

Article

A Reconstructed Global Daily Seamless SIF Product at 0.05 Degree Resolution Based on TROPOMI, MODIS and ERA5 Data

Jiaochan Hu ¹, Jia Jia ¹, Yan Ma ², Liangyun Liu ^{2,*}  and Haoyang Yu ³ 

¹ College of Environmental Sciences and Engineering, Dalian Maritime University, Dalian 116026, China; hujc@dmlu.edu.cn (J.H.); jiaj15@dmlu.edu.cn (J.J.)

² Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; mayan2017@radi.ac.cn

³ Information Science and Technology College, Dalian Maritime University, Dalian 116026, China; yuhy@dmlu.edu.cn

* Correspondence: liuly@radi.ac.cn; Tel.: +86-10-8217-8163

Abstract: Satellite-derived solar-induced chlorophyll fluorescence (SIF) has been proven to be a valuable tool for monitoring vegetation's photosynthetic activity at regional or global scales. However, the coarse spatiotemporal resolution or discrete space coverage of most satellite SIF datasets hinders their full potential for studying carbon cycle and ecological processes at finer scales. Although the recent TROPospheric Monitoring Instrument (TROPOMI) partially addresses this issue, the SIF still has drawbacks in spatial insufficiency and spatiotemporal discontinuities when gridded at high spatiotemporal resolutions (e.g., 0.05°, 1-day or 2-day) due to its nonuniform sampling sizes, swath gaps, and clouds contaminations. Here, we generated a new global SIF product with Seamless spatiotemporal coverage at Daily and 0.05° resolutions (SDSIF) during 2018–2020, using the random forest (RF) approach together with TROPOMI SIF, MODIS reflectance and meteorological datasets. We investigated how the model accuracy was affected by selection of explanatory variables and model constraints. Eventually, models were trained and applied for specific continents and months given the similar response of SIF to environmental variables within closer space and time. This strategy achieved better accuracy ($R^2 = 0.928$, RMSE = 0.0597 mW/m²/nm/sr) than one universal model ($R^2 = 0.913$, RMSE = 0.0653 mW/m²/nm/sr) for testing samples. The SDSIF product can well preserve the temporal and spatial characteristics in original TROPOMI SIF with high temporal correlations (mean R^2 around 0.750) and low spatial residuals (less than ± 0.081 mW/m²/nm/sr) between them two at most regions (80% of global pixels). Compared with the original SIF at five flux sites, SDSIF filled the temporal gaps and was better consistent with tower-based SIF at the daily scale (the mean R^2 increased from 0.467 to 0.744. Consequently, it provided more reliable 4-day SIF averages than the original ones from sparse daily observations (e.g., the R^2 at Daman site was raised from 0.614 to 0.837), which resulted in a better correlation with 4-day tower-based GPP. Additionally, the global coverage ratio and local spatial details had also been improved by the reconstructed seamless SIF. Our product has advantages in spatiotemporal continuities and details over the original TROPOMI SIF, which will benefit the application of satellite SIF for understanding carbon cycle and ecological processes at finer spatial and temporal scales.

Keywords: solar-induced chlorophyll fluorescence; TROPOMI; global daily seamless; MODIS; flux-towers; spatiotemporal resolution enhancement



Citation: Hu, J.; Jia, J.; Ma, Y.; Liu, L.; Yu, H. A Reconstructed Global Daily Seamless SIF Product at 0.05 Degree Resolution Based on TROPOMI, MODIS and ERA5 Data. *Remote Sens.* **2022**, *14*, 1504. <https://doi.org/10.3390/rs14061504>

Academic Editors: Peiqi Yang and Xinjie Liu

Received: 24 January 2022

Accepted: 15 March 2022

Published: 20 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Solar-induced chlorophyll fluorescence (SIF) is an emitted optical signal by plant chlorophyll during the photosynthesis process under natural sunlight, which is the main process governing the global carbon cycle [1]. As a byproduct of photosynthesis, SIF can

serve as a probe into a plant's physiological state and photosynthetic capacity, unlike the vegetation indices (VIs) that just reflect the plant's apparent 'greenness' [2,3]. With successful SIF retrieval from satellite sensors, SIF opens up an innovative insight in monitoring the spatial and temporal patterns of the terrestrial carbon fixation by plants (i.e., the Gross Primary Productivity (GPP)) based on remote sensing data. Numerous studies have verified the strong relationship between SIF and GPP and thus adopt SIF in the GPP estimation at regional or global scales [4–9]. Moreover, satellite SIF is widely used in other ecological, agricultural, and forestry fields, such as crop yield estimation [10,11], drought or water stress detecting [12,13], or phenology monitoring [14,15]. These applications promote SIF to being a research hotspot and put forward a strong demand for more refined SIF global products in space and time.

Over the past several decades, numerous satellite SIF datasets have been successfully retrieved from different spaceborne instruments, including: GOSAT [16], SCIAMACHY [17], GOME-2 [18], OCO-2 [19], and TanSat [20]. However, these sensors are either spatially sparse sampling (e.g., GOSAT, OCO-2, and Tansat) or have coarse spatial resolutions (e.g., SCIAMACHY: $30 \times 240 \text{ km}^2$, GOME-2: $40 \times 40 \text{ km}^2$), and always with low revisit frequencies (e.g., GOME-2: 1.5 days, and OCO-2/Tansat: 16 days). In addition, the effects of cloud contamination on retrieved SIF signal further reduce the number of valid observations. Moreover, in order to reduce the large noises inherent in individual SIF retrievals, raw SIF soundings need to be aggregated to gridded products with spatial and temporal averaging for further applications. This results in coarser spatiotemporal resolutions for SIF datasets, such as the GOME-2 SIF products at 0.5° grids and monthly/biweekly scales, and the global contiguous OCO-2 products at 1° grids and monthly scales [21]. The arrival of Tropospheric Monitoring Instrument (TROPOMI) on board Sentinel-5P in October 2017 substantially improves the ability in spatiotemporal sampling of earlier sensors. It can provide SIF retrievals at a spatial continuous sampling of around 3.5–14 km across track and 7 km along track (5 km since August 2019) with a revisit frequency of nearly one day. Several TROPOMI SIF datasets are successively retrieved and published by [22–24].

However, the TROPOMI SIF still has spatiotemporal limitations when generated at high resolutions (e.g., 0.05° , 1-day or 2-day interval). First, the size of TROPOMI sampling footprint is varied across track. It approaches to 0.05° (denoting the size in both latitude and longitude in this study) when at nadir, and frequently is larger than 0.05° when away from the nadir. Figure 1a displays an enlarged image of one region far away from the nadir for example. Referring to [22], each grid cell is the average of those footprints covering the center of this grid. The size of TROPOMI footprints (black solid frames) are 2–3 times larger than the 0.05° grids (gray dotted frames). In this case, the composite 0.05° SIF cannot achieve the real expected resolution, which contains large uncertainties in heterogeneous area. Second, the TROPOMI SIF has remarkable discontinuities at 1-day or 2-day scales. Lots of blank terrestrial areas exist within one day due to the swath gaps and thick clouds (Figure 1b), which results in a global coverage of less than 75% and 88%, respectively, within one day and two days during 2019 (Figure 1c). Meanwhile, the resulting temporal gaps are also noteworthy. Taking the Aurora site in north America (42.7228°N , 76.6628°W) for an example, almost a half and a quarter of the one-day and two-day periods during 2019 have none valid observations (see the upper subgraph of Figure 1d). Even at 4-day scales, the gaps are still present and more than half of the 4-day periods have less than two observations. Note that, cloudy-sky SIF retrievals are retained (cloudy fraction less than 0.8) and no screening of observation numbers for data averaging is conducted in this original SIF (called TROSIF⁰⁰⁵ in this study). With more strict screening of cloudy fraction and observation numbers to reduce the noises, the valid cells will have a huge decrease.

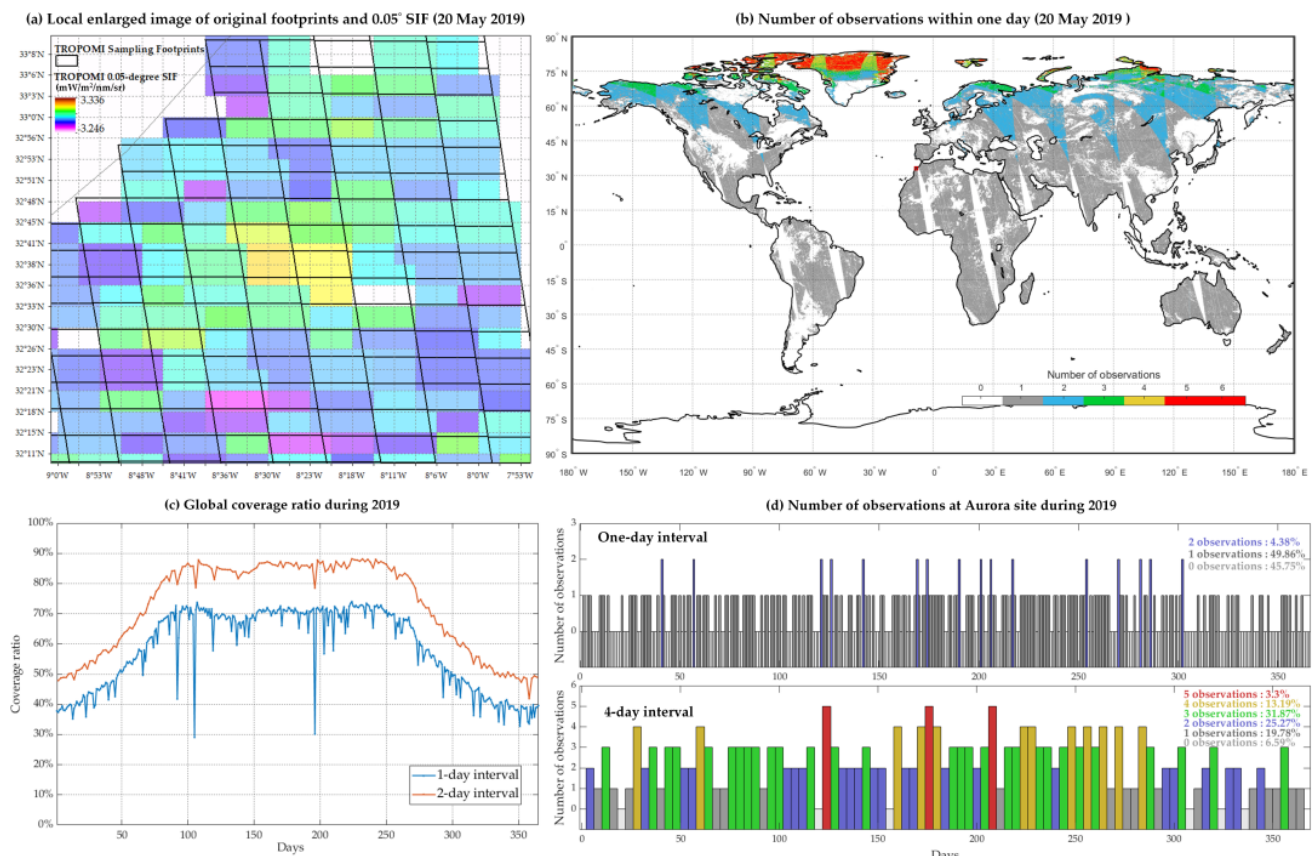


Figure 1. Visualizations of the spatiotemporal limitations in original TROPOMI SIF including spatial resolution insufficiency (a), spatial gaps (b), and temporal discontinuities (c,d) at 0.05° resolution.

The above two drawbacks in spatial insufficiency and spatiotemporal discontinuities will hinder the potential of TROPOMI SIF for finer-scale applications, especially for the resulting spatiotemporal inconsistency with tower-based GPP or other continuous datasets. First of all, the insufficiency with 0.05° will cause uncertainties in heterogeneous regions, e.g., the spatial inconsistent with tower flux measurements will introduce errors in SIF-GPP relationships and SIF-based GPP estimations. On the other hand, since SIF is a highly changing signal with time, the multi-day SIF averages from sparse days (like 4-day SIF in Figure 1d) cannot represent the real SIF signal at this time scale. It will cause a temporal mismatch with continuous carbon flux measurements and introduce uncertainties in relationships between SIF and GPP, particularly for the cases in many studies [22,25] that only instantaneous clear-sky SIF values or their n-day averages were used for GPP estimations [26]. Therefore, developing a spatiotemporal enhancement method for TROPOMI SIF to generate a 0.05° global daily seamless SIF dataset is quite valuable.

