Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL076294

Key Points:
- A new machine learning-based vegetation product, Reconstructed Solar-Induced Fluorescence (RSIF), is developed using visible and near-infrared MODIS channels
- RSIF improves SIF with better correlation with in situ eddy covariance and remote sensing-based photosynthesis and lower noise
- RSIF has high spatial resolution and a long record and does not saturate, unlike optical vegetation indices

Supporting Information:
- Supporting Information S1

Correspondence to:
P. Gentine,
pg2328@columbia.edu

Citation:
Gentine, P., & Alemohammad, S. H. (2018). Reconstructed Solar-Induced Fluorescence: A machine learning vegetation product based on MODIS surface reflectance to reproduce GOME-2 solar-induced fluorescence. *Geophysical Research Letters*, *45*. https://doi.org/10.1002/2017GL076294

Received 4 NOV 2017
Accepted 13 MAR 2018
Accepted article online 23 MAR 2018

Reconstructed Solar-Induced Fluorescence: A Machine Learning Vegetation Product Based on MODIS Surface Reflectance to Reproduce GOME-2 Solar-Induced Fluorescence

P. Gentine[1] **and S. H. Alemohammad**[1,2]

[1]Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA, [2]Radiant.Earth, Washington, DC, USA

Abstract Solar-induced fluorescence (SIF) observations from space have resulted in major advancements in estimating gross primary productivity (GPP). However, current SIF observations remain spatially coarse, infrequent, and noisy. Here we develop a machine learning approach using surface reflectances from Moderate Resolution Imaging Spectroradiometer (MODIS) channels to reproduce SIF normalized by clear sky surface irradiance from the Global Ozone Monitoring Experiment-2 (GOME-2). The resulting product is a proxy for ecosystem photosynthetically active radiation absorbed by chlorophyll ($fAPAR_{Ch}$). Multiplying this new product with a MODIS estimate of photosynthetically active radiation provides a new MODIS-only reconstruction of SIF called Reconstructed SIF (RSIF). RSIF exhibits much higher seasonal and interannual correlation than the original SIF when compared with eddy covariance estimates of GPP and two reference global GPP products, especially in dry and cold regions. RSIF also reproduces intense productivity regions such as the U.S. Corn Belt contrary to typical vegetation indices and similarly to SIF.

Plain Language Summary A new proxy for photosynthesis is developed using Moderate Resolution Imaging Spectroradiometer observations and a machine learning approach. The new product is able to effectively reproduce observations from eddy covariance towers and more sophisticated photosynthesis models that rely on more information (such as weather information).

1. Introduction

Estimating gross primary production (GPP), or photosynthesis, at the global scale is essential for various applications ranging from yield prediction (Guan et al., 2016) to evaluating and predicting the impact of regional and global environmental changes (Friend et al., 2007; Le Quéré et al., 2009; McDowell et al., 2015; Poulter et al., 2014). To correctly evaluate the impact of environmental changes, such as land use land cover changes, remote sensing estimates of GPP require both fine spatial resolution, to capture the diversity of ecosystem response to environmental drivers, and long-term record, to assess interannual variability and long-term trends.

In the last decades, there has been substantial development in remote sensing proxies for GPP. GPP can be written as $GPP = LUE \times fPAR_{Ch} \times PAR$, with LUE the light use efficiency, $fPAR_{Ch}$ the fraction of photosynthetically active radiation absorbed by chlorophyll, and PAR the incoming photosynthetically active radiation (Monteith, 1972). Plants have distinct spectral signatures with low reflectance in the visible range and high reflectance in the near infrared, which is directly related to fPAR (Sellers, 2007; Tucker, 1979). One natural proxy was thus the Normalized Difference Vegetation Index (NDVI), which takes advantage of such spectral signature to monitor vegetation changes (Tucker, 1979). Later, the Enhanced Vegetation Index (EVI) was developed to reduce the saturation seen in NDVI with high biomass and to reduce the canopy background signal (Huete et al., 2002, 2006). We note that the original EVI algorithm included a blue channel information, but a simplified two-band version was then developed so that it could be used with the formed Advanced Very High Resolution Radiometer (AVHRR) sensor (Jiang et al., 2008).

