Filling gaps in global daily TROPOMI solar-induced chlorophyll fluorescence data from 2018 to 2021

Jingbo Li, Qunming Wang, and Peter M. Atkinson

***Abstract*—Solar-induced chlorophyll fluorescence (SIF) is a crucial variable towards timely and effective monitoring of vegetation productivity, as well as physiological and biochemical parameters, across extensive areas. Among these advances, the TROPOspheric Monitoring Instrument (TROPOMI) SIF has significantly increased the spatiotemporal resolution and data coverage compared to previous sensors. However, TROPOMI SIF data suffer from nonuniform sampling, swath gaps and cloud contamination, resulting in numerous instances of missing data. In this paper, we proposed a physical and spatial information-aided gap filling (PSGF) method, which addresses effectively the missing data problem, generating a spatially seamless, 0.05°, daily SIF (S^2-SIF) dataset globally at a spatial resolution of 0.05° from 2018 to 2021. Through missing data simulation experiments conducted in six regions worldwide, we demonstrated consistency between the reference SIF and the filled SIF, with a correlation coefficient (CC) of 0.659. Furthermore, validation using *in situ* data from 35 SIF and gross primary productivity (GPP) ground sites yielded a CC of approximately 0.70 for the SIF sites and CC values above 0.60 between the ground GPP and filled SIF. Additionally, consistency was observed between the filled SIF datasets and two other SIF products across 11 vegetation types, confirming the reliability of the filled SIF data and the efficacy of the PSGF method. The produced filled SIF data are made publicly available and should increase greatly the applicability of the daily SIF data for a wide range of applications, including quantifying the photosynthesis of vegetation and accurately estimating GPP globally.**

***Index Terms*—solar-induced chlorophyll fluorescence (SIF), TROPOspheric Monitoring Instrument (TROPOMI), gross primary productivity (GPP), photosynthesis, machine learning.**

I. INTRODUCTION

Solar-induced chlorophyll fluorescence (SIF) is a complementary product of vegetative photosynthesis that can mirror directly the dynamics of plant photosynthesis [1] and discern the physiological intricacies of photosynthesis in vegetation [2]. Vegetation photosynthesis constitutes a pivotal element of the carbon cycle within terrestrial ecosystems. Thus, precise characterization of the physiological responses of vegetation energized by SIF is imperative to enable accurate estimation of Gross Primary Productivity (GPP) [3, 4] and, thereby, enhance predictive capacity pertaining to the global carbon cycle [5].

Currently, SIF remote sensing offers novel technological and data acquisition approaches spanning from leaf, to plant canopy, to the global scale [6]. Up to now, several SIF satellites have been launched, and ultra-fine-spectral satellite sensors such as GOSAT, GOME-2, OCO-2, OCO-3, and TanSat have been employed widely for SIF detection [6-10]. This expanding suite of sensors enables a comprehensive understanding of the Earth's biosphere at a macroscopic level, elucidating global seasonal variation [11], vegetation phenology [12, 13] and evapotranspiration [14], and enabling drought forecasting [15].

Most of the earlier satellite sensors with SIF detection capabilities typically offer discrete sampling and long revisit intervals, leading to global SIF products with coarse temporal resolution. Since October 2017, the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Copernicus Sentinel-5 Precursor satellite has acquired global SIF (TROPOSIF) data with a daily revisit cycle, presenting new potential for SIF remote sensing. However, owing to the narrow swath width of the TROPOMI sensor, large orbital intervals and its susceptibility to cloud cover [16], gaps in the daily TROPOSIF data persist, hindering its wider applicability, including the capture of real-time fluctuations in photosynthetic capacity during sudden environmental changes, analysis of finely refined vegetation functional traits [17] and regional-scale crop yield estimation [18].

To fill the gaps in the daily TROPOSIF SIF data, machine learning algorithms have been applied due to their robust performance [19-22]. For example, several studies [8, 23, 24] have utilized neural networks to model the MODIS reflectance-SIF relationship, resulting in the development of SIF global data products such as contiguous solar-induced fluorescence (CSIF) with 4-day frequency and 0.05° spatial resolution, and the global "OCO-2" SIF dataset (GOSIF) with 8-day frequency and 0.05°spatial resolution. Camps-Valls et al. indicated that remotely sensed vegetation indices (VIs), derived from spectral reflectance and ancillary data, can effectively capture vegetation cover, biochemistry, structure and function [25]. For example, the normalized difference vegetation index is effective in assessing chlorophyll content [26], while the near-infrared reflectance of vegetation demonstrates time-specific properties and a large correlation with SIF and GPP [27]. Integration of these indices into SIF data filling is beneficial to increase accuracy [16, 28]. However, the MODIS-based vegetation index products used in these studies exhibit significant gaps due to cloud contamination. Although these gaps can be reduced to some extent by 4-day and 8-day composites, it is still challenging to obtain spatially seamless MODIS data.

Manuscript received ***; revised ***; accepted ***. This work was supported by the National Natural Science Foundation of China under Grants 42222108 and 42171345. *(Corresponding author*: *Q. Wang*.)

J. Li is with the College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China.

Q. Wang is with the College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China and also with Key Laboratory of Ethnic Language Intelligent Analysis and Security Governance of MOE, Minzu University of China, Beijing 100081, China (e-mail: wqm11111@126.com).

P.M. Atkinson is with the Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, UK and also with Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

Vegetation photosynthesis is closely related to SIF, but existing studies have not utilized effectively the covariates controlling vegetation photosynthesis to fill the SIF gaps. In this paper, covariates providing information on vegetation photosynthesis were used for gap filling daily TROPOSIF data. We classified the variables affecting vegetation photosynthesis into five types: topographic factors, solar radiation, extreme climate, weather condition, and vegetation status. Moreover, we also integrated spatial features (characterized by longitude and latitude). As a result, we proposed a physical and spatial information-aided gap filling (PSGF) method. Based on this method, we generated a spatially seamless, 0.05°, daily SIF (S^2-SIF) dataset. The availability of this new spatially seamless SIF dataset holds significant potential for global and regional-scale GPP estimation and global ecosystem monitoring.

II. Materials

A. *In-situ data*

In this research, SIF station data from two locations were utilized: Xiaotangshan station (XTS) and Shangqiu station (SQ) in China, which are publicly available through the ChinaSpec network (https://chinaspec.nju.edu.cn/). The XTS and SQ stations are situated in cropland areas with two crop rotations annually; specifically winter wheat and summer maize [29, 30]. The GPP data employed in this research were sourced from the Ameriflux GPP site (http://ameriflux.ornl.gov/), which were derived by upscaling GPP based on flux observatory sites (including a total of 34 sites). Fig. 1 shows the MCD12Q1 classification data, from which vegetation data were extracted across various regions including Evergreen Needleleaf Forests (ENF), Evergreen Broadleaf Forests (EBF), Deciduous Needleleaf Forests (DNF), Deciduous Broadleaf Forests (DBF), Mixed Forests (MF), Closed Shrublands (CS), Open Shrublands (OS), Woody Savannas (WS), Savannas (SA), Grasslands (GL), Permanent Wetlands (PW), Croplands (CL), and Cropland/Natural Vegetation Mosaics (CNVM).

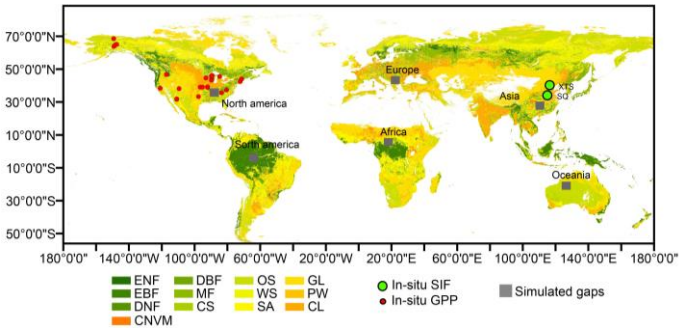

Fig. 1. Locations of *in situ* SIF and GPP data and simulated gaps. See text for definition of vegetation classes.

B. *Satellite sensor data*

1) SIF data. Daily TROPOSIF data from May 2018 to December 2021 [9] were downloaded. Given that Hu et al. [16] demonstrated greater accuracy in the validation of $TROPOSIF_{743}$, we selected the daily average version for use in this paper. However, the data were stored as discrete points and, thus, we gridded the TROPOSIF data with geospatial attributes into SIF images with a spatial resolution of 0.05°. We also calculated the percentage of missing TROPOSIF data for each date. As shown in Fig. 2, the average missing ratio exceeds 0.5.

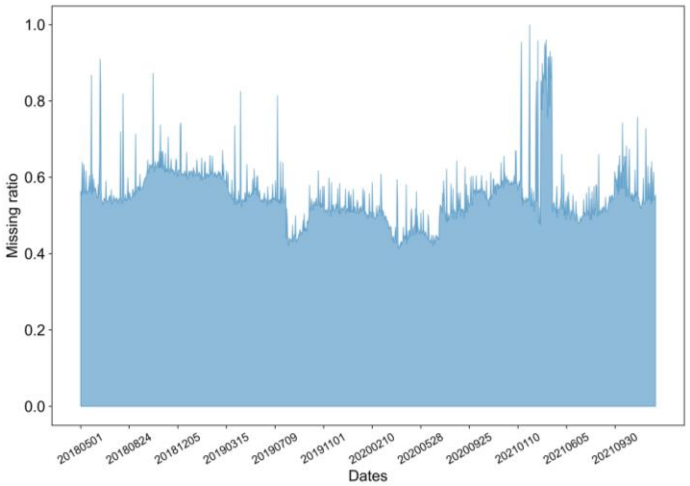

Fig. 2. Missing ratio of the TROPOSIF data on each date.

2) ERA 5 data. The used ERA5 data were generated by reproducing the land component of the ECMWF ERA5 climate reanalysis. This dataset offers fine temporal resolution and a comprehensive range of open data on terrestrial ecosystems. From this dataset, we selected key variables including temperature, precipitation, solar radiation, forecast albedo, leaf area index and skin reservoir content in the vegetation canopy (0.1° × 0.1°). We averaged all meteorological and chemical composition simulations at the hourly level to obtain daily values. The variables were interpolated from the original spatial resolution to approximately 5 km resolution (0.05° × 0.05°) using the bilinear interpolation method [31] to match that of the SIF data.

3) DEM data. The ASTER Global Digital Elevation Model is a topographic dataset released collaboratively by the National Aeronautics and Space Administration and Japan's Ministry of Economy, Trade, and Industry in 2009. The dataset uses the UTM/WGS84 projection coordinate system. In Google Earth Engine, the ID of the DEM data used in this research is "NASA/ASTER_GED/AG100_003". We used elevation, slope and aspect as the key terrain features.

III. Methods

A. *Input physical and spatial features*

Based on the mechanisms affecting vegetation photosynthesis, the physical features of the PSGF model are divided into five components: topographic factors, solar radiation, extreme climate, weather condition and vegetation status. Collectively, they capture the influence of environment on vegetation photosynthesis, as illustrated in Fig. 3 and detailed in Table 1.

