



# Global Retrieval of 24-hourly Solar-Induced Chlorophyll Fluorescence and Evapotranspiration from OCO-2, OCO-3 and ECOSTRESS over 1982–2022

Zhuoying Deng<sup>1,2,#</sup>, Tingyu Li<sup>1,2,#</sup>, Jinghua Chen<sup>3</sup>, Shaoqiang Wang<sup>1,2,3,4,\*</sup>, Kun Huang<sup>5,6</sup>, Peng Gu<sup>1,2</sup>, Haoyu Peng<sup>1,2</sup>, Zhihui Chen<sup>1,2</sup>

<sup>1</sup>Hubei Key Laboratory of Regional Ecology and Environmental Change, School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei, China

<sup>2</sup>Engineering Technology Innovation Center for Intelligent Monitoring and Spatial Regulation of Land Carbon Sinks, Ministry of Natural Resources, Wuhan, Hubei, China

<sup>3</sup>Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, China

<sup>4</sup>College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China

<sup>5</sup>Zhejiang Zhoushan Island Ecosystem Observation and Research Station, School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China

<sup>6</sup>Institute of Eco-Chongming (IEC), East China Normal University, Shanghai, China

<sup>#</sup>These authors contributed equally to this work.

\*Correspondence to: Shaoqiang Wang (sqwang@igsnr.ac.cn)

**Abstract.** Solar-induced chlorophyll fluorescence (SIF) and evapotranspiration (ET) have been widely recognized as proxies for carbon gain and water loss at the ecosystem level. However, most SIF and ET products on the global scale are generally at coarse temporal resolutions of at least one day, which limits their ability to characterize diurnal carbon and water cycles. In this study, we extended the spatiotemporal scale of satellite SIF (from OCO-2 and OCO-3) and ET (from ECOSTRESS) data using machine learning methods, resulting in a global hourly SIF and ET dataset (HOUR\_SIF<sub>OCO</sub> and HOUR\_ET<sub>ECO</sub>) spanning from 1982 to 2022, with a spatial resolution of 0.1°. Our product also provides photosystem-level SIF derived from direct estimation and simulation of Soil Canopy Observation, Photochemistry and Energy fluxes (SCOPE) model, aiming to offer a more accurate description of photosynthesis. Our satellite-derived products show good correlations with in-situ flux tower measurements from the FLUXNET2015 community (hourly-scale median  $R^2$  for SIF: 0.72, and ET: 0.53; daily-scale median  $R^2$  for SIF: 0.73, and ET: 0.63). Globally, our product shows good consistency with popular SIF and ET gridded products: the mean proportions of pixels with monthly  $R^2$  exceeding 0.7 are 69.5% and 68.1% when compared with four popular products, respectively. The causal-based attribution analysis revealed significant spatial heterogeneity in the lagged effects of different environmental factors on SIF, ET, and water use efficiency based on SIF and ET on the global scale. Overall, our dataset will provide new insights for monitoring the diurnal variations of carbon and water cycles and deepen our understanding of their changes over the past 40 years. The global hourly SIF and ET dataset (1982–2022) at 0.1° spatial resolution produced in this study is available at <https://doi.org/10.57760/sciencedb.ecodb.00177> (Deng et al., 2025b).



## 1 Introduction

35 Carbon and water cycles are the key processes within terrestrial ecosystems (Piao et al., 2020; Regnier et al., 2022). In this cycle, photosynthesis of vegetation forms the gross primary production (GPP), which serves as a major carbon sink in the ecosystem (Miller et al., 2023; Yu et al., 2022); transpiration from vegetation and evaporation from soil and water (referred to as evapotranspiration, ET) is essential for regulating surface energy balance (Cheng et al., 2017; Yang et al., 2023). Therefore, a deeper understanding of photosynthesis and ET is needed especially under global climate change (Rockström et al., 2023; 40 Wankmüller et al., 2024; Zhang K. et al., 2024). However, accurately quantifying these processes on the global scale still poses severe challenges (Fuentes et al., 2024; Lai et al., 2024; Ryu et al., 2019).

In recent years, huge advancements in remote sensing have made it possible to monitor carbon and water cycles on the global scale (Huang et al., 2018; Rodell et al., 2023; Xiao et al., 2019). Solar-induced chlorophyll fluorescence (SIF) and ET are two 45 representative advancements (Xiao et al., 2021). SIF is the faint signal re-emitted by vegetation after absorbing light energy during photosynthesis (Jiao et al., 2019; Porcar-Castell et al., 2021; Van der Tol et al., 2014), which addresses the limitations of traditional optical vegetation indices and has shown great potential in estimating GPP (Sun et al., 2023a), monitoring environmental stress (Mohammed et al., 2019) and plant phenology (Huang et al., 2023). The integration of SIF and ET can further enhance our understanding of plants' functional characteristics, such as analyzing vegetation's water use strategies 50 under various environments through water use efficiency (Zhang Z. et al., 2023a).

Despite numerous studies leveraging satellite-based SIF and ET data for large-scale ecosystem monitoring, data availability remains a major limitation (Elnashar et al., 2021; Sun et al., 2023b). For SIF, this limitation arises primarily from the lack of satellites specifically designed to detect SIF due to its faint signal, resulting in current observations being characterized by 55 coarse spatiotemporal resolution and sparse sampling (Quiros-Vargas et al., 2022). To overcome this, machine learning models and light use efficiency (LUE) models have been developed to upscale low-resolution and temporally discontinuous SIF remote sensing products into high-resolution and spatiotemporally continuous datasets, such as CSIF (Zhang Y. et al., 2018), GOSIF (Li and Xiao, 2019), and SIF005 (Wen et al., 2020). The temporal resolution of these global products typically ranges from 4 days to 1 month, with the highest resolution reaching one day (such as TROPOMI SIF (Lorente et al., 2021)), which can meet 60 the commonly need for monitoring seasonal vegetation photosynthesis and estimating GPP. However, these products typically represent fixed-time SIF values, determined by satellite overpass schedules—for instance, around 9:30 AM local time for the MetOp-A satellite (August et al., 2012), and approximately 1:30 PM local time for TROPOMI (Lorente et al., 2021), OCO-2 (Sun et al., 2018), and TanSat satellites (Liu et al., 2018). Most products convert instantaneous SIF values to daily averages using scaling factors such as the ratio of the instantaneous cosine of the solar zenith angle ( $\cos(\text{SZA})$ ) to its daily average (Li 65 and Xiao, 2019) or ratios derived from atmospheric radiative transfer models (Zhang Y. et al., 2018). However, all these methods fail to capture the diurnal variations in SIF. Compared to SIF datasets, spatiotemporally expanded ET datasets are



even scarce (Leng et al., 2024), limiting the ability to monitor diurnal changes in water fluxes and the joint application of SIF and ET (e.g. water use efficiency) on the hourly scale.

70 Recent satellite missions, such as Orbiting Carbon Observatory-3 (OCO-3) and the Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) have provided new opportunities to enhance spatiotemporal resolution for SIF and ET (Xiao et al., 2021). OCO-3 and ECOSTRESS, mounted on the International Space Station, enable observations of the same region at different times of the day (Taylor et al., 2020). This unique feature offers the potential to construct diurnal SIF and ET datasets for specific regions (Xiao et al., 2021). Recent studies have utilized OCO-3 SIF and  
75 ECOSTRESS ET data to investigate the Amazon rainforest's response to global warming (Zhang Z. et al., 2023a; Zhang Z. et al., 2023b). However, these missions provide SIF and ET data with sparse or coarse spatial sampling, necessitating spatiotemporal upscaling for broader application in terrestrial ecosystem monitoring (Zhang Y. et al., 2023). Recent studies have started to address this issue and have attempted to produce hourly-scale SIF or ET datasets. However, current methods still face limitations in spatial resolution and extent ( $0.5^\circ$  or non-global scale) due to computational power or data availability,  
80 which restrict their broader and more in-depth applications (Deng et al., 2025a; Jeong et al., 2024; Zhang Z. et al., 2023b). Specifically, producing the dataset at an hourly resolution would lead to an order-of-magnitude increase in computational demand and require more efficient algorithms. In addition, most SIF retrieval methods are based on satellite-observed optical reflectance (from satellites such as MODIS), which limits the temporal range and continuity of the input data, thereby affecting the production of SIF datasets. Finally, some studies have pointed out that satellite-observed SIF may be greatly affected by  
85 hotspot effects, resulting in substantial measurement biases (Zeng et al., 2023). Converting satellite-derived canopy SIF into photosystem level SIF may help enhance its correlation with GPP (Zhang Z. et al., 2023b), while most SIF datasets do not take this into account.

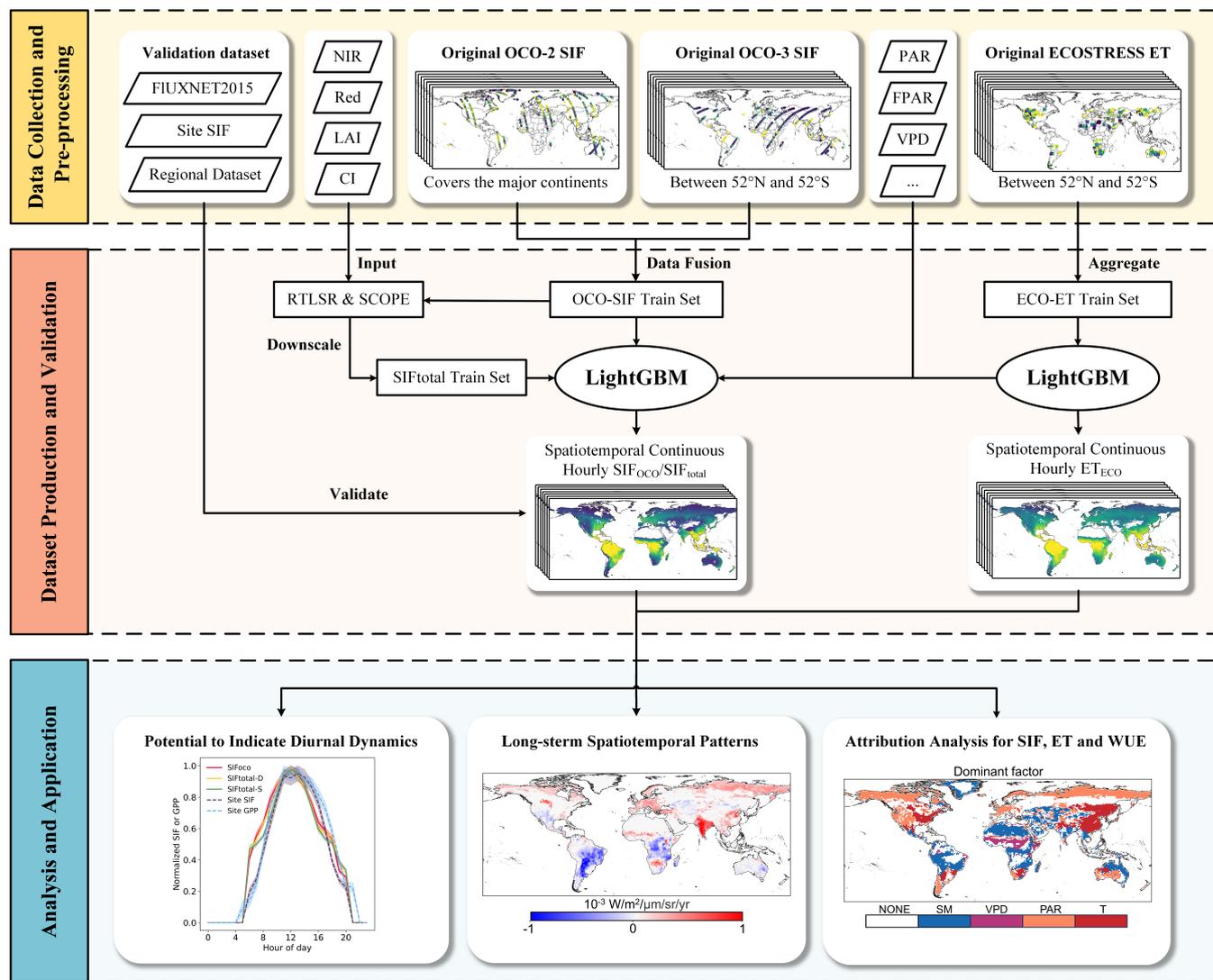
In this study, we aim to develop a long-term (from 1982 to 2022), high-temporal-resolution (one hour) global SIF and ET  
90 dataset with an efficient algorithm based on the OCO-2 and OCO-3 satellites and the ECOSTRESS mission (HOUR\_SIF<sub>OCO</sub> and HOUR\_ET<sub>ECO</sub>). We also attempted to convert satellite-observed SIF into the total SIF emitted by vegetation, to achieve a more accurate modeling of actual photosynthesis. All output data are produced with a spatial resolution of  $0.1^\circ$  and validated by site-level observations and popular regional-scale products. Additionally, we conducted a preliminary long-term analysis to explore how carbon and water fluxes varied and how they were regulated by environmental factors in the past 40 years.

## 95 **2 Data and methodology**

The technical workflow of this study, as shown in Fig. 1, involves several critical steps. First, raw satellite observation data are collected and pre-processed (Sect. 2.1 to 2.3). Next, photosystem-level SIF is estimated (Sect. 2.4), and the spatiotemporal



upscaling model is constructed (Sect. 2.5). Following the production of our dataset, a rigorous validation process is conducted before being used for attribution analysis (Sect. 2.6).



**Figure 1.** Technical flowchart of this study. The upper section displays the main data used in this study, the middle section represents the production portion of the dataset, and the lower section shows the analysis and application based on the production dataset.

## 2.1 Data for training and production

### 2.1.1 OCO satellite SIF data

OCO-2 and OCO-3 satellites are designed to explore carbon dynamics on the Earth's surface. Although their primary mission is to measure atmospheric CO<sub>2</sub> concentrations rather than detect SIF, their advanced sensors allow them to capture SIF signals



with some of the highest spatial resolutions among current SIF-focused satellites (1.3 km × 2.25 km for OCO-2 and 1.6 km × 2.2 km for OCO-3). This high spatial resolution provides a strong foundation for further spatial upscaling. OCO-2 is onboard  
110 a sun-synchronous orbit satellite, with an overpass time of approximately 1:30 PM local time and a revisit cycle of about 16 days. In contrast, OCO-3 is mounted on the International Space Station, resulting in less predictable revisit cycles and overpass times (Doughty et al., 2021). This unique characteristic of OCO-3 enables monitoring of SIF at various times of the day, enabling the study of diurnal SIF variations (Xiao et al., 2021). We utilized the latest version (V11r) of the OCO-2 and OCO-3 SIF datasets for our study. Since both datasets were derived using nearly identical retrieval algorithms and quality control  
115 standards, the SIF data provided by OCO-2 and OCO-3 exhibit strong consistency. Both satellites provide SIF values at the O<sub>2</sub>-A band (757 nm and 771 nm), but the SIF signal at 757 nm is stronger and has relatively lower uncertainty (Doughty et al., 2021; Sun et al., 2018). This makes it particularly advantageous for monitoring sparsely vegetated regions, and thus the 757 nm SIF values were adopted in our study. Although the official OCO team also provides SIF values at the 740 nm band (near the emission peak in the near-infrared spectrum) calculated using measurements at 757 nm and 771 nm, these values are not  
120 directly retrieved and inherent uncertainties from both bands. Consequently, we excluded them from our analysis, consistent with common practices for OCO SIF data usage (Li X. et al., 2020; Zhang Y. et al., 2023).

To ensure the quality of our training samples, we implemented a rigorous quality control process. Specifically, (1) only observations with a quality flag of 0 (indicating the highest overall retrieval quality) were used; (2) the strictest cloud fraction  
125 thresholds were applied to minimize the presence of clouds in the observed areas, even though clouds typically have minimal impact on SIF retrieval (Frankenberg et al., 2012); (3) only nadir mode observations were included to eliminate variations due to viewing zenith angle (VZA), ensuring maximum signal-to-noise ratio and highest spatial resolution; (4) following official recommendations, negative SIF values were retained if their adjusted values obtained by adding two times the uncertainty were greater than zero (Doughty et al., 2021); To further reduce the impact of observational uncertainties, we aggregated the  
130 raw SIF observations into 0.1° geographic grid cells and retained only those grid cells with at least 15 observations. This aggregation reduces uncertainty to approximately  $1/\sqrt{15}$  of the original value (Frankenberg et al., 2014). For the analysis, we selected OCO-2 SIF data spanning September 2014 to March 2024 and OCO-3 SIF data spanning August 2019 to November 2023. Due to the availability of model-driven data (Sect. 2.1.3), only data up to 2023 were used for training.