To enhance spatial details of SIF datasets, a few studies have developed spatial downscaling or gap-filling approaches for coarse GOME-2 [27–30] or sparse OCO-2/Tansat SIF datasets [31–34], respectively. The first of such models depends on the semi-empirical function from the GPP light-used efficiency (LUE) concept [27]. More recently, machine-learning (ML) algorithms such as Neural Networks (NN), regression tree, or Random Forest (RF) are adopted due to its more flexible fitting, which have been proven to be effective in SIF reconstruction [28,30–34]. Previous reconstructed SIF products are adequate in spatial resolutions (mostly 0.05°) but not meticulous in temporal resolutions (mostly monthly or 8-day intervals), which cannot provide continuous daily SIF to track the temporal dynamics. They are insufficient with the ideal temporal resolution proposed by [29] for a SIF product truly applied in the Earth system science community as a proxy for GPP. In addition, in

those studies for GOME-2, OCO-2, or Tansat SIF reconstruction, different strategies in selection of explanatory variables and model constraints are employed. However, for the reconstruction of TROPOMI SIF, the strategies need to be specifically designed and further verified with this new dataset.

The objective of this work is to reconstruct a new global Seamless SIF product at Daily and 0.05° resolutions (SDSIF) for tackling the spatiotemporal limitations of TROPOMI SIF for application. We first used the RF approach to train and test models between TROPOMI SIF retrievals and explanatory variables from Moderate Resolution Imaging Spectroradiometer (MODIS) and ERA5. In the process of model development, we investigated if the accuracy was affected by the selection of explanatory variables and the model constraints. Then, the specified models were applied to reconstruct the SDSIF product during 2018–2020 using the explanatory variables at fine scale. Validation of this product was conducted by comparisons with (i) original TROPOMI SIF retrievals at both 0.2° , daily scales and 0.1° , 16-day scales to respectively verify the preservation of temporal and spatial characteristics in original SIF, and (ii) continuous tower-based SIF retrievals from five tower sites to verify its capability of tracking SIF variations and advantages over TROSIF⁰⁰⁵ in temporal continuity and consistency at daily and 4-day scales. The spatiotemporal patterns of SDSIF were also analyzed at global and regional scales, which emphasized its advantages in data coverage and spatial details over TROSIF⁰⁰⁵ and its included physiological information compared with VIs data.

2. Materials and Methods

2.1. Datasets from Space and Ground

2.1.1. Satellite SIF Data from TROPOMI

The global original TROPOMI SIF used in this study was derived from the datasets published by [24], in which a data-driven approach [35] was applied to retrieve SIF and the reliability of this dataset has been verified. We acquired the ungridded L2B SIF dataset at daily scale covering the period from May 2018 to December 2020, which has passed routine data screening, i.e., removing the data over water bodies, with a cloud fraction (hereafter named CF) larger than 0.8, and a quality assurance (QA) value larger than 0.5 (referring to [24] for details). This dataset included SIF retrievals from two fitting windows (743–758 nm and 735–758 nm). According to [24], the 735–758 nm SIF retrievals still have difficulties in dealing with spectrally-steep radiance spectra, and the 743–758 nm window has the best compromise between effects of clouds and retrieval precision errors. Thus, the 743–758 nm SIF was selected in this work. Further, we adopted the day-length corrected SIF file based on cosine of the solar zenith angle ($\cos(\text{SZA})$) [4], which was also provided in the L2B datasets.

The raw L2B SIF dataset was aggregated to gridded maps with data screening at different spatiotemporal scales in this work. First, for model development, we conducted data screening to reduce the uncertainties in SIF retrievals due to clouds and noises in individual SIF soundings. The screening strategies included: (i) only SIF soundings acquired under a CF less than 0.2 were remained for the averaging of SIF grids (named 0.2-CF screening in this study); and (ii) only SIF grids that averaged from more than 5 SIF soundings (named 5-N screening in this study) were involved in the model development, given that the uncertainty in SIF retrievals can be reduced by a factor $\sqrt{n}$, in which n is the number of observations in each aggregated grid [19]. On this basis, we aggregated the SIF soundings to 0.1° and 8-day scales with 0.2-CF and 5-N screening (named 8-day TROSIF⁰¹) for model development (Section 2.2.2). The 0.1° resolution was selected since it was comparable to the size of TROPOMI footprint that was often larger than 0.05° when away from the nadir (see Figure 1). In this case, the 8-day interval was a compromise between temporal details and number of SIF samples. Second, as reference products for SDSIF validation, a 16-day TROSIF⁰¹ was used to provide spatial details, and a daily 0.2° SIF with 0.2-CF and 5-N screening (named daily TROSIF⁰²) was used to provide temporal details. Third, the 0.05°

gridded SIF with none 0.2-CF and 5-N screening (TROSIF⁰⁰⁵) was generated as a contrast to verify the advantages of SDSIF in correlations with tower-based SIF and spatial details.

2.1.2. MODIS and ERA5 Datasets

Part of the potential explanatory variables (e.g., reflectance and VIs) involved in the SIF reconstructed models were obtained from the MODIS product. Thereinto, the reflectance datasets were derived from the MCD43C4 v006 product, which provided the BRDF-corrected seven-band reflectance at 0.05°, daily resolution from 2000 to present [36]. In order to ensure the quality of the training and testing samples, the reflectance with the QA flag less than 3 was discarded. The VIs maps, including the Normalized Difference Water Index (NDWI), Near-Infrared Radiance of vegetation (NIRv), Normalized Difference Vegetation Index (NDVI), and the Enhanced Vegetation Index (EVI), were also calculated. The 0.05°, daily reflectance and VIs datasets were used for the global SIF prediction, and were resampled to 0.1°, 8-day resolution for the model development.

The meteorological variables possibly involved in SIF reconstruction, including air temperature (Ta), vapor pressure deficit (VPD), photosynthetically active radiation (PAR), and PAR under clear-sky conditions, were derived or calculated from the fifth ECMWF Reanalysis (ERA5) dataset. This dataset provides related variables at 0.1°, hourly resolutions from 1950 to present. Thereinto, the VPD was calculated using Ta and the dewpoint temperature (DTa) based on the formula proposed by [37]. The provided variables were aggregated to 8-day scale at 0.1° resolution for model development, and then resampled to 0.05° at daily scale for global SIF prediction.

In addition, the land cover products at 0.05° from MCD12C1 were used to evaluate the performance of our model at each biome. This dataset was based on the International Geosphere Biosphere Programme (IGBP) classification scheme [38] with 17 land cover types. We merged the 17 types into smaller classes based on the strategy in [4]. According to our testing results, the performance of models had differences between Closed Shrublands (CS) and Open Shrublands (OS), as well as between Evergreen Needleleaf Forest (ENF) and Deciduous Needleleaf Forests (DNF). Eventually, nine land cover types were divided across the globe, including ENF, DNF, CS, OS, Evergreen Broadleaf Forest (EBF), Deciduous Broadleaf Forests and Mixed Forests (DBF), Woody Savannas and Savannas (SAV), Grasslands (GRA) and Croplands and Cropland/Natural Vegetation Mosaic (CRO). Figure 2 displays the land cover map in 2019. The unvegetated types such as “Water”, “Permanent Wetland”, “Urban and Built-up”, “Snow and ice”, and “Barren or sparsely vegetated” were excluded in this study.

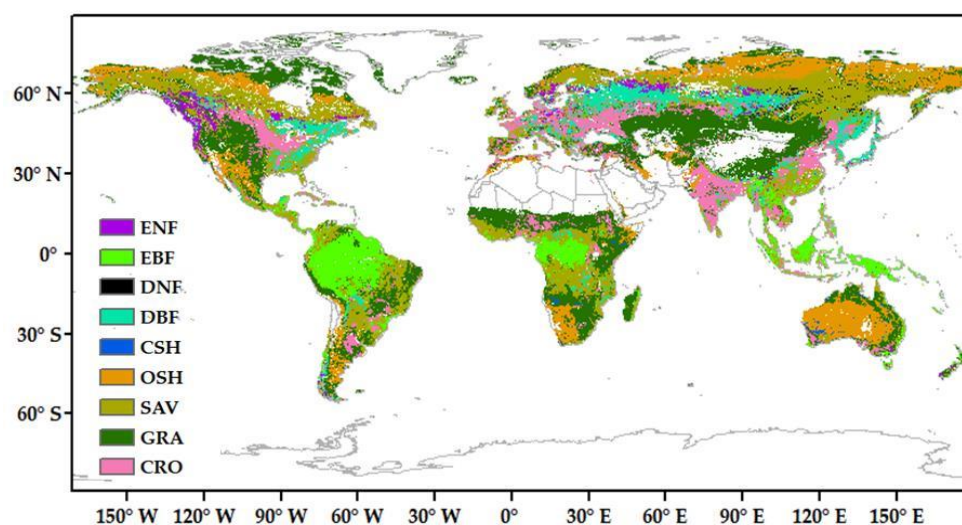


Figure 2. The land cover map in 2019 from MCD12C1.

2.1.3. Tower-Based Datasets

Long-term and continuous flux tower experiments can provide ideal references for the satellite SIF validation. In this study, we used tower-based datasets at five flux sites to validate the reliability of our reconstructed SDSIF. The details of five sites were shown in Table 1. Thereinto, four sites were components of the ChinaSpec (i.e., a network for long-term measurements of SIF and reflectance in China) [39], including the HL and GC sites in Hebei province at North China plain, the DM site in Daman, Gansu province at northwest China, and the AR site in Qinghai province at northwest China. In addition, we used the SIF retrievals at Aurora site in New York, northeast United State to supplement the validation datasets, which were obtained from the California Institute of Technology (<https://ecommons.cornell.edu/handle/1813/69711>, accessed on 21 January 2022) [40]. The five sites were predominated by different vegetation types, i.e., single-cropping maize at HL, DM, and Aurora sites, rotation cultivation of winter wheat and maize at GC site, and the Alpine meadow at AR site. All sites were located with homogeneous underlying surface, so that they were representative to compare with the 0.05° satellite cells. Periods for data collection cover most of the vegetation growing season at five sites. In spite of power outage conditions, we collected a total of 95, 213, 234, 74, and 78 days of valid datasets for the five sites in Table 1, respectively. More detailed information about the site conditions, measurement system, and data process can be found in [39,40] and our previous studies [26,41,42].

Table 1. Details of the flux tower sites.

Land Cover Type	Site Name	ID	Latitude	Longitude	Period	Height
CRO	HuaiLai	HL	40.3489°N	115.7882°E	May to October in 2018	4 m
	DaMan	DM	38.8555°N	100.3722°E	June to October in 2018 & 2019	25 m
	GuCheng	GC	39.1487°N	115.7350°E	May to December in 2020	25 m
	Aurora	-	42.7228°N	76.6628°W	July to October in 2018	7 m
GRA	Arou	AR	38.0473°N	100.4643°E	June to September in 2019	25 m

The SIF values were retrieved using the hyperspectral down-welling and up-welling radiances from automatically observation system [43,44]. Before the process of retrieving SIF, an atmospheric correction method proposed in [45] was conducted for the HL, DM, GC, and AR sites with relatively large tower height. The SIF retrievals at 760 nm were based on three-band Fraunhofer Line Depth (3FLD) [46] and Singular Value Decomposition (SVD) Method [47]. In order to compare with the TROPOMI satellite SIF at around 740 nm, we converted the 760 nm tower-based SIF into 740 nm by multiplying a wavelength scaling coefficient of 1.5 based on the literatures [48,49]. The SIF retrievals at several-minute intervals were first averaged into half-hourly values and then to daily values for comparison with the daily SDSIF. Furthermore, we also averaged the daily SIF values into 4-day values if there were four daily values in a 4-day period. At the 4-day scale, due to the limited number of valid days at HL, AR, and Aurora, we only showed the results at GC and DM sites. In addition, we calculated the half-hourly GPP data at DM site using the meteorological data from an automatic weather station (AWS) [50] and the flux data from an eddy covariance (EC) system. The process included the gap filling [51] and day-time partitioning method [52], which was integrated on the online tool available on the Max Planck Institute for Biogeochemistry (MPI-BGC) website. Similar to the SIF calculation, we averaged the half-hourly GPP into daily scales, and also calculated the 4-day GPP averages if there were four daily values in a 4-day period.

2.2. Data-Driven Method for SIF Reconstruction

2.2.1. Explanatory Variable Selection

Similar to the LUE concept for estimating GPP, SIF can be expressed as follows [2,10,53]:

$$\text{SIF} = \text{PAR} \times \text{fPAR} \times \varepsilon \times \Phi_{\text{SIF}} \quad (1)$$

where fPAR is the fraction of PAR absorbed by vegetation canopies, ε represents the fraction of SIF photons escaping from the photosystem level to the canopy level, and Φ_{SIF} is fluorescence quantum yield.