The last few years have seen tremendous development in the remote sensing observations of GPP. Indeed, global observations of solar-induced fluorescence (SIF) have been shown to be feasible on various platforms, first with the Greenhouse Gases Observing SATellite (Frankenberg et al., 2011; Joiner et al., 2011), then with Global Ozone Monitoring Experiment-2 (GOME-2) and SCanning Imaging Absorption SpectroMeter for

Atmospheric CHartographY (Köhler et al., 2014), and then later with the Orbiting Carbon Observatory (OCO)-2 (Frankenberg et al., 2014). SIF has a small-amplitude signal, so it was not possible to observe it until very recently. SIF observations have been shown to be directly pertinent to estimate crop photosynthesis (Guanter et al., 2014) and yield (Guan et al., 2016), GPP across ecosystems (Lee et al., 2015; Yang et al., 2015; Zhang, Xiao, Jin, et al., 2016), water stress (Guan et al., 2015; Konings et al., 2017; Sun et al., 2015; Zhang, Xiao, Guanter, et al., 2016), biosphere-atmosphere interactions (Green et al., 2017), surface turbulent fluxes (Alemohammad et al., 2017), and phenology, especially in northern latitudes where vegetation indices and their seasonality are polluted by the snow albedo (Jeong et al., 2017). One other advantage of SIF is that it responds to only the PAR absorbed by chlorophyll of the canopy, whereas typical optical (absorbed photosynthetic active radiation) APAR or fPAR products reflect the PAR absorbed by the entire canopy (nonphotosynthetic and photosynthetic; Song et al., 2013).

Similar to GPP, SIF can be written using a LUE strategy as follows: SIF = $fesc \times \varepsilon \times \mathrm{fPAR}_{\mathrm{Ch}} \times \mathrm{PAR}$ (Joiner et al., 2014; Lee et al., 2013), with *fesc* a parameter related to the optical properties of the canopy and sensor angle (accounting for fraction of SIF photons escaping from leaf level to the canopy level) and ε as SIF yield. SIF variations at the leaf level and at short time scales may reflect changes in LUE (Agati et al., 1995; Corp et al., 2003; Long & Bernacchi, 2003; van der Tol et al., 2009); however, it is less clear whether the signal seen by coarse satellite observations over longer term period (several days) reflects mainly changes in LUE. Indeed, it appears that most of the variability of the SIF signal at the ecosystem might be explained by $\mathrm{APAR}_{\mathrm{Ch}}$, that is, the product of $\mathrm{fPAR}_{\mathrm{Ch}}$ and PAR, thus reflecting changes in canopy structure and light absorption (Badgley et al., 2017; Li et al., 2017), at least for the far-red signal (Du et al., 2017). In addition, downscaling using vegetation indices improves the correlation of SIF with eddy covariance towers, further emphasizing the importance of fPAR (Duveiller & Cescatti, 2016). Variations in LUE appear to be happening at short time scales and might not be easy to detect given the inherent noise in the small-amplitude SIF remote sensing observations. As a result, a new index, NIRv, was developed that is the product of near infrared (NIR) and NDVI, as a way to emphasize the NIR reflectance, which is less sensitive to background contamination. This new index performed well compared to eddy covariance towers (Badgley et al., 2017). Nonetheless, such approach directly still relies on NDVI and any vegetation index is ultimately empirical. In addition, it still exhibits some interference of background reflectance due to NDVI, especially in low light conditions and snowy conditions (supporting information Figure S1), as it cannot directly have information on PAR.

To alleviate those issues, our primary goal here is to define an objective vegetation product, which extracts as much information content available in the Moderate Resolution Imaging Spectroradiometer (MODIS) data, to predict GPP. Based on the realization that much of the variability in GPP and SIF might be due to changes in $\mathrm{fPAR}_{\mathrm{Ch}}$ and PAR, the MODIS observations might be able to yield nearly as much information as the SIF data. In addition, the MODIS data have the advantage that it is available over a longer period of time (since July 2002) and at much higher resolution (500 m) than the current SIF data streams, so that ecosystem responses to environmental changes can be better assessed. To do so, we use the reflectance data from MODIS on board the Aqua satellite to predict GOME-2 SIF data, normalized by clear-sky surface irradiance, at a scale of 0.5° resolution using a neural network algorithm. The resulting product, which is a proxy for $\mathrm{fPAR}_{\mathrm{Ch}}$, is then multiplied by a recent remote sensing PAR product (Breathing Earth System Simulator, BESS), which is also based on MODIS (Ryu et al., 2017). The new MODIS-only based product, called Reconstructed Solar-Induced Fluorescence (RSIF), can then be applied to the 0.05° data to estimate GPP at fine spatial scale. We note that the 0.05° RSIF product downscaled to 0.5° does not exactly match the 0.5° product computed on the coarse-scale reflectances because of the nonlinearity of the neural network, but the differences are very small (~1e−3 mW m^{-2} sr^{-1} nm^{-1}).

