Tighten the bolts and nuts on GPP estimations from sites to the globe: an assessment of LUE models and supporting data fields

Zhao Wang[1], Shuguang Liu[1], Ying-Ping Wang[2], Ruben Valbuena[3], Yiping Wu[4], Mykola Kutia[4], Shuqing Zhao[5], Wenping Yuan[4], Yi Zheng[4], Yu Zhu[1], Weizhi Lu[1], Meifang Zhao[1], Xi Peng[1], Haiqiang Gao[1], Shuailong Feng[1], and Yi Shi[1]

[1]Central South University of Forestry and Technology
[2]CSIRO
[3]Bangor University
[4]Affiliation not available
[5]Peking University

July 2, 2020

Abstract

Gross primary production (GPP) determines the amounts of carbon and energy that enter terrestrial ecosystems. However, the tremendous uncertainty of the GPP still hinders the reliability of the GPP estimates and therefore understanding of the global carbon cycle. In this study, using observations from global eddy covariance (EC) flux towers, we appraised the performance of 22 widely used GPP models and quality of major spatial data layers that drive the models. Results show that the global GPP products generated by the 22 models varied greatly in the means (from 92.7 to 178.9 Pg C yr-1), trends (from -0.25 to 0.84 Pg C yr-1). Model structures (i.e., light use efficiency models, machine learning models, and process-based biophysical models) are an important aspect contributing to the large uncertainty. In addition, various biases in currently available spatial datasets have found (e.g., only 57% of the observed variation in photosynthetically active radiation was explained by the spatial dataset), which contributed greatly affects global GPP estimates. Our analysis indicates that the model development did not converge GPP simulations with the advance of time. Moving forward, research into efficacy of model structures and the precision of input data may be more important than the development of new models for global GPP estimation.

Significance Statement

Gross primary production (GPP), the amount of carbon fixed during photosynthesis in a given length of time, is the fuel of life and one of the key components of the global carbon cycle. Although numerous efforts have been made to estimate GPP at the global scale, its uncertainty has remained high for decades. It is necessary to evaluate the performance of relevant models and supporting datasets against high-quality field measurements to find out where the loose parts are, and then tighten the loose ends. Our study reveals that large biases exist in some of the models and driving spatial data products, and addressing these biases is a high priority in modeling GPP from site to global scales.

1

Introduction

Terrestrial gross primary production (GPP) or the total photosynthetic uptake of carbon by plants plays a critical role in maintaining the global carbon balance between the biosphere and the atmosphere. However, the estimation of terrestrial GPP by existing models remains highly uncertain, with global estimates ranging widely between 92.7 to 168.7 Pg C yr^{-1} (1, 2). This large uncertainty poses a serious obstacle to quantifying and understanding the global carbon cycle (3). It is broadly agreed that, to reduce the uncertainty of GPP estimation and advance carbon cycle science, it is crucial to consider: (1) the impacts of model structure, (2) the determination of parameter values, and (3) the quality of data feeding in GPP models (4, 5).

Model structure has been considered as one of the most important factors that affect model performance (6). Yet, large structural differences can be observed among GPP models. For example, the fraction of photosynthetic active radiation (FPAR), an important parameter in the light use efficiency (LUE) models, has been treated in disparate ways, either approximated by enhanced vegetation index (EVI) in the vegetation photosynthesis model (VPM) (7), as a linear function of normalized difference vegetation index (NDVI) in the eddy covariance-light use efficiency (EC-LUE) model (9), or as a nonlinear function of leaf area index (LAI) according to Beer's Law, among others(10). A similar situation exists for representing temperature stress (T_S), water stress (W_S), and their interactions among models. The moderate resolution imaging spectroradiometer (MODIS) model and the VPM model adopt a multiplicative structure to represent the collective influences of W_S and T_S on GPP (12, 7). The EC-LUE model, on the other hand, considers that the Liebig's law is ecologically more reasonable in representing the effects of W_S and T_S (8).

The estimation of model parameters often affect the simulation accuracy of the model, thus rigorous model parameterization and calibration should be adopted in GPP modeling (13). Variation in the values of the same biophysical parameters among different models is a major concern in GPP estimation. For example, the maximum light use efficiency ($LUE_{(max)}$), a parameter used in LUE-based GPP models, represents the maximum efficiency of unit vegetation converting energy to photosynthates and therefore should be relatively stable (14). However, it has taken many different values in LUE-based GPP models. In the MODIS model, $LUE_{(max)}$ values are biome-specific, varying from 0.604 to 1.259 g C MJ^{-1} (12, 15), and similar approaches can be found in other models (16). The EC-LUE model, on the other hand, takes a constant value at $LUE_{(max)}$=2.25 g C MJ^{-1}, that was derived from many flux tower observations, and the authors later advocated the use of different constant $LUE_{(max)}$ values for C3 and C4 plants (17). However, another study has suggested that fixed $LUE_{(max)}$ value would lead to increased GPP uncertainty (18). Analysis of the parameters of the diagnostic carbon flux model (DCFM) showed that cross-site estimation improved the representativeness and robustness of parameter estimates (19). Studies considering a wider number of flux towers are thus necessary for more reliable tuning of GPP model parameters.

Regional to global simulations inevitably employ spatiotemporal data for initialization or as driving forces (20). How spatial data products affect GPP simulation has rarely been assessed because users of the data products tend to take a leap of faith by assuming the quality of data has met the accuracy requirement, and limited findings regarding the importance of data quality have been ignored frequently. For example, it was found that the widely used average of eight-days MODIS satellite FPAR data was unable to reflect the reality effectively (22). Other studies have revealed that MODIS satellite FPAR are systematically lower than ground-measured FPAR observations in winter and spring (22). Clouds seriously affect satellite observations in humid regions such as the Amazon (23). Even when applying a cloud correction by the CFMask algorithm (24) only 70% of PAR can be satisfactorily simulated (25). Failure to reproduce the driving data of the models faithfully would affect the simulation of GPP, resulting in the uncertainty of GPP.