### 2.1.2 ECOSTRESS satellite ET data

135 We used the ECO3ETPTJPLv001 instantaneous latent heat flux (ET<sub>inst</sub>) product from the ECOSTRESS (ECOsysteM Spaceborne Thermal Radiometer Experiment on Space Station) mission (Fisher et al., 2008). Mounted on the International Space Station, ECOSTRESS is a high-precision thermal radiation imaging system designed to monitor evapotranspiration at the Earth's surface. Operating in an inclined polar orbit, the mission provides global coverage between latitudes 52°N and 52°S. By measuring vegetation surface temperature, ECOSTRESS provides valuable insights into vegetation water  
140 requirements and their responses to climate stress. It offers ET data at a spatial resolution of 70 meters and a various temporal



resolution (Anderson et al., 2021), which performs well compared with in situ measurements at 82 EC sites globally (Fisher et al., 2020). We used only high-quality ET data for training and made every effort to ensure the training distribution covers the globe. Compared to SIF data, ET data are at a much higher spatial resolution per observation. Consequently, we did not aggregate the ET data in the same way as SIF. Instead, we retained grid cells where ET detection coverage exceeded 99%  
145 within a 0.1° grid cell, ensuring the ET within these cells accurately represented detected ET.

### 2.1.3 Auxiliary data for production

We selected the hourly meteorological dataset provided by ERA5-Land (Muñoz-Sabater et al., 2021), which has a spatial resolution of 0.1°, matching the resolution of our training label data and output products. To estimate photosynthetically active radiation (PAR), we used the hourly surface solar radiation downwards in ERA5-Land, multiplying it by 0.46. The 2m  
150 temperature and 2m dewpoint temperature were used to estimate temperature and vapour pressure deficit (VPD). Additionally, the soil moisture from the top three layers, weighted appropriately, was incorporated into our study.

We utilized the Global Inventory Monitoring and Modeling System (GIMMS) FPAR product, GIMMS FPAR4g (Zhao et al., 2024), which features a high spatial resolution of 1/12° and a half-month temporal resolution. As the latest FPAR product in  
155 this series, GIMMS FPAR4g demonstrates excellent spatiotemporal consistency over the 41 years from 1982 to 2022, which is crucial for constructing long-term SIF and ET datasets.

In addition, we utilized the MODIS land cover product MOD12Q1, which is at 500m spatial resolution. We reclassified it to 0.1° grid cells based on the dominant plant types. Since MOD12Q1 does not provide products before the 21st century, we  
160 utilized the most frequent land cover categories from 2001 to 2003 as proxies for the land cover types before 2001, minimizing potential errors (Li M. et al., 2023). Finally, we used the National Oceanic and Atmospheric Administration (NOAA) DEM data to simulate the potential impact of elevation on atmospheric pressure (Cheng et al., 2025).

### 2.2 Data for validation

We utilized site-level observational data from FLUXNET2015 to validate the accuracy of our products at an hourly timescale.  
165 Based on data quality, we selected 136 sites for SIF validation and 146 sites for ET validation (Table S1). Gross Primary Production derived from the Daytime Partitioning Method (GPP\_DT\_VUT\_REF) and latent heat flux (LE\_F\_MDS) were employed to validate SIF and ET, respectively, while only high-quality data with QC flags less than 2 were included. Although site-scale validation inherently faces spatial scale mismatch issues and may reduce validation accuracy, we directly paired and validated our products with the 0.1° grid values corresponding to the site locations due to the lack of effective correction  
170 methods. Since FLUXNET2015 does not provide SIF observations, we selected SIF data from eight publicly available studies to evaluate the capacity of predicted SIF to capture diurnal variations (Table S2). Furthermore, we selected four widely recognized gridded SIF products—CSIF (Zhang Y. et al., 2018), GOSIF (Li and Xiao, 2019), LT\_SIFc (Wang et al., 2022),



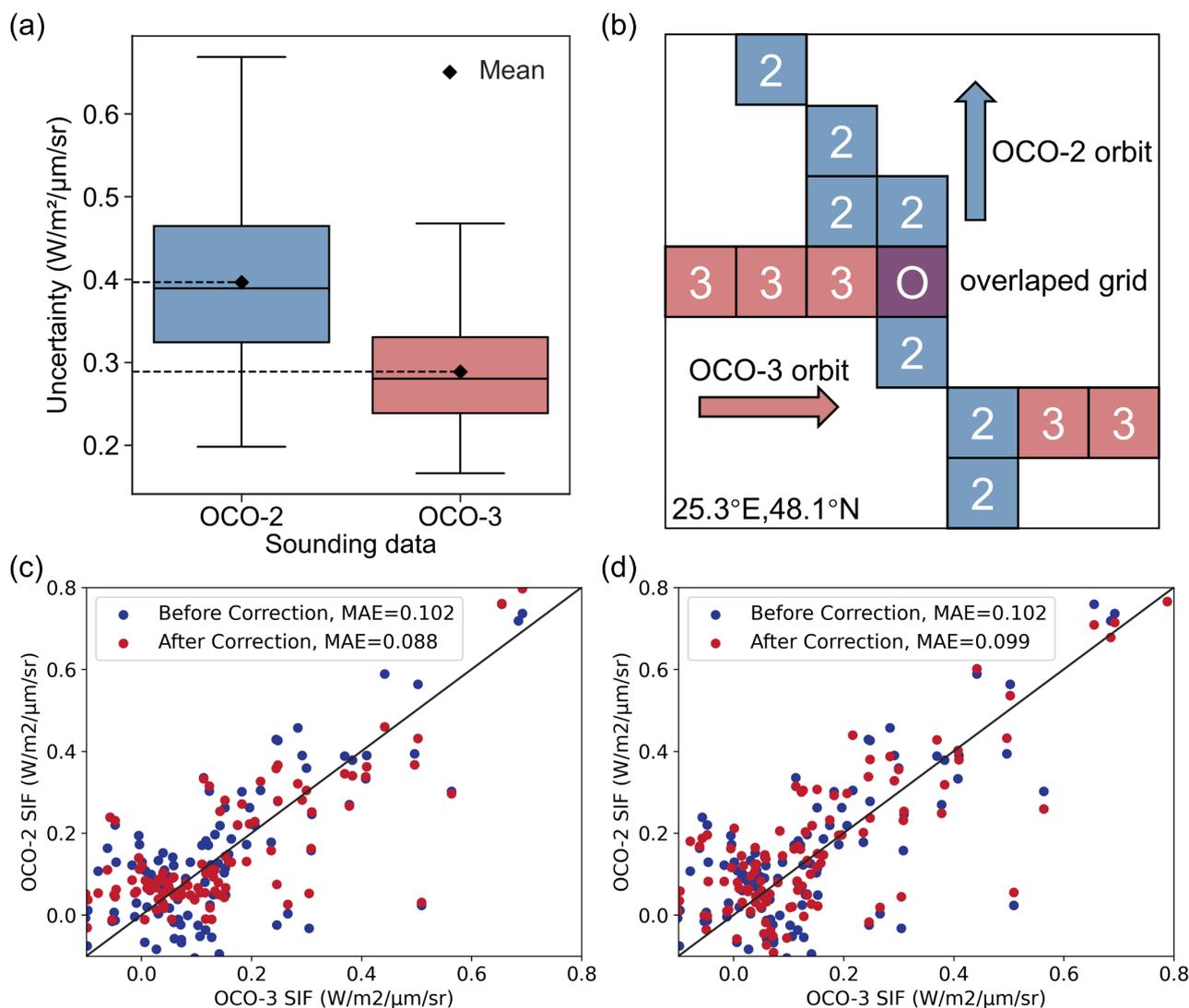
and SIF005 (Wen et al., 2020)—along with four gridded ET products—BESSv2.0 (Li B. et al., 2023b), BEPS-DP (Leng et al., 2024), REA (Lu et al., 2021), and ERA5-Land (Muñoz-Sabater et al., 2021) to validate our products. All comparisons were  
175 conducted at a spatial resolution of 1° and at a monthly timescale.

### 2.3 Data fusion of OCO-2 and OCO-3 SIF sounding data

Although OCO-2 and OCO-3 SIF sounding data are derived using nearly identical retrieval algorithms, differences exist due to variations in observation altitudes and the new Pointing Mirror Assembly (PMA) provided for OCO-3 on the International Space Station, which affect radiometric reception characteristics (Eldering et al., 2019). Consequently, SIF data from the two  
180 satellites cannot be directly combined (Deng et al., 2025a). To address this, we employed a LightGBM-based (Light Gradient Boosting Machine, Ke et al., 2017) fusion method to establish a quantitative relationship between OCO-2 and OCO-3 SIF data using observational parameters from OCO-2. This relationship was then applied to all OCO-2 SIF grid cells, ensuring their characteristics align with those of OCO-3:

$$SIF_{OCO-3} = F(SIF_{OCO-2}, Param1_{OCO-2}, Param2_{OCO-2} \dots) \quad (1)$$

185 Specifically, we used 20 variables—including OCO-2 SIF data and 17 observational parameters such as solar zenith angle and latitude/longitude—as input features for the quantitative correction model. The full list of variables is detailed in Table S3. Each grid cell's properties were represented by the average values of all observation points within that cell. For training samples, we included only grid cells where OCO-2 and OCO-3 observations had identical geographical locations and observation times (to the hour), ensuring comparable observation conditions (Fig. 2).



190

**Figure 2** (a) The original uncertainty distribution map of OCO SIF. (b) A schematic of the fusion of OCO-2 and OCO-3 SIF (UTC: August 18, 2019, 11:00 AM; Location: 25.3°E, 48.1°N). The blue represents SIF grid cells aggregated from OCO-2 detection points, the red represents OCO-3, and the purple “O” indicates the overlapping grid cells between both, which serve as our training samples. (c) Comparison of OCO-2 and OCO-3 grid cells before and after fusion based on the LightGBM model (on the test set). (d) Fusion results based on a multiple linear regression model.

195

We randomly selected 70% of the fusion samples as training data and used the optimized parameters to train the LightGBM model with five-fold cross-validation. On the test set, our fusion approach reduced the mean absolute error (MAE) between OCO-2 and OCO-3 SIF grid-cell values from 0.102 W/m<sup>2</sup>/μm/sr to 0.088 W/m<sup>2</sup>/μm/sr, compared to a linear correction model which only reduced the error to 0.099 W/m<sup>2</sup>/μm/sr (see Fig. 2c, d). The average MAE of raw SIF soundings at the observation

200



point level is 0.358 W/m<sup>2</sup>/μm/sr (Fig. 2a). After grid aggregation, the theoretical uncertainty is approximately 0.0924 W/m<sup>2</sup>/μm/sr (calculated as 0.358√15, Frankenberg et al., 2014). This indicates that the inherent uncertainty in SIF observations accounts for most of the residual error after fusion, demonstrating the fusion process's effectiveness. Following data fusion, we obtained 4,661,615 SIF grid data points, of which 2,881,096 (approximately 62%) came from OCO-2, sharply increasing the sample size. In addition, SIF retrievals from OCO-2 effectively compensate for the observational gaps of OCO-3 in high-latitude regions, as the latter is limited to observations below 53° latitude.

## 2.4 Estimation of photosystem-level SIF

We employed two methods to convert canopy-level SIF (SIF<sub>OCO</sub>) observed by OCO-2 and OCO-3 into photosystem-level SIF (SIF<sub>total</sub>), both of which are included in our final product. SIF<sub>total</sub> refers to the fluorescence emitted on the microscopic scale (i.e., from Photosystem I and Photosystem II) during photosynthesis. This fluorescence undergoes reabsorption within leaf tissues and multiple scattering and absorption processes within the canopy before reaching the canopy top, where it can be observed. To achieve this conversion, it is necessary to estimate the escape ratio of SIF from the photosystem level to the canopy top. First, we applied a direct estimation method based on radiative transfer models (Yang and van der Tol, 2018; Zeng et al., 2018):

$$f_{esc} = \frac{NIR_V}{\pi \times i_0 \times K_\lambda} \quad (2)$$

$$SIF_{total} = \frac{SIF_{OCO}}{f_{esc}} \quad (3)$$

Here, NIR<sub>v</sub> represents the near-infrared reflectance of vegetation, which can be obtained by multiplying the Normalized Difference Vegetation Index (NDVI) by the near-infrared reflectance (NIR). The denominator includes π, which is used to convert directional fluorescence to canopy hemispherical fluorescence. K<sub>λ</sub> refers to the ratio of leaf albedo to the escape ratio of fluorescence from the photosystem to the leaf surface, with a value of 1.2 adopted in this study (Zhang Z. et al., 2021). i<sub>0</sub> denotes canopy interception:

$$i_0 = 1 - \exp\left(\frac{-G(\theta) \times LAI \times CI}{\cos(SZA)}\right) \quad (4)$$

$$G(SZA) = \phi_1 + \phi_2 \times \cos(SZA) \quad (5)$$

$$\phi_1 = 0.5 - 0.663\chi_L - 0.33\chi_L^2 \quad (6)$$

Here, SZA represents the solar zenith angle. χ<sub>L</sub> refers to the departure of leaf angles from a random distribution. In this study, χ<sub>L</sub> values for different vegetation types were obtained from the Community Land Model 4.5 (Oleson et al., 2023). LAI represents the Leaf Area Index, with data sourced from the MODIS MCD15A3H product. CI is the clumping index, for which



we utilized a set of seasonal CI products (Fang et al., 2021; Wei et al., 2019). It is important to note that OCO-2 and OCO-3 do not directly provide bidirectional reflectance in the red and near-infrared bands for the observed regions. To eliminate the influence of observation geometry, we used the RossThick-LiSparseR (RTLSR) BRDF model to retrieve the corresponding reflectance values (Lucht et al., 2000):

$$R(\theta_i, \theta_v, \varphi, \Lambda) = f_{iso}(\Lambda)K_{iso} + f_{vol}(\Lambda)K_{vol}(\theta_i, \theta_v, \varphi) + f_{geo}(\Lambda)K_{geo}(\theta_i, \theta_v, \varphi) \quad (7)$$

Here,  $R(\theta_i, \theta_v, \varphi, \Lambda)$  represents the bidirectional reflectance distribution function (BRDF) at wavelength  $\Lambda$ .  $\theta_i$ ,  $\theta_v$  and  $\varphi$  are the solar zenith angle, sensor zenith angle, and relative azimuth angle (the difference between the solar azimuth angle and the observation azimuth angle), respectively.  $K_{iso}$ ,  $K_{vol}$  and  $K_{geo}$  are the isotropic scattering, volumetric scattering, and geometric-optical scattering kernels, respectively. We obtained daily kernel parameters from the MCD43A1 product and used satellite-provided observation parameters to drive the RTLSR model. All data retrieved from MODIS were of the highest quality level. The  $SIF_{total}$  derived from this method is referred to as  $SIF_{total-D}$ .

We also employed a newly proposed method to estimate the canopy escape ratio on the satellite scale (Li H. et al., 2024). Specifically, we utilized the Soil Canopy Observation, Photochemistry and Energy fluxes (SCOPE) model to simulate escape probabilities under various scenarios (Van der Tol et al., 2009). In version 2.1 of the SCOPE model, it is possible to simulate SIF transmission under different conditions by using varied input datasets (Yang et al., 2021). In this study, we used the input parameters listed in Table 1 to simulate as many scenarios as possible (Li H. et al., 2024; Liu et al., 2019), resulting in a total of 1,469,664 samples. Among these, 70% were randomly selected as the training set, while the remaining data were used as the test set. We employed a LightGBM model (detailed in Sect. 2.5) to establish a quantitative relationship between red-band reflectance, near-infrared reflectance, NIRv, SZA, and LAI with escape ratio. Ultimately, we achieved an  $R^2$  close to 1 on the test set (Fig. S1). This model was then applied to the extrapolation of satellite-observed SIF. To match SCOPE-simulated reflectance with MODIS satellite-provided reflectance as closely as possible, we used bidirectional reflectance outputs from SCOPE at 648 nm and 858 nm for training. Additionally, we used reflectance outputs from the RTLSR model to drive our framework, enabling the conversion of  $SIF_{OCO}$  to photosystem-level  $SIF_{total}$ . The  $SIF_{total}$  derived from this method is referred to as  $SIF_{total-S}$ .