1) Topographic factors: Factors such as elevation, slope and aspect relative to solar radiation flux reach the vegetation-covered surface. They influence the amount of solar radiation absorbed by the vegetation.
2) Solar radiation: The principles of SIF estimation are as follows:

$$SIF = APAR \times \Phi_F \times f_{esc} \quad (1)$$

where APAR is the photosynthetically active radiation (PAR) absorbed by the canopy, Φ_F is the physiological total SIF emission yield of the canopy photosynthetic light response, and f_{esc} is the escape probability. In this paper, we used surface solar radiation downwards, surface net solar radiation sum and forecast albedo from ERA5 as the relevant variables for PAR.

3) Extreme climate: Research has shown that extreme weather events lead to decreased vegetation productivity due to incurred damage [32] and because vegetation has an optimal temperature threshold for photosynthesis. Specifically, excessively high surface temperatures cause vegetation stomata to close, halting photosynthesis, while excessively low surface temperatures impede vegetation activity and hinder growth [33]. In this paper, we used temperature max and temperature min as covariates representing extreme climate variables.
4) Weather condition: Vegetation photosynthesis is also controlled by weather condition. The factors characterizing weather condition include precipitation and temperature. These two factors govern the water and temperature stress experienced by vegetation. Both significantly impact plant respiration and soil organic carbon decomposition, ultimately influencing plant photosynthesis [34].
5) Vegetation status: The state of vegetation affects the ability to undertake photosynthesis. We used leaf area index (LAI) and skin reservoir content to characterize vegetation status. LAI delineates the structure of vegetation, directly impacting the degree of sunlight absorption [35]. Skin reservoir content is closely linked to the transpiration and photosynthesis processes of vegetation.

Table 1. Detailed physical characteristics used in the proposed PSGF method.

Factors	Characteristics	Detailed characteristics	Abbreviation
Topographic factor	Elevation	The height above sea level.	ELE
	Slope	The angle at which it slopes.	SLO
	Aspect	The surface orientation.	ASP
Solar radiation	Surface net solar radiation	Amount of solar radiation reaching the surface of the Earth (both direct and diffuse) minus the amount reflected by the Earth's surface.	SNSR
	Surface solar radiation downwards	Amount of solar radiation reaching the surface of the Earth.	SSRD
	Forecast albedo	The reflectivity of the Earth's surface.	FA
Extreme climate	Skin temperature max	Daily maximum skin temperature value.	*Tmax*
	Skin temperature min	Daily minimum skin temperature value.	*Tmin*
Weather condition	Skin temperature	Temperature of the surface of the Earth.	*T*
	Total precipitation sum	Accumulated liquid and frozen water, including rain and snow that fall to the Earth's surface.	TPS
Vegetation status	Leaf area index high vegetation	One-half of the total green leaf area per unit horizontal ground surface area for high vegetation type.	LAI-H
	Leaf area index low vegetation	One-half of the total green leaf area per unit horizontal ground surface area for low vegetation type.	LAI-L
	Skin reservoir content	Amount of water in the vegetation canopy.	SRC

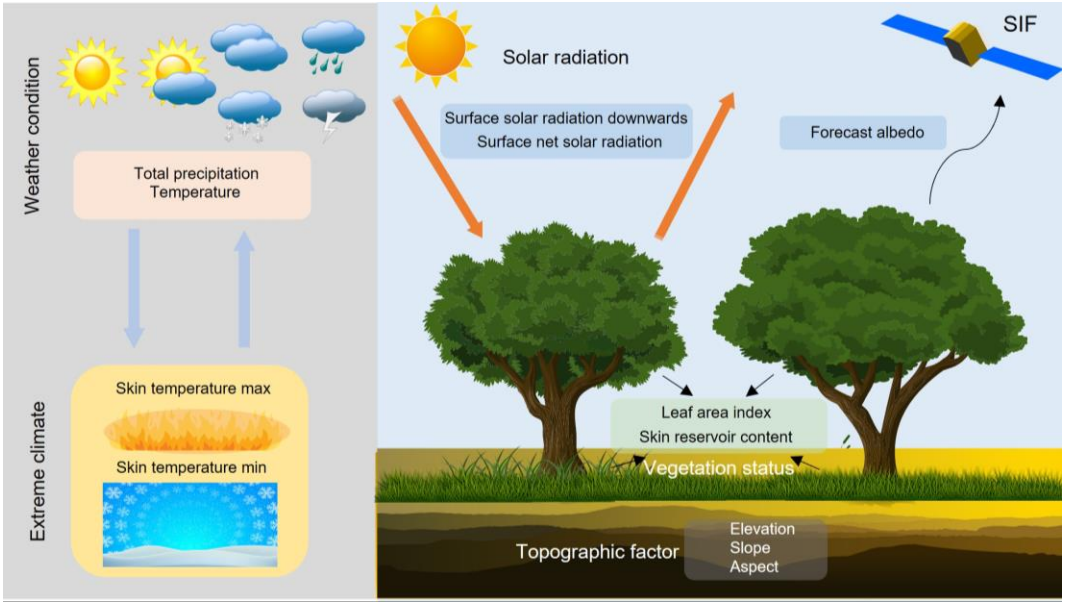

Fig. 3. Physical features driving photosynthesis in vegetation.

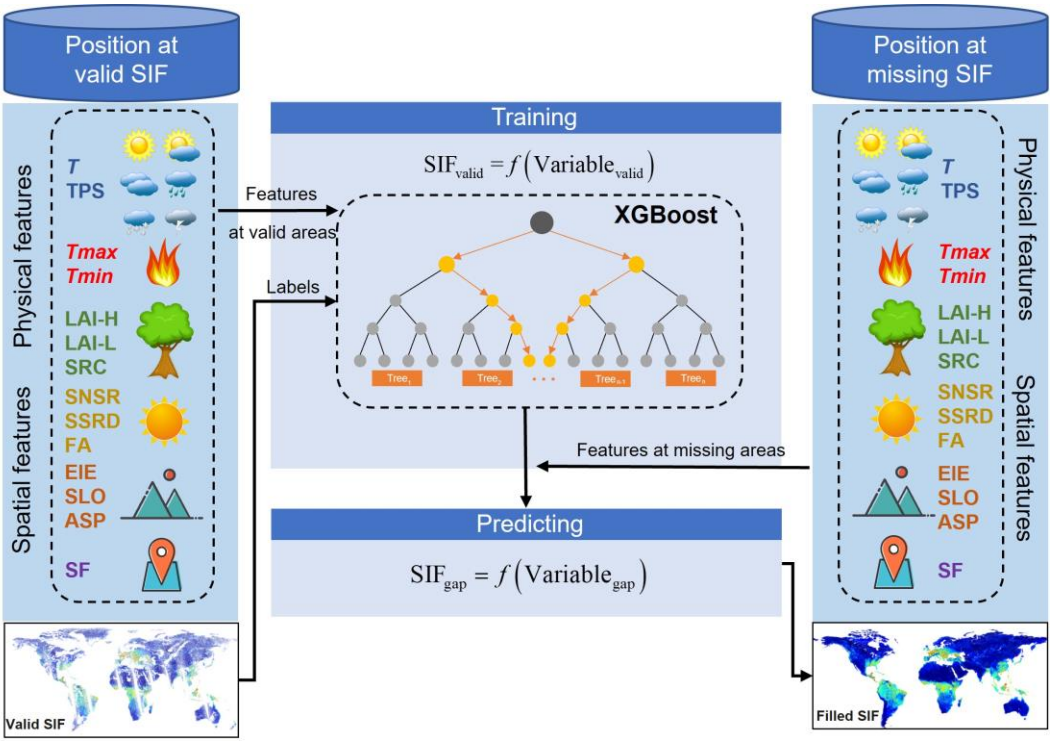

Fig. 4. Framework of the proposed PSGF model for gap filling of SIF data.

As acknowledged widely, the spatial distribution of SIF is related closely to geographical position on the sphere. Regarding spatial features, we used the weighting effect of polar coordinates [36] for characterization, as shown in Eq. (2):

$$\mathrm{SF}=\begin{bmatrix}\cos\left(2\pi\frac{lon}{360}\right)\cos\left(2\pi\frac{lat}{180}\right),\cos\left(2\pi\frac{lon}{360}\right)\sin\left(2\pi\frac{lat}{180}\right),\\ \sin\left(2\pi\frac{lon}{360}\right)\sin\left(2\pi\frac{lat}{180}\right),\sin\left(2\pi\frac{lat}{360}\right)\cos\left(2\pi\frac{lat}{180}\right),\\ \sin\left(2\pi\frac{lon}{360}\right),\cos\left(2\pi\frac{lat}{180}\right)\end{bmatrix} \quad (2)$$

where SF denotes the spatial features, *lon* represents longitude with a range of [0, 360] and *lat* represents latitude with a range of [0, 180].

We calculated the correlation coefficient (CC) between SIF data and each physical variable for each date from May 2018 to December 2021, as shown in the Appendix. The results demonstrate relatively strong correlation between SIF and the physical variables. These findings highlight the importance of using these physical variables in SIF gap filling.

B. Model training

Suppose that the SIF with gaps on date *n* is to be filled. As shown in Fig. 4, the basic principle of the PSGF model is as follows:

$$\mathrm{Variable}=\{\mathrm{TF},\mathrm{SR},\mathrm{EC},\mathrm{WC},\mathrm{VS},\mathrm{SF}\} \quad (3)$$

$$\mathrm{SIF}_{\mathrm{valid}}^{n}=f\left(\mathrm{Variable}_{\mathrm{valid}}^{n}\right) \quad (4)$$

where $\mathrm{SIF}_{\mathrm{valid}}^{n}$ is the valid value of the non-missing SIF zone for date *n*. TF, SR, EC, WC and VS denote topographic factors, solar radiation, extreme climate, weather condition, and vegetation status. *f* characterizes the relation between the valid SIF data and the input variables.

To fill the global SIF data, we chose a machine learning method to fit the complex relationship (i.e., function *f*) between the valid SIF and the input variables (i.e., labels of the training data and input data). Specifically, we selected XGBoost due to its robustness to multicollinearity and default values, noise tolerance and capability to handle large datasets. It consists of a series of decision trees, where each decision tree is generated based on the previous one, and the spatial gradient of the loss function is continuously reduced by weighting multiple decision trees. Operations such as regularization, parallel processing, missing value processing and feature subsampling are introduced to reduce overfitting and increase robustness. To ensure the stability of the model, five-fold cross-validation was employed in the training process.

C. Model prediction

With the XGBoost-based training model in Section III-B, the relation between the valid SIF and the input physical and spatial features can be fitted by the function *f*, which can then be used to predict the missing SIF data on the corresponding date *n*. That is, the SIF for the gaps is filled using the following Eq. (5):

$$\mathrm{SIF}_{\mathrm{gap}}^{n}=f\left(\mathrm{Variable}_{\mathrm{gap}}^{n}\right) \quad (5)$$

where $\mathrm{SIF}_{\mathrm{gap}}^{n}$ is the predicted SIF for the gaps and $\mathrm{Variable}_{\mathrm{gap}}^{n}$ represents the input physical and spatial features at the corresponding gap areas.