The four terms on the right-hand side of this equation can be further interpreted to available factors as follows. First, since PAR is linearly correlated with $\cos(\text{SZA})$ under clear-sky conditions [4,54], we used $\cos(\text{SZA})$ as the proxy for clear-sky incoming PAR in the process of model development. Second, fPAR is mainly related to the leaf optics and canopy structure, of which variations can be mostly denoted by the remote-sensed reflectance. Many previous studies have demonstrated that fPAR can be quantified by the vegetation indices such as NDVI and EVI [42,55–57]. Third, recent studies have provided the mechanistic equation for ε based on the spectral invariant theory [58–61], in which the ε at the NIR band can be expressed as a function of several variables including the NIR reflectance, the canopy structures (leaf area index and the leaf inclination distribution), and the $\cos(\text{SZA})$. Liu et al., (2020) [42] provided a simplification of this expression in which the ε at the NIR band can be approximately calculated by NDVI multiplied with the NIR band reflectance (i.e., the NIRv) and divided by fPAR. In many previous studies of satellite SIF downscaling or gap-filling, the seven or first four MODIS reflectance bands were used to express the optical and structural information in ε and fPAR [28–34]. Further, Φ_{SIF} was mostly affected by the plant's physiology that governed by the plant species and the environmental factors such as sunlight, temperature, and water [62].

Based on above evidences, several variables were selected as the potential explanatory variables for SIF reconstructed models in this work, including $\cos(\text{SZA})$, Ta, VPD, NDWI, seven bands of MODIS reflectance, and NIRv. The first four variables were used to denote the environmental effects on Φ_{SIF} in SIF variations. Seven-band reflectance and NIRv were used to express the leaf optical and canopy structure information involved in ε and fPAR for SIF variations. Other vegetation indices were also tested in the model but no improved effects for SIF predictions had been observed.

According to the accuracies for SIF reconstruction models, we explored how the potential variables affected the model results and thus determined the final explanatory variables. Apart from the necessary variable $\cos(\text{SZA})$, we investigated two issues in the selection of variables: (i) whether and how the environmental factors Ta, VPD, and NDWI affected the performance of models; (ii) what was the difference in results between using different variable combinations to denote the vegetation optical and structural information (i.e., the single NIRv, 4-band reflectance, 7-band reflectance, or the first two of seven bands substituted with NIRv). Eventually, the model accuracies for eight combinations of explanatory variables (as listed in Figure 3) were evaluated by three statistical measurements (i.e., the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE)) based on the testing samples in 2019. According to the statistical metrics of model accuracy, we determined the final 8 explanatory variables (i.e., $\cos(\text{SZA})$, the first four reflectance bands, Ta, VPD, and NDWI) for further model development.

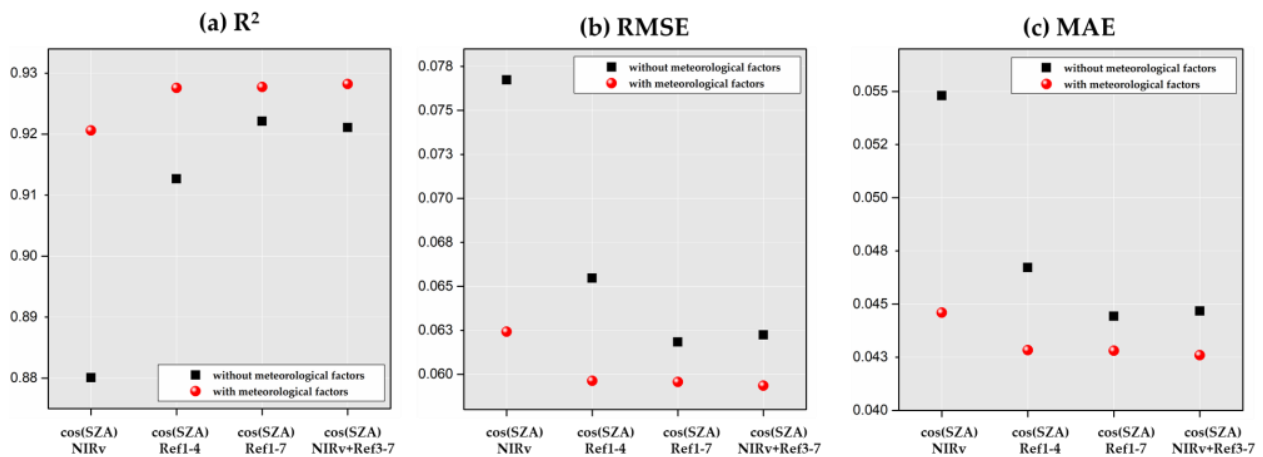


Figure 3. The statistical metrics for the accuracy of SIF reconstruction models using different combinations of explanatory variables based on the testing samples at 0.1° , 8-day resolutions in 2019. (a) coefficient of determination (R^2); (b) Root Mean Square Error (RMSE, $\text{mW}/\text{m}^2/\text{nm}/\text{sr}$); (c) Mean Absolute Error (MAE, $\text{mW}/\text{m}^2/\text{nm}/\text{sr}$). Ref1–4 and Ref1–7 refer to MODIS bands 1–4 and MODIS bands 1–7, respectively.

2.2.2. Model Development

The key process of SIF reconstruction was to establish the relationship between clear-sky satellite SIF and the selected explanatory variables as follows:

$$\text{SIF} = F(\cos(\text{SZA}), \text{Ref1-4}, T_a, \text{VPD}, \text{NDWI}) \quad (2)$$

where F is the model representing the relationship between clear-sky SIF and explanatory variables. Ref1–4 denote the first four bands of MODIS reflectance.

In this study, we used a random forest (RF) approach to establish the model F using the samples derived from the aggregated 8-day TROSIF_s^{01} and explanatory variables. RF is a tree-based model first proposed by [63] and has been widely used in remote sensing application. It has been demonstrated that RF has outstanding performance in regression tasks for large and multi-dimensional datasets [64,65]. For generating spatially contiguous high-resolution SIF product, RF has also been tested and found to be efficient in previous studies [30,33]. Since RF contains many decision trees, and for each tree the samples and features are completely random, RF is less sensitive to overfitting [66] than several other machine learning methods and relatively robust to the noises in input datasets. Moreover, an RF model has better physical interpretability than several other ML algorithms such as neural networks, which can better resolve the explicit physical relationship between the predictor (SIF) and the explanatory variables used in this study.

The whole samples for model development were divided into the training and testing samples based on the strategy similar to a three-fold cross validation. Specifically, each of the three experiments individually selected 30% of the whole samples for testing and the remaining 70% for training. The testing samples of each experiment did not overlap with each other. Both the training and testing samples were normalized by the averages and standardized deviation of the training data. For each experiment, the RF built multiple decision trees, each of which is a tree-like model with multiple nodes. Then, the training samples were divided into different subsets using a bootstrapping method [67] (i.e., selecting random samples from the whole datasets repeatedly). Each training subset was segmented at each node using a random subset of features (i.e., $\cos(\text{SZA})$, the first four MODIS reflectance bands, T_a , VPD, and NDWI) through Gini index, information gain or other methods to construct the splitting rules. Each decision tree grew up based on random subsets of both training samples and features. The final SIF results were obtained from all trees by averaging. In this study, we set 100 trees and five minimum leaf size (i.e., the

minimum number of samples in each training subsets used to split the decision tree at each node) as the RF parameters, which was determined by comparison experiments about the prediction accuracy and computing time.

Three statistical metrics (i.e., R^2 , RMSE, and MAE) were used to evaluate the performance of the model for testing samples. We compared the statistical metrics both at each biome and all biomes. Moreover, we compared the performances of model testing with different strategies of model constraints. Compared with the one single “universal” model across the globe for one year that was used in [31,33], we first added the spatial constraints for specific continents (i.e., continent-specific model), in view of the similar response of SIF to variables within closer space. Five continents were divided across the globe, including North America, South America, Africa, Oceania, and the combination of Asia and Europe. In this case, five models were obtained in one year. Then, we further added the temporal constraints for specific months (i.e., continent- and monthly-specific model), considering the similar response of SIF to variables within closer time. As a result, 60 models for all of five continents and 12 months were constructed in one year. Additionally, we also compared the performance between the continent-specific and the biome-specific models. The two models had almost equal performance with testing samples, but the biome-specific one produced unnatural boundaries at the junction of different biomes in reconstructed SIF products. Thus, the biome-specific model was not employed in this work.

2.2.3. Global-Scale SIF Reconstruction

Based on the established models with spatial and temporal constraints, we first predicted the 0.05° , daily seamless global SIF at clear-sky conditions (i.e., $\text{SDSIF}_{\text{clear-daily}}$) using eight explanatory variables at 0.05° , with daily resolution. The explanatory variables were normalized by the averages and standardized deviation of the training data. In order to tackle the temporal mismatch between the clear-sky SIF and the all-sky GPP, similar to the previous studies [26,31,68], we used a temporal upscaling factor (i.e., PAR) to convert the SDSIF from the clear-sky daily scale to all-sky daily scale ($\text{SDSIF}_{\text{all-daily}}$). Specifically, since the diurnal variations in SIF was mainly governed by the PAR [9], and many in-situ experiments had verified an approximately linear relationship between SIF and PAR at the diurnal scale [53,57], the $\text{SDSIF}_{\text{all-daily}}$ can be expressed as:

$$\text{SDSIF}_{\text{all-daily}} = \frac{\text{SDSIF}_{\text{clear-daily}}}{\text{PAR}_{\text{clear-daily}}} \times \text{PAR}_{\text{all-daily}} \quad (3)$$

where $\text{PAR}_{\text{clear-daily}}$ and $\text{PAR}_{\text{all-daily}}$ represent the daily PAR averages assuming clear-sky conditions and under natural situation, respectively, which are both derived from the ERA5 dataset. Thus, the final global 0.05° SDSIF product were reconstructed over the period May 2018 to December 2020.

2.3. Validation Approaches

The validation of SDSIF was conducted in three ways. First, we validated the capability of reconstructed SIF for preserving the temporal and spatial characteristics of original TROPOMI SIF, similar to the validation strategy in several studies [28,29,31]. On one hand, by using the pixel-wise R^2 and slope of linear regressions as the statistical metrics, we calculated the temporal consistency between the re-aggregated clear-sky SDSIF and the daily TROSIF_s⁰² to verify the preservation of day-to-day variations in original SIF. On the other hand, we calculated residuals between the re-aggregated clear-sky SDSIF and the 16-day TROSIF_s⁰¹ as well as the latitudinal averages of both two products to assess the capability of SDSIF for the preservation of spatial characteristics in original SIF. The correlations between these two products were assessed using the performance of linear regressions. Second, we used the continuous tower-based SIF retrievals at five sites to assess the reliability of SDSIF and its advantage over the original TROSIF⁰⁰⁵ in temporal continuity and consistency at daily and 4-day scales. Third, we explored the advantage

of SDSIF over TROSIF⁰⁰⁵ with different cloud fractions in spatial continuity and details at global and regional scales.

3. Results

3.1. Performance of the SIF Reconstruction Models

To determine the explanatory variables involved in modeling, we explored how the model accuracy was affected by the selection of explanatory variables. Figure 3 shows the statistical metrics of model accuracy for eight combinations of explanatory variables based on the testing samples in 2019. On one hand, adding the meteorological factors to the explanatory variables can significantly improve the performance of model with higher R^2 , lower RMSE, and MAE values. On the other hand, for vegetation optical and structural related variables, the single use of NIR_v (the first case in axis) produced the lowest accuracy, whereas using seven reflectance bands or the first four bands instead of NIR_v (the second and third cases in axis) performed better, both with and without meteorological factors. In addition, substituting the first two of the seven bands with NIR_v (the fourth case in axis) did not result in an obvious improvement in the statistical metrics, probably because they contained similar information. Note that the performances of the last three cases had no significant differences with a small range of R^2 (0.927–0.929), RMSE (0.059–0.060), and MAE (0.042–0.043) when adding the meteorological factors. Therefore, to reduce the complexity of the model features, we selected the second case (eight variables including cos(SZA), the first four reflectance bands, Ta, VPD, and NDWI) as the final explanatory variables for further model development and analysis in this work.

We further evaluated the performances of models with different strategies of model constraints. Table 2 shows the statistical results for model testing in 2019 with three strategies of model constraints: a single “universal” model, continent-specific model, and continent and monthly-specific model. It can be seen that the continent- and monthly-specific models produced the best results, probably because it considered the similar response of SIF to environmental variables within closer space and time. More specifically, the model accuracy for all biomes had been increased by adding the spatial constraints, and further improved by adding the temporal constraints for specific months. Consequently, in this work, we developed 60 models for all of five continents and 12 months in one year for reconstructing the corresponding SIF values across space and time.

