Journal of Remote Sensing
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

CMLR: A Mechanistic Global GPP Dataset Derived from TROPOMIS SIF Observations

Ruonan Chen[1,2], Liangyun Liu[1,2*], Xinjie Liu[1], and Uwe Rascher[3]

[1]Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China. [2]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China. [3]Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52428 Jülich, Germany.

*Address correspondence to: liuly@radi.ac.cn

Citation: Chen R, Liu L, Liu X, Rascher U. CMLR: A Mechanistic Global GPP Dataset Derived from TROPOMIS SIF Observations. *J. Remote Sens.* 2024;4:Article 0127. https://doi.org/10.34133/remotesensing.0127

Submitted 23 May 2023
Accepted 7 February 2024
Published 13 March 2024

Solar-induced chlorophyll fluorescence (SIF) has shown promise in estimating gross primary production (GPP); however, there is a lack of global GPP datasets directly utilizing SIF with models possessing clear expression of the biophysical and biological processes in photosynthesis. This study introduces a new global 0.05° SIF-based GPP dataset (CMLR GPP, based on Canopy-scale Mechanistic Light Reaction model) using TROPOMI observations. A modified mechanistic light response model was employed at the canopy scale to generate this dataset. The canopy q_L (opened fraction of photosynthesis II reaction centers), required by the CMLR model, was parameterized using a random forest model. The CMLR GPP estimates showed a strong correlation with tower-based GPP ($R^2 = 0.72$) in the validation dataset, and it showed comparable performance with other global datasets such as Boreal Ecosystem Productivity Simulator (BEPS) GPP, FluxSat GPP, and GOSIF (global, OCO-2-based SIF product) GPP at a global scale. The high accuracy of CMLR GPP was consistent across various normalized difference vegetation index, vapor pressure deficit, and temperature conditions, as well as different plant functional types and most months of the year. In conclusion, CMLR GPP is a novel global GPP dataset based on mechanistic frameworks, whose availability is expected to contribute to future research in ecological and geobiological regions.

Introduction

Terrestrial gross primary production (GPP) is the largest carbon cycle flux [1]. Therefore, quantifying GPP is important for tracking carbon budgets, projecting vegetation–climate interactions, and supporting the policy decisions regarding future land management. However, mapping GPP at a global scale remains a challenging task, which involves complex biophysical and biochemical processes across different spatiotemporal scales [2].

To date, numerous attempts have been made to produce global GPP datasets. Traditional GPP products are usually generated using light use efficiency (LUE) framework, machine learning methods, or process-based models. Models based on the LUE framework represent GPP as the product of absorbed photosynthetically active radiation (APAR) and LUE. In these models, estimating LUE is crucial but challenging work because this parameter cannot be directly measured at a global scale. As a result, various approaches to model LUE have been developed, leading to different global GPP products [3,4]. However, these methods mainly rely on environmental conditions or surface reflectance information to simulate LUE, lacking direct physiological information and thereby introducing uncertainties [5]. GPP estimation based on machine learning methods avoids the estimation of LUE and can easily reach high accuracies when the models are fed with large, unbiased datasets. However, due to the biased distribution of flux towers, the reliability of GPP products generated in this way is often doubted in regions that are poorly represented by the samples [2]. In addition, machine learning models are hard to explain in a mechanistic way. In comparison, GPP estimated using process-based models are supported by solid mechanistic foundations. However, process-based models are usually very complicated. They often require substantial time and computing resources to generate a long-term global dataset. Due to the lack of direct photosynthetic signals as inputs, all of the models above estimate GPP somewhat indirectly, indicating the necessity to find a more direct and simple way to estimate GPP.

Recently, the emergence of solar-induced chlorophyll fluorescence (SIF) retrieval techniques has provided opportunities for the estimation of global GPP. SIF is an electromagnetic signal emitted by excited photosynthetic pigments during plant photosynthesis [6]. It exhibits a strong correlation with APAR and GPP at multiple spatial and temporal scales [7,8]. Due to the physiological information it contains, SIF outperforms many vegetation indices in tracking the photosynthetic dynamics, particularly under stressed conditions [9] or in evergreen ecosystems of which the vegetation structure shows less variation [10]. Consequently, SIF is considered a proxy for photosynthesis and widely used in regional and global studies. Furthermore, a study has shown that SIF alone can provide a rough estimate of global GPP, leading to the generation of a global GPP dataset (a GPP dataset based on global, OCO-2-based SIF product

[GOSIF] GPP) based on the empirical linear relationship between GOSIF (a reconstructed SIF dataset based on machine learning methods) and GPP [5].

However, the variability in the SIF–GPP relationship introduces uncertainty to GPP estimation in empirical SIF-based models. An increasing number of studies have demonstrated that SIF and photosynthesis do not always change in sync, and their quantitative relationship is not constant [11]. Changes in environmental and physiological factors can modulate the relationship between SIF and GPP [12,13]. SIF typically shows high consistency with APAR [14], indicating that SIF can be used to avoid the estimation of APAR and yield good performance in GPP estimation. Mechanistically, SIF and GPP are produced during different periods of photosynthesis. SIF is emitted during the light reaction, while GPP depends on the Calvin cycle in the carbon reaction. This discrepancy may be the physiological basis for their dynamic quantitative relationship. At the canopy scale, observations from flux towers demonstrate that SIF is less sensitive than GPP to stomatal control and changes in ambient environments [13,15]. Additionally, there are discrepancies in the peak time and early spring onset between SIF and GPP [13,16].

During the fall transition and winter, the quantitative relationship between SIF and GPP changes with temperature, showing nonlinearity at the seasonal scale [17,18]. Satellite-based studies have also demonstrated that the observed SIF-to-GPP ratio is seasonally dependent and influenced by moisture conditions on a global scale [19,20]. The canopy structure plays a role in the link between SIF and GPP due to its impact on the escape possibility of SIF photons (fesc) [21], thus partially explaining the biome-specific SIF–GPP relationship. While studies have shown a more consistent correlation between total SIF emitted by the photosystems and GPP across biomes [22], further quantitative study research has indicated variations in the relationship between total SIF and GPP across different developmental stages [12]. Consequently, relying solely on empirical models and SIF is not sufficient for accurate GPP estimation. To improve the precision of GPP estimation using SIF, a step forward should be taken from traditional empirical models by integrating SIF with auxiliary environmental information within a state-of-the-art mechanistic framework.

Chang et al. [23] have proposed a mechanistic light response (MLR) framework to quantitatively describe the coupling between SIF and GPP, enabling a more mechanistic estimation of GPP using SIF. According to this framework, the electron transport rate (ETR), which serves as a bridge between SIF and GPP, can be estimated using SIF emitted by the photosystem II complex (SIF_{PSII}) and active fluorescence parameters such as the opened fraction of PSII reaction centers (q_L) and the maximum photochemical quantum yield of PSII ($\Phi_{PSIImax}$). Based on the estimation of ETR, GPP can be further estimated using the description of carbon reaction in the Farquhar–von Caemmerer–Berry (FvCB) model [24,25].

In this study, our objective was to generate a mechanistic global GPP dataset based on the MLR framework using the TROPOMI SIF dataset (TROPOSIF) [26]. We employed the random forest model to incorporate environmental factors (such as air temperature, vapor pressure deficit [VPD], etc.) and canopy structure information (including leaf area index [LAI], clumping index, etc.) into the parameterization process. The resulting dataset, called CMLR GPP (GPP based on canopy-scale MLR model), was generated at a global scale with a resolution of 0.05°, covering the period from 2018 to 2021.

To assess the accuracy of the CMLR GPP dataset, we conducted validation and comparison exercises against GPP values obtained from eddy covariance towers and other global products. The performance of the model was also evaluated across different plant functional types (PFTs), months, and varying conditions of normalized difference vegetation index (NDVI), temperature, and water availability. Additionally, for the purpose of global mapping, we provided a composite GPP dataset generated using an 8-day moving window. This novel dataset offers a new opportunity to investigate global photosynthesis based on actual SIF observations (as opposed to spatially downscaled or reconstructed SIF data), thereby ensuring more reliable physiological interpretations in future applications.

Materials and Methods

Dataset

TROPOMI SIF and GPP datasets

In this study, we utilized the global SIF product obtained from the Copernicus Sentinel-5P TROPOMI mission, referred to as TROPOSIF, covering the period from May 2018 to December 2021 [26]. The sampling strategy of the TROPOMI satellite supports SIF retrieval with a nadir mode footprint size of 3.5 km × 5.5 km every day (before August 2019, 3.5 km × 7.5 km). The TROPOSIF dataset employed a data-driven approach for SIF retrieval and provided SIF values within the 743- to 758-nm (SIF743) and 735- to 758-nm (SIF735) fitting windows. Alongside SIF retrievals, the dataset also included corresponding SIF errors, top-of-atmosphere radiance (within the fitting window), reflectance, land cover information, cloud fraction, and imaging conditions (such as acquisition time, location, solar zenith angle, view zenith angle, and solar azimuth angle) for each SIF sounding. In line with recommendations in the literature [26], we selected SIF743 for this study.

GPP observations from eddy covariance flux towers were utilized for model parameterization and validation. The flux tower data needed to overlap with the temporal coverage of TROPOMI. Hence, we gathered and filtered flux data shared under CC-BY 4.0 policy from AmeriFlux (https://ameriflux.lbl.gov/), ICOS (https://www.icos-cp.eu/), TERN (https://www.tern.org.au/), and TPDC (https://data.tpdc.ac.cn/) [27,28] to obtain GPP observations from 293 towers (sites are listed in Supplementary file E1). For sites where only raw data (net ecosystem exchange data and other environmental observations) were provided, we employed gap filling and daytime partitioning methods [29] using the REddyProcWeb tool [30] to estimate GPP. As we aimed to estimate GPP based on instantaneous SIF observations, the GPP estimation in this study represents the photosynthesis intensity at the overpass time of TROPOMI (unit: μmol m^{-2} s^{-1}). Thus, we matched half-hourly GPP with SIF soundings within a 3-km radius around the flux tower and selected the closest GPP records based on the SIF acquisition time. These matched TROPOMI SIF and tower GPP observations were separated into a training dataset (70% of the total dataset, used in the model establishment process) and a validation dataset (30% of the total dataset, used in the model validation process). In addition, data from a process-based GPP product (Boreal Ecosystem Productivity Simulator [BEPS] GPP) [31], a data-driven GPP product (MODIS FluxSat GPP) [32], and an empirical SIF-based GPP product (GOSIF GPP, v2) [5] were also collected for comparison with CMLR GPP. Specifically, we extracted the values in these global GPP products based on

the locations of 293 flux sites and compared their accuracy and time series. We converted the units (from gC m^{-2} day^{-1} to μmol m^{-2} s^{-1}) of the 3 reference datasets and transformed their data into instantaneous values at the TROPOMI overpass time using the cosine of the solar zenith angle, as it predominantly determines diurnal changes in GPP and APAR [33,34]. Further details regarding this conversion can be found in Text S1.