To address the three issues mentioned above and to improve the estimation accuracy of GPP at the global scale, we comprehensively appraised the structure and performance of existing LUE models against GPP estimates from 151 eddy covariance (EC) flux towers worldwide, and assessed the impacts of using currently available data products on the estimates of global GPP using newly developed LUE models. The specific

objectives were: (1) to compare existing models and address the model structure deficiency, if necessary, (2) to find the optimal parameter values using GPP observed at the EC towers, (3) to develop a new model to taking advantage of remote sensing (RS) data as directly as possible, reducing errors of intermediate data products and algorithms, (4) to evaluate errors of spatial RS data and their impacts on GPP estimation, and (5) to generate new GPP model and subsequent global product, after correcting biases in spatial data layers.

Results

3.1 Comparison of model performance

Globally, the LUE-EF model had the highest correlation coefficient of r = 0.86, followed by the LUE-NDWI model with r = 0.82 (Fig.1a). The LUE-EF model had the smallest normalized RMSE (0.50), indicating that the difference between the LUE-EF model and EC-towers GPP was the smallest among all models. The LUE-NDWI model had the second smallest normalized RMSE (0.52). The normalized RMSE for the other models was larger than the two new models. Model LUE-EF simulated the amplitude of the variations close to the data amplitude of EC-towers (SD ratio=0.93). When grouped by latitudinal zone (Fig.1b-e) the improvements of the new models were more apparent for tropical and northern temperate zones. In the temperate zones, the distribution of models in the Taylor diagram was relatively concentrated, whereas in the tropical and boreal areas there were larger differences among models. When grouped by biome (Fig.S2), the new LUE-EF model showed advantages of fit in simulating daily GPP for most biomes, both in terms of correlation coefficients and RMSE. For example, correlation coefficients were highest for LUE-EF in deciduous broadleaved and evergreen needleleaf forests, wetlands and grasslands (DBF, ENF, WET and GRA), with LUE-NDWI being the second highest in the latter three of those. It is important to note that these biomes also have the largest number of EC-towers. It is relevant to note that many biomes are underrepresented in the current EC-tower network, such as closed shrublands or deciduous needle-leaved forests (CSH and DNF), each represented by only one and two EC-towers. The Taylor diagram shows that the models in CSH, DNF and SAV are more dispersed, which means that model performances in these biomes vary greatly.
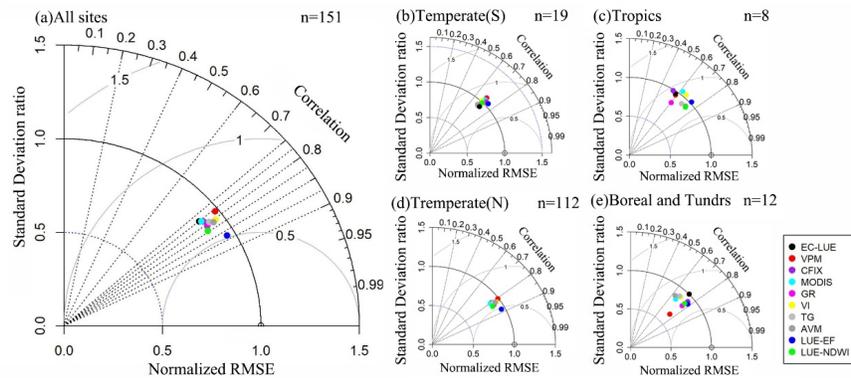
<Fig 1 roughly here>

Fig.1. The Taylor diagram showing (a) the performance of all model data from EC-towers were used.; (b-e) the performance of all model under different latitudes, where N and S represent the northern and southern hemispheres, respectively. Color dots represent the models in the corresponding legend. Taylor diagram is a polar graph in which the cosine of the angle between the X-axis is the correlation coefficient between the GPP of the model and EC-tower. The radial direction is the ratio of model to EC-tower GPP standard deviation. The grey arcs represent RMSE normalized by standard deviation for each model. The n is number of EC-towers.

3

The ability of a model to capture temporal changes is one of the keys to evaluating the performance of GPP models. The double mass curve (cumulative GPP predicted by model versus cumulative GPP observed at EC-towers) computed per site can reflect the ability of models to simulate the temporal changes of GPP (Fig.2). From the distribution of double mass curve, results show that, among the 10 models compared, the distribution cumulated GPP for the LUE-EF was the most concentrated around the 1:1 correspondence against the cumulated GPP from EC-towers, which indicates its greater ability to simulate the patterns in temporal variability in GPP. Fig. 2 also shows the distributions of relative bias in ratio (PB), which ranged ±0.4 for all models, indicating that the models had a large heterogeneity in simulating the temporal change of GPP. The biases were however narrower for the LUE-EF and LUE-NDWI models, respectively containing 120 and 110 of the EC-towers within ±0.2 in their PB, the largest number of EC-towers amongst all models. This indicates that these two models have a strong comprehensive ability to capture the temporal changes of GPP.
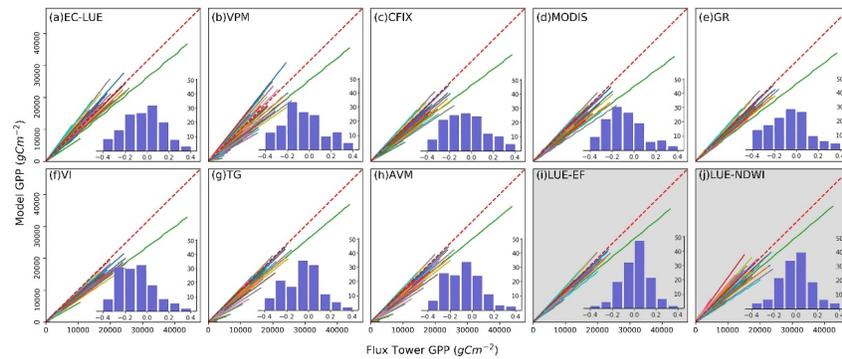
<Fig 2 roughly here>

Fig.2. Comparison of cumulative GPP estimates from the flux towers and the models. The color lines represent the GPP value of cumulative comparison between the EC-tower and model for each site. The red dashed line is the 1:1 reference to the differences of modeled GPP and EC-tower. Inset histogram shows the frequency distribution of the percentage biases (PB). The two shadowed plots are two new models developed in this study.

