Global analysis of the relationship between reconstructed solar induced chlorophyll fluorescence (SIF) and gross primary production (GPP)

Gao, H.; Liu, Shuguang; Lu, W.; Smith, Andy; Valbuena, Ruben; Yan, Wende; Wang, Zhao; Xiao, Li; Peng, Xi; Li, Qinyuan; Feng, Yujun; McDonald, Morag; Pagella, Tim; Liao, Juyang; Wu, Zhenming; Zhang, Gui

Remote Sensing

Accepted/In press: 27/06/2021

Publisher's PDF, also known as Version of record *remote sensing*

MDPI

Article

Global analysis of the relationship between reconstructed solar induced chlorophyll fluorescence (SIF) and gross primary production (GPP)

Haiqiang Gao [1,2], Shuguang Liu [1,2,*], Weizhi Lu [1,2], Andrew R. Smith [3], Rubén Valbuena [3], Wende Yan [1,2], Zhao Wang [1,2], Li Xiao [1,2], Xi Peng [1,2], Qinyuan Li [1,2], Yujun Feng [4,5], Morag McDonald [3], Tim Pagella [3], Juyang Liao [6], Zhenming Wu [7], Gui Zhang [8]

1 College of Life Science and Technology, Central South University of Forestry and Technology, Changsha 410004, China; 20200100045@csuft.edu.cn (H.G.); weizhilu@csuft.edu.cn (W.L.); t20001421@csuft.edu.cn (W.Y.); 20171200061@csuft.edu.cn (Z.W.); lixiao@csuft.edu.cn (L.X.); 20180100029@csuft.edu.cn (X.P.); 20181100139@csuft.edu.cn (Q.L.)
2 National Engineering Laboratory for Applied Technology of Forestry & Ecology in South China, Central South University of Forestry and Technology, Changsha 410004, China
3 School of Natural Sciences, Bangor University, Gwynedd, LL57 2UW, UK; a.r.smith@bangor.ac.uk (A.S.); r.valbuena@bangor.ac.uk (R.V.); m.mcdonald@bangor.ac.uk (M.M.); t.pagella@bangor.ac.uk (T.P.)
4 Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, 510650, China; fengyujun20@mails.ucas.ac.cn (Y.F.)
5 University of Chinese Academy of Sciences, Beijing 100049, China
6 Hunan Forest Botanical Garden, Changsha 410116, China; liaojuyang@163.com (J.L.)
7 Hunan Academy of Forestry, Changsha 410007, China; wzm417@163.com (Z.W.)
8 College of Science, Central South University of Forestry and Technology, Changsha 410004, China; zhanggui@csuft.edu.cn (G.Z.)

* Correspondence: Shuguang.liu@csuft.edu.cn

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Remote Sens.* **2021**, *13*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

Abstract: Solar-induced chlorophyll fluorescence (SIF) is increasingly known as an effective proxy for plant photosynthesis and therefore has great potential in monitoring gross primary production (GPP). However, the relationship between SIF and GPP remains highly uncertain across space and time. Here, we analyzed the SIF (reconstructed, SIFc)-GPP relationships and their spatiotemporal variability, using GPP estimates from FLUXNET2015 and two spatiotemporally contiguous SIFc datasets (CSIF and GOSIF). Results showed that SIFc had significant positive correlations with GPP at the spatiotemporal scales investigated ($p<0.001$). The generally linear SIFc-GPP relationships were substantially affected by spatial and temporal scales and SIFc datasets. GPP/SIFc slope of the evergreen needleleaf forest (ENF) biome was significantly higher than those of several other biomes ($p<0.05$), while the other 11 biomes showed no significant differences in GPP/SIFc slope between each other ($p>0.05$). We therefor propose a two-slope scheme to differentiate ENF from non-ENF biome and synopsize spatiotemporal variability of GPP/SIFc slope. The relative biases were 7.14% and 11.06% in the estimated cumulative GPP across all EC towers, respectively, for GOSIF and CSIF using two-slope scheme. The significantly higher GPP/SIFc slopes of the ENF biome in the two-slope scheme are intriguing and deserve further study. In addition, there was still considerable dispersion in the comparisons of CSIF/GOSIF and GPP at both site and biome levels, calling for discriminatory analysis backed by higher spatial resolution to systematically address issues related to landscape heterogeneity and mismatch between SIFc pixel and the footprints of flux towers and their impacts on the SIF and GPP relationship.

Keywords: SIF-GPP conversion coefficient, eddy covariance flux towers, land cover type, GOSIF, CSIF, evergreen needleleaf forest

1. Introduction

GPP is the largest flux in the global carbon cycle [1], yet accurate estimation of GPP at regional and global scales is still a major challenge [2]. Solar-induced chlorophyll fluorescence (SIF) has recently emerged as process that can be detected using Earth observation technologies, thus having potential to radically improve terrestrial GPP estimation [3, 4]. SIF is the energy emitted directly from the core of photosynthetic machinery during the return photosystem II from excited to non-excited states nanoseconds after light absorption with the wavelength range from 600 to 800 nm [5, 6]. Light energy absorbed by the leaf chlorophyll molecules has three different pathways: photochemistry, non-photochemical quenching (NPQ, i.e., heat dissipation), and a small fraction re-emitted as SIF [6]. SIF is highly correlated with photosynthesis when NPQ dominates at high light levels [6], and it shows stronger capability in general in characterizing the temporal and spatial dynamics of photosynthesis or gross primary productivity in terrestrial ecosystems than traditional vegetation indices (e.g. NDVI and EVI) [7] as it is directly related to actual photosynthetic rate [8].

Constructing a direct relationship between satellite-derived SIF and eddy covariance (EC) flux tower based GPP is crucial for using SIF to estimate GPP at large scales [2], but has been hindered by the spatial and temporal coverage of SIF datasets [9]. Current SIF products are derived from Greenhouse Gases Observing Satellite (GOSAT) [10], SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) [11], Global Ozone Monitoring Instrument (GOME) [12] and Global Ozone Monitoring Mission Experiment-2 (GOME-2) [13], Orbiting Carbon Observatory-2 (OCO-2) [14], TanSat [15] and TROPOspheric Monitoring Instrument (TROPOMI) [16]. Among these products, SIF retrieved from OCO-2 showed the smallest footprints (1.30 × 2.25 km) and slightly higher signal-to-noise ratios than others, and provided new opportunities to directly link satellite-derived SIF to flux tower GPP at the ecosystem scale [17]. Many studies have reported the relationship between SIF derived from various satellite missions with GPP derived from EC flux tower [18] and gridded Moderate-resolution Imaging Spectroradiometer (MODIS) products [19] at different spatiotemporal scales.

The relationship between SIF measurements obtained with remote passive techniques (i.e., remote sensing SIF signal (OCO-2 SIF)) and photosynthesis (i.e., GPP) is not well understood [20] due to large uncertainties when establishing the relationship between SIF and EC flux tower GPP across different ecosystems [21]. Wood, Griffis [22] found a linear SIF-GPP relationship that is sensitive to crop type (corn *vs.* soybean) and invariant across spatiotemporal scales in the Corn Belt. This study only investigated two types of crops in a small part of the United States, therefore it is not a systematic study of the SIF-GPP relationship and more studies should be conducted regarding of the differences of C3 and C4 crops [21]. It was found that the strength of this linear relationship in temperate forest was scale-dependent, and its linearity was stronger at the midday time scale [23]. Similar results have been found across several vegetated biomes, especially for OCO-2 SIF at 757nm [21, 24]. Li, Xiao [21] reported a nearly universal linear SIF-GPP relationship between OCO-2 SIF and EC-GPP from a total of 64 sites across eight major biomes. Recently, Wang, Chen [25] improved the SIF-GPP relationship using photochemical reflectance index. However, some studies based on GOSAT and GOME-2 analysis indicated that the SIF-GPP relationship varied across biomes [19]. Indeed, Sun, Frankenberg [26] found the linear SIF-GPP relationship diverges somewhat across 10 biomes at the global scale. The main reasons for the uncertainty in the SIF-GPP relationship across sites and biomes are spatiotemporal mismatches and data uncertainties among the SIF and GPP products, which can be traced back into at least three major issues. First, the spatial mismatch of EC flux tower sites and OCO-2 orbit is the general limitations of satellite SIF application [5]. Second, the temporal inconsistent between the short lifetime of OCO-2 SIF (available from 6th September 2014 to present) and GPP estimated from EC flux towers (i.e., FLUXNET data is only updated to 2015 (FLUXNET2015)) is not relevant for the development or validation of the SIF-GPP relationship [2]. Third, uncertainties in estimating

GPP from EC towers [27] and SIF sampling instrument and retrieval methodologies [21]. Thus, the amount (spatial and temporal coverage) of data available from the satellite SIF at present is insufficient to support comprehensive analysis the SIF-GPP relationship [28]. Therefore, more studies tackling with these issues are required to truly address the complexities and drivers of variability in the SIF-GPP relationships across biomes.

Several global spatially contiguous SIF datasets (hereafter referred to as SIFc) developed recently can contribute to address the above issues. Zhang, Joiner [29] generated global spatially contiguous SIF dataset (hereafter referred to as CSIF, i.e., clear-sky instantaneous and all-sky daily average) at moderate spatiotemporal resolutions (0.05° and 4-day) by training a neural network with surface reflectance from MODIS and OCO-2 SIF soundings. Yu, Wen [30] developed another spatially contiguous global SIF product (hereafter referred to as GCSIF) at 0.05° and 16-day resolutions using machine learning with time-and-biome-specific model. Li and Xiao [31] further developed a global OCO-2 SIF dataset (GOSIF) with a similar spatiotemporal resolutions (0.05° and 8-day) based on discrete OCO-2 SIF soundings, EVI and land cover type data from MODIS, and meteorological reanalysis data from Modern-Era Retrospective analysis for Research and Applications (MERRA-2) [32]. And Duveiller, Filipponi [33] presented a new SIF dataset (hereafter referred to as GOMESIF) based on GOME-2 satellite observations with an enhanced spatial resolution covering the period 2007–2018. In general, differences exist among SIFc products due to different reconstruction methods in this study (**see Supplementary Material**), and there is an urgent need to recognize and, if possible, reconcile the differences of SIFc datasets and understand their potential impacts on GPP estimation.