**Table 1:** Input parameters in SCOPE model for simulating different scenarios of SIF transmission.

Parameter	Symbol	Values
Leaf Chlorophyll Content	Cab	20, 50, 80 ( $\mu\text{g}/\text{cm}^2$ )
Maximum Carboxylation Capacity	Vcmax25	20, 60, 100 ( $\mu\text{mol}/\text{m}^2 \cdot \text{s}$ )
Soil Spectra	/	Column number [1, 2, 3]
Leaf Area Index	LAI	0.5–7 with steps of 0.5



Leaf Angle Distribution	(LIDFa, LIDFb)	(1, 0), (-1, 0), (0, -1), (0, 1), (-0.35, -0.15), (0, 0)
Shortwave Incoming Radiation	Rin	100, 400, 800 (W/m <sup>2</sup> )
Air Temperature	Ta	15, 25, 35 (°C)
Solar Zenith Angle	SZA	0–70 with steps of 10
View Zenith Angle	VZA	0, 0.2, 0.5
Relative Azimuth Angle	RAA	0, 90, 180

## 2.5 Description of continuous spatiotemporal scale-up model

255 We used an efficient machine learning model, LightGBM, to produce our data products. The LightGBM model is capable of efficiently handling large-scale datasets while capturing complex nonlinear relationships and maintaining high accuracy (Ke et al., 2017). As a result, LightGBM has been widely applied in recent years in geoscience and ecological studies (Guo et al., 2023; Li B. et al., 2023a; Shen et al., 2022). We used the merged SIF and ET grid cells as prediction labels and selected different variables for modelling SIF and ET. Specifically, for SIF, we relied on the following light use efficiency model:

$$260 \quad SIF = PAR \times FPAR \times \phi_f \times \varepsilon \quad (8)$$

In this model, PAR represents photosynthetically active radiation, FPAR is the fraction of PAR absorbed by vegetation,  $\phi_f$  denotes fluorescence efficiency, and  $\varepsilon$  is the canopy escape ratio of SIF. PAR and FPAR can be represented by the ERA5-Land and GIMMS FPAR products.  $\phi_f$  primarily depends on photosynthetic physiological factors and has a strong correlation with environmental variables such as temperature and moisture. The canopy escape ratio is closely related to canopy structure and geometry. Therefore, we ultimately selected land cover types (Landcover), the cosine of the solar zenith angle ( $\cos(SZA)$ ), absorbed photosynthetically active radiation ( $APAR = PAR \times FPAR$ ), temperature ( $T_{2M}$ ), soil moisture (SM), vapour pressure deficit (VPD), latitude and longitude (Lon and Lat), day of year (DOY), and elevation (DEM) as the input variables for the LightGBM model:

$$SIF = f(Landcover, APAR, T_{2M}, SM, VPD, Lon, Lat, DOY, DEM) \quad (9)$$

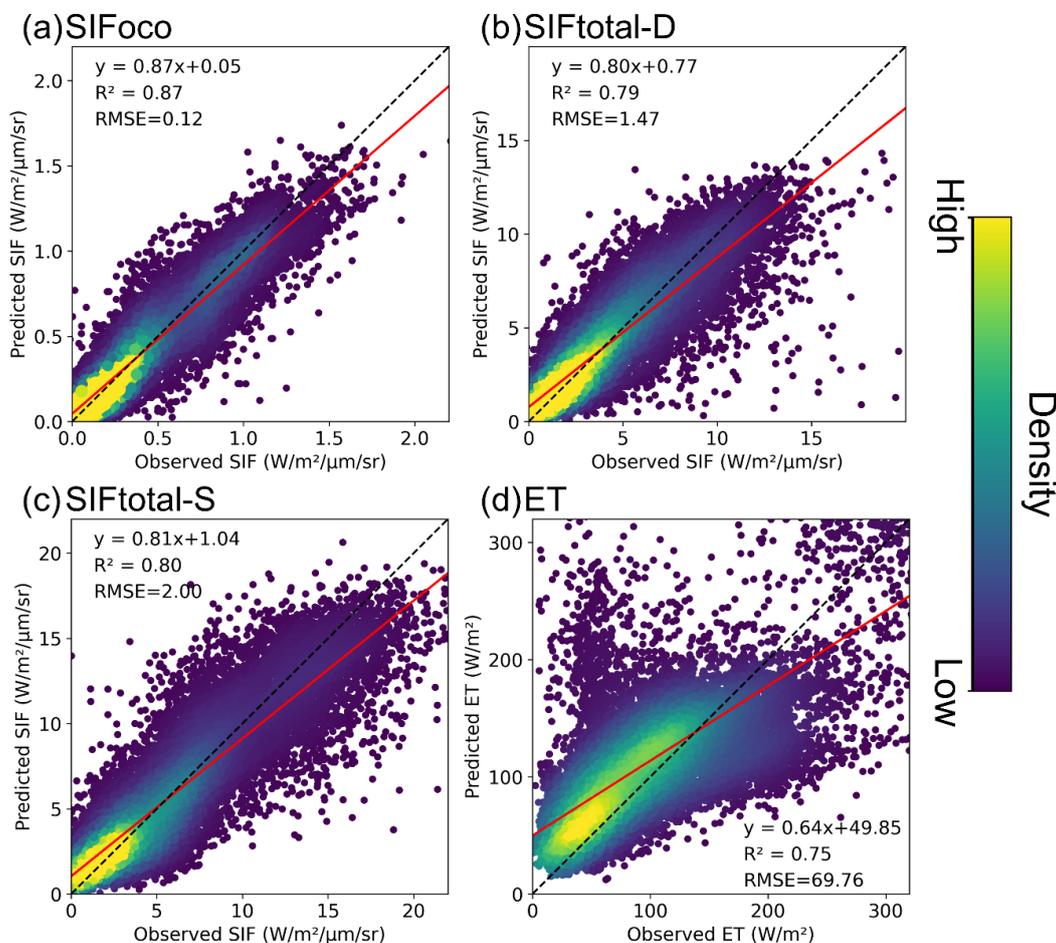
270 When considering the input variables for ET, we relied on a well-known ET calculation model, the Penman-Monteith equation (Allen et al., 2006):

$$ET = \frac{0.408\Delta(R_n - G) + r \frac{900}{T_{2M} + 273} u_2 VPD}{\Delta + r(1 + 0.34u_2)} \quad (10)$$



275 In this model,  $R_n$  represents the net radiation flux density at the surface,  $G$  is the sensible heat flux density from the surface to the soil,  $r$  denotes the psychrometric constant, and  $u_2$  is the wind speed at 2 meters above the surface. Given data availability, most of the variables used to predict ET are almost similar to those used for SIF. The main difference is that we add the ratio of vapour pressure deficit to temperature (i.e.,  $VPD/T_{2M}$ ) as the model-driven data according to Eq. 10.

280 Finally, we used a Bayesian optimization-based hyperparameter tuning framework, Optuna (Akiba et al., 2019), to obtain the optimal combination of LightGBM parameters, balancing accuracy and prediction time. To mitigate the negative effects of spatial autocorrelation and overfitting, we split the quality-controlled dataset into three parts: training set, validation set, and test set. The training and validation sets were used to optimize parameters and adjust model performance within the Optuna framework, while the test set was reserved for final performance evaluation. For  $SIF_{OCO}$  and  $SIF_{total}$ , we selected data from 2021 as the validation set, 2022 data as the test set, and data from other years as the training set. For the ET data, which covered fewer years (2018–2022), we chose odd months of 2022 as the validation set, even months of 2022 as the test set, and the remaining data as the training set. We performed 100 iterations of parameter tuning for training the  $SIF_{OCO}$  and  $SIF_{total}$  models and 1000 iterations for the ET model due to more complex input data and fewer training samples. The final model achieved an  $R^2$  of 0.87 (for  $SIF_{OCO}$ ), 0.79 (for  $SIF_{total-D}$ ), 0.80 (for  $SIF_{total-S}$ ), and 0.75 (for ET) on the independent test dataset (Fig. 3). For SIF, both model training and subsequent production are conducted at a spatial resolution of  $0.1^\circ$ , whereas the training phase for ET is performed at a spatial resolution of  $0.05^\circ$  to increase the number of samples.



290

**Figure 3** The accuracy validation of the (a) SIF<sub>oco</sub>, (b) SIF<sub>total-D</sub> (directly calculated SIF<sub>total</sub>), (c) SIF<sub>total-S</sub> (SIF<sub>total</sub> derived from SCOPE simulations), and (d) ET prediction models on the test set. The black dashed line represents the 1:1 line, while the red line represents the linear regression line. The points are coloured from deep blue to yellow, indicating increasing point density. A random selection of 20,000 sample points was used for plotting.

295

In addition to using the LightGBM model for the spatiotemporal scaling of SIF and ET data, we also applied it for the fusion of OCO-2 and OCO-3 SIF data (see Sect. 2.3) and for the conversion from SIF<sub>oco</sub> to SIF<sub>total-S</sub> based on the SCOPE simulation (see Sect. 2.4). The parameter optimization for these models was also conducted within the Optuna parameter optimization framework.

## 300 2.6 Attribution analysis of SIF, ET, and water use efficiency

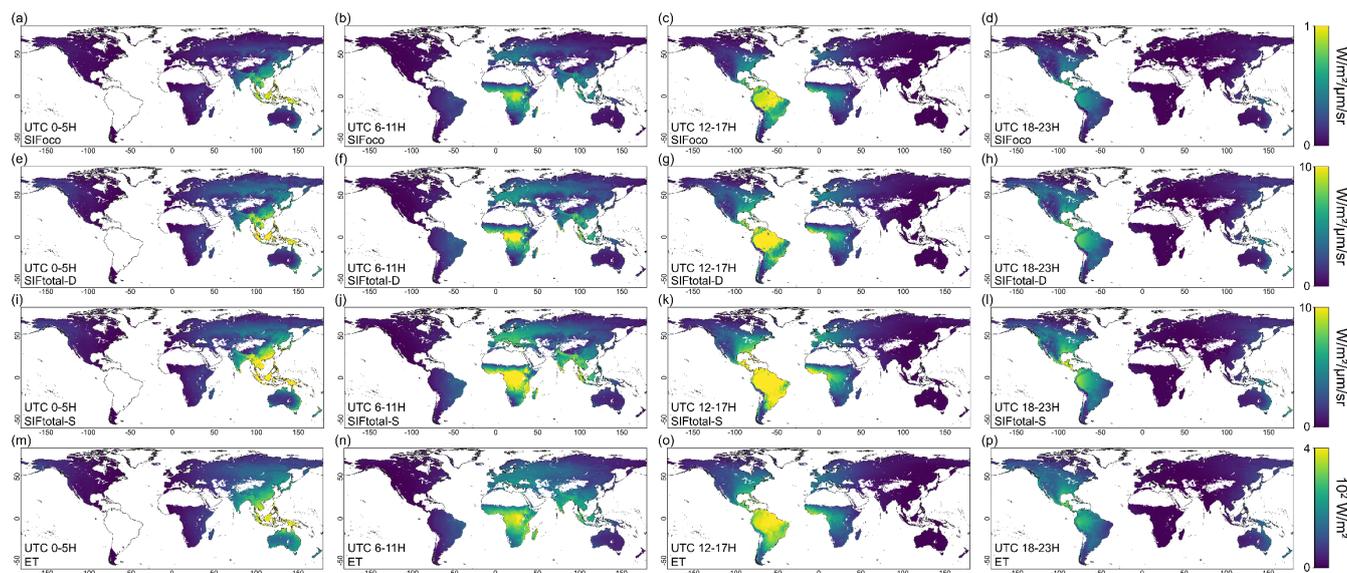
Based on daily-scale SIF, ET, and environmental variables (PAR, VPD, SM, and air temperature), we investigated the influence of environmental factors on the regulation of vegetation photosynthesis and transpiration. We applied a novel causal



analysis algorithm, SURD (Synergistic-Unique-Redundant Decomposition of Causality, Martínez-Sánchez, et al., 2024), which provides a robust framework for decomposing causal relationships into three components: redundant, unique, and synergistic. This decomposition helps to reveal the complex interactions between variables in a system. Through fine-grained causal decomposition, SURD can distinguish synergistic and redundant causal effects that traditional methods may not capture. It enables the decomposition and quantitative analysis of causal relationships and that's crucial for understanding the interactions between variables in complex systems. For each grid, we tested lag periods ranging from 1 to 60 days and iteratively identified the lag period that maximized the combined causal contributions of the four environmental factors. When the contribution of a single environmental factor exceeded 1.5 times that of any other factor, it was classified as the dominant factor; otherwise, no dominant factor was assigned. We also use the feature importance provided by the LightGBM model with an interpretability tool SHAP (Broeck., 2022) as a supplementary analysis. We didn't use it as the main method because it failed to capture the potential lag effects of environmental variables. Finally, we calculated water use efficiency (WUE) based on the ratio of SIF to ET, which can be considered a proxy for GPP-derived WUE (Zhang Z. et al., 2023a). WUE was further utilized in long-term temporal analyses and attribution analyses based on SURD.

### 3 Results

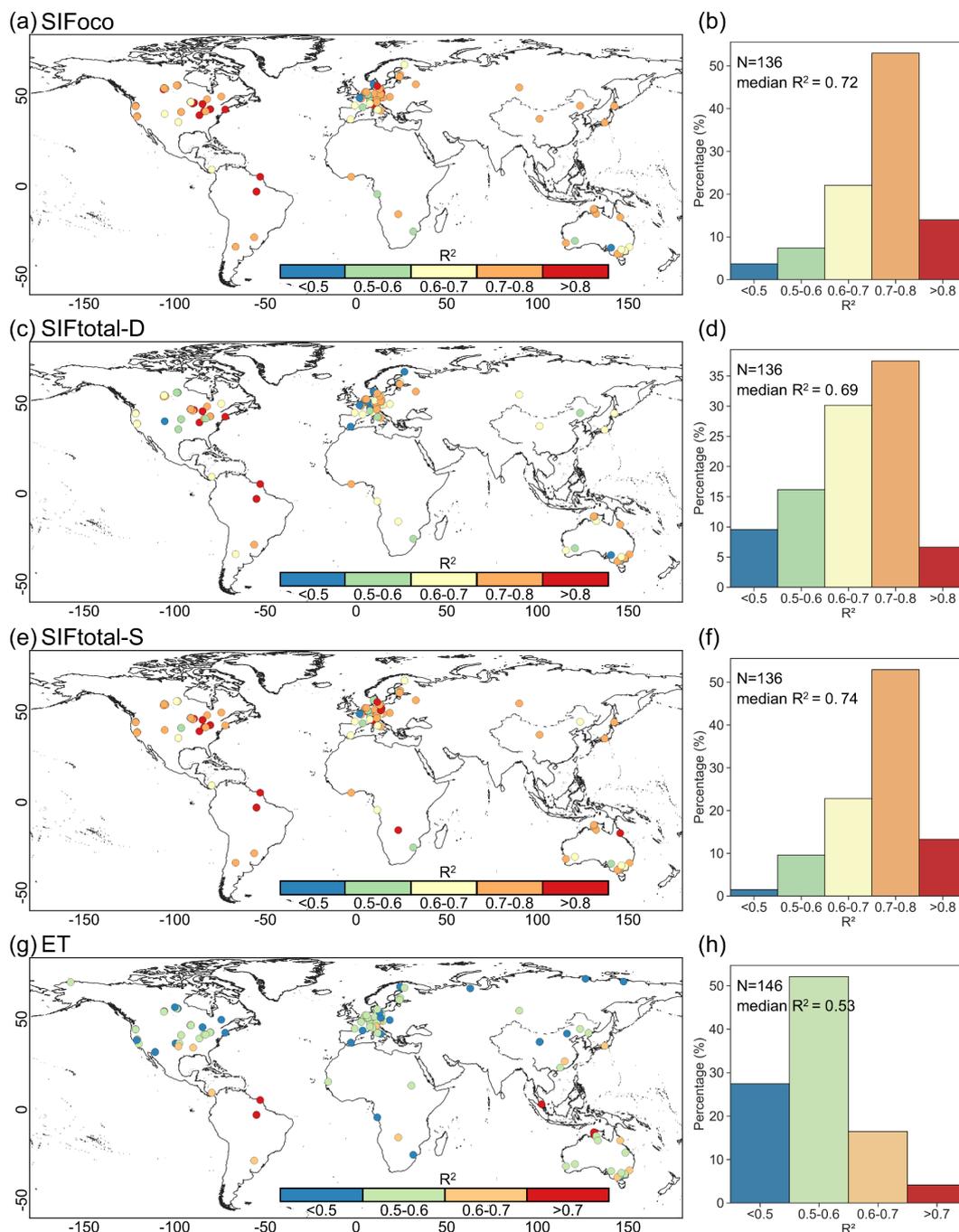
#### 3.1 Diurnal dynamic validation of SIF and ET