2. Data

2.1. MODIS Surface Reflectance

We use the bidirectional reflectance distribution function corrected surface reflectance product (MYD09A1, https://doi.org/10.5067/modis/myd09a1.006) of the Aqua satellite covering the following frequency ranges: 459–479 nm (blue), 545–565 nm (green), 620–670 nm (orange/red) and 841–876 nm (NIR). We also test the impact of the infrared channels IR1 (1230,–1,250 nm), IR2 (1,628–1,652 nm), and IR3 (2,105–2,155 nm). We

use the 500 m resolution data at 8-day temporal resolution. The data are then coarse grained to 0.5°, similar to the GOME-2 SIF data, using a spatial average across the finer-scale pixels. The data cover the period from July 2002 to present.

2.2. GOME-2 SIF

We used the GOME-2 version 26 (V26) far-red 740 nm (near O_2 A-band) terrestrial chlorophyll fluorescence data retrieval, which was provided by Dr. Joanna Joiner (Joiner et al., 2013, 2014, 2016). The data are available at 0.5° spatial resolution and at biweekly temporal resolution which covers the period January 2007 to September 2017.

2.3. FLUXNET 2015 GPP

To validate our fine-scale retrieval, we use GPP estimates from 74 eddy covariance sites based on the tier 1 product of the recent FLUXNET 2015 data set (http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/). We specifically use the estimated GPP product using the nighttime respiration partitioning method (Papale et al., 2017). Furthermore, data were filtered using procedures used in previous studies (Zhou et al., 2015).

2.4. FLUXNET-Multiple Tree Ensemble GPP

FLUXNET-MTE (Multiple Tree Ensemble) is a machine learning global upscaling eddy covariance estimate of monthly GPP available from 1983 to 2011 at 0.5° resolution (Beer et al., 2010; Jung et al., 2009). The model uses as input weather information from the Climatic Research Units data set along with fPAR information based on MODIS. FLUXNET-MTE has been shown to perform well compared to eddy covariance estimates of GPP, especially in terms of seasonal cycle (Beer et al., 2010; Jung et al., 2009), but underestimates the interannual variability.

2.5. MODIS GPP

In addition to the FLUXNET-MTE GPP product we use the GPP MODIS (MOD17) estimate. The MOD17 algorithm is based on a light-use efficiency approach (Monteith, 1977). We use the monthly MOD17A2 GPP product (Running, 2004; Zhao et al., 2005, 2006). MOD17A2 is available from 2000 on and is provided at 1 km resolution.

2.6. Water, Energy, and Carbon With Artificial Neural Networks GPP

Water, Energy, and Carbon with Artificial Neural Networks (WECANN) is a machine learning retrieval of GPP, which uses only remote sensing observations and in particular GOME-2 SIF to estimate surface fluxes at the global scale (Alemohammad et al., 2017). The retrieval is available at 1° and monthly resolution, from January 2007 to present. This retrieval has been shown to outperform other retrievals, especially in terms of the representation of seasonal and interannual variability (Alemohammad et al., 2017). The WECANN data set is used as a baseline comparison for RSIF, as GOME-2 SIF is used as input data of the neural network; hence, there is strong correlation between WECANN GPP and SIF. An increase in correlation of WECANN GPP with RSIF would indicate that the new product improves the regular SIF measurements (e.g., due to noise reduction and higher spatial resolution reflecting various surface conditions).

3. Methodology

In this work, we use a feedforward neural networks (NN) architecture using the reflectances of four MODIS Aqua channels as input to fit the GOME-2 SIF normalized by solar angle data (a proxy for clear-sky irradiance). To validate and test the NN, we divide the data into 80% for training, 10% for test, and 10% for validation. Such divide allows for testing for overfitting and avoidance of using data that were seen in the training data set for validation (similar to what is typically done for regression). A correctly fitted and nonoverfitting NN is able to generalize instead of just learning a functional relationship on existing data (Wan et al., 2013). There have been multiple successful applications of NN in the geosciences in recent years ranging from soil moisture retrievals (Jimenez et al., 2013; Kolassa et al., 2013, 2016; Kolassa, Gentine, et al., 2017; Kolassa, Reichle, & Draper, 2017) and surface temperature and emissivity retrievals (Aires et al., 2001) to surface flux retrievals (Alemohammad et al., 2017; Beer et al., 2010; Jimenez et al., 2009; Jung et al., 2009; Koirala et al., 2017).