Table 2. The statistical metrics for the accuracy of SIF reconstruction models with three strategies of model constraints based on the testing samples at 0.1°, 8-day resolutions in 2019.

Biome	Universal Model			Continent-Specific Model			Continent- and Monthly-Specific Model		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
ENF	0.804	0.0681	0.0502	0.819	0.0652	0.0481	0.829	0.0632	0.0463
EBF	0.725	0.0851	0.0636	0.755	0.0800	0.0596	0.778	0.0760	0.0563
DNF	0.886	0.0654	0.0486	0.889	0.0640	0.0476	0.892	0.0631	0.0468
DBF	0.928	0.0735	0.0533	0.933	0.0709	0.0512	0.938	0.0685	0.0491
CSH	0.864	0.0464	0.0329	0.879	0.0440	0.0309	0.886	0.0420	0.0296
OSH	0.775	0.0491	0.0356	0.793	0.0470	0.0340	0.807	0.0454	0.0328
SAV	0.892	0.0702	0.0517	0.902	0.0666	0.0488	0.911	0.0635	0.0464
GRA	0.883	0.0577	0.0417	0.892	0.0554	0.0400	0.899	0.0535	0.0385
CRO	0.937	0.0678	0.0493	0.943	0.0643	0.0468	0.948	0.0610	0.0441
All	0.913	0.0653	0.0472	0.921	0.0622	0.0449	0.928	0.0596	0.0428

Our selected models performed pretty well for reconstructing SIF with testing samples at each of nine biomes, with the R^2 higher than 0.8, and the RMSE lower than 0.07 mW/m²/nm/sr (except for the EBF type). The accuracies for EBF and OSH types were lower than other biomes, which was probably because (i) the EBF was mainly distributed in tropical rain forest area where the seasonal variations of meteorology were relatively

weak, so that the SIF's responses to driven variables were more complicated to be modeled; and (ii) the SIF emissions for the OSH type were weak throughout the whole year which caused higher levels of noise in the original SIF products. The scatterplots between the TROPOMI SIF and predicted SIF for the testing samples of three cross-validation experiments in 2019 were shown in Figure 4. We can see that the predicted SIF values were highly consistent with the original ones: the scatters were both distributed closely to the 1:1 line and produced satisfactory accuracy (the averaged R^2 of three experiments is around 0.928, and the averaged RMSE is around $0.0597 \text{ mW/m}^2/\text{nm/sr}$).

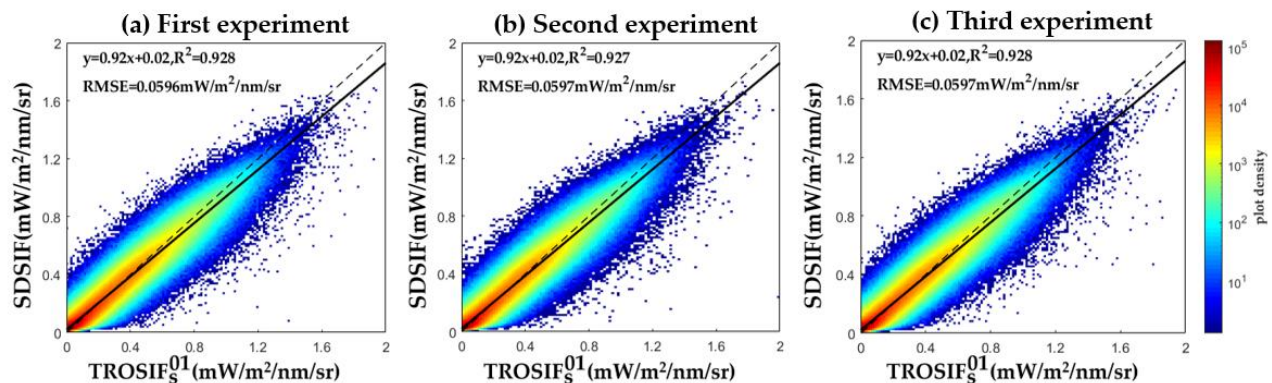


Figure 4. Scatter diagrams between the TROPOMI SIF and the SIF predicted by RF models for the testing samples of three cross-validation experiments: first (a), second (b), and third (c) at 0.1° , 8-day resolutions in 2019. The density of points in logarithmic scale is represented by the colorbar. The black dash line represents the 1:1 line.

3.2. Validation of SDSIF with Original TROPOMI SIF

We validated the capability of reconstructed SIF for preserving day-to-day variations in original SIF. The pixel-wise R^2 and slope values of the linear regressions with intercept between the re-aggregated daily clear-sky SDSIF and daily TROSIF_s^{02} during 2019 are displayed in Figure 5. Most pixels across the globe achieved the significance level of 0.05, the rest ones (only around 5% of all pixels) were discarded for our analysis. It can be observed that the temporal variations between SDSIF and daily TROSIF_s^{02} was highly consistent at most regions: the mean values of regression slope and R^2 were around 0.750 and 0.760 with 80% of all pixels, respectively. Referring to the Landcover map in Figure 2, relatively weaker consistencies occurred in two cases: the EBF in tropical rainforests with weak seasonality (including Amazon, Indonesia and Congo areas) and the OSH in central Australia, South Africa and Argentina, southwestern North America and North Asia, with the mean R^2 and slope values of 0.415 and 0.424, respectively. This phenomenon agreed with the model accuracies shown in Table 2 and the results in previous studies [32].

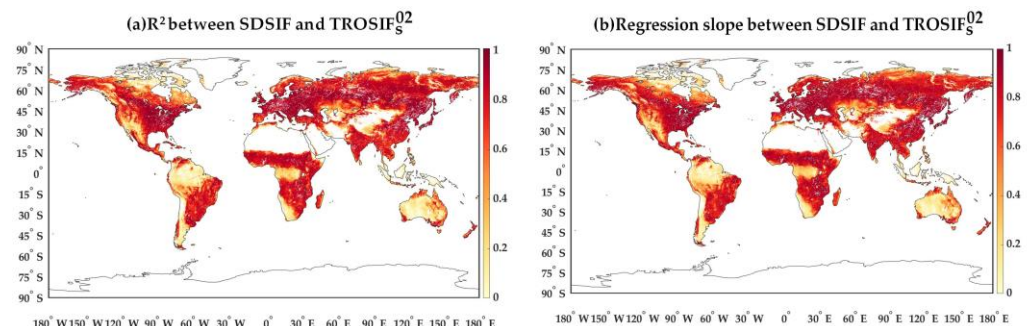


Figure 5. The pixel-wise correlations between day-to-day SIF values from SDSIF and TROSIF_s^{02} in 2019 at 0.2° , daily scales in terms of the coefficient of determination (R^2) (a) and regression slope (b). All pixels in this figure achieved the significance level of 0.05.

On the other hand, we validated the capability of reconstructed SIF for preserving spatial variations in original SIF. Figure 6 displays the re-aggregated SDSIF product (the left column) and its residuals with 16-day TROSIF_s⁰¹ (i.e., the difference between the former and the latter, the middle column) for the first 16 days of four months in 2019. In general, SDSIF values were highly consistent with the original ones: a majority of pixels (around 80%) exhibit minimal residuals with the extent lower than $\pm 0.081 \text{ mW/m}^2/\text{nm/sr}$ for all months (see red numbers in the middle column). In the northern hemisphere, the temporal variations in the absolute value of residuals kept pace with the seasonal variation of SIF magnitude. For example, the absolute values of residuals for GRA, DBF, and CRO biomes in North America were approximately zeros at the start or end of the growing season (January or October) and increased with the growing of vegetation, reaching their highest levels in July. For the tropical rainforest areas (Amazon, Indonesia, and Congo), relatively large values were exhibited during the whole year due to the high SIF signal from long-active rainforest. We can see low absolute value of residuals for the OSH types in central Australia, South Africa and Argentina, southwestern North America, and North Asia, owing to the low SIF magnitude throughout the year. The latitudinal averages of both products (the right column) also showed that the re-aggregated SDSIF can well preserve the latitudinal variations of the original TROPOMI SIF.

Additionally, the quantitative assessment of the consistency between these two products (Figure 7) further verified that the SDSIF can well preserve the spatial variability of the original SIF, with high spatial correlations between them two: the scatters fell on closely to the 1:1 line with the R^2 larger than 0.890 and the RMSE less than $0.072 \text{ mW/m}^2/\text{nm/sr}$ for all four months in 2019.

3.3. Validation of SDSIF with Tower-Based SIF

The comparison between the time series of tower-based SIF and the two satellite SIF products (SDSIF and original TROSIF⁰⁰⁵) at 0.05° , daily scales was shown in Figure 8. The linear regression between satellite SIF and tower-based SIF was also presented. All regressions in this figure achieved the significance level of 0.05. For better comparison, the same number of scatters were remained for the regressions of two products both in Figures 8 and 9. In general, the SDSIF (red curves), TROSIF⁰⁰⁵ (blue curves), and tower-based SIF (gray curves) gave similar seasonal trend with the growing of vegetation at five sites. Thereinto, it was obviously that SDSIF had more continuous daily observations, whereas the original TROSIF⁰⁰⁵ was highly sparse and noisy due to its swath gaps, clouds contaminations, and individual sampling errors. Moreover, SDSIF can greatly capture the continuous tower-based SIF, which gave better consistencies than the daily TROSIF⁰⁰⁵ did at all five sites. Specifically, the R^2 values for SDSIF (red texts in the right panel) was all larger than 0.7 except for the Aurora, which was much higher than those for daily TROSIF⁰⁰⁵ (blue texts in the right panel). Further, the scatters for SDSIF (red dots) distributed closer to the 1:1 line than those for daily TROSIF⁰⁰⁵ (blue dots), except for the GC site where the overestimated noises in TROSIF⁰⁰⁵ coincidentally reduced the differences between satellite and tower-based SIF.

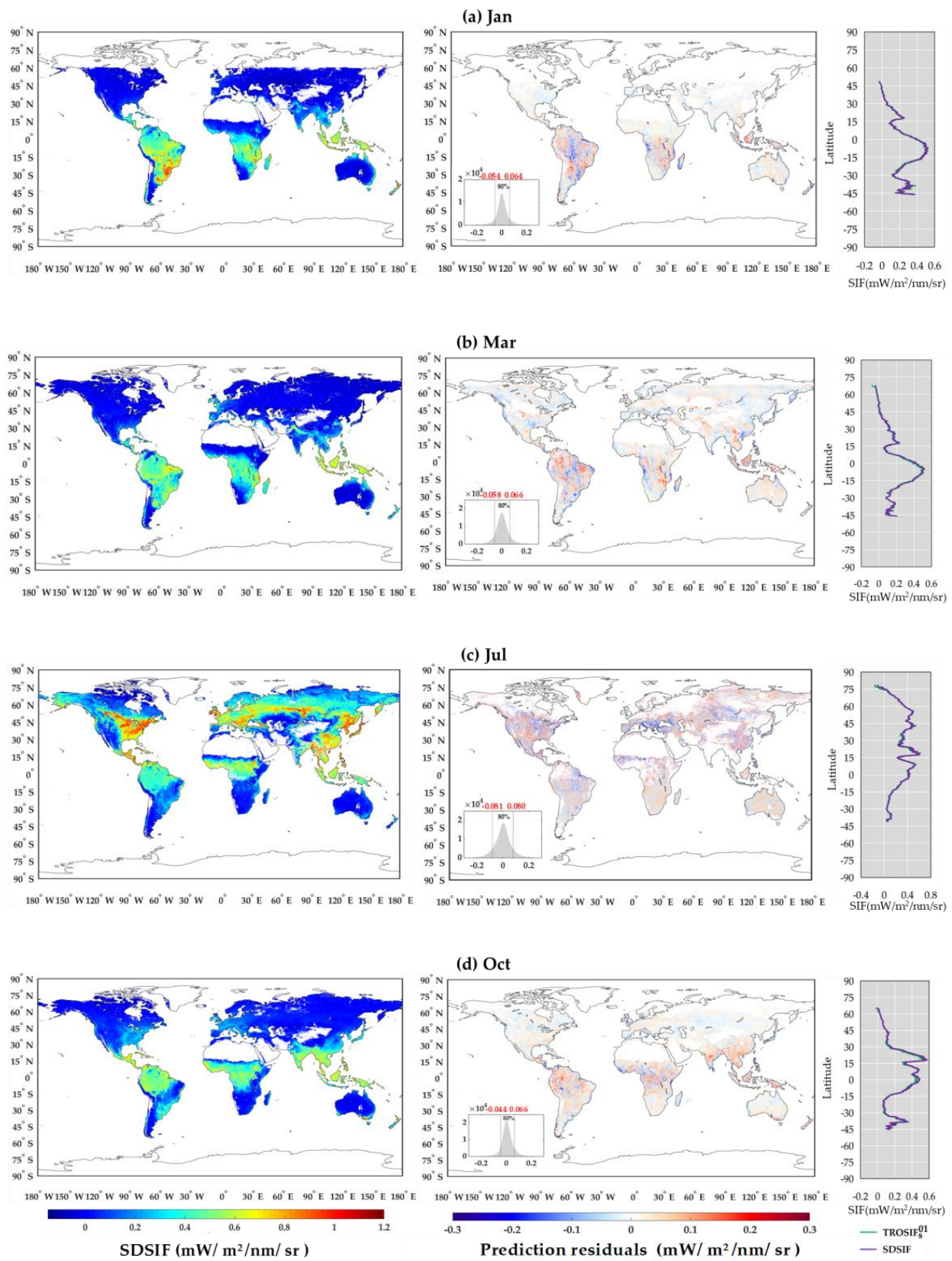


Figure 6. Spatial patterns of the 16-day, 0.1° re-aggregated SDSIF product (leaf column), as well as its residuals (middle column) and latitudinal averages (right column) compared with the original 16-day TROSIFs⁰¹ in January (a), March (b), July (c), and October (d) 2019. For each month, the first 16-day maps are shown here.