To improve the quantification of terrestrial photosynthesis at various spatial and temporal scales using the recently available remotely sensed spatially contiguous SIFc datasets, further efforts should focus on the application of expanded SIFc datasets to test the robustness of the SIF-GPP relationship across all vegetated biomes [2]. Here, we use two global spatially contiguous SIFc datasets (CSIF [29] and GOSIF [31]), coupled with GPP obtained by EC flux tower from the worldwide network FLUXNET2015 [34], to address the following objectives: (1) to explore the commonality and differences of the SIFc–GPP relationship across 12 vegetated IGBP biomes; (2) to examine the variability of SIFc-GPP relationships over a range of spatial and temporal scales; (3) to elucidate the application prospects and limitations of existing spatially contiguous SIFc datasets.

2. Materials and Methods

2.1. Datasets

Two available spatially contiguous SIFc datasets (unit in W m^{-2} μm^{-1} sr^{-1}) based on OCO-2 SIF (V8r) at 757nm were used in this study. First, the CSIF dataset, generated by Zhang, Joiner [29], has two global spatially contiguous SIFc data layers at moderate spatiotemporal resolutions (0.05° spatial resolution, and 4-day temporal resolution, obtained upon request from the author Zhang Yao): one from instantaneous measurements obtained on clear-sky conditions (2000–2017) and the other from daily averages including all-sky conditions (2000–2016) (referred to as CSIFall-daily). They are generated based on the SIF retrievals from OCO-2, interpolated by artificial neural networks (ANN) to a grid using the surface reflectance from MODIS aboard the Terra and Aqua satellites [29]. The ANN with one layer and five neurons exhibited the highest model performance with a good performance in validation (R^2 = 0.79, RMSE = 0.18 W m^{-2} μm^{-1} sr^{-1}). The errors of CSIF in 9 of 14 biomes to OCO-2 SIF were less than 10%, and most of them were lower than 5% [29]. To better match with the GPP data, the all-sky daily average CSIF dataset (CSIFall-daily) was used (referred to as CSIF), which exhibited strong spatial, seasonal, and interannual dynamics that were consistent with daily SIF from OCO-2 and GOME-2 [29].

Second, we employed the global 'OCO-2' SIF dataset (referred to as GOSIF) (0.05° spatial resolution, 8-day temporal resolution, freely available at http://globalecology.unh.edu) [31]. The dataset was based on a data-driven model developed based on discrete OCO-2 SIF data, EVI and land cover data from MODIS, and meteorological reanalysis data. Similar to CSIF, the GOSIF dataset has extended the start date of data record of OCO-2 SIF to March 2000 and at daily time scale. The dataset also performed fairly well in SIF validation ($R^2 = 0.79$, RMSE = 0.07 W m^{-2} μm^{-1} sr^{-1}). These two reconstructed SIF products (i.e., CSIF and GOSIF) offer opportunities to examine the synergy between satellite SIF and photosynthesis at consistent spatial scales globally [29, 31, 35].

GPP data was extracted from the global network FLUXNET2015 (http://fluxnet.fluxdata.org//data/fluxnet2015-dataset/), which contains terrestrial ecosystem carbon flux data from 212 EC flux towers worldwide [34]. Considering small differences between different GPP partitioning methods [36] (Table A1 and Figs A1 and A2), daily average GPP estimates (GPP_M, unit in g C m^{-2} d^{-1}) were calculated as the mean of GPP estimates from both daytime respiration (GPP_D) and nighttime respiration (GPP_N) [37] and used to analyze the SIFc-GPP relationship globally. Four sites (i.e., IT-SRo, NO-Blv, US-LWW and US-Me4 sites) were removed due to the limited data and large landscape heterogeneity in the SIFc pixel, after visually examining the landscape composition of all flux tower footprints and associated SIFc pixels using Google Earth images. Consequently, 208 EC flux tower sites distributed across 12 vegetated biomes were used, which was different from some previous researches [21, 29] (Fig. 1, Table A2). In addition, all the daily data used for analysis were extracted at 8-day time interval (i.e., CSIF from 4-day to 8-day, GOSIF 8-day and GPP from 1-day to 8-day).

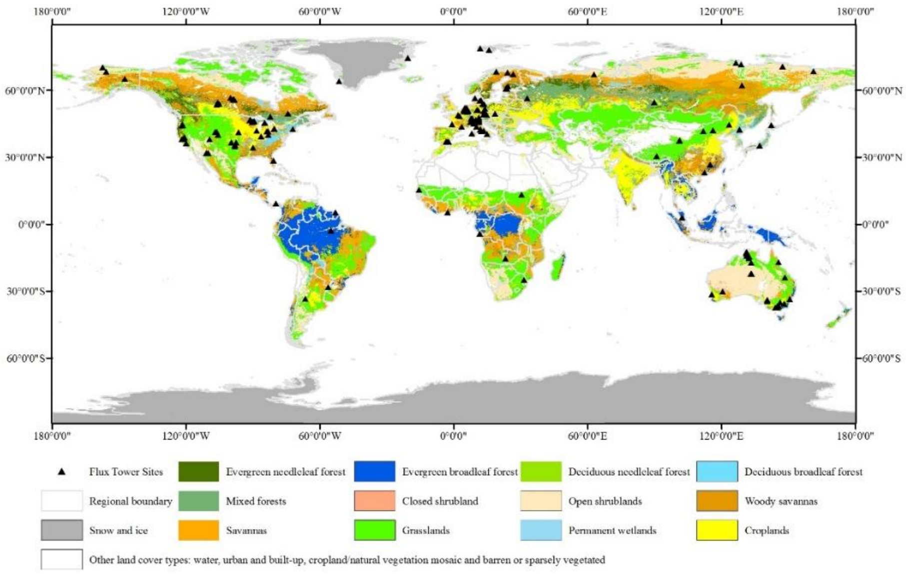

Figure 1. The spatial distribution of all the 212 eddy-covariance (EC) flux tower sites from the FLUXNET 2015 Tier 1 dataset, depicted by black triangles. Biomes in the legend are from a static land cover map (MCD12C1 Land Cover Type 1: IGBP global vegetation classification scheme for 2007) referred to Friedl, McIver [38].

2.2. *SIFc-GPP relationship analysis*

All analyses were performed using programming environments in R language version 3.6.1 [39]. All significance tests were performed with an alpha of 0.05 by default unless specified otherwise. All mean values presented in the paper were accompanied by corresponding Standard Error (SE) values unless otherwise stated. This study covered 12 vegetated biomes according to IGBP [40, 41] classification: croplands (CRO), closed shrublands (CSH), deciduous broadleaf forests (DBF), deciduous needleleaf forests (DNF), evergreen broadleaf forests (EBF), evergreen needleleaf forests (ENF), grasslands (GRA), mixed Forests (MF), open shrublands (OSH), savannas (SAV), permanent wetlands (WET), and woody savannas (WSA) (see Table A3 for more details).

First, the correlation and differences between the SIFc and GPP dataset have been analyzed. Specifically, the differences between two SIFc datasets (CSIF and GOSIF) and the differences between two GPP datasets (GPP_D and GPP_N) were tested based on daily (8-day temporal resolution based on GPP data; all the daily GPP/SIF data were extracted at 8-day interval) data using confidence interval (CI) approach. Second, the relationships between two reconstructed SIFc products and GPP_M were investigated across six combinations of temporal scales (i.e., daily: mean of half-hour GPP data for each day, yearly: mean of daily GPP and SIF for each year, and multi-year: mean of the whole observation period) and spatial scales (i.e., site and biome), using major axis regression [42] to account for data uncertainties in both x and y in the analysis of SIFc-GPP relationship. In the analysis of the SIFc-GPP relationship, we forced trend lines to pass through the origin by setting intercept to zero based on the logic that zero SIF would suggest zero photosynthesis or GPP approximately [2]. Whether significant differences existed among biomes in the SIFc-GPP conversion coefficients at site-year and site-multi-year scales were evaluated using *wilcox.test()* in *ggsignif* package. Third, the SIFc-GPP relationships at six spatial (site and biome) and temporal (daily, yearly and multi-year) scales were analyzed to examine the change of the SIFc-GPP relationship with scales.

Abovementioned analyses led to the conclusion that it is necessary to synopsize the inter-biome variability of GPP/SIFc slopes using a two-slope scheme. To develop the two-slope scheme, we first reclassified all sites into ENF and Non-ENF biomes, and then analyzed and compared the site-scale GPP/SIFc slopes within the ENF and Non-ENF groups. The mean and standard error (SE) were calculated from site-scale GPP/SIFc slopes within the ENF and Non-ENF biomes, respectively, to represent the two GPP/SIFc slopes and their uncertainty of the two-slope scheme. Similarly, the adequacy of using median and median absolute deviation (MAD) of site-scale GPP/SIFc slopes within the ENF and Non-ENF biomes to represent the two-slope scheme was also investigated. The two-sided Students t-test, the *t.test(two.sided)* function in R, was applied to test the difference between ENF and Non-ENF groups. The performance of two-slope scheme was measured with correlation coefficient (r), standard deviation (SD), root mean square error (RMSE), and percentage bias (PB) between flux GPP and SIFc_GPP.