The NN is trained on the GOME-2 SIF data set at 0.5° and biweekly resolution. Surface reflectance from four MODIS channels are used as input of the NN to predict the GOME-2 SIF normalized by solar angle data, using

all pixels and all times in the training data set. A mean square error cost function is used for training, and an Adam stochastic gradient descent algorithm is used to update the weights and biases of the NN. Activation functions for the hidden layers are the typical rectified linear unit functions (ReLU)—that is, 0 output below 0 input and a 1:1 relationship for positive values. The minimum size NN reaching good performance is then chosen to reduce overfitting. A summary of the different model performances, correlation and root-mean-square error, is presented in Figure S2 for four bands and in Figure S3 when using all bands, with only negligible improvements using infrared channels. To obtain the daily integrated RSIF, the NN-fitted normalized SIF is then multiplied by the daily average BESS PAR. The final RSIF product has an 8-day time scale consistent with MODIS MYD09A1 product, and it has units of mW m^{-2} sr^{-1} nm^{-1}.

4. Results

Overall, a relatively shallow feedforward NN with five neurons and one hidden layer (Figure S2) is able to correctly reproduce solar angle-normalized SIF (Figure 1), with a temporal and spatial correlation of 0.78 and 0.78, and root-mean-square error of 0.37 and 0.35 mW m^{-2} sr^{-1} nm^{-1} for the training and validation data sets (Figure 1), respectively. We note that this performance includes all periods of the year and all pixels so that it samples various climatic conditions and ecosystems. Given the inherent low signal-to-noise ratio of the SIF data, we should not expect a perfect match between NN-based and actual solar angle-normalized SIF data. In fact, given the high noise level in the target data, it is in our interest not to perfectly fit the observed solar angle-normalized SIF as otherwise the NN could learn the noise structure as well. In other words, we want the NN to learn the overall, mean, deterministic relationship between MODIS surface reflectances and normalized SIF, while reducing the inherent SIF noise. An important result is that the normalized RSIF and RSIF products do not saturate unlike NDVI and EVI. Indeed, normalized RSIF maintains a linear relationship with normalized GOME-2 SIF, even at the peak observed values of SIF (Figure 1). This is of specific interest as it demonstrates that RSIF performs well even in the presence of high biomass and thus can be used to study GPP variability over high biomass ecosystems. The sensitivity of the retrievals to the different channels can be diagnosed using the derivative of the NN to the input channels. This sensitivity analysis of the normalized RSIF shows that as expected NIR and red channels have an important positive contribution to the normalized RSIF signal similarly to NDVI (Table S1, which presents the mean sensitivity across input data as it is nonlinear). But, interestingly, a large contribution comes from the green channel. Since normalized RSIF is a close proxy for $fPAR_{Ch}$, the importance of the green channel likely reflects the abundance of chlorophyll pigments and thus vegetation health. Red to far red reflectances are also strongly correlated with chlorophyll content (Gitelson et al., 1999).

An important point is that the relationship between GPP and SIF varies between C3 and C4 species at least at the field scale (Liu et al., 2017) and is higher for C4 species; that is, GPP is higher for the same SIF amplitude (Liu et al., 2017). To account for this effect, we included a map of C3/C4 fraction (Meiyappan & Jain, 2012) during the training of an alternative NN, but this strongly deteriorates the retrieval compared to both eddy covariance data and other global retrievals. There could be several reasons for this. First, the different LUE relationships between C3 and C4 species might not be as clear at the coarse satellite scale. In addition, the C3/C4 characterization map might be too uncertain to yield useful information to the NN. We thus chose not to retain this additional information in the final product.

In a second set of experiments, we included the MODIS infrared channels as input to the NN. This allows testing the added information due to the infrared channels. The reduced four reflectance channels are able to explain variabilities similar to estimates with all-channel MODIS in the input (Figures S3 and S4, for the best performing configuration of the all-channel MODIS NN). Therefore, we decide to only use the first four channels (as described in section 2.1) as inputs, as the thermal infrared channels could be used for independent estimates of evapotranspiration, for instance.