3.2 Biases in remote sensing data products and consequences on global GPP estimation

Evaluating the quality of input data and understanding the impact of data biases on GPP simulation are prerequisites for improving GPP simulation accuracy. First, various biases were found when the spatial datasets that feed the models for global GPP simulations were evaluated at the site scale (Fig.3; Fig.S3). For example, the spatial PAR dataset only explained 57% of the observed PAR variation at the EC-towers, and the slope and intercept were 1.2 and 0.57, respectively, indicating that the PAR data fields overestimated PAR as a whole and slightly underestimated PAR at the low value. The determination coefficients of the global datasets of CO_2, LE, and H at the EC-towers were less than 20% ($R^2<0.2$), only that of the temperature data was efficient in representing site-scale variation ($R^2=0.89$).

Second, the biases in the spatial datasets had a significant impact on GPP simulations. Before correcting these biases, the simulated GPP by the LUE-EF and LUE-NDWI models explained only 49% and 61% of the EC-tower GPP variation, and the slopes of the linear regression between simulated and tower-estimated GPP were 1.54 and 1.31, respectively, and the corresponding intercepts were -2.09 and -1.23. These results indicate that both models overestimated GPP as a whole, but underestimated low GPP values (Fig.4b and

4

f). After correcting the biases in the spatial datasets, the R^2 of LUE-EF and LUE-NDWI models improved to 0.80 and 0.79, with the slopes closer to 1 (1.20 and 1.18 values, respectively) and the intercepts closer to 0 (-0.59 and -0.91, respectively) (Fig.3c and g). The results also indicated that the LUE-NDWI model was less sensitive to the biases in the spatial data fields than the LUE-EF model, as shown by the smaller differences in R^2 before and after data correction, probably attributed to the fact that the LUE-NDWI relies on NDWI, a factor that can be derived directly from remote sensing data and thereby less prone to error propagation than the LUE-EF model.
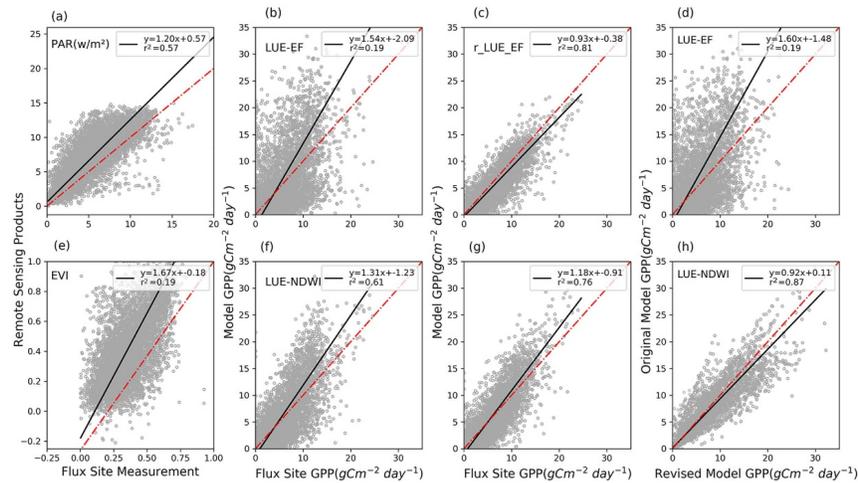
<Fig 3 roughly here>

Fig.3 Comparison of (a) spatial PAR and site PAR, (b) tower GPP and LUE-EF GPP (uncorrected spatial data), (c) tower GPP and LUE-EF GPP (corrected spatial data), (d) LUE-EF GPP with uncorrected and LUE-EF GPP corrected spatial data, (e)spatial EVI (500m resolution) and spatial EVI (10km resolution), (f) tower GPP and LUE-NDWI GPP (uncorrected spatial data), (g) tower GPP and LUE-NDWI GPP (corrected spatial data), and (h) LUE-NDWI GPP with uncorrected and LUE-NDWI GPP corrected spatial data. All comparisons are based on site scale.

At the global scale, the biases in spatial data inputs had a great impact on the simulated GPP even for the less sensitive model LUE-NDWI. Fig.4a shows the global average annual GPP distribution from 2000 to 2018, simulated by the LUE-NDWI model using corrected input data layers. The spatial pattern of GPP agrees well with previous studies. However, the impact of data biases on the spatial pattern of simulated GPP was obvious and not uniform across space (Fig.4b). The area overestimated is much larger than underestimated area when the data biases were not attended, and the area fractions with GPP biases at (-50%)-(-30%), (-30%)-(-10%),10%-30%, and 30%-50%, were 8%, 19%, 27%, 31% and 15%, respectively. After data correction, area of GPP serious reduction occurs in the mountain systems of the Tibetan plateau in Asia, northern Africa and South America region. Area of serious growth were observed in Australia, northwest North America and Siberia. The global annual average GPP estimated by the LUE-NDWI, after input data correction, was about 125.6Pg C yr-1. Without data correction, the LUE-NDWI model would overestimate global GPP by 18% (Fig 4c and 4d). The corresponding global growth rate of GPP decreased from 0.34 to 0.17 Pg C yr-1 after input data correction (Fig. 4e).