3. Results

3.1. Correlation between SIFc and GPP

SIFc (both CSIF and GOSIF) showed significant positive correlations with GPP (GPP_D, GPP_N, and GPP_M) worldwide across all the 12 biome types and available years (from 2001 to 2014) (Fig. 2). Among biomes, the highest GPP-SIFc correlation was manifested in DBF, and the lowest was in EBF (Fig. 2A). The correlations between GOSIF and GPP (i.e., GPP_D, GPP_N, and GPP_M) were higher than those from CSIF in general. However, the r values for GOSIF-GPP were lower than those for CSIF for OSH and SAV biomes (Fig. 2A). Strong positive correlations were also observed between SIFc and GPP across all 14 years (r > 0.71, $p < 0.001$; Fig. 2B). Among all the SIFc-GPP correlation coefficients, those of SIFc-GPP_M were the highest: concentrated at 0.76 ± 0.01 and 0.77 ± 0.00 for CSIF and GOSIF, respectively, across 14 years (Fig. 2B).

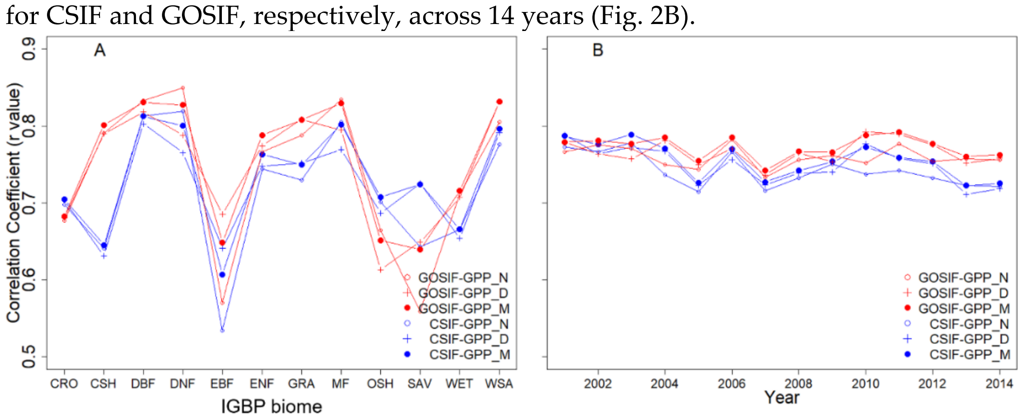

Figure 2. Pearson correlation coefficients (r value) between SIFc (CSIF and GOSIF) and GPP from two partitioning methods (GPP_D and GPP_N) and the mean of them (GPP_M) across (A) all 12 biomes (daily data at specific biome of all the years) and (B) all 14 years (daily data from whole year) ($p < 0.001$). Biomes are: croplands (CRO), closed shrublands (CSH), deciduous broadleaf forest (DBF), deciduous needleleaf forest (DNF), evergreen broadleaf forest (EBF), evergreen needleleaf forest (ENF), grasslands (GRA), mixed forests (MF), open shrublands (OSH), savannas (SAV), permanent wetlands (WET) and woody savannas (WSA). Years are from 2001 to 2014.

3.2. SIFc-GPP relationship across sites, biomes, and years

Fig. 3 showed the distributions of GPP/SIFc slopes in individual biomes at site-multi-year (A and B) and site-year (C and D) levels. GPP/SIFc slopes varied greatly across sites and biomes. The CSIF-GPP relationships at the site-multi-year scale (Fig. 3A) indicated that CSH had the largest inter-site variability with the biggest interquartile ranges (the height of the boxes). In addition, there were significant differences in GPP/CSIF slopes between ENF and several other biomes (i.e., DBF, EBF, GRA and OSH) ($p < 0.001$), and no significant difference was found among all other biome pairs ($p > 0.05$). Although the GOSIF-GPP relationships at the site-multi-year scale (Fig. 3B) look similar to the GOSIF-GPP relationships (Fig. 3A), there were substantial differences. First, the GOSIF-GPP inter-site variabilities were smaller than those of GOSIF-GPP in most biomes. Second, with less inter-site variability, the GOSIF-GPP data showed that the number of biomes significantly different from ENF was one more than the CSIF-GPP data (ENF *vs.* MF) and the significance level (*p* value) generally increased as well. In addition to these differences, it is important to notice that there was still no significant difference between any non-ENF biome pairs according to GOSIF-GPP ($p > 0.05$), consistent with CSIF-GPP. The SIFc-GPP relationships at the site-year scale (Fig. 3C and D), as expected, showed larger variability than those at the site-multi-year scale. ENF biome showed significant differences with all other biomes ($p < 0.001$), except for CSH biome with GPP/CSIF slopes ($p > 0.05$) (Fig. 3C and D). However, there were no significant differences of SIFc-GPP relationship between OSH and DNF biome with other biomes at site-multi-year scale ($p > 0.05$), although the slopes from OSH and DNF are lower than others (Figs. 3A, B and 6).

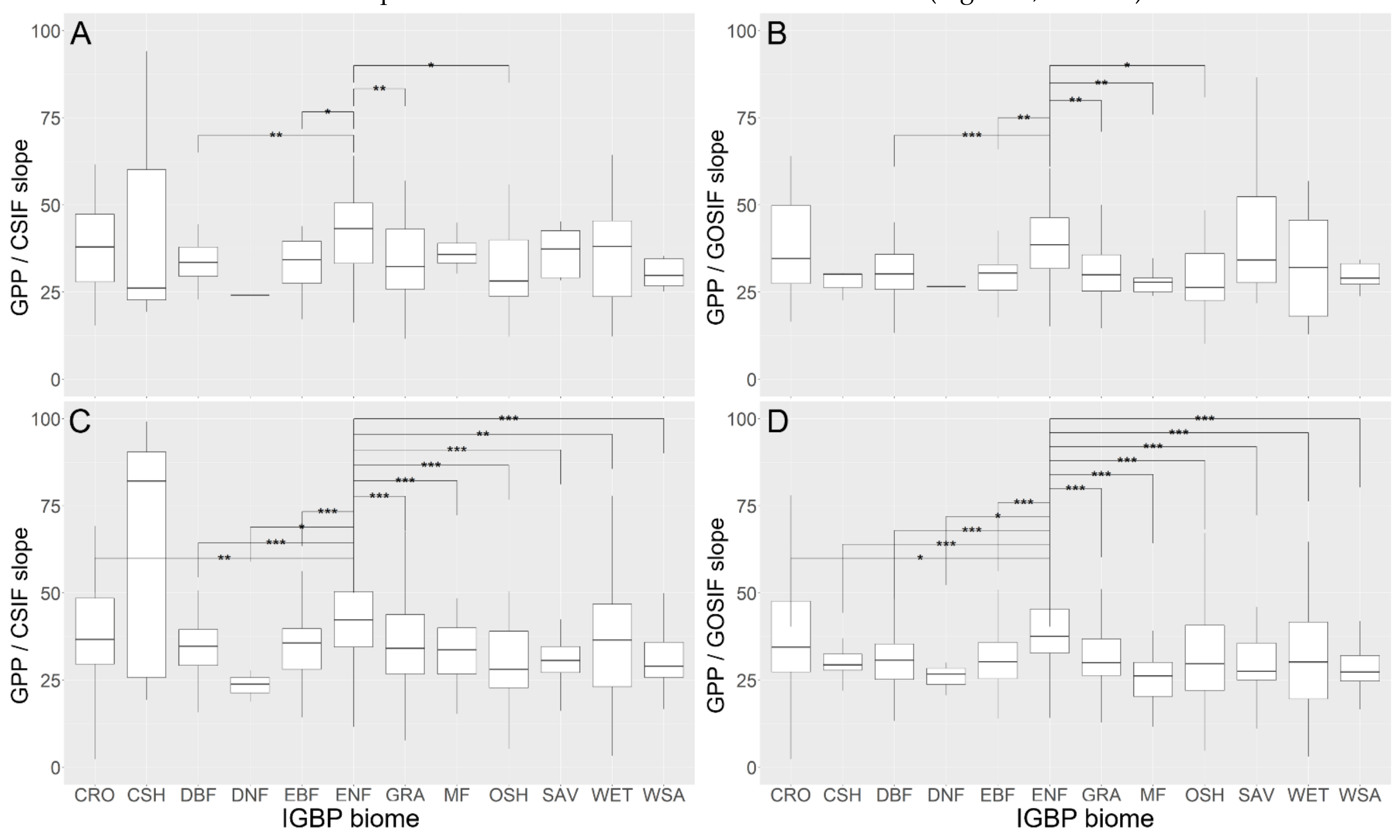

Figure 3. Boxplots and comparison of GPP/SIFc slopes between 12 biomes that similar with figure 1. (A, B) site-multi-year slopes (basis of slopes calculated for whole time); (C, D) site-year slopes (basis of slopes calculated for each year). Each boxplot represents the distribution of GPP/SIFc slopes in corresponding biome. The top and bottom of the boxes represent 75 and 25 percentiles (i.e., Q3 and Q1), respectively; the solid line in the box is median value of the box; the whole box is the interquartile range (IQR = Q3 – Q1); the top and bottom whiskers represent the maximum and minimum values (i.e., Q3 – 1.5 * IQR, Q1 – 1.5 * IQR), respectively; the data outside of the maximum and minimum are shown as points beyond the whiskers. Site-year and site-multi-year slopes were derived from daily SIFc-GPP data obtained in a year and in the whole observation period for each site, respectively, using major axis regression. Black star points indicate that the difference of mean GPP/SIFc slopes between two connected biomes is significant (***: $p < 0.001$; **: $0.001 < p < 0.01$; *: $0.01 < p < 0.05$).