To evaluate the RSIF retrieval, we use several approaches. The first approach is to compare the retrieval to two global estimates of GPP: FLUXNET-MTE and WECANN. This comparison sheds light on the spatial and temporal agreement of the retrieval with other products, which use more data input streams, such as meteorological information. The second approach is to compare the retrieval to in situ eddy covariance estimates of GPP across climates and plant functional types. This gives us further confidence in the quality of the retrieval. Finally, we investigate whether RSIF is able to capture the global maxima in GPP, observed in the Northern

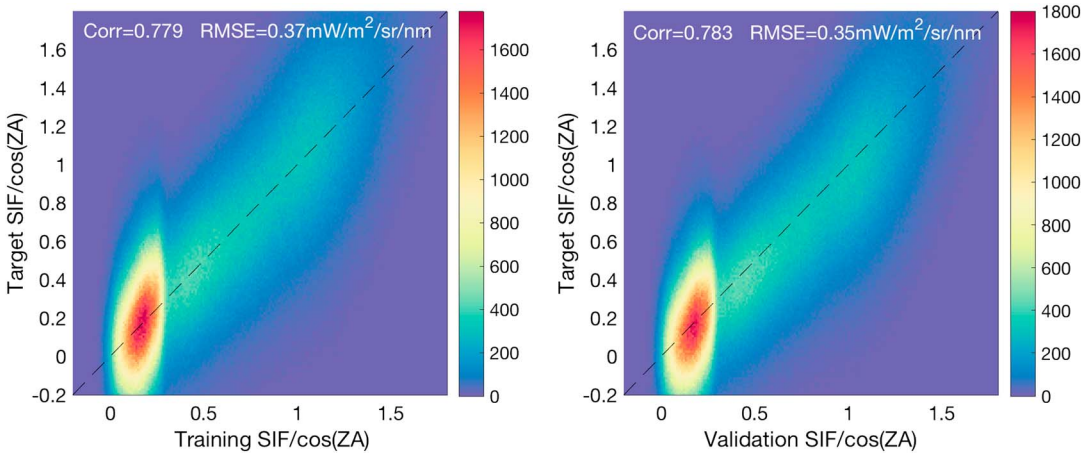

Figure 1. Performance of the neural network retrieval of Global Ozone Monitoring Experiment-2 SIF normalized by the cosine of Sun solar zenith angle using first four Moderate Resolution Imaging Spectroradiometer reflectance channels, at 0.5°, 16-day resolution over the training and validation test. The density of scatter points is represented by the shading color. The diagonal black dashed line depicts the 1:1 relationship.

Hemisphere agricultural regions, in particular in the U.S. Corn Belt regions, which are typically missed by vegetation indices but observed using CO_2 or carbonyl sulfide (COS) flux inversions (Campbell, Carmichael, & Chai, 2008; Hilton et al., 2017).

RSIF is highly temporally correlated with GOME-2 SIF in most regions with a typical temporal correlation above 0.8 (Figure 2), a mean temporal correlation of 0.62 across pixels, and a spatial correlation of 0.86. Lower temporal correlations are nonetheless observed in tropical rainforests, that is, the Amazon, Congo and Indonesia, where seasonality is weak (Anber et al., 2015; Xu et al., 2015), where cloud interference affects the retrievals and therefore the noise level, and where light structure and leaf photosynthetic capacity changes are important for the seasonality (Morton & Cook, 2016; Saleska et al., 2016). In addition, in those tropical rainforests variations in vegetation structure is small (Huete et al., 2006; Saleska et al., 2016) so that changes in LUE might be more important and not observable with the MODIS data. Correlations are also lower in very dry regions, because seasonality is weak and the noise level in the SIF product is high compared to the actual signal and compared to MODIS reflectances, which are less noisy. In high northern latitudes, reduced temporal correlations between SIF and RSIF is also noticeable. This decorrelation is due to the small SIF signal which results in reduced signal-to-noise ratio. Snow could potentially reduce the quality of the RSIF retrievals because it affects the MODIS reflectance signal, and related vegetation indices such as NDVI and EVI, but it is not the case as we show below.

Comparison with two reference global GPP products, FLUXNET-MTE and WECANN, show that RSIF performs well in terms of temporal correlation throughout the entire globe with a mean of 0.88 correlation with FLUXNET-MTE and 0.81 with WECANN (Figure 2). Tropical rainforests and the Amazon, in particular, display the lowest correlations (0.3 to 0.6) but are also regions where seasonality is the weakest (Huete et al., 2006). Those are also regions where representation of GPP is the most challenging and thus where we expect the most divergence across GPP retrievals and models. Notably, correlation of RSIF with different products is very good at high latitudes, which is usually a challenge for vegetation indices which are oversensitive to snow cover and tend to mix snow melt (and thus an apparent greening) with photosynthetic activity (Jeong et al., 2017).