The comparison of global GPP products simulated by 22 models is shown in Fig 4c. The GPP products were from TRENDY and other studies. Large differences can be seen from these models with long-term GPP averages varying from 92.7 to 178.9 Pg C yr-1 with more GPP estimates concentrated in the 120-130 Pg C yr-1 (Fig 4d). The interannual variabilities simulated by these models were also quite different. In

5

general, the inter-annual variabilities of LUE and biophysical models were greater than those of the machine learning models. It is worth noting that the GPP simulations from the Revised-ECLUE, PR, LPJ-GUESS, and data LUE-NDWI (after correcting input data biases) models were all within the interquartile range (IQR) of the 22 GPP products throughout all years. In addition, the GPP from the LUE-NDWI model, after data correction, was the closest to the median GPP value of the 22 global models. The trends of GPP simulated by the 22 models also varied greatly from -0.25 to 0.84 Pg C yr-1 (Fig 4e). The trends of the machine learning models were smaller than those of other models. Most model showed positive trends, only the revised-ECLUE and CLASS-CTEM models showed downward trends, and some models demonstrated no significant trends (i.e., MODIS, FLUXCOM_ANN, FLUXCOM_MARS, and FLUXCOM_RF).

<Fig 4 roughly here>

Fig.4. Consequences of biased input data on LUE-NDWI GPP at global scale. (a) the mean annual GPP from 2000 to 2018 after data correction. (b) the difference of mean annual GPP from 2000 to 2018, $GPP_{uncorrected} - GPP_{corrected}$. Symbols indicate various levels of difference: '−': -50%-(-30%), '-': -30%-(-10%), '±': -10%-10%, '+': 10%-30%, '++': 30%-50%. The inset bar chart shows the global distribution of the difference proportions. (c) Comparison of annual global GPP estimates from various models. The number after each model's name in parentheses is the number of years of the model in the interquartile range (IQR), indicated by the shaded region. LUE-NDWI and LUE-NDWI1 are GPP estimated by LUE-NDWI with corrected and uncorrected input data, respectively. (d) Boxplot of annual GPP values during the study period for each model. (e) Trends of annual GPP (Pg C yr-1) by model. Symbols of LUE models are in black, machine learning models in green, biophysical models in blue.

Discussion

4.1 Adequacy of model structure in representing processes

The understanding of the processes involved in GPP is fundamental to building a reliable GPP model. For example, we found that failed incorporation of the effect of clouds on GPP in some existing models significantly underestimated GPP in areas with frequent cloudy cover. Under clear sky conditions, the upper canopy leaves are close to light saturation, while the lower canopy leaves are shaded and have limited light (26). In contrast, under cloudy conditions, a higher proportion of the light in the form of diffuse

radiation can reach the lower parts of the canopy, thus increasing the total photosynthetic use of PAR by the vegetation (27). Some studies indicated that a 1% increase in diffuse radiation induces a 0.94% increase in GPP (29, 30). In addition, many studies have shown that CO_2 fertilization has a significant effect on vegetation production, a dominant factor contributing to the 31% increase in global GPP since 1990 (30). Nevertheless, many LUE models have not explicitly accounted for the effect of increasing atmospheric CO_2 concentration(32, 33). In our study, after incorporating the impacts of both cloud cover and CO_2, the performance of the LUE-EF and LUE-NDWI models improved compared with the original EC-LUE model: R^2 improved from 0.61 to 0.68 for LUE-NDWI and from 0.61 to 0.74 for LUE-EF, respectively.

Water availability is an important factor that affects GPP(33). In this study, we adopted two alternative parameterized expressions to represent the impact of water stress on GPP. First, the evaporative fraction (EF) of total energy, closely related to the Bowen ratio, was used in the LUE-EF model. The relevance of LUE-EF is grounded on the fact that less energy used for ecosystem evaporation (i.e., smaller evaporative fraction) implies a stronger water limitation (35), which has largely been verified using flux tower measurements(35). Nevertheless, the application of the LUE-EF model is hindered by the derivation of EF that required multiple steps and input data layers and therefore prone to error propagation (36). In order to get a direct measure of water stress, and therefore minimize error propagation, from satellite observations we replaced the EF using NDWI in the LUE-NDWI model, based on the evidence that NDWI is closely related to the plant water content and thus a good proxy for plant water stress (37). The direct use of NDWI in the LUE-NDWI model makes it ideal for mapping GPP at the regional to global scales.

4.2 input data biases and possible impacts on GPP simulations

In this study, we found that the spatial data at the EC-tower sites had various systematic deviations, which seriously affected GPP estimates. For example, the spatial data fields of PAR explained only 57% of the PAR variation observed at the EC-tower sites. Reasons for the data biases are mainly rooted in the influence of sensor errors and atmospheric factors (e.g., cloud and snow) (39, 40, 41). In addition, attempts unifying data from different spatial and temporal resolutions also bring biases as we found that data at different resolutions sometimes had poor correlations. The reason for the existence of resolution mismatch is mainly caused by mixed pixels and/or different time scales (e.g., daily, 8-day, or monthly data), compared to the spatio-temporal resolution that applies to ground conditions (42, 43). In general, spatial data errors are an important cause of the uncertainty of GPP simulation, which might have contributed to the large differences in GPP estimates among existing GPP models (44). Thus, improving the quality of input data fields should be a major research component in reducing the uncertainty in GPP simulations at regional to global scales.