ERA5 reanalysis dataset

We utilized the fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis dataset, ERA5 hourly data on single levels from 1940 to present [35], for model parameterization and evaluation under various conditions. This dataset provided variables such as surface solar radiation downwards, 2 m air temperature, and 2 m dewpoint temperature. There variables were selected and resampled to a 0.05° resolution using linear interpolation. Subsequently, the environmental variables were matched with SIF soundings based on the location and acquisition time. Photosynthetically active radiation (PAR) was derived from the surface solar radiation downwards (unit: J m^{-2}) by multiplying it by 0.46 and converting the unit to μmol m^{-2} s^{-1} [36]. VPD was calculated using the air temperature and dew temperature with the following equation [37,38]:

$$\mathrm{VPD} = 0.611 \times e^{\frac{17.5\times T}{240.978+T}} - 0.611 \times e^{\frac{17.5\times T_\mathrm{d}}{240.978+T_\mathrm{d}}} \quad (1)$$

where T is the air temperature (unit: °C) and T_d is the dew temperature (unit: °C). The unit of VPD is kPa, and it is converted into hPa in this study. In summary, we matched the ERA5 data with SIF soundings to describe the light, thermal, and moisture condition for each data record.

Soil background albedo, nadir surface reflectance, and LAI/fPAR datasets

We used a soil background albedo dataset sourced from MODIS satellite products [39] in conjunction with the MODIS Nadir Bidirectional Reflectance Distribution Function-Adjusted Reflectance dataset (MCD43C4 v6.1) [40] to consider the impact of soil background and compute vegetation indices in our study. In comparison to the reflectance data provided by TROPOSIF, the MCD43C4 reflectance data underwent atmospheric correction and was adjusted to the nadir direction. Here, we followed a previous study [41] to get NIRv (the near-infrared reflectance of vegetation). NIRv is the product of NDVI and near-infrared reflectance, so we calculated NDVI using the MCD43C4 reflectance data and then multiplied it with reflectance data from TROPOMI in the SIF fitting window (provided by TROPOSIF).. The MODIS Leaf Area Index/FPAR 8-Day L4 dataset (MOD15A2H) was employed to acquire fraction of photosynthetically active radiation (fPAR) for estimating fesc (calculated as the ratio of NIRv to fPAR) [42]. Additionally, the LAI information contained in this dataset was utilized to further account for the impact of canopy structure.

Other datasets

In this study, we utilized a Köppen–Geiger climate classification map [43] to determine the general climate condition associated with each observation. To account for the spatiotemporal variations in CO_2 concentration, we also used the latest gridded monthly OCO-2 carbon dioxide assimilated dataset (OCO2_GEOS_L3CO2_MONTH) [44]. Additionally, we employed a MODIS clumping index dataset [45] to calculate the effective LAI (LAI multiplied by the clumping index). All of these datasets were resampled to a resolution of 0.05° and extracted to match the SIF soundings.

To assist in distinguishing observations taken under clear-sky and overcast conditions for data quality control, we calculated the clearness index (CI). The CI is influenced by the diffused light fraction and reflects the clear-sky condition corresponding to each measurement [46]. It is defined as the ratio of the observed PAR to theoretical PAR at the top of the atmosphere ($\mathrm{PAR_{TOA}}$). The $\mathrm{PAR_{TOA}}$ is estimated by multiplying the solar constant (S_0, 1,367 W m^{-2}) by the cosine of the solar zenith angle.

Determining GPP using an MLR model

MLR model and its application at the canopy scale

The MLR model is a mechanistic model describing the quantitative relationship between SIF and GPP [25]. According to this framework, GPP can be estimated using the following equation:

$$\mathrm{GPP} = \frac{C_\mathrm{c} - \Gamma^*}{4C_\mathrm{c} + 8\Gamma^*} \frac{\left(1+k_\mathrm{DF}\right) qL \times \Phi_\mathrm{PSIImax}}{1 - \Phi_\mathrm{PSIImax}} \mathrm{SIF_{PSII}} \quad (2)$$

where C_c is the CO_2 partial pressure in the chloroplast, Γ^* is the photorespiratory compensation point of CO_2, $\mathrm{SIF_{PSII}}$ is the total full band SIF emitted by photosystem II, and k_DF is the ratio of k_D to k_F. In this study, we adopted a value of 9 for k_DF based on the work of Liu et al. [47]. The parameter Φ_PSIImax, representing the maximum quantum efficiency of the photochemical reaction in healthy plants under unstressed conditions (i.e., the maximum value during summer), was taken as 0.83. Correspondingly, q_L denoted the fraction of open PSII reaction centers (the denominator includes both active and inactivated components).

Equation 2 illustrates the method for estimating GPP in the original leaf-level MLR model [25]. To apply the MLR model at the canopy scale, a conversion is necessary from the directional observed SIF at a specific wavelength ($\mathrm{SIF_{obs}}$) to $\mathrm{SIF_{PSII}}$. This conversion is accomplished using the following equations [47]:

$$\mathrm{SIF_{PSII}} = \sum_{\lambda=640}^{850} \left(\mathrm{SIF_{PSII\_ps}} \times fc(\lambda) \times \frac{\lambda \times 10^6}{h \times c \times N_\mathrm{A} \times 10^3 \times 10^9} \right) \quad (3)$$

$$\mathrm{SIF_{PSII\_ps}} = \frac{\mathrm{SIF_{obs}} \times \mathrm{fPSII}}{\mathrm{fesc} \times f_\mathrm{Lp}} \quad (4)$$

where $\mathrm{SIF_{PSII\_ps}}$ is the total PSII SIF at the photosystem level at the observed wavelength, fesc is the escape probability from leaf surface to the top of canopy, f_Lp is the escape probability from the photosystems to the leaf surface, fc(λ) is the factor converting SIF at observed wavelength to SIF at the λ wavelength, and fPSII is the ratio of PSII SIF to the total SIF at the observed wavelength. In Eq. 3, h, c, λ, and N_A represent the Planck constant, the speed of light, the wavelength, and the Avogadro constant, respectively. According to previous studies [48,49], f_Lp is regarded as 0.9. The determination of fc(λ) uses the first principal component of 6,720 photosystem-level SIF simulations from the Soil Canopy Observation, Photochemistry

and Energy fluxes (SCOPE) model, as described in the literature [47]. Additionally, fPSII is determined using the SCOPE simulations, which encompass various leaf properties, canopy structures, and geometry conditions. By dividing the PSII SIF simulations by the total SIF (including both PSII and PSI SIF) simulations, a range of fPSII values at the observation wavelength are obtained and subsequently averaged. In this study, we used the SIF743, and therefore, fPSII was set as 0.6819 to derive the PSII SIF at the 743-nm wavelength. By combining Eqs. 2, 3, and 4, GPP was estimated based on top-of-canopy SIF observations.

Parameterization of q_L

As a highly dynamic variable, q_L plays a crucial role in the MLR model [25]. Leaf-scale measurements have shown that q_L is mainly influenced by ambient light conditions and exhibits a negative exponential relationship with PAR [50]. However, unlike leaf-scale studies, q_L in this research represents a canopy-scale concept that reflects the community-level outcome within a pixel. Consequently, its environmental response may differ from leaf-scale findings, posing challenges in the direct determination of q_L values under various conditions.

In this study, q_L is linked not only to the redox state of active PSII reaction centers (representing the fraction of open PSII reaction centers to the active PSII reaction centers), but also to the fraction of active PSII reaction centers in relation to the total number of PSII reaction centers. For this reason, environmental factors that induce growth pressure or even stress can impact the magnitude of q_L. For a pixel in satellite products, both canopy structure and soil background can provide insights into the vegetation coverage and density in a given region, which potentially influence light interception and redistribution within the canopy. Because the canopy is not a simple 2-dimensional plane, the actual light conditions driving overall canopy photosynthesis differ from the light at the canopy top, although the former is constrained by the latter. Thus, the soil albedo and canopy structure have potential impacts on canopy q_L, making its response different from leaf-scale results. SIF can also reflect light interception within the canopy and aid in parameterizing canopy q_L due to its high correlation with APAR. Furthermore, SIF contains physiological information about the vegetation, which can be utilized to account for the physiological-related variation in q_L. All of the factors above contribute to the variation in q_L and should be considered in its parameterization.

Given the limited knowledge and complexity in determining canopy q_L, we employed a machine learning method for its parameterization. First, we calculated canopy q_L values using Eq. 2 based on the matched dataset comprising TROPOMI SIF and tower GPP. Then, we took 70% of this matched dataset to train a random forest model, and the remaining 30% data served as the validation dataset. Considering the potential contributions of environmental factors, canopy structure, land cover, geospatial location, and climate zones, we tested various input combinations and ultimately selected variables including total SIF, solar zenith angle, air temperature, PAR, VPD, CO_2 concentration, NDVI, soil albedo, effective LAI (eLAI), land cover, Köppen-Geiger climate class, and the position of soundings as the model inputs to estimate canopy q_L. Notably, these variables were selected to avoid underfitting and improve the estimation accuracy, rather than for explanatory purposes. Therefore, this machine learning model is not suitable for explaining the individual contributions of each factor to q_L. To ensure the model learned from data under different vegetation cover conditions, we also randomly augmented the data with both high NDVI and low NDVI before training.

Parameterization of C_c in the carbon reaction

For simplicity, C_c was assumed to be equivalent to C_i (intercellular CO_2 concentration), and the parameterization of C_i was accomplished using method based on eco-evolutionary optimality theory [51–53]:

$$C_i = \frac{\Gamma^*}{C_a} + \left(1 - \frac{\Gamma^*}{C_a}\right)\frac{\xi}{\xi + \sqrt{\mathrm{VPD}}}\,C_a, \text{ with } \xi = \sqrt{\frac{\beta\left(K + \Gamma^*\right)}{1.6\eta^*}} \tag{5}$$

In this equation, C_a represents the atmospheric CO_2 concentration (unit: Pa), K represents the Michaelis–Menten coefficient for photosynthesis, β is a constant representing the unit cost ratio (taken as 146), and η^* is the relative water viscosity. Both K and Γ^* were estimated using air temperature, and their units were converted into Pa. The calculation was performed with the help of functions in the rpmodel package of *R* language [53].