In contrast, the GOME-2 SIF temporal correlation with the FLUXNET-MTE and WECANN GPP products are systematically lower than the RSIF product, especially in cold and midlatitude regions, with a mean of 0.675 and 0.73, respectively, across flux tower sites. We further emphasize that the WECANN product uses SIF as an input of the retrieval. It is therefore very encouraging to see that the RSIF product better captures the GPP variability of WECANN compared to the original SIF data. This important result is likely the result of the noise reduction when using the MODIS channels as opposed to the SIF data for GPP estimate.

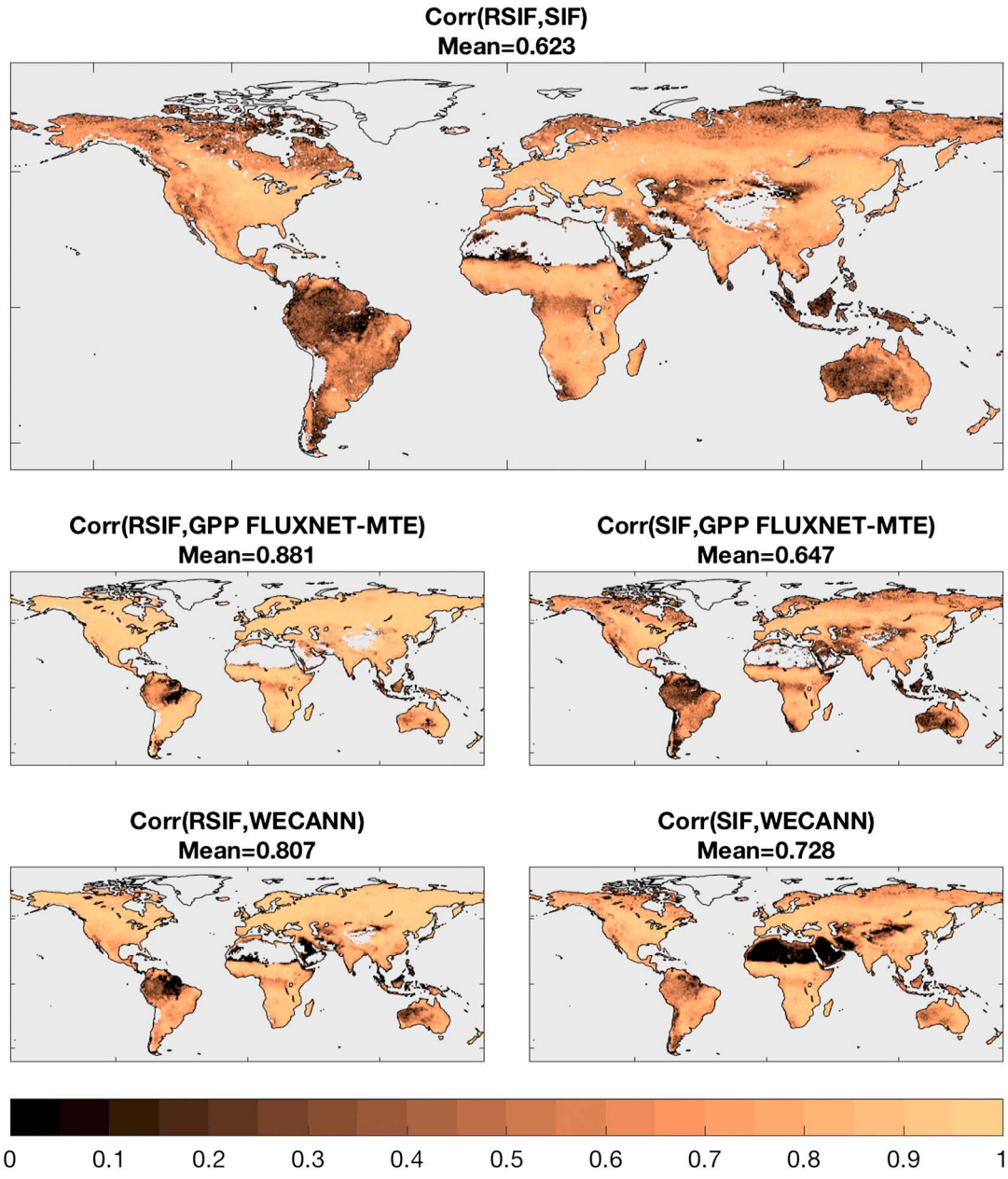

Figure 2. Pixel-wise temporal correlations between Reconstructed Solar-Induced Fluorescence (RSIF) based on rectified linear unit activation function with four input reflectance channels, five neurons through one hidden layer and solar-induced fluorescence (SIF; top), and SIF/RSIF with two global estimates of gross primary productivity (GPP): FLUXNET-Multiple Tree Ensemble (MTE) and Water, Energy, and Carbon with Artificial Neural Networks (WECANN). The correlations are computed over the longest period possible based on data limitations (Global Ozone Monitoring Experiment-2 SIF 2007–2016, MTE GPP 2007–2011, and WECANN GPP 2007–2015).