The comparison of existing global GPP products shows a huge variation from 92.7 to 178.9 Pg C yr^{-1}, which might be related to data biases (45). One should realize that the sensitivities of models to data biases are usually different, depending on model structure. In our study, it was found that the LUE-EF model is more susceptible to data errors, while the LUE-NDWI model is less affected by data biases (Fig.3d). There are many GPP models in the world, and the sensitivity of each model to data biases has not been effectively evaluated and compared. The sensitivities of the models to data deviations should be systematically investigated in future research. Another research area that deserves more attention is how to improve data quality. Data bias should be a primary concern in modeling GPP from site to region scales. The existence bias of spatial data not only effects the GPP simulation, but more importantly it hinders the observation and understanding of the earth system. Therefore, effective correction of spatial data is critical for reducing the uncertainty in GPP simulations, and research on improving data quality should be encouraged. In this regard, our research cautions using currently available spatial data for relevant research.

4.4 Improving GPP simulation capability: the ways forward

In this study, after data correction, the coefficients of determination for the LUE-EF and LUE-NDWI models at site scale reached 0.80 and 0.79, respectively. The evaporative fraction parameter of the LUE-EF

7

model involves more steps than the NDWI of the LUE-NDWI model, and consequently the risk of the error accumulation increases. The direct observation and continuous recording of NDWI as a remote sensing product is one of the main advantages of the LUE-NDWI model. The risk of error propagation of the LUE-NDWI model at the regional to global scale is small. Therefore, the LUE-NDWI model is more practical and attractive in spatially-explicit simulations of GPP. In addition, many studies now show that Sun-Induced chlorophyll Fluorescence (SIF) has a strong near linear relationship with GPP (43), indicating SIF can be used as a direct indicator of GPP. In principle, the simple forms of GPP models and widely available inputs, as compared with more complex global models, make them more practical for applications over large areas and better suited for attribution and uncertainty analysis (46, 47).

Should a balance or combination of multiple approaches be considered in the study of GPP simulation improvement? The comparison of global GPP products shows that there are differences among the three types of GPP models. The biophysical models showed the greatest inter-annual variability, followed by LUE models, and the machine learning models had minimal. Some research shows the machine learning models rely on empirical relationships between forcing variables and fluxes, which causes them barely captured the inter-annual variation of GPP (45). In contrast, LUE models assume that GPP is only related to environmental factors (e.g., temperature, water, etc.) and light use efficiency in relatively straightforward forms. Compared with the biophysical models, LUE models have less parameters, and simpler structures. This means that biophysical models have higher probability of error expansion in the processes of GPP simulations. In addition, studies have shown that the combination of machine learning models with other type of models have more potential to improve GPP estimation, which implies a balance or combination of multiple approaches might be the choice to reduce the uncertainty in GPP simulation.

Vastly different GPP products, as shown by the means, trends, and interannual variabilities of GPP, generated by the 22 models suggest our current ability in simulating global GPP is not encouraging (Fig.4c). For nearly 40 years committed to the global simulation of GPP, there does not seem to have a clear direction of the improvement of GPP simulation. This is mainly reflected in the fact that the models did not converge in GPP simulations with the advance of time. GPP model development is not explicitly directed, despite the constant emergence of new models. Our research indicates that neither the structure of the model nor the quality of the input data are error prone. Therefore, specific standards need to be developed to optimize model structures as well as sufficient validation and calibration of the input data. Research into efficacy of model structures and the precision of input data may be more important than the development of new models for global GPP estimation.

Materials and Methods

Description of the LUE-EF model

This model was developed mainly based on the principles of the EC-LUE model (48). Specifically, the regulation of water on GPP is represented by the evaporative fraction (EF), taking advantage of the newly available EF products (49). In addition, two new modifiers of GPP were added to the original EC-LUE model. The first modifier considers the impact of cloudiness on GPP. The other modifier addresses the fertilization effect of increased CO_2 concentration in the atmosphere.

The LUE-EF model can be expressed as follows:

$$\text{GPP} = \text{PAR} \times \text{FPAR} \times \mathbf{F}_{\text{CI}} \times \mathbf{F}_{\text{CO2}} \times \text{LUE}_{(\text{MAX})} \times \min\left(\mathbf{T_S}, \mathbf{W_S}\right) \qquad (1)$$

Where PAR is incident photosynthetic active radiation (MJ/m^2) over a period of time; $FPAR$ is the fraction of PAR absorbed by the vegetation; F_{CI} is regulation of cloudiness on GPP; F_{CO2} is the regulation scalar of atmospheric CO_2 concentration; $LUE_{(MAX)}$ is maximum light use efficiency; T_S and W_S are regulation scalars

respectively for temperature and water stress on GPP, from which the minimum value is taken, following the Leibig law (9). The determination of models parameters was done as:

FPAR is in practice approximated by EVI (7), since photosynthetically active of vegetation is estimated as a ratio *a* of *EVI* , set to be *a* =1:

$$\text{FPAR}_{\text{EVI}} = \alpha \times \textbf{EVI} \tag{2}$$

Most previous models underestimates of GPP on cloudy days mainly because photosynthesis can be increased by diffuse radiation under cloudy conditions (28). The regulating effect of cloud cover on GPP was expressed by cloudiness index (CI) as follow:

$$\textbf{F}_{\text{CI}} = \textbf{a} \times \text{CI} + \textbf{b} \tag{3}$$

Where *CI* is the ratio of PAR to potential PAR (PPAR) (33). Using the FLUXNET2015 dataset, the coefficients were determined to be *a* (=2.9) and *b* (=1.2), using the parameter optimization for nonlinear least-squares (NLS) regression using the 'nls' function in R. The robustness of the NLS method was verified by Weibull function sensitivity analysis (4).