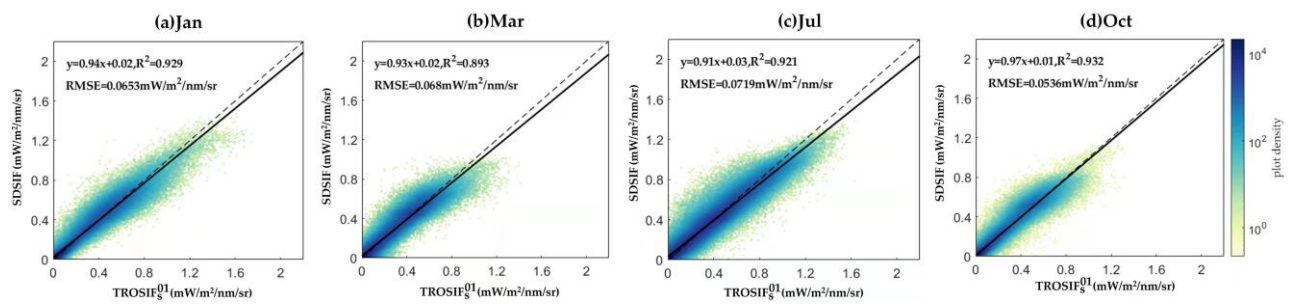


Figure 7. Scatter diagrams between the re-aggregated SDSIF and the original TROSIF⁰¹ at 16-day, 0.1° scales for the first 16 days in January (a), March (b), July (c), and October (d) 2019. The density of points in logarithmic scale is represented by the colorbar. The black dash line represents the 1:1 line.

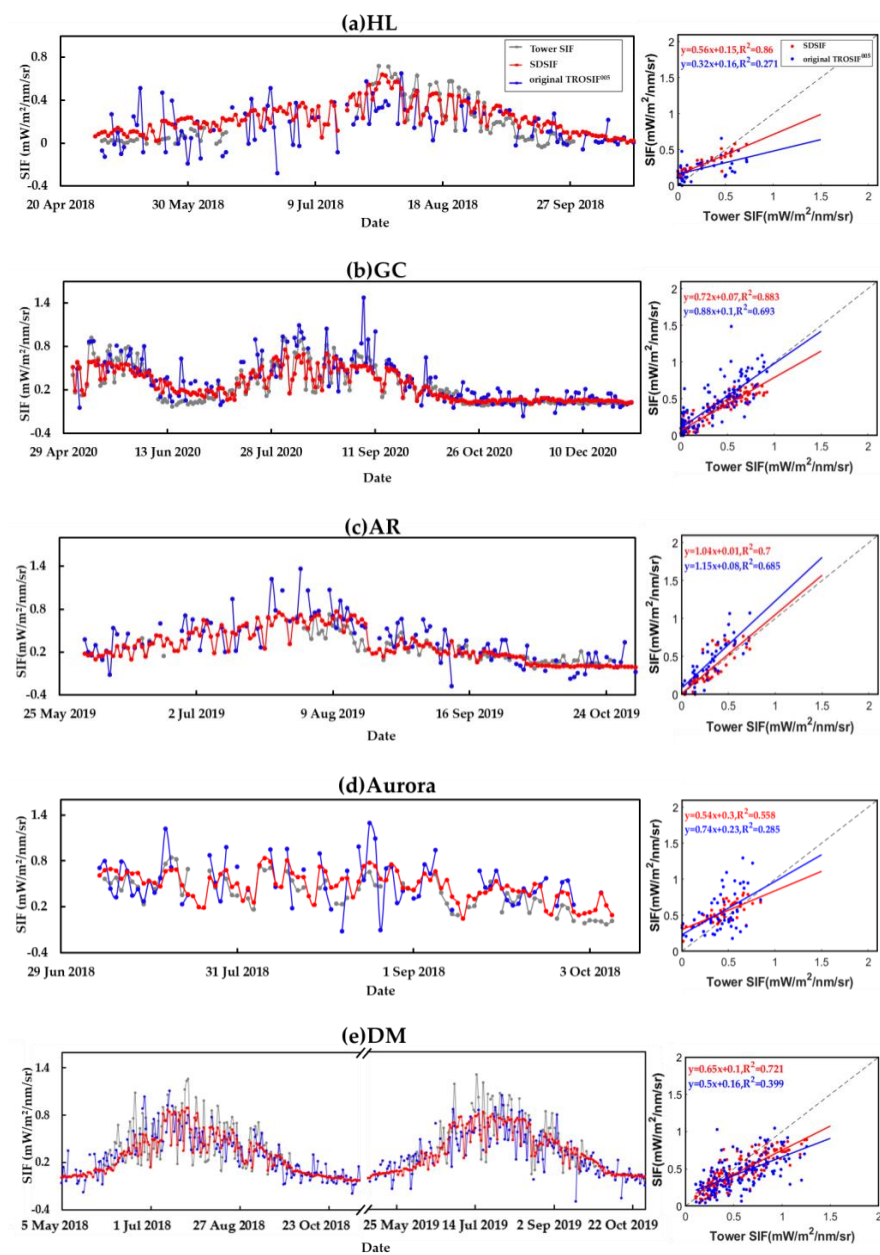


Figure 8. Comparison between the time series of tower-based SIF and the two satellite SIF products (SDSIF and original TROSIF⁰⁰⁵) at daily scale for (a–e) sites. All regressions in the right panel achieved the significance level of 0.05.

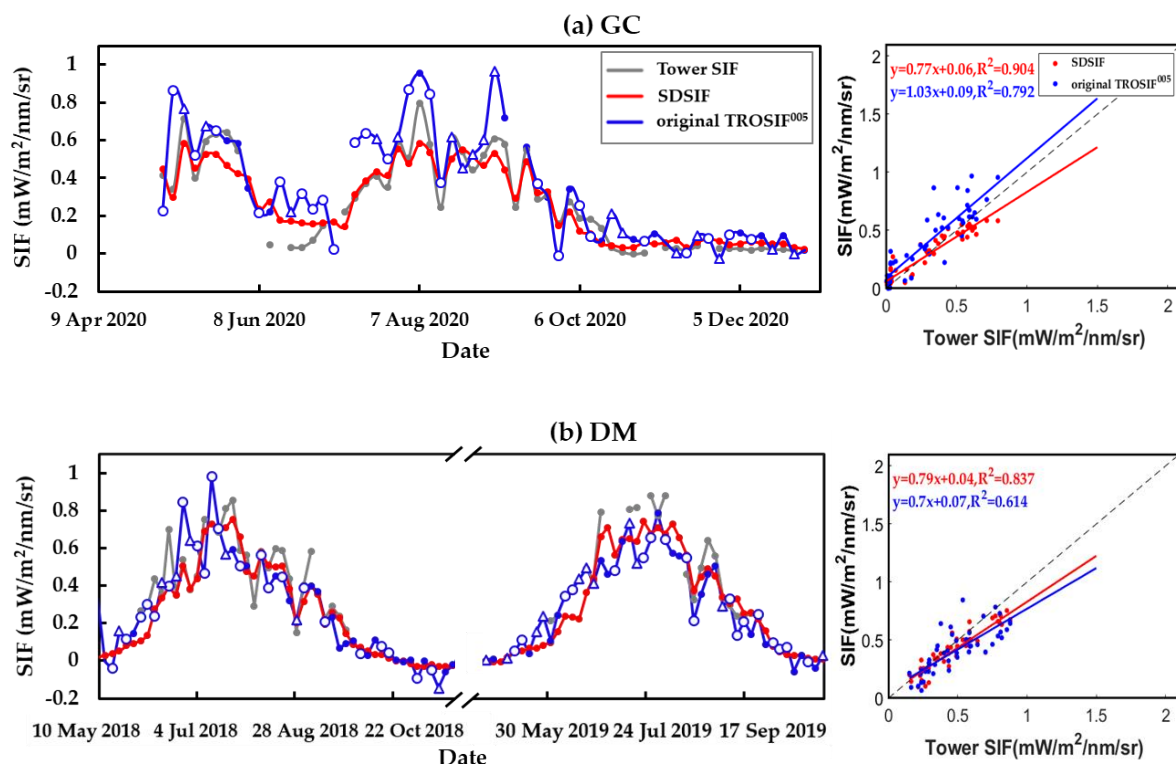


Figure 9. Comparison between the time series of tower-based SIF and two satellite SIF products (SDSIF and original TROSIF⁰⁰⁵) at the 4-day scale for (a,b) sites. All regressions in the right panel achieved the significance level of 0.05. The blue hollow dots, hollow triangles and solid dots represent the 4-day averages with valid observations from no more than one or two days, three days and four days, respectively.

To investigate the performance of SDSIF at larger temporal scales, we conducted the similar comparison at the 4-day scale, as shown in Figure 9. Due to the limited number of valid days at HL, AR and Aurora, we only displayed the results at GC and DM sites. We can see that, although the original TROSIF⁰⁰⁵ was relatively continuous at the 4-day scale, it still exhibited lower consistency with the tower-based SIF than the SDSIF does: the R^2 value for SDSIF was much larger than that for TROSIF⁰⁰⁵ at the 4-day scale. One reason was that the 4-day TROSIF⁰⁰⁵ averages still had errors from the noisy satellite observations at daily scale (as shown in Figure 8). More importantly, since SIF was highly changing over time, the 4-day averages from sparse daily TROSIF⁰⁰⁵ values cannot represent the real SIF at this time scale. As shown in the blue curves, most 4-day TROSIF⁰⁰⁵ averages had only valid observations no more than one or two days (hollow dots). However, SDSIF filled the temporal gaps of the daily TROSIF⁰⁰⁵, which provided the real 4-day SIF and thus improved the linear relationship with 4-day tower-based SIF. Overall, SDSIF had advantages over the original TROSIF⁰⁰⁵ in both temporal continuity and temporal consistency with in-situ SIF at the daily and 4-day scales.

3.4. Spatial Patterns of the Global SIF Product

The spatial pattern of global annual averages and the 90 percentile of SDSIF in 2019 are displayed in Figure 10a,b. The high values of annual mean SDSIF were observed in tropical forests, such as Amazon, Indonesia and Congo, which was consistent with the patterns of OCO-2 SIF in [31]. Annual mean SIF characterized the regions that transit from dry to wet, such as the Sahel and the gradient of eastern-western United States. The 90th percentile of SDSIF, which represented the maximal productivity, exhibited hotspots in the U.S. Corn Belt region, south Asia, center of Europe, and the tropical regions, consistent with high productive regions shown in [10]. We also compared the spatial patterns of

SDSIF with TROSIF⁰⁰⁵ at daily scale and found that SDSIF not only preserved the spatial pattern but also improved the spatial coverage of the original SIF retrievals. The results on 3 August 2019 is taken as an example (Figure 10c,d). These two products had the same spatial patterns. SDSIF filled the discontinuities in the original daily SIF due to swath gaps and clouds contaminations, and exhibited more continuous global coverage.

