Lawrence Berkeley National Laboratory
Climate & Ecosystems

Title
A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks

Permalink

Journal
Biogeosciences, 15(19)

ISSN
1726-4170

Authors
Zhang, Yao
Joiner, Joanna
Alemohammad, Seyed Hamed
et al.

Publication Date
2018

DOI
10.5194/bg-15-5779-2018

Peer reviewed

A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks

Yao Zhang[1], Joanna Joiner[2], Seyed Hamed Alemohammad[3], Sha Zhou[1], and Pierre Gentine[1,4]

[1]Department of Earth and Environmental Engineering, Columbia University, New York, NY 10027, USA

[2]NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

[3]Radiant Earth, Washington, DC 20005, USA

[4]Earth Institute, Columbia University, New York, NY 10027, USA

Correspondence: Yao Zhang (zy2309@columbia.edu)

Abstract

Satellite-retrieved solar-induced chlorophyll fluorescence (SIF) has shown great potential to monitor the photosynthetic activity of terrestrial ecosystems. However, several issues, including low spatial and temporal resolution of the gridded datasets and high uncertainty of the individual retrievals, limit the applications of SIF. In addition, inconsistency in measurement footprints also hinders the direct comparison between gross primary production (GPP) from eddy covariance (EC) flux towers and satellite-retrieved SIF. In this study, by training a neural network (NN) with surface reflectance from the MODerate-resolution Imaging Spectroradiometer (MODIS) and SIF from Orbiting Carbon Observatory-2 (OCO-2), we generated two global spatially contiguous SIF (CSIF) datasets at moderate spatiotemporal ($0.05°$ 4-day) resolutions during the MODIS era, one for clear-sky conditions (2000–2017) and the other one in all-sky conditions (2000–2016). The clear-sky instantaneous CSIF ($CSIF_{clear-inst}$) shows high accuracy against the clear-sky OCO-2 SIF and little bias across biome types. The all-sky daily average CSIF ($CSIF_{all-daily}$) dataset exhibits strong spatial, seasonal and interannual dynamics that are consistent with daily SIF from OCO-2 and the Global Ozone Monitoring Experiment-2 (GOME-2). An increasing trend (0.39 %) of annual average $CSIF_{all-daily}$ is also found, confirming the greening of Earth in most regions. Since the difference between satellite-observed SIF and CSIF is mostly caused by the environmental down-regulation on SIF_{yield}, the ratio between OCO-2 SIF and $CSIF_{clear-inst}$ can be an effective indicator of drought stress that is more sensitive than the normalized difference vegetation index and enhanced vegetation index. By comparing $CSIF_{all-daily}$ with GPP estimates from 40 EC flux towers across the globe, we find a large cross-site variation (c.v. = 0.36) of the GPP–SIF relationship with the highest regression slopes for evergreen needleleaf forest. However, the cross-biome variation is relatively limited (c.v. = 0.15). These two contiguous SIF datasets and the derived GPP–SIF relationship enable a better understanding of the spatial and temporal variations of the GPP across biomes and climate.

1 Introduction

Obtaining a spatiotemporal continuous photosynthetic carbon fixation or gross primary production (GPP) dataset is crucial to food security, ecosystem service and health evaluation, and global carbon cycle studies (Beer et al., 2010). However, this is not possible without remote sensing data, since in situ carbon flux measurements, such as FLUXNET (Baldocchi et al., 2001), are usually costly and have limited spatial and temporal coverage (Schimel et al., 2015). Many remote-sensing-based productivity efficiency models (PEMs) have been built, but the model structure and parameterizations differ from each other and the performance of most models is not satisfactory in terms of simulated interannual variability and trends (Anav et al., 2015; Chen et al., 2017).

Müller (1874) found that the chlorophyll fluorescence (Chl F) from a dilute chlorophyll solution was much stronger than the Chl F from a green leaf, suggesting that an alternative energy pathway exists for leaves in vivo. In the 1980s, scientists found that plant photosynthesis and heat dissipation are two alternatives to quench the excited chlorophyll, and there is a close linkage between Chl F and carbon assimilation rate (Genty et al., 1989; Krause and Weis, 1991). Leaf-level photosynthesis (A_{leaf}) and fluorescence (Chl F) share the same source of energy originating from photosynthetically active radiation (PAR) absorbed by chlorophyll ($APAR_{chl}$), which can be written using a light-use efficiency (LUE) approach (Monteith, 1972):

$$Chl\ F = PAR \times fPAR_{chl} \times \phi_F \qquad (1)$$
$$A_{leaf} = PAR \times fPAR_{chl} \times \phi_P, \qquad (2)$$

where ϕ_F and ϕ_P represent the efficiencies for Chl F emission and photochemistry, respectively. $fPAR_{chl}$, being different from the conventional definition of fraction of photosynthetically active radiation absorption, only considers the fractions absorbed by chlorophyll pigments where the photosynthesis and fluorescence originate (Zhang et al., 2018c). However, Chl F measurements have been mostly conducted at the leaf level, using pulse amplitude modulation (PAM) fluorometers (Porcar-Castell et al., 2008; Roháček and Barták, 1999). In this case, the measured Chl F intensity is not induced by the Sun but by the modulated light source. Although the absolute value of the Chl F intensity does not directly link to A_{leaf}, it can still be used to calculate the fluorescence yield and investigate the reaction mechanism of the energy partitioning during the light reaction, and to calculate the quantum yield for photochemistry or as a tool to detect plant reactions under stress (Adams and Demmig-Adams, 2004; Flexas et al., 2002).

The successful retrieval of solar-induced (steady-state) chlorophyll fluorescence (SIF) from satellites has made it possible for vegetation photosynthetic activities to be observed at the global scale (Frankenberg et al., 2011; Guanter et al., 2012; Joiner et al., 2011, 2013). Satellite SIF can be expressed as a function similar to the Chl F at the leaf level but with extra

terms considering the radiative transfer within the canopy and through the atmosphere (Joiner et al., 2014):

$$\mathrm{SIF_{sat}}(\lambda) = \mathrm{PAR} \times f\mathrm{PAR_{chl}} \times \Theta_F(\lambda) \times f_{esc}(\lambda, \theta_s, \theta_v, \phi)$$
$$\times \tau_{atm}(\lambda, \theta_s, \theta_v, \phi), \qquad (3)$$

where the satellite-retrieved SIF ($\mathrm{SIF_{sat}}$), fluorescence yield (Θ_F), f_{esc} and τ_{atm} are all functions of the wavelength (λ); in addition, f_{esc} and τ_{atm} are also affected by sun-sensor geometry characterized by Sun zenith angle (SZA; θ_s), view zenith angle (θ_v) and relative azimuth angle (ϕ). f_{esc} is a factor describing how much SIF emitted by the chloroplast leaves the canopy, and τ_{atm} is a function of atmospheric optical depth, which indicates how much SIF that leaves the canopy top passes through the atmosphere before it is captured by the satellite sensors. It should be noted that the fraction of PAR for fluorescence ($f\mathrm{PAR}_F$) may have a different activation spectrum than that for photosynthesis ($f\mathrm{PAR_{chl}}$), but this difference is ignored here for simplicity. Although additional factors come into play during this process, satellite-retrieved SIF shows high consistency with GPP using both model simulations and ground-based measurements from eddy covariance (EC) flux towers, at least at the monthly timescale (Guanter et al., 2014; Li et al., 2018a; Zhang et al., 2016c, b). In addition, recent studies suggest that the GPP–SIF relationship is consistent across biome types (Sun et al., 2017). This finding, if valid across all biomes, would greatly benefit the usage of SIF for model benchmarking (Luo et al., 2012) and global GPP estimation.

However, several issues hinder exploring the relationship between SIF and in situ GPP estimates. Since the SIF signal is very small and sensors used to retrieve SIF were not initially built to estimate SIF, the satellite-retrieved SIF usually has a large footprint and large uncertainties in individual retrievals (Frankenberg et al., 2014; Joiner et al., 2013, 2016). For instance, the SIF retrieval from the Global Ozone Monitoring Experiment-2 (GOME-2) has a footprint of 40 km×40 km or larger, and the SIF from the Greenhouse gases Observing SATellite (GOSAT) has a circular footprint with 10.5 km in diameter. Direct comparison between the satellite-retrieved SIF signal and GPP estimates from EC flux tower sites thus faces the problem of spatial inconsistency except in areas of large homogenous landscape, e.g., the US Midwest cropland (Zhang et al., 2014) or boreal evergreen forests (Walther et al., 2016). However, corn (C4 pathway) and soybean (C3 pathway) in SIF footprints have different electron use efficiencies (Guan et al., 2016), which should affect the relationship between SIF and GPP. The low precision of SIF measurements also leads to a need for averaging multiple pixels either in space or time before being used.

SIF retrieved from the Orbiting Carbon Observatory-2 (OCO-2) satellite partially solved this issue with a much smaller footprint size (1.3 km×2.25 km), higher signal-to-noise ratio compared to GOSAT (relatively higher SIF retrieval accuracy) and much larger numbers of observations per day (Frankenberg et al., 2014; Sun et al., 2018). However, due to the sparse

sampling strategy and long revisit cycle, the OCO-2 SIF data have large gaps between nearby swaths, and the average sampling frequency for each flux tower site is only $3.21\,year^{-1}$ during 2015–2016 (Lu et al., 2018). In addition, OCO-2 is often aggregated to a monthly dataset at relatively coarse spatial resolution, typically at $1° \times 1°$, which limits its application in small regions. Although several statistical methods have been proposed to downscale satellite observations to finer spatial–temporal resolutions (Tadić et al., 2015, 2017), considering the large land surface heterogeneity and wide gaps between OCO-2 swaths (~100 km), it could be challenging to apply these methods to OCO-2 SIF.

A high spatiotemporal resolution SIF dataset is needed to improve our understanding of the relationship between SIF and GPP and provide accurate GPP estimates at the global scale. As discussed previously, the satellite-observed SIF contains signals from $APAR_{chl}$, fluorescence yield, and canopy and atmospheric attenuation. $APAR_{chl}$ is considered to be the first-order approximation of SIF as it exhibits high correlation with SIF at the canopy scale (Du et al., 2017; Rossini et al., 2016; Verrelst et al., 2015; Zhang et al., 2018c). Previous studies have shown that fPAR$_{chl}$ can be inversely estimated using the surface reflectances and radiative transfer models (Zhang et al., 2005, 2016a). The canopy structure information that affects the SIF reabsorption within canopy is also embedded in the near-infrared reflectance (Badgley et al., 2017; Knyazikhin et al., 2013; Yang and van der Tol, 2018). Many previous studies have shown high correlation between SIF and vegetation indices (VIs), especially VIs related to the chlorophyll concentration (Frankenberg et al., 2011; Guanter et al., 2012). Therefore, broadband surface reflectances may have the potential to be used to estimate vegetation information and reconstruct global SIF (Duveiller and Cescatti, 2016; Gentine and Alemohammad, 2018a). However, physical models that can predict SIF (e.g., the Soil Canopy Observation, Photochemistry and Energy fluxes, SCOPE; van der Tol et al., 2009) often require many parameters, making it difficult to use reflectance and modeling to predict SIF at a larger scale.

Neural networks (NNs), together with many other machine learning algorithms, have been used with remote sensing datasets in the Earth sciences, especially for carbon and water fluxes estimation (Alemohammad et al., 2017; Jung et al., 2011; Tramontana et al., 2016), land cover mapping (Kussul et al., 2017; Zhu et al., 2017), soil moisture retrievals and downscaling (Alemohammad et al., 2018; Kolassa et al., 2018) or to bypass parameterization (Gentine et al., 2018). These studies mostly attempted to link the satellite signals with limited in situ observation or model simulations for model training, while taking advantage of the large amount of data in remote sensing observations; they applied the trained algorithm to generate a regional or global dataset. Reconstructing SIF from surface reflectance, on the other hand, uses no in situ observations but faces more problems related to the satellite data quality assurance. The SIF–reflectance relationship is

complicated, and the NN benefits from the fact that an explicit physical and radiative transfer relationship is not required.