For calculating the influence of atmospheric CO_2 on GPP, we employed the algorithm in the Frankfurt biosphere model (FBM) (46):

$$\textbf{F}_{\text{CO2}} = \textbf{f}\,(\text{CO2}, \textbf{T}) = \frac{\text{CCL} - (\textbf{T})}{\text{CCL} + \textbf{2}\,(\textbf{T})} \tag{4}$$

where *CCL* is the internal CO_2 concentration of leaves, and it assumed to be 70% of atmospheric CO_2 concentration. *Δ(T)* is the CO_2 compensation point for gross photosynthesis and photorespiration at temperature T(°C) (47):

$$(\textbf{T}) = \textbf{40.6e}^{\frac{(\textbf{9.46} \times (\textbf{T}-\textbf{25}))}{(\textbf{T}+\textbf{273.2})}} \tag{5}$$

Similar argument for eq (4), it always is <1.0 at all T and CO2.

The regulation scalar of water on GPP, W_S , was expressed as the evaporative fraction (EF) of the total sensible and latent heat (8):

$$\textbf{W}_{\textbf{S}} = \text{EF} = \frac{\text{LE}}{\text{LE} + \textbf{H}} \tag{6}$$

where *LE* is latent heat flux (W m^{-2}), and *H* is sensible heat flux (W m^{-2}).

Description of the LUE-NDWI model

The NDWI, strongly related to vegetation water content (Murphy et al., 2018;)(50), can be a very good proxy for vegetation water stress. After examining measurements from many flux towers, we found that the following nonlinear function can be used to represent W_S , the regulation scalar of water stress on GPP:

$$\textbf{W}_{\textbf{S\_NDWI}} = \textbf{a} * (-\text{NDWI} + \textbf{0.5})^{\textbf{b}} + \textbf{c} \tag{7}$$

Using the NLS method of parameter optimization adjustment, the coefficients were determined to be *a* (=0.35), *b* (=2.14) and c (=0.086). The W_{S_NDWI} values vary between 0 and 1, with values beyond the

bounds set to 0 or 1, respectively. The use of W_{S_NDWI} can be very convenient for applications from local to global scales as the NDWI fields can be directly derived from satellite observations.

Data for Evaluation of Models and Spatial Data Products

For evaluation of models and remotely sensed data products, we used eddy covariance (EC) flux tower data from the FLUXNET2015 dataset (https://fluxnet.fluxdata.org). Our study included data from 151 EC tower sites which belonged to the following 12 terrestrial biomes. We used two criteria to filter the data, following (33): (1) if more than 20% of the data in a given year was missing, the whole year was discarded, and (2) after this first step of processing, EC-tower sites with records for less than two years were completely discarded. EC-towers details are in the Table.S2.

To simulate global GPP, the following spatial data products were used: (1) meteorological data fields – radiation, air temperature, latent heat flux, and sensible heat flux – derived from the second Modern-Era Retrospective analysis for Research and Applications (MERRA-2), (2) MODIS satellite products including enhanced vegetation index (EVI) and normalized difference wetness index (NDWI), and (3) atmospheric CO_2 concentration from the Earth System Research Laboratory Global Monitoring Division (https://www.esrl.noaa.gov/gmd/dv/site/). In addition, global GPP products were selected for comparison, including: EC-LU, revised EC-LUE, MODIS, MOD17 C6, PR, VPM, FLUXCOM, SVR, BESS, BEPS, and 10 TRENDY products.

Acknowledgments

References

1. J. E. Campbell, *et al.* , Large historical growth in global terrestrial gross primary production. *Nature* **544** , 84–87 (2017).

2. Y. Zheng, *et al.* , Improved estimate of global gross primary production for reproducing its long-term variation, 1982-2017.*Earth Syst. Sci. Data Discuss.* (2019).

3. G. Lasslop, *et al.* , Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate. *Science (80-.).***329** , 834–838 (2010).

4. J. Xia, *et al.* , Joint control of terrestrial gross primary productivity by plant phenology and physiology. *Proc. Natl. Acad. Sci. U. S. A.* **112** , 2788–2793 (2015).

5. Wang, H., Prentice, I.C., Keenan, T.F. et al. Towards a universal model for carbon dioxide uptake by plants. Nature Plants **3** , 734–741 (2017).

6. G. B. Bonan, *et al.* , Model Structure and Climate Data Uncertainty in Historical Simulations of the Terrestrial Carbon Cycle (1850–2014). *Global Biogeochem. Cycles* **33** , 1310–1326 (2019).

7. X. Xiao, *et al.* , Satellite-based modeling of gross primary production in an evergreen needleleaf forest. *Remote Sens. Environ.* **89** , 519–534 (2004).

8. W. Yuan, *et al.* , Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. *Agric. For. Meteorol.* **143** , 189–207 (2007).

9. W. Yuan, *et al.* , Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. **143** , 189–207 (2007).

10. Z. Xiao, *et al.* , Long-Time-Series Global Land Surface Satellite Leaf Area Index Product Derived From MODIS and AVHRR Surface Reflectance. 1–18 (2016).

11. S. W. Running, M. Zhao, Daily GPP and Annual NPP (MOD17A2/A3) products NASA Earth Observing System MODIS Land Algorithm - User's guide V3. 28 (2015).

12. P. Wagle, P. H. Gowda, X. Xiao, K. C. Anup, Parameterizing ecosystem light use efficiency and water use efficiency to estimate maize gross primary production and evapotranspiration using MODIS EVI. *Agric. For. Meteorol.* **222** , 87–97 (2016).

13. Y. Zheng, *et al.* , Agricultural and Forest Meteorology Sources of uncertainty in gross primary productivity simulated by light use efficiency models: Model structure, parameters, input data, and spatial resolution. *Agric. For. Meteorol.* **263** , 242–257 (2018).

14. W. Yuan, *et al.* , Estimating crop yield using a satellite-based light use efficiency model. *Ecol. Indic.***60** , 702–709 (2016).