Quality control, model validation, and data format

Stringent quality control measures were implemented prior to model training. Given the inherent uncertainties in both satellite SIF and tower-based GPP observations, errors in the calculated canopy q_L from these sources are inevitable. To ensure the reliability of the data, records with q_L (calculated using TROPOMI SIF and tower GPP) exceeding 1 and or falling below 0 were excluded from the training dataset, based on the prior knowledge that q_L values typically range from 0 to 1. During the training process, additional filtering criteria were applied to the data to enhance their reliability. These criteria included the following: (1) cloud fraction < 0.05, (2) CI > 0.7, (3) SIF error < 0.4, (4) GPP > 0, and (5) SIF > 0. In the production of the dataset, soundings with cloud fraction higher than 0.3, fesc lower than 0.1, and SIF lower than −0.1 were excluded. The quality control measures in this step were slightly relaxed to prevent significant spatial gaps in the output caused by excessively strict filtering. The estimated GPP values were ultimately aggregated into 0.05° grids, and the original CMLR GPP dataset was generated on a daily basis.

To assess the CMLR model, we initially compared its performance (the accuracy of q_L and GPP estimations) against empirical models using the validation dataset. Subsequently, we matched pixels extracted from BEPS GPP, MODIS FluxSat GPP, GOSIF GPP, and CMLR GPP, and conducted validation to evaluate the model performance and patterns under different conditions on a global scale.

For the global mapping application, an additional composite dataset was generated using an 8-day temporal moving window. In this dataset, pixels with LAI = 0 were assigned GPP = 0. To address spatial gaps resulting from the absence of model inputs or poor data quality in certain pixels, we filled them using a 2-dimensional Gaussian function within a 3×3 moving window, following a previous study [54]. These filling values were marked in the quality control layer (Table).

These 2 types of datasets (the original CMLR GPP and the 8-day composite CMLR GPP) are all organized in the GeoTiff format, and the composited version contains 2 bands: the first

Table. Flag values and their meaning in the quality control layer

	Types of the pixels	Value
0	Original estimates	0
1	Filling data	1
2	Missing data	255

one is the GPP layer and the second one is the quality control layer.

Results

Spatial pattern of the CMLR GPP dataset

Figure 1 illustrates the spatial pattern of CMLR GPP on 2019 July 29. In Fig. 1A, the global map highlights regions with high productivity, such as the USA Corn Belt, Amazon Rainforest, Congo Rainforest, and islands in Southeast Asia among others. In Fig. 1B and C, although there are some noisy pixels, the dataset effectively captures the elevated GPP in the Nile Delta and regions along the Mississippi River. This overall spatial pattern aligns with the findings of other global GPP datasets.

Comparing CMLR and empirical methods using the validation dataset

We conducted tests on the CMLR model and compared it with empirical methods using the validation dataset. As displayed in Fig. 2A, the GPP estimated by the CMLR model exhibits a high correlation with tower-based observations ($R^2 = 0.72$), and the GPP estimation demonstrates an unbiased performance in both high and low ranges, without significant overestimation or underestimation. The majority of GPP data fall within the range of 0 to 25 μmol m^{-2} s^{-1}, and the densest cluster of points in Fig. 2A is located closest to the 1:1 line. In Fig. 2B, there is a relatively high correlation between the estimated q_L using the random forest model and the calculated q_L based on the tower GPP and TROPOMI SIF observations ($R^2 = 0.69$). This suggests that, for most TROPOMI overpass times, the q_L parameter falls within the range of 0 to 0.2. Compared to the estimation of GPP in Fig. 2A, the estimation of q_L in Fig. 2B is more biased. Therefore, it appears that a less satisfying R^2 in the q_L estimation does not necessarily limit the R^2 in the GPP estimation.

Figure S1 illustrates the performances of different empirical models in the validation dataset. GPP estimates from all of these models exhibit low correlations with the tower-based observations, despite their regression lines being close to the 1:1 line. However, compared to the CMLR GPP estimates in Fig. 2A, the GPP estimates from these empirical models display more bias. They demonstrate significant overestimations in the low GPP range and underestimations in the high GPP range. Although this issue is less pronounced in the model based on the nonlinear total SIF–GPP relationship (Fig. S1D), it still exhibits a lower R^2 than the models based on the observed SIF–GPP relationship. Regarding the model comparison in different PFTs, Figs. S2 to S8 and Table S1 reveal that methodology employed in CMLR GPP shows the most significant improvement in R^2 for evergreen forest (EF) and grassland (GRA). Additionally, this approach leads to the most substantial reduction in the root mean square error (RMSE) for cropland (CRO).

Comparing CMLR GPP with other global GPP products

In this section, we conducted a comparison between the CMLR GPP, BEPS GPP, FluxSat GPP, and GOSIF GPP by extracting their pixel values based on the locations of 293 flux towers. Figure 3 presents the validation results for these 4 datasets. Among these models, BEPS GPP demonstrates the highest RMSE (8.48 μmol m^{-2} s^{-1}) and the lowest R^2 value (0.36), with its regression line noticeably biased from the 1:1 line. On the other hand, FluxSat GPP exhibits a higher R^2 (0.51) and a lower RMSE (7.26 μmol m^{-2} s^{-1}) compared to BEPS GPP. GOSIF GPP yields similar results to FluxSat GPP, slightly outperforming FluxSat GPP with an R^2 of 0.51 and an RMSE of 7.09 μmol m^{-2} s^{-1}. Among the 4 datasets, CMLR GPP exhibits the highest R^2 (0.55), the lowest RMSE (6.69 μmol m^{-2} s^{-1}), and a regression line that is closest to the 1:1 line (slope: 1.05, intercept: −0.21 μmol m^{-2} s^{-1}).

To assess the performance of the 4 global GPP datasets in temporal analysis, we compared their GPP time series and presented the results for a deciduous broadleaf forest site (US-Ha1) and a CRO site (US-Bi2) in Fig. 4. For the US-Ha1 site, all of the GPP datasets successfully captured the dynamics of tower-based GPP observations. However, GOSIF GPP showed significant overestimation during the peak growing season from June to September with value surpassing above 35 μmol m^{-2} s^{-1}. At the US-Bi2 site, both GOSIF GPP and FluxSat GPP underestimated photosynthesis during summer, with GOSIF GPP showing a more pronounced underestimation. In contrast, CMLR GPP exhibited a high consistency with tower-based GPP values. We also compared these 4 datasets at the high-latitude evergreen needleleaf forest (ENF) sites (Fig. S9) and tropical sites (Fig. S10). CMLR generally shows comparable performance with the reference datasets.

Figure S11 and Fig. 5 illustrate the global averages of GPP and their time series across different PFTs. As shown in Fig. S11, CMLR GPP and GOSIF GPP generally exhibit similar patterns across many PFTs, except for the EF. FluxSat GPP tends to yield higher values compared to CMLR GPP and GOSIF GPP, except for the GRA and shrubland (SH) PFTs. In CROs and needleleaf forests, GOSIF GPP and CMLR GPP demonstrate high consistency, while BEPS GPP and FluxSat GPP tend to exhibit higher values during peak periods (Fig. 5A, C, and E). In broadleaf forests including both evergreen and deciduous forests, CMLR GPP shows relatively low values, while BEPS GPP, FluxSat GPP, and GOSIF GPP show similar patterns (Fig. 5B, D, and F). However, this observation may be attributed to the overestimation of GOSIF GPP, as evidenced by its overestimation (especially in peaks) at the US-Ha1 site (Fig. 4C).

Model performance in different conditions

Figure 6 shows the pattern of model performance in the temperature–VPD space. Both CMLR GPP and GOSIF GPP exhibit an increase in adjusted R^2 (between estimated and observed GPP) as temperature and VPD increase. The highest adjusted R^2 values are observed when temperature ranges from 30 to 35 °C and VPD ranges from 30 to 35 hPa (Fig. 6A and B). The result for CMLR GPP also indicates a slight decrease in adjusted R^2 when VPD exceeds 30 hPa, whereas this pattern is not discovered

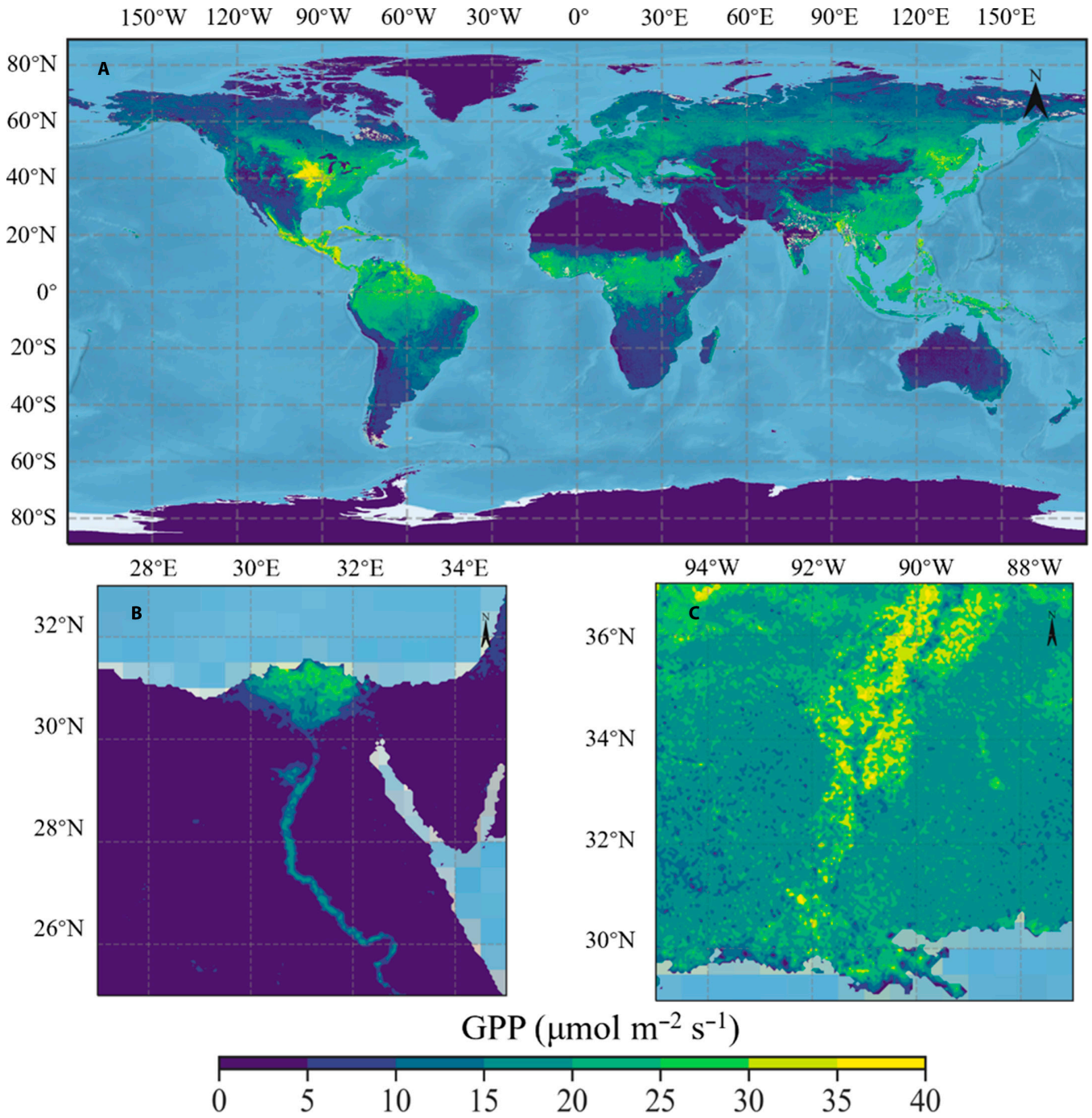

Fig. 1. Spatial pattern of CMLR GPP on 29 July 2019 at a 0.05° resolution. (A) The global map of CMLR GPP. (B) The GPP pattern in the Nile Delta. (C) The GPP pattern in the Mississippi River. The 8-day composited dataset was used for global mapping.

in GOSIF GPP. In Fig. 6C and D, as the temperature and VPD increase, RMSE initially increases and then decreases, which may be attributed to higher vegetation productivity under mild temperature and moisture conditions. Comparing Fig. 6A and C with Fig. 6B and D, it can be inferred that the CMLR GPP achieves the highest adjusted R^2 (0.64), but it does not show superiority over GOSIF GPP under cold and wet conditions (temperature lower than 10 °C and VPD lower than 10 hPa).