Temporal variability of site-level GPP/CSIF slopes remained relative stable for most biomes except for a few biomes with very limited number of flux towers (i.e., CSH, OSH, SAV, and WET) (Fig. 4). Medians of slopes were more similar than means in different variants. The interannual variability of forest biomes were in general the smallest, followed by grassland and cropland. It is interesting to see that the interannual variability of forest, grassland and cropland biomes remained relatively stable, not affected by the increase of number of flux towers over time in general. In contrast, other biomes showed different as the number of sites were small and the number of towers in normal operation fluctuated across years, which led to large interannual variability in the slopes within each of these biomes.

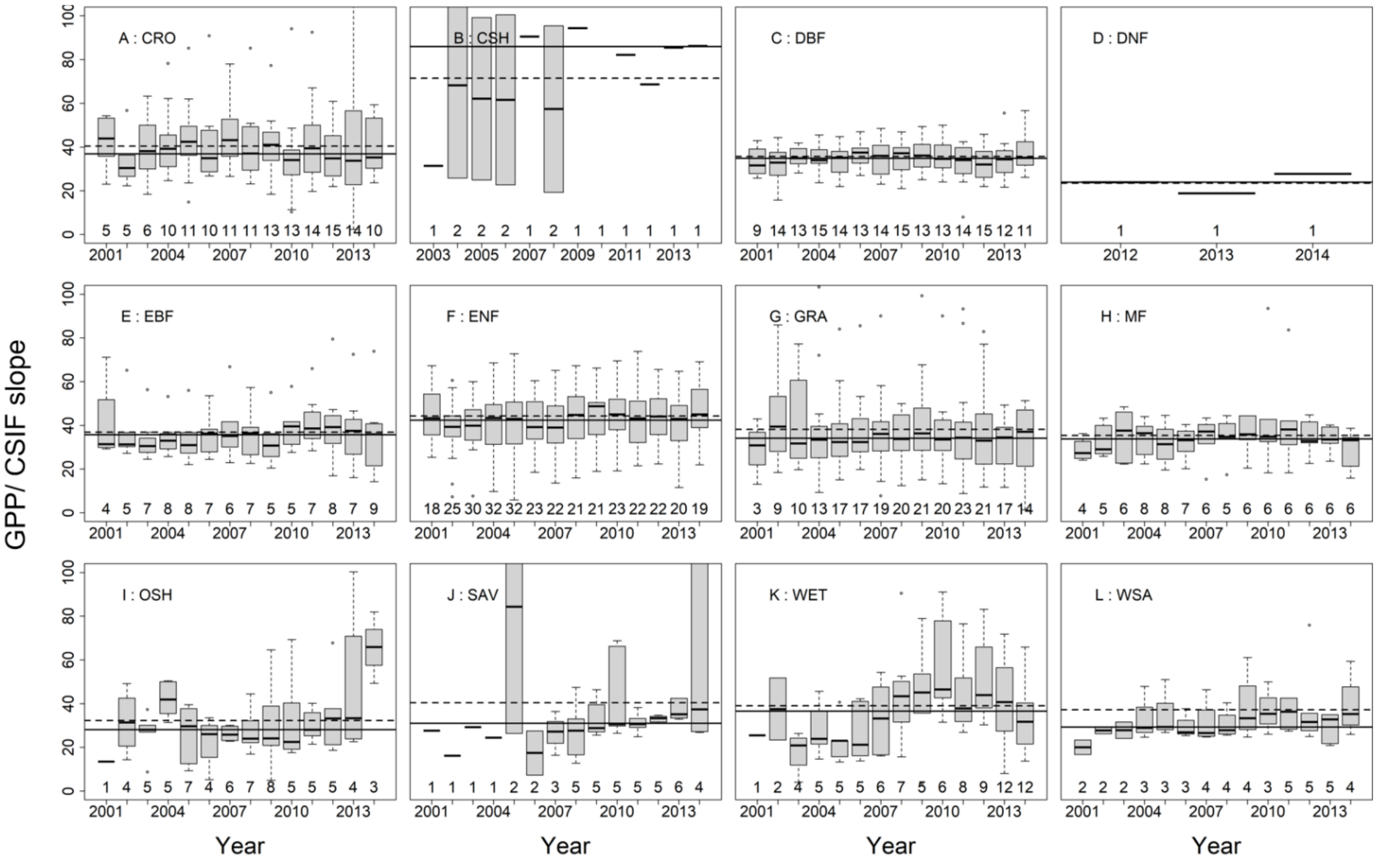

Figure 4. Interannual variability of the slopes between CSIF and GPP across 12 biomes. 12 biomes are similar with Fig. 1. The explanation of the boxplot symbols is given in Fig. 3. The solid line and dash line across each boxplot are median value and mean value for the whole boxplot (biome), respectively. The number of flux tower sites for each year is shown on top of the x axis.

Compared with GPP/CSIF slopes, the temporal variabilities of GPP/GOSIF slopes were smaller for most biomes (Fig. 5). The reduction of variability was most in those biomes that showed large interannual variabilities in GPP/CSIF slopes (i.e., CSH, OSH and

SAV). The site-variability of wetland (WET) biome expanded greatly from 2009 to 2011 compared with surrounding years and those of GPP/CSIF, and the variability of the grassland (GRA) biome also increased in 2002 and 2003. The enlarged variabilities were probably caused by underestimated GOSIF at a few flux sites in these two biomes in the given years as the median slope was higher than the median and lower than the mean from all years (Fig. 5).

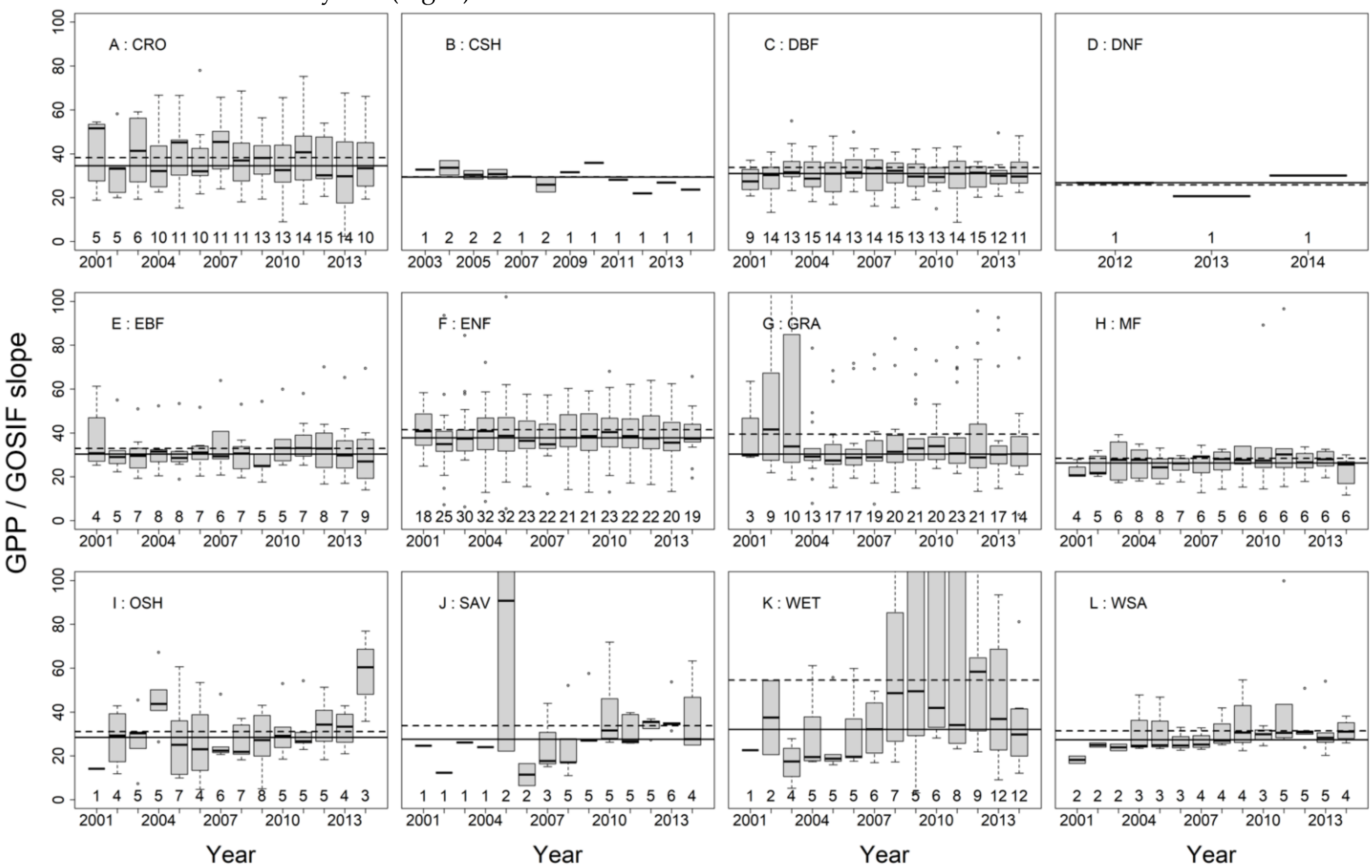

Figure 5. Interannual variability of the slopes between GOSIF and GPP across 12 biomes. The explanation of the boxplot symbols is given in Fig. 4.

Averaged all SIFc and GPP together by biome and ignoring the inter-site differences, strong linear relationships between GPP and SIFc were found consistently across 12 biomes (Fig. 6) and 14 years (Fig. 7). Although SIFc-GPP relationship varied among sites (see Fig. 3), the SIFc-GPP relationships at the biome scale were strongly linear ($p < 0.001$) for both CSIF and GOSIF. However, the large dispersion of the data points also suggests the large temporal (across years) and spatial (across sites) variability (see Figs. 3, 4 and 5). For example, the diverging relationship found in CSH was caused by the low CSIF values at IT.Noe site (Figs. A3 and 6).