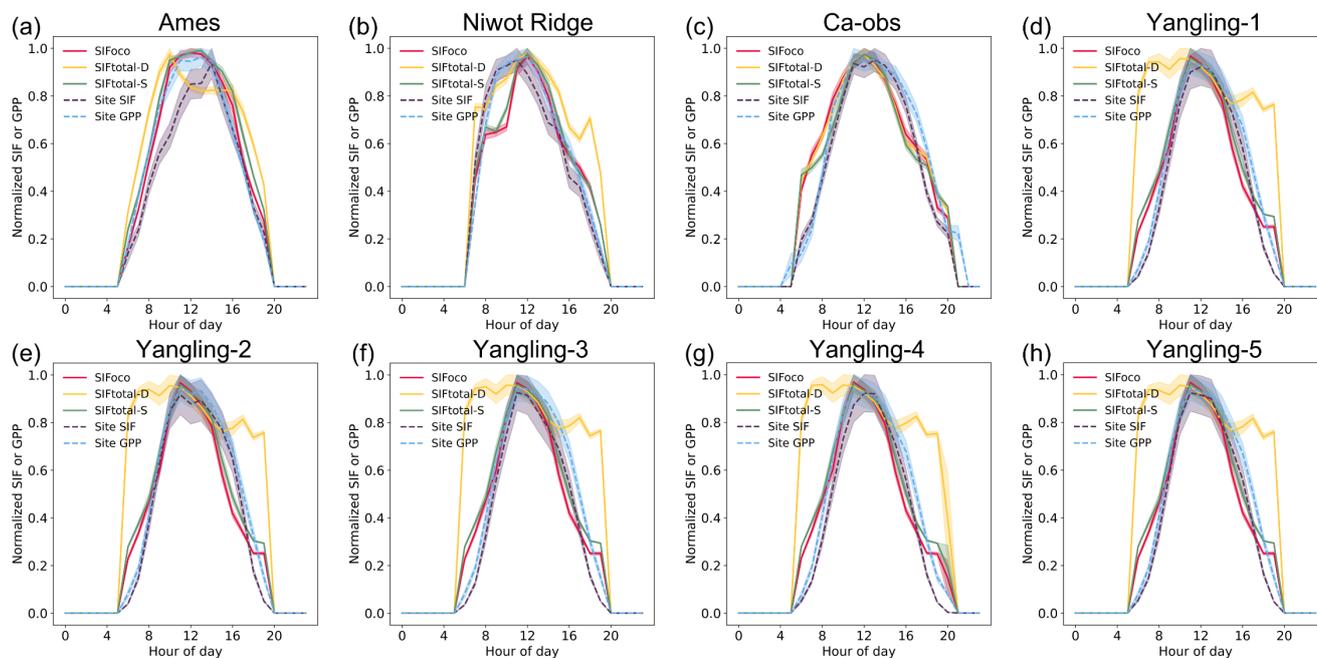
**Figure 4** The diurnal dynamic pattern of global SIF<sub>oco</sub>, SIF<sub>total-D</sub>, SIF<sub>total-S</sub>, and ET shown at 6-hour intervals. Each map is plotted using the corresponding 6-hour mean values of SIF or ET. The first row represents the results for SIF<sub>oco</sub> (a to d), the second row represents the results for SIF<sub>total-D</sub> (e to h), the third row represents the results for SIF<sub>total-S</sub> (i to l), and the fourth row represents the results for ET (m to p). All results are plotted based on the average values from 1982 to 2022. Pixels with a value of exactly 0 are not plotted.



325 The dataset we produced captures the diurnal dynamics of SIF and ET on the global scale (Fig. 4). There is a strong correlation between SIF  
and ET variations throughout the day. From UTC 0:00 to 23:00, high values gradually shift from the Eastern Hemisphere to the Western  
Hemisphere, which aligns with the variation in solar radiation at the Earth's surface. In addition, the value distribution for each time interval  
exhibits a pattern of diffusion from the centre outward. For instance, during the UTC 6-11 interval, high-value regions are located in central  
Africa, with a clear trend of decreasing SIF or ET values as the distance from the centre increases. This pattern aligns with the distribution  
characteristics of the superimposed diurnal variations. Since SIF is only generated under solar radiation, we retained the SIF values only in  
330 areas where the radiation is greater than 0 in ERA5-Land, setting other regions to 0. This can be easily observed in panels (a), (e), and (i) of  
Fig. 4. For consistency, and because most effective observations from ECOSTRESS are during the daytime rather than at night, we similarly  
retained ET values only in areas where solar radiation is greater than 0, setting ET in other pixels to 0 (Fig.4m).



335 **Figure 5** Validation of SIF and ET based on FLUXNET2015 sites. (a), (c), (e), and (g) show the global distribution of the coefficient of determination ( $R^2$ ) for the comparisons of SIF<sub>oco</sub>, SIF<sub>total-D</sub>, SIF<sub>total-S</sub>, and ET, respectively. Panels (b), (d), (f), and (h) present the statistical distribution of  $R^2$  values for SIF<sub>oco</sub>, SIF<sub>total-D</sub>, SIF<sub>total-S</sub>, and ET across these sites. All comparisons were conducted on the hourly timescale.



340 **Figure 6** Comparison of SIF with observed diurnal variations at eight SIF sites. (a) Ames site during July–August 2017 (Magney et al., 2019a); (b) Niwot Ridge site during July–August 2017 (Magney et al., 2019b); (c) Ca-obs site during July–August 2019 (Chen et al., 2024); (d)–(h) five sites in Yangling during April–May 2021 (Liu et al., 2022); The shaded areas represent one standard error of mean. To ensure consistency in units, normalized results were used for plotting.

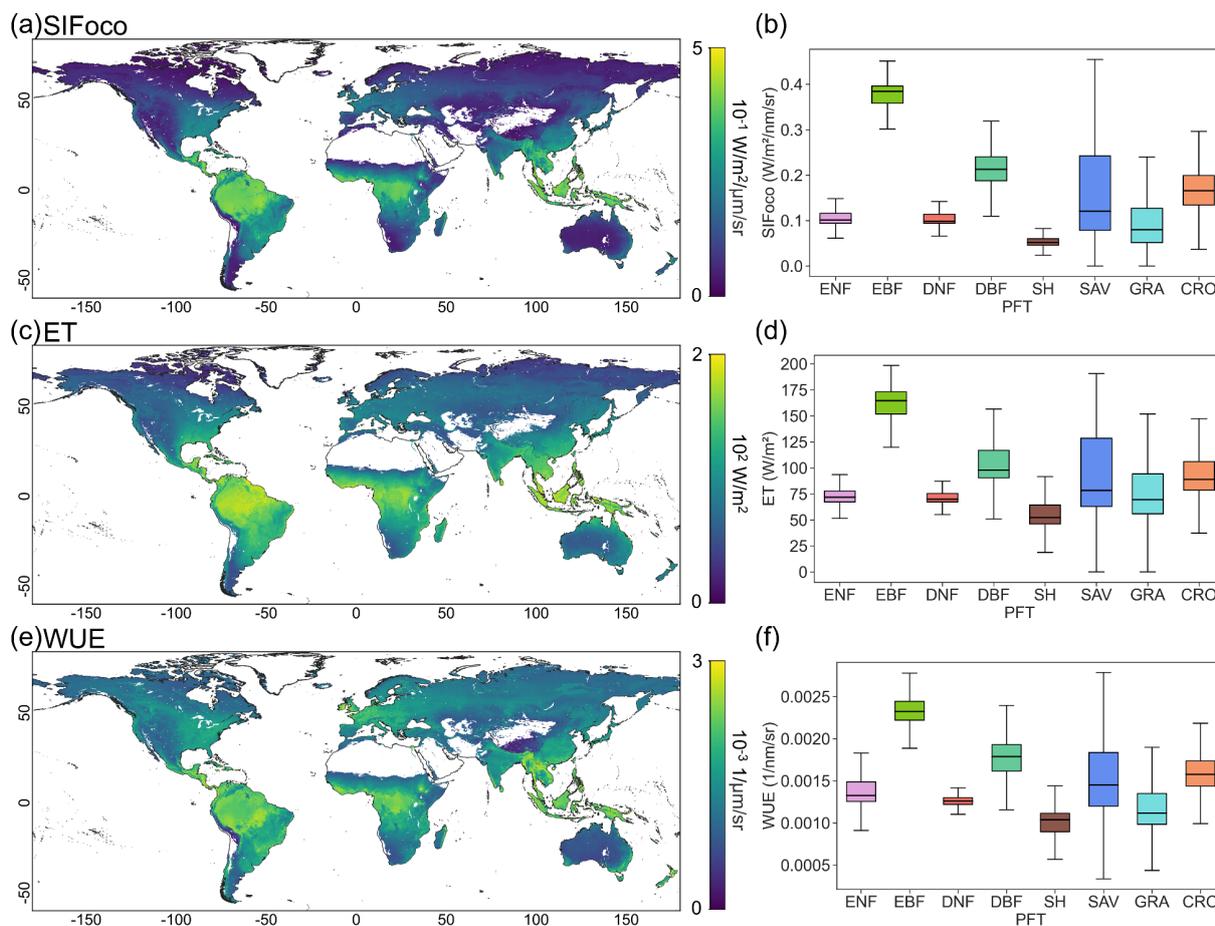
345 Our SIF and ET products demonstrated strong consistency with FLUXNET2015 sites on the global scale (Fig. 5). The median coefficient of determination ( $R^2$ ) was 0.72 for  $SIF_{OCO}$ , 0.69 for  $SIF_{total-D}$ , 0.74 for  $SIF_{total-S}$  (in 136 sites), and 0.53 for ET (in 146 sites) on a daily timescale. On the hourly timescale, the corresponding  $R^2$  values were 0.73 for  $SIF_{OCO}$ , 0.69 for  $SIF_{total-D}$ , 0.73 for  $SIF_{total-S}$ , and 0.63 for ET (Fig. S2). The consistency of the three SIF products remained relatively stable across hourly and daily scales, while ET showed a notable improvement ( $\Delta R^2=0.1$ ) on the daily scale. At these sites, the  $R^2$  values did not exhibit a clear spatial distribution pattern and appeared to be randomly distributed. It should be noted that SIF and GPP are not inherently linearly correlated, and these comparisons were conducted at different spatial scales ( $0.1^\circ$  and site-scale). Furthermore, we compared our products with observations from eight sites that feature concurrent SIF and GPP measurements (Fig. 6). The results showed that both  $SIF_{OCO}$  and  $SIF_{total-S}$  exhibited good agreement with observed SIF and GPP across all eight sites.  $SIF_{total-D}$  performed well at the Ca-obs site but showed significantly lower accuracy at the other seven sites compared to  $SIF_{OCO}$  and  $SIF_{total-S}$ , and this pattern is consistent with the validation results from the FLUXNET2015 sites.

350

355

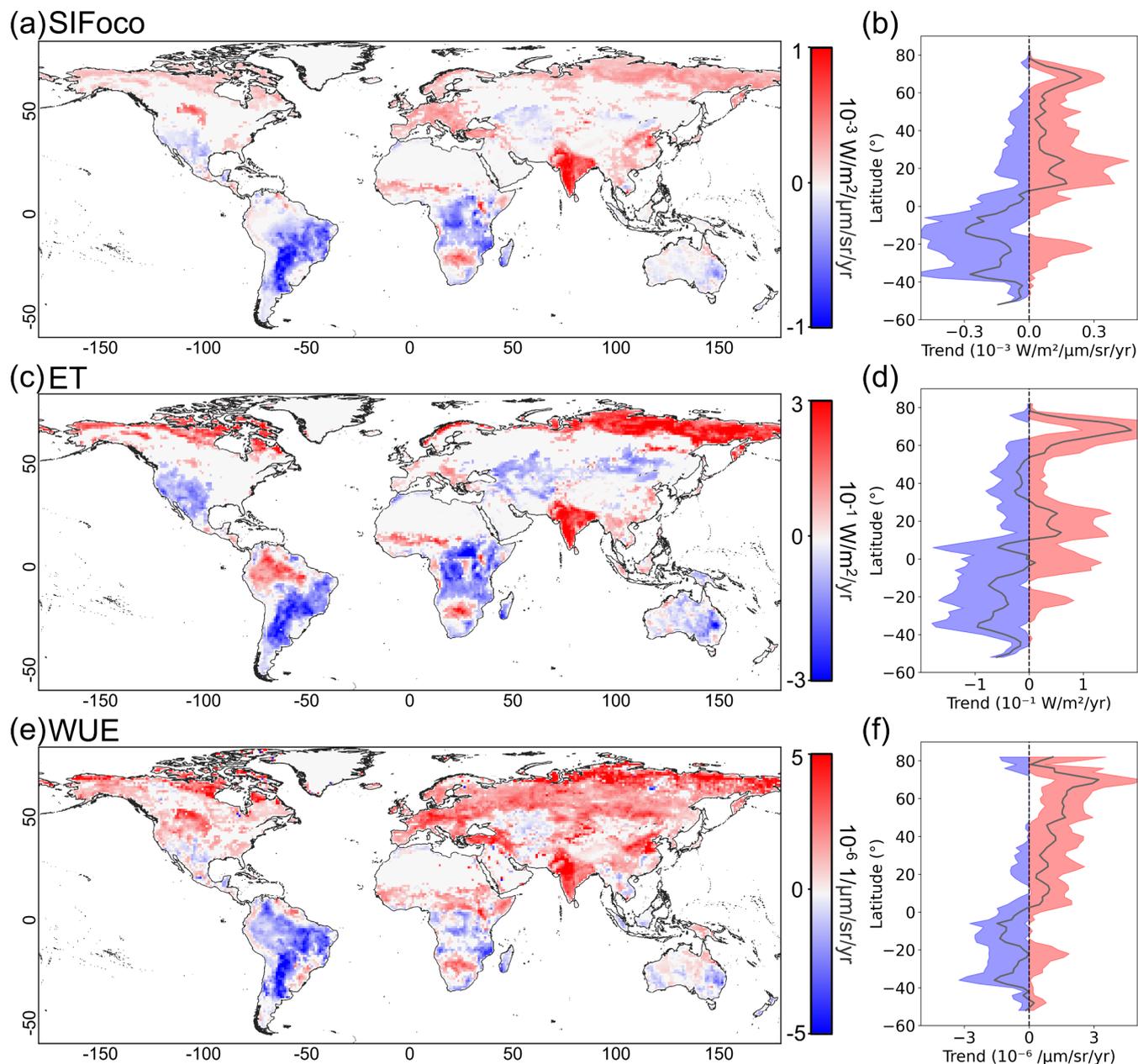


### 3.2 Spatial and temporal patterns of SIF, ET, and WUE



**Figure 7** The spatial patterns of the mean values for global SIF<sub>oco</sub> (a), ET (c), and WUE (e) from 1982 to 2022. The mean values of SIF<sub>oco</sub> (b), ET (d), and WUE (f) are presented by plant functional types (PFTs): ENF represents evergreen needleleaf forests, EBF represents evergreen broadleaf forests, DNF represents deciduous needleleaf forests, DBF represents deciduous broadleaf forests, SH represents shrublands, SAV represents savannas, GRA represents grasslands, and CRO represents croplands.