Figures 7, 8, and 9 present the model performance across different canopy structures (indicated by NDVI), ecosystems (indicated by PFT), and periods (indicated by month), respectively. In Fig. 7, CMLR GPP consistently exhibits the highest R^2 and the lowest RMSE among all NDVI conditions, while BEPS GPP shows the lowest R^2 and the highest RMSE compared to the other GPP datasets. The R^2 values of FluxSat GPP and GOSIF GPP show similar patterns, initially increasing and then decreasing as NDVI rises from 0.3 to 0.8. In Fig. 8A, CMLR GPP demonstrates the highest R^2 among the 4 GPP datasets for EF, GRA, and CRO. In Fig. 8B, CMLR GPP shows the lowest RMSE compared to other GPP datasets for EFs,

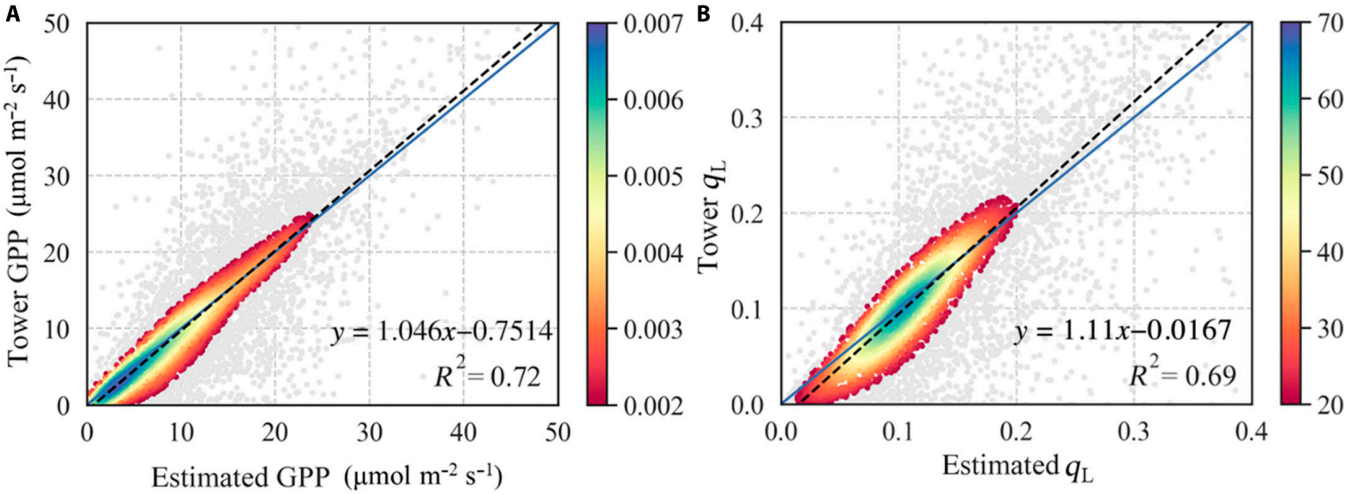

Fig. 2. Comparison of the CMLR estimations with the tower observations in the validation dataset. (A) Scatterplot of the estimated GPP and tower-based GPP. (B) Scatterplots of the q_L estimated by random forest model and q_L calculated using tower GPP and TROPOM SIF. The color of dots represents the data density, and dots with a density lower than 0.002 in (A) and 20 in (B) are marked in gray. The dashed black and solid blue lines represent the regression line and the 1:1 line, respectively.

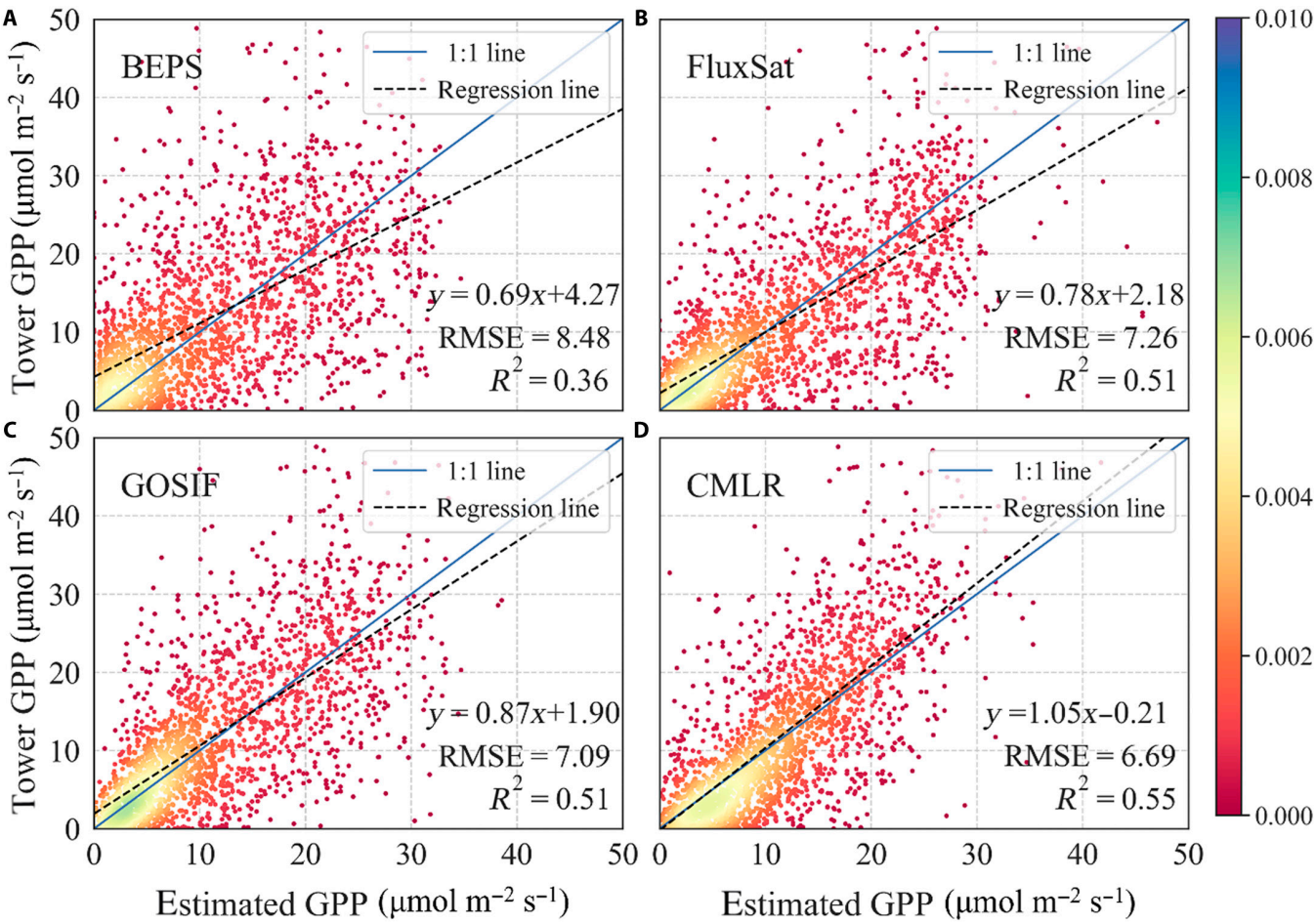

Fig. 3. Validation of GPP estimates using all matched tower-based measurements. (A) BEPS GPP, (B) FluxSat GPP, (C) GOSIF GPP, and (D) CMLR GPP. The color of dots represents the data density. The dashed black and solid blue lines represent the regression line and the 1:1 line, respectively.