The linear SIFc-GPP relationships were SIFc dataset dependent (Figs. 6 and 7). The GPP/CSIF slopes were generally higher than GPP/GOSIF slopes for all biomes except DNF (Fig. 6). Similar differences existed between GPP/CSIF slopes and GPP/GOSIF slopes for all 14 years (Fig. 7). The GPP/CSIF slopes ranged from 34.86 ($R^2 = 0.82$, $p < 0.001$) to 39.29 ($R^2 = 0.79$, $p < 0.001$) across 14 years with a mean value of 37.64 ± 0.32, and the value of R^2 ranged from 0.77 to 0.83. In contrast, the GPP/GOSIF slopes ranged from 30.65 ($R^2 = 0.81$, $p < 0.001$) to 35.19 ($R^2 = 0.82$, $p < 0.001$) with a mean value of 33.14 ± 0.32, and the value of R^2 ranged from 0.78 to 0.83. It should be noticed that there were significant differences between CSIF and GOSIF products across all 12 biomes except CSH (Figs. A4, A7) as well as across all 14 years (Fig. A5).

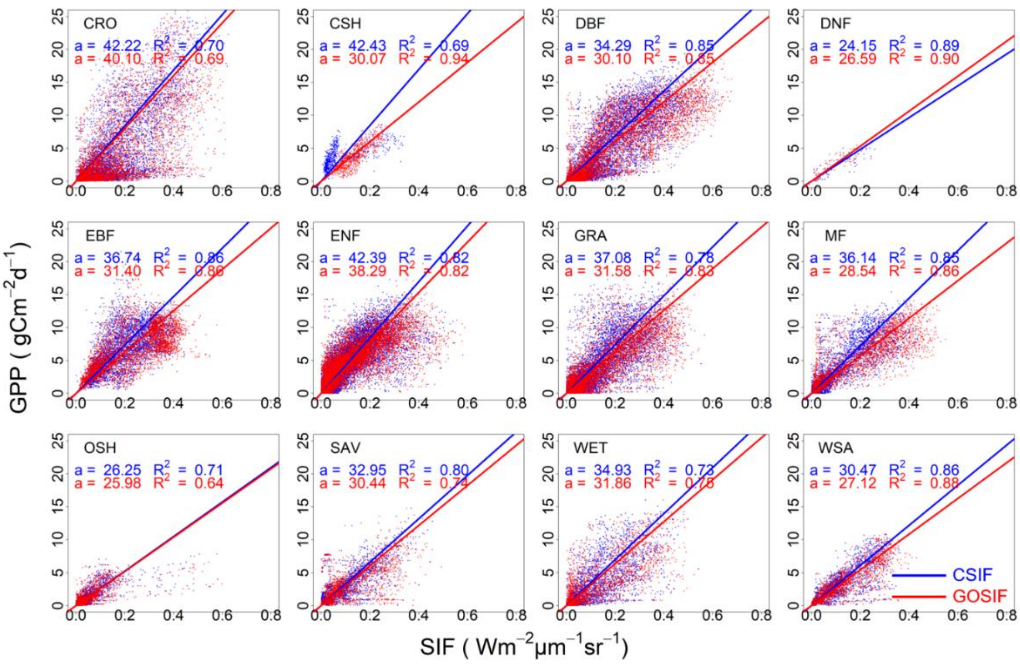

Figure 6. Scatter plots and linear regression of GPP and SIFc (CSIF and GOSIF) for 12 individual biomes at daily scale (p < 0.001). The statistical measures for linear regression listed in the top left corner corresponding to different color. All the linear regressions were forced to through origin. 12 biomes are the similar with Fig. 2.

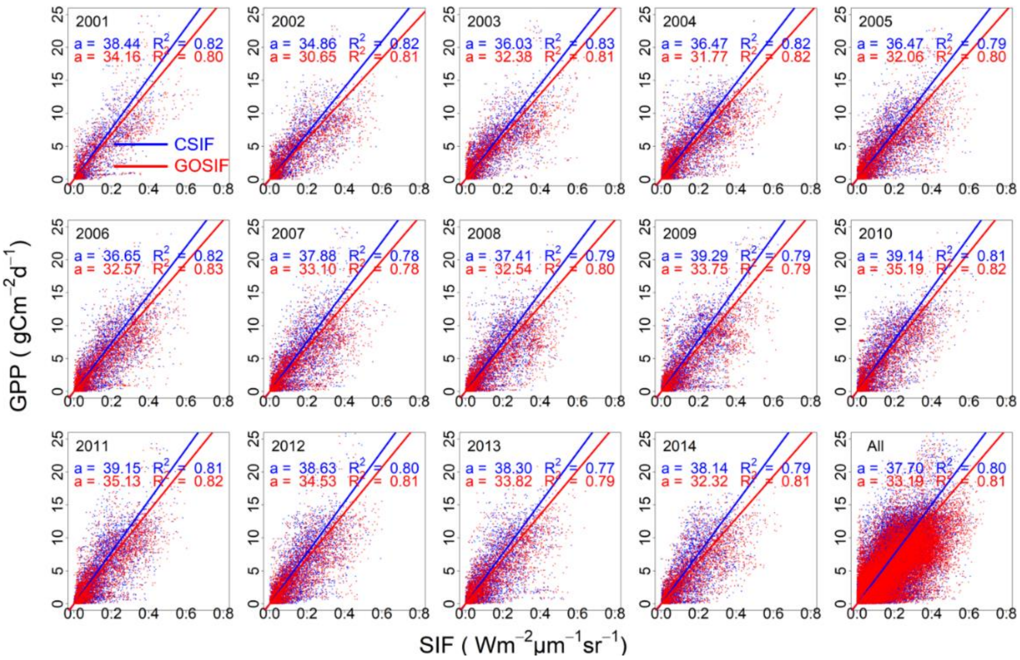

Figure 7. Scatter plots and linear regression of GPP and SIFc (CSIF and GOSIF) for all 12-biome types year by year at daily scale (p < 0.001). The statistical measures for linear regression listed in the top left corner corresponding to different color. All the linear regressions were forced to through origin. Years range from 2001 to 2014, and the last figure means all the matched daily SIFc-GPP data.

3.3. Variation of the linear SIFc-GPP relationship across spatiotemporal scales

The robustness of linear SIFc-GPP relationship increased with spatiotemporal upscaling generally (i.e., site-daily, site-yearly, site-multi-year, biome-daily, biome-yearly and biome-multi-year scale) (Fig. 8). From site to biome level, the R^2 values increased while slopes of the linear SIFc-GPP relationship significantly decreased regardless of time scale. For example, at daily scale, the R^2 value of CSIF-GPP relationship increased from 0.80 to

0.96 from site to biome level, and the corresponding slope decreased from 37.70 to 32.60. Similar changes of R^2 values and slopes of linear CSIF- GPP relationship can also be found at yearly and multi-year time scales (Fig. 8).

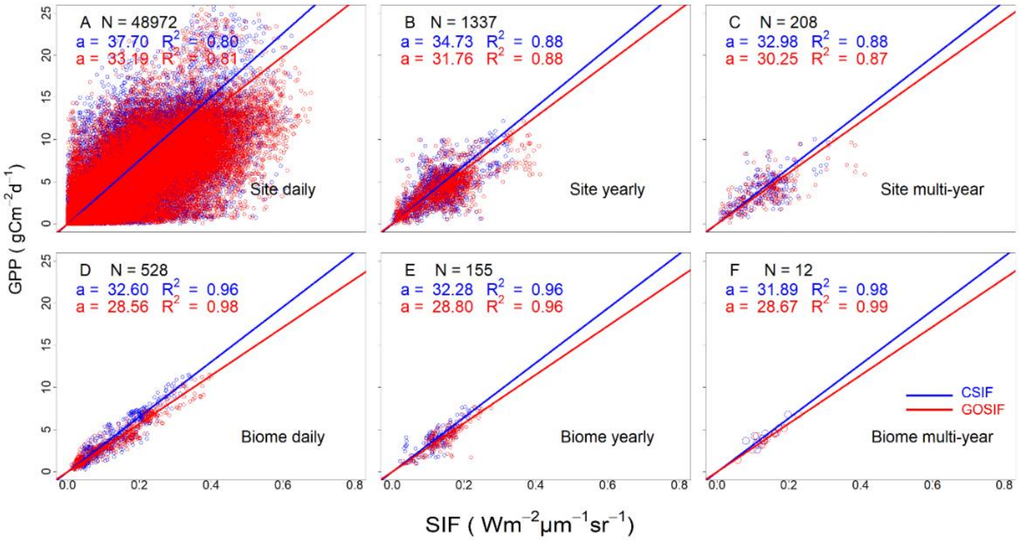

Figure 8. Scatter plots and linear regression of SIFc (both CSIF and GOSIF) and GPP for all biomes across six spatiotemporal scales ($p < 0.001$). The statistical measures for linear regression listed in the top left corner corresponding to different color. All the linear regressions were forced to through origin. The solid lines represent the fitted major axis regression models for different SIFc-GPP combinations: CSIF- GPP (blue) and GOSIF- GPP (red). (A) site-daily; (B) site-yearly; (C) site-multi-year; (D) biome-daily; (E) biome-yearly and (F) biome-multi-year.