360



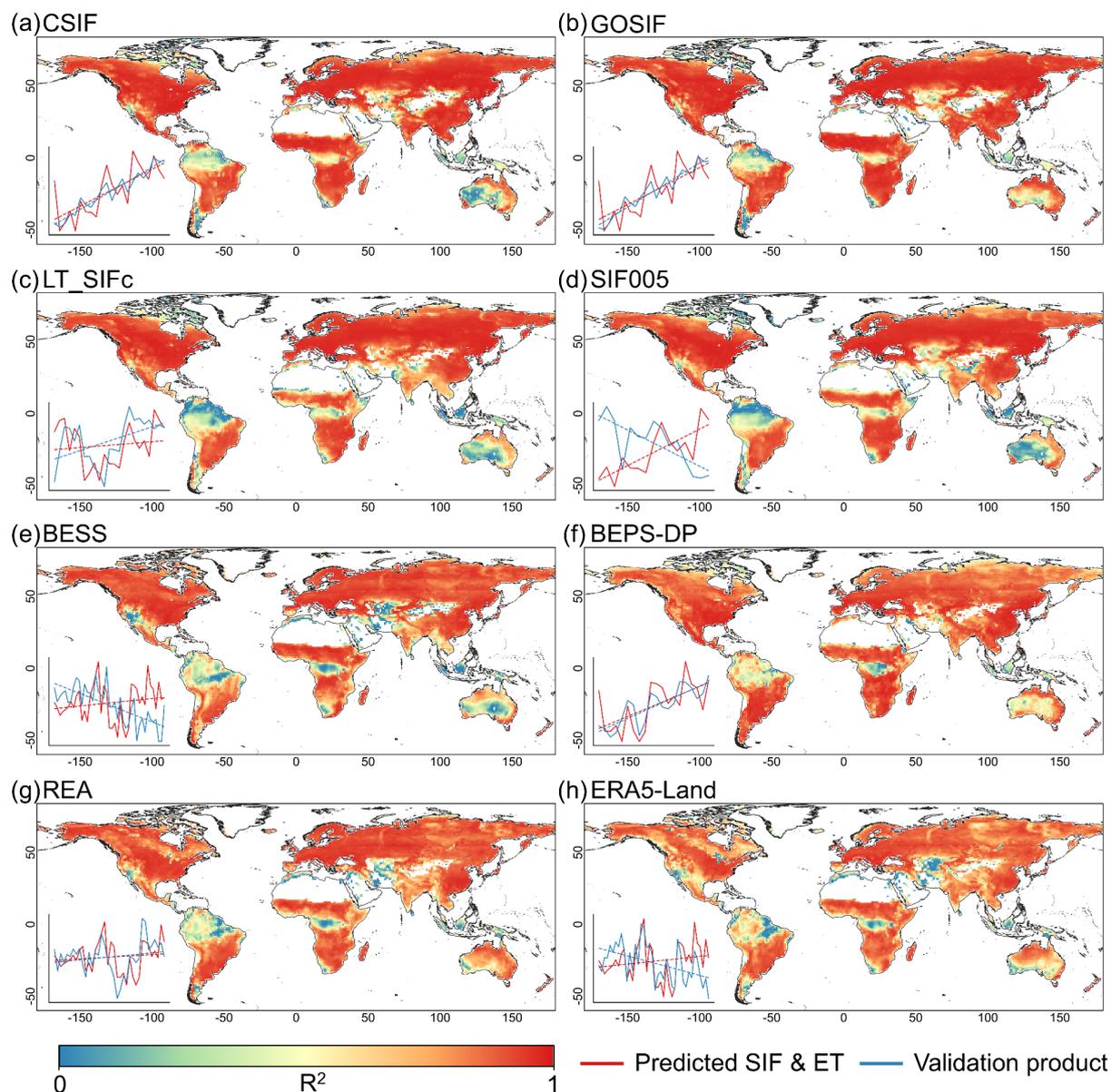
365 **Figure 8** Trends of global SIFoco (a), ET (c), and WUE (e) from 1982 to 2022. Trends were calculated on a per-pixel using linear regression. Pixels were retained for visualization only if they passed the 95% significance level of the Mann-Kendall trend test. The latitudinal profiles of the global average trends of SIFoco (b), ET (d), and WUE (f) from 1982 to 2022, with shaded areas representing  $\pm 1$  standard deviation.

The spatial distribution patterns of SIF, ET, and WUE are largely consistent with the global distribution of vegetation (Fig. 7, Fig. S3). High-value regions are primarily concentrated in the Amazon Rainforest, Central Africa, and the Malay Archipelago. In North America, the SIF, ET, and WUE values in the eastern region are significantly higher than those in the west. The high



values in the northern hemisphere's high-latitude regions are predominantly found in Europe. Notably, along the central mountain range of New Guinea, stretching from the northwest to the southeast at approximately 140°E, 10°S, both SIF, ET, and WUE exhibit markedly lower values in comparison to the northern and southern regions. This observation underscores the high spatial detail capture capability of our product at the global scale. From the perspective of different vegetation types, SIF, ET, and WUE values in broadleaf forests (EBF and DBF) are significantly higher than those in needleleaf forests (ENF and DNF), with SIF showing the most pronounced difference—ranging from 2 to 4 times higher. It is followed by ET (1.2 to 2.2 times) and, to a lesser extent, WUE (1.2 to 1.7 times). In addition, the disparity between the two broadleaf types is considerably greater than that between the two needleleaf types (Fig. 7b, d, f). Shrublands exhibit the lowest average SIF, ET, and WUE, while their heterogeneity in savannas is the highest.

The long-term trends from 1982 to 2022 reveal significant spatial heterogeneity in photosynthesis (Fig. 8). SIF generally shows an increasing trend in the Northern Hemisphere and a decreasing trend in the Southern Hemisphere. For example, regions such as northern Eurasia, northern America, the Indian subcontinent, and southern Africa exhibit a clear increasing trend in SIF (0 to 0.001 W/m<sup>2</sup>/μm/sr per year), whereas regions like western North America, southern South America, and central Africa display a declining trend (0 to -0.001 W/m<sup>2</sup>/μm/sr per year). Results from SIF at the photosystem level confirm this pattern as well (0 to 0.01 W/m<sup>2</sup>/μm/sr per year, Fig. S4a, b). Similar to SIF, ET also shows increasing trends in the aforementioned regions, with a notable exception being the Amazon rainforest, where ET also increases significantly (0.1 to 0.2 W/m<sup>2</sup> per year, Fig. 8c). For WUE calculated from SIF and ET, a distinct spatial heterogeneity and hemispheric symmetry emerge (Fig. 8f): WUE increases remarkably in the Northern Hemisphere while decreasing in the Southern Hemisphere, with most trends being statistically significant. This hemispheric asymmetry is also supported by the results from the two photosystem-level SIF calculations (Fig. S4c, d).



**Figure 9** Validation based on global gridded products. (a) to (d) represent the validation of SIF<sub>oco</sub> using CSIF, GOSIF, LT\_SIFc, and SIF005, respectively. (e) to (h) illustrate the validation of ET using BESS, BEPS-DP, LT\_SIFc, and SIF005, respectively. All comparisons were conducted based on the  $R^2$  of the monthly time series for each grid cell. The plot in the bottom-left corner of each map illustrates the long-term global grid-averaged normalized trends for predicted data (red) and the validation products (blue). The global mean was calculated using area-weighted averaging based on grid sizes. The comparison periods are as follows: (a) CSIF, 2001–2022; (b) GOSIF, 2001–2022; (c) LT\_SIFc, 1996–2018; (d) SIF005, 2003–2017; (e) BESS, 1982–2022; (f) BEPS-DP, 2001–2020; (g) REA, 1982–2017; and (h) ERA5-Land, 1982–2021.



In the validation of global gridded products on the monthly timescale, SIF<sub>OCO</sub> demonstrated strong consistency with most products across the Northern Hemisphere (Fig. 9). Specifically, the proportions of monthly pixels with  $R^2$  values exceeding 0.7 for SIF<sub>OCO</sub> and the four products are 71.2% (CSIF), 72.9% (GOSIF), 67.2% (LT\_SIFc), and 66.8% (SIF005), respectively.

405 However, significant uncertainties ( $R^2$  less than 0.7) were observed in regions such as the Amazon rainforest, Central Africa, the Indonesian archipelago, and the desert areas of Australia. These discrepancies are likely associated with differences in the methods used to calculate daily SIF: in our study, daily SIF was obtained by aggregating SIF over all 24 hours of the day, whereas the four validation SIF products derived daily SIF using diurnal correction factors applied to instantaneous SIF observations. This approach typically accounts only for variations in solar radiation and viewing angles. Our SIF product

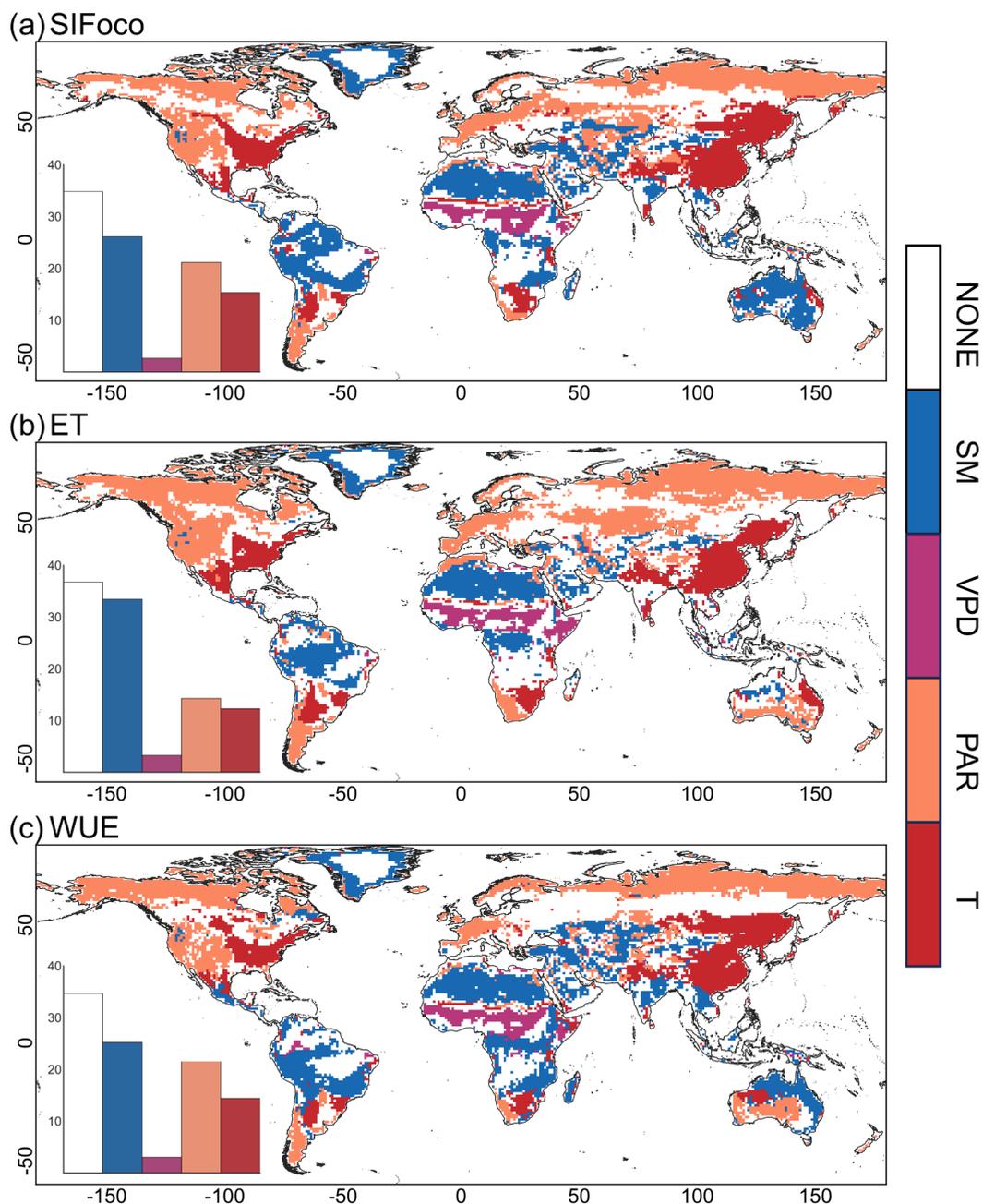
410 showed good long-term consistency with CSIF and GOSIF. However, it showed some discrepancies with LT\_SIFc and larger differences with SIF005. This disparity may be attributed to the fact that LT\_SIFc and SIF005 are both fused datasets integrating SIF observations from multiple satellites, and their fusion methods may not fully preserve the original long-term trends. As a result, SIF005 displayed a decreasing trend in SIF from 2003 to 2017, which deviates from the trends observed in the other SIF products. For ET, the proportions of monthly pixels with  $R^2$  values greater than 0.7 for the four products are

415 67.0% (BESS), 70.4% (BEPS-DP), 70.0% (REA), and 65.1% (ERA5-Land), respectively. Similar uncertainties were primarily found in the aforementioned regions, while good consistency was observed in other areas. This could also be related to the methods used for calculating daily ET. One supporting evidence is that our ET product exhibited the best spatial and long-term consistency with BEPS-DT, which, like our approach, predicts ET on the hourly scale and then aggregates it to obtain daily ET. Another factor contributing to the uncertainties is the observational characteristics of ECOSTRESS. Our ET product

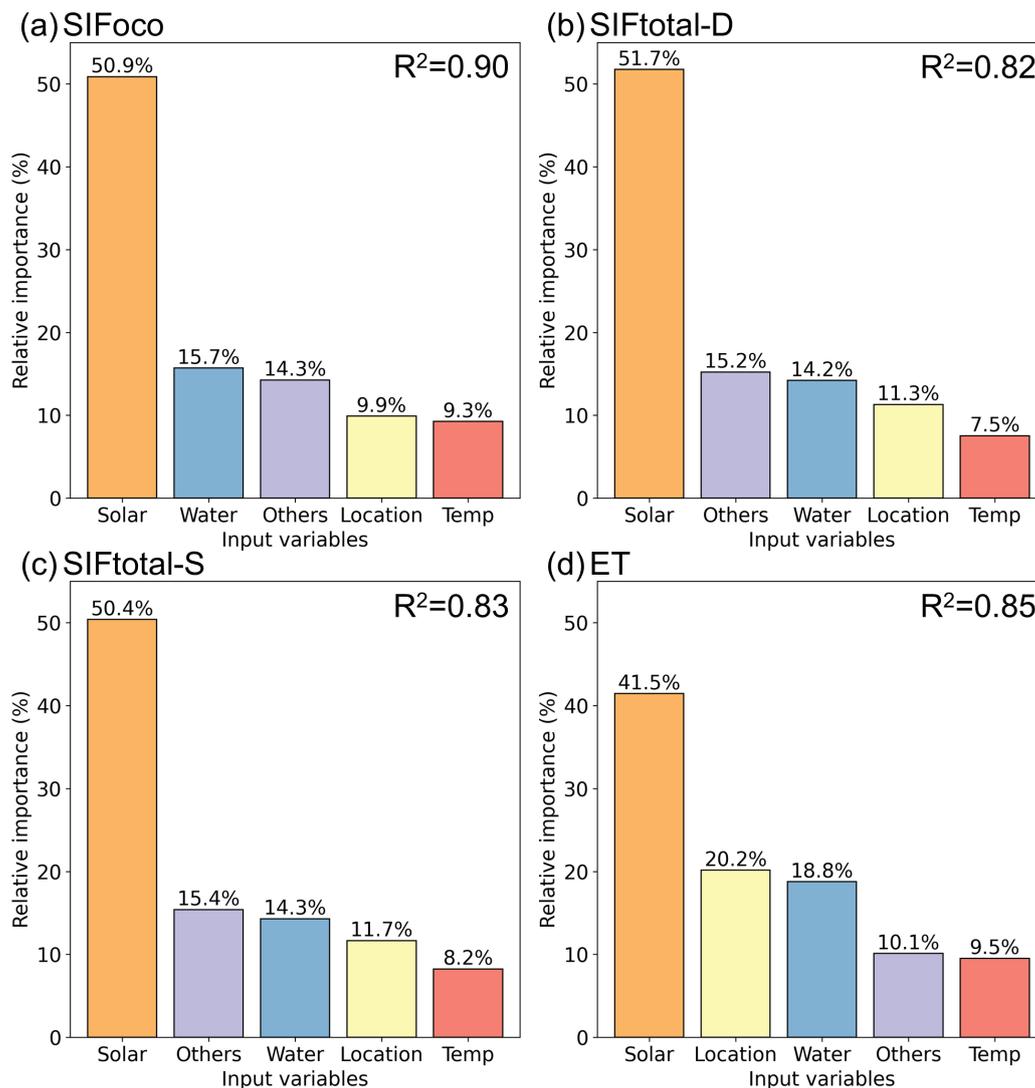
420 is limited to daytime values, whereas nighttime ET is not zero in reality. Overall, our SIF and ET product effectively capture their diurnal and seasonal variation characteristics.



### 3.3 Attribution of drivers for SIF, ET, and WUE



**Figure 10** Dominant factor analysis for SIFoco (a), ET (b), and WUE (c) based on the SURD algorithm. Blue pixels represent dominance by soil moisture (SM), purple pixels indicate dominance by VPD, orange pixels represent dominance by PAR, red pixels indicate dominance by air temperature (T), and white pixels indicate there is no dominant factor. Calculations were performed using daily-aggregated data from 1982 to 2022. The bar chart in each map represents the proportion of pixels dominated by each factor.