15. S. W. RUNNING, *et al.* , A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. *Bioscience* **54** , 547 (2006).

16. X. Xie, *et al.* , Assessments of gross primary productivity estimations with satellite data-driven models using eddy covariance observation sites over the northern hemisphere. *Agric. For. Meteorol.* **280** , 107771 (2020).

17. W. Yuan, *et al.* , Agricultural and Forest Meteorology Uncertainty in simulating gross primary production of cropland ecosystem from satellite-based models. *Agric. For. Meteorol.* **207** , 48–57 (2015).

18. H. Wang, *et al.* , Deriving maximal light use efficiency from coordinated flux measurements and satellite data for regional gross primary production modeling. *Remote Sens. Environ.* **114** , 2248–2258 (2010).

19. J. Xiao, K. J. Davis, N. M. Urban, K. Keller, N. Z. Saliendra, Upscaling carbon fluxes from towers to the regional scale: Influence of parameter variability and land cover representation on regional flux estimates. **116** , 1–15 (2011).

20. G. Badgley, L. D. L. Anderegg, J. A. Berry, C. B. Field, Terrestrial gross primary production: Using NIRV to scale from site to globe.*Glob. Chang. Biol.* **25** , 3731–3740 (2019).

21. S. Kang, *et al.* , A regional phenology model for detecting onset of greenness in temperate mixed forests, Korea: an application of MODIS leaf area index. **86** , 232–242 (2003).

22. T. Sasai, K. Okamoto, T. Hiyama, Y. Yamaguchi, Comparing terrestrial carbon fluxes from the scale of a flux tower to the global scale.*Ecol. Modell.* **208** , 135–144 (2007).

23. J. Gomis-cebolla, J. C. Jimenez, J. A. Sobrino, Remote Sensing of Environment LST retrieval algorithm adapted to the Amazon evergreen forests using MODIS data. *Remote Sens. Environ.* , 0–1 (2017).

24. S. Foga, *et al.* , Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* **194** , 379–390 (2017).

25. Stillinger, T., Roberts, D. A., Collar, N. M., & Dozier, J. Cloud masking for Landsat 8 and MODIS Terra over snow-covered terrain: Error analysis and spectral similarity between snow and cloud. Water Resources Research, **55** , 6169 – 6184 (2019).

26. S. J. Cheng, Z. Butterfield, G. Keppel-aleks, A. L. Steiner, The Global Influence of Cloud Optical Thickness on Terrestrial Carbon Uptake. **23** (2019).

27. L. Gu, *et al.* , Response of a Deciduous Forest to the Mount Pinatubo Eruption: Enhanced Photosynthesis. **299** , 2035–2038 (2003).

28. Keppel-Aleks, G., and Washenfelder, R. A., The effect of atmospheric sulfate reductions on diffuse radiation and photosynthesis in the United States during 1995–2013, Geophys. Res. Lett., **43** , 9984– 9993, (2016).

29. M. S. Lee, D. Y. Hollinger, T. F. Keenan, A. P. Ouimette, S. V Ollinger, Agricultural and Forest Meteorology Model-based analysis of the impact of di ff use radiation on CO_2 exchange in a temperate deciduous forest. *Agric. For. Meteorol.* **249** , 377–389 (2018).

30. V. Haverd, *et al.* , Higher than expected CO_2 fertilization inferred from leaf to global observations. 1–13 (2020).

31. Z. Sun, *et al.* , Science of the Total Environment Evaluating and comparing remote sensing terrestrial GPP models for their response to climate variability and CO_2 trends. *Sci. Total Environ.* **668** , 696–713 (2019).

32. A. Bastos, P. Ciais, F. Chevallier, C. Rödenbeck, A. P. Ballantyne, Contrasting effects of CO_2 fertilization, land-use change and warming on seasonal amplitude of Northern Hemisphere CO_2 exchange. 12361–12375 (2019).

33. W. Yuan, *et al.* , Agricultural and Forest Meteorology Global comparison of light use efficiency models for simulating terrestrial vegetation gross primary production based on the LaThuile database. *Agric. For. Meteorol.* **192** –**193** , 108–120 (2014).

34. S. A. Kurc, E. E. Small, Dynamics of evapotranspiration in semiarid grassland and shrubland ecosystems during the summer monsoon season, central New Mexico. **40** , 1–15 (2004).

35. Y. Chen, *et al.* , Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. *Remote Sens. Environ.* **140** , 279–293 (2014).

36. H. Y. Ma, *et al.* , CAUSES: On the Role of Surface Energy Budget Errors to the Warm Surface Air Temperature Error Over the Central United States. *J. Geophys. Res. Atmos.* **123** , 2888–2909 (2018).

37. B. Gao, NDWI-A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. *Remote Sens. Environ.* **266** , 257–266 (1996).

38. Z. Liu, L. Wang, S. Wang, Comparison of different GPP models in China using MODIS image and ChinaFLUX data. *Remote Sens.* **6** , 10215–10231 (2014).

39. C. Wu, J. W. Munger, Z. Niu, D. Kuang, Remote Sensing of Environment Comparison of multiple models for estimating gross primary production using MODIS and eddy covariance data in Harvard Forest. *Remote Sens. Environ.* **114** , 2925–2939 (2010).

40. Morton, D., Nagol, J., Carabajal, C. et al. Amazon forests maintain consistent canopy structure and greenness during the dry season. Nature **506** , 221–224 (2014).

41. E. Borbas, G. Hulley, R. Knuteson, M. Feltz, S. Science, MEaSUREs Unified and Coherent Land Surface Temperature and Emissivity (LST & E) Earth System Data Record (ESDR): The Combined ASTER and MODIS Emissivity database over Land (CAMEL) Version 2 Users ' Guide (2019).

42. C. Jin, *et al.* , Agricultural and Forest Meteorology Effects of in-situ and reanalysis climate data on estimation of cropland gross primary production using the Vegetation Photosynthesis Model. *Agric. For. Meteorol.* **213** , 240–250 (2015).