The change of R^2 values and slopes of linear SIFc-GPP relationship with time scale varied with spatial scale. For example, at site level, the R^2 values of linear CSIF-GPP relationship increased from daily (slope = 37.70, R^2 = 0.80) to yearly (slope = 34.73, R^2 = 0.88) scale, but did not increase to multi-year (slope = 32.98, R^2 = 0.88) scale (Fig. 8). Similarly, the R^2 values and slopes of linear CSIF-GPP relationship had little changes with the temporal scale at biome level (daily: slope = 32.60, R^2 = 0.96; yearly: slope = 32.28, R^2 = 0.96; multi-year: slope = 31.89, R^2 = 0.98) (Fig. 8).

4. Discussion

4.1. Dataset dependence of the SIFc-GPP relationship

The linear SIFc-GPP relationship forcing to pass original point developed from GPP and two contiguous SIFc datasets is SIFc dataset dependent (Fig. 2). This SIFc dependency can be explained by the fact that GOSIF is generally higher in value than CSIF across biomes and years (Figs. A3 and A4), which can be traced back to their reconstruction methods and the uncertainty of the SIFc products. GOSIF, generated from discrete OCO-2 SIF soundings, EVI and land cover type data from MODIS and meteorological reanalysis data, had a RMSE of only 0.07 W m^{-2} μm^{-1} sr^{-1} [31]. In contrast, CSIF, generated using a machine learning approach (trained by discrete OCO-2 SIF soundings and MODIS surface reflectance), had a RMSE of 0.18 W m^{-2} μm^{-1} sr^{-1} [29], more than doubled that of GOSIF. The stronger SIFc-GPP relationship derived from GOSIF than that from CSIF was consistent with previous studies [31]. Zhang, Joiner [29] found the R^2 value of linear relationship between GPP derived from 40 EC flux towers and CSIF ranges from 0.01 to 0.93 with a median value of 0.64. And Li and Xiao [31] reported a higher linear relationship between GOSIF and flux GPP (R^2 = 0.73, $p < 0.001$) based on GPP from 91 EC flux towers. However, GOSIF did not always performed better than CSIF as shown in some years and some places (Figs. 4 and 5), which might be influenced by the meteorological conditions input. On the other hand, as the differences existing between both SIFc datasets should include variability ranges, systematical bias between different SIFc datasets would lead to different offsets for the linear relationship. Hence, further efforts in improving SIFc-GPP relationship must reconcile differences from different SIFc datasets and further understand their implications [29-31].

4.2. Variability of the SIFc-GPP relationship

Our results show that the linear SIFc-GPP relationship is significantly affected by the spatial and temporal scales (Fig. 8). SIFc-GPP shows the strongest linear relationship at the coarsest scales (i.e., biome-multi-year). The R^2 value of the linear SIFc-GPP relationship increased with spatial upscaling from site to biome at all temporal scales (i.e., daily, year and multi-year). In contrast, the R^2 value did not necessarily increase with temporal upscaling at different spatial scales (i.e., site and biome). This suggests that SIFc (both CSIF and GOSIF) is not effective in capturing temporal variabilities of GPP, particularly the inter-annual variability. Overall, the reconstructed SIFc performs well in tracking long-term biome-wide GPP (Fig. 8F), consistent with other studies [9]. The reduced ability of SIFc in capturing the short-term changes of GPP might largely attributed to the errors in SIFc and GPP products, as well as the footprint mismatch between SIFc and GPP, especially at finer resolutions and during the reconstruction period (from 2001 to 2014) [29, 31].

It should be noticed that, despite moderate to strong R^2 values, there was considerable dispersion in the comparisons of CSIF/GOSIF and GPP at both site and biome levels (Figs. 3 and A6). The scattered distribution might be caused by SIFc and flux_GPP data quality as well as the non-universality of the linear GPP-SIFc relationship [43, 44], particularly at CRO and DBF biomes (Figs. A3 and A6). For example, SIFc-GPP points scattered around the linear regression lines widely at the DE.Geb site (CRO biome) (Fig. A6) while the aggregated annual change of flux_GPP synchronized well with those of SIFc in addition to many scattered points caused by interannual variability (Fig. A5), suggesting interannual variability of cropping practices (e.g., rotation of crops, fallow, and fertilization) may contribute substantially to the pronounced scattering of points around the regression lines in Fig. A6. On the other hand, there were clearly two clusters at the IT.Noe site (CSH biome), which signifies major difference between CSIF and GOSIF there. Apparently, future efforts are required to investigate the variability in the SIF-GPP relationship systematically to answer a suite of important questions: where/when does a linear SIF-GPP relationship break down? Where/when does it change in slope and why?

Li, Xiao [21] has reported that C4-dominated grasslands and crops, albeit only two C4 sites, had a significantly higher slope than C3-dominated grasslands and crops (29.42 *vs.* 20.98, $p < 0.001$). and Wood, Griffis [22] found a linear SIF-GPP relationship that is sensitive to crop type (corn *vs.* soybean) as well. However, our results suggested that there was no significant difference ($p > 0.05$) in the slopes between C4 ($n = 3$) and C3 ($n = 8$) crops at site-multi-year scale, respectively, for CSIF (C4 *vs.* C3: 42.71 ± 1.33 (mean ± SE) *vs.*38.87 ± 5.38) and for GOSIF (C4 *vs.*C3: 38.70 ± 3.48 *vs.*40.32 ± 6.11). The difference between our study and Li, Xiao [21] may be due to the limited number of C4 crop sites and the different approaches used for analysis. We compared the difference in the means of the slopes from individual C3 and C4 sites while Li, Xiao [21] compared the difference in the overall slopes of the C3 and C4 crops after pooling SIF and GPP data from all C3 and C4 sites. Apparently, further research is needed to understand the differences in the SIF-GPP relationships for C3/C4 plants with more C4 sites.

The number of EC flux towers is not balanced among the biomes, and some biomes only include one or a few sites. This leads to large uncertainties in the linear GPP/SIFc slopes in some biomes (e.g., CSH, OSH and WET). For example, the GPP/CSIF slopes of OSH and DNF were lower than other biomes; there are clearly two clusters of data captured within CSH (Fig. 3). The main reason may be the limited GPP data at OSH (72 site-year) and DNF (3 site-year) biomes. Thus, increasing the number of EC flux towers, particularly in some underrepresented biomes (e.g., OSH and DNF), is necessary to make our global analysis more representative and robust to support GPP modeling using SIF.

4.3 A generic two-slope scheme SIFc-GPP relationship

Our study found that at site-multi-year scale there was no significant difference between any biome pairs in GPP/SIFc slopes except a few pairs between ENF and others (Fig. 3). Specifically, GPP/SIFc slopes in ENF biome were significantly higher than those in four biomes (DBF, EBF, GRA and OSH) according to CSIF or five biomes (DBF, EBF, GRA, OSH and MF) according to GOSIF, and the slopes between any other non-ENF biome pairs were not significantly different. To summarize these findings, we therefore propose a two-slope scheme to differentiate ENF from non-ENF and synopsize the GPP/SIFc slope variability across all biomes and years. It should be noted that the two-slope scheme is SIFc dataset specific (Table 1), resulted from the systematic differences between these two SIFc datasets.

Table 1. Two-slope scheme of linear SIFc-GPP relationship based on GPP/SIFc slopes at site-multi-year scale. The SIFc dataset dependent scheme, divides all sites into two groups: ENF and non-ENF (11 biomes) according to the significance of the GPP/SIFc slopes (see Fig. 3). All slopes are represented as mean ± SE or median (MAD). N is the number of sites included in the specific biome.

Biome	N	CSIF mean	GOSIF mean	CSIF median	GOSIF median
CRO	20	39.94±3.39	39.14±3.45	37.95(15.45)	34.68(13.99)
CSH	3	46.51±23.86	27.70±2.54	26.19(10.25)	30.09(0.44)
DBF	26	38.25±4.79	40.96±10.88	33.87(6.27)	30.75(7.90)
DNF	1	24.15	26.59	24.15	26.59
EBF	15	34.49±2.91	30.15±2.76	34.30(9.85)	30.48(5.88)
GRA	37	36.20±2.86	35.15±3.23	32.31(14.43)	30.38(7.91)
MF	9	42.36±7.60	35.83±8.91	37.18(4.2)	28.05(4.24)
OSH	14	30.69±4.47	28.64±3.99	28.05(13.57)	25.54(15.12)
SAV	8	56.90±17.40	42.49±7.77	38.63(14.04)	34.21(15.87)
WET	21	41.23±5.25	61.66±16.05	38.44(22.35)	33.89(22.00)
WSA	6	33.36±4.31	31.04±2.83	29.77(5.78)	29.04(5.52)
ENF	**48**	**42.75±2.04**	**40.36±2.28**	**43.21(13.39)**	**38.61(10.65)**
Non_ENF	**160**	**38.41±1.74**	**39.09±3.02**	**34.36(11.66)**	**30.39(9.65)**

To sift the statistics (mean or median slope) building the two-slope scheme, we compared the SIFc-derived GPP with flux tower GPP. It can be seen that the two-slope scheme derived from median values (median PB were 7.14% and 11.06% for GOSIF and CSIF, respectively) outperformed the one from the mean values (median PB were 31.65% and 20.67% for GOSIF and CSIF, respectively) in estimating GPP across all EC towers (Figs. 9, A8 and A9), probably the median-based scheme effectively avoided the impacts of slope outliers. We thus used the median values of slopes to develop the two-slope scheme in this study. The median slopes for the GPP/CSIF were 43.21 (13.39) and 34.36 (11.66) with corresponding mean ± SE as 42.75 ± 2.04 and 38.41 ± 1.74, respectively for ENF and other biomes (Table 1), resulting in the following two-slope scheme for converting CSIF into GPP:

$$\mathrm{GPP} = \begin{cases} 43.21 \times \mathrm{SIF}, & \text{ENF biome} \\ 34.36 \times \mathrm{SIF}, & \text{other biomes} \end{cases} \quad (1)$$