D. Model validation

In this paper, the accuracy and robustness of the PSGF model were validated using (i) a complete SIF dataset in which the gaps to be filled were simulated and (ii) a real experiment using the global, daily TROPOSIF time-series from 2018 to 2021. For the simulated gaps, the filled SIF predictions were compared with the original real SIF. Through this scheme, the performance of

the proposed PSGF method can be evaluated objectively, as the reference is known perfectly. The accuracy of the filled S^2-SIF at the global scale was validated using *in situ* SIF and GPP data, and also compared with other SIF and GPP products. Quantitative assessment was conducted using the CC, root mean square error (RMSE) and mean absolute error (MAE).

IV. Results

The validation results for the simulated and real data experiments are provided in Sections IV-A and IV-B, respectively.

A. Validation based on simulated gaps

Fig. 5 shows a comparison between the reference (i.e., with gaps) SIF data and the filled SIF data for six simulated regions. The six simulated regions were assumed to have completely missing SIF data. The first line of Fig. 5(a) and Fig. 5(b) presents the reference SIF data that are originally available. The references were used to evaluate the performance of gap filling for the simulated regions. Note there are also missing data in the references, which is an inherent characteristic of satellite observations of TROPOSIF data. These simulated regions are distributed across each continent, with the locations depicted in Fig. 1. Furthermore, the reference SIF data display distinct global variation in distribution. For example, on Oct 3, 2018, the southern hemisphere experiences spring while the northern hemisphere is in autumn, resulting in notably larger SIF values in Africa compared to other regions. Conversely, Oceania exhibits consistently smaller SIF values due to sparse vegetation, as depicted in Fig. 1, where the predominant vegetation type is OS. On May 23, 2020, during the northern hemisphere's spring season, elevated SIF values are observed in Europe and North America. These observations highlight distinct global differences between the six regions. Following the application of the filling model, the previously missing regions are effectively filled.

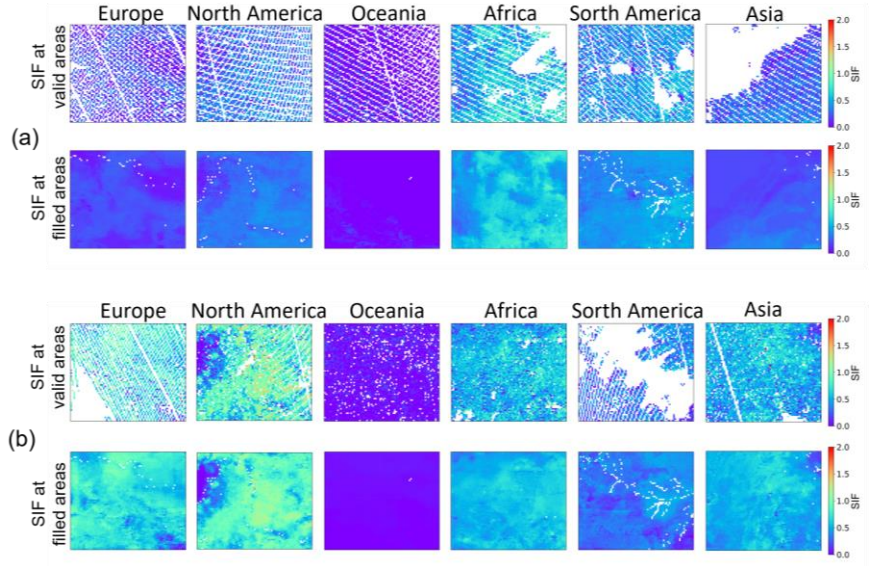

Fig. 5. Experimental results of the six regions of the gap filling simulation. (a) Oct 3, 2018. (b) May 23, 2020.

Table 2 presents a quantitative assessment of prediction accuracy following the filling process. The results indicate that on October 3, 2018, Africa exhibited the most accurate filling result, with a CC of 0.797, accompanied by RMSE and MAE values of 0.076 and 0.061, respectively. On May 23, 2020, North America demonstrated the highest filling accuracy, attaining a CC of 0.904, with RMSE and MAE values of 0.110 and 0.086. However, certain regions exhibited lower accuracies, notably Europe on October 3, 2018 and Oceania on May 23, 2020. This was likely caused by senescing or sparse vegetation and lower solar radiation during this period. Additionally, weaker sunlight during winter contributes to reduced SIF, as depicted in Fig. 5. Furthermore, noise in sparsely vegetated areas, such as Oceania, adds complexity to the filling process, as illustrated in Fig. 5.

Table 2. Accuracy of the SIF gap filling model simulations.

Date	October 3, 2018			May 23, 2020		
Accuracy	CC	RMSE	MAE	CC	RMSE	MAE
Asia	0.594	0.047	0.037	0.744	0.066	0.051
South America	0.683	0.050	0.039	0.751	0.062	0.049
Africa	0.797	0.076	0.061	0.697	0.057	0.045
Oceania	0.605	0.044	0.035	0.294	0.023	0.019
North America	0.605	0.055	0.042	0.904	0.110	0.086
Europe	0.388	0.024	0.019	0.842	0.079	0.061

B. Validation based on real gaps

1) Visual results of the filled global SIF data

We filled the global daily SIF data from May 2018 to December 2021, and five resulting maps are depicted, together with the original reference data, in Fig. 6. By selecting different dates for each year (May 1, 2018, February 26, 2019, July 20, 2021, October 18, 2021, and April 18, 2020), we ensured to cover different seasons throughout the year.

Spatially, the filled gaps in the SIF data exhibit a consistent distribution with the valid SIF values of the surrounding vegetation. However, some patterns resembling salt-and-pepper noise are observed, which we further explain in Section V-D. The tropical rainforest regions (Amazon, Indonesia, and Congo) exhibit consistently large SIF values throughout the year, attributed to the vigorous photosynthesis in tropical rainforest

vegetation and the region's tropical climate zone. The regions with large SIF values on at least one date are predominantly concentrated in the western United States, South America, southwestern Europe, central Africa, southeastern China, and Southeast Asia. In contrast, areas dominated by OS vegetation types, such as Australia, southern Africa, Argentina, southwestern North America, and northern Asia, display consistently small SIF values. Overall, the SIF data exhibit seasonal variability, with smaller values observed in winter, and peak values occurring in summer, a trend closely linked to the amount of short-wave radiation from the Sun.

2) Accuracy evaluation based on the in situ SIF data

The SIF values filled by the PSGF model were validated against ground station reference data. As depicted by the scatterplots in Fig. 7, the scatterplots for the filled SIF data (against the ground reference) demonstrate substantial concordance with the scatterplots for the valid SIF data (against the ground reference), with the goodness-of-fits and the fitted slopes being consistent between the two. Specifically, the CC between the ground reference SIF at the XTS station and the valid SIF data is 0.505, slightly larger than for the filled SIF data, which is 0.486. At the SQ station, the CC between the *in situ* SIF and the valid SIF data and between the *in situ* SIF and the filled SIF data were nearly identical, measuring 0.752 and 0.759, respectively.

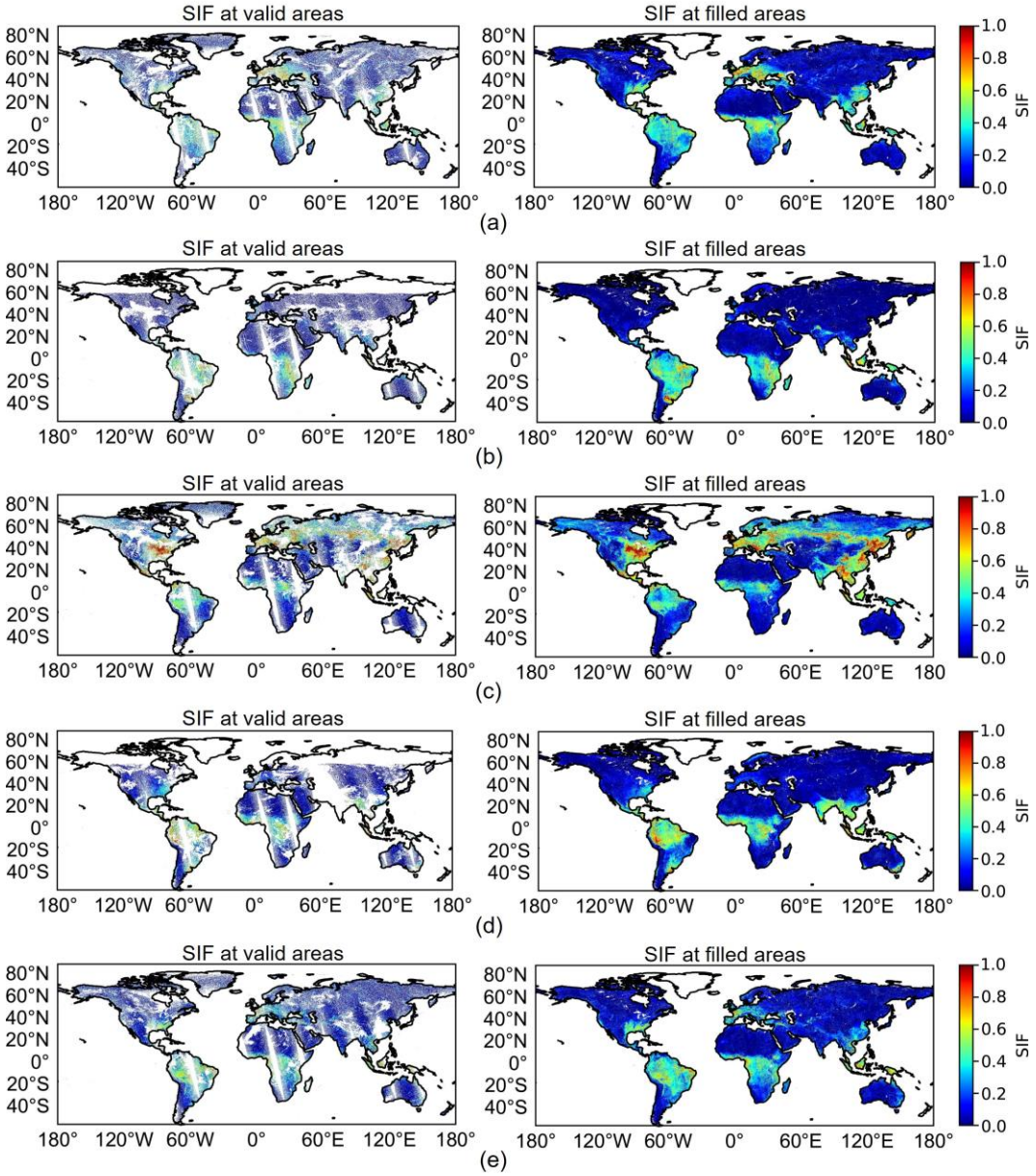

Fig. 6. Maps of (left) reference and (right) gap-filled global SIF. (a) May 1, 2018, (b) February 26, 2019, (c) July 20, 2021, (d) October 18, 2021, (e) April 18, 2020.