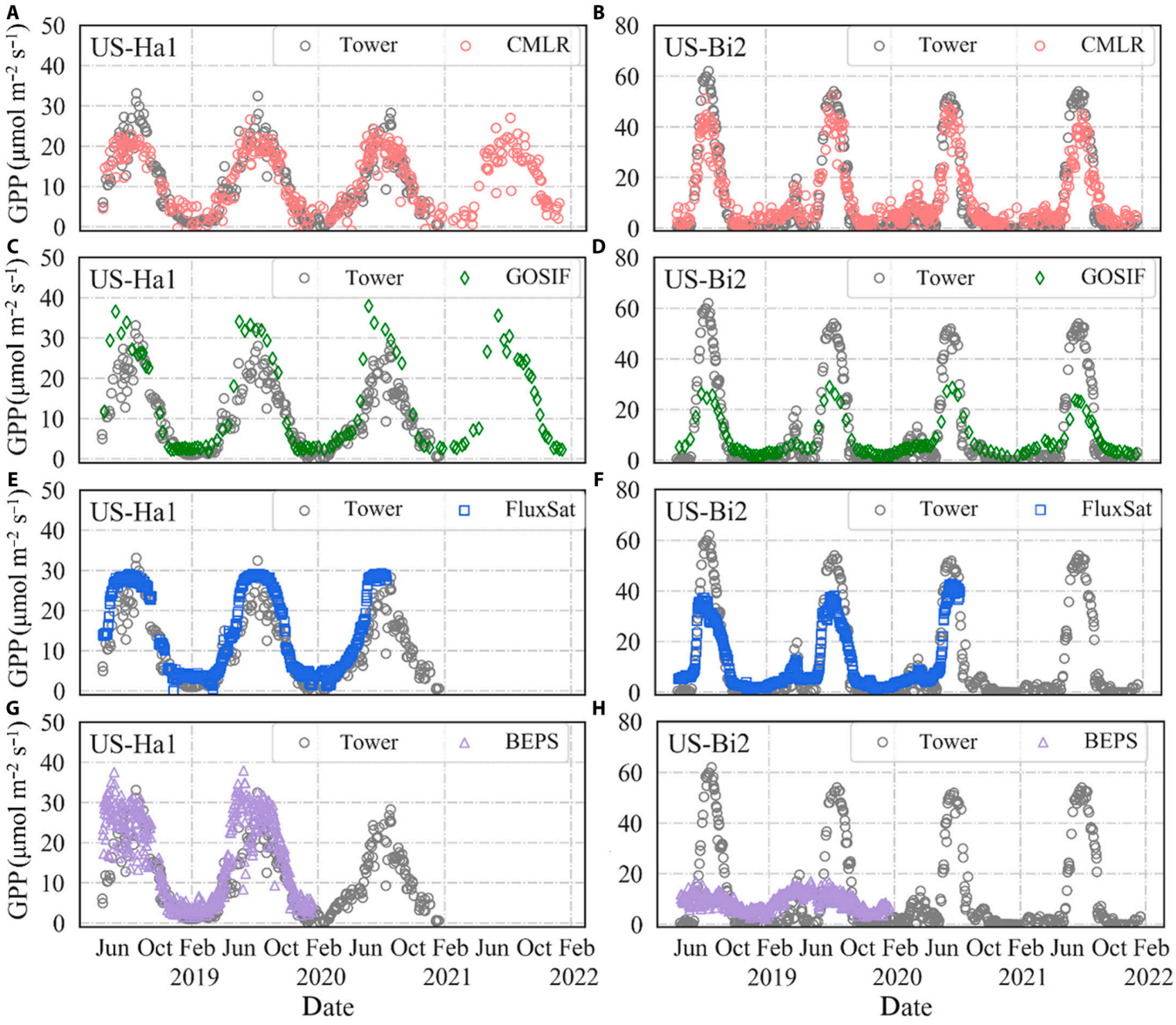

Fig. 4. Time series of CMLR GPP (A and B), GOSIF GPP (C and D), FluxSat GPP (E and F), and BEPS (G and H) in comparison with long-term continuous GPP observations at 2 representative towers. The first column displays the results at the US-Ha1 tower (deciduous broadleaf forest, DBF), and the second column displays the results at the US-Bi2 tower (cropland, CRO).

deciduous forests, GRA, and CRO. FluxSat GPP proves to be the most accurate dataset for mixed forests, while GOSIF GPP performs best for SH. Finally, upon analyzing the model performances in different months in Fig. 9, we observe that CMLR GPP consistently outperforms the other datasets, except for April and September, which typically represent transition periods for natural vegetation.

Discussion

Difference between CMLR and existing SIF-based GPP estimations

Although there are numerous GPP models and datasets available [3,4], only a small fraction of them incorporate SIF. However, SIF holds great promise as a tool for global photosynthetic mapping and other applications in Earth system science [55]. The results of this study ("Comparing CMLR and empirical methods using the validation dataset" and "Comparing CMLR GPP with other global GPP products" sections) demonstrate that the 2 SIF-based datasets, CMLR GPP and GOSIF GPP, outperform BEPS GPP and FluxSat GPP in most cases. GOSIF GPP is an empirical dataset that sorely utilizes SIF to estimate GPP, but many studies have shown that the relationship between SIF and GPP is dependent on ecosystem and environmental factors. Additionally, GOSIF is a reconstructed dataset generated from reflectance and environmental variables using machine learning methods. Despite GOSIF GPP exhibiting a high signal-to-noise ratio and appearing spatially smooth, it is important to note that it is based on SIF simulations rather than direct observations [41]. This limitation may impede the in-depth interpretation of results in subsequent studies. Using these data to analyze the physiological response of vegetation patterns may inadvertently lead to over-interpretation of the data, due to uncertainties in the physiological information

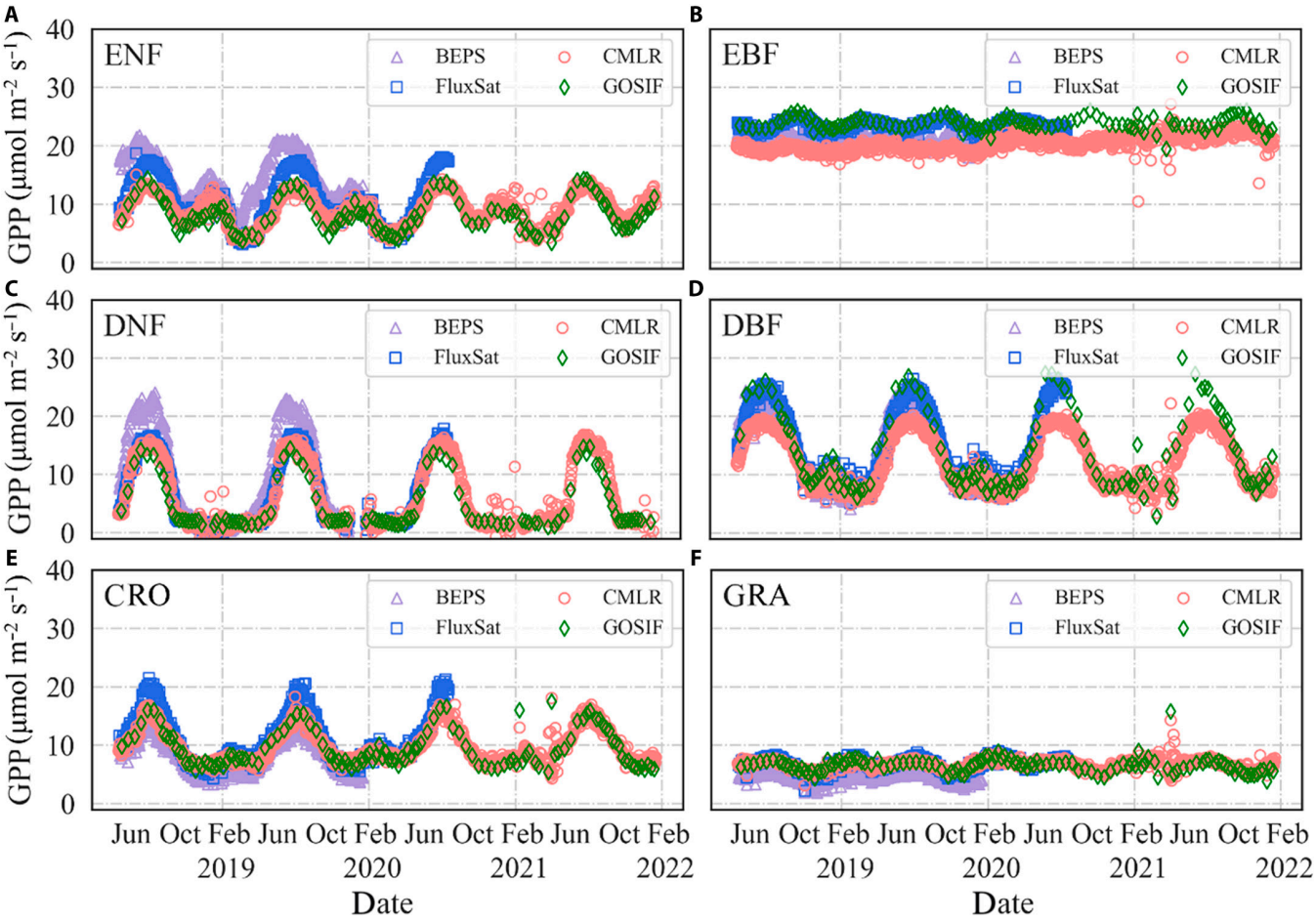

Fig. 5. Global averaged time series of BEPS GPP, CMLR GPP, GOSIF GPP, and FluxSat GPP in various plant functional types (PFTs). (A) ENF: evergreen needleleaf forest, (B) EBF: evergreen broadleaf forest, (C) DNF: deciduous needleleaf forest, (D) DBF: deciduous broadleaf forest, (E) CRO: cropland, and (F) GRA: grassland.

it contains. In contrast, the application of CMLR GPP avoids this concern because it directly used TROPOMI SIF observations. For the comparison of the new GPP dataset with reference GPP products, we used the cosine of solar zenith angle to get the instantaneous value. This conversion might undermine the accuracy of these reference GPP products, for it assumes a clear-sky condition. However, the uncertainties differ according to the algorithms to estimate GPP. For example, in the model training process of GOSIF, OCO-2 SIF at the overpass time was converted to daily value using pure geometric incoming light scaling. This means the conversion method used in this study is exactly an inverse process, and thereby causing limited impacts to the results.

There are also studies employing SIF observations as a constraint to optimize process-based models or calculate essential parameters [56,57]. In these studies, SIF is usually an output used for data assimilation, and such models simulate SIF and GPP by incorporating descriptions of many complex processes. However, the mechanisms and processes through which NPQ affects SIF and GPP in reality may be more intricate than those described in the models. In other words, the accuracy of GPP obtained in this way is constrained by the completeness of the process description and the accuracy of SIF simulation in these models. In contrast, our model concentrates solely on modeling the quantitative relationship between SIF and GPP. We directly utilize SIF as an input and avoid modeling SIF, NPQ, and their responses. Therefore, our approach involves fewer parameters and is somewhat less computationally intensive compared to complex assimilation models.

Uncertainties and prospects

The CMLR GPP product is derived from TROPOMI SIF observations and represents instantaneous GPP at the time of TROPOMI overpass (unit: μmol m^{-2} s^{-1}). It has not been converted to a daily product because both SIF and canopy q_L exhibit diurnal patterns, which are influenced by changes in the light intensity and possibly incident angle throughout the day. Consequently, converting instantaneous GPP to a daily value would require a comprehensive understanding of the diurnal patterns of canopy q_L, which is a topic to be studied in the future. In this study, we used the big-leaf assumption, which is suitable for the investigation of the spatial and seasonal patterns [58]. However, the uncertainty introduced by this assumption becomes significant if this approach is applied to satellite products capable of tracking daily changes in SIF (such as OCO-3) in the future.