The median slopes for the GPP/GOSIF were 38.61 (10.65) and 30.39 (9.65) with corresponding mean ± SE as 40.36 ± 2.28 and 39.09 ± 3.02, respectively for ENF and other biomes, and the corresponding two-slope scheme was:

$$\mathrm{GPP} = \begin{cases} 38.61 \times \mathrm{SIF}, & \text{ENF biome} \\ 30.39 \times \mathrm{SIF}, & \text{other biomes} \end{cases} \quad (2)$$

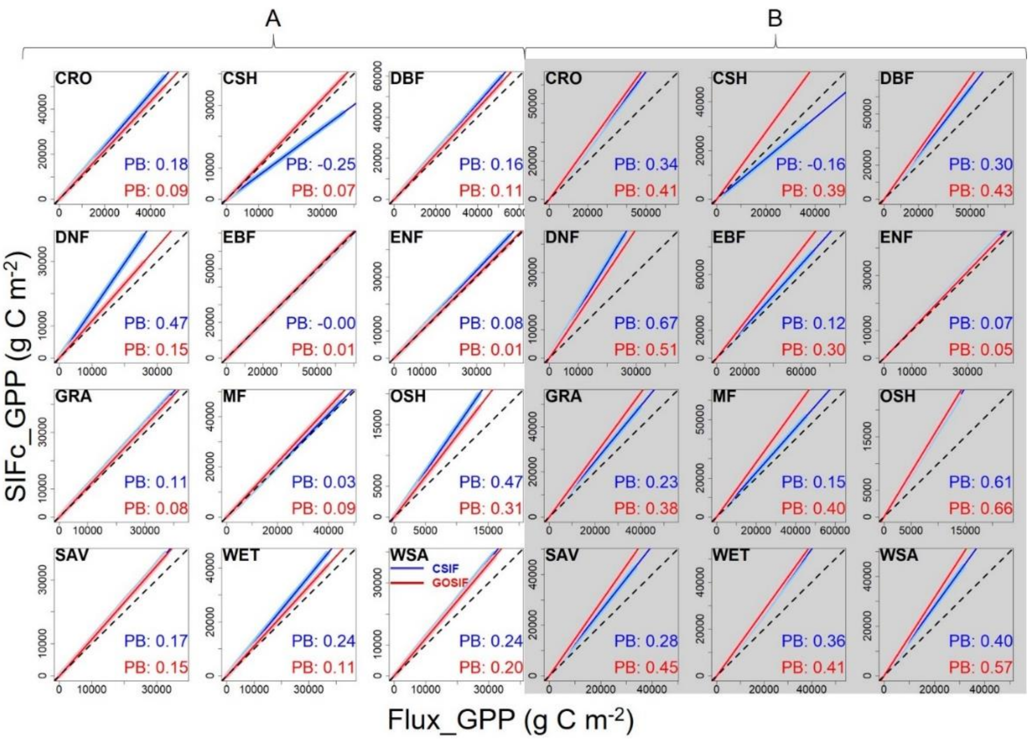

Figure 9. Comparison of the cumulative Flux_GPP and SIFc_GPP across 12 biomes. Median-based two-slope scheme (A) and Mean-based two-slope scheme (B). The two two-slope schemes are listed in table 1. The dash line is the 1:1 line. Blue and red lines are the fitted linear regression lines of cumulative CSIF_GPP and GOSIF_GPP with Flux_GPP, respectively, and their associated PB values are provided as well. To avoid the impacts of large unbalance in the number of GPP values across biomes on accumulative GPP, 10,000 daily Flux_GPP and SIFc_GPP value pairs were sampled with replacement for each biome for this comparison.

The two-slope scheme provides a very convenient and effective tool for converting SIFc to GPP and monitoring GPP dynamics in time and space as it is almost land cover independent (only the distribution of ENF needs to be identified). The significantly higher slopes for ENF biome in the two-slope scheme are intriguing and deserve further study. This phenomenon is in line with the observations reported by Gamon, Huemmrich [45] and Zhang, Joiner [29] who pointed out that the lower SIF (therefore higher GPP/SIFc slope) for ENF is mainly caused by a stronger canopy reabsorption and/or scattering of SIF for needle leaf forest, and the core mechanism is the high dependency between SIF and APAR, chlorophyll content [46] and photosynthetic light-use efficiency [47]. However, it is still a great challenge to measure (in field), observe (from satellite) and model (based mechanism) photosynthesis in boreal forests, especially at ENF biome [47, 48]. More in-depth research is still needed to expand our understanding of the effects of needle leaf clumping index [49]), leaf chlorophyll content [50], and chlorophyll/carotenoid index [45] on plant photosynthesis (especially for SIF as an agents) from canopy to global scale [9].

The coefficients for the two SIF data sets are different from each other (Equations 1 and 2). It can be explained by the fact that GOSIF is generally higher in value than CSIF across biomes and years (Figs. A4 and A5), which can be traced back to their reconstruction methods and the uncertainty of the SIFc products. GOSIF, generated from discrete OCO-2 SIF soundings, EVI and land cover type data from MODIS and meteorological reanalysis data, had a RMSE of only 0.07 W m^{-2} μm^{-1} sr^{-1} [31]. In contrast, CSIF, generated using a machine learning approach (trained by discrete OCO-2 SIF soundings and MODIS surface reflectance), had a RMSE of 0.18 W m^{-2} μm^{-1} sr^{-1} [29], more than doubled that of GOSIF.

Previous studies have highlighted that the linear SIF-GPP relationship is either biome-dependent [22, 26] or ecosystem-specific (e.g., Sun, C. Frankenberg [2]). To our knowledge, none of them has really examined the discriminatory power of their datasets on the observed differences of SIF-GPP relationship across ecosystems or biomes. In other words, previous studies often studied the uniqueness of the SIF-GPP relationships, and none addressed their commonality or the discriminatory power of their datasets across biomes. Our two-slope scheme represents a major step forward in this direction. In addition, it provides a practicable method for estimating GPP from SIFc with a greatly reduced need on land cover specificity, which should benefit the reduction of GPP uncertainty from land cover classification. This general scheme may have reconciled the differences among previous studies that were either restricted to small regions [51], few flux towers and/or biomes [21], or with low spatiotemporal resolutions [28].

4.4. *Potential caveats and uncertainties*

Landscape heterogeneity and inconsistency between the flux-tower footprint and SIFc pixel should have contributed to the uncertainty of our results. We acknowledge that the landscape heterogeneity at the EC flux towers is an important obstacle to analyzing SIF-GPP relationship. Although we have visually checked landscape conditions around all EC-flux tower sites using Google Earth images, and removed four flux towers from our analysis. A more robust approach to address the issue would be using a footprint model to obtain the footprints of all the sites. Our manual examination approach resulted in 208 sites, which was different from that of Zhang, Joiner [29] who selected only 40 sites using an automated NDVI-based approach. Our results therefore might have higher uncertainties than Zhang, Joiner [29] but at the same time encompassed more spatial variability of sites globally which might enhance the robustness of our results. The latter is demonstrated by the good performance of the median-based two-slope scheme at many flux-tower sites and biomes (Fig. 9). Retrospectively, our two-slope scheme suggests that the impact of landscape heterogeneity and inconsistency between the flux tower footprint and SIFc pixel might not as severe as we previously thought.

In addition, for CRO and GRA biomes, the site selection is in particularly very important as the flux towers at these biomes are of a height of 2-6 meters and the footprint area is like a lot when visualized with the SIFc pixel. On top of that the GPP from these sites very much depend on the crop and management practice (e.g., rotation of crops, fallow, grazing, and fertilization), which can change in every few hundred meters thereby making the satellite-ground comparison challenging.

Moreover, whether the regression method used in this study works well also introduced uncertainty. Just from the mathematical view, one of the fundamental problems of forcing the intercept to zero is getting much higher R^2 (Fig. 8), which will lead to a large portion of bias in results interpretating, especially at daily scale [2]. While at the point of vegetation physiology, the zero-intercept logic provides a unique perspective for the SIFc-GPP relationship analysis. Xiao, Li [52] reported that low daily SIF/GPP measurements are not available in some areas/biomes, such as EBF. In such occasions, the application of our zero-intercept logic makes more sense, which may have greater predictability under unseen conditions (e.g., low SIF/GPP). As this study focused on the comparison of GPP/SIFc slopes across different biomes, we finally applied the zero-intercept method rather than the free intercept. Nevertheless, some free intercept regression model would provide more information about the SIFc-GPP relationship.

5. Conclusions

Our work is a global analysis investigating the relationship between SIFc and GPP at various spatial and temporal scales, which expands previous research on this topic particularly in the following two areas. First, we used all GPP data in the FLUXNET2015 collection and two global SIFc products for the analysis, provided the most comprehensive coverage so far (208 flux towers and the longest study period from 2001 to 2014) to explore SIFc-GPP relationship. Second, we used Major Axis regression to account for uncertainties in both SIFc and GPP estimates in the analysis of SIFc-GPP relationship which produced higher GPP/SIFc slopes than OLS. Our research expands several pioneering works which have reported the relationship between OCO-2 SIF and tower GPP at individual sites and few biomes. We propose a two-slope scheme to differentiate ENF from non-ENF biome and synopsize the GPP/SIFc slope variability across biomes and years. The relative biases were 7.14% and 11.06% in the estimated cumulative GPP across all EC towers, respectively, for GOSIF and CSIF using the two-slope scheme. Nevertheless, our results suggested some major issues related to SIFc-GPP relationship including dataset dependency of the SIFc-GPP relationship, variability of the SIFc-GPP relationship across spatial and temporal scales, and a two-slope scheme that was distilled from SIFc-GPP relationships across biomes. We thus call for more research in these issues mentioned above, and offered a few thoughts on caveats, uncertainty of our research, and future research directions to provide clues for further research.