430 **Figure 11** SHAP-based attribution analysis for (a) SIFobs, (b) SIF<sub>total-D</sub>, (c) SIF<sub>total-S</sub>, and (d) ET. The input variables used for explanation are divided into five categories: Solar, Water, Temp, Location, and Others. Among them, Solar represents radiation factors (PAR and cos(SZA)), Water represents moisture factors (VPD and soil moisture), Temp represents air temperature, Location represents longitude and latitude, and Others include DOY, DEM, and land use types. For the models in (a), (b), and (c), the same model used for retrieval was directly applied. For the model in (d), retraining was conducted using the variables from (a), (b), and (c) to ensure consistency.

435

Through a driver analysis that accounts for lag effects, we created maps of the dominant factors and found that for SIF, ET, and WUE, over 60% of the global area exhibits distinct dominant factors, with a certain degree of spatial similarity (Fig. 10). For example, temperature plays a strongly dominant role in eastern Eurasia, and this pattern is similarly evident in eastern North America (Fig. 10a, c). In high-latitude regions, PAR emerges as the absolute dominant factor. VPD-dominated areas are



440 primarily concentrated in sub-Saharan Africa, while VPD does not show dominance in other regions. Similar conclusions were  
obtained for the two photosystem-level SIF datasets and the corresponding WUE (Fig. S5). Soil moisture has a strong dominant  
influence on SIF, ET, and WUE in South America. To supplement this analysis, we conducted a SHAP-based attribution  
analysis (Fig. 11). The result reveals that radiation factors (Solar) are the absolute dominant factors for both SIF and ET,  
contributing over 50% of the total influence on SIF and 41.5% on ET. The main inconsistency lies in the fact that SHAP-based  
445 attribution analysis struggles to account for the lag effects of environmental factors, instead primarily reflecting the  
instantaneous effects of environmental drivers. Nevertheless, water and temperature factors also show significant importance,  
contributing between 20% and 30% (Fig. 11).

## 4 Discussion

### 4.1 Advantages of our dataset and rationality of production

450 Hourly-scale SIF and ET data have been scarce in previous studies. Popular SIF datasets such as CSIF, GOSIF, and SIF005  
(Li and Xiao, 2019; Wen et al., 2020; Zhang Y. et al., 2018) are only available on a daily scale due to fixed overpass times.  
This limitation restricts the ability to conduct diurnal research on carbon and water cycles. Recent missions like OCO-3 and  
ECOSTRESS are able to monitor SIF and ET data at varying times of day (Xiao et al., 2021). This feature provides a  
fundamental feasibility for constructing the hourly SIF and ET datasets. Based on the hourly SIF and ET datasets, we can  
455 capture diurnal vegetation physiological activities on the global scale that were previously difficult to explore in the past. For  
instance, we calculated pixel-level changes in the peak timing during the day based on predicted SIF (Fig. S6) and found a  
clear delaying trend (0 to 0.5 minutes per year) is evident in high-latitude regions of both hemispheres, while clustered  
advancing trends are observed in regions such as the Amazon rainforest, central Africa, eastern Asia, and northern North  
America. Research on related phenomena has recently been proven to encapsulate some profound vegetation physiological  
460 implications (Li X. et al., 2023; Liu et al., 2024; Zhang Z. et al., 2023a). Moreover, by providing both hourly SIF and ET data,  
our dataset makes it possible to analyze global WUE on a diurnal rather than a daily scale (Li F. et al., 2023).

While recent studies have produced hourly SIF products leveraging these advancements (Deng et al., 2025a; Jeong et al., 2024;  
Zhang Z. et al., 2023b), limitations persist. For instance, Jeong et al.'s study utilized OCO-3 SIF data combined with  
465 geostationary satellite observations to reconstruct the spatial distribution of SIF from August 2019 to July 2021, covering  
Eastern Asia and Oceania (Jeong et al., 2024). However, this dataset's restricted spatial coverage and reliance on specific input  
data limit its utility for global and long-term applications. In fact, long-term temporal analysis has become a key focus in  
current research, such as the weakening of the CO<sub>2</sub> fertilization effect (Wang et al., 2020) and drought stress (Deng et al.,  
2025a). Additionally, the SIF dataset produced by Zhang Z. et al. is at a coarser spatial resolution (0.5°) and does not offer  
470 hourly datasets for each day, which may limit its use in accurately estimating GPP or analysing sudden disaster events. In  
contrast, our study produced a long-term, hourly SIF, and ET dataset covering the entire globe, offering new insights into



475 large-scale vegetation photosynthesis in the context of climate change (Li X. et al., 2023; Zhang Z. et al., 2023a). A potential limitation of the above studies is computational power, as simulating hourly data products requires much more computational resources compared to commonly used 4-day or 8-day data products. In response, we developed an efficient method to enable rapid sampling of input raster data (see code in Code Availability for details). Additionally, the efficient machine learning model LightGBM we used improved accuracy while maintaining high performance (Cheng et al., 2025; Zhao et al., 2024).

480 Previous studies on scale-up predominantly relied on reflectance data from optical satellites such as MODIS and Landsat, or vegetation indices derived from them (Li and Xiao, 2019; Tao et al., 2024; Zhang Y. et al., 2018). However, these satellite observations frequently suffer from severe data gaps under strict quality control conditions. While these gaps may be manageable during the model training phase—where only high-quality data are utilized—they can lead to numerous low-quality predictions during the production phase. Although various data imputation methods (e.g., Zhang Y. et al., 2017) are commonly employed to address such gaps, this study sought to minimize reliance on optical remote sensing satellite data wherever possible. Instead, we relied on complete and temporally consistent input datasets, namely ERA5-Land and GIMMS  
485 FPAR4g. This approach ensures temporal consistency in the dataset, covering the long-term period from 1982 to 2022 (Zhao et al., 2024). We used these two data as the main input to drive our LightGBM model, ensuring that the produced dataset maintained the highest level of consistency over the 41 years. This consistency supports long-term analyses of photosynthesis and transpiration variations. In addition, the OCO-2 SIF data product after calibration also demonstrates excellent consistency (Doughty et al., 2021; Rosenberg et al., 2020), which justifies the fusion of OCO-2 and OCO-3 SIF data and the expansion of  
490 the training samples.

#### 4.2 Comparison of the two methods to estimate $SIF_{total}$

To this day, almost all SIF products directly obtained from satellite spectral inversions are at the canopy level. This limitation arises because atmospheric correction cannot account for transmission losses within the canopy and from the photosystem to the leaf surface. In other words, it is currently not possible to directly estimate the escape ratio of SIF or the SIF emitted at the  
495 photosystem level from satellite observations (Regaieg et al., 2025). However, the photosystem-level SIF is likely more directly related to actual photosynthesis (Liu et al., 2021; Guo et al., 2024). Therefore, exploring methods to obtain photosystem-level SIF is both beneficial and necessary. We used two methods to estimate the canopy escape ratio: one directly based on a radiative transfer model and the other based on radiative transfer probabilities from the SCOPE simulation. Both methods share the common feature of considering the vegetation reflectance in the near-infrared band as a crucial parameter  
500 (Eq. 2). The method that directly calculates the escape ratio involves the use of the Canopy Index (CI), but relevant products on the global scale over long periods are still lacking. From the retrieval of  $SIF_{total}$  estimated by both methods using the LightGBM model, the accuracy difference between them appears to be minimal but their absolute magnitudes show notable differences (Fig. 3). Our results show that  $SIF_{total}$  simulated based on the SCOPE model ( $SIF_{total-S}$ ) appears to have a higher



505 correlation with site-level GPP (Figs. 3 and 4) and gridded GPP products (Fig. S7), but it remains uncertain which product better reflects the true SIF<sub>total</sub>. Therefore, we provide both SIF<sub>total</sub> products for users to choose from.

In our study, we found that the estimation accuracy of both SIF<sub>total</sub> for GPP does not appear to be significantly higher than that of SIF<sub>OCO</sub> (Fig. 5, 6, and S2). This may be because the canopy information carried by SIF<sub>OCO</sub> can also, to some extent, explain the variation in GPP, particularly in terms of seasonal changes (Liu et al., 2019). Moreover, the input data used to calculate the two escape probabilities inherently contain certain uncertainties, such as the MODIS LAI and BRDF parameter products. Although we have used the highest-quality data as input, uncertainties remain unavoidable (Zhang Z. et al., 2021). For SIF<sub>total</sub> directly calculated based on the radiative transfer model (SIF<sub>total</sub>-D), the mixing of vegetation types within the 0.1° grid inevitably affects parameters set according to vegetation types, such as leaf angle distribution. For the canopy escape probability simulated using the SCOPE model, previous studies have shown that it has a higher correlation with GPP in sparsely vegetated areas compared to results directly calculated from the radiative transfer model (Li H. et al., 2024). However, the SCOPE model struggles to fully represent the vertical heterogeneity of vegetation (Yang et al., 2017), and systematic differences may exist between the reflectance in the SCOPE model and the satellite-scale input reflectance. Recent studies have simulated the transmission process of SIF using three-dimensional models (e.g., DART) (Liu et al., 2019; Regaieg et al., 2023, 2025), but there is still a lack of validation regarding the feasibility of these methods at the satellite scale.

#### 520 4.3 Uncertainties and limitations

In addition to the uncertainties associated with the two photosystem-level SIF estimation methods mentioned in Sect. 4.2, our study also faces other uncertainties and limitations. First, although we incorporated OCO-2 SIF data as a supplement to the SIF training samples, the training samples only cover the period from 2014 to 2022. This means that we lack stable reference samples for earlier SIF data, which has rarely been considered in current studies. Second, since the primary input data come from the ERA5-Land dataset with a spatial resolution of 0.1°, it is challenging to achieve higher spatial resolution in our dataset. A recent study focused on the spatial resolution (500m) of the product rather than the temporal resolution (8 days) (Tao et al., 2024), primarily due to the use of MODIS products (500m, 8 days) as input data for retrieval. This also indicates the limitation of input data, rather than computational power, is the main reason why it is difficult to produce products that simultaneously feature high spatial and temporal resolutions.

530 Regarding the dataset validation, the lack of publicly available SIF site datasets similar to FLUXNET2015 makes it difficult to conduct comparable validation. In fact, as shown in the validation results from the eight SIF sites used in this study, the predicted SIF appears to have a stronger correlation with site-level GPP than with site-level SIF. This is mainly because SIF sites typically only capture SIF emissions over relatively small areas (Hao et al., 2022), whereas GPP is often estimated based on CO<sub>2</sub> concentrations measured over several kilometres (Chen et al., 2024), making it better aligned with the 0.1° grid resolution. Moreover, differences in the spectral bands used for observed SIF measurements may lead to significant differences



in magnitude, further hindering the direct comparison between produced gridded SIF and site-level observed SIF. Future work should focus on finding a more compatible validation approach that could enhance the credibility of regional-scale hourly dataset validation and promote its development. The retrieval of ET exhibits greater uncertainties compared to SIF, especially in regions with high ET values (Fig. 3). This is primarily due to the more complex origins of ET, which include not only vegetation transpiration but also soil water evaporation. In future work, more accurate ET retrieval may require consideration of additional influencing factors, such as nighttime processes and canopy conductance (Zheng et al., 2025).

## 5 Data availability

The global hourly SIF and ET dataset (1982–2022) at a  $0.1^\circ$  spatial resolution produced in this study is available at <https://doi.org/10.57760/sciencedb.ecodb.00177> (Deng et al., 2025b). All products are provided in NetCDF4 format. Our dataset contains four main fields: SIF<sub>OCO</sub>, which represents continuous SIF data obtained by extending canopy-level SIF from OCO-2 and OCO-3 observations; SIF<sub>total-D</sub>, which represents photosystem-level SIF directly calculated based on the radiative transfer method; SIF<sub>total-S</sub>, which represents photosystem-level SIF obtained from the simulation of SCOPE model; and ET<sub>ECO</sub>, which represents continuous ET data obtained by scaling ECOSTRESS observations. Detailed descriptions of these four main fields, including units, coverage, and other relevant information, are included in the NetCDF4 files.

## 6 Code availability

The code for training the models and producing the dataset can be directly downloaded from <https://doi.org/10.57760/sciencedb.ecodb.00177> (Deng et al., 2025b).

## 7 Conclusion

In this study, we employed machine learning-based data fusion and scale-up methods to propose, for the first time, a long-term (1982–2022) and high-temporal resolution (1-hour) dataset of SIF and ET with a spatial resolution of  $0.1^\circ$  (HOUR\_SIF<sub>OCO</sub> and HOUR\_ET<sub>ECO</sub>). At the same temporal span and resolution, we also provided two comparable photosystem-level SIF datasets using two different methods. We thoroughly evaluated the produced products using both site-level and regional-scale data and demonstrated the high accuracy of these datasets. The long-term analysis highlighted their unique advantages in capturing the diurnal dynamics of carbon and water cycles. We also applied an advanced causal analysis method (SURD) to investigate the regulatory effects of four environmental factors (PAR, VPD, soil moisture, and air temperature) on SIF, ET, and water use efficiency (WUE). The results revealed strong spatial clustering and variability on the global scale. In conclusion, our dataset offers great potential for advancing our understanding of terrestrial ecosystem responses to climate change and improving the monitoring of diurnal carbon and water cycles on the global scale.



## 565 **Author contribution**

ZD, TL, JC, and SW devised the conceptual ideas. ZD and TL processed the data. ZD and TL conducted the investigation and performed formal analysis. JC and SW acquired the funding., JC, SW, and KH verified the results and supervised the findings of this work. ZD and TL drafted the original manuscript. JC, SW, KH, PG, HP, and ZC authors reviewed and commented on the manuscript.

## 570 **Acknowledgements**

This work was financially supported by the National Natural Science Foundation of China [grant number 42250205]. We thank Weiqing Zhao for providing the GIMMS FPAR4g dataset; Troy S. Magney, Ruonan Chen, and Xiaoliang Lu for providing the site-level SIF data; Yao Zhang, Xing Li, Songhan Wang, and Jiaming Wen for providing the gridded SIF data; Bolun Li, Jiye Leng, Jiao Lu, and Joaquín Muñoz-Sabater for providing the gridded ET data; the European Centre for Medium-  
575 Range Weather Forecasts for providing the ERA5-Land data; the National Aeronautics and Space Administration for providing the OCO-2, OCO-3, and ECOSTRESS data; and the FLUXNET community for providing the site-level GPP and ET data.

## **Competing interests**

The authors declare no competing interests.

## **References**

- 580 Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A Next-generation Hyperparameter Optimization Framework, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, <https://doi.org/10.1145/3292500.3330701>, 2019.
- Allen, R. G., Pruitt, W. O., Wright, J. L., Howell, T. A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J., Yrisarry, J. B., Smith, M., Pereira, L. S., Raes, D., Perrier, A., Alves, I., Walter, I., and Elliott, R.: A recommendation on  
585 standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method, *Agricultural Water Management*, 81, 1-22, <https://doi.org/10.1016/j.agwat.2005.03.007>, 2006.
- Anderson, M. C., Yang, Y., Xue, J., Knipper, K. R., Yang, Y., Gao, F., Hain, C. R., Kustas, W. P., Cawse-Nicholson, K., Hulley, G., Fisher, J. B., Alfieri, J. G., Meyers, T. P., Prueger, J., Baldocchi, D. D., and Rey-Sanchez, C.: Interoperability of ECOSTRESS and Landsat for mapping evapotranspiration time series at sub-field scales, *Remote Sensing of  
590 Environment*, 252, 112189, <https://doi.org/10.1016/j.rse.2020.112189>, 2021.