In this study, we aim to generate a global contiguous SIF (CSIF) product based on the SIF retrievals from OCO-2 and surface reflectances from Moderate-resolution Imaging Spectroradiometer (MODIS) aboard the Terra and Aqua satellites. The CSIF dataset aims to fill the spatial gaps between the OCO-2 swaths and temporal gaps due to the long revisit cycle of OCO-2. Specifically, we first trained and validated the NN using the satellite-observed instantaneous SIF under clear-sky conditions so that the relationship is not affected by cloud-related artifacts. We further generated two SIF products, namely the clear-sky instantaneous SIF ($CSIF_{clear-inst}$) and the all-sky daily SIF ($CSIF_{all-daily}$). The spatiotemporal variations of these CSIF products were analyzed and compared with SIF from OCO-2 and three other GOME-2 SIF datasets. Finally, we showed two applications of CSIF datasets: (1) monitoring drought impact using $CSIF_{clear-inst}$ and OCO-2 SIF; (2) evaluating the GPP–SIF relationship by comparing CSIF with GPP estimates from 40 flux tower sites.
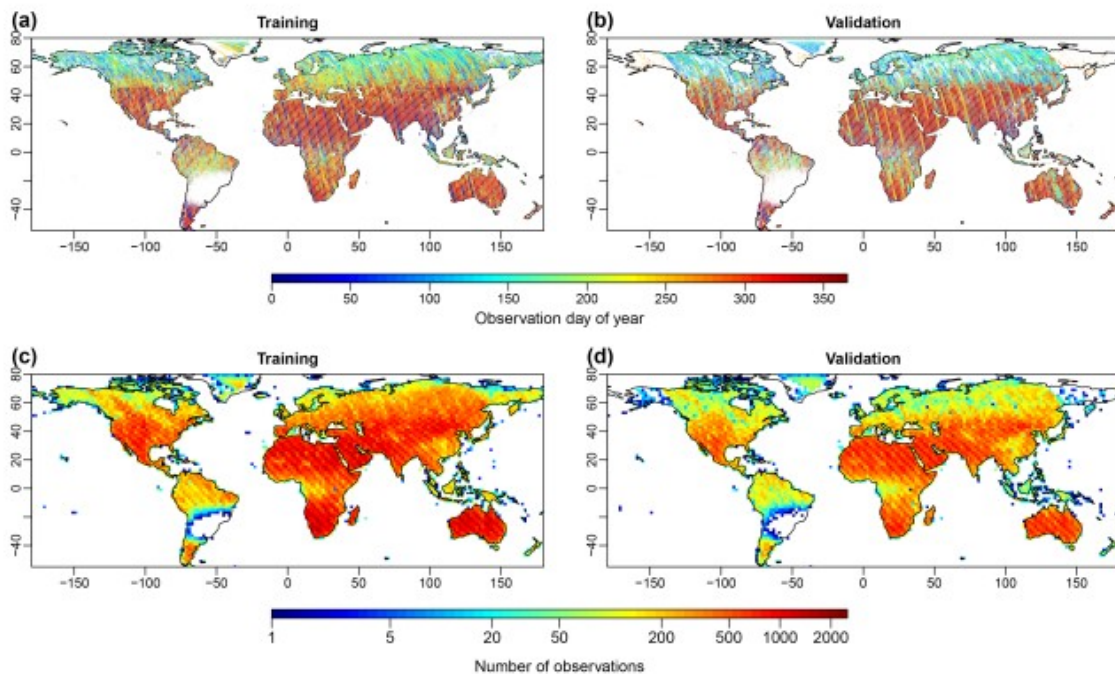
Figure 1. Samples that were used for NN training (years 2015 and 2016) and validation (2014 and 2017). Panels **(a)** and **(b)** show the spatial distribution of observation day of year (DOY) and panels **(c)** and **(d)** show the spatial distribution of the sample density. Each point in panels **(a)** and **(c)** represents a 0.05° training grid cell. Limited observations in South America were caused by the South Atlantic Anomaly (Sun et al., 2018).

2 Materials and methods

2.1 OCO-2 solar-induced chlorophyll fluorescence dataset

The 8100r OCO-2 SIF data between September 2014 and December 2017 were used for NN training and evaluation (Frankenberg, 2015; Frankenberg

et al., 2014; Sun et al., 2018). The daily sounding-based SIF retrievals at 757 nm were first aggregated to 0.05° (around 5.6 km×5.6 km at the Equator), consistent with MODIS Climate Modeling Grid (CMG) resolution. The reasons for using this resolution include the following: (1) it is directly comparable (of the same order of magnitude) to the OCO-2 SIF footprint size (around 1.3 km×2.25 km) and the samples within each grid cell can be more evenly distributed and thus more representative of the grid cell SIF values than using much coarser 1°×1° or 2°×2° grids; (2) by averaging multiple observations, the uncertainty in the SIF signal can be approximately reduced by a factor of $\sqrt{n}$ (n is the number of observations within this grid cell), assuming independent estimates and homogeneous SIF value within each grid cell (Frankenberg et al., 2014). During this aggregation, we only used cloud-free observations indicated by the OCO-2 cloud flag. For each 0.05° grid cell, the SIF value was only calculated when it contained more than five cloud-free SIF soundings. Although several studies have shown that SIF at different wavelengths has different sensitivity to stress and leaf and canopy reabsorption (Porcar-Castell et al., 2014; Rossini et al., 2015, 2016), we only use SIF at 757 nm since it showed superior performance to SIF at 771 nm in predicting GPP (Li et al., 2018a). The years 2015 and 2016 were used for training and 2014 and 2017 were used for validation. Altogether, 2 947 819 SIF grid cells passed quality check during 2014–2017. Figure 1 shows the spatial distribution of the SIF grid cells used for training and validation (test). It should be noted that the OCO-2 satellite started obtaining data from September 2014 and experienced some malfunctioning during August and September in 2017, causing lower coverage for validation samples in boreal regions.

In addition to these cloud-free observations, we also calculated the all-sky SIF at 0.05° resolution. All SIF retrievals that passed the suggested quality checks (documented in detailed by Sun et al., 2018) were used for the aggregation. The aggregated all-sky instantaneous SIF retrievals were converted to daily values based on the solar zenith angle (Zhang et al., 2018a). We used this dataset to validate the all-sky daily SIF ($CSIF_{all-daily}$) (see Sect. 2.5). In both cloud-free and all-sky aggregations, only observations from the nadir mode were used since glint mode tends to underestimate SIF (Sun et al., 2018).

2.2 MODIS reflectance dataset (MCD43C4 V006)

We used the 0.05° daily nadir bidirectional reflectance distribution adjusted reflectance (NBAR) product from MODIS (MCD43C4 V006) during 2000–2017 as input variables for the NN. The NBAR product computed the reflectance at a nadir viewing angle for each pixel at local solar noon. Compared to MOD09 or MYD09 surface reflectance product, it removed the angle effects and therefore should be more stable and consistent (Schaaf et al., 2002). This dataset was processed in two different ways for training and prediction. For the training process, following Gentine and Alemohammad (2018a), we extracted the reflectance from the first four bands of MODIS (centered at

645, 858, 469 and 555 nm, respectively) for the corresponding pixels and days when the cloud-free SIF observations were obtained. It should be noted that although the MCD43C4 is generated for each day and can match the daily SIF observations, the MCD43C4 NBAR uses 16 days worth of inputs and so that the reflectances includes the information on other days than the day of interest. However, we consider this to have limited effects since (1) the vegetation growth/changes are continuous in time, (2) the NBAR product uses 16-day data but also emphasizes the specific day of interest (Schaaf, 2018). These four bands were selected because the visible and near-infrared bands included most of the vegetation information and drives the variation of SIF (Verrelst et al., 2015). We also tested using all seven bands with/without the meteorological variables (temperature and vapor pressure deficit, obtained from the OCO-2 SIF lite files) to train the NN, but the improvements in training and validation were very minor (R^2 increased by less than 0.01; data not shown), and thus we decided not to use it. Since SIF is very sensitive to the incoming solar radiation, using cloud-free training samples can minimize the uncertainty of using cosine of the solar zenith angle as the proxy of incoming PAR. It should be noted that the training dataset may contain snow-affected samples, but these were not removed to get a more realistic prediction of SIF during winter.

For prediction, we first aggregated the daily reflectance to 4 days. The 4-day temporal resolution is selected to reach a balance among application requirements, information redundancy and dataset sizes. During this process, we used a gap-filling and smoothing algorithm to reconstruct the surface reflectance for the four bands. The detailed description of the gap-filling algorithm can be found in Zhang et al. (2017a). In this study, we slightly modified the algorithm by not applying the best index slope extraction (BISE) algorithm and Savitzky–Golay (SG) filter. The reconstructed 4-day 0.05° reflectance together with other datasets allowed us to predict SIF at 4-day 0.05° resolution during 2000–2017. Since this processing does not involve any extra information and only uses the reflectance observations from the successful model inversion, it should be comparable to the reflectance used for NN training.

2.3 Machine learning algorithms

A feed-forward NN is a number of computational nodes (called neurons) structured in a single or multi-layer architecture. Each neuron is connected with all neurons in the previous layer and next layer. The neuron values are calculated using an activation function with a pre-activated value, i.e., the weighted sum of all neurons in previous layer plus biases. The training of the NN attempts to optimize these weights and biases so that the differences between the output variable in the training data and NN prediction are minimized. In this study, we used Tensorflow (https://www.tensorflow.org, last access: 27 September 2018) and built feed-forward networks with one to three layers and two to nine neurons for each layer. After training models with data from 2015 and 2016, we validated the models using the test

dataset from the years 2014 and 2017. We then picked the one with best performance and simplest structure for SIF prediction. The rectified linear unit (ReLU) was used as the activation function since it has shown better performance in our application, and the cost function used is the root mean square error (RMSE). We used 50 epochs with a batch size of 1024. Before training, each variable was normalized by its mean and standardized deviation. Since the NN is not deep and there is no sign of overfitting, we did not use any regularization methods during the training.

2.4 Reconstructing the clear-sky instantaneous SIF and daily SIF

During the NN training process, we only used the SIF and reflectance data in clear-sky conditions, and therefore cos(SZA) was used as a proxy of the incoming photosynthetically active radiation at top of canopy. In the prediction process, we also used the calculated cos(SZA) based on the satellite overpass local solar time and latitude. Since we did not consider the cloud and aerosol attenuation of the PAR, this product was referred to as the "clear-sky instantaneous SIF ($CSIF_{clear-inst}$)".

In addition to the clear-sky instantaneous SIF, we also calculated two daily SIF data by assuming that the incoming solar radiation is the only factor that drives the diurnal cycle (Zhang et al., 2018a). All-sky daily SIF ($CSIF_{all-daily}$) can be calculated using the clear-sky top-of-canopy radiation ($PAR_{clear-inst}$) and the daily average radiation from the Breathing Earth System Simulator (BESS) (Ryu et al., 2018):

$$CSIF_{all-daily} = \frac{CSIF_{clear-inst}}{PAR_{clear-inst}} \times PAR_{daily}^{BESS}, \qquad (4)$$

where $PAR_{clear-inst}$ was calculated following previous studies that only considered atmospheric scattering (see Appendix A1). Clear-sky daily SIF ($CSIF_{clear-daily}$) assumes no cloud throughout the day and can be calculated by multiplying $CSIF_{clear-inst}$ with a daily correction factor (γ) (Zhang et al., 2018a):

$$CSIF_{clear-daily} = CSIF_{clear-inst} \times \gamma. \qquad (5)$$

γ is calculated as the ratio between the cos(SZA) during the satellite overpass and the daily averaged cos(SZA).

2.5 GOME-2 SIF (SIF_{GOME-2}), reconstructed SIF from GOME-2 ($RSIF_{GOME-2}$) and SIF^* datasets

In this study, we also used the GOME-2 SIF (SIF_{GOME-2}), reconstructed SIF from GOME-2 ($RSIF_{GOME-2}$) using machine learning and the SIF^* dataset in comparison with our contiguous SIF from OCO-2. The GOME-2 SIF V27 was retrieved using a principle component analysis algorithm in the wavelength range 734–758 nm (Joiner et al., 2013, 2016). The V27 version, compared to the widely used V26, provides daily correction factor and improved bias correction and calibration (https://avdc.gsfc.nasa.gov/, last access: 27

September 2018). The level-3 monthly 0.5° daily average SIF was used to compare with $CSIF_{all-daily}$.

$RSIF_{GOME-2}$ (Gentine and Alemohammad, 2018a) uses a similar machine learning technique approach to CSIF but the training is based on the biweekly gridded SIF product from GOME-2 and the 8-day MYD09A1 reflectance dataset. Both clear-sky and cloudy-sky SIF are used for NN training. This dataset has a spatial resolution of 0.05° and 8-day temporal resolution. Both $RSIF_{GOME-2}$ and $CSIF_{all-daily}$ were aggregated to the 0.5° and semi-monthly to facilitate the comparison.
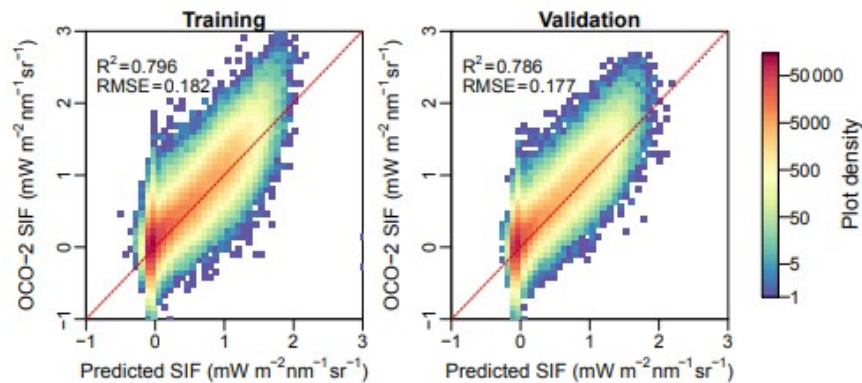

Figure 2. Predicted SIF in comparison with the OCO-2 SIF. Red lines represent the regression slope and the black dotted lines represent the 1 : 1 line.

The SIF^* dataset (Duveiller and Cescatti, 2016) applies a statistical method and calibrates a model that links monthly 0.5° SIF to the normalized difference vegetation index (NDVI), evapotranspiration (ET) and land surface temperature (LST) dataset for each moving window. The model and its spatiotemporally varied parameters were then applied to finer resolution dataset (NDVI, ET, LST) with a weighted average to generate SIF at 0.05° resolution. In this study, we used the 0.5° monthly SIF^* dataset during 2007–2013 to compare with CSIF.