Author Contributions: Conceptualization, S.L.; methodology, S.L. and H.G.; writing—original draft preparation, H.G.; writing—review and editing, S.L., W.L., A.S., W.Y., Z.W., L.X., Q.L., Y.F., M.M., T.P., J.L., Z.W., G.Z., and H.G.; supervision, S.L.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 41971152); and the Hunan Innovative Talent Program (Grant No. 2019RS1062) to S.L.

Acknowledgments: This work used EC flux data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET EC flux data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices. The CSIF dataset is provided by Yao Zhang, and the GOSIF dataset is provided by Xing Li and Jingfeng Xiao. We thank the anonymous reviewers for their valuable comments.

Conflicts of Interest: We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service, and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled "Global analysis of the relationship between reconstructed solar induced chlorophyll fluorescence (SIF) and GPP".

References

1. Beer, C., et al., *Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate.* Sci, 2010. **329**(5993): p. 834-8.
2. Sun, Y., et al., *OCO-2 advances photosynthesis observation from space via solarinduced chlorophyll fluorescence.* Sci, 2017. **358**: p. 189-197.
3. Frankenberg, C., et al., *New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity.* Geophys. Res. Lett., 2011. **38**(17).
4. Wei, X., et al., *Use of Sun-Induced Chlorophyll Fluorescence Obtained by OCO-2 and GOME-2 for GPP Estimates of the Heihe River Basin, China.* Remote Sensing, 2018. **10**(12): p. 2039.
5. Eldering, A., et al., *The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes.* Sci, 2017. **358**(6360).
6. Baker, N.R., *Chlorophyll fluorescence: a probe of photosynthesis in vivo.* Annu. Rev. Plant Biol., 2008. **59**: p. 89-113.
7. Zhang, L., et al., *Monitoring Drought Effects on Vegetation Productivity Using Satellite Solar-Induced Chlorophyll Fluorescence.* Remote Sensing, 2019. **11**(4).
8. Magney, T.S., et al., *Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence.* Proceedings of the National Academy of Sciences, 2019: p. 201900278.
9. Ryu, Y., J.A. Berry, and D.D. Baldocchi, *What is global photosynthesis? History, uncertainties and opportunities.* Remote Sens. Environ., 2019. **223**: p. 95-114.
10. Yokota, T., et al., *Global concentrations of CO2 and CH4 retrieved from GOSAT: First preliminary results.* SOLA, 2009. **5**: p. 160-163.
11. Joiner, J., et al., *Filling-in of near-infrared solar lines by terrestrial fluorescence and other geophysical effects: simulations and space-based observations from SCIAMACHY and GOSAT.* Atmospheric Measurement Techniques, 2012. **5**(4): p. 809-829.
12. Joiner, J., et al., *Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2.* Atmospheric Measurement Techniques, 2013. **6**(10): p. 2803-2823.
13. Köhler, P., L. Guanter, and J. Joiner, *A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data.* Atmospheric Measurement Techniques, 2015. **8**(6): p. 2589-2608.
14. Wen, J., et al., *A framework for harmonizing multiple satellite instruments to generate a long-term global high spatial-resolution solar-induced chlorophyll fluorescence (SIF).* Remote Sens. Environ., 2020. **239**.
15. Du, S.S., et al., *Retrieval of global terrestrial solar-induced chlorophyll fluorescence from TanSat satellite.* Science Bulletin, 2018. **63**(22): p. 1502-1512.
16. Köhler, P., et al., *Global Retrievals of Solar-Induced Chlorophyll Fluorescence With TROPOMI: First Results and Intersensor Comparison to OCO-2.* Geophys. Res. Lett., 2018.
17. Frankenberg, C., et al., *Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2.* Remote Sens. Environ., 2014. **147**: p. 1-12.
18. Zuromski, L.M., et al., *Solar-Induced Fluorescence Detects Interannual Variation in Gross Primary Production of Coniferous Forests in the Western United States.* Geophys. Res. Lett., 2018. **45**(14): p. 7184-7193.
19. Guanter, L., et al., *Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements.* Remote Sens. Environ., 2012. **121**: p. 236-251.
20. Porcar Castell, A., et al., *Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges.* J Exp Bot, 2014. **65**(15): p. 4065-95.

21. Li, X., et al., *Solar-induced chlorophyll fluorescence is strongly correlated with terrestrial photosynthesis for a wide variety of biomes: First global analysis based on OCO-2 and flux tower observations.* Global Change Biol., 2018.
22. Wood, J.D., et al., *Multiscale analyses of solar-induced florescence and gross primary production.* Geophys. Res. Lett., 2017. **44**(1): p. 533-541.
23. Li, X., J. Xiao, and B. He, *Chlorophyll fluorescence observed by OCO-2 is strongly related to gross primary productivity estimated from flux towers in temperate forests.* Remote Sens. Environ., 2018. **204**: p. 659-671.
24. Zhang, Y., 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**: p. 276-289.
25. Wang, X., J.M. Chen, and 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., 2020. **246**: p. 111888.
26. Sun, Y., et al., *Overview of Solar-Induced chlorophyll Fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP.* Remote Sens. Environ., 2018. **209**: p. 808-823.
27. Chen, B., et al., *Assessing Tower Flux Footprint Climatology and Scaling Between Remotely Sensed and Eddy Covariance Measurements.* BoLMe, 2009. **130**(2): p. 137-167.
28. Zhang, Y., et al., *Model-based analysis of the relationship between sun-induced chlorophyll fluorescence and gross primary production for remote sensing applications.* Remote Sens. Environ., 2016. **187**: p. 145-155.
29. Zhang, Y., et al., *A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks.* Biogeosciences, 2018. **15**(19): p. 5779-5800.
30. Yu, L., et al., *High Resolution Global Contiguous Solar‐Induced Chlorophyll Fluorescence (SIF) of Orbiting Carbon Observatory‐2 (OCO‐2).* Geophys. Res. Lett., 2018.
31. Li, X. and J. Xiao, *A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS, and Reanalysis Data.* Remote Sensing, 2019. **11**(5): p. 517-539.
32. Gelaro, R., et al., *The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2).* J. Clim., 2017. **30**.
33. Duveiller, G., et al., *A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity.* Earth System Science Data, 2020. **12**(2): p. 1101-1116.
34. Pastorello, G., et al., *A new data set to keep a sharper eye on land-air exchanges.* Eos, Transactions American Geophysical Union (Online), 2017. **98**(8).
35. Li and Xiao, *Mapping Photosynthesis Solely from Solar-Induced Chlorophyll Fluorescence: A Global, Fine-Resolution Dataset of Gross Primary Production Derived from OCO-2.* Remote Sensing, 2019. **11**(21).
36. Reichstein, M., et al., *On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm.* Global Change Biol., 2005. **11**(9): p. 1424-1439.
37. Zhang, Y., et al., *A global moderate resolution dataset of gross primary production of vegetation for 2000-2016.* Sci Data, 2017. **4**: p. 170165.
38. Friedl, M.A., et al., *Global land cover mapping from MODIS: Algorithms and early results.* Remote Sens. Environ., 2002. **83**: p. 287.
39. R Core Team, *R: A language and environment for statistical computing*. 2019: R Foundation for Statistical Computing.
40. Loveland, T., et al., *Development of a global land characteristics database and IGBP DISCover from 1 km AVHRR data.* Int. J. Remote Sens., 2000. **21**: p. 1303-1330.
41. Belward, A.S., *The IGBP-DIS global 1 km land cover data set" DISCover": Proposal and implementation plans: Report of the Land Cover Working Group of IGBP-DIS*. 1992: IGBP-DIS Office.
42. Warton, D.I., et al., *smatr 3- an R package for estimation and inference about allometric lines.* Methods Ecol. Evol., 2012. **3**(2): p. 257-259.

43. Marrs, J.K., et al., *Solar-induced fluorescence does not track photosynthetic carbon assimilation following induced stomatal closure.* Geophys. Res. Lett., 2020. **n/a**(n/a): p. e2020GL087956.
44. Wang, Z., et al., *Tighten the Bolts and Nuts on GPP Estimations from Sites to the Globe: An Assessment of Remote Sensing Based LUE Models and Supporting Data Fields.* Remote Sensing, 2021. **13**(2).
45. Gamon, J.A., et al., *A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers.* Proc Natl Acad Sci U S A, 2016. **113**(46): p. 13087-13092.
46. Louis, J., et al., *Remote sensing of sunlight-induced chlorophyll fluorescence and reflectance of Scots pine in the boreal forest during spring recovery.* Remote Sens. Environ., 2005. **96**(1): p. 37-48.
47. Walther, S., et al., *Satellite chlorophyll fluorescence measurements reveal large-scale decoupling of photosynthesis and greenness dynamics in boreal evergreen forests.* Glob. Chang. Biol., 2016. **22**(9): p. 2979-96.
48. Pisek, J., J.M. Chen, and T. Nilson, *Estimation of vegetation clumping index using MODIS BRDF data.* Int. J. Remote Sens., 2011. **32**(9): p. 2645-2657.
49. He, L., et al., *Inter- and intra-annual variations of clumping index derived from the MODIS BRDF product.* IJAEO, 2016. **44**: p. 53-60.
50. Croft, H., et al., *Leaf chlorophyll content as a proxy for leaf photosynthetic capacity.* Glob. Chang. Biol., 2017. **23**(9): p. 3513-3524.
51. Zhang, Y., et al., *Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America.* Remote Sens. Environ., 2016. **183**: p. 154-169.
52. Xiao, J., et al., *Solar-induced chlorophyll fluorescence exhibits a universal relationship with gross primary productivity across a wide variety of biomes.* Glob. Chang. Biol., 2019.