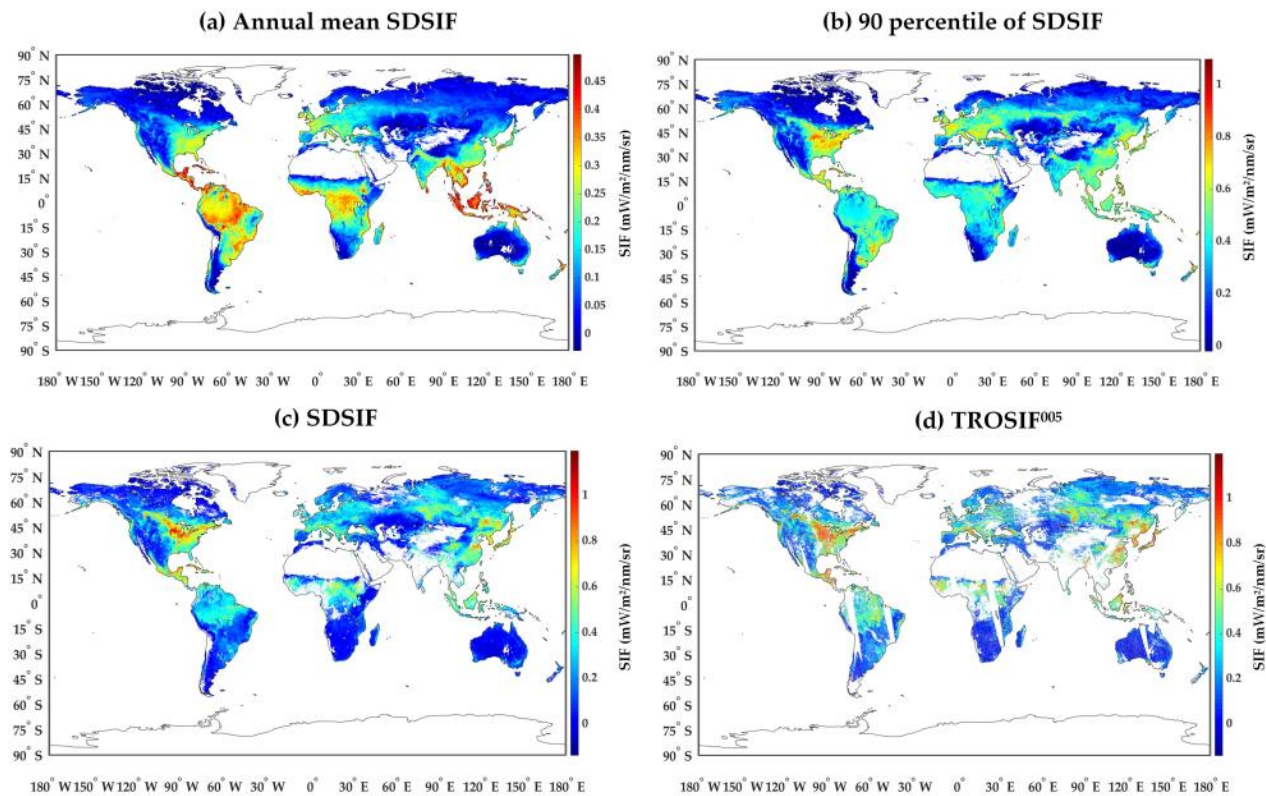


Figure 10. Spatial patterns of annual mean (a) and maximum (90th percentile) (b) of re-aggregated SDSIF in 2019, as well as the spatial comparison between SDSIF (c) and TROSIF⁰⁰⁵ (d) on 3 August 2019.

To further illustrate the advantages of SDSIF in spatial details, we display the comparison of SDSIF with the original daily TROSIF⁰⁰⁵ as well as the corresponding VI in a local enlarged image of the Mideastern United States region on 3 August 2019 (Figure 11). In general, the SDSIF well captured the spatial distributions of different biomes (Figure 11f), which had the consistent spatial patterns with the results in [32]. Specifically, in August 2019, the highest SDSIF values was observed for the CRO type, including the corn belt in the Mideastern regions and rice along the Mississippi river. While the DBF and SAV types showed moderate SIF magnitude and low SIF values were observed in OSH and GRA. Compared with the original daily TROSIF⁰⁰⁵, SDSIF had the similar SIF values and spatial variations. More importantly, it resolved the spatial problems of original daily TROSIF⁰⁰⁵ described in Figure 1, and thus exhibited finer spatial details at regional scales, such as the identification of the Mississippi river (black ellipse in Figure 11). In contrast, the daily TROSIF⁰⁰⁵ loss many spatial details, especially for the frequently-used clear-sky products with CF less than 0.2 (Figure 11c). For the comparison with VIs maps, SDSIF can clearly figure out the differences between agricultural regions with other biomes, while the spatial extent of the high-productivity regions was much oversize in NDVI and EVI, such as the DBF type in the eastern America (blue ellipse in Figure 11).

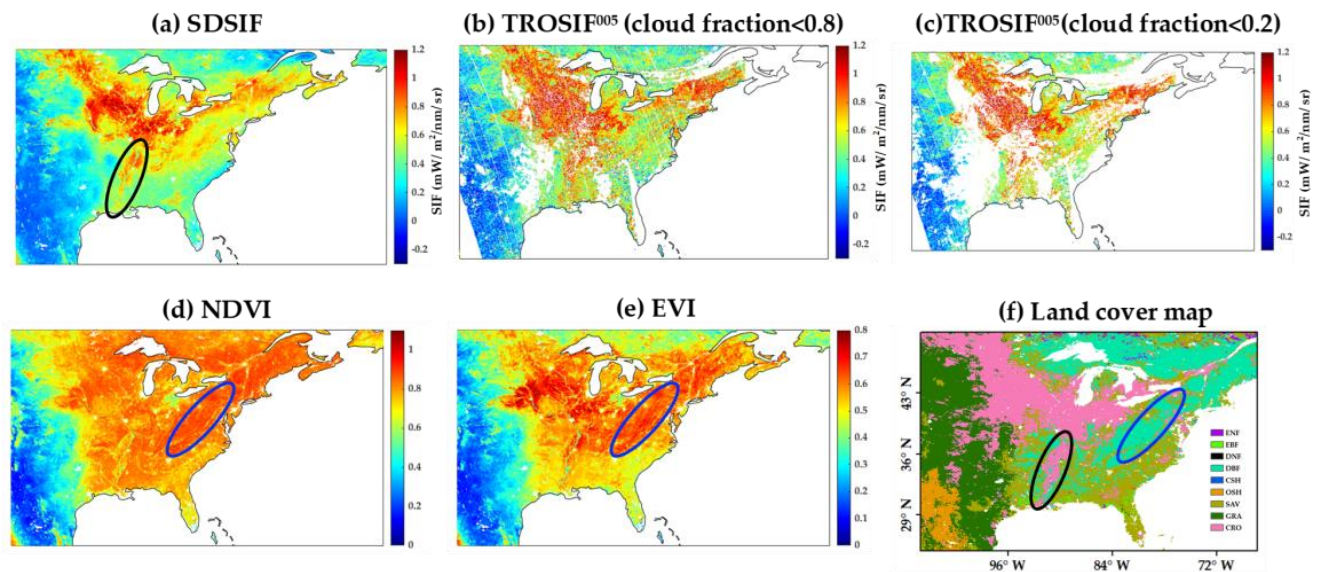


Figure 11. Local enlarged images of the Mideastern United States region on 3 August 2019 in terms of different products: (a–e). All maps are at 0.05° , daily resolution.

4. Discussions

4.1. Benefits of the Reconstructed SDSIF

The spatiotemporal limitations of the gridded TROPOMI SIF datasets at high spatiotemporal resolutions (e.g., 0.05° , 1-day or 2-day interval) have been described in Figure 1, including spatial insufficiency and spatiotemporal discontinuities. Our reconstructed SDSIF can tackle these drawbacks and benefit applications of satellite SIF product in carbon cycle and ecological studies.

The advantages of SDSIF over original TROPOMI product in terms of spatiotemporal enhancement can be summarized as follows. First, the spatial gaps due to swath gaps and clouds contaminations can be filled, and the spatial details away from the nadir have been improved (Figures 10 and 11). More importantly, the SDSIF largely improved the product's ability to track SIF temporal changes at daily scale by enhancing temporal continuity and reducing noises in original individual observations (Figure 8), and thus corrected the bias in multi-day SIF averages from sparse days (Figure 9). Consequently, it will tackle the temporal mismatch between original multi-day SIF averages and continuous GPP measurements, thereby reducing the accompanying uncertainties in SIF-GPP analysis or SIF-based GPP estimations. Particularly, for the cases in many studies that only clear-sky observations are remained for analysis, the averages of them cannot represent the real SIF to estimate the all-sky GPP [26,60]. Figure 12 displays a comparison between the time series of tower-based GPP and three satellite SIF products (SDSIF, clear-sky TROSIF⁰⁰⁵ (CF < 0.2), and cloud-contaminated TROSIF⁰⁰⁵ (CF < 0.8)) at 4-day scale for the DM site. For better visualization, the clear-sky TROSIF⁰⁰⁵ is multiply by 0.5 in Figure 12a to reach the same magnitude with other SIF products. It can be seen that the 4-day SDSIF exhibits higher temporal consistency and provides a stronger linear correlation with the 4-day GPP ($R^2 = 0.811$). Whereas, the correlation for two original TROPOMI products are much lower ($R^2 = 0.680$ and 0.701 , respectively) due to their fewer valid observations during a 4-day period. This advantage of SDSIF is caused by its reducing errors in original TROSIF⁰⁰⁵ and its enhancing temporal continuity for real 4-day SIF other than the 4-day averages from sparse daily TROSIF⁰⁰⁵ values.

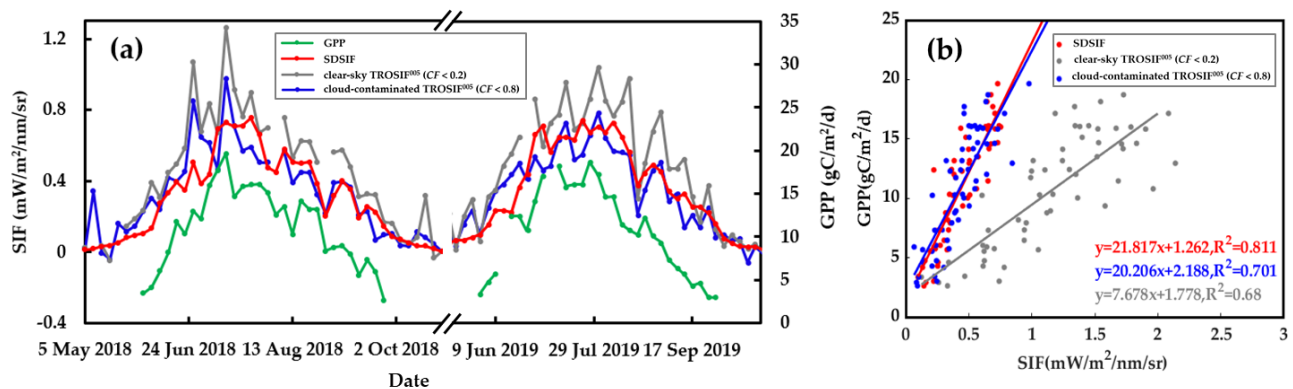


Figure 12. Comparison between the time series of tower-based GPP with SDSIF and original TROSIF⁰⁰⁵ (a), as well as the corresponding correlations (b) at the 4-day scale for the DM site.

Further opportunities are available to apply the SDSIF product as a proxy for global-scale GPP estimation and provide a new SIF-based GPP product into existing GPP datasets. The improvement of SDSIF in the relationships with tower-based GPP measurements need to be further comprehensively investigated based on more tower-based sites at various landcover types. In addition, SDSIF and the reconstructed method may also provide references for the generation of other daily seamless products, and bring new perspectives for our better understanding and application of the forthcoming higher-resolution SIF data (300 m) from the ESA's Earth Explorer Fluorescence Explorer (FLEX) mission.

4.2. Reliability and Uncertainties in SIF Reconstruction Method

In this study, a data-driven method using ML approach was designed to reconstruct a spatiotemporal enhanced TROPOMI product, by using $\cos(\text{SZA})$, MODIS reflectance, NDWI, T_a , and VPD at fine scale to indicate the variations in SIF at space and time. In theory, this strategy is reasonable since these explanatory variables governed mostly information in SIF, including leaf optical, canopy structure, and plant physiological parameters, which has been illustrated in Section 2.2 based on the LUE concept. Meanwhile, the similar data-driven approach has been affirmed by the downscaling of GOME-2 and gap-filling of OCO-2/Tansat in previous literatures [27–34]. Moreover, the performance of model testing in this work (Figure 4) is comparable with that in similar previous literatures, such as the RMSE of OCO-2 SIF predictions ranges from 0.065 to 0.08 mW/m²/nm/sr in literatures [31–33], and the RMSE of GOME-2 predictions is around 0.06–0.07 mW/m²/nm/sr over 12 months [30].