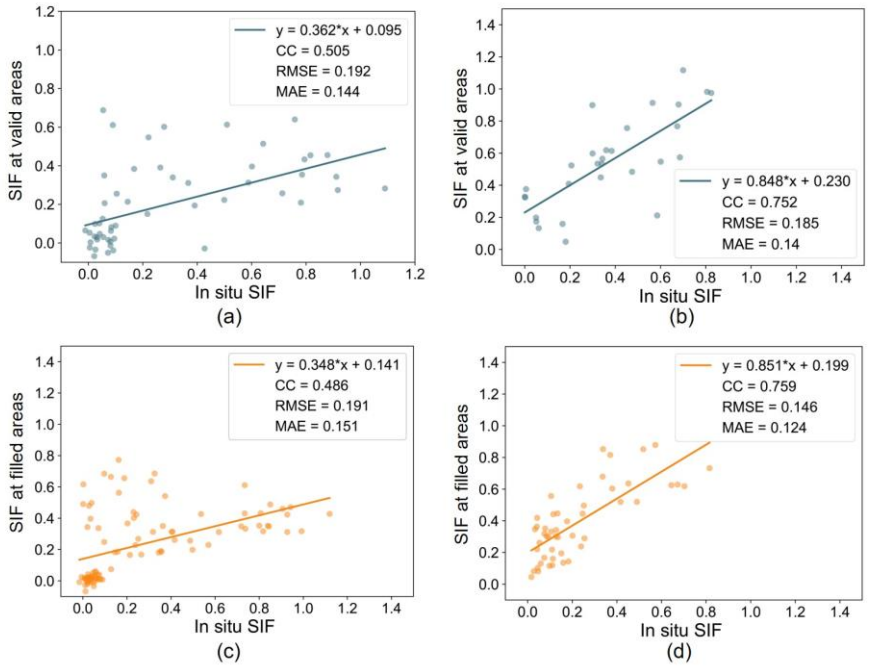

Fig. 7. Scatterplots of SIF at ground stations against the (upper) SIF reference data and (lower) SIF predictions. (a) and (c) are results for XTS site. (b) and (d) are results for SQ site.

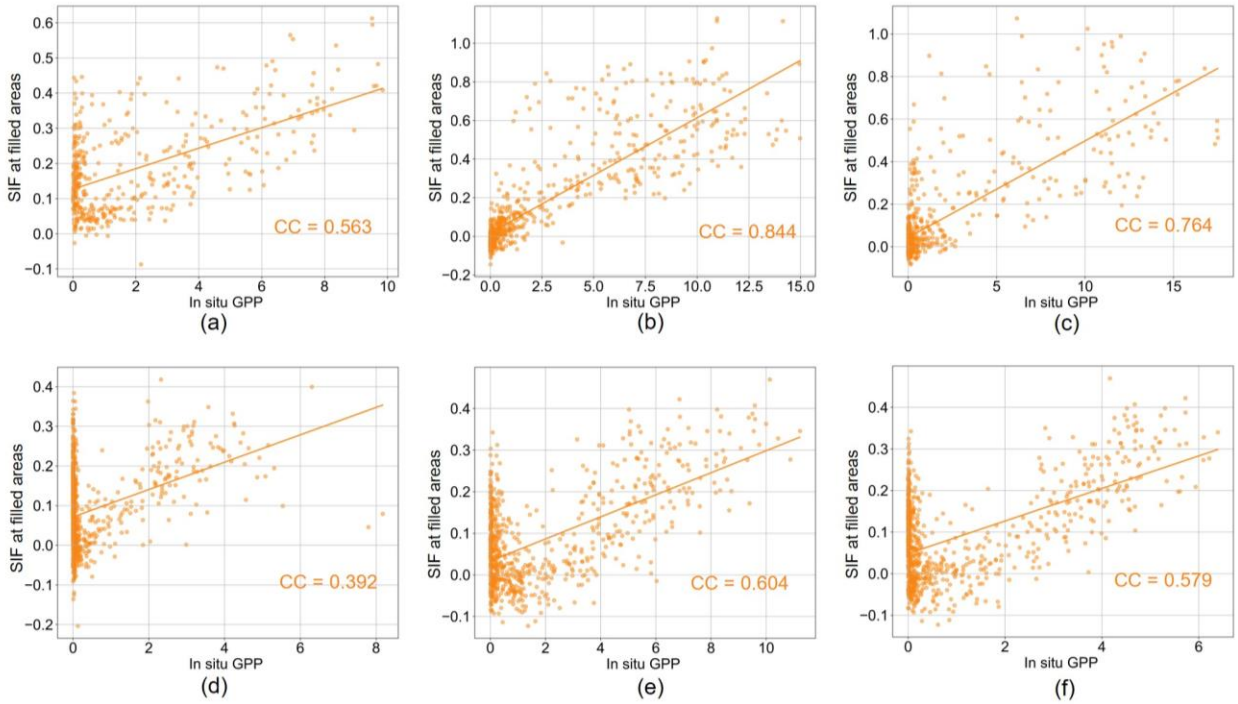

Fig. 8. Accuracy of the GPP sites compared to the SIF at filled areas (P-value<0.005). (a) US-Var. (b) US-xTR. (c) US-CF4. (d) US-Ich. (e) US-BZS. (f) US-BZB.

C. Comparison with in situ GPP data

GPP represents the total carbon absorbed by green plants through photosynthesis. Research suggests that SIF exhibits great sensitivity to vegetative photosynthesis and, thus, maintains a robust linear relationship with GPP [2, 37]. Because of the wide availability of data on GPP, GPP can be used to validate methods of predicting SIF [4]. Fig. 8 shows scatterplots of measured GPP against filled SIF for six sites selected from 34 sites. The CC between *in situ* GPP and the filled SIF at the US-Var, US-xTR, US-CF4, US-Ich, US-BZS, and US-BZB sites was 0.563, 0.844, 0.764, 0.393, 0.604 and 0.579, respectively, demonstrating the reliability of the PSGF model.

Fig. 9 presents detailed results for each of the available 33 GPP sites. The mean CC between the site GPP and filled SIF is 0.688, with a good degree of consistency, further highlighting the efficacy and robustness of the PSGF. To further assess the accuracy of the SIF predictions, Fig. 10 shows a time-series plot of the valid SIF and filled SIF together with the site GPP for two

ground sites. It is evident that the filled SIF maintains the same temporal trend as the GPP time-series, similar to the valid SIF.

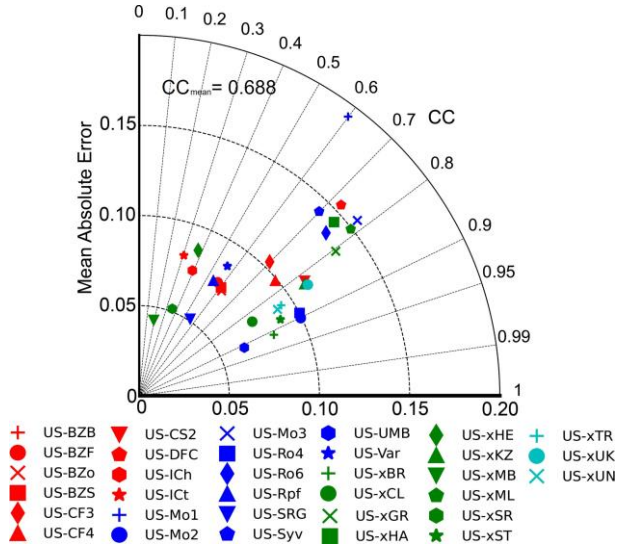

Fig. 9. The correlation between *in situ* GPP at each site and the SIF at filled area (P-value<0.005).

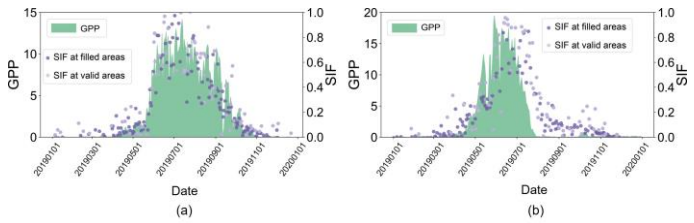

Fig. 10. Time-series plots of filled and valid SIF, and GPP, for two GPP sites: (a) US-xTR and (b) US-CF4.

D. *Ablation experiments*

Each of the used physical variables has the potential to influence vegetation photosynthesis significantly. Hence, it is crucial to evaluate the influence of each variable (e.g., by adding spatial features to the physical variables in the model) to ensure maximum accuracy and robustness when compensating for missing SIF data. Each physical variable was entered into the model individually, before including all five physical variables together and then all variables (i.e., adding spatial features to the physical variables). The dataset comprises 80% for training and 20% for validation. The prediction accuracy when using only the topographic factors are notably low, with CC of 0.254 and a larger RMSE than for other physical variables. The CC is largest for the extreme climate and vegetation status parameters, 0.604 and 0.603, respectively, underscoring their significance in compensating for the missing SIF data. Moreover, by adding the spatial features (ALL in Fig. 11), the prediction accuracy is increased by approximately 2% compared to using only the five physical variables. The individual contribution of each variable will be further explained in Section V-A.

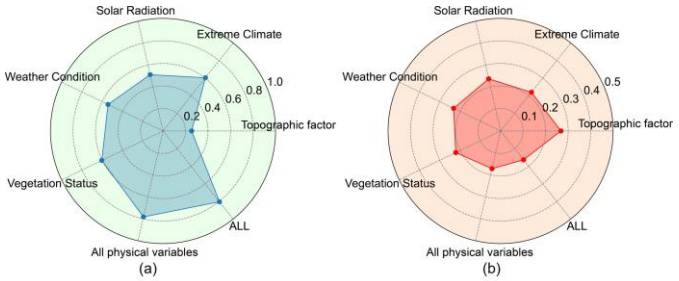

Fig. 11. Accuracy results for individual, and combinations of, covariates in the ablation experiments. (a) CC. (b) RMSE.

V. Discussion

A. *Factors affecting the filled SIF data*

We conducted an analysis of the importance scores of the variables used in SIF filling, as depicted in Fig. 12. Overall, spatial features exhibit the highest contribution, at 0.38. This prominence can be attributed to spatial features encompassing geographic characteristics, including seasonal and climate-specific attributes. Additionally, both the extreme climate and vegetation status contribute approximately 0.2 to the filled SIF. Studies have demonstrated the significance of extreme climate, such as cold, drought and freezing events, in regulating photosynthesis [32]. Extreme climate influences vegetation photosynthesis primarily through temperature variation [34]. We utilized *Tmin* and *Tmax* to characterize specific climatic temperature conditions, such as cold and drought, finding their contributions to be slightly higher than those of *T* within weather condition. Notably, *Tmin* exhibits a significantly higher contribution, indicating its sensitive control on vegetation photosynthesis across different geographic locations. This conclusion further demonstrates that temperature influences vegetation activity and transpiration, indirectly affecting SIF [38]. This underscores the importance of considering both *Tmin* and *Tmax* variation for SIF filling.