The accuracy of CMLR GPP is affected by the noise present in TROPOMI SIF. Due to the relatively large swath, low signal-to-ratio, and spectral resolution limitations of TROPOMI, there are uncertainties in SIF inputs, which are fully inherited by our

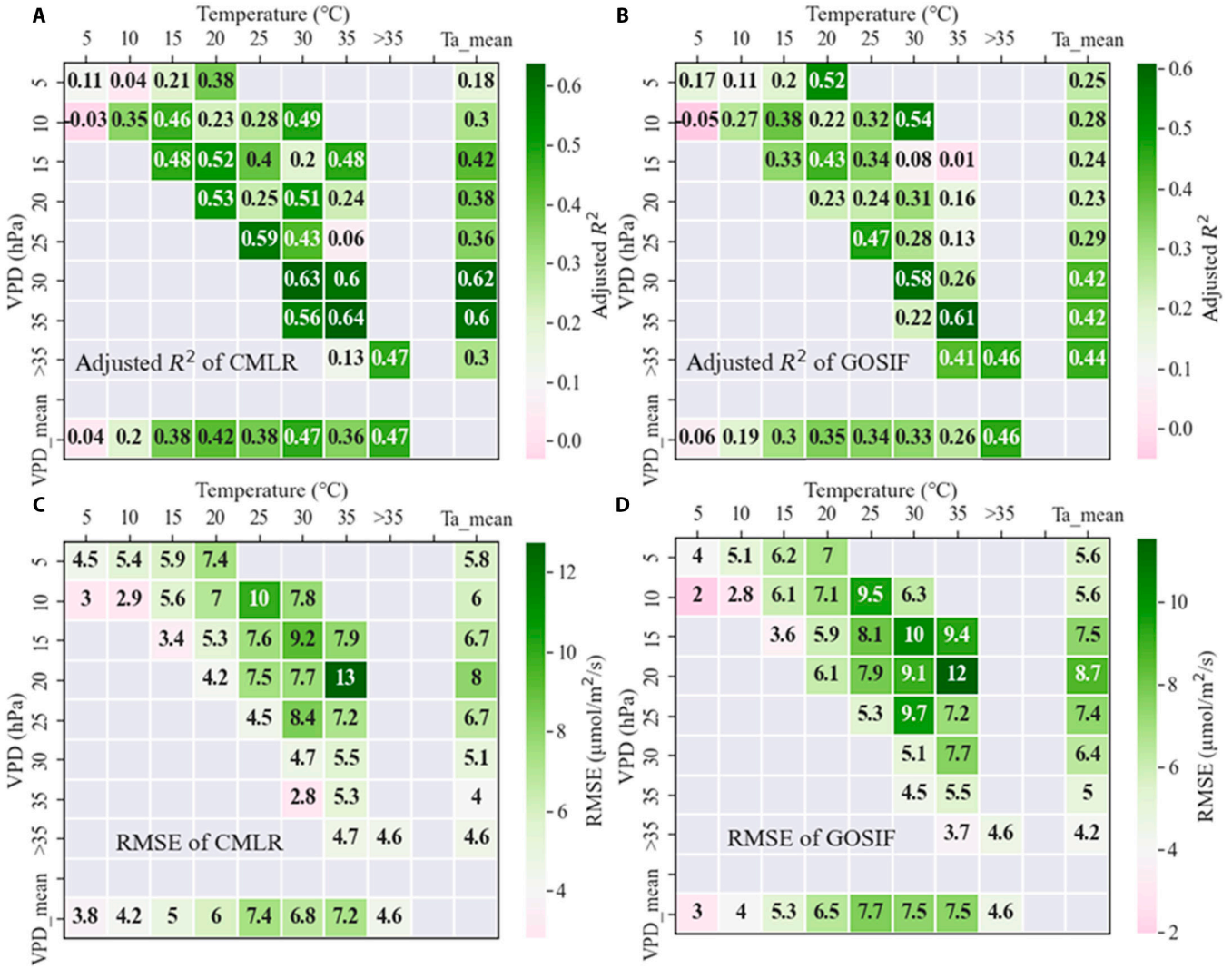

Fig. 6. Model performances under different combination of temperature and VPD conditions. (A) Adjusted R^2 of CMLR GPP in the temperature–VPD space; (B) adjusted R^2 of GOSIF model in the temperature–VPD space; (C) RMSE of CMLR GPP in the temperature–VPD space; (D) RMSE of GOSIF model in the temperature–VPD space. VPD_mean represents the mean adjusted R^2 or RMSE along the VPD axis, and Ta_mean represents the mean adjusted R^2 or RMSE along the temperature axis. On account of different data volumes in each temperature and VPD bin, we used adjusted R^2 instead of R^2 and filtered the bins with less than 15 records.

data product. The sun-sensor geometry and cloud cover can also influence the accuracy of SIF retrievals and thereby CLMR GPP values, although we tried to use strict rules to filter many unreliable records. In comparison, GOSIF GPP uses GOSIF, which is a reconstruction dataset with high signal-to-noise ratio. Therefore, the spatial pattern of GOSIF GPP appears smoother than that of CMLR GPP (although validation results show higher accuracy for CMLR GPP in most cases). When replacing the dataset with its monthly average, we observed an increase in GPP accuracy but a decrease in q_L accuracy (Fig. S12). This discrepancy is likely because q_L is an instantaneous variable that cannot be well modeled using a monthly average, while the estimation accuracy of GPP improves mainly due to reduced uncertainty in monthly SIF. Therefore, it is reasonable to expect that the accuracy of GPP calculated using this method will increase as the accuracy of the satellite SIF product improves or as better satellite instruments designed for SIF measurement are launched. Although the CMLR GPP is currently available only for the period from 2018 to 2021, it is likely to contribute to the study of long-term photosynthesis as satellite SIF data accumulate in the future.

While the CMLR model is supported by mechanistic frameworks, our limited knowledge of canopy q_L necessitated its parameterization using a machine learning method, making the CMLR model essentially semi-empirical. As with many machine-learning-based studies, the uneven distribution of samples (in this case, flux sites) may also impact our results. However, because we employ the random forest algorithm solely to estimate canopy q_L, SIF still predominantly explains the variation in CMLR GPP. Hence, the potential bias in canopy q_L has limited impact on CMLR GPP. Although machine learning is often considered a black box, the CMLR model is not a black box because the mechanistic link between SIF and GPP is described within the MLR framework. Nonetheless, further investigation of canopy q_L is warranted in the future. Additionally, incorporating the latest theories in photosynthetic physiology and improving the modeling of the link between light and carbon reactions should be considered to enhance the model

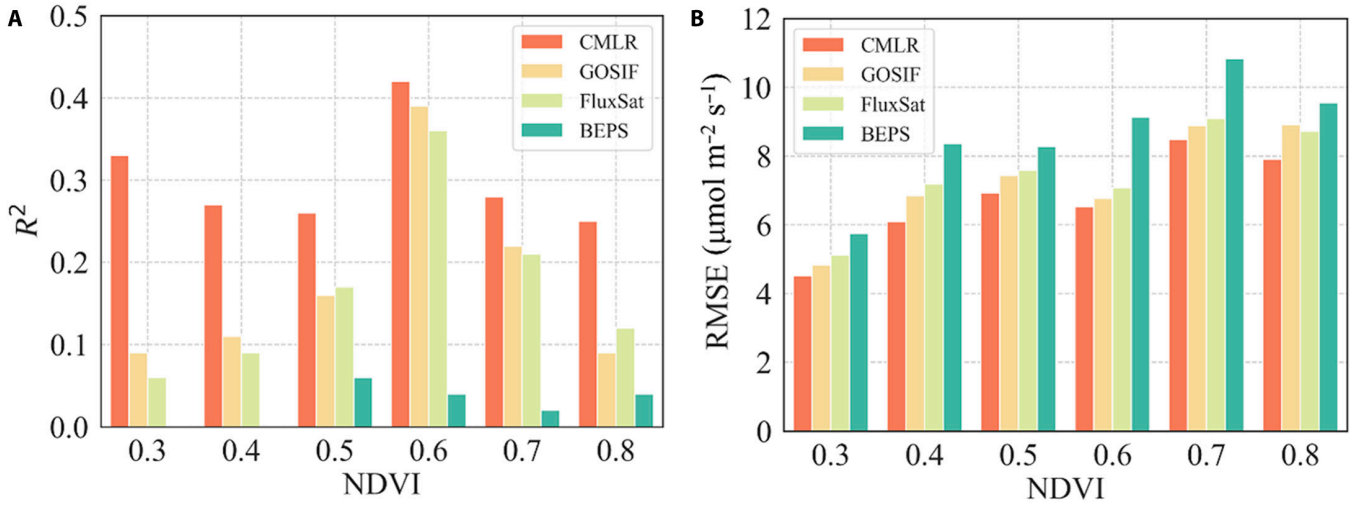

Fig. 7. Model performances under different NDVI conditions (which is related to canopy structure and reflects the vegetation density in each pixel). (A) The R^2 of GPP estimates from each dataset with the tower-based GPP measurements in different NDVI conditions; (B) the RMSE of GPP estimates from each dataset with the tower-based GPP measurements in different NDVI conditions.

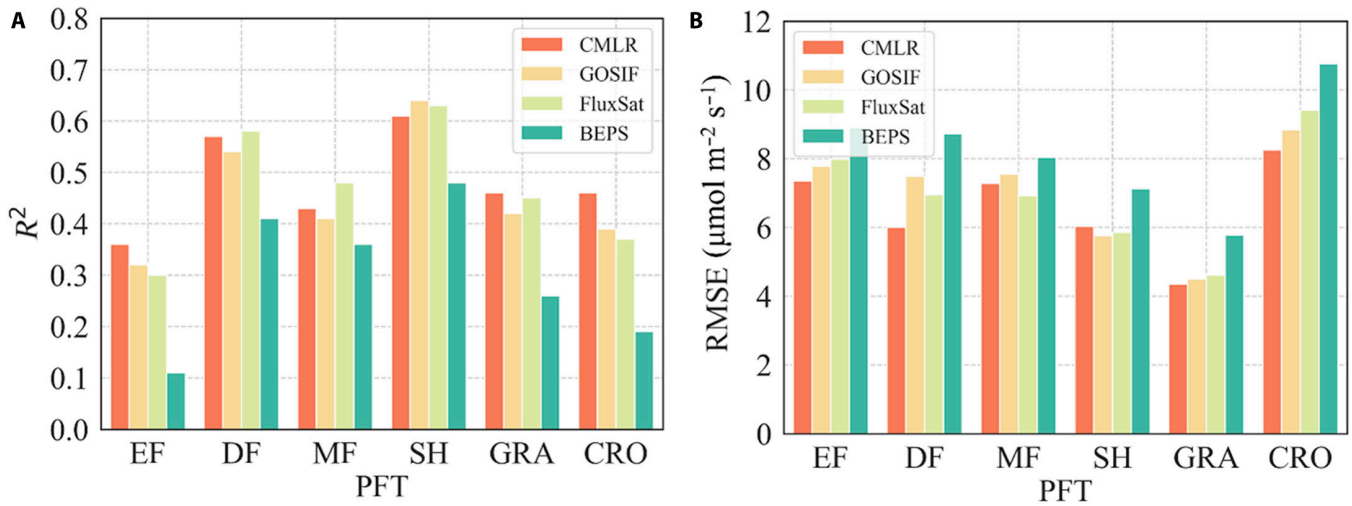

Fig. 8. Model performances in different PFTs. CRO: cropland, DF: deciduous forest, including both deciduous broadleaf and needleleaf forests, EF: evergreen forest, including both evergreen broadleaf and needleleaf forests, GRA: grassland, MF: mixed forest, SH: shrubland, including open shrubland, closed shrubland, savanna, and woody savanna. (A) The R^2 of GPP estimates from each dataset with the tower-based GPP measurements in different PFTs; (B) the RMSE of GPP estimates from each dataset with the tower-based GPP measurements in different PFTs.

performance and extend its applicability for GPP estimation in stressful environments.