The strategy of using multiple models with continent- and monthly- constrains is also reasonable and reliable for SIF reconstruction. Compared with the one “universal” model strategy, it can better describe the different relationships between Φ_{SIF} and meteorological parameters (i.e., explanatory variables including PAR, T_a , VPD, and NDWI) among various regions and phenological stages with space and time. A universal model is more convenient for application but multiple models can achieve more accurate results for predictions (Table 2) due to more refined training samples for model development. If there is no systematic error in the training samples between different models, the usage of multiple models will not introduce uncertainties for SIF prediction since the models were completely data-driven and each model was trained using samples with large volume and sufficient representativeness. However, the usage of multiple models has a limitation that they can only be applied to periods and regions of which a certain number of original SIF samples is available for model training. The trained models cannot be extrapolated to other ranges with completely vacant original SIF.

The uncertainties involved in data sources of explanatory variables are parts of error propagated to the reconstructed SIF products. The first one comes from the original errors in TROPOMI SIF retrievals, even though the errors have been reduced by the averaging from five samples and the screening process by cloudy fractions (less than 0.2) before model

development in this work. The second one comes from the errors in 0.05° MODIS reflectance products, but it is relatively robust and smaller than the SIF noises. The third uncertainty results from the errors in ERA5 meteorological product, which includes original data errors in T_a and VPD for model development process and the errors from the interpolation of PAR data from 0.1° to 0.05° for SIF prediction. Moreover, high cloud density at tropical regions may also affect the model accuracy. For model development, we conduct data screening and only remained SIF soundings with 0.2-CF and 5-N. Due to high cloud density at tropical regions as South America, the number of the model training and testing samples in these regions was less than other regions after data screening, which maybe degrade the model accuracy. In addition, the number of observations in each grid was smaller, so the errors in gridded SIF samples might be less reduced by averaging than other regions. They are probably the reasons why the model accuracy for EBF which was mainly distributed in tropical rain forest was lower than other biomes in Table 2.

Despite the effects of data errors, the assumptions in SIF reconstructing approach also introduce some uncertainties. For the all-sky SIF prediction at daily scale, we assume that the diurnal SIF is mainly driven by diurnal PAR variations. First, the effects of diurnal variations in FPAR and the fraction of SIF escaping the canopy (i.e., ϵ) on the daily SIF are not considered, since the effects of these two variables on SIF diurnal variations is much smaller than that of PAR [9] and current satellite products cannot provide the daily variation information of these two parameters. Second, the effects of the diurnal variations in the fluorescence quantum yield (i.e., Φ_{SIF}) are neglected. This simplification can be supported by evidences from previous studies that the Φ_{SIF} has a weak diurnal variation with the light level and temperature due to the opposite trends in the fraction of open PSII reaction centers (qL) and non-photochemical quenching (NPQ) [62], and this effect are further weakened at the whole canopy scale [69].

More accurate semi-empirical, entirely machine learning, or knowledge-based ML models need to be developed for the spatiotemporal enhancement of SIF datasets in the future, like the generation of SIF datasets with daily seamless and sub-kilometer resolutions for finer applications at different spatial and temporal scales. Our future work will focus on better utilizing both the spatial and temporal response information of SIF to explanatory variables into the model development.

5. Conclusions

Due to the nonuniform sampling sizes, swath gaps, and cloud contamination, the TROPOMI SIF product has drawbacks in spatial insufficiency and spatiotemporal discontinuities when gridded at high spatiotemporal resolutions (e.g., 0.05° , 1-day or 2-day interval). In this study, a data-driven method based on RF was designed to reconstruct a spatiotemporal enhanced SIF product, by using MODIS reflectance, VIs, and ERA5 meteorological datasets as explanatory variables at a fine scale to indicate the variations in SIF over space and time. The reconstructed seamless daily product, namely SDSIF, can provided a global daily SIF with seamless spatiotemporal coverage at 0.05° resolution. Through the assessment of model accuracy with different explanatory variables and model constraints, the final 8 explanatory variables (i.e., $\cos(\text{SZA})$, the first four reflectance bands, T_a , VPD, and NDWI) and continent- and monthly-specific models were selected, which produced a fairly high accuracy ($R^2 = 0.928$, $\text{RMSE} = 0.0597 \text{ mW/m}^2/\text{nm/sr}$) and outperformed one universal model across the globe over a whole year.

Our validation results show that the SDSIF can well preserve the temporal and spatial variations in original TROPOMI SIF and reliably track the tower-based SIF retrievals at daily and 4-day scales. Compared with the original TROPOMI SIF, SDSIF had advantages in terms of the spatial details and the spatiotemporal continuities. It filled the spatial gaps in original product and exhibited more details at the regional scale like the Mideastern United States region. In addition, it improved the temporal continuities and reduced the individual noises in original product at daily scale, which resulted in higher consistency with the daily tower-based SIF at five sites (the mean R^2 increased from 0.467 to 0.744), and

thus corrected the bias in 4-day SIF averages from sparse daily observations with stronger correlations with the 4-day tower-based SIF and GPP. This study provided an effective strategy for SIF reconstruction with TROPOMI datasets and offer preferable SIF datasets for understanding SIF-GPP relationships and processes at finer spatial and temporal scales.

Author Contributions: Conceptualization, J.H.; Data curation, L.L.; Investigation, J.J.; Methodology, J.H. and J.J.; Resources, Y.M.; Software, Y.M.; Supervision, L.L.; Validation, J.J.; Visualization, J.J.; Writing—original draft, J.H.; Writing—review & editing, L.L. and H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the financial support provided by the National Natural Science Foundation of China (42001280, 41825002), and the China Postdoctoral Science Foundation (2021M690497).

Data Availability Statement: The reconstructed seamless daily and 4-day global SIF datasets at 0.05° resolution from 2018 to 2020 is free to access at <https://doi.org/10.5281/zenodo.5888283> (accessed on 21 January 2022).

Acknowledgments: The authors would like to thank the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences for generously providing the long-term data from flux tower sites.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Porcar-Castell, A.; Tyystjärvi, E.; Atherton, J.; van der Tol, C.; Flexas, J.; Pfündel, E.E.; Moreno, J.; Frankenberg, C.; Berry, J.A. Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: Mechanisms and challenges. *J. Exp. Bot.* **2014**, *65*, 4065–4095. [\[CrossRef\]](#)
2. Berry, J.A.; Frankenberg, C.; Wennberg, P.; Baker, I.; Bowman, K.W.; Castro-Contreras, S.; Cendrero-Mateo, M.P.; Damm, A.; Drewry, D.; Ehlmann, B. New methods for measurement of photosynthesis from space. *Geophys. Res. Lett.* **2012**, *38*, L17706.
3. Zarco-Tejada, P.; Catalina, A.; González, M.; Martín, P. Relationships between net photosynthesis and steady-state chlorophyll fluorescence retrieved from airborne hyperspectral imagery. *Remote Sens. Environ.* **2013**, *136*, 247–258. [\[CrossRef\]](#)
4. Frankenberg, C.; Fisher, J.; Worden, J.; Badgley, G.; Saatchi, S.S.; Lee, J.-E.; Toon, G.C.; Butz, A.; Jung, M.; Kuze, A.; 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*, L17706. [\[CrossRef\]](#)
5. Voigt, M.; Guanter, L.; Zhang, Y.; Walther, S.; Kohler, P.; Jung, M. Global analysis of the relationship between canopy-scale chlorophyll fluorescence and GPP. In Proceedings of the 5th International Workshop on Remote Sensing of Vegetation Fluorescence, Paris, France, 22–24 April 2014.
6. Lee, J.-E.; Berry, J.A.; van der Tol, C.; Yang, X.; Guanter, L.; Damm, A.; Baker, I.; Frankenberg, C. Simulations of chlorophyll fluorescence incorporated into the Community Land Model version 4. *Glob. Chang. Biol.* **2015**, *21*, 3469–3477. [\[CrossRef\]](#)
7. Yang, X.; Tang, J.; Mustard, J.F.; Lee, J.-E.; Rossini, M.; Joiner, J.; Munger, J.W.; Kornfeld, A.; Richardson, A.D. Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest. *Geophys. Res. Lett.* **2015**, *42*, 2977–2987. [\[CrossRef\]](#)
8. Sun, Y.; Frankenberg, C.; Wood, J.D.; Schimel, D.S.; Jung, M.; Guanter, L.; Drewry, D.T.; Verma, M.; Porcar-Castell, A.; Griffis, T.J.; et al. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science* **2017**, *358*, 6360. [\[CrossRef\]](#)
9. Li, X.; Xiao, J.; He, B.; Arain, M.A.; Beringer, J.; Desai, A.R.; Emmel, C.; Hollinger, D.Y.; Krasnova, A.; Mammarella, I.; 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. *Glob. Chang. Biol.* **2018**, *24*, 3990–4008. [\[CrossRef\]](#)
10. Guanter, L.; Zhang, Y.; Jung, M.; Joiner, J.; Voigt, M.; Berry, J.A.; Frankenberg, C.; Huete, A.R.; Zarco-Tejada, P.; Lee, J.-E.; et al. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1327–1333. [\[CrossRef\]](#)
11. Guan, K.; Berry, J.A.; Zhang, Y.; Joiner, J.; Guanter, L.; Badgley, G.; Lobell, D. Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. *Glob. Chang. Biol.* **2016**, *22*, 716–726. [\[CrossRef\]](#)
12. 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*, 112555. [\[CrossRef\]](#)
13. De Cannière, S.; Herbst, M.; Vereecken, H.; Defourny, P.; Jonard, F. Constraining water limitation of photosynthesis in a crop growth model with sun-induced chlorophyll fluorescence. *Remote Sens. Environ.* **2021**, *267*, 112722. [\[CrossRef\]](#)