2.6 Comparing CSIF with GPP at flux tower sites

We further compared the CSIF dataset to GPP estimates from the tier 1 FLUXNET2015 datasets (http://fluxnet.fluxdata.org, last access: 27 September 2018) to investigate the SIF–GPP relationship. Since the CSIF dataset is continuous in space and time, it provides many more samples pairs compared to the original OCO-2 SIF data (Lu et al., 2018). However, because of the landscape heterogeneity and inconsistency between the flux tower footprint and CSIF pixel size, a rigorous site selection is needed. We took the vegetation growth condition into consideration during this process: (1) the annual average, minimum, maximum and seasonal variability (represented by standard deviation) of NDVI (from MOD13Q1 C6) for the target pixel (where the flux tower is located, 250 m by 250 m) need to be similar (within 20 % difference or 0.05 NDVI) to the neighboring (5 km by 5

km) area; (2) the maximum NDVI value for target pixel and neighboring area needs to be greater than 0.2 (not barren). The daily GPP estimates, estimated using nighttime method (Reichstein et al., 2005), were averaged and aggregated into 4-day values to compare with CSIF. The 4-day GPP based on more than 80 % of half-hourly valid (not gap-filled) net ecosystem exchange was retained. Only sites that have at least 92 valid observations (1 year) were used. Only 40 out of 166 sites passed these criteria and were grouped into different biome types (Table S1). In addition to $CSIF_{all\text{-}daily}$, we also calculated $CSIF_{clear\text{-}daily}$ and $CSIF_{site}$ which used flux-tower-observed radiation instead of PAR_{daily}^{BESS} in Eq. (4).
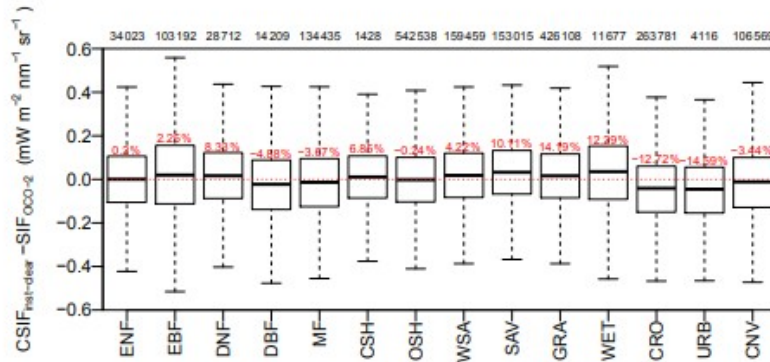

Figure 3. Difference between $CSIF_{clear\text{-}inst}$ and $SIF_{OCO\text{-}2}$ for major biome types during 2014–2017. The MODIS land cover dataset for 2010 was used to identify the land cover type for each 0.05° grid cell (Friedl et al., 2010). The red percentages above each box represent the mean relative error, and the numbers on top of the figure frame represent the total sample numbers for each biome type. Abbreviations are as follows: ENF, evergreen needleleaf forest; EBF, evergreen broadleaf forest; DNF, deciduous needleleaf forest; DBF, deciduous broadleaf forest; MF, mixed forest; CSH, closed shrubland; OSH, open shrubland; WSA, woody savannas; SAV, savannas; GRA, grassland; WET, wetland; CRO, cropland; URB, urban; CNV, cropland or natural vegetation mosaics.

3 Results

3.1 NN training and validation

The NN with one layer and five neurons generally predicts the OCO-2 SIF during the training with a coefficient of determination (R^2) around 0.8 and an RMSE of 0.18 mW m^{-2} nm^{-1} sr^{-1} (Fig. 2). The model also performs well in the validation (R^2=0.79, RMSE = 0.18) and does not show effects of overfitting. Using a variety of layer (one to three) and neuron (two to nine) combinations, we found that one layer with five neurons exhibited slightly higher model performance during the validation compared with a more complex NN (Fig. A1 in Appendix). Therefore, we chose to use the four-band reflectances to feed the one-layer-five-neuron NN to generate the contiguous SIF for 2000 to 2017 when MCD43C4 NBAR dataset is available.
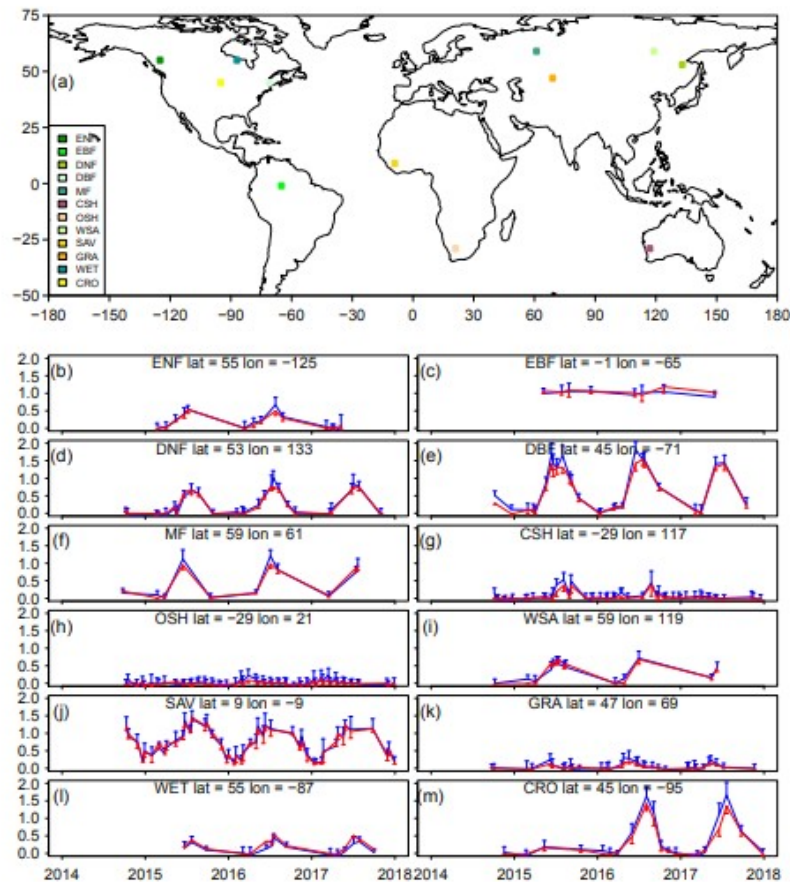
Figure 4. Comparison of predicted SIF by NN and OCO-2-observed SIF for 12 samples ($2° \times 2°$) of major vegetated land cover types during 2014 to 2017. All samples in the training and validation are used. The blue color represents the observed SIF by OCO-2, and the red color represents the SIF prediction by NN. The error bars represent the standard deviation of all $0.05° \times 0.05°$ samples used to generate the $2° \times 2°$ grid boxes. MODIS MOD12C1 V6 land cover dataset is used to select these sample grid boxes.

We also investigated the bias of our prediction among different biome types in Fig. 3. For 9 out of 14 biome types, the differences between the $CSIF_{clear-inst}$ and the satellite-retrieved SIF are less than 10 %, and most of the biases were within 5 %. Wetlands and urban ecosystem show a 15 % bias compared to the satellite-retrieved SIF, which may be caused by the water or built-up contamination on the reflectance signal and the relatively small sample numbers. For savannas and grassland, the changes in fluorescence yield due to seasonal drought may be important, which cannot be considered in the NN based on reflectances only. Over croplands, CSIF exhibits a 12 % underestimation. The croplands usually have high nitrogen/chlorophyll concentration that may not be fully captured by the four broadband reflectances (Wu et al., 2008). Because we did not build biome-specific NNs for the training, we do not expect biome-specific (especially needleleaf vs. broadleaf) relationships between SIF and reflectance. Interestingly, we still reproduced SIF with very high accuracy regardless of the plant function traits (PFTs), i.e., leaf types and canopy characteristics (leaf clumping, etc.). This

suggests that the escape factor and long-term changes in mean fluorescence yield might be correctly accounted for by the NN across PFTs, through the information available in the reflectances only. However, it should be noted that this does not suggest that the NN and reflectances can fully replicate the fluorescence yield variations due to short-term variations caused by stresses.
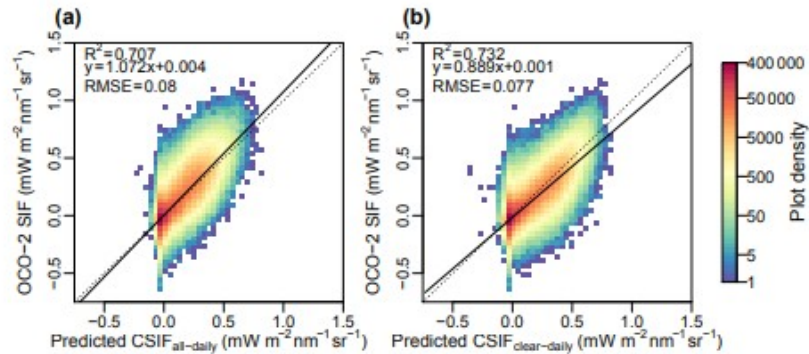
Figure 5. Comparison between the retrieved SIF and the **(a)** predicted all-sky daily CSIF and **(b)** clear-sky daily CSIF. The instantaneous SIF retrievals from OCO-2 were converted to daily average values for comparison.

We also compared the time series of predicted CSIF and OCO-2 SIF for 12 typical biome types (Fig. 4). The predicted CSIF accurately captures the seasonal and interannual variation for most biome types, while the standard deviation for each DOY is usually smaller than OCO-2 SIF. This may suggest that the uncertainty of SIF is smaller in CSIF dataset. For some ecosystems, e.g., DBF, MF and CRO, CSIF shows slight underestimation during the peak growing season.

When comparing the daily average SIF from satellite retrievals with the predicted all-sky daily CSIF (CSIF$_{all-daily}$) dataset (Fig. 5), the predicted SIF exhibits ~7 % underestimation, with an R^2 of 0.71 and a RMSE of 0.08 mW m^{-2} nm^{-1} sr^{-1}. The clear-sky daily CSIF (CSIF$_{clear-daily}$) shows ~11 % overestimation, with a slightly higher R^2 and lower RMSE. Considering the uncertainty in SIF retrievals and the inconsistency in time of the comparison (satellite SIF was based on instantaneous PAR at the time of satellite overpass and converted to daily values assuming the atmospheric condition did not change within a day; predicted CSIF was based on 4-day average PAR), the all-sky daily CSIF performs reasonably well.
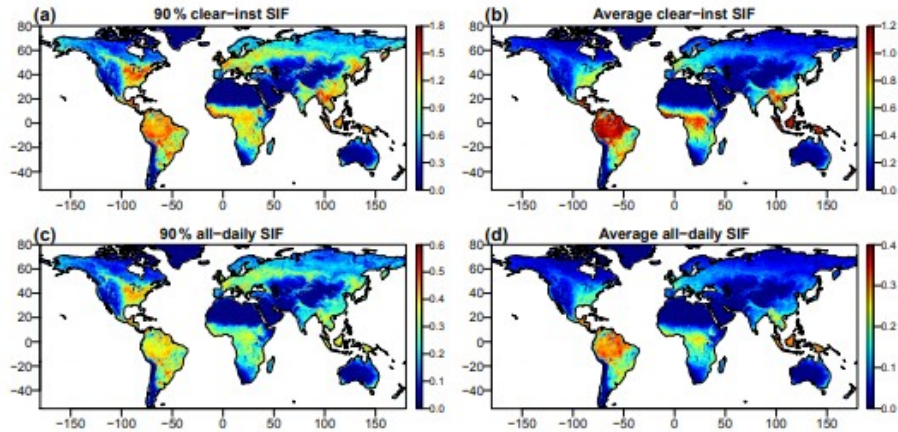
Figure 6. Spatial pattern of maximum (90th percentile) and average daily values for instantaneous clear-sky SIF and all-sky daily SIF. All values are in units of mW m^{-2} nm^{-1} sr^{-1}.

## 3.2 Spatial–temporal variation of the global 0.05° SIF datasets

Using the trained NN with the gap-filled reflectance datasets, we produced two global CSIF datasets at 4-day temporal and 0.05° spatial resolution. Figure 6 shows the spatial patterns of the 90th percentile for each pixel and the annual average for both clear-sky instantaneous CSIF (CSIF$_{clear-inst}$) and the all-sky daily average CSIF (CSIF$_{all-daily}$). For the 90th percentile, CSIF$_{clear-inst}$ exhibits hotspots in the tropical rainforest, south Asia and the North American corn belt, consistent with regions with high peak productivity (Guanter et al., 2014); CSIF$_{all-daily}$ shows similar spatial patterns but with relatively lower values in the tropical forest, due to the persistent cloud coverage. For the annual average SIF, tropical forests exceed temperate cropland and show very high values for instantaneous clear-sky SIF. In all conditions, African tropical forests exhibit lower values than Amazonian and Southeast Asian tropical forests.