43. X. Wang, J. M. Chen, W. Ju, Photochemical reflectance index (PRI) can be used to improve the relationship between gross primary productivity (GPP) and sun-induced chlorophyll fluorescence (SIF). *Remote Sens. Environ.* **246** , 111888 (2020).

44. M. Chiesi, *et al.* , Testing the applicability of BIOME-BGC to simulate beech gross primary production in Europe using a new continental weather dataset. *Ann. For. Sci.* **73** , 713–727 (2016).

45. M. M. Rienecker, *et al.* , MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Clim.* **24** , 3624–3648 (2011).

46. David W. Kicklighter, Michele Bruno, Silke DZönges *et al.* , A first-order analysis of the potential rôle of CO2 fertilization to affect the global carbon budget: a comparison of four terrestrial biosphere models, Tellus B: Chemical and Physical Meteorology,**51** :2, 343-366 (1999).

47. A. J. W. Raich, *et al.* , Potential Net Primary Productivity in South America: Application of a Global Model Published by : Ecological Society of America Stable URL : http://www.jstor.org/stable/1941899 . POTENTIAL NET PRIMARY PRODUCTIVITY IN SOUTH AMERICA: APPLICATION OF A GLOB. **1** , 399–429 (2013).

48. R. J. Murphy, B. Whelan, A. Chlingaryan, S. Sukkarieh, Quantifying leaf - scale variations in water absorption in lettuce from hyperspectral imagery: a laboratory study with implications for measuring leaf water content in the context of precision agriculture.*Precis. Agric.* (2018).

49. C. Ding, X. Liu, F. Huang, Y. Li, X. Zou, Onset of drying and dormancy in relation to water dynamics of semi-arid grasslands from MODIS NDWI. *Agric. For. Meteorol.* **234** –**235** , 22–30 (2017).

50. Y. Zhang, N. C. Parazoo, A. P. Williams, S. Zhou, P. Gentine, Large and projected strengthening moisture limitation on end-of-season photosynthesis. *Proc. Natl. Acad. Sci.* , 201914436 (2020).

Figures
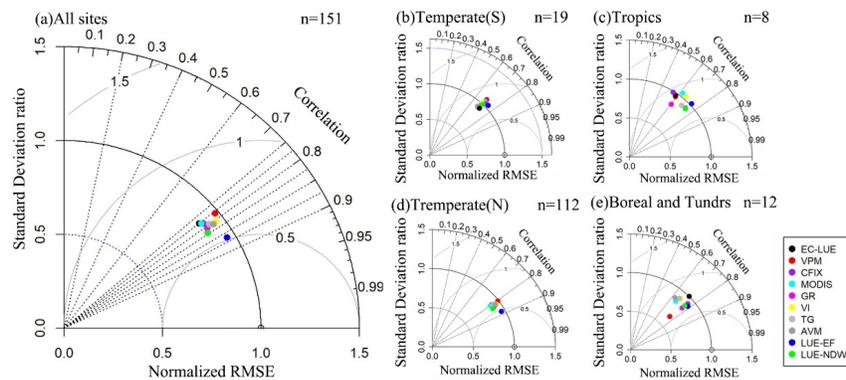
Fig.1. The Taylor diagram showing (a) the performance of all model data from EC-towers were used.; (b-e) the performance of all model under different latitudes, where N and S represent the northern and southern hemispheres, respectively. Color dots represent the models in the corresponding legend. Taylor diagram is a polar graph in which the cosine of the angle between the X-axis is the correlation coefficient between the GPP of the model and EC-tower. The radial direction is the ratio of model to EC-tower GPP standard deviation. The grey arcs represent RMSE normalized by standard deviation for each model. The n is number of EC-towers.

Fig.2. Comparison of cumulative GPP estimates from the flux towers and the models. The color lines represent the GPP value of cumulative comparison between the EC-tower and model for each site. The red dashed line is the 1:1 reference to the differences of modeled GPP and EC-tower. Inset histogram shows the frequency distribution of the percentage biases (PB). The two shadowed plots are two new models developed in this study.

Fig.3 Comparison of (a) spatial PAR and site PAR, (b) tower GPP and LUE-EF GPP (uncorrected spatial data), (c) tower GPP and LUE-EF GPP (corrected spatial data), (d) LUE-EF GPP with uncorrected and LUE-EF GPP corrected spatial data, (e)spatial EVI (500m resolution) and spatial EVI (10km resolution), (f) tower GPP and LUE-NDWI GPP (uncorrected spatial data), (g) tower GPP and LUE-NDWI GPP (corrected spatial data), and (h) LUE-NDWI GPP with uncorrected and LUE-NDWI GPP corrected spatial data. All comparisons are based on site scale.

Fig.4. Consequences of biased input data on LUE-NDWI GPP at global scale. (a) the mean annual GPP from 2000 to 2018 after data correction. (b) the difference of mean annual GPP from 2000 to 2018, $GPP_{uncorrected} - GPP_{corrected}$. Symbols indicate various levels of difference: '−': -50%-(-30%), '-': -30%-(-10%), '±': -10%-10%, '+': 10%-30%, '++': 30%-50%. The inset bar chart shows the global distribution of the difference proportions. (c) Comparison of annual global GPP estimates from various models. The number after each model's name in parentheses is the number of years of the model in the interquartile range (IQR), indicated by the shaded region. LUE-NDWI and LUE-NDWI1 are GPP estimated by LUE-NDWI with corrected and uncorrected input data, respectively. (d) Boxplot of annual GPP values during the study period for each model. (e) Trends of annual GPP (Pg C yr^{-1}) by model. Symbols of LUE models are in black, machine learning models in green, biophysical models in blue.