The RSIF product is then compared to local flux observations form the FLUXNET 2015 data set. RSIF has a mean correlation of 0.73 with GPP estimates from flux towers compared to 0.69 for the FLUXNET-MTE GPP product and 0.68 for the MODIS GPP product across 74 sites from the Tier 1 product. Given that the RSIF product does not use additional weather/climate information unlike these other GPP products, this result is very encouraging and shows that the raw information content within the MODIS shortwave channels provides directly pertinent information regarding GPP.

We further split the temporal correlation into a climatology of the seasonal cycle and the interannual variability. RSIF outperforms other products in terms of seasonal cycle with a mean correlation of 0.83 with

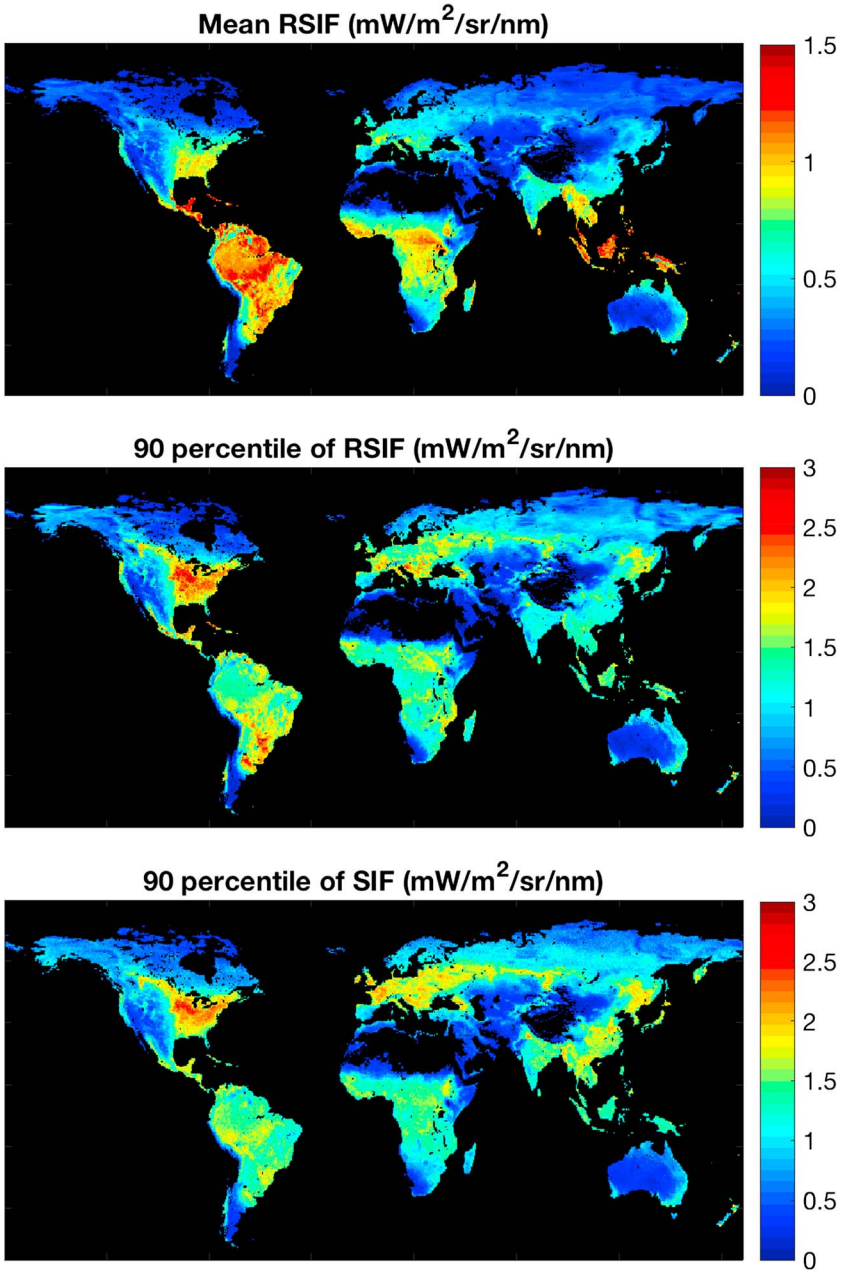

Figure 3. Temporal mean (top) and 90 percentile value (middle) of Reconstructed Solar-Induced Fluorescence (RSIF; $mW^2\ m^{-4}\ sr^{-1}\ nm^{-1}$) and 90 percentile of solar-induced fluorescence (SIF; $mW^2\ m^{-4}\ sr^{-1}\ nm^{-1}$; bottom), emphasizing agricultural regions.