We further investigated the seasonal and interannual variations of the all-sky daily SIF across the latitudes. The tropical regions show continuous high SIF values across seasons, and the northern mid- to high-latitude regions also exhibit recurrent high values during the Northern Hemisphere summers (Fig. 7a). Near 40° S, a hot spot is present in austral summer, with high interannual variability. Low SIF values can be found in dry years (2006–2007, 2009–2010), while high values were observed in wet or normal years (2010–2011, 2012–2015). The global average SIF also displays a strong seasonality coinciding with the Northern Hemisphere growing season (Fig. 7b). For the annual total SIF values, a statistically significant increasing trend (Mann–Kendall test, $p<0.0001$) is found with around 0.39 % increase per year. The year 2015 exhibited a low anomaly after detrending, which may be caused by the El Niño events (Fig. 7c).
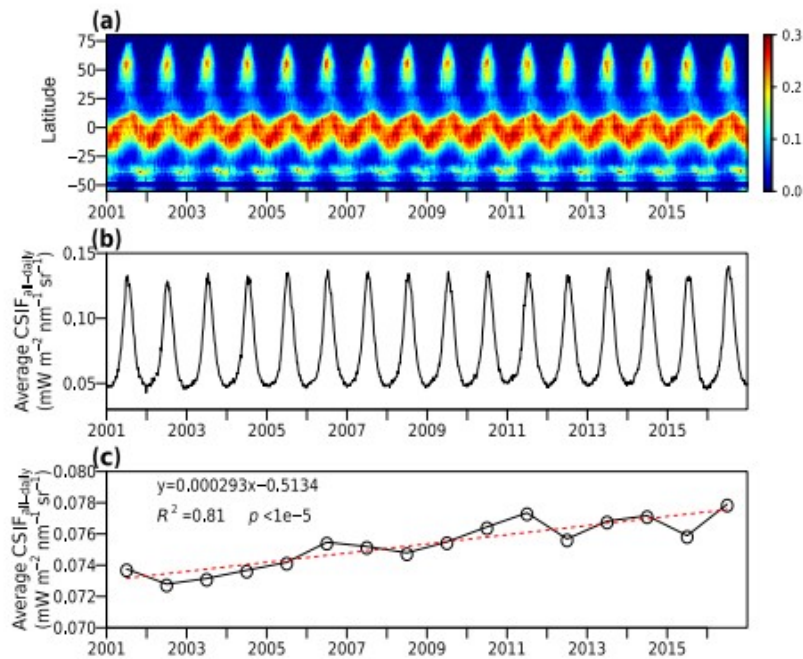
Figure 7. Seasonal and interannual variation of all-sky condition daily CSIF (CSIF$_{all\text{-}daily}$). **(a)** The latitudinal averages of CSIF$_{all\text{-}daily}$ for each 4-day period (in mW m^{-2} nm^{-1} sr^{-1}). **(b)** Global average of CSIF$_{all\text{-}daily}$ for each 4-day period. **(c)** The annual average CSIF$_{all\text{-}daily}$ between 2001 and 2016 (black line) with linear fit (red dashed line).
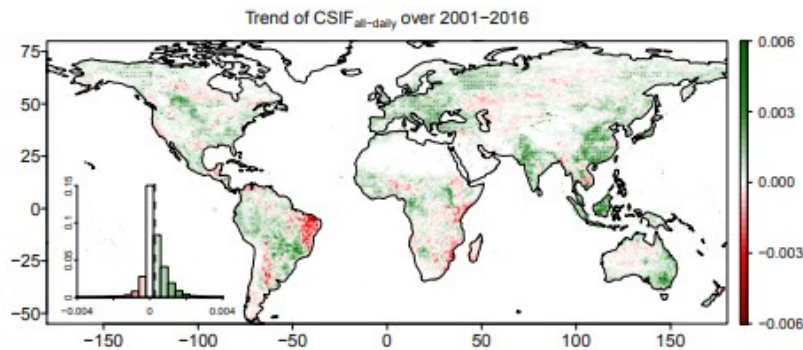
Figure 8. Trend of annual average CSIF$_{all\text{-}daily}$ during 2003–2016. The trend is calculated by the Sen's slope estimator. Dots represent the trend is significant ($p<0.05$) through a Mann–Kendall test. The inset in the bottom left shows the histogram of the CSIF$_{all\text{-}daily}$ trend. Dashed vertical line represents the average trend. Barren areas with an annual average CSIF$_{all\text{-}daily}$ smaller than 0.006 mW m^{-2} nm^{-1} sr^{-1} are screened from analysis. Trends are in units of mW m^{-2} nm^{-1} sr^{-1} yr^{-1}.

The spatial pattern of the trend in CSIF$_{all\text{-}daily}$ is displayed in Fig. 8. An increasing trend dominates Europe, southeast Asia and south Amazon. A decreasing trend is mostly found in east Brazil, east Africa and some areas of inland Eurasia. The histogram also shows a positive shift with a magnitude (0.00027 mW nm^{-1} sr^{-1} yr^{-1}) similar to the average global trend in Fig. 7c. The spatial pattern of CSIF$_{all\text{-}daily}$ is very similar to the trend pattern of MODIS enhanced vegetation index (EVI) (C6) (Zhang et al., 2017b), but the south Brazilian Amazon forest shows a more positive trend than that of EVI.
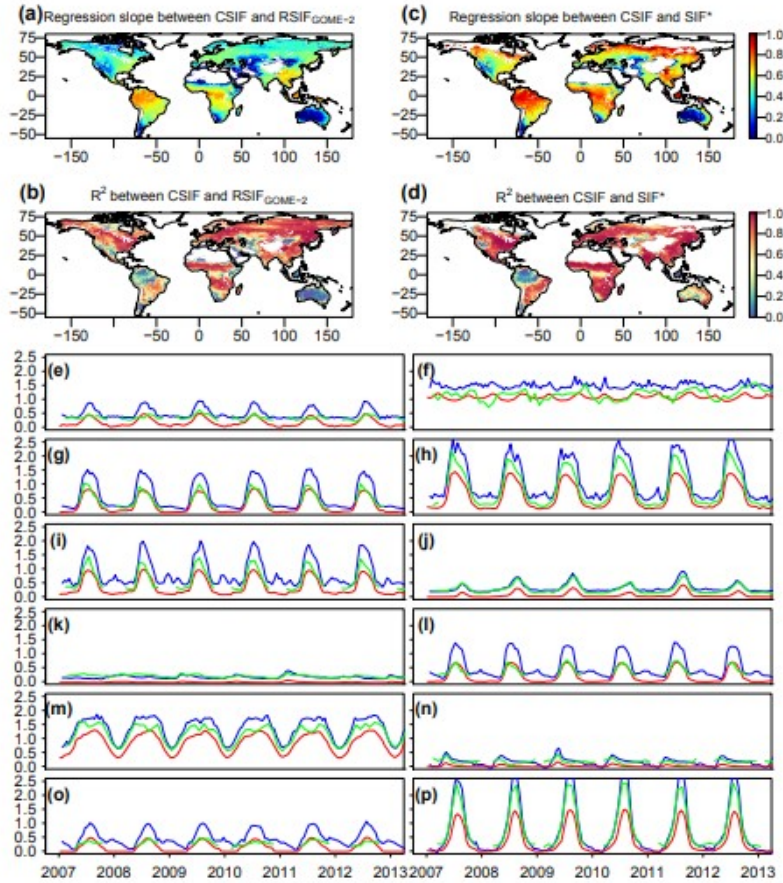
Figure 9. Comparison between CSIF, $RSIF_{GOME-2}$ and SIF^* dataset. Regression slopes and coefficient of determination (R^2) between the contiguous clear-sky condition instantaneous SIF from OCO-2 ($CSIF_{inst-clear}$) and the reconstructed SIF from GOME-2 ($RSIF_{GOME-2}$ **a, b**) or SIF^* (**c, d**) dataset. The regressions are forced to pass the origin. The $CSIF_{clear-inst}$ is aggregated to semi-monthly and $0.5° \times 0.5°$ spatial resolution to be consistent with $RSIF_{GOME-2}$. Comparison uses the data between 2007 and 2016 (RSIF) or 2007 to 2013 (SIF^*). White regions are barren regions. (**e–p**) Time series comparison among CSIF (red), $RSIF_{GOME-2}$ (blue) and SIF^* (green) for pixels in 12 major land cover types shown in Fig. 4.

3.3 Comparison between SIF from GOME-2 and CSIF

We then compared the CSIF datasets with the reconstructed SIF (RSIF) and SIF^* based on coarser-scale and all-sky GOME-2. Although these datasets were trained based on different satellites, the relationship between CSIF and RSIF or CSIF and SIF^* is consistent across most regions across the globe (Fig. 9). The R^2 values are generally high (> 0.8) for most regions except over tropical rainforests, barren regions in western US, northwestern China, and northern Canada and Russia. The low R^2 values are mostly due to the relatively low variability in the temporal domain in the tropics but are also indicative of regions strongly polluted by cloud cover in which CSIF might have a competitive advantage, as the training OCO-2 data better observe the surface due to smaller footprint and with higher signal-to-noise ratio. The regression slopes are higher for regions with persistent cloud cover (e.g., tropical forest). In the time series comparison (Fig. 9e–p), all three SIF datasets show similar seasonal patterns, while GOME-2-based RSIF and SIF^* generally show higher values than CSIF. In addition, RSIF exhibits larger

fluctuation during the non-growing season for some sites, which may be caused by snow contamination.
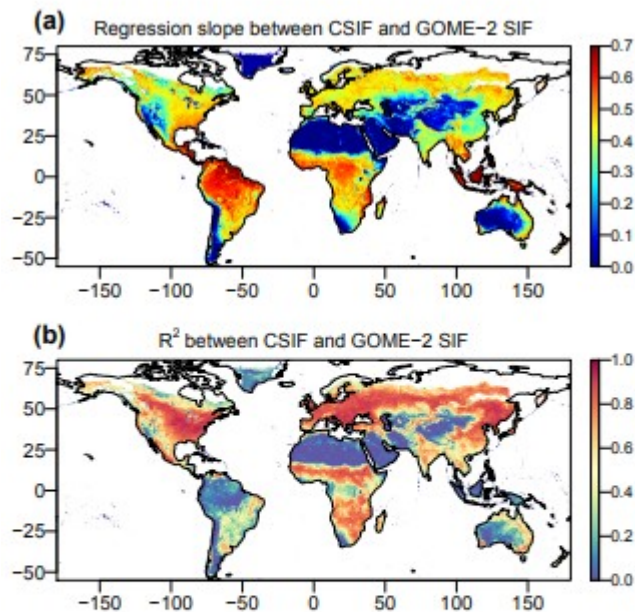
Figure 10. Regression slopes and coefficient of determination (R^2) between the contiguous all-sky condition daily SIF from OCO-2 (CSIF_all-daily) and the satellite-retrieved daily SIF from GOME-2 (SIF_GOME-2). The regressions are forced to pass through the origin. The CSIF_all-daily is aggregated to monthly and $0.5° \times 0.5°$ spatial resolution to be consistent with SIF_GOME-2. Comparison uses the data between 2007 and 2016.

We further compared the CSIF_all-daily with GOME-2 daily average SIF (Fig. 10). In general, the correlation is much lower as compared with RSIF for most regions. For regions with high variability in temporal domain, the CSIF_all-daily still shows high R^2 values with respect to GOME-2 SIF. The regression slopes exhibit smaller variation except for the Amazonian tropical rainforests, southeast Asia and barren regions in the Sahara, western US, northwestern China, central Australia and the Andes mountains in South America. In general, considering the various uncertainties and different satellite overpass times, sensors used and retrieval algorithms, CSIF_all-daily well captured the GOME-2 SIF variations both in space and time. In addition, since GOME-2 SIF in most Argentina is affected by the South Atlantic Anomaly (SAA), the coefficient of determination values are also lower as compared with Fig. 9.
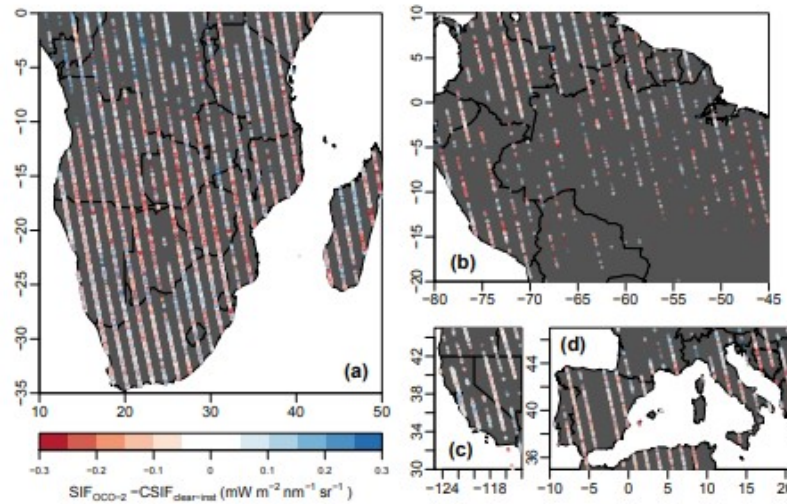
Figure 11. Difference between the OCO-2 SIF and $CSIF_{clear-inst}$ for four specific drought events during 2014–2017. **(a)** Southern Africa drought between October 2015 and February 2016. **(b)** Northeast Amazon drought between January and March 2016. **(c)** California drought between January and March 2015. **(d)** Southern Europe drought between July and August 2017.

