximum value of ERF_GPP in some corn growing regions, such as the North American Corn Belt (Figure 5c), which is supported by previous studies showing that C4 crops have much higher GPP peaks than other vegetation types (Yuan et al., 2015; Chen et al., 2011).

Due to the drought trend, the constraining~~constraint~~ effect of water on vegetation is gradually increasing, and some studies have reported the decoupling phenomenon of LAI and GPP under some specific conditions (Jiao et al., 2021; Hu et al., 2022). However, in China and India that two regions with significant greening, GPP ontinues to increase~~is still increasing~~ in most datasets, and ERF_GPP supports this view. This phenomenon may be due to the low drought pressure on ~~farm~~croplands in China and India due to irrigation, which poses less constraint on GPP~~is less of a constraint on GPP~~ (Ambika and Mishra, 2020; Ai et al., 2020). The global estimate of ERF_GPP ~~was~~is 132.7 ± 2.8~~131.2~~ PgC yr^{-1}, which is close to estimates from most previous studies (Wang et al., 2021; Badgley et al., 2019). ~~Some~~A stud~~ies~~y have suggested that the global GPP may reach 150-175 PgC yr^{-1} (Welp et al., 2011), however, there is no further evidence to support this view.

ERF_GPP exhibit~~show~~~~s~~ed high uncertainty in ~~the ~~tropical regions, similar reports have been made in previously published GPP datasets (Badgley et al., 2019; Guo et al., 2023). The scarcity of flux observations in these regions (Pastorello et al., 2020), coupled with the well-known issue of cloud pollution and saturation in remote sensing data in the tropics~~There are very few observations of flux in these regions, so both in terms of annual totals and long-term trends, and tropical regions are currently the most controversial areas in global GPP estimates. In addition, the problem of cloud pollution in remote sensing data in the tropics is well known~~ (Badgley et al., 2019), exacerbates the uncertainty in GPP estimates for these regions. Therefore, in future studies, on the one hand, more flux observations in tropical regions are needed, and on the other hand, attempts can be made to combine optical and microwave data to improve the estimation of GPP.~~which further exacerbates the uncertainty in GPP estimates for the regions.~~

4.3 Limitations and uncertainties

In this study, we improved GPP estimat~~es~~ion based on the ~~ensemble~~ERF model. However, there are still some limitations and uncertainties due to the availability of data and methods. First, C4 crop distribution maps were used in our study to improve estimates of cropland GPP. However, it is important to note that this dataset only represents the spatial distribution of crops around the year 2000, which may add uncertainty to GPP simulations of cropland in a few C3 and C4 alternating areas. Secondly~~,~~, the ERF model considers six GPP simulations~~only the GPP simulations of four remote sensing models were considered in our model~~, and it is not clear whether adding more GPP simulations to the model can further improve the GPP estimation. Finally, our model did not consider the effect of soil moisture on GPP, and some previous studies have highlighted the importance of incorporating~~considering~~ soil moisture in GPP estimates, especially for dry years (Stocker et al., 2019; Stocker et al., 2018).

5 Conclusion

In this study, we compared the performance of the ERF model with other GPP models at the site scale, especially for the phenomenon of "high value underestimation and low value overestimation", and further developed the ERF_GPP dataset.~~In this study, we evaluated the performance of five remote sensing models and one~~an ~~ensemble model to simulate GPP.~~ Overall, GPP$_{ERF}$ had higher model accuracy, explaining 8~~3.7~~5.1% of the monthly GPP variations~~ in GPP~~, and demonstrated~~showed~~ reliable ~~good~~ accuracy in different months, vegetation types and subvalues~~different vegetation types, different months and different extreme regions~~. Over the period from 2001 to 2022, the global estimate of ERF_GPP was 132.7 ± 2.8 PgC yr^{-1}, corresponding to a trend of 0.42 PgC yr^{-2}.~~The global GPP of ERF_GPP for 2001-2022 is 131.2 PgC yr^{-1}.~~ ~~The~~ Validation results from ChinaFlux indicate~~show~~d that ERF_GPP ha~~s~~d good generalization. For the current emerging GPP estimation models, ~~machine learning-based ensemble models~~ERF model provides an alternative GPP estimation method that lead to better model accuracy~~ another method of GPP estimation, and this may lead to higher model accuracy and more reliable global GPP estimation~~.

Data and code availability

The ~~global GPP dataset based on the ensemble model~~ERF_GPP for 2001-2022 is available at https://doi.org/10.6084/m9.figshare.24417649~~10.6084/m9.figshare.24417649~~ (Chen et al., 2023). The spatial resolution of ERF_GPP is 0.05° and the temporal resolution is monthly. Code is available from the author upon reasonable request.

Author contributions

X.C. and T.X.C. conceived the scientific ideas and designed this research framework. X.C. compiled the data, conducted 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 suggestions for improving the manuscript.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (No. 42130506, 42161144003 and 31570464) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX23_1322).

Declaration of interests

The authors have not disclosed any competing interests.

References

Ai, Z., Wang, Q., Yang, Y., Manevski, K., Yi, S., and Zhao, X.: Variation of gross primary production, evapotranspiration and water use efficiency for global croplands, Agricultural and Forest Meteorology, 287, 10.1016/j.agrformet.2020.107935, 2020.