eddy covariance GPP across sites, compared to 0.81 for FLUXNET-MTE and 0.807 for MODIS-GPP. RSIF largely outperforms other products in terms of its interannual correlation with a correlation of 0.29, compared to 0.12 for FLUXNET-MTE and a negative correlation for MODIS-GPP (−0.03). We note that we cannot expect perfect correlation given the inherent noise and data gap filling in eddy covariance data, but this higher correlation is comforting and shows the potential of RSIF to assess interannual variability in GPP.

Finally, global averages and the 90th percentile of RSIF, in order to assess the peak seasonal value, are analyzed (Figure 3). The highest values of annual average RSIF are observed in tropical forests, as expected, with a peak over the Amazon and Maritime Continent, similarly to more complex GPP products using more input sources, such as WECANN (Alemohammad et al., 2017). Dry to wet transitional regions are correctly

characterized, such as in Portugal-Spain, North Africa, the Sahel, the east-west gradient in the continental United States, and in India. Deserts are also correctly characterized with the lowest RSIF values, whereas the original SIF data tend to be noisy over deserts and more erratic due to the inherent noise in the small-amplitude SIF data (Frankenberg et al., 2014; Joiner et al., 2011; Köhler et al., 2014).

The 90th percentile of RSIF (Figure 3, bottom), which shows the seasonal peak of RSIF, exhibits a peak in the Corn Belt region, where corn and soybean are the major crops, with a lower high GPP region in the eastern U.S. forests and along the Mississippi river. These results are similar to recent observations, using carbonyl sulfide (COS) atmospheric inversion, which demonstrates that this region had the highest GPP during the growing season in the continental United States and is believed to be the highest worldwide (Hilton et al., 2017) or using SIF directly (Guanter et al., 2014; Figure 3c) and emphasize that contrary to previous thinking (e.g., Guanter et al., 2014), MODIS channels may have sufficient information to determine high-productivity regions. The second highest region is located in the wet regions of Argentina and Paraguay, where soybean and cotton are the major crops and which have similar reflectance signature as the U.S. crop region (where soybean is one of the major crops). NDVI and EVI, on the other hand, are unable to depict those high-productivity regions (Figure S5) and tend to overestimate the spatial extent of the regions of high productivity to the entire eastern United States and over the entire Amazon. This emphasizes that RSIF is a better GPP product to study seasonal and interannual variability and extremes compared to either NDVI or EVI and that it is better correlated with GPP than the original SIF GOME-2 data.

5. Conclusions

A new proxy for GPP, called RSIF, is developed using a machine learning approach based on MODIS reflectance inputs. The product predicts GOME-2 SIF normalized by clear-sky irradiance, a proxy for $fPAR_{Ch}$, which is then multiplied by an estimate of PAR (BESS) also based on MODIS observations to produce the reconstructed SIF product. This approach performs well across biomes and climates and largely improves the original GOME-2 SIF product. The main reason for this improvement is that the MODIS data are less noisy than the original GOME-2 product and because the far-red GOME-2 SIF seems to largely reflect changes in $APAR_{Ch}$, observable with MODIS, at least at the time and spatial scales considered here (biweekly, 0.5°).

RSIF exhibits strong correlation with GPP eddy covariance observations both on interannual and seasonal time scales and outperforms other global GPP retrievals, which use additional information in their retrieval (e.g., temperature), thus demonstrating that the MODIS information content to retrieve GPP might be higher than previously thought. In addition, MODIS reflectance may carry information on LUE as $V_{c,max}$ correlates with chlorophyll and nitrogen content, both of which modify the spectral reflectance.

In addition to reducing the noise in the original data, RSIF has an advantage in providing a longer-term record (since July 2002) and higher spatial resolution (500 m) than the GOME-2 SIF data so that it can be used to address important questions regarding land use land cover changes or response of GPP to environmental changes. RSIF also does not saturate