Within the vegetation status group, LAI-H makes the largest contribution to SIF filling, aligning with previous studies on the contribution (and correlation) of LAI to SIF [39]. LAI reflects the photosynthetic capacity of canopy and may also indicate the contributions of grasslands or shrubs to the SIF signal [40]. Additionally, the skin reservoir content contributes approximately 0.04, underscoring its crucial role in vegetation photosynthesis [41, 42]. Furthermore, forecast albedo exhibiting the highest contribution due to its large correlation with reflectance. This prominence is attributed to forecast albedo serving as a representative of reflectance, complementing SIF measurements by capturing narrow solar Fraunhofer lines using fine-resolution spectrometers [6]. In weather condition, besides *T*, precipitation accounts for nearly 0.02. Precipitation enhances soil moisture and atmospheric humidity, thereby promoting vegetation photosynthesis. Conversely, low precipitation levels lead to reduced surface and deep soil moisture, resulting in decreased photosynthesis. Although topographic factors contribute the least to filled SIF, topographic changes impact shortwave solar radiation and are, therefore, worth considering.

B. *Comparison with other SIF products*

To further assess the reliability and robustness of the PSGF and further confirm the data accuracy of the S^2-SIF, we compared the S^2-SIF with (i) the CSIF product [24], featuring a

four-day temporal resolution and 0.05 °spatial resolution and (ii) the reconstructed TROPOMI SIF (RTSIF) product [28], with an eight-day temporal resolution and 0.05 ° spatial resolution. First, we averaged the S^2-SIF to match the temporal resolution of the other SIF products. To validate the data points, we utilized the MCD12Q1 ground vegetation cover type data (Fig. 1), where we categorized ENF and EBF as evergreen forest (EF), and DBF and DNF as deciduous forest (DF). We extracted 11 vegetation types including EF, DF, MF, CS, OS, WS, SA, GL, PW, CL and CNVM, each with 10,000 pixels. Accuracy was assessed by computing the CC, RMSE and MAE between these vegetation validation points and the corresponding CSIF and RTSIF product data. The results of this comparison are presented in Table 3.

From Table 3, we observe that the S^2-SIF exhibits consistency with both the four-day temporal resolution CSIF product and the eight-day temporal resolution RTSIF product. Notably, the correlation with the RTSIF product is the largest, with most vegetation types achieving a CC above 0.8. To elucidate the greater association with RTSIF, we analyze two factors. Firstly, both the S^2-SIF and RTSIF datasets originate from TROPOMI satellite sensor data, ensuring data source consistency. Conversely, CSIF is generated by OCO-2 SIF reconstruction. This distinction is evident in Fig. 13, illustrating that the S^2-SIF exhibits greater visual consistency with RTSIF, whereas some discrepancies exist with CSIF. Secondly, the literature suggests that increasing the number of dates for SIF averaging reduces in-grid SIF inversion errors by a factor of $\sqrt{n}$ [28, 43]. With increased averaging over multiple dates, noise in the SIF data diminishes, leading to more stable and reliable results. Therefore, the RTSIF product data with an 8-day time resolution is expected to be more robust.

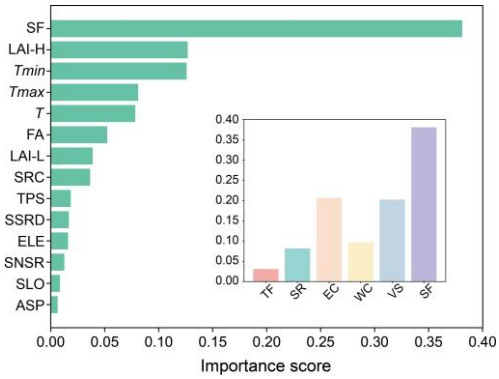

Fig. 12. Importance scores of the variables used for SIF filling.

Table 3. Comparison of S^2-SIF with the CSIF and RTSIF products in terms of the CC, RMSE and MAE, under different vegetation covers.

Vegetation	CSIF			RTSIF		
	CC	RMSE	MAE	CC	RMSE	MAE
CL	0.873	0.097	0.073	0.911	0.079	0.060
CNVM	0.831	0.107	0.081	0.888	0.087	0.066
CS	0.824	0.069	0.05	0.865	0.058	0.043
DF	0.534	0.079	0.059	0.620	0.062	0.047
EF	0.606	0.089	0.068	0.736	0.069	0.054
GL	0.834	0.089	0.067	0.870	0.076	0.058
MF	0.780	0.082	0.063	0.867	0.061	0.047
OS	0.468	0.076	0.059	0.472	0.067	0.053
PW	0.679	0.102	0.077	0.827	0.077	0.059
Sa	0.866	0.098	0.075	0.913	0.079	0.061
WS	0.868	0.095	0.072	0.924	0.073	0.055

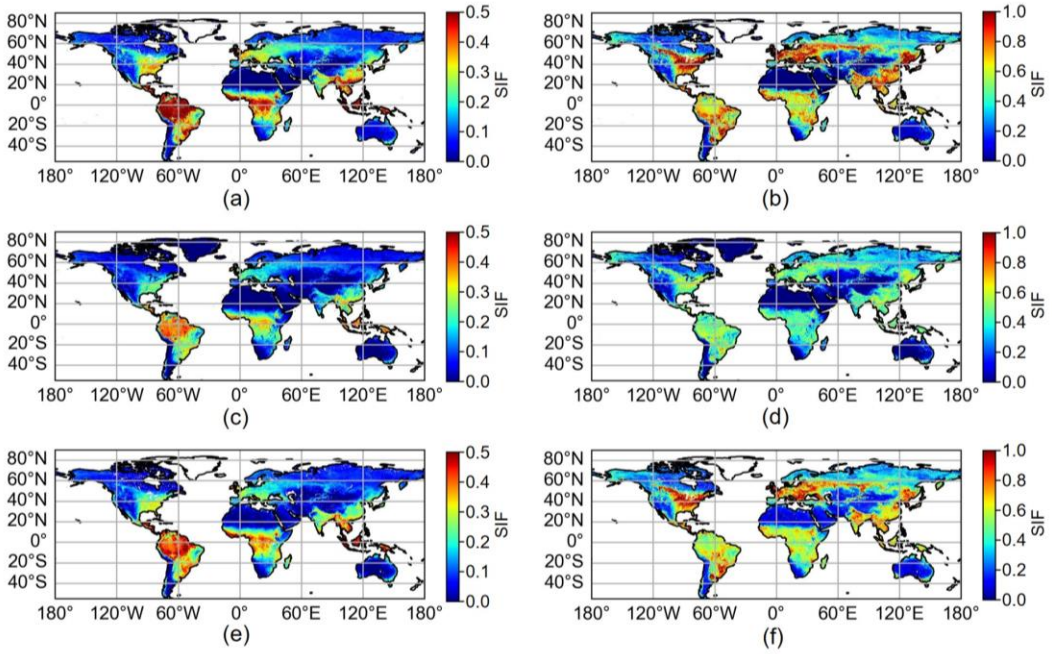

Fig. 13. Global spatial presentation of RTSIF, CSIF and filled SIF datasets showing (left) annual mean and (right) annual maximum values. (a) and (b) RTSIF. (c) and (d) CSIF. (e) and (f) S^2-SIF.

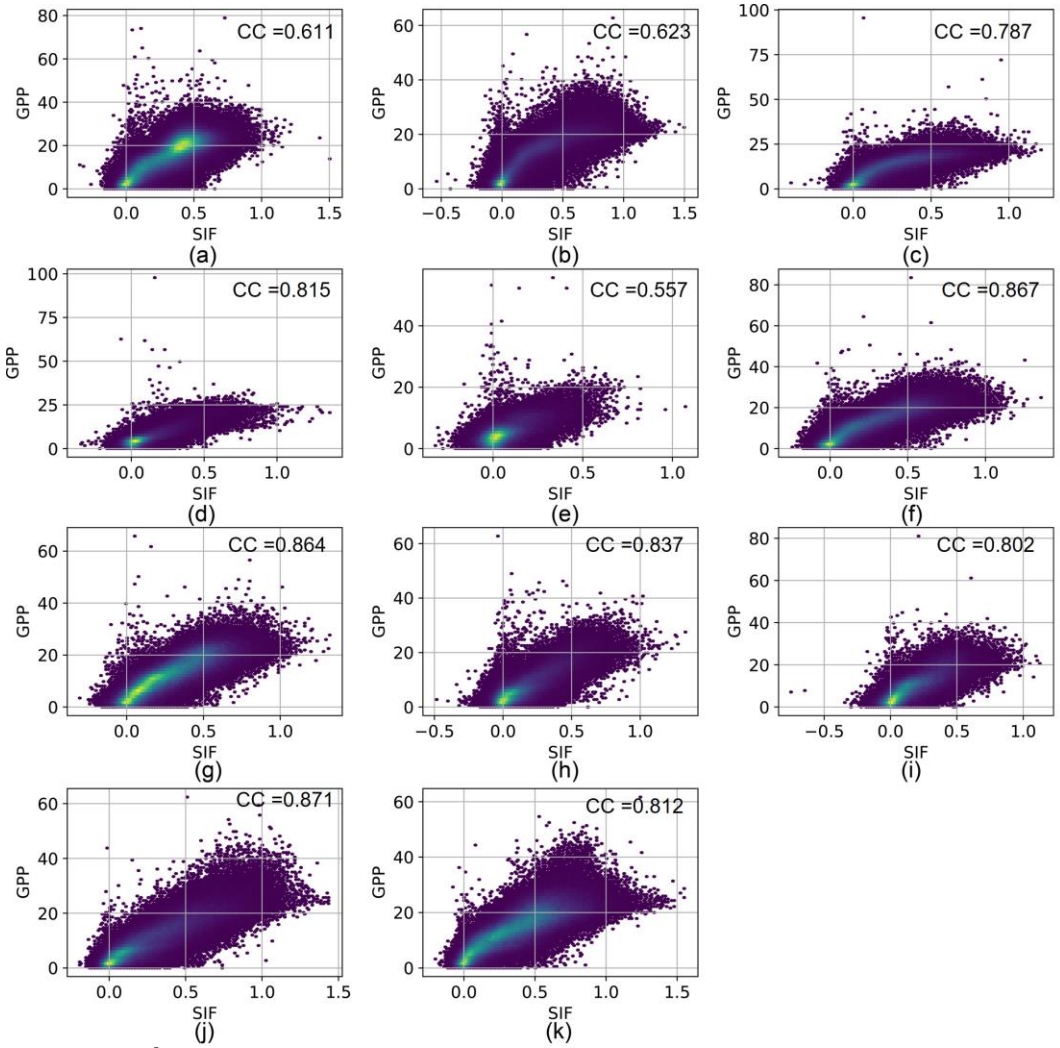

Fig. 14. Correlation results between S^2-SIF and the GPP product under different vegetation cover types. (a) EF. (b) DF. (c) MF. (d) CS. (e) OS. (f) WS. (g) Sa. (h) GL. (i) PW. (j) CL. (k) CNVM.