Overall, CMLR GPP performed well in EFs and GRA but showed lower performance in SH. The accuracy of the dataset should be improved specifically in April and September, as well as under particularly wet and cold conditions. Therefore, caution is advised when analyzing data in these conditions.

Summary

In this study, we developed a global 0.05° SIF-based GPP dataset (CMLR GPP) using the MLR model and TROPOMI SIF observations. To parameterize q_L (fraction of open photosystem II reaction centers) in the CMLR model, we trained a random forest model using several variables that represent environmental conditions, canopy structure, land cover, geospatial location, and climate zones. The GPP estimates obtained using the CMLR method showed a high correlation with tower GPP ($R^2 = 0.72$) in the validation dataset. On a global scale, CMLR GPP successfully captured spatial and temporal patterns of GPP and exhibited similar global averages to GOSIF GPP across different PFTs. At the global scale, CMLR GPP also showed R^2 (0.55) and RMSE (6.69 μmol m^{-2} s^{-1}) comparable with those in BEPS GPP, FluxSat GPP, and GOSIF GPP datasets. This superior performance was consistent across different time periods and under various NDVI, temperature, and VPD conditions.

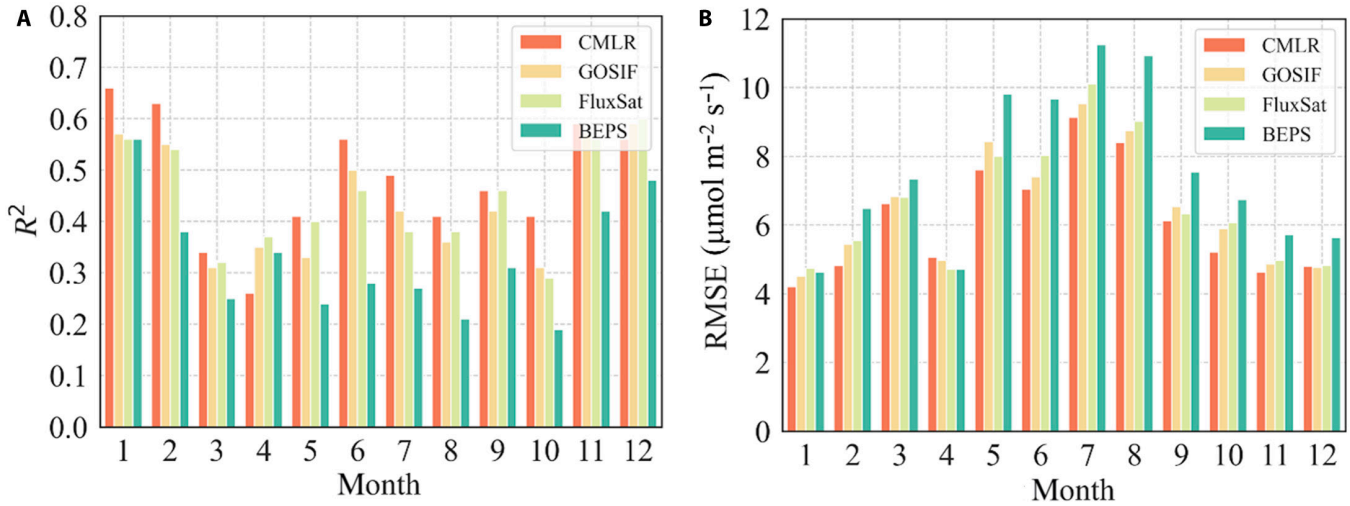

Fig. 9. Model performances in different months. (A) The R^2 of GPP estimates from each dataset with the tower-based GPP measurements in different months; (B) the RMSE of GPP estimates from each dataset with the tower-based GPP measurements in different months.

Acknowledgments

We appreciate Prof. Juliana Bendig from Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich GmbH, for providing valuable computing resources during the data processing progress. We are also grateful to the AmeriFlux (https://ameriflux.lbl.gov), ICOS (https://www.icos-cp.eu/), TERN (https://www.tern.org.au/), and TPDC (https://data.tpdc.ac.cn/) platforms for providing valuable flux datasets, and to the data producers of BEPS GPP, FluxSat GPP, and Prof. Jingfeng Xiao and Prof. Xing Li for sharing the GOSIF GPP dataset. We also appreciate Prof. Luis Guanter and many other scholars related to the TROPOSIF and TROPOMI program for providing TROPOSIF and making it available, and Prof. Ziti Jiao and the related group for providing the clumping index dataset. We thank the PIs for sharing their data through the AmeriFlux, ICOS, TERN, and TPDC program (https://ameriflux.lbl.gov), and their names are listed in the Supplementary E1. The PI for all the NEON sites (site ID starting with "US-x" and "PR-x") is National Ecological Observatory Network neon-ameriflux@battelleecology.org - NEON Program, Battelle. Funding for the AmeriFlux data portal was provided by the U.S. Department of Energy Office of Science. The BEPS GPP Dataset is provided by the National Ecosystem Science Data Center, National Science & Technology Infrastructure of China. (http://www.nesdc.org.cn); the FluxSat GPP dataset was downloaded from https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1835; the GOSIF GPP dataset was downloaded from https://globalecology.unh.edu/data/GOSIF-GPP.html. The Köppen–Geiger Global 1-km climate classification map is obtained from https://www.gloh2o.org/koppen/; the OCO2_GEOS_L3CO2_MONTH data were downloaded from https://disc.gsfc.nasa.gov/. The MCD43C4 product is available at https://earthdata.nasa.gov/. ERA5 datasets are available at https://cds.climate.copernicus.eu/.

Funding: This research was supported by the National Natural Science Foundation of China (grant numbers 41825002 and 42071310). This work has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2070 – 390732324.

Author contributions: R.C. and L.L. conceived the research; R.C. performed the data preprocessing and analyzed the data. R.C wrote the manuscript; L.L., X.L., and U.R. discussed the results, and reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Competing interests: The authors declare that they have no competing interests.

Data Availability

The data supporting the findings of this study are available from dataset "CMLR: A mechanistic global 0.05° Gross Primary Production dataset using TROPOMI Solar-induced fluorescence observations" (link: https://doi.org/10.5281/zenodo.7954622) at the Zenodo website.

Supplementary Materials

Text S1
Table S1
Figs. S1 to S12
File E1

References

1. Smith NG, Malyshev SL, Shevliakova E, Kattge J, Dukes JS. Foliar temperature acclimation reduces simulated carbon sensitivity to climate. *Nat Clim Chang*. 2016;6(4):407–411.
2. Ryu Y, Berry JA, Baldocchi DD. What is global photosynthesis? History, uncertainties and opportunities. *Remote Sens Environ*. 2019;223:95–114.
3. Zheng Y, Shen R, Wang Y, Li X, Liu S, Liang S, Chen J, Ju W, Zhang L, Yuan W. Improved estimate of global gross primary production for reproducing its long-term variation, 1982–2017. *Earth Syst Sci Data*. 2020;12(4):2725–2746.
4. Pei Y, Dong J, Zhang Y, Yuan W, Doughty R, Yang J, Zhou D, Zhang L, Xiao X. Evolution of light use efficiency models: Improvement, uncertainties, and implications. *Agric For Meteorol*. 2022;317:Article 108905.
5. Li X, Xiao J. Mapping photosynthesis solely from solar-induced chlorophyll fluorescence: A global, fine-resolution