Ambika, A. K. and Mishra, V.: Substantial decline in atmospheric aridity due to irrigation in India, Environmental Research Letters, 15, 10.1088/1748-9326/abc8bc, 2020.

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N. C., Peylin, P., Piao, S., Sitch, S., Viovy, N., Wiltshire, A., and Zhao, M.: Spatiotemporal patterns of terrestrial gross primary production: A review, Reviews of Geophysics, 53, 785-818, 10.1002/2015rg000483, 2015.

Badgley, G., Field, C. B., and Berry, J. A.: Canopy near-infrared reflectance and terrestrial photosynthesis, Science advances, 3, e1602244, 2017.

Badgley, G., Anderegg, L. D., Berry, J. A., and Field, C. B.: Terrestrial gross primary production: Using NIRV to scale from site to globe, Global change biology, 25, 3731-3740, 2019.

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 based ensemble approaches to improve evapotranspiration estimates from croplands across a wide environmental gradient, Agricultural and Forest Meteorology, 298, 108308, 2021.

Belgiu, M. and Drăguţ, L.: Random forest in remote sensing: A review of applications and future directions, ISPRS journal of photogrammetry and remote sensing, 114, 24-31, 2016.

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., and Guanter, L.: A unified vegetation index for quantifying the terrestrial biosphere, Science Advances, 7, eabc7447, 2021.

Chang, Q., Xiao, X. M., Doughty, R., Wu, X. C., Jiao, W. Z., and Qin, Y. W.: Assessing variability of optimum air temperature for photosynthesis across site-years, sites and biomes and their effects on photosynthesis estimation, Agricultural and Forest Meteorology, 298, 10.1016/j.agrformet.2020.108277, 2021.

Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R. K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R., Tommervik, H., Bala, G., Zhu, Z., Nemani, R. R., and Myneni, R. B.: China and India lead in greening of the world through land-use management, Nature Sustainability, 2, 122-129, 10.1038/s41893-019-0220-7, 2019.

Chen, X., Chen, T., Li, X., Chai, Y., Zhou, S., Guo, R., Dai, J.: 2001-2022 global gross primary productivity dataset using an ensemble model based on random forest. figshare. Dataset. https://doi.org/10.6084/m9.figshare.24417649.v2, 2023

Chen, T., van der Werf, G. R., Dolman, A. J., and Groenendijk, M.: Evaluation of cropland maximum light use efficiency using eddy flux measurements in North America and Europe, Geophysical Research Letters, 38, 10.1029/2011gl047533, 2011.

Chen, T., Van Der Werf, G., Gobron, N., Moors, E., and Dolman, A.: Global cropland monthly gross primary production in the year 2000, Biogeosciences, 11, 3871-3880, 2014.

Chen, X., Chen, T., Liu, S., Chai, Y., Guo, R., Dai, J., Wang, S., Zhang, L., and Wei, X.: Vegetation Index-Based Models Without Meteorological Constraints Underestimate the Impact of Drought on Gross Primary Productivity, Journal of Geophysical Research: Biogeosciences, 129, e2023JG007499, 2024.

Chen, Y., Yuan, H., Yang, Y., and Sun, R.: Sub-daily soil moisture estimate using dynamic Bayesian model averaging, Journal of Hydrology, 590, 125445, 2020.

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.: Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites, Agricultural and Forest Meteorology, 301, 108350, 2021.

Dechant, B., Ryu, Y., Badgley, G., Köhler, P., Rascher, U., Migliavacca, M., Zhang, Y., Tagliabue, G., Guan, K., and Rossini, M.: NIRVP: A robust structural proxy for sun-induced chlorophyll fluorescence and photosynthesis across scales, Remote Sensing of Environment, 268, 112763, 2022.

Dechant, B., Ryu, Y., Badgley, G., Zeng, Y., Berry, J. A., Zhang, Y., Goulas, Y., Li, Z., Zhang, Q., and Kang, M.: Canopy structure explains the relationship between photosynthesis and sun-induced chlorophyll fluorescence in crops, Remote Sensing of Environment, 241, 111733, 2020.

Doughty, R., Xiao, X. M., Qin, Y. W., Wu, X. C., Zhang, Y., and Moore, B.: Small anomalies in dry-season greenness and chlorophyll fluorescence for Amazon moist tropical forests during El Nino and La Nina, Remote Sensing of Environment, 253, 10.1016/j.rse.2020.112196, 2021.

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

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 Global GPP From the Plant Functional Type Perspective Using a Machine Learning Approach, Journal of Geophysical Research-Biogeosciences, 128, 10.1029/2022jg007100, 2023.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Welp, L. R., Keeling, R. F., Meijer, H. A. J., Bollenbacher, A. F., Piper, S. C., Yoshimura, K., Francey, R. J., Allison, C. E., and Wahlen, M.: Interannual variability in the oxygen isotopes of atmospheric CO₂ driven by El Nino, Nature, 477, 579-582, 10.1038/nature10421, 2011.

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

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

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

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

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 evapotranspiration estimation using support vector machine by integrating three process-based algorithms, Agricultural and Forest Meteorology, 242, 55-74, 2017.

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

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

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

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

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

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

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 global gross primary production for reproducing its long-term variation, 1982–2017, Earth System Science Data, 12, 2725-2746, 2020.