We conducted a comparative analysis of the accuracy of the S^2-SIF with CSIF and RTSIF across various vegetation types. Overall, the S^2-SIF produced a large correlation with the CSIF and RTSIF products across different vegetation types, with CC exceeding 0.8 for the CL, CNVM, CS, GL, SA and WS vegetation types. Specifically, in the correlation analysis with RTSIF products, correlations surpass 0.9 for the CL, SA and WS vegetation types. However, we observed slightly lower accuracies for two vegetation types: DF and OS. The DF vegetation region, mainly found in tropical rainforest regions, exhibits relatively weak seasonal variation, necessitating more complex explanatory variables in model construction. Conversely, the OS vegetation region displays a weaker SIF signal and, thus, experiences greater levels of signal noise interference throughout the year. In the OS, climatic factors, such as low precipitation, result in reduced water availability in both the soil and atmosphere. This leads to decreased stomatal conductance, leaf surface water potential, and xylem water transport in vegetation leaves, subsequently reducing photochemical and fluorescence efficiency [35].

C. S^2-SIF comparison with GPP product data

SIF was utilized effectively for calculating global GPP data products [37]. To further validate the application value of the S^2-SIF dataset generated by the PSGF model, we examined the correlation between S^2-SIF and the global GPP dataset [4] across various vegetation cover types: EF, DF, MF, CS, OS, WS, Sa, GL, PW, CL and CNVM. As illustrated in Fig. 14, the CC between S^2-SIF and CMLR-GPP exceeds 0.8 for the CS, WS, Sa, GL, PW, CL and CNVM vegetation types. The smallest CC, with a coefficient of 0.557, is observed for OS vegetation types. Overall, these results demonstrate the large correlation between S^2-SIF and GPP.

To further assess the physiological relevance of the reconstructed SIF, we validated it by comparing the difference in temporal profiles between S^2-SIF and GPP. We selected four vegetation types (forest, shrub, grassland, and farmland) and plotted the time-series for both filled SIF and GPP in Fig. 15. The results show strong correlations and similar seasonal trends, indicating that the gap-filled SIF effectively retains important physiological information across these vegetation types.

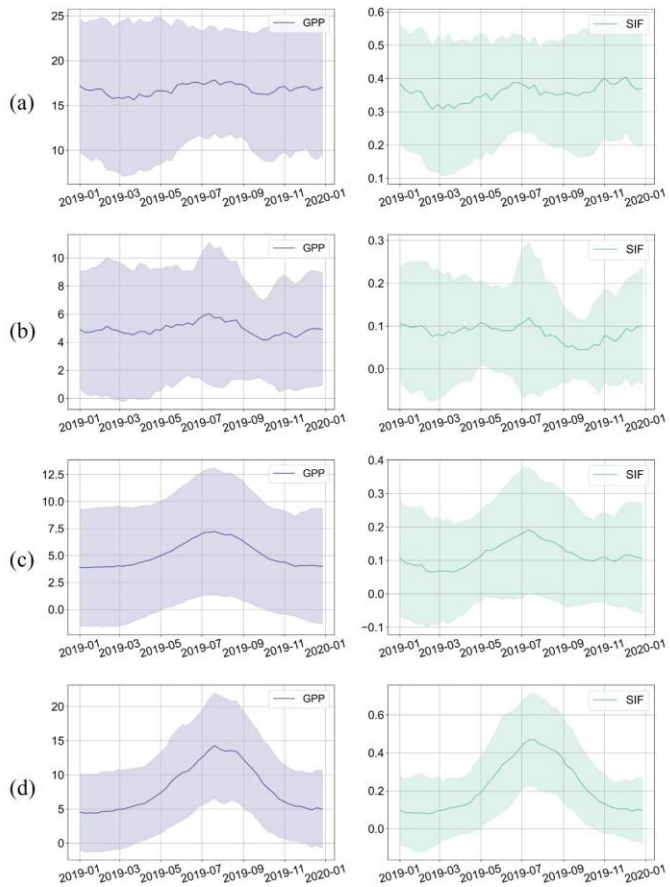

Fig. 15. The temporal profiles of the GPP and filled SIF for different vegetation cover types. (a) Forest. (b) Shrub. (c) Grassland. (d) Farmland.

D. Advantages of the proposed PSGF model and applications in the future

To address the pervasive problem of gaps in global daily SIF data, we introduced the PSGF model for SIF gap filing that incorporates both physical variables and spatial features. The set of physical variables represent comprehensively the factors influencing vegetation photosynthesis, split into five groups: topographic factors, solar radiation, extreme climate, weather condition and vegetation status. These factors were shown to contribute effectively to the PSGF model (Fig. 12). Considering the global-scale computation and the complex relationship between SIF and the independent variables, we introduced the machine learning model (XGBoost), which has advantages in handling large-scale datasets and complex model training. The model also excels in balancing processing efficiency and accuracy, as evidenced by its strong performance in various applications [44, 45]. We provided the training time cost of the model for each date in Fig. 16. It is seen that the average training time is about 12 s. Moreover, the model effectively captures complex nonlinear relationships between input variables and SIF. Meanwhile, the regularization operation effectively reduces the overfitting phenomenon [28].

The independent variables used in the PSGF include additional features related to weather condition and extreme climate , compared to previous work [28]. As discussed in Section V-A, the combined proportion of these features amounts to 0.3, highlighting the superiority of our model. LAI has always been considered in SIF-filling models, typically as an important factor [35, 46]. However, LAI saturates in areas with high vegetation cover [47]. Therefore, we introduced variables for LAI in high vegetation (LAI-H) and low vegetation (LAI-L) to quantify LAI in greater detail. Due to the absence of daily mean scale LAI data, it is difficult to explore the distinction between LAI-H/LAI-L and LAI, or their respective contributions to SIF.

We employed the PSGF model to address missing SIF data and, thereby generated a seamless 0.05 ° daily SIF dataset (S^2-SIF). It is important to elucidate the significance of this achievement in relation to potential applications. Previous studies have utilized SIF datasets scaled primarily at intervals of four days, eight days and monthly [23, 24, 28]. However, gaps exist in these products, with a notable lack of SIF products providing seamless daily temporal resolution. The PSGF model proposed herein effectively resolves this limitation. Moreover, the introduction of S^2-SIF with seamless daily temporal resolution will facilitate advances in addressing various challenges. For example, the current coarse spatiotemporal resolution of SIF data hampers the identification of "sudden drought" events lasting less than eight days, as well as the accurate extraction of vegetation phenology over extended timeframes, and in local areas. Furthermore, it inhibits the assessment of short-term stressors on vegetation photosynthetic capacity. Therefore, the availability of a seamless daily temporal resolution SIF dataset holds significant promise for researchers and policymakers engaged in environmental and climate change research, vegetation monitoring and agronomy.

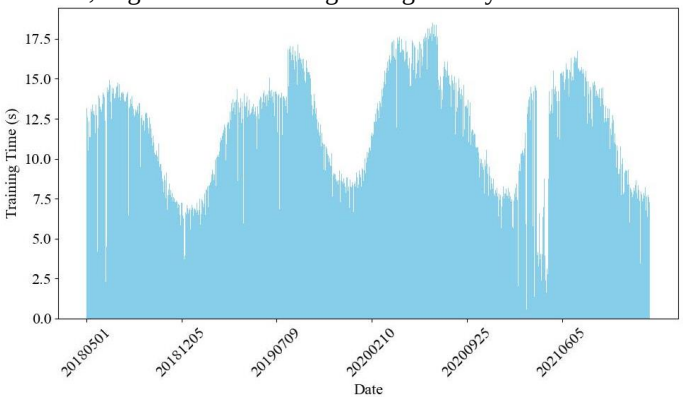

Fig. 16. The training time cost of the model for each date.

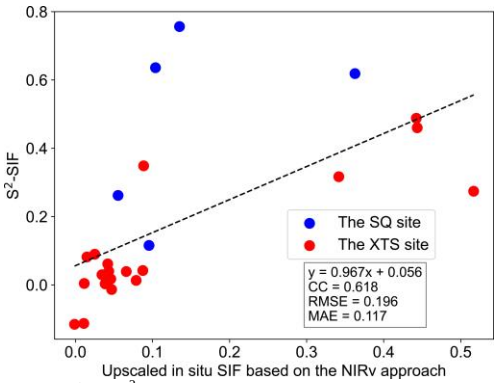

Fig. 17. Accuracy of the S^2-SIF validated using the upslaced *in situ* SIF based on the NIRv approach.

E. Uncertainty in SIF gap filling

The gap filling of SIF greatly expands the application potential of SIF, and the generated daily scale SIF product is of great significance for large-scale vegetation monitoring and carbon sink calculations. However, the following uncertainties also exist in the gap filling:

1) We observed that the SIF products exhibited some noise during the experiments, as illustrated in Fig. 5. Consequently, the presence of noise introduces errors into the model during the gap filling process. In our future research, we plan to reconstruct the SIF to produce a data product with enhanced precision and spatiotemporal resolution.
2) The model utilizes the available SIF value from the same day within the valid area to fill gaps in the missing area. Consequently, our gap filling model becomes ineffective when faced with completely missing dates, which is a limitation of the gap filling model itself.
3) Topographic factors are incorporated into the model because the solar radiation flux varies across different locations due to terrain elevation and surface orientation relative to sunlight, influencing vegetation photosynthesis and the SIF values. However, the DEM data utilized in our research is unavailable for the North and South Poles regions, leading to the inability to fill SIF gaps in these areas.
4) *In situ* SIF data are widely regarded as "true values" for validating satellite SIF products [23, 24, 28]. However, due to inconsistencies in spatial scale and geographical locations between satellite pixels and *in situ* points, unavoidable uncertainty exists in the evaluation scheme [16, 28]. Spatial upscaling is an effective solution to increase the reliability of ground measurements used for evaluation, by aggregating the *in situ* data within a single pixel to produce pixel-scale ground data. For example, Du et al. [48] established a relationship between Sentinel-2 NIRv and tower-based NIRv to upscale *in situ* SIF to satellite-scale SIF. Based on this approach, we upscaled the *in situ* SIF using Sentinel-2 data, and the accuracy evaluation is shown in Fig. 17. However, it should be pointed out that this method requires temporal alignment between Sentinel-2 and *in situ* data, making it suitable for validating instantaneous SIF values. Additionally, satellite overpasses require revisit cycles, which significantly limits the number of valid validation points. As a result, it remains challenging to validate daily SIF products using this approach. With the expansion of flux tower networks and the accumulation of more *in situ* data, this limitation is expected to be overcome to some extent.

VI. Conclusion

We proposed the PSGF model to fill the gaps in TROPOSIF data and generate a spatially seamless 0.05° global daily SIF dataset (S^2-SIF) from 2018 to 2021. For validation, the SIF images for six simulated regions were gap-filled. By referring to the reference data, the overall CC, RMSE and MAE were 0.659, 0.058, and 0.045. Additionally, we validated the model predictions using *in situ* data from 36 SIF and GPP sites, with an overall CC between GPP and filled SIF of 0.688. Comparison with other SIF products (CSIF and RTSIF) yielded consistent results, particularly in the CL, CNVM, CS, GL, Sa and WS vegetation types, with a CC above 0.8. Furthermore, we conducted ablation experiments using physical and spatial information to analyze the contribution of SIF-related factors to SIF filling, providing insights for future studies. In summary, the PSGF effectively addresses missing SIF data problem, yielding spatiotemporally complete datasets crucial for global and regional-scale GPP estimation and global carbon assessment. The produced S^2-SIF data can be accessed via the link available at https://doi.org/10.5281/zenodo.11918785.