dataset of gross primary production derived from OCO-2. *Remote Sens*. 2019;11(21):2563.
6. Porcar-Castell A, Tyystjärvi E, Atherton J, Van Der Tol C, Flexas J, Pfündel EE, Moreno J, Frankenberg C, Berry JA. Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: Mechanisms and challenges. *J Exp Bot*. 2014;65(15):4065–4095.
7. Sun Y, Frankenberg C, Wood JD, Schimel DS, Jung M, Guanter L, Drewry DT, Verma M, Porcar-Castell A, Griffis TJ, et al. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science*. 2017;358(6360):eaam5747.
8. Mohammed GH, Colombo R, Middleton EM, Rascher U, van der Tol C, Nedbal L, Goulas Y, Pérez-Priego O, Damm A, Meroni M, et al. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote Sens Environ*. 2019;231:Article 111177.
9. Song L, Guanter L, Guan K, You L, Huete A, Ju W, Zhang Y. Satellite sun-induced chlorophyll fluorescence detects early response of winter wheat to heat stress in the Indian indo-Gangetic Plains. *Glob Chang Biol*. 2018;24(9):4023–4037.
10. Pierrat Z, Magney T, Parazoo NC, Grossmann K, Bowling DR, Seibt U, Johnson B, Helgason W, Barr A, Bortnik J, et al. Diurnal and seasonal dynamics of solar-induced chlorophyll fluorescence, vegetation indices, and gross primary productivity in the boreal Forest. *J Geophys Res Biogeosci*. 2022;127(2):e2021JG006588.
11. Magney TS, Barnes ML, Yang X. On the covariation of chlorophyll fluorescence and photosynthesis across scales. *Geophys Res Lett*. 2020;47(23):e2020GL091098.
12. Chen R, Liu L, Liu X. Leaf chlorophyll contents dominates the seasonal dynamics of SIF/GPP ratio: Evidence from continuous measurements in a maize field. *Agric For Meteorol*. 2022;323:Article 109070.
13. Yang JC, Magney TS, Albert LP, Richardson AD, Frankenberg C, Stutz J, Grossmann K, Burns SP, Seyednasrollah B, Blanken PD, et al. Gross primary production (GPP) and red solar induced fluorescence (SIF) respond differently to light and seasonal environmental conditions in a subalpine conifer forest. *Agric For Meteorol*. 2022;317:Article 108904.
14. Miao G, Guan K, Yang X, Bernacchi CJ, Berry JA, DeLucia EH, Wu J, Moore CE, Meacham K, Cai Y, et al. Sun-induced chlorophyll fluorescence, photosynthesis, and light use efficiency of a soybean field from seasonally continuous measurements. *J Geophys Res Biogeo*. 2018;123(2):610–623.
15. Van Der Tol C, Berry JA, Campbell PKE, Rascher U. Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence. *J Geophys Res Biogeo*. 2014;119(12):2312–2327.
16. Wu G, Jiang C, Kimm H, Wang S, Bernacchi C, Moore CE, Suyker A, Yang X, Magney T, Frankenberg C, et al. Difference in seasonal peak timing of soybean far-red SIF and GPP explained by canopy structure and chlorophyll content. *Remote Sens Environ*. 2022;279:113104.
17. Kim J, Ryu Y, Dechant B, Lee H, Seok H, Kornfeld A, Berry JA. Solar-induced chlorophyll fluorescence is non-linearly related to canopy photosynthesis in a temperate evergreen needleleaf forest during the fall transition. *Remote Sens Environ*. 2021;258:Article 112362.
18. Chen R, Liu X, Chen J, Du S, Liu L. Solar-induced chlorophyll fluorescence imperfectly tracks the temperature response of photosynthesis in winter wheat. *J Exp Bot*. 2022;73(22):7596–7610.
19. Chen A, Mao J, Ricciuto D, Xiao J, Frankenberg C, Li X, Thornton PE, Gu L, Knapp AK. Moisture availability mediates the relationship between terrestrial gross primary production and solar-induced chlorophyll fluorescence: Insights from global-scale variations. *Glob Chang Biol*. 2021;27(6):1144–1156.
20. Chen A, Mao J, Ricciuto D, Lu D, Xiao J, Li X, Thornton PE, Knapp AK. Seasonal changes in GPP/SIF ratios and their climatic determinants across the northern hemisphere. *Glob Chang Biol*. 2021;27(20):5186–5197.
21. Xu S, Atherton J, Riikonen A, Zhang C, Oivukkamäki J, MacArthur A, Honkavaara E, Hakala T, Koivumäki N, Liu Z, et al. Structural and photosynthetic dynamics mediate the response of SIF to water stress in a potato crop. *Remote Sens Environ*. 2021;263:Article 112555.
22. Zhang Z, Chen JM, Guanter L, He L, Zhang Y. From canopy-leaving to total canopy far-red fluorescence emission for remote sensing of photosynthesis: First results from TROPOMI. *Geophys Res Lett*. 2019;46(21):12030–12040.
23. Chang CY, Guanter L, Frankenberg C, Köhler P, Gu L, Magney TS, Grossmann K, Sun Y. Systematic assessment of retrieval methods for canopy far-red solar-induced chlorophyll fluorescence using high-frequency automated field spectroscopy. *J Geophys Res Biogeo*. 2020;125(7):e2019JG005533.
24. Gu L, Sun Y. Artefactual responses of mesophyll conductance to CO_2 and irradiance estimated with the variable J and online isotope discrimination methods. *Plant Cell Environ*. 2014;37(5):1231–1249.
25. Gu L, Han J, Wood JD, Chang CYY, Sun Y. Sun-induced Chl fluorescence and its importance for biophysical modeling of photosynthesis based on light reactions. *New Phytol*. 2019;223(3):1179–1191.
26. Guanter L, Bacour C, Schneider A, Aben I, Van Kempen TA, Maignan F, Retscher C, Köhler P, Frankenberg C, Joiner J, et al. The TROPOSIF global sun-induced fluorescence dataset from the sentinel-5P TROPOMI mission. *Earth Syst Sci Data*. 2021;13(11):5423–5440.
27. Liu S, Li X, Xu Z, Che T, Xiao Q, Ma M, Liu Q, Jin R, Guo J, Wang L, et al. The Heihe integrated observatory network: A basin-scale land surface processes observatory in China. *Vadose Zone J*. 2018;17(1):180072.
28. Li X, Cheng G, Liu S, Xiao Q, Ma M, Jin R, Che T, Liu Q, Wang W, Qi Y, et al. Heihe watershed allied telemetry experimental research (HiWATER): Scientific objectives and experimental design. *Bull Am Meteorol Soc*. 2013;94(8):1145–1160.
29. Keenan TF, Migliavacca M, Papale D, Baldocchi D, Reichstein M, Torn M, Wutzler T. Widespread inhibition of daytime ecosystem respiration. *Nat Ecol Evol*. 2019;3(3):407–415.
30. Wutzler T, Lucas-Moffat A, Migliavacca M, Knauer J, Sickel K, Šigut L, Menzer O, Reichstein M. Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*. 2018;16:5015–5030.
31. He Q, Ju W, Dai S, He W, Song L, Wang S, Li X, Mao G. Drought risk of global terrestrial gross primary productivity over the last 40 years detected by a remote sensing-driven process model. *J Geophys Res Biogeosci*. 2021;126(6):e2020JG005944.
32. Joiner J, Yoshida Y. Global MODIS and FLUXNET-derived Daily Gross Primary Production, V2. ORNL Distributed Active Archive Center; 2021.
33. Zhang Y, Xiao X, Zhang Y, Wolf S, Zhou S, Joiner J, Guanter L, Verma M, Sun Y, Yang X, et al. On the relationship between

sub-daily instantaneous and daily total gross primary production: Implications for interpreting satellite-based SIF retrievals. *Remote Sens Environ*. 2018;205:276–289.
34. Frankenberg C. Solar Induced Chlorophyll Fluorescence: OCO-2 Lite Files (B7000) User Guide. Pasadena (CA): California Institute of Technology; 2015.
35. Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, et al. The ERA5 global reanalysis. *Q J R Meteorol Soc*. 2020;146(730):1999–2049.
36. Ryu Y, Jiang C, Kobayashi H, Detto M. MODIS-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000. *Remote Sens Environ*. 2018;204:812–825.
37. Seager R, Hooks A, Williams AP, Cook B, Nakamura J, Henderson N. Climatology, variability, and trends in the U.S. vapor pressure deficit, an important fire-related meteorological quantity. *J Appl Meteorol Climatol*. 2015;54(6):1121–1141.
38. Barkhordarian A, Saatchi SS, Behrangi A, Loikith PC, Mechoso CR. A recent systematic increase in vapor pressure deficit over tropical South America. *Sci Rep*. 2019;9(1):15331.
39. Liu L, Zhang X. Dynamic mapping of broadband visible albedo of soil background at global 500-m scale from MODIS satellite products. In: *Land Surface and Cryosphere Remote Sensing IV*. SPIE; 2018. p. 31–34.
40. Schaaf C. MCD43C4 MODIS/Terra+ Aqua BRDF-Adjusted Nadir Reflectance Daily L3 Global 0.05 Deg CMG V006. Sioux Falls (SD): NASA EOSDIS Land Processes DAAC; 2015.
41. Liu X, Liu L, Bacour C, Guanter L, Chen J, Ma Y, Chen R, Du S. A simple approach to enhance the TROPOMI solar-induced chlorophyll fluorescence product by combining with canopy reflected radiation at near-infrared band. *Remote Sens Environ*. 2023;284:113341.
42. Zeng Y, Badgley G, Dechant B, Ryu Y, Chen M, Berry JA. A practical approach for estimating the escape ratio of near-infrared solar-induced chlorophyll fluorescence. *Remote Sens Environ*. 2019;232:111209.
43. Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci Data*. 2018;5: Article 180214.
44. Weir, B., Ott L and O-2 ST. OCO-2 GEOS Level 3 monthly, 0.5x0.625 assimilated CO2 V10r. Greenbelt (MD): Goddard Earth Sciences Data and Information Services Center (GES DISC); 2022.
45. Jiao Z, Dong Y, Schaaf CB, Chen JM, Román M, Wang Z, Zhang H, Ding A, Erb A, Hill MJ, et al. An algorithm for the retrieval of the clumping index (CI) from the MODIS BRDF product using an adjusted version of the kernel-driven BRDF model. *Remote Sens Environ*. 2018;209:594–611.
46. Kumar R, Umanand L. Estimation of global radiation using clearness index model for sizing photovoltaic system. *Renew Energy*. 2005;30(15):2221–2233.
47. Liu Z, Zhao F, Liu X, Yu Q, Wang Y, Peng X, Cai H, Lu X. Direct estimation of photosynthetic CO_2 assimilation from solar-induced chlorophyll fluorescence (SIF). *Remote Sens Environ*. 2022;271:Article 112893.
48. Lu X, Liu Z, Zhao F, Tang J. Comparison of total emitted solar-induced chlorophyll fluorescence (SIF) and top-of-canopy (TOC) SIF in estimating photosynthesis. *Remote Sens Environ*. 2020;251:112083.
49. Liu X, Liu L, Hu J, Guo J, Du S. Improving the potential of red SIF for estimating GPP by downscaling from the canopy level to the photosystem level. *Agric For Meteorol*. 2020;281:Article 107846.
50. Liu X, Liu Z, Liu L, Lu X, Chen J, Du S, Zou C. Modelling the influence of incident radiation on the SIF-based GPP estimation for maize. *Agric For Meteorol*. 2021;307:Article 108522.
51. Harrison SP, Cramer W, Franklin O, Prentice IC, Wang H, Brännström Å, de Boer H, Dieckmann U, Joshi J, Keenan TF, et al. Eco-evolutionary optimality as a means to improve vegetation and land-surface models. *New Phytol*. 2021;231(6):2125–2141.
52. Wang H, Prentice IC, Keenan TF, Davis TW, Wright IJ, Cornwell WK, Evans BJ, Peng C. Towards a universal model for carbon dioxide uptake by plants. *Nat Plants*. 2017;3(9): 734–741.
53. Stocker BD, Wang H, Smith NG, Harrison SP, Keenan TF, Sandoval D, Davis T, Prentice IC. P-model v1.0: An optimality-based light use efficiency model for simulating ecosystem gross primary production. *Geosci Model Dev*. 2020;13(3):1545–1581.
54. Duveiller G, Cescatti A. Spatially downscaling sun-induced chlorophyll fluorescence leads to an improved temporal correlation with gross primary productivity. *Remote Sens Environ*. 2016;182:72–89.
55. Porcar-Castell A, Malenovský Z, Magney T, Van Wittenberghe S, Fernández-Marín B, Maignan F, Zhang Y, Maseyk K, Atherton J, Albert LP, et al. Chlorophyll a fluorescence illuminates a path connecting plant molecular biology to earth-system science. *Nat Plants*. 2021;7(8):998–1009.
56. Bacour C, Maignan F, MacBean N, Porcar-Castell A, Flexas J, Frankenberg C, Peylin P, Chevallier F, Vuichard N, Bastrikov V. Improving estimates of gross primary productivity by assimilating solar-induced fluorescence satellite retrievals in a terrestrial biosphere model using a process-based SIF model. *J Geophys Res Biogeo*. 2019;124(11):3281–3306.
57. He L, Chen JM, Liu J, Zheng T, Wang R, Joiner J, Chou S, Chen B, Liu Y, Liu R, et al. Diverse photosynthetic capacity of global ecosystems mapped by satellite chlorophyll fluorescence measurements. *Remote Sens Environ*. 2019;232:Article 111344.
58. Chen JM, Liu J, Cihlar J, Goulden ML. Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications. *Ecol Model*. 1999;124(2–3): 99–119.