3.4 Using CSIF for drought monitoring

Since the CSIF dataset only uses broadband reflectances, it should not contain the SIF_{yield} information. Compared to the SIF retrieved from OCO-2, the difference can be mostly attributed to the SIF_{yield}. Therefore, the difference or ratio between SIF_{OCO-2} and CSIF can reflect the environmental stress on SIF_{yield}. Figure 11 shows the difference between instantaneous clear-day OCO-2 SIF and $CSIF_{clear-inst}$. Except for Fig. 11c, the difference mostly captures the physiological limitation of drought on energy partitioning after being absorbed by chlorophyll. The spatial extent of drought is also well-captured by the difference, where the most severe drought-impacted places also exhibited the largest decline (e.g., Namibia, Botswana, Zimbabwe in Fig. 11a, northeast Amazon in Fig.11b and southern Spain, southernmost France, central Italy, Croatia and Bosnia and Herzegovina). The drought impact on California is less pronounced, possibly because of the irrigation systems and sparse sampling points.
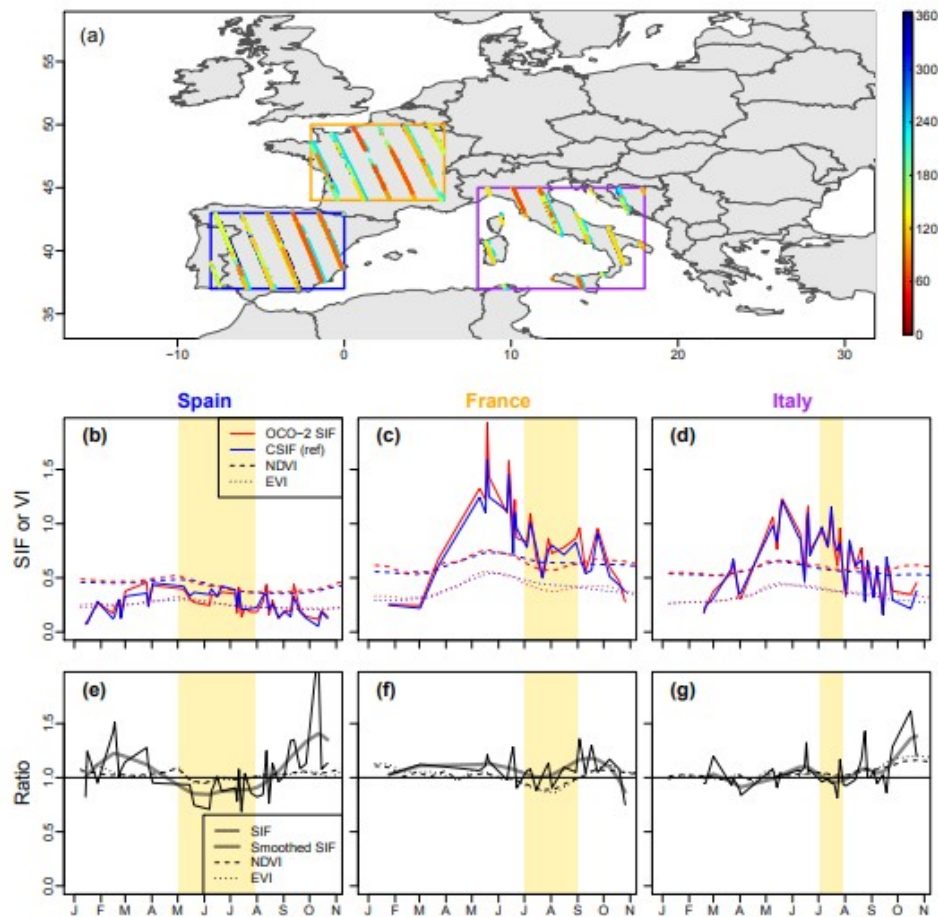
Figure 12. (a) Spatial distribution of OCO-2 SIF observations during 1 January to 1 November in 2015. Different colors represent the observation DOY. **(b–d)** Average OCO-2 SIF, CSIF NDVI and EVI for the three countries as indicated by three boxes in panel **(a)**. For two vegetation indices, the red color represents the observations in 2015 and blue color represents multi-year average (2000–2014). **(e–g)** The ratio between OCO-2 SIF and CSIF (SIF) or vegetation indices in 2015 and multi-year average. The thick grey line presents the splines' smoothed SIF ratio.

We further focused on the 2015 European drought to compare the drought response of CSIF and two vegetation indices (NDVI and EVI). Because the OCO-2 samples were not collected at the same swath for each DOY, a large fluctuation can be found in OCO-2 SIF and on the CSIF (which are using the same pixels for a fair comparison) (Fig. 12a–d). However, when calculating the ratio between CSIF and OCO-2 SIF, its variation can be mostly attributed to the variation in SIF_{yield}, which can quantify the drought stress on plant physiology. In all three regions, the ratio between OCO-2 SIF and CSIF experienced a decrease during the drought period, but the signal is only obvious after applying a smoothing filter. The two vegetation indices, NDVI and EVI, on the other hand, show a reduced response in Spain and Italy, perhaps due to the plants' adaption or very short drought duration.
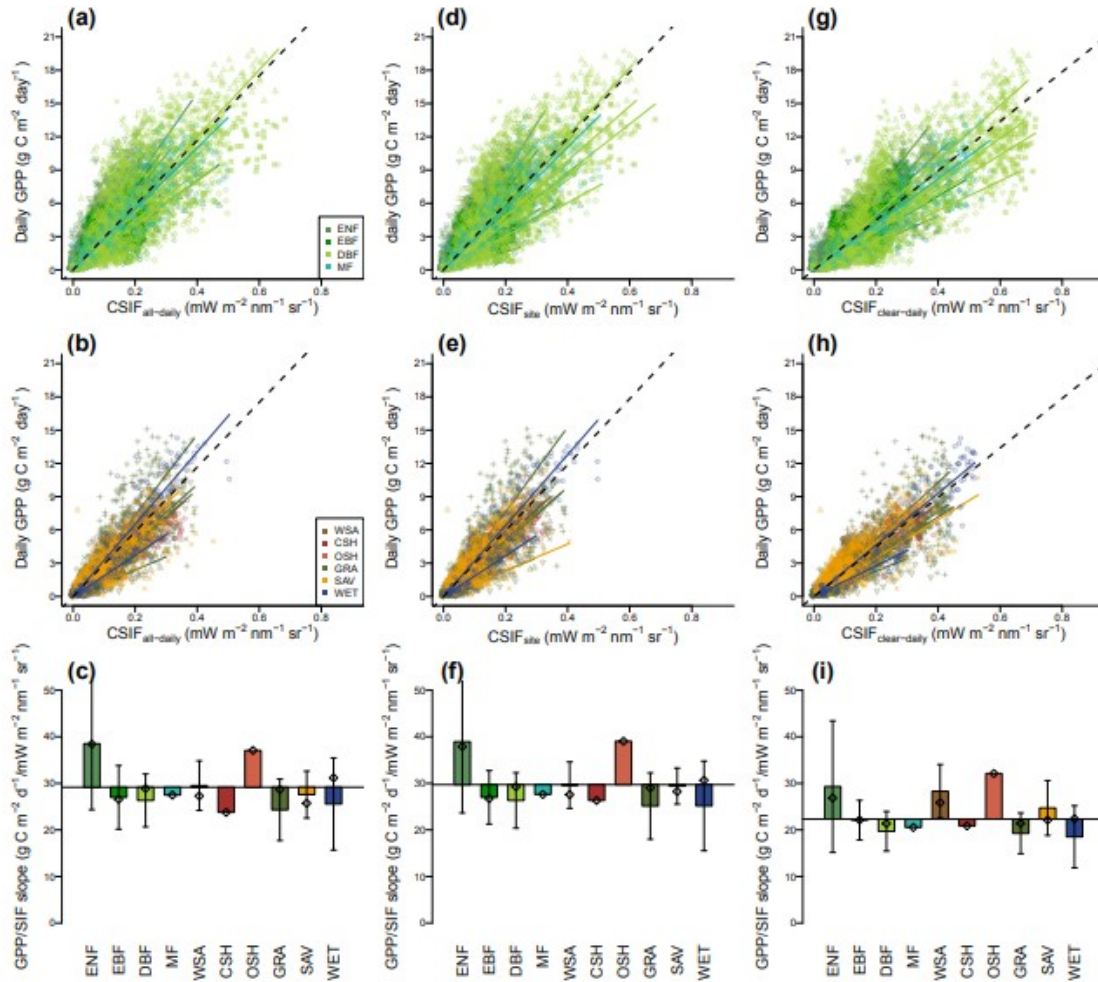
Figure 13. Comparison between GPP estimates from 40 EC flux towers and CSIF$_{all-daily}$ **(a–c)** that uses BESS PAR, CSIF$_{site}$ **(d–f)** that uses site-measured radiation and CSIF$_{clear-daily}$ **(g–i)** that assumes clear-sky condition. The 40 sites were grouped into forests **(a, d, g)** and non-forests **(b, e, h)**. Color–symbol combinations represent different sites. Summary of the regression slopes between GPP and CSIF for different land cover types **(c, f, i)**. The baseline (dashed black lines) was calculated using all samples (29.71 for CSIF$_{all-daily}$, 29.18 for CSIF$_{site}$ and 22.33 for CSIF$_{clear-daily}$). Error bars represent the standard deviation of slopes across sites within this biome type. Rhombuses represent regression for each biome type when data from all sites were combined.

3.5 GPP–CSIF relationship across biome types

With this contiguous SIF$_{all-daily}$ dataset, we finally evaluated the GPP–CSIF relationship using GPP estimates from 40 flux tower sites from FLUXNET tier 1 dataset. The regression slope between GPP and CSIF ($a_{GPP/CSIF}$) spreads across sites with a regression slope ranging from 11.91 to 68.59 (g C m^{-2} day^{-1}/mW m^{-2} nm^{-1} sr^{-1}) for CSIF$_{all-daily}$, 11.61 to 72.10 (g C m^{-2} day^{-1}/mW m^{-2} nm^{-1} sr^{-1}) for CSIF$_{site}$ and 11.37 to 62.75 (g C m^{-2} day^{-1}/mW m^{-2} nm^{-1} sr^{-1}) for CSIF$_{clear-daily}$. The R^2 value for each individual site ranges from 0.01 to 0.93 with a median value of 0.64, 0.62 and 0.69 for all-daily, site and clear-daily CSIF, respectively. The RMSE is 1.67 g C m^{-2} day^{-1} on average.

Although the CSIF–GPP relationship varies across 40 sites, when lumping all observations within each biome type, the variation is smaller (c.v. = 0.16,

rhombus in Fig. 13c, f, i). Specifically, ENF exhibited a significant larger $a_{GPP/CSIF}$ (two-tailed Student's t test, $p=0.036$), which is caused by a stronger canopy reabsorption/scattering of SIF. OSH only has one site and also showed very high value. If both biomes are eliminated, the $a_{GPP/CSIF}$ for the other biomes exhibited smaller variation (c.v. = 0.08).

The CSIF–GPP relationship not only varies across biomes but also varies within each biome type, especially for evergreen needleleaf forest (ENF, nine sites), grassland (GRA, eight sites) and wetland (WET, two sites) (Fig. 13c, f). For $CSIF_{all\text{-}daily}$, the average within-biome variation of $a_{GPP/CSIF}$ (c.v. = 0.26±0.08) is comparable to cross-site variations (c.v. = 0.34) but larger than the cross-biome variations (c.v. = 0.16, using the biome-specific CSIF–GPP factor). A similar pattern can be found using $CSIF_{site}$ or $CSIF_{clear\text{-}daily}$.

4 Discussion

4.1 Information in contiguous SIF produced by machine learning

Vegetation photosynthetic activity has variations in several respects controlled by vegetation type, phenology, coverage and interactions with the environment. These variations can be expressed in the spatial, seasonal, diurnal and/or interannual domains (Zhang et al., 2018a). Machine learning algorithms try to minimize the differences between the predicted SIF and the satellite-observed SIF. For OCO-2 SIF and the MODIS reflectance used for NN training, the variance in the spatial and seasonal domains is largest. Therefore, the NN generally predicts SIF well in these two domains. The interannual variations (i.e., the variations caused by year-to-year anomalies, e.g., due to drought) typically have much smaller variance and are more difficult to capture. This is why some machine learning products fail to reproduce interannual variability accurately (Jung et al., 2011). Using additional variables that are sensitive to this interannual anomaly in the model training can improve the model performance (Alemohammad et al., 2017; Gentine and Alemohammad, 2018b; Tramontana et al., 2016).

In this study, since the variations in SIF_{yield} are relatively small (Lee et al., 2015) and cannot be detected by broadband surface reflectances, the SIF_{yield} information may not be reproduced by our CSIF data. Because the environmental limitation on SIF_{yield} may be complicated (may not be a linear combination of temperature, vapor pressure deficit (VPD) or surface reflectance in the shortwave infrared) and biome specific (van der Tol et al., 2014), inclusion of other environmental variables and reflectances in shortwave bands during NN training did not greatly increase the SIF prediction accuracy. It should also be noted that SIF_{yield} is relatively stable when no strong environmental limitation is present (Zhang et al., 2018c). Therefore, the CSIF product should be considered as a good proxy of OCO-2 SIF.

The satellite-retrieved SIF has a relatively large uncertainty for each individual sounding, typically ranging between 0.3 and 0.5 $mW\,m^{-2}\,nm^{-1}\,sr^{-1}$

(Frankenberg et al., 2014). Previous site-level studies usually use SIF averaged over a large buffered area (Li et al., 2018a; Verma et al., 2017) to reduce the uncertainty. Assuming the uncertainty is unbiased and has a Gaussian distribution, machine learning algorithms are designed to reproduce SIF with lower uncertainty. Compared with previous studies that use light-use efficiency models to downscale SIF to higher resolution (Duveiller and Cescatti, 2016), this study does not rely on multiple modeled input (evapotranspiration, for example) that may introduce additional uncertainties.

We also found a significant increasing trend ($0.39\,\%\,\text{yr}^{-1}$) in the global annual $\text{CSIF}_{\text{all-daily}}$ (Fig. 7). This trend is close to the GPP trend derived from the satellite-data-driven vegetation photosynthesis model (VPM) ($0.32\,\%\,\text{yr}^{-1}$) (Zhang et al., 2017a) but much greater than GPP derived from other remote sensing data-driven models – FLUXCOM ($0.01\,\%\,\text{yr}^{-1}$; Tramontana et al., 2016), BESS GPP ($0.22\,\%\,\text{yr}^{-1}$; Jiang and Ryu, 2016), MODIS C6 ($0.26\,\%\,\text{yr}^{-1}$; Zhao et al., 2005) and WECANN ($-0.8\,\%\,\text{yr}^{-1}$, affected by the decreasing GOME-2 SIF trend; Zhang et al., 2018b; Alemohammad et al., 2017). Considering there is no significant trend ($-0.02\,\%\,\text{yr}^{-1}$, $p>0.1$) in BESS PAR (Ryu et al., 2018), this increase is likely caused by the greening of the Earth (Zhang et al., 2017b; Zhu et al., 2016) as captured in the MODIS reflectance data. This increasing trend is also within the range of most Earth system models' predictions (Anav et al., 2015). We also observed a more pronounced increasing trend in the southern Amazon than when using MODIS EVI (Zhang et al., 2017b). This may suggest that CSIF is less likely to suffer from high biomass saturation than optical vegetation indices and can more effectively detect changes in tropical rainforests or over high leaf area regions such as croplands.