14. Wen, L.; Guo, M.; Yin, S.; Huang, S.; Li, X.; Yu, F. Vegetation phenology in permafrost regions of Northeastern China based on MODIS and solar-induced chlorophyll fluorescence. *Chin. Geogr. Sci.* **2021**, *31*, 459–473. [\[CrossRef\]](#)
15. Lu, X.; Liu, Z.; Zhou, Y.; Liu, Y.; An, S.; Tang, J. Comparison of phenology estimated from reflectance-based indices and solar-induced chlorophyll fluorescence (SIF) observations in a temperate forest using GPP-based phenology as the standard. *Remote Sens.* **2018**, *10*, 932. [\[CrossRef\]](#)
16. Joiner, J.; Yoshida, Y.; Vasilkov, A.P.; Yoshida, Y.; Corp, L.A.; Middleton, E.M. First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences* **2011**, *8*, 637–651. [\[CrossRef\]](#)
17. Köhler, P.; Guanter, L.; Joiner, J. A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data. *Atmos. Meas. Tech.* **2015**, *8*, 2589–2608. [\[CrossRef\]](#)
18. Joiner, J.; Guanter, L.; Lindstrot, R.; Voigt, M.; Vasilkov, A.; Middleton, E.; Huemmrich, K.; Yoshida, Y.; Frankenberg, C. Global monitoring of terrestrial chlorophyll fluorescence from moderate spectral resolution near-infrared satellite measurements: Methodology, simulations, and application to GOME-2. *Atmos. Meas. Tech.* **2013**, *6*, 2803–2823. [\[CrossRef\]](#)
19. Frankenberg, C.; O'Dell, C.; Berry, J.; Guanter, L.; Joiner, J.; Köhler, P.; Pollock, R.; Taylor, T.E. Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2. *Remote Sens. Environ.* **2014**, *147*, 1–12. [\[CrossRef\]](#)
20. Du, S.; Liu, L.; Liu, X.; Zhang, X.; Zhang, X.; Bi, Y.; Zhang, L. Retrieval of global terrestrial solar-induced chlorophyll fluorescence from TanSat satellite. *Sci. Bull.* **2018**, *63*, 1502–1512. [\[CrossRef\]](#)
21. Sun, Y.; Frankenberg, C.; Jung, M.; Joiner, J.; Guanter, L.; Köhler, P.; Magney, T. 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*, 808–823. [\[CrossRef\]](#)
22. Köhler, P.; Frankenberg, C.; Magney, T.S.; Guanter, L.; Joiner, J.; Landgraf, J. Global retrievals of solar-induced chlorophyll fluorescence with TROPOMI: First results and intersensor comparison to OCO-2. *Geophys. Res. Lett.* **2018**, *45*, 10–456. [\[CrossRef\]](#)
23. Köhler, P.; Behrenfeld, M.J.; Landgraf, J.; Joiner, J.; Magney, T.S.; Frankenberg, C. Global retrievals of solar-induced chlorophyll fluorescence at red wavelengths with TROPOMI. *Geophys. Res. Lett.* **2020**, *47*, e2020GL087541. [\[CrossRef\]](#)
24. Guanter, L.; Bacour, C.; Schneider, A.; Aben, I.; van Kempen, T.A.; 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*, 5423–5440. [\[CrossRef\]](#)
25. Zhang, Y.; Kong, D.; Gan, R.; Chiew, F.H.S.; McVicar, T.R.; Zhang, Q.; Yang, Y. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. *Remote Sens. Environ.* **2019**, *222*, 165–182. [\[CrossRef\]](#)
26. Hu, J.; Liu, L.; Guo, J.; Du, S.; Liu, X. Upscaling solar-induced chlorophyll fluorescence from an instantaneous to daily scale gives an improved estimation of the gross primary productivity. *Remote Sens.* **2018**, *10*, 1663. [\[CrossRef\]](#)
27. 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. [\[CrossRef\]](#)
28. Gentile, P.; Alemohammad, S.H. Reconstructed solar-induced fluorescence: A machine learning vegetation product based on MODIS surface reflectance to reproduce GOME-2 solar-induced fluorescence. *Geophys. Res. Lett.* **2018**, *45*, 3136–3146. [\[CrossRef\]](#)
29. Duveiller, G.; Filippini, F.; Walther, S.; Köhler, P.; Frankenberg, C.; Guanter, L.; Cescatti, A. A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity. *Earth Syst. Sci. Data* **2020**, *12*, 1101–1116. [\[CrossRef\]](#)
30. Wen, J.; Köhler, P.; Duveiller, G.; Parazoo, N.; Magney, T.; Hooker, G.; Yu, L.; Chang, C.; Sun, Y. 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*, 111644. [\[CrossRef\]](#)
31. Zhang, Y.; Joiner, J.; Alemohammad, S.H.; Zhou, S.; Gentile, P. A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. *Biogeosciences* **2018**, *15*, 5779–5800. [\[CrossRef\]](#)
32. Yu, L.; Wen, J.; Chang, C.Y.; Frankenberg, C.; Sun, Y. High-resolution global contiguous SIF of OCO-2. *Geophys. Res. Lett.* **2018**, *26*, 1449–1458. [\[CrossRef\]](#)
33. Li, X.; Xiao, J.; He, B. 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*, 659–671. [\[CrossRef\]](#)
34. Ma, Y.; Liu, L.; Chen, R.; Du, S.; Liu, X. Generation of a global spatially continuous TanSat solar-induced chlorophyll fluorescence product by considering the impact of the solar radiation intensity. *Remote Sens.* **2020**, *12*, 2167. [\[CrossRef\]](#)
35. Guanter, L.; Aben, I.; Tol, P.; Krijger, J.M.; Hollstein, A.; Köhler, P.; Damm, A.; Joiner, J.; Frankenberg, C.; Landgraf, J. Potential of the TROPospheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence. *Atmos. Meas. Tech.* **2015**, *8*, 1337–1352. [\[CrossRef\]](#)
36. Schaaf, C.; Wang, Z. MCD43C4 MODIS/Terra + Aqua BRDF/Albedo Nadir BRDF-Adjusted Ref Daily L3 Global 0.05 Deg CMG V006 [Data Set]. *NASA EOSDIS Land Process. DAAC* **2015**. Available online: <https://catalog.data.gov/dataset/modis-terraaqua-brdf-albedo-nadir-brdf-adjusted-ref-daily-l3-global-0-05deg-cmg-v006> (accessed on 21 January 2022).
37. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration—Guidelines for Computing Crop Water Requirements*—FAO Irrigation and Drainage Paper 56; FAO: Rome, Italy, 1998; Volume 300, p. D05109.
38. Friedl, M.A.; McIver, D.K.; Hodges, J.C.F.; Zhang, X.Y.; Muchoney, D.; Strahler, A.H.; Woodcock, C.E.; Gopal, S.; Schneider, A.; Cooper, A.; et al. Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens. Environ.* **2002**, *83*, 287–302. [\[CrossRef\]](#)

39. Zhang, Y.; Zhang, Q.; Liu, L.; Zhang, Y.; Wang, S.; Ju, W.; Zhou, G.; Zhou, L.; Tang, J.; Zhu, X.; et al. ChinaSpec: A network for long-term ground-based measurements of solar-induced fluorescence in China. *J. Geophys. Res.* **2020**, *126*, e2020JG006042. [\[CrossRef\]](#)
40. Chang, C.Y.; Guanter, L.; Frankenberg, C.; Köhler, P.; Gu, L.; Magney, T.S.; 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. Biogeosci.* **2020**, *125*, e2019JG005533. [\[CrossRef\]](#)
41. Du, S.; Liu, L.; Liu, X.; Guo, J.; Hu, J.; Wang, S.; Zhang, Y. SIFSpec: Measuring solar-induced chlorophyll fluorescence observations for remote sensing of photosynthesis. *Sensors* **2019**, *19*, 3009. [\[CrossRef\]](#)
42. Liu, L.; Liu, X.; Chen, J.; Du, S.; Ma, Y.; Qian, X.; Chen, S.; Peng, D. Estimating maize GPP using near-infrared radiance of vegetation. *Sci. Remote Sens.* **2020**, *2*, 100009. [\[CrossRef\]](#)
43. Du, S.; Liu, L.; Liu, X.; Hu, J. Response of canopy solar-induced chlorophyll fluorescence to the absorbed photosynthetically active radiation absorbed by chlorophyll. *Remote Sens.* **2017**, *9*, 911. [\[CrossRef\]](#)
44. Grossmann, K.; Frankenberg, C.; Magney, T.; Hurlock, S.C.; Seibt, U.; Stutz, J. PhotoSpec: A new instrument to measure spatially distributed red and far-red solar-induced chlorophyll fluorescence. *Remote Sens. Environ.* **2018**, *216*, 311–327. [\[CrossRef\]](#)
45. Liu, X.; Guo, J.; Hu, J.; Liu, L. Atmospheric correction for tower-based solar-induced chlorophyll fluorescence observations at O2-A band. *Remote Sens.* **2019**, *11*, 355. [\[CrossRef\]](#)
46. Maier, S.W.; Günther, K.P.; Stellmes, M. Sun-induced fluorescence: A new tool for precision farming. In *Digital Imaging and Spectral Techniques: Applications to Precision Agriculture and Crop Physiology*; McDonald, M., Schepers, J., Tartly, L., Toai, T.V., Major, D., Eds.; American Society of Agronomy: Madison, WI, USA, 2004; Volume 66, pp. 207–222.
47. Guanter, L.; Frankenberg, C.; Dudhia, A.; Lewis, P.; Gómez-Dans, J.; Kuze, A.; Suto, H.; Grainger, R. Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. *Remote Sens. Environ.* **2012**, *121*, 236–251. [\[CrossRef\]](#)
48. Köhler, P.; Guanter, L.; Kobayashi, H.; Walther, S.; Yang, W. Assessing the potential of sun-induced fluorescence and the canopy scattering coefficient to track large-scale vegetation dynamics in Amazon forests. *Remote Sens. Environ.* **2018**, *204*, 769–785. [\[CrossRef\]](#)
49. Parazoo, N.C.; Frankenberg, C.; Köhler, P.; Joiner, J.; Yoshida, Y.; Magney, T.; Sun, Y.; Yadav, V. Towards a harmonized long-term spaceborne record of far-red solar-induced fluorescence. *J. Geophys. Res. Biogeosci.* **2019**, *124*, 2518–2539. [\[CrossRef\]](#)
50. Liu, S.M.; Xu, Z.W.; Wang, W.Z.; Jia, Z.Z.; Zhu, M.J.; Bai, J.; Wang, J.M. A comparison of eddy-covariance and large aperture scintillometer measurements with respect to the energy balance closure problem. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1291–1306. [\[CrossRef\]](#)
51. Falge, E.; Baldocchi, D.; Olson, R.; Anthoni, P.; Aubinet, M.; Bernhofer, C.; Burba, G.; Ceulemans, R.J.; Clement, R.; Dolman, H.; et al. Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agric. For. Meteorol.* **2001**, *107*, 43–69. [\[CrossRef\]](#)
52. Reichstein, M.; Falge, E.; Baldocchi, D.; Papale, D.; Aubinet, M.; Berbigier, P.; Bernhofer, C.; Buchmann, N.; Gilmanov, T.; Granier, A.; et al. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm. *Glob. Chang. Biol.* **2005**, *11*, 1424–1439. [\[CrossRef\]](#)
53. Damm, A.; Guanter, L.; Paul-Limoges, E.; van der Tol, C.; Hueni, A.; Buchmann, N.; Eugster, W.; Ammann, C.; Schaepman, M. Far-red sun-induced chlorophyll fluorescence shows ecosystem-specific relationships to gross primary production: An assessment based on observational and modeling approaches. *Remote Sens. Environ.* **2015**, *166*, 91–105. [\[CrossRef\]](#)
54. Yoshida, Y.; Joiner, J.; Tucker, C.; Berry, J.; Lee, J.-E.; Walker, G.; Reichle, R.; Koster, R.; Lyapustin, A.; Wang, Y. The 2010 Russian drought impact on satellite measurements of solar-induced chlorophyll fluorescence: Insights from modeling and comparisons with parameters derived from satellite reflectances. *Remote Sens. Environ.* **2015**, *166*, 163–177. [\[CrossRef\]](#)
55. Sellers, P.J.; Tucker, C.J.; Collatz, G.J.; Los, S.O.; Justice, C.O.; Dazlich, D.A.; Randall, D. A global 1° by 1° NDVI data set for climate studies. Part 2: The generation of global fields of terrestrial biophysical parameters from the NDVI. *Int. J. Remote Sens.* **1994**, *15*, 3519–3545. [\[CrossRef\]](#)
56. Jiang, D.; Wang, N.; Yang, X.; Liu, H. Dynamic properties of absorbed photosynthetic active radiation and its relation to crop yield. *Syst. Sci. Compr. Stud. Agric.* **2002**, *18*, 51–54.
57. Liu, L.; Liu, X.; Hu, J.; Guan, L. Assessing the wavelength-dependent ability of solar-induced chlorophyll fluorescence to estimate the GPP of winter wheat at the canopy level. *Int. J. Remote Sens.* **2017**, *38*, 4396–4417. [\[CrossRef\]](#)
58. Yang, P.; van der Tol, C. Linking canopy scattering of far-red sun-induced chlorophyll fluorescence with reflectance. *Remote Sens. Environ.* **2018**, *209*, 456–467. [\[CrossRef\]](#)
59. Liu, L.; Zhang, X.; Xie, S.; Liu, X.; Song, B.; Chen, S.; Peng, D. Global white-sky and black-sky FAPAR retrieval using the energy balance residual method: Algorithm and validation. *Remote Sens.* **2019**, *11*, 1004. [\[CrossRef\]](#)
60. Zhang, Z.; Chen, J.M.; 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*, 12030–12040. [\[CrossRef\]](#)
61. Zhang, Z.; Zhang, Y.; Chen, J.M.; Ju, W.; Migliavacca, M.; El-Madany, T.S. Sensitivity of estimated total canopy SIF emission to remotely sensed LAI and BRDF products. *J. Geophys. Res. Biogeosci.* **2021**, *2021*, 9795837. [\[CrossRef\]](#)
62. Gu, L.; Han, J.; Wood, J.D.; Chang, C.Y.; Sun, Y. Sun-induced Chl fluorescence and its importance for biophysical modeling of photosynthesis based on light reactions. *New Phytol.* **2019**, *223*, 1179–1191. [\[CrossRef\]](#) [\[PubMed\]](#)

-
63. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
 64. Rossini, M.; Nedbal, L.; Guanter, L.; Ac, A.; Alonso, L.; Burkart, A.; Cogliati, S.; Colombo, R.; Damm, A.; Drusch, M.; et al. Red and far red sun-induced chlorophyll fluorescence as a measure of plant photosynthesis. *Geophys. Res. Lett.* **2015**, *42*, 1632–1639. [[CrossRef](#)]
 65. Mutanga, O.; Adam, E.; Cho, M.A. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 399–406. [[CrossRef](#)]
 66. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
 67. Booth, J.G.; Hall, P.; Wood, A.T.A. Balanced importance resampling for the bootstrap. *Ann. Stat.* **1993**, *21*, 286–298. [[CrossRef](#)]
 68. Zhang, Z.; Zhang, Y.; Zhang, Y.; Chen, J.M. Correcting clear-sky bias in gross primary production modeling from satellite solar-induced chlorophyll fluorescence data. *J. Geophys. Res. Biogeosci.* **2020**, *125*, e2020JG005822. [[CrossRef](#)]
 69. Chang, C.Y.; Wen, J.; Han, J.; Kira, O.; LeVonne, J.; Melkonian, J.; Riha, S.J.; Skovira, J.; Ng, S.; Gu, L.; et al. Unpacking the drivers of diurnal dynamics of sun-induced chlorophyll fluorescence (SIF): Canopy structure, plant physiology, instrument configuration and retrieval methods. *Remote Sens. Environ.* **2021**, *265*, 112672. [[CrossRef](#)]