- August, T., Klaes, D., Schlüssel, P., Hultberg, T., Crapeau, M., Arriaga, A., O'Carroll, A., Coppens, D., Munro, R., and Calbet, X.: IASI on Metop-A: Operational Level 2 retrievals after five years in orbit, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 113, 1340-1371, <https://doi.org/10.1016/j.jqsrt.2012.02.028>, 2012.
- Broeck, G. V. d., Lykov, A., Schleich, M., and Suci, D.: On the Tractability of SHAP Explanations, *J. Artif. Int. Res.*, 74, 36, <https://doi.org/10.1613/jair.1.13283>, 2022.
- 595 Brunke, M. A., Broxton, P., Pelletier, J., Gochis, D., Hazenberg, P., Lawrence, D. M., Leung, L. R., Niu, G.-Y., Troch, P. A., and Zeng, X.: Implementing and Evaluating Variable Soil Thickness in the Community Land Model, Version 4.5 (CLM4.5), *Journal of Climate*, 29, 3441-3461, <https://doi.org/10.1175/JCLI-D-15-0307.1>, 2016.
- Chen, B., Wang, P., Wang, S., Liu, Z., and Croft, H.: Evaluation of Leaf-To-Canopy Upscaling Approaches for Simulating  
600 Canopy Carbonyl Sulfide Uptake and Gross Primary Productivity, *Journal of Geophysical Research: Biogeosciences*, 129, e2023JG007521, <https://doi.org/10.1029/2023JG007521>, 2024.
- Chen, R., Liu, L., Liu, X., Liu, Z., Gu, L., and Rascher, U.: Improving estimates of sub-daily gross primary production from solar-induced chlorophyll fluorescence by accounting for light distribution within canopy, *Remote Sensing of Environment*, 300, 113919, <https://doi.org/10.1016/j.rse.2023.113919>, 2024.
- 605 Cheng, F., Li, Z., Yang, Z., Li, R., Wang, D., Jia, A., Li, K., Zhao, B., Wang, S., Yin, D., Li, S., Xue, W., Cribb, M., and Wei, J.: First retrieval of 24-hourly 1-km-resolution gapless surface ozone (O<sub>3</sub>) from space in China using artificial intelligence: Diurnal variations and implications for air quality and phytotoxicity, *Remote Sensing of Environment*, 316, 114482, <https://doi.org/10.1016/j.rse.2024.114482>, 2025.
- Deng, Z., Chen, J., Wang, S., Li, T., Huang, K., Gu, P., Peng, H., and Chen, Z.: Response of Vegetation Photosynthesis to the  
610 2022 Drought in Yangtze River Basin by Diurnal OCO-2/3 Satellite Observations, *Journal of Remote Sensing*, <https://doi.org/10.34133/remotesensing.0445>, 2025.
- Deng, Z., Li, T., Chen, J., Wang, S., Huang, K., Gu, P., Peng, H., and Chen, Z.: A Global 24-hourly Retrieval of Solar-Induced Chlorophyll Fluorescence and Evapotranspiration from OCO-2, OCO-3 and ECOSTRESS over 1982–2022 [DS/OL], V1, *Science Data Bank*, <https://doi.org/10.57760/sciencedb.ecodb.00177>, 2025.
- 615 Doughty, R., Kurosu, T. P., Parazoo, N., Köhler, P., Wang, Y., Sun, Y., and Frankenberg, C.: Global GOSAT, OCO-2, and OCO-3 solar-induced chlorophyll fluorescence datasets, *Earth Syst. Sci. Data*, 14, 1513-1529, <https://doi.org/10.5194/essd-14-1513-2022>, 2022.
- Elnashar, A., Wang, L., Wu, B., Zhu, W., and Zeng, H.: Synthesis of global actual evapotranspiration from 1982 to 2019, *Earth Syst. Sci. Data*, 13, 447-480, <https://doi.org/10.5194/essd-13-447-2021>, 2021.
- 620 Fang, H., Li, S., Zhang, Y., Wei, S., and Wang, Y.: New insights of global vegetation structural properties through an analysis of canopy clumping index, fractional vegetation cover, and leaf area index, *Science of Remote Sensing*, 4, 100027, <https://doi.org/10.1016/j.srs.2021.100027>, 2021.
- Fisher, J. B., Lee, B., Purdy, A. J., Halverson, G. H., Dohlen, M. B., Cawse-Nicholson, K., Wang, A., Anderson, R. G., Aragon, B., Arain, M. A., Baldocchi, D. D., Baker, J. M., Barral, H., Bernacchi, C. J., Bernhofer, C., Biraud, S. C., Bohrer, G.,



- 625 Brunsell, N., Cappelaere, B., Castro-Contreras, S., Chun, J., Conrad, B. J., Cremonese, E., Demarty, J., Desai, A. R., De  
Ligne, A., Foltýnová, L., Goulden, M. L., Griffis, T. J., Grünwald, T., Johnson, M. S., Kang, M., Kelbe, D., Kowalska,  
N., Lim, J.-H., Mainassara, I., McCabe, M. F., Missik, J. E. C., Mohanty, B. P., Moore, C. E., Morillas, L., Morrison, R.,  
Munger, J. W., Posse, G., Richardson, A. D., Russell, E. S., Ryu, Y., Sanchez-Azofeifa, A., Schmidt, M., Schwartz, E.,  
Sharp, I., Šigut, L., Tang, Y., Hulley, G., Anderson, M., Hain, C., French, A., Wood, E., and Hook, S.: ECOSTRESS:  
630 NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station, *Water Resources  
Research*, 56, e2019WR026058, <https://doi.org/10.1029/2019WR026058>, 2020.
- Fisher, J. B., Tu, K. P., and Baldocchi, D. D.: Global estimates of the land–atmosphere water flux based on monthly AVHRR  
and ISLSCP-II data, validated at 16 FLUXNET sites, *Remote Sensing of Environment*, 112, 901-919,  
<https://doi.org/10.1016/j.rse.2007.06.025>, 2008.
- 635 Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Köhler, P., Pollock, R., and Taylor, T. E.: Prospects for  
chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2, *Remote Sensing of Environment*, 147,  
1-12, <https://doi.org/10.1016/j.rse.2014.02.007>, 2014.
- Frankenberg, C., O'Dell, C., Guanter, L., and McDuffie, J.: Remote sensing of near-infrared chlorophyll fluorescence from  
space in scattering atmospheres: implications for its retrieval and interferences with atmospheric CO<sub>2</sub>  
640 retrievals, *Atmos. Meas. Tech.*, 5, 2081-2094, <https://doi.org/10.5194/amt-5-2081-2012>, 2012.
- Fuentes, I., Vervoort, R. W., and McPhee, J.: Global evapotranspiration models and their performance at different spatial  
scales: Contrasting a latitudinal gradient against global catchments, *Journal of Hydrology*, 628, 130477,  
<https://doi.org/10.1016/j.jhydrol.2023.130477>, 2024.
- Guo, C., Liu, Z., Jin, X., and Lu, X.: Improved estimation of gross primary productivity (GPP) using solar-induced chlorophyll  
645 fluorescence (SIF) from photosystem II, *Agricultural and Forest Meteorology*, 354, 110090,  
<https://doi.org/10.1016/j.agrformet.2024.110090>, 2024.
- Guo, X., Gui, X., Xiong, H., Hu, X., Li, Y., Cui, H., Qiu, Y., and Ma, C.: Critical role of climate factors for groundwater  
potential mapping in arid regions: Insights from random forest, XGBoost, and LightGBM algorithms, *Journal of  
Hydrology*, 621, 129599, <https://doi.org/10.1016/j.jhydrol.2023.129599>, 2023.
- 650 Hao, D., Zeng, Y., Zhang, Z., Zhang, Y., Qiu, H., Biriukova, K., Celesti, M., Rossini, M., Zhu, P., Asrar, G. R., and Chen, M.:  
Adjusting solar-induced fluorescence to nadir-viewing provides a better proxy for GPP, *ISPRS Journal of  
Photogrammetry and Remote Sensing*, 186, 157-169, <https://doi.org/10.1016/j.isprsjprs.2022.01.016>, 2022.
- Huang, K., Xia, J., Wang, Y., Ahlström, A., Chen, J., Cook, R. B., Cui, E., Fang, Y., Fisher, J. B., Huntzinger, D. N., Li, Z.,  
Michalak, A. M., Qiao, Y., Schaefer, K., Schwalm, C., Wang, J., Wei, Y., Xu, X., Yan, L., Bian, C., and Luo, Y.:  
655 Enhanced peak growth of global vegetation and its key mechanisms, *Nature Ecology & Evolution*, 2, 1897-1905,  
<https://doi.org/10.1038/s41559-018-0714-0>, 2018.
- Huang, Z., Zhou, L., and Chi, Y.: Spring phenology rather than climate dominates the trends in peak of growing season in the  
Northern Hemisphere, *Global Change Biology*, 29, 4543-4555, <https://doi.org/10.1111/gcb.16758>, 2023.



- 660 Hoek van Dijke, A. J., Mallick, K., Schlerf, M., Machwitz, M., Herold, M., and Teuling, A. J.: Examining the link between  
vegetation leaf area and land–atmosphere exchange of water, energy, and carbon fluxes using FLUXNET data,  
Biogeosciences, 17, 4443–4457, <https://doi.org/10.5194/bg-17-4443-2020>, 2020.
- Jiao, W., Chang, Q., and Wang, L.: The Sensitivity of Satellite Solar-Induced Chlorophyll Fluorescence to Meteorological  
Drought, *Earth's Future*, 7, 558–573, <https://doi.org/10.1029/2018EF001087>, 2019.
- 665 Jeong, S., Ryu, Y., Li, X., Dechant, B., Liu, J., Kong, J., Choi, W., Fang, J., Lian, X., and Gentine, P.: GEOSIF: A continental-  
scale sub-daily reconstructed solar-induced fluorescence derived from OCO-3 and GK-2A over Eastern Asia and Oceania,  
*Remote Sensing of Environment*, 311, 114284, <https://doi.org/10.1016/j.rse.2024.114284>, 2024.
- Lai, J., Kooijmans, L. M. J., Sun, W., Lombardozzi, D., Campbell, J. E., Gu, L., Luo, Y., Kuai, L., and Sun, Y.: Terrestrial  
photosynthesis inferred from plant carbonyl sulfide uptake, *Nature*, 634, 855–861, [10.1038/s41586-024-08050-3](https://doi.org/10.1038/s41586-024-08050-3), 2024.
- 670 Leng, J., Chen, J. M., Li, W., Luo, X., Xu, M., Liu, J., Wang, R., Rogers, C., Li, B., and Yan, Y.: Global datasets of hourly  
carbon and water fluxes simulated using a satellite-based process model with dynamic parameterizations, *Earth Syst. Sci.*  
*Data*, 16, 1283–1300, <https://doi.org/10.5194/essd-16-1283-2024>, 2024.
- Li, B., Liu, K., Wang, M., Wang, Y., He, Q., Zhuang, L., and Zhu, W.: High-spatiotemporal-resolution dynamic water  
monitoring using LightGBM model and Sentinel-2 MSI data, *International Journal of Applied Earth Observation and*  
*Geoinformation*, 118, 103278, <https://doi.org/10.1016/j.jag.2023.103278>, 2023.
- 675 Li, B., Ryu, Y., Jiang, C., Dechant, B., Liu, J., Yan, Y., and Li, X.: BESSv2.0: A satellite-based and coupled-process model  
for quantifying long-term global land–atmosphere fluxes, *Remote Sensing of Environment*, 295, 113696,  
<https://doi.org/10.1016/j.rse.2023.113696>, 2023.
- Li, F., Xiao, J., Chen, J., Ballantyne, A., Jin, K., Li, B., Abraha, M., and John, R.: Global water use efficiency saturation due  
to increased vapor pressure deficit, *Science*, 381, 672–677, <https://doi.org/10.1126/science.adf5041>, 2023.
- 680 Li, H., Zhang, H., Wang, Y., Zhao, J., Feng, Z., Chen, H., Guo, X., Xiong, T., Xiao, J., and Li, X.: Evaluation of photosynthesis  
estimation from machine learning-based solar-induced chlorophyll fluorescence downscaling from canopy to leaf level,  
*Ecological Indicators*, 166, 112439, <https://doi.org/10.1016/j.ecolind.2024.112439>, 2024.
- Li, M., Cao, S., Zhu, Z., Wang, Z., Myneni, R. B., and Piao, S.: Spatiotemporally consistent global dataset of the GIMMS  
Normalized Difference Vegetation Index (PKU GIMMS NDVI) from 1982 to 2022, *Earth Syst. Sci. Data*, 15, 4181–4203,  
685 <https://doi.org/10.5194/essd-15-4181-2023>, 2023.
- Li, X., Ryu, Y., Xiao, J., Dechant, B., Liu, J., Li, B., Jeong, S., and Gentine, P.: New-generation geostationary satellite reveals  
widespread midday depression in dryland photosynthesis during 2020 western U.S. heatwave, *Science Advances*, 9,  
[eadi0775](https://doi.org/10.1126/sciadv.adi0775), <https://doi.org/10.1126/sciadv.adi0775>, 2023.
- 690 Li, X. and Xiao, J.: A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS,  
and Reanalysis Data, *Remote Sensing*, 11, 517, <https://doi.org/10.3390/rs11050517>, 2019.



- Li, X., Xiao, J., Kimball, J. S., Reichle, R. H., Scott, R. L., Litvak, M. E., Bohrer, G., and Frankenberg, C.: Synergistic use of SMAP and OCO-2 data in assessing the responses of ecosystem productivity to the 2018 U.S. drought, *Remote Sensing of Environment*, 251, 112062, <https://doi.org/10.1016/j.rse.2020.112062>, 2020.
- 695 Liu, X., Guanter, L., Liu, L., Damm, A., Malenovsky, Z., Rascher, U., Peng, D., Du, S., and Gastellu-Etchegorry, J.-P.: Downscaling of solar-induced chlorophyll fluorescence from canopy level to photosystem level using a random forest model, *Remote Sensing of Environment*, 231, 110772, <https://doi.org/10.1016/j.rse.2018.05.035>, 2019.
- Liu, Y., Peñuelas, J., Cescatti, A., Zhang, Y., and Zhang, Z.: Atmospheric Dryness Dominates Afternoon Depression of Global Terrestrial Photosynthesis, *Geophysical Research Letters*, 51, e2024GL110954, <https://doi.org/10.1029/2024GL110954>, 2024.
- 700 Liu, Y., Wang, J., Yao, L., Chen, X., Cai, Z., Yang, D., Yin, Z., Gu, S., Tian, L., Lu, N., and Lyu, D.: The TanSat mission: preliminary global observations, *Science Bulletin*, 63, 1200-1207, <https://doi.org/10.1016/j.scib.2018.08.004>, 2018.
- Liu, Z., Zhao, F., Liu, X., Yu, Q., Wang, Y., Peng, X., Cai, H., and Lu, X.: Direct estimation of photosynthetic CO<sub>2</sub> assimilation from solar-induced chlorophyll fluorescence (SIF), *Remote Sensing of Environment*, 271, 112893, <https://doi.org/10.1016/j.rse.2022.112893>, 2022.
- 705 Lorente, A., Borsdorff, T., Butz, A., Hasekamp, O., aan de Brugh, J., Schneider, A., Wu, L., Hase, F., Kivi, R., Wunch, D., Pollard, D. F., Shiomi, K., Deutscher, N. M., Velasco, V. A., Roehl, C. M., Wennberg, P. O., Warneke, T., and Landgraf, J.: Methane retrieved from TROPOMI: improvement of the data product and validation of the first 2 years of measurements, *Atmos. Meas. Tech.*, 14, 665-684, <https://doi.org/10.5194/amt-14-665-2021>, 2021.
- Lu, J., Wang, G., Chen, T., Li, S., Hagan, D. F. T., Kattel, G., Peng, J., Jiang, T., and Su, B.: A harmonized global land evaporation dataset from model-based products covering 1980–2017, *Earth Syst. Sci. Data*, 13, 5879-5898, <https://doi.org/10.5194/essd-13-5879-2021>, 2021.
- Lucht, W., Schaaf, C. B., and Strahler, A. H.: An algorithm for the retrieval of albedo from space using semiempirical BRDF models, *IEEE Transactions on Geoscience and Remote Sensing*, 38, 977-998, <https://doi.org/10.1109/36.841980>, 2000.
- Magney, T. S., Bowling, D. R., Logan, B. A., Grossmann, K., Stutz, J., Blanken, P. D., Burns, S. P., Cheng, R., Garcia, M. A., 715 Köhler, P., Lopez, S., Parazoo, N. C., Raczka, B., Schimel, D., and Frankenberg, C.: Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence, *Proceedings of the National Academy of Sciences*, 116, 11640-11645, <https://doi.org/10.1073/pnas.1900278116>, 2019.
- Magney, T. S., Frankenberg, C., Köhler, P., North, G., Davis, T. S., Dold, C., Dutta, D., Fisher, J. B., Grossmann, K., Harrington, A., Hatfield, J., Stutz, J., Sun, Y., and Porcar-Castell, A.: Disentangling Changes in the Spectral Shape of 720 Chlorophyll Fluorescence: Implications for Remote Sensing of Photosynthesis, *Journal of Geophysical Research: Biogeosciences*, 124, 1491-1507, <https://doi.org/10.1029/2019JG005029>, 2019.
- Martínez-Sánchez, Á., Arranz, G., and Lozano-Durán, A.: Decomposing causality into its synergistic, unique, and redundant components, *Nature Communications*, 15, 9296, <https://doi.org/10.1038/s41467-024-53373-4>, 2024.