4.2 The use of satellite SIF for drought monitoring

Drought can be categorized into different stages. At an early stage, when plants sense water deficit in the soil and higher vapor pressure deficit in the atmosphere, they reduce water loss through stomatal closure. This, in turn, also reduces the CO_2 exchange from stomatal closure and inhibits photosynthesis. The quantum yield for heat dissipation will increase accompanied by a decrease in quantum yield for photochemical quenching and fluorescence (Genty et al., 1989; Porcar-Castell et al., 2014). This should allow satellites to potentially capture this decrease in the SIF signal (especially during the mid-afternoon when stress is more pronounced) as an indicator of vegetation stress. In the second stage, with prolonged dry conditions, plants will recycle the nitrogen in the leaves as represented by a decrease of the greenness (chlorophyll content) of leaves. In the third stage, if the drought continues, leaf senescence and vegetation mortality may follow. SIF can potentially detect changes during all those drought stages, whereas broadband reflectance-based indices (NDVI, EVI) should only see the second and third stages.

Previous drought monitoring studies have mostly used vegetation indices (VIs) as a indictor of drought stress (Ji and Peters, 2003; Zhang et al., 2013). However, vegetation indices can only respond to drought changes in the plants' optical properties (mostly during the second and third stages). For most plants, there might be a tipping point where plants will not recover from drought-induced xylem cavitation (Urli et al., 2013). Since most VIs (e.g., NDVI, EVI) are most sensitive to the canopy changes, drought monitoring based on VIs may not be useful for drought mitigation and agricultural irrigation management. SIF retrievals from satellite, compared to optical reflectance signals, carry the information not only about the PAR absorption by chlorophyll but also about the drought stress on plant physiology. Although previous studies used satellite-based SIF datasets for post-drought impact assessment (Lee et al., 2013; Yoshida et al., 2015; Sun et al., 2015; Wang et al., 2016), these studies did not separate the contribution of decreased $APAR_{chl}$ or deceased SIF_{yield}. A more recent study compared the SIF and VIs in India during heat stress (Song et al., 2018) and found that SIF is more sensitive to heat stress than VIs. Similarly, since NDVI and EVI cannot well capture the change in chlorophyll concentration, heat stress on $APAR_{chl}$ and SIF_{yield} cannot be fully separated. This study developed a new method to compare the difference between SIF signals and the reflectances, which can be applied for early drought warning at global scale. Although daily OCO-2 data have large gaps between swaths, combining several days of observation can provide enough spatial coverage considering the spatial extent for most drought events. The spatial coverage issues could be further improved using geostatistical-based methods (Tadić et al., 2017), but this may need further investigation. Compared to other meteorological drought indices, this drought monitoring technique uses only near real-time data and avoids the interannual anomalies caused by other factors (land cover change, crop rotation, etc.). The MCD43C4 dataset uses 16 days of inputs for the model inversion, and although this may lead to temporal inconsistencies for the comparison between CSIF and OCO-2 SIF, it may have limited effect due to the higher data quality during drought because of the reduced cloud coverage.

4.3 Cross-biome and within-biome GPP–CSIF relationship

In contrast to Sun et al. (2017), we found a large variation of GPP–CSIF relationship across sites. Compared to previous studies, our study gave higher $a_{GPP/CSIF}$ estimates, probably due to a much higher $a_{GPP/CSIF}$ value for evergreen needleleaf forest (10 out of 40 sites are ENF) (Table S1 in the Supplement) and slight underestimation of $CSIF_{all-daily}$ dataset. This higher $a_{GPP/CSIF}$ value for ENF was also suggested by the comparison between OCO-2 SIF and FLUXCOM GPP dataset (Sun et al., 2018) and other comparisons using GOSAT SIF (Guanter et al., 2012). Consistent with Li et al. (2018b), we also found small cross-biome variation of the GPP–SIF relationship. However, a large within-biome variation of $a_{GPP/CSIF}$ is also found, which contributes to a large proportion of the observed cross-site variations rather than the cross-

biome variation. Compared to studies that use OCO-SIF within a large buffering area (e.g., 40 km diameter circle in Verma et al., 2017), we made the comparison over a much smaller area and much higher temporal frequency.

There are several explanations for the observed site-specific GPP–SIF relationship. (1) Leaf morphology may directly affect the reabsorption and scattering of SIF that leaves the foliage (Atherton et al., 2017); however, this factor is not considered in current SIF modeling (van der Tol et al., 2009; Verrelst et al., 2015) and will directly affect the model simulation of the GPP–SIF relationship at the ecosystem scale (Verrelst et al., 2016; Zhang et al., 2016c). (2) Vegetation canopy characteristics also affect the reabsorption and scattering of SIF before leaving the canopy (Romero et al., 2018; Yang and van der Tol, 2018). (3) Atmospheric condition may attenuate and bias satellite SIF retrievals to some extent, but this effect is assumed to be small unless thick clouds are present (Frankenberg and Berry, 2017). (4) SIF and GPP likely have different sensitivities to environmental stresses (Flexas et al., 2002); therefore, ecosystems with frequent environmental stresses (e.g., drought) during the growing season tend to have relatively lower GPP-to-SIF ratio. (5) Since light saturations have less effect on SIF than GPP (Damm et al., 2015; Zhang et al., 2016c), the growing-season averaged light intensity (affected by latitude, average cloud coverage), vegetation canopy structure and leaf characteristics that relate to the light saturation will also affect GPP–SIF relationship. For example, the evergreen needleleaf forests have much higher specific leaf area and usually lower Sun zenith angle, making them less prone to light saturation. These factors may vary not only across biomes but also across sites. Therefore, within one biome type, the GPP–SIF relationship can also be different.

It is also noteworthy that clear-sky daily SIF exhibited stronger correlation with GPP (Fig. 13); a possible explanation would be that the light-use efficiency increases with diffused radiation, which partly compensates for the decrease in incoming PAR when clouds are present (Gu et al., 2002; Turner et al., 2006). Because the satellite SIF retrieval algorithm discarded observations that were affected by thick clouds (Sun et al., 2018), the SIF retrievals from OCO-2 are more positively biased than the actual SIF emission of the plants. However, during periods when thick clouds are present, the LUE also increases and so does the GPP / SIF ratio. The positive SIF retrieval biases compensated the increase in the GPP / SIF ratio and therefore contributed to a stronger correlation between satellite-retrieved SIF (rather than the actual SIF emission) and GPP.

4.4 Uncertainties and caveats

Although our CSIF$_{clear-inst}$ showed good performance as supported by the comparison with the clear-sky instantaneous SIF retrievals from OCO-2, the CSIF$_{all-daily}$ exhibits a slight underestimation. A possible explanation is that most SIF retrievals during overcast conditions did not pass the quality

checks, such that OCO-2 SIF are more likely obtained during clear-sky conditions. This is supported by the fact that if we compare OCO-2 SIF with clear-daily SIF, the R^2 is even higher (Fig. 6).

The canopy structure and sun-sensor geometry were not explicitly considered in our modeling and only implicitly embedded in the machine learning retrieval. Several recent studies suggest that canopy structure will affect the PAR absorption and re-absorption of SIF before leaving the canopy (f_{esc} in Eq. 3) (Knyazikhin et al., 2013; Liu et al., 2016; Yang and van der Tol, 2018) and further affect the GPP–SIF relationship (He et al., 2017; Migliavacca et al., 2017; Zhang et al., 2016c). However, most of these studies made assumptions requiring either a dense canopy or non-reflecting soil and thus cannot be easily applied at the global scale. In addition, OCO-2 SIF data used in this study are from nadir observations, while both the MODIS and GOME-2 sensors acquire images both at nadir and near nadir. Such discrepancy in observation angles may induce bidirectional effects. Since CSIF is trained based on the satellite-observed SIF instead of the canopy SIF emission, and as previously discussed, it did not consider the atmospheric attenuation of SIF signal in the presence of clouds. The CSIF values are expected to be closer to the canopy SIF emission than the satellite-observed SIF at the top of atmosphere.

The BESS PAR 4-day dataset has high overall accuracy (RRMSE of 15.2 %) and very little bias (1.4 %). For different climate zones, the uncertainties are typically under 20 %. These uncertainties do not affect the CSIF$_{clear-inst}$ data but will propagate to CSIF$_{all-daily}$.

5 Conclusions

In this study, using the surface reflectance from the MODIS instrument and a NN algorithm, we developed two spatially contiguous and high-temporal-resolution SIF datasets (CSIF). These two SIF products not only show high accuracy when validated against the satellite-retrieved OCO-2 SIF but also exhibit reasonably high consistency with both reconstructed and satellite-retrieved GOME-2 SIF. CSIF$_{all-daily}$ exhibits an increasing trend globally during 2001–2016, which is attributed to the Earth greening and not to changes in PAR. Since the CSIF dataset includes most information on PAR absorption of chlorophyll, the difference between OCO-2 SIF and CSIF mostly contains the information on physiological stress on fluorescence yield. This indicator is found to be more effective for early drought warning than vegetation indices. By comparing CSIF$_{all-daily}$ with GPP estimates across 40 EC flux tower sites, the GPP–SIF relationship is found to vary across sites, and a large proportion of this comes from within-biome variation. However, this finding still requires further examination using SIF from both new satellites instruments (e.g., TROPOMI) and ground-based measurements. The high-resolution CSIF dataset can be further used for regional to global carbon and water flux analysis.

Code availability.

The code used to generate the CSIF dataset is available at https://github.com/zhangyaonju/ (Zhang et al., 2018).

Data availability.

The CSIF dataset (CSIF$_{clear-inst}$, CSIF$_{clear-daily}$ and CSIF$_{all-daily}$) with a 0.5° spatial resolution and 4-day temporal resolution can be accessed through Figshare: https://doi.org/10.6084/m9.figshare.6387494 (Zhang et al., 2018). The 0.05° 4-day dataset can be obtained upon request, given the large size. The MCD43C4 dataset can be accessed through NASA EARTHDATA (https://earthdata.nasa.gov, Schaaf et al., 2002). The BESS PAR product can be accessed through the Environmental Ecology Lab at Seoul National University (http://environment.snu.ac.kr/bess_rad/, Ryu et al., 2018).

Appendix A: Calculation of clear-sky radiation

We calculated the clear-sky radiation following previous studies (Duffie and Beckman, 2013; Ryu et al., 2018). The total surface shortwave radiation R_T is the summation of direct surface beam radiation (R_{sb}) and diffused radiation (R_{sd}):

$$R_T = R_{sb} + R_{sd}. \tag{A1}$$

R_{sb} and R_{sd} are calculated as the product of the top-of-atmosphere shortwave radiation (R_{TOA}) and the atmospheric transmittance for beam radiation (τ_b) and that for diffused radiation (τ_d):

$$R_{sb} = R_{TOA} \times \tau_b \tag{A2}$$

$$R_{sd} = R_{TOA} \times \tau_d, \tag{A3}$$

where R_{TOA} is calculate as a function of solar constant (S_0=1360.8 W m^{-2}), the proportion of solar irradiance within shortwave range (α=0.98), the day of year (n) and the cosine of the solar zenith angle ($\cos\theta_s$):

$$R_{TOA} = S_0 \times \alpha \times \left[1 + 0.033\cos\left(\frac{2\pi n}{365}\right)\right] \times \cos\theta_s, \tag{A4}$$

and τ_b is calculated as

$$\tau_b = a_0 + a_1\exp\left(\frac{-k}{\cos\theta_s}\right), \tag{A5}$$

where a_0, a_1 and k are coefficients that consider the atmospheric attenuation based on the atmosphere path length and abundance of the gases or particles that need to be adjusted for elevation:

$$a_0 = 0.4237 - 0.00821(6 - A)^2 \tag{A6a}$$

$$a_1 = 0.5055 + 0.00595(6.5 - A)^2 \tag{A6b}$$

$$k = 0.2711 + 0.01858(2.5 - A)^2, \tag{A6c}$$

where A is the elevation in kilometers. The ETOPO1 Global Relief Model was used to provide the elevation information. This dataset was downloaded from National Oceanic and Atmospheric Administration (https://data.nodc.noaa.gov/, last access: 27 September 2018) and aggregated to 0.05°. In this study, we did not consider the variation of these parameters for different climate and latitudinal zones since those effects are less important. The transmittance for diffused radiation (τ_d) is calculated as a function of τ_b:

$$\tau_d = 0.271 - 0.294\tau_b. \qquad (A7)$$
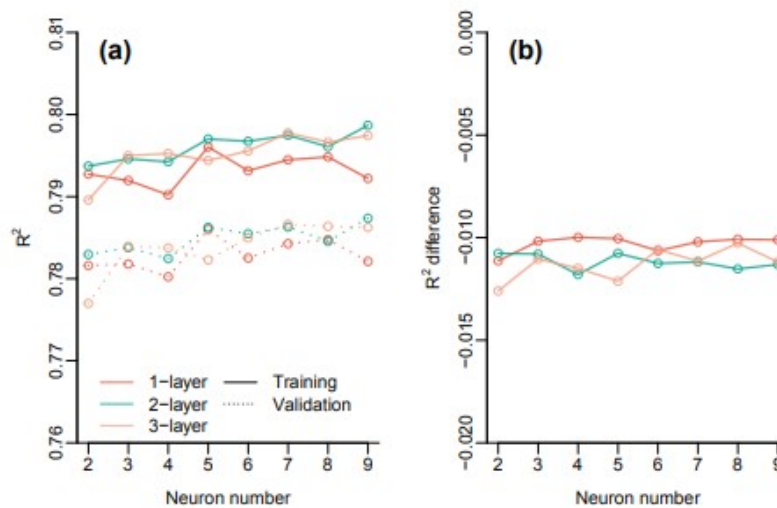
Figure A1. (a) Comparison of model performance (R^2) during training and validation with a variety of NN layers (one to three) and neuron numbers for each layer (two to nine). **(b)** Difference of model performance between the training and validation for different layer and neuron combinations.

Supplement.

The supplement related to this article is available online at: https://doi.org/10.5194/bg-15-5779-2018-supplement.

References

Adams, W. W. and Demmig-Adams, B.: Chlorophyll Fluorescence as a Tool to Monitor Plant Response to the Environment, in: Chlorophyll a Fluorescence, Springer, Dordrecht, 583–604, 2004.

Alemohammad, S. H., Fang, B., Konings, A. G., Aires, F., Green, J. K., Kolassa, J., Miralles, D., Prigent, C., and Gentine, P.: Water, Energy, and Carbon with Artificial Neural Networks (WECANN): a statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence, Biogeosciences, 14, 4101–4124, https://doi.org/10.5194/bg-14-4101-2017, 2017.

Alemohammad, S. H., Kolassa, J., Prigent, C., Aires, F., and Gentine, P.: Global Downscaling of Remotely-Sensed Soil Moisture using Neural Networks, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-680, in review, 2018.

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N. C., Peylin, P., Piao, S., Sitch, S., Viovy, N., Wiltshire, A., and Zhao, M.: Spatiotemporal patterns of terrestrial gross primary production: A review, Rev. Geophys., 53, 785–818, https://doi.org/10.1002/2015RG000483, 2015.

Atherton, J., Olascoaga, B., Alonso, L., and Porcar-Castell, A.: Spatial Variation of Leaf Optical Properties in a Boreal Forest Is Influenced by Species and Light Environment, Front. Plant Sci., 8, p. 309, https://doi.org/10.3389/fpls.2017.00309, 2017.

Badgley, G., Field, C. B., and Berry, J. A.: Canopy near-infrared reflectance and terrestrial photosynthesis, Science Advances, 3, e1602244, https://doi.org/10.1126/sciadv.1602244, 2017.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw U, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities, Bull. Am. Meteorol. Soc., 82, 2415–2434, https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2, 2001.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rodenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A.,

Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F. I., and Papale, D.: Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate, Science, 329, 834–838, https://doi.org/10.1126/science.1184984, 2010.

Chen, M., Rafique, R., Asrar, G. R., Bond-Lamberty, B., Ciais, P., Zhao, F., Reyer, C. P. O., Ostberg, S., Chang, J., Ito, A., Yang, J., Zeng, N., Kalnay, E., Tristram West, Leng, G., Francois, L., Munhoven, G., Henrot, A., Tian, H., Pan, S., Kazuya Nishina, Viovy, N., Morfopoulos, C., Betts, R., Schaphoff, S., Steinkamp, J., and Hickler, T: Regional contribution to variability and trends of global gross primary productivity, Environ. Res. Lett., 12, 105005, https://doi.org/10.1088/1748-9326/aa8978, 2017.

Damm, A., Guanter, L., Paul-Limoges, E., van der Tol, C., Hueni, A., Buchmann, N., Eugster, W., Ammann, C., and Schaepman, M. E.: Far-red sun-induced chlorophyll fluorescence shows ecosystem-specific relationships to gross primary production: An assessment based on observational and modeling approaches, Remote Sens. Environ., 166, 91–105, https://doi.org/10.1016/j.rse.2015.06.004, 2015.

Du, S., Liu, L., Liu, X., and Hu, J.: Response of Canopy Solar-Induced Chlorophyll Fluorescence to the Absorbed Photosynthetically Active Radiation Absorbed by Chlorophyll, Remote Sens., 9, p. 911, https://doi.org/10.3390/rs9090911, 2017.

Duffie, J. A. and Beckman, W. A.: Solar Engineering of Thermal Processes, John Wiley & Sons, 2013.

Duveiller, G. and Cescatti, A.: Spatially downscaling sun-induced chlorophyll fluorescence leads to an improved temporal correlation with gross primary productivity, Remote Sens. Environ., 182, 72–89, https://doi.org/10.1016/j.rse.2016.04.027, 2016.

Flexas, J., Escalona, J. M., Evain, S., Gulías, J., Moya, I., Osmond, C. B., and Medrano, H.: Steady-state chlorophyll fluorescence (F_s) measurements as a tool to follow variations of net CO_2 assimilation and stomatal conductance during water-stress in C_3 plants, ESA SP-European Space Agency (Special Publication), 527, 26–29, https://doi.org/10.1034/j.1399-3054.2002.1140209.x, 2002.

Frankenberg, C. and Berry, J. A.: Solar Induced Chlorophyll Fluorescence: Origins, Relation to Photosynthesis and Retrieval, in: Comprehensive Remote Sensing, Elsevier, 143–162, 2017.

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J. E., Toon, G. C., Butz, A., Jung, M., Kuze, A., and Yokota, T.: New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity, Geophys. Res. Lett., 38, 1–6, https://doi.org/10.1029/2011GL048738, 2011.

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 Sens. Environ., 147, 1–12, https://doi.org/10.1016/j.rse.2014.02.007, 2014.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, Remote Sens. Environ., 114, 168–182, https://doi.org/10.1016/j.rse.2009.08.016, 2010.

Gentine, P. and Alemohammad, S. H.: Reconstructed Solar Induced Fluorescence: a machine-learning vegetation product based on MODIS surface reflectance to reproduce GOME-2 solar-induced fluorescence, Geophys. Res. Lett., 45, 3136–3146, 2018a.

Gentine, P. and Alemohammad, S. H.: RSIF (Reconstructed Solar Induced Fluorescence): a machine-learning vegetation product based on MODIS surface reflectance to reproduce GOME-2 solar induced fluorescence, Geophys. Res. Lett., 45, 3136–3146, https://doi.org/10.1002/2017GL076294, 2018b.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G.: Could Machine Learning Break the Convection Parameterization Deadlock?, Geophys. Res. Lett., 45, 5742–5751, https://doi.org/10.1029/2018GL078202, 2018.

Genty, B., Briantais, J. M., and Baker, N. R.: The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence, Biochimica et Biophysica Acta – General Subjects, 990, 87–92, https://doi.org/10.1016/S0304-4165(89)80016-9, 1989.

Gu, L., Baldocchi, D., Verma, S. B., Black, T. A., Vesala, T., Falge, E. M., and Dowty, P. R.: Advantages of diffuse radiation for terrestrial ecosystem productivity, J. Geophys. Res.-Atmos., 107, ACL 2-1–ACL 2-23, https://doi.org/10.1029/2001JD001242, 2002.

Guan, K., Berry, J. A., Zhang, Y., Joiner, J., Guanter, L., Badgley, G., and Lobell, D. B.: Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence, Glob. Change Biol., 22, 716–726, https://doi.org/10.1111/gcb.13136, 2016.

Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P. E., Gómez-Dans, J., Kuze, A., Suto, H., and Grainger, R. G.: Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements, Remote Sens. Environ., 121, 236–251, https://doi.org/10.1016/j.rse.2012.02.006, 2012.

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J. A., Frankenberg, C., Huete, A. R., Zarco-Tejada, P., Lee, J.-E., Moran, M. S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J. M., and Griffis, T. J.: Global and time-resolved monitoring of crop

photosynthesis with chlorophyll fluorescence, P. Natl. Acad. Sci. USA, 111, E1327–E1333, https://doi.org/10.1073/pnas.1320008111, 2014.

He, L., Chen, J. M., Liu, J., Mo, G., and Joiner, J.: Angular normalization of GOME-2 Sun-induced chlorophyll fluorescence observation as a better proxy of vegetation productivity, Geophys. Res. Lett., 44, 2017GL073708, https://doi.org/10.1002/2017GL073708, 2017.

Ji, L. and Peters, A. J.: Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices, Remote Sens. Environ., 87, 85–98, https://doi.org/10.1016/S0034-4257(03)00174-3, 2003.

Jiang, C. and Ryu, Y.: Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS), Remote Sens. Environ., 186, 528–547, https://doi.org/10.1016/j.rse.2016.08.030, 2016.

Joiner, J., Yoshida, Y., Vasilkov, A. P., Yoshida, Y., Corp, L. A., and Middleton, E. M.: First observations of global and seasonal terrestrial chlorophyll fluorescence from space, Biogeosciences, 8, 637–651, https://doi.org/10.5194/bg-8-637-2011, 2011.

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., Huemmrich, K. F., Yoshida, Y., and Frankenberg, C.: Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2, Atmos. Meas. Tech., 6, 2803–2823, https://doi.org/10.5194/amt-6-2803-2013, 2013.

Joiner, J., Yoshida, Y., Vasilkov, A. P., Schaefer, K., Jung, M., Guanter, L., Zhang, Y., Garrity, S., Middleton, E. M., Huemmrich, K. F., Gu, L., and Belelli Marchesini, L.: The seasonal cycle of satellite chlorophyll fluorescence observations and its relationship to vegetation phenology and ecosystem atmosphere carbon exchange, Remote Sens. Environ., 152, 375–391, https://doi.org/10.1016/j.rse.2014.06.022, 2014.

Joiner, J., Yoshida, Y., Guanter, L., and Middleton, E. M.: New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: simulations and application to GOME-2 and SCIAMACHY, Atmos. Meas. Tech., 9, 3939–3967, https://doi.org/10.5194/amt-9-3939-2016, 2016.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, J. Geophys. Res.-Biogeo., 116, 1–16, https://doi.org/10.1029/2010JG001566, 2011.

Knyazikhin, Y., Schull, M. A., Stenberg, P., Mottus, M., Rautiainen, M., Yang, Y., Marshak, A., Latorre Carmona, P., Kaufmann, R. K., Lewis, P., Disney, M. I., Vanderbilt, V., Davis, A. B., Baret, F., Jacquemoud, S., Lyapustin, A., and Myneni, R. B.: Hyperspectral remote sensing of foliar nitrogen content, P. Natl. Acad. Sci. USA, 110, E185–E192, https://doi.org/10.1073/pnas.1210196109, 2013.

Kolassa, J., Reichle, R. H., Liu, Q., Alemohammad, S. H., Gentine, P., Aida, K., Asanuma, J., Bircher, S., Caldwell, T., Colliander, A., Cosh, M., Holifield Collins, C., Jackson, T. J., Martínez-Fernández, J., McNairn, H., Pacheco, A., Thibeault, M., and Walker, J. P.: Estimating surface soil moisture from SMAP observations using a Neural Network technique, Remote Sens. Environ., 204, 43–59, https://doi.org/10.1016/j.rse.2017.10.045, 2018.

Krause, G. H. and Weis, E.: Chlorophyll Fluorescence and Photosynthesis: The Basics, Annu. Rev. Plant Phys., 42, 313–349, https://doi.org/10.1146/annurev.pp.42.060191.001525, 1991.

Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A.: Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data, IEEE Geosci. Remote Sens. Lett., 14, 778–782, https://doi.org/10.1109/LGRS.2017.2681128, 2017.

Lee, J.-E., Frankenberg, C., van der Tol, C., Berry, J. A., Guanter, L., Boyce, C. K., Fisher, J. B., Morrow, E., Worden, J. R., Asefi, S., Badgley, G., and Saatchi, S.: Forest productivity and water stress in Amazonia: observations from GOSAT chlorophyll fluorescence, P. Roy. Soc. B, 280, 20130171–20130171, https://doi.org/10.1098/rspb.2013.0171, 2013.

Lee, J.-E., Berry, J. A., van der Tol, C., Yang, X., Guanter, L., Damm, A., Baker, I., and Frankenberg, C.: Simulations of chlorophyll fluorescence incorporated into the Community Land Model version 4, Glob. Change Biol., 21, 3469–3477, https://doi.org/10.1111/gcb.12948, 2015.

Li, X., Xiao, J., and He, B.: Chlorophyll fluorescence observed by OCO-2 is strongly related to gross primary productivity estimated from flux towers in temperate forests, Remote Sens. Environ., 204, 659–671, https://doi.org/10.1016/j.rse.2017.09.034, 2018a.

Li, X., Xiao, J., He, B., Arain, M. A., Beringer, J., Desai, A. R., Emmel, C., Hollinger, D. Y., Krasnova, A., Mammarella, I., Noe, S. M., Ortiz, P. S., Rey-Sanchez, C., Rocha, A. V., and Varlagin, A.: Solar-induced chlorophyll fluorescence is strongly correlated with terrestrial photosynthesis for a wide variety of biomes: First global analysis based on OCO-2 and flux tower observations, Glob. Change Biol., 24, 3990–4008, https://doi.org/10.1111/gcb.14297, 2018b.

Liu, L., Liu, X., Wang, Z., and Zhang, B.: Measurement and Analysis of Bidirectional SIF Emissions in Wheat Canopies, IEEE Trans. Geosci. Remote Sens., 54, 2640–2651, https://doi.org/10.1109/TGRS.2015.2504089, 2016.

Lu, X., Cheng, X., Li, X., and Tang, J.: Opportunities and challenges of applications of satellite-derived sun-induced fluorescence at relatively high spatial resolution, Sci. Total Environ., 619/620, 649–653, https://doi.org/10.1016/j.scitotenv.2017.11.158, 2018.

Luo, Y. Q., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., Ciais, P., Dalmonech, D., Fisher, J. B., Fisher, R., Friedlingstein, P., Hibbard, K., Hoffman, F., Huntzinger, D., Jones, C. D., Koven, C., Lawrence, D., Li, D. J., Mahecha, M., Niu, S. L., Norby, R., Piao, S. L., Qi, X., Peylin, P., Prentice, I. C., Riley, W., Reichstein, M., Schwalm, C., Wang, Y. P., Xia, J. Y., Zaehle, S., and Zhou, X. H.: A framework for benchmarking land models, Biogeosciences, 9, 3857–3874, https://doi.org/10.5194/bg-9-3857-2012, 2012.