References

[1] N. R. Baker, "Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo," *Annual Review of Plant Biology*, vol. 59, no. 1, pp. 89-113, 2008, 2008.

[2] J. K. Green, J. Berry, P. Ciais, Y. Zhang, and P. Gentine, "Amazon rainforest photosynthesis increases in response to atmospheric dryness," *Science Advances*, vol. 6, no. 47, pp. eabb7232, 2020-11-20, 2020.

[3] J. Bai, H. Zhang, R. Sun, X. Li, J. Xiao, and Y. Wang, "Estimation of global GPP from GOME-2 and OCO-2 SIF by considering the dynamic variations of GPP-SIF relationship," *Agricultural and Forest Meteorology*, vol. 326, pp. 109180, 11/2022, 2022.

[4] R. Chen, L. Liu, X. Liu, and U. Rascher, "CMLR: A Mechanistic Global GPP Dataset Derived from TROPOMIS SIF Observations," *Journal of Remote Sensing*, vol. 4, pp. 0127, 2024-02-13, 2024.

[5] W. Li, J. Pacheco-Labrador, M. Migliavacca, D. Miralles, A. Hoek van Dijke, M. Reichstein, M. Forkel, W. Zhang, C. Frankenberg, A. Panwar, Q. Zhang, U. Weber, P. Gentine, and R. Orth, "Widespread and complex drought effects on vegetation physiology inferred from space," *Nature Communications*, vol. 14, no. 1, pp. 4640, 2023-08-15, 2023.

[6] Y. Sun, C. Frankenberg, J. D. Wood, D. S. Schimel, M. Jung, L. Guanter, D. T. Drewry, M. Verma, A. Porcar-Castell, T. J. Griffis, L. Gu, T. S. Magney, P. Köhler, B. Evans, and K. Yuen, "OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence," *Science*, vol. 358, no. 6360, pp. eaam5747, 2017-10-13, 2017.

[7] S. Du, L. Liu, X. Liu, X. Zhang, X. Zhang, Y. Bi, and L. Zhang, "Retrieval of global terrestrial solar-induced chlorophyll fluorescence from TanSat satellite," *Science Bulletin*, vol. 63, no. 22, pp. 1502-1512, 11/2018, 2018.

[8] P. Gentine, and S. H. Alemohammad, "Reconstructed Solar-Induced Fluorescence: A Machine Learning Vegetation Product Based on MODIS Surface Reflectance to Reproduce GOME-2 Solar-Induced Fluorescence," *Geophysical Research Letters*, vol. 45, no. 7, pp. 3136-3146, 2018, 2018.

[9] L. Guanter, C. Bacour, A. Schneider, I. Aben, T. A. Van Kempen, F. Maignan, C. Retscher, P. Köhler, C. Frankenberg, J. Joiner, and Y. Zhang, "The TROPOSIF global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission," *Earth System Science Data*, vol. 13, no. 11, pp. 5423-5440, 2021-11-19, 2021.

[10] L. Yu, J. Wen, C. Y. Chang, C. Frankenberg, and Y. Sun, "High-Resolution Global Contiguous SIF of OCO-2," *Geophysical Research Letters*, vol. 46, no. 3, pp. 1449-1458, 2019-02-16, 2019.

[11] A. Chen, J. Mao, D. Ricciuto, D. Lu, J. Xiao, X. Li, P. E. Thornton, and A. K. Knapp, "Seasonal changes in GPP/SIF ratios and their climatic determinants across the Northern Hemisphere," *Global Change Biology*, vol. 27, no. 20, pp. 5186-5197, 2021, 2021.

[12] X. Li, and J. Xiao, "TROPOMI observations allow for robust exploration of the relationship between solar-induced chlorophyll fluorescence and terrestrial gross primary production," *Remote Sensing of Environment*, vol. 268, pp. 112748, 01/2022, 2022.

[13] J. Yang, X. Xiao, R. Doughty, M. Zhao, Y. Zhang, P. Köhler, X. Wu, C. Frankenberg, and J. Dong, "TROPOMI SIF reveals large uncertainty in estimating the end of plant growing season from vegetation indices data in the Tibetan Plateau," *Remote Sensing of Environment*, vol. 280, pp. 113209, 10/2022, 2022.

[14] Q. Zhang, X. Liu, K. Zhou, Y. Zhou, P. Gentine, M. Pan, and G. G. Katul, "Solar-induced chlorophyll fluorescence sheds light on global evapotranspiration," *Remote Sensing of Environment*, vol. 305, pp. 114061, 2024-05-01, 2024.

[15] S. Xu, J. Atherton, A. Riikonen, C. Zhang, J. Oivukkamäki, A. MacArthur, E. Honkavaara, T. Hakala, N. Koivumäki, Z. Liu, and A. Porcar-Castell, "Structural and photosynthetic dynamics mediate the response of SIF to water stress in a potato crop," *Remote Sensing of Environment*, vol. 263, pp. 112555, 09/2021, 2021.

[16] J. Hu, J. Jia, Y. Ma, L. Liu, and H. Yu, "A Reconstructed Global Daily Seamless SIF Product at 0.05 Degree Resolution Based on TROPOMI, MODIS and ERA5 Data," *Remote Sensing*, vol. 14, no. 6, pp. 1504, 2022/1, 2022.

[17] L. Guanter, Y. Zhang, M. Jung, J. Joiner, M. Voigt, J. A. Berry, C. Frankenberg, A. R. Huete, P. Zarco-Tejada, J.-E. Lee, M. S. Moran, G. Ponce-Campos, C. Beer, G. Camps-Valls, N. Buchmann, D. Gianelle, K. Klumpp, A. Cescatti, J. M. Baker, and T. J. Griffis, "Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence," *Proceedings of the National Academy of Sciences*, vol. 111, no. 14, 2014-04-08, 2014.

[18] X. Kang, C. Huang, L. Zhang, H. Wang, Z. Zhang, and X. Lv, "Regional-scale cotton yield forecast via data-driven spatio-temporal prediction (STP) of solar-induced chlorophyll fluorescence (SIF)," *Remote Sensing of Environment*, vol. 299, pp. 113861, 2023-12-15, 2023.

[19] J. Gensheimer, A. J. Turner, P. Köhler, C. Frankenberg, and J. Chen, "A convolutional neural network for spatial downscaling of satellite-based solar-induced chlorophyll fluorescence (SIFnet)," *Biogeosciences*, vol. 19, no. 6, pp. 1777-1793, 2022-03-31, 2022.

[20] O. Kira, and Y. Sun, "Extraction of sub-pixel C3/C4 emissions of solar-induced chlorophyll fluorescence (SIF) using artificial neural network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 161, pp. 135-146, 03/2020, 2020.

[21] H. Shen, Y. Wang, X. Guan, W. Huang, J. Chen, D. Lin, and W. Gan, "A Spatiotemporal Constrained Machine Learning Method for OCO-2 Solar-Induced Chlorophyll Fluorescence (SIF) Reconstruction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-17, 2022, 2022.

[22] J. Wen, P. Köhler, G. Duveiller, N. C. Parazoo, T. S. Magney, G. Hooker, L. Yu, C. Y. Chang, and Y. Sun, "A framework for harmonizing multiple satellite instruments to generate a long-term global high spatial-resolution solar-induced chlorophyll fluorescence (SIF)," *Remote Sensing of Environment*, vol. 239, pp. 111644, 03/2020, 2020.

[23] X. Li, and J. Xiao, "A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS, and Reanalysis Data," *Remote Sensing*, vol. 11, no. 5, pp. 517, 2019/1, 2019.

[24] Y. Zhang, J. Joiner, S. H. Alemohammad, S. Zhou, and P. Gentine, "A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks," *Biogeosciences*, vol. 15, no. 19, pp. 5779-5800, 2018-10-02, 2018.

[25] G. Camps-Valls, M. Campos-Taberner, Á. Moreno-Martínez, S. Walther, G. Duveiller, A. Cescatti, M. D. Mahecha, J. Muñoz-Marí, F. J. García-Haro, L. Guanter, M. Jung, J. A. Gamon, M. Reichstein, and S. W. Running, "A unified vegetation index for quantifying the terrestrial biosphere," *Science Advances*, vol. 7, no. 9, pp. eabc7447, 2021-02-26, 2021.

[26] R. B. Myneni, F. G. Hall, P. J. Sellers, and A. L. Marshak, "The interpretation of spectral vegetation indexes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 2, pp. 481-486, 1995-03, 1995.

[27] G. Badgley, C. B. Field, and J. A. Berry, "Canopy near-infrared reflectance and terrestrial photosynthesis," *Science Advances*, vol. 3, no. 3, pp. e1602244, 2017-03-22, 2017.

[28] X. Chen, Y. Huang, C. Nie, S. Zhang, G. Wang, S. Chen, and Z. Chen, "A long-term reconstructed TROPOMI solar-induced fluorescence dataset using machine learning algorithms," *Scientific Data*, vol. 9, no. 1, pp. 427, 2022-07-20, 2022.

[29] S. Du, X. Liu, W. Duan, and L. Liu, "Validation of solar-induced chlorophyll fluorescence products derived from OCO-2/3 observations using tower-based in situ measurements," *Remote Sensing Letters*, vol. 14, no. 7, pp. 713-721, 2023-07-03, 2023.

[30] Z. Li, Q. Zhang, J. Li, X. Yang, Y. Wu, Z. Zhang, S. Wang, H. Wang, and Y. Zhang, "Solar-induced chlorophyll fluorescence and its link to canopy photosynthesis in maize from continuous ground measurements," *Remote Sensing of Environment*, vol. 236, pp. 111420, 2020-01-01, 2020.

[31] J. Wei, Z. Li, A. Lyapustin, J. Wang, O. Dubovik, J. Schwartz, L. Sun, C. Li, S. Liu, and T. Zhu, "First close insight into global daily gapless 1 km PM2.5 pollution, variability, and health impact," *Nature Communications*, vol. 14, no. 1, pp. 8349, 2023-12-15, 2023.

[32] R. Beigaitė, H. Tang, A. Bryn, O. Skarpaas, F. Stordal, J. W. Bjerke, and I. Žliobaitė, "Identifying climate thresholds for dominant natural vegetation types at the global scale using machine learning: Average climate versus extremes," *Global Change Biology*, vol. 28, no. 11, pp. 3557-3579, 2022, 2022.

[33] S. Zhou, B. Medlyn, S. Sabaté, D. Sperlich, I. C. Prentice, and D. Whitehead, "Short-term water stress impacts on stomatal, mesophyll and biochemical limitations to photosynthesis differ consistently among tree species from contrasting climates," *Tree Physiology*, vol. 34, no. 10, pp. 1035-1046, 2014-10-01, 2014.

[34] P. B. Reich, K. M. Sendall, A. Stefanski, R. L. Rich, S. E. Hobbie, and R. A. Montgomery, "Effects of climate warming on photosynthesis in boreal tree species depend on soil moisture," *Nature*, vol. 562, no. 7726, pp. 263-267, 2018-10, 2018.