- 725 Miller, D. L., Wolf, S., Fisher, J. B., Zaitchik, B. F., Xiao, J., and Keenan, T. F.: Increased photosynthesis during spring drought  
in energy-limited ecosystems, *Nature Communications*, 14, 7828, <https://doi.org/10.1038/s41467-023-43430-9>, 2023.
- Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., and Moreno, J.: Remote sensing of solar-induced  
chlorophyll fluorescence: Review of methods and applications, *Remote Sensing of Environment*, 113, 2037-2051,  
<https://doi.org/10.1016/j.rse.2009.05.003>, 2009.
- 730 Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M.,  
Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo,  
C., and Thépaut, J. N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, *Earth Syst. Sci. Data*,  
13, 4349-4383, <https://doi.org/10.5194/essd-13-4349-2021>, 2021.
- Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., Ciais, P., Friedlingstein, P., and Sitch, S.: Interannual variation  
of terrestrial carbon cycle: Issues and perspectives, *Global Change Biology*, 26, 300-318,  
735 <https://doi.org/10.1111/gcb.14884>, 2020.
- Porcar-Castell, A., Malenovský, Z., Magney, T., Van Wittenberghe, S., Fernández-Marín, B., Maignan, F., Zhang, Y., Maseyk,  
K., Atherton, J., Albert, L. P., Robson, T. M., Zhao, F., Garcia-Plazaola, J.-I., Ensminger, I., Rajewicz, P. A., Grebe, S.,  
Tikkanen, M., Kellner, J. R., Ihalainen, J. A., Rascher, U., and Logan, B.: Chlorophyll a fluorescence illuminates a path  
740 connecting plant molecular biology to Earth-system science, *Nature Plants*, 7, 998-1009, <https://doi.org/10.1038/s41477-021-00980-4>, 2021.
- Quiros-Vargas, J., Siegmann, B., Damm, A., Krieger, V., Muller, O., and Rascher, U.: Spatial dependency of Solar-induced  
Chlorophyll Fluorescence (SIF)-emitting objects in the footprint of a FLuorescence EXplorer (FLEX) pixel: a SIF-  
downscaling perspective, *EGU General Assembly Conference Abstracts*, EGU22-12671,  
<https://doi.org/10.5194/egusphere-egu22-12671>, 2022.
- 745 Regaieg, O., Lauret, N., Wang, Y., Guilleux, J., Chavanon, E., and Gastellu-Etchegorry, J.-P.: Bi-directional Monte-Carlo  
modelling of solar-induced chlorophyll fluorescence images for 3D vegetation canopies in the DART model, *International  
Journal of Applied Earth Observation and Geoinformation*, 118, 103254, <https://doi.org/10.1016/j.jag.2023.103254>, 2023.
- Regaieg, O., Malenovský, Z., Siegmann, B., Buffat, J., Krämer, J., Lauret, N., and Le Dantec, V.: DART-based temporal and  
750 spatial retrievals of solar-induced chlorophyll fluorescence quantum efficiency from in-situ and airborne crop  
observations, *Remote Sensing of Environment*, 319, 114636, <https://doi.org/10.1016/j.rse.2025.114636>, 2025.
- Regnier, P., Resplandy, L., Najjar, R. G., and Ciais, P.: The land-to-ocean loops of the global carbon cycle, *Nature*, 603, 401-  
410, <https://doi.org/10.1038/s41586-021-04339-9>, 2022.
- Rockström, J., Gupta, J., Qin, D., Lade, S. J., Abrams, J. F., Andersen, L. S., Armstrong McKay, D. I., Bai, X., Bala, G., Bunn,  
S. E., Ciobanu, D., DeClerck, F., Ebi, K., Gifford, L., Gordon, C., Hasan, S., Kanie, N., Lenton, T. M., Loriani, S.,  
755 Liverman, D. M., Mohamed, A., Nakicenovic, N., Obura, D., Ospina, D., Prodani, K., Rammelt, C., Sakschewski, B.,  
Scholtens, J., Stewart-Koster, B., Tharammal, T., van Vuuren, D., Verburg, P. H., Winkelmann, R., Zimm, C., Bennett,  
E. M., Bringezu, S., Broadgate, W., Green, P. A., Huang, L., Jacobson, L., Ndehedehe, C., Pedde, S., Rocha, J., Scheffer,



- M., Schulte-Uebbing, L., de Vries, W., Xiao, C., Xu, C., Xu, X., Zafra-Calvo, N., and Zhang, X.: Safe and just Earth system boundaries, *Nature*, 619, 102-111, <https://doi.org/10.1038/s41586-023-06083-8>, 2023.
- 760 Rodell, M. and Reager, J. T.: Water cycle science enabled by the GRACE and GRACE-FO satellite missions, *Nature Water*, 1, 47-59, <https://doi.org/10.1038/s44221-022-00005-0>, 2023.
- Rosenberg, R., Chapsky, L., Crisp, D., Keller, G., Lee, R., Marchetti, Y., Yu, S., and Eldering, A.: OCO-2 Calibration Refinement Across Versions and Plans for OCO-3, IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, 26 Sept.-2 Oct. 2020, 6381-6384, <https://doi.org/10.1109/IGARSS39084.2020.9324511>, 2020.
- 765 Ryu, Y., Berry, J. A., and Baldocchi, D. D.: What is global photosynthesis? History, uncertainties and opportunities, *Remote Sensing of Environment*, 223, 95-114, <https://doi.org/10.1016/j.rse.2019.01.016>, 2019.
- Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., Miller, C., Frankenberg, C., Hibbard, K., and Cox, P.: Observing terrestrial ecosystems and the carbon cycle from space, *Global Change Biology*, 21, 1762-1776, <https://doi.org/10.1111/gcb.12822>, 2015.
- 770 Shen, H., Wang, Y., Guan, X., Huang, W., Chen, J., Lin, D., and Gan, W.: A Spatiotemporal Constrained Machine Learning Method for OCO-2 Solar-Induced Chlorophyll Fluorescence (SIF) Reconstruction, *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-17, <https://doi.org/10.1109/TGRS.2022.3204885>, 2022.
- Sun, Y., Frankenberg, C., Jung, M., Joiner, J., Guanter, L., Köhler, P., and 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 Sensing of Environment*, 209, 808-823, <https://doi.org/10.1016/j.rse.2018.02.016>, 2018.
- 775 Sun, Y., Gu, L., Wen, J., van der Tol, C., Porcar-Castell, A., Joiner, J., Chang, C. Y., Magney, T., Wang, L., Hu, L., Rascher, U., Zarco-Tejada, P., Barrett, C. B., Lai, J., Han, J., and Luo, Z.: From remotely sensed solar-induced chlorophyll fluorescence to ecosystem structure, function, and service: Part I—Harnessing theory, *Global Change Biology*, 29, 2926-2952, <https://doi.org/10.1111/gcb.16634>, 2023.
- 780 Sun, Y., Wen, J., Gu, L., Joiner, J., Chang, C. Y., van der Tol, C., Porcar-Castell, A., Magney, T., Wang, L., Hu, L., Rascher, U., Zarco-Tejada, P., Barrett, C. B., Lai, J., Han, J., and Luo, Z.: From remotely-sensed solar-induced chlorophyll fluorescence to ecosystem structure, function, and service: Part II—Harnessing data, *Global Change Biology*, 29, 2893-2925, <https://doi.org/10.1111/gcb.16646>, 2023.
- Tao, S., Chen, J. M., Zhang, Z., Zhang, Y., Ju, W., Zhu, T., Wu, L., Wu, Y., and Kang, X.: A high-resolution satellite-based solar-induced chlorophyll fluorescence dataset for China from 2000 to 2022, *Scientific Data*, 11, 1286, <https://doi.org/10.1038/s41597-024-04101-6>, 2024.
- 785 van der Tol, C., Berry, J. A., Campbell, P. K. E., and Rascher, U.: Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence, *Journal of Geophysical Research: Biogeosciences*, 119, 2312-2327, <https://doi.org/10.1002/2014JG002713>, 2014.



- 790 van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., and Su, Z.: An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance, *Biogeosciences*, 6, 3109-3129, <https://doi.org/10.5194/bg-6-3109-2009>, 2009.
- Wang, S., Zhang, Y., Ju, W., Chen, J. M., Ciais, P., Cescatti, A., Sardans, J., Janssens, I. A., Wu, M., Berry, J. A., Campbell, E., Fernández-Martínez, M., Alkama, R., Sitch, S., Friedlingstein, P., Smith, W. K., Yuan, W., He, W., Lombardozzi, D.,  
795 Kautz, M., Zhu, D., Lienert, S., Kato, E., Poulter, B., Sanders, T. G. M., Krüger, I., Wang, R., Zeng, N., Tian, H., Vuichard, N., Jain, A. K., Wiltshire, A., Haverd, V., Goll, D. S., and Peñuelas, J.: Recent global decline of CO<sub>2</sub> fertilization effects on vegetation photosynthesis, *Science*, 370, 1295-1300, <https://doi.org/10.1126/science.abb7772>, 2020.
- Wankmüller, F. J. P., Delval, L., Lehmann, P., Baur, M. J., Cecere, A., Wolf, S., Or, D., Javaux, M., and Carminati, A.: Global  
800 influence of soil texture on ecosystem water limitation, *Nature*, 635, 631-638, <https://doi.org/10.1038/s41586-024-08089-2>, 2024.
- Wei, S., Fang, H., Schaaf, C. B., He, L., and Chen, J. M.: Global 500 m clumping index product derived from MODIS BRDF data (2001–2017), *Remote Sensing of Environment*, 232, 111296, <https://doi.org/10.1016/j.rse.2019.111296>, 2019.
- Wen, J., Köhler, P., Duveiller, G., Parazoo, N. C., Magney, T. S., Hooker, G., Yu, L., Chang, C. Y., and Sun, Y.: A framework  
805 for harmonizing multiple satellite instruments to generate a long-term global high spatial-resolution solar-induced chlorophyll fluorescence (SIF), *Remote Sensing of Environment*, 239, 111644, <https://doi.org/10.1016/j.rse.2020.111644>, 2020.
- Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J. A., Huete, A. R., Ichii, K., Ni, W., Pang, Y., Rahman, A. F., Sun, G., Yuan, W., Zhang, L., and Zhang, X.: Remote sensing of the terrestrial carbon cycle: A review of advances over 50  
810 years, *Remote Sensing of Environment*, 233, 111383, <https://doi.org/10.1016/j.rse.2019.111383>, 2019.
- Xiao, J., Fisher, J. B., Hashimoto, H., Ichii, K., and Parazoo, N. C.: Emerging satellite observations for diurnal cycling of ecosystem processes, *Nature Plants*, 7, 877-887, <https://doi.org/10.1038/s41477-021-00952-8>, 2021.
- Yang, P., Prikaziuk, E., Verhoef, W., and van der Tol, C.: SCOPE 2.0: a model to simulate vegetated land surface fluxes and satellite signals, *Geosci. Model Dev.*, 14, 4697-4712, <https://doi.org/10.5194/gmd-14-4697-2021>, 2021.
- 815 Yang, P. and van der Tol, C.: Linking canopy scattering of far-red sun-induced chlorophyll fluorescence with reflectance, *Remote Sensing of Environment*, 209, 456-467, <https://doi.org/10.1016/j.rse.2018.02.029>, 2018.
- Yang, P., Verhoef, W., and van der Tol, C.: The mSCOPE model: A simple adaptation to the SCOPE model to describe reflectance, fluorescence and photosynthesis of vertically heterogeneous canopies, *Remote Sensing of Environment*, 201, 1-11, <https://doi.org/10.1016/j.rse.2017.08.029>, 2017.
- 820 Yang, Y., Roderick, M. L., Guo, H., Miralles, D. G., Zhang, L., Fatichi, S., Luo, X., Zhang, Y., McVicar, T. R., Tu, Z., Keenan, T. F., Fisher, J. B., Gan, R., Zhang, X., Piao, S., Zhang, B., and Yang, D.: Evapotranspiration on a greening Earth, *Nature Reviews Earth & Environment*, 4, 626-641, <https://doi.org/10.1038/s43017-023-00464-3>, 2023.



- 825 Yu, P., Zhou, T., Luo, H., Liu, X., Shi, P., Zhang, Y., Zhang, J., Zhou, P., and Xu, Y.: Global Pattern of Ecosystem Respiration  
Tendencies and Its Implications on Terrestrial Carbon Sink Potential, *Earth's Future*, 10, e2022EF002703,  
<https://doi.org/10.1029/2022EF002703>, 2022.
- Zeng, Y., Badgley, G., Dechant, B., Ryu, Y., Chen, M., and Berry, J. A.: A practical approach for estimating the escape ratio  
of near-infrared solar-induced chlorophyll fluorescence, *Remote Sensing of Environment*, 232, 111209,  
<https://doi.org/10.1016/j.rse.2019.05.028>, 2019.
- 830 Zeng, Y., Hao, D., Park, T., Zhu, P., Huete, A., Myneni, R., Knyazikhin, Y., Qi, J., Nemani, R. R., Li, F., Huang, J., Gao, Y.,  
Li, B., Ji, F., Köhler, P., Frankenberg, C., Berry, J. A., and Chen, M.: Structural complexity biases vegetation greenness  
measures, *Nature Ecology & Evolution*, 7, 1790-1798, <https://doi.org/10.1038/s41559-023-02187-6>, 2023.
- Zhang, K., Chen, H., Ma, N., Shang, S., Wang, Y., Xu, Q., and Zhu, G.: A global dataset of terrestrial evapotranspiration and  
soil moisture dynamics from 1982 to 2020, *Scientific Data*, 11, 445, <https://doi.org/10.1038/s41597-024-03271-7>, 2024.
- 835 Zhang, Y., Fang, J., Smith, W. K., Wang, X., Gentine, P., Scott, R. L., Migliavacca, M., Jeong, S., Litvak, M., and Zhou, S.:  
Satellite solar-induced chlorophyll fluorescence tracks physiological drought stress development during 2020 southwest  
US drought, *Global Change Biology*, 29, 3395-3408, <https://doi.org/10.1111/gcb.16683>, 2023.
- Zhang, Y., Joiner, J., Alemohammad, S. H., Zhou, S., and Gentine, P.: A global spatially contiguous solar-induced fluorescence  
(CSIF) dataset using neural networks, *Biogeosciences*, 15, 5779-5800, <https://doi.org/10.5194/bg-15-5779-2018>, 2018.
- 840 Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., and Dong, J.: A global moderate resolution dataset of gross primary  
production of vegetation for 2000–2016, *Scientific Data*, 4, 170165, <https://doi.org/10.1038/sdata.2017.165>, 2017.
- Zhang, Z., Cescatti, A., Wang, Y.-P., Gentine, P., Xiao, J., Guanter, L., Huete, A. R., Wu, J., Chen, J. M., Ju, W., Peñuelas, J.,  
and Zhang, Y.: Large diurnal compensatory effects mitigate the response of Amazonian forests to atmospheric warming  
and drying, *Science Advances*, 9, eabq4974, <https://doi.org/10.1126/sciadv.abq4974>, 2023.
- 845 Zhang, Z., Guanter, L., Porcar-Castell, A., Rossini, M., Pacheco-Labrador, J., and Zhang, Y.: Global modeling diurnal gross  
primary production from OCO-3 solar-induced chlorophyll fluorescence, *Remote Sensing of Environment*, 285, 113383,  
<https://doi.org/10.1016/j.rse.2022.113383>, 2023.
- Zhang, Z., Zhang, Y., Chen, J. M., Ju, W., Migliavacca, M., and El-Madany, T. S.: Sensitivity of Estimated Total Canopy SIF  
Emission to Remotely Sensed LAI and BRDF Products, *Journal of Remote Sensing*,  
<https://doi.org/10.34133/2021/9795837>, 2021.
- 850 Zhao, W., Zhu, Z., Cao, S., Li, M., Zha, J., Pu, J., and Myneni, R. B.: A global dataset of the fraction of absorbed  
photosynthetically active radiation for 1982–2022, *Scientific Data*, 11, 707, <https://doi.org/10.1038/s41597-024-03561-0>, 2024.
- 855 Zheng, C., Wang, S., Chen, J. M., Xiao, J., Chen, J., Zhang, Z., and Forzieri, G.: Estimating global transpiration from  
TROPOMI SIF with angular normalization and separation for sunlit and shaded leaves, *Remote Sensing of Environment*,  
319, 114586, <https://doi.org/10.1016/j.rse.2024.114586>, 2025.