Migliavacca, M., Perez-Priego, O., Rossini, M., El-Madany, T. S., Moreno, G., van der Tol, C., Rascher, U., Berninger, A., Bessenbacher, V., Burkart, A., Carrara, A., Fava, F., Guan, J. H., Hammer, T. W., Henkel, K., Juarez-Alcalde, E., Julitta, T., Kolle, O., Martín, M. P., Musavi, T., Pacheco-Labrador, J., Pérez-Burgueño, A., Wutzler, T., Zaehle, S., and Reichstein, M.: Plant functional traits and canopy structure control the relationship between photosynthetic CO_2 uptake and far-red sun-induced fluorescence in a Mediterranean grassland under different nutrient availability, New Phytol., 214, 1078–1091, https://doi.org/10.1111/nph.14437, 2017.

Monteith, J. L.: Solar Radiation and Productivity in Tropical Ecosystems, J. Appl. Ecol., 9, 747–766, https://doi.org/10.2307/2401901, 1972.

Müller, J. N. C.: Untersuchungen über die diffusion der atmosphärischen gase und die gasausscheidung unter verschiedenen beleuchtungs-bedingungen, Jahrbucher fur Wissenschaftliche Botanik, 9, 36–49, 1874.

Porcar-Castell, A., Pfündel, E., Korhonen, J. F. J., and Juurola, E.: A new monitoring PAM fluorometer (MONI-PAM) to study the short- and long-term acclimation of photosystem II in field conditions, Photosynthesis Research, 96, 173–179, https://doi.org/10.1007/s11120-008-9292-3, 2008.

Porcar-Castell, A., Tyystjärvi, E., Atherton, J., Van Der Tol, C., Flexas, J., Pfündel, E. E., Moreno, J., Frankenberg, C., and Berry, J. A.: Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: Mechanisms and challenges, J. Exp. Bot., 65, 4065–4095, https://doi.org/10.1093/jxb/eru191, 2014.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J. M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review

and improved algorithm, Glob. Change Biol., 11, 1424–1439, https://doi.org/10.1111/j.1365-2486.2005.001002.x, 2005.

Roháček, K. and Barták, M.: Technique of the modulated chlorophyll fluorescence: Basic concepts, useful parameters, and some applications, Photosynthetica, 37, 339–363, 1999.

Romero, J. M., Cordon, G. B., and Lagorio, M. G.: Modeling re-absorption of fluorescence from the leaf to the canopy level, Remote Sens. Environ., 204, 138–146, https://doi.org/10.1016/j.rse.2017.10.035, 2018.

Rossini, M., Nedbal, L., Guanter, L., Ač, A., Alonso, L., Burkart, A., Cogliati, S., Colombo, R., Damm, A., Drusch, M., Hanus, J., Janoutova, R., Julitta, T., Kokkalis, P., Moreno, J., Novotny, J., Panigada, C., Pinto, F., Schickling, A., Schüttemeyer, D., Zemek, F., and Rascher, U.: Red and far red Sun-induced chlorophyll fluorescence as a measure of plant photosynthesis, Geophys. Res. Lett., 42, 1632–1639, https://doi.org/10.1002/2014GL062943, 2015.

Rossini, M., Meroni, M., Celesti, M., Cogliati, S., Julitta, T., Panigada, C., Rascher, U., van der Tol, C., and Colombo, R.: Analysis of red and far-red sun-induced chlorophyll fluorescence and their ratio in different canopies based on observed and modeled data, Remote Sens., 8, p. 412, https://doi.org/10.3390/rs8050412, 2016.

Ryu, Y., Jiang, C., Kobayashi, H., and Detto, M.: MODIS-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000, Remote Sens. Environ., 204, 812–825, https://doi.org/10.1016/j.rse.2017.09.021, 2018.

Schaaf, C.: MODIS BRDF/Albedo User Guide, available from: https://www.umb.edu/spectralmass/, last access: 27 September 2018.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J.-P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R. P., Hu, B., Liang, S., Privette, J. L., and Roy, D.: First operational BRDF, albedo nadir reflectance products from MODIS, Remote Sens. Environ., 83, 135–148, https://doi.org/10.1016/S0034-4257(02)00091-3, 2002.

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, Glob. Change Biol., 21, 1762–1776, https://doi.org/10.1111/gcb.12822, 2015.

Song, L., Guanter, L., Guan, K., You, L., Huete, A., Ju, W., and Zhang, Y.: Satellite sun-induced chlorophyll fluorescence detects early response of winter wheat to heat stress in the Indian Indo-Gangetic Plains, Glob. Change Biol., 24, 4023–4037, https://doi.org/10.1111/gcb.14302, 2018.

Sun, Y., Fu, R., Dickinson, R., Joiner, J., Frankenberg, C., Gu, L., Xia, Y., and Fernando, N.: Drought onset mechanisms revealed by satellite solar-induced chlorophyll fluorescence: Insights from two contrasting extreme events, J.

Geophys. Res.-Biogeo., 120, 2427–2440, https://doi.org/10.1002/2015JG003150, 2015.

Sun, Y., Frankenberg, C., Wood, J. D., Schimel, D. S., Jung, M., Guanter, L., Drewry, D. T., Verma, M., Porcar-Castell, A., Griffis, T. J., Gu, L., Magney, T. S., Köhler, P., Evans, B., and Yuen, K.: OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence, Science, 358, 6360, https://doi.org/10.1126/science.aam5747, 2017.

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 Sens. Environ., 209, 808–823, https://doi.org/10.1016/j.rse.2018.02.016, 2018.

Tadic, J. M., Qiu, X., Yadav, V., and Michalak, A. M.: Mapping of satellite Earth observations using moving window block kriging, Geosci. Model Dev., 8, 3311–3319, https://doi.org/10.5194/gmd-8-3311-2015, 2015.

Tadić, J. M., Qiu, X., Miller, S., and Michalak, A. M.: Spatio-temporal approach to moving window block kriging of satellite data v1.0, Geosci. Model Dev., 10, 709–720, https://doi.org/10.5194/gmd-10-709-2017, 2017.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13, 4291–4313, https://doi.org/10.5194/bg-13-4291-2016, 2016.

Turner, D. P., Ritts, W. D., Styles, J. M., Yang, Z., Cohen, W. B., Law, B. E., and Thornton, P. E.: A diagnostic carbon flux model to monitor the effects of disturbance and interannual variation in climate on regional NEP, Tellus B, 58, 476–490, https://doi.org/10.1111/j.1600-0889.2006.00221.x, 2006.

Urli, M., Porté, A. J., Cochard, H., Guengant, Y., Burlett, R., and Delzon, S.: Xylem embolism threshold for catastrophic hydraulic failure in angiosperm trees, Tree Physiol., 33, 672–683, https://doi.org/10.1093/treephys/tpt030, 2013.

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.

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, J. Geophys. Res.-Biogeo., 119, 2312–2327, https://doi.org/10.1002/2014JG002713, 2014.

Verma, M., Schimel, D., Evans, B., Frankenberg, C., Beringer, J., Drewry, D. T., Magney, T., Marang, I., Hutley, L., Moore, C., and Eldering, A.: Effect of

environmental conditions on the relationship between solar-induced fluorescence and gross primary productivity at an OzFlux grassland site, J. Geophys. Res.-Biogeo., 122, 716–733, https://doi.org/10.1002/2016JG003580, 2017.

Verrelst, J., Rivera, J. P., van der Tol, C., Magnani, F., Mohammed, G., and Moreno, J.: Global sensitivity analysis of the SCOPE model: What drives simulated canopy-leaving sun-induced fluorescence?, Remote Sens. Environ., 166, 8–21, https://doi.org/10.1016/j.rse.2015.06.002, 2015.

Verrelst, J., van der Tol, C., Magnani, F., Sabater, N., Rivera, J. P., Mohammed, G., and Moreno, J.: Evaluating the predictive power of sun-induced chlorophyll fluorescence to estimate net photosynthesis of vegetation canopies: A SCOPE modeling study, Remote Sens. Environ., 176, 139–151, https://doi.org/10.1016/j.rse.2016.01.018, 2016.

Walther, S., Voigt, M., Thum, T., Gonsamo, A., Zhang, Y., Köhler, P., Jung, M., Varlagin, A., and Guanter, L.: Satellite chlorophyll fluorescence measurements reveal large-scale decoupling of photosynthesis and greenness dynamics in boreal evergreen forests, Glob. Change Biol., 22, 2979–2996, https://doi.org/10.1111/gcb.13200, 2016.

Wang, S., Huang, C., Zhang, L., Lin, Y., Cen, Y., and Wu, T.: Monitoring and Assessing the 2012 Drought in the Great Plains: Analyzing Satellite-Retrieved Solar-Induced Chlorophyll Fluorescence, Drought Indices, and Gross Primary Production, Remote Sens., 8, p. 61, https://doi.org/10.3390/rs8020061, 2016.

Wu, C., Niu, Z., Tang, Q., and Huang, W.: Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation, Agr. Forest Meteorol., 148, 1230–1241, https://doi.org/10.1016/j.agrformet.2008.03.005, 2008.

Yang, P. and van der Tol, C.: Linking canopy scattering of far-red sun-induced chlorophyll fluorescence with reflectance, Remote Sens. Environ., 209, 456–467, https://doi.org/10.1016/j.rse.2018.02.029, 2018.

Zhang, Q., Xiao, X., Braswell, B., Linder, E., Baret, F., and Mooreiii, B.: Estimating light absorption by chlorophyll, leaf and canopy in a deciduous broadleaf forest using MODIS data and a radiative transfer model, Remote Sens. Environ., 99, 357–371, https://doi.org/10.1016/j.rse.2005.09.009, 2005.

Zhang, Q., Middleton, E. M., Cheng, Y.-B., Huemmrich, K. F., Cook, B. D., Corp, L. A., Kustas, W. P., Russ, A. L., Prueger, J. H., and Yao, T.: Integrating chlorophyll fAPAR and nadir photochemical reflectance index from EO-1/Hyperion to predict cornfield daily gross primary production, Remote Sens. Environ., 186, 311–321, https://doi.org/10.1016/j.rse.2016.08.026, 2016a.

Zhang, Y., Peng, C., Li, W., Fang, X., Zhang, T., Zhu, Q., Chen, H., and Zhao, P.: Monitoring and estimating drought-induced impacts on forest structure,

growth, function, and ecosystem services using remote-sensing data: recent progress and future challenges, Environ. Rev., 21, 103–115, 2013.

Zhang, Y., Guanter, L., Berry, J. A., Joiner, J., van der Tol, C., Huete, A., Gitelson, A., Voigt, M., and Köhler, P.: Estimation of vegetation photosynthetic capacity from space-based measurements of chlorophyll fluorescence for terrestrial biosphere models, Glob. Change Biol., 20, 3727–3742, https://doi.org/10.1111/gcb.12664, 2014.

Zhang, Y., Xiao, X., Jin, C., Dong, J., Zhou, S., Wagle, P., Joiner, J., Guanter, L., Zhang, Y., Zhang, G., Qin, Y., Wang, J., and Moore, B.: Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America, Remote Sens. Environ., 183, 154–169, https://doi.org/10.1016/j.rse.2016.05.015, 2016b.

Zhang, Y., Guanter, L., Berry, J. A., van der Tol, C., Yang, X., Tang, J., and Zhang, F.: Model-based analysis of the relationship between sun-induced chlorophyll fluorescence and gross primary production for remote sensing applications, Remote Sens. Environ., 187, 145–155, https://doi.org/10.1016/j.rse.2016.10.016, 2016c.

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, Sci. Data, 4, 170165, https://doi.org/10.1038/sdata.2017.165, 2017a.

Zhang, Y., Song, C., Band, L. E., Sun, G., and Li, J.: Reanalysis of global terrestrial vegetation trends from MODIS products: Browning or greening?, Remote Sens. Environ., 191, 145–155, https://doi.org/10.1016/j.rse.2016.12.018, 2017b.

Zhang, Y., Xiao, X., Zhang, Y., Wolf, S., Zhou, S., Joiner, J., Guanter, L., Verma, M., Sun, Y., Yang, X., Paul-Limoges, E., Gough, C. M., Wohlfahrt, G., Gioli, B., van der Tol, C., Yann, N., Lund, M.. and de Grandcourt, A.: On the relationship between sub-daily instantaneous and daily total gross primary production: Implications for interpreting satellite-based SIF retrievals, Remote Sens. Environ., 205, 276–289, https://doi.org/10.1016/j.rse.2017.12.009, 2018a.

Zhang, Y., Joiner, J., Gentine, P., and Zhou, S.: Reduced solar-induced chlorophyll fluorescence from GOME-2 during Amazon drought caused by dataset artifacts, Glob. Change Biol., 24, 2229–2230, https://doi.org/10.1111/gcb.14134, 2018b.

Zhang, Y., Xiao, X., Wolf, S., Wu, J., Wu, X., Gioli, B., Cescatti, A., Van Der Tol, C., Zhou, S., Gough, C., Gentine, P., Zhang, Y., Steinbrecher, R., and Ardö, J.: Spatio-temporal convergence of maximum daily light-use efficiency based on radiation absorption by canopy chlorophyll, Geophys. Res. Lett., 45, 3508–3519, https://doi.org/10.1029/2017GL076354, 2018c.

Zhao, M., Heinsch, F. A., Nemani, R. R., and Running, S. W.: Improvements of the MODIS terrestrial gross and net primary production global data set, Remote Sens. Environ., 95, 164–176, https://doi.org/10.1016/j.rse.2004.12.011, 2005.

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., and Fraundorfer, F.: Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources, IEEE Geoscience and Remote Sensing Magazine, 5, 8–36, https://doi.org/10.1109/MGRS.2017.2762307, 2017.

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peñuelas, J., Poulter, B., Pugh, T. A. M., Stocker, B. D., Viovy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S., and Zeng, N.: Greening of the Earth and its drivers, Nat. Clim. Change, 6, 791–795, https://doi.org/10.1038/nclimate3004, 2016.