[35] Y. Zhang, and J. Peñuelas, "Combining Solar-Induced Chlorophyll Fluorescence and Optical Vegetation Indices to Better Understand Plant Phenological Responses to Global Change," *Journal of Remote Sensing*, vol. 3, pp. 0085, 2023-09-19, 2023.

[36] J. Wei, Z. Li, X. Chen, C. Li, Y. Sun, J. Wang, A. Lyapustin, G. P. Brasseur, M. Jiang, L. Sun, T. Wang, C. H. Jung, B. Qiu, C. Fang, X. Liu, J. Hao, Y. Wang, M. Zhan, X. Song, and Y. Liu, "Separating Daily 1 km PM2.5 Inorganic Chemical Composition in China since 2000 via Deep Learning Integrating Ground, Satellite, and Model Data," *Environmental Science & Technology*, 2023-04-28, 2023.

[37] R. Cheng, "Solar-Induced Chlorophyll Fluorescence (SIF): Towards a Better Understanding of Vegetation Dynamics and Carbon Uptake in Arctic-Boreal Ecosystems," *Current Climate Change Reports*, vol. 10, no. 2, pp. 13-32, 2024-04-04, 2024.

[38] W. H. Maes, B. R. Pagán, B. Martens, P. Gentine, L. Guanter, K. Steppe, N. E. C. Verhoest, W. Dorigo, X. Li, J. Xiao, and D. G. Miralles, "Sun-induced fluorescence closely linked to ecosystem transpiration as evidenced by satellite data and radiative transfer models," *Remote Sensing of Environment*, vol. 249, pp. 112030, 2020/11/01/, 2020.

[39] G. Wu, C. Jiang, H. Kimm, S. Wang, C. Bernacchi, C. E. Moore, A. Suyker, X. Yang, T. Magney, C. Frankenberg, Y. Ryu, B. Dechant, and K. Guan, "Difference in seasonal peak timing of soybean far-red SIF and GPP explained by canopy structure and chlorophyll content," *Remote Sensing of Environment*, vol. 279, pp. 113104, 2022-09-15, 2022.

[40] J. Bendig, Z. Malenovský, B. Siegmann, J. Krämer, and U. Rascher, "Comparing methods for solar-induced fluorescence efficiency estimation using radiative transfer modelling and airborne diurnal measurements of barley crops," *Remote Sensing of Environment*, vol. 317, pp. 114521, 2025/02/01/, 2025.

[41] A. Gupta, A. Rico-Medina, and A. I. Caño-Delgado, "The physiology of plant responses to drought," *Science*, vol. 368, no. 6488, pp. 266-269, 2020-04-17, 2020.

[42] K. Hufkens, T. F. Keenan, L. B. Flanagan, R. L. Scott, C. J. Bernacchi, E. Joo, N. A. Brunsell, J. Verfaillie, and A. D. Richardson, "Productivity of North American grasslands is increased under future climate scenarios despite rising aridity," *Nature Climate Change*, vol. 6, no. 7, pp. 710-714, 2016-07, 2016.

[43] C. Frankenberg, C. O'Dell, J. Berry, L. Guanter, J. Joiner, P. Köhler, R. Pollock, and T. E. Taylor, "Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2," *Remote Sensing of Environment*, vol. 147, pp. 1-12, 05/2014, 2014.

[44] Y. Shang, K. Song, F. Lai, L. Lyu, G. Liu, C. Fang, J. Hou, S. Qiang, X. Yu, and Z. Wen, "Remote sensing of fluorescent humification levels and its potential environmental linkages in lakes across China," *Water Research*, vol. 230, Feb 15, 2023.

[45] S.-B. Duan, Y. Lian, E. Zhao, H. Chen, W. Han, and Z. Wu, "A Novel Approach to All-Weather LST Estimation Using XGBoost Model and Multisource Data," *Ieee Transactions on Geoscience and Remote Sensing*, vol. 61, 2023, 2023.

[46] Z. Zhang, W. Jin, R. Dou, Z. Cai, H. Wei, T. Wu, S. Yang, M. Tan, Z. Li, C. Wang, G. Yin, and B. Xu, "Improved Estimation of Leaf Area Index by Reducing Leaf Chlorophyll Content and Saturation Effects Based on Red-Edge Bands," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-14, 2023, 2023.

[47] K. Yan, J. Wang, R. Peng, K. Yang, X. Chen, G. Yin, J. Dong, M. Weiss, J. Pu, and R. B. Myneni, "HiQ-LAI: a high-quality reprocessed MODIS leaf area index dataset with better spatiotemporal consistency from 2000 to 2022," *Earth System Science Data*, vol. 16, no. 3, pp. 1601-1622, 2024-03-26, 2024.

[48] S. Du, X. Liu, J. Chen, W. Duan, and L. Liu, "Addressing validation challenges for TROPOMI solar-induced chlorophyll fluorescence products using tower-based measurements and an NIRv-scaled approach," *Remote Sensing of Environment*, vol. 290, pp. 113547, 05/2023, 2023.

Appendix

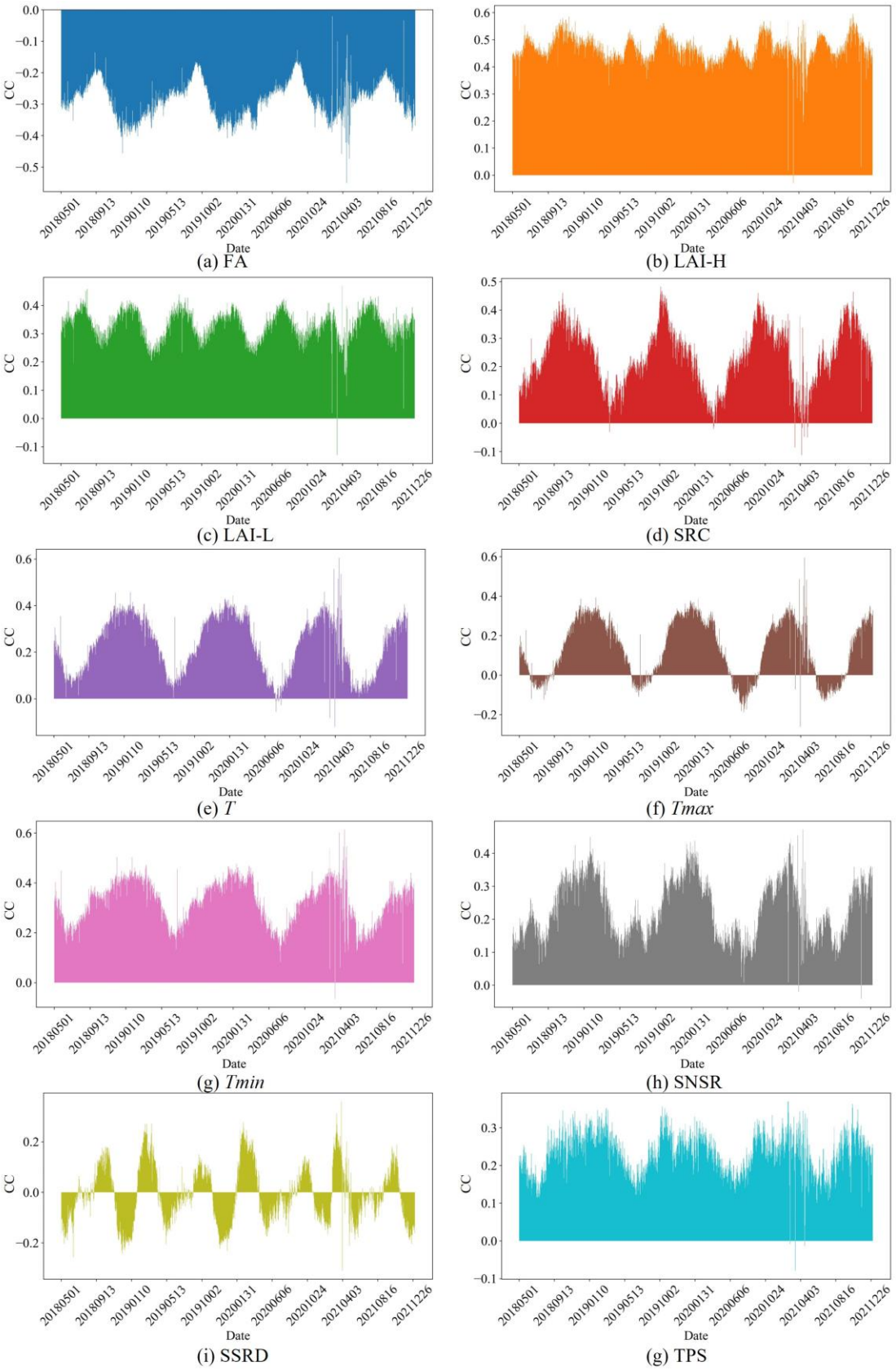

Fig. A1. The CC between SIF and each physical variable.

Jingbo Li received the B.S. and M.S. degrees in surveying and mapping engineering from Henan Polytechnic University, Jiaozuo, China, in 2020 and 2023, respectively. He is currently pursuing the Ph.D. degree with Tongji University, Shanghai, China.

His research interests include multisource and multiscale data fusion, spatiotemporal data reconstruction, and vegetation remote sensing.

Qunming Wang received the Ph.D. degree from the Hong Kong Polytechnic University, Hong Kong, in 2015.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. He was a Lecturer (Assistant Professor) with Lancaster Environment Centre, Lancaster University, Lancaster, U.K., from 2017 to 2018. His 3-year Ph.D. study was supported by the hypercompetitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was awarded as the Outstanding Thesis in the Faculty. He has authored or coauthored over 100 peer-reviewed articles in international journals such as *Remote Sensing of Environment*, *IEEE Transactions on Geoscience and Remote Sensing*, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing, image processing, and geostatistics.

Dr. Wang serves as Associate Editor for *Science of Remote Sensing* and *Photogrammetric Engineering & Remote Sensing*, and was Associate Editor for *Computers and Geosciences* (2017−2020).

Peter M. Atkinson received the Ph.D. degree from the University of Sheffield (NERC CASE award with Rothamsted Experimental Station) in 1990. More recently, he received the MBA degree from the University of Southampton in 2012.

He is currently Distinguished Professor of Spatial Data Science and Executive Dean of the Faculty of Science and Technology at Lancaster University, UK. He was previously Professor of Geography at the University Southampton, where he is currently Visiting Professor. He is also Visiting Professor at the Chinese Academy of Sciences, Beijing and previously held the Belle van Zuylen Chair at Utrecht University, the Netherlands.

The main focus of Peter's research is in remote sensing, geographical information science and spatial (and space-time) statistics applied to a range of environmental science and socio-economic problems. He has published over 400 peer-reviewed articles on these topics in international scientific journals and over 50 refereed book chapters. He has also edited 14 journal special issues and eight books. Peter is currently listed as an ISI highly-cited researcher.

Peter has received several awards for his research including the Cuthbert Peek Award of the Royal Geographical Society-Institute of British Geographers and the Peter Burrough Award of the International Spatial Accuracy Research Association. He is also a Fellow of the Learned Society of Wales.

Professor Atkinson is Editor-in-Chief of *Science of Remote Sensing*, a sister journal of *Remote Sensing of Environment*. He is also Associate Editor for *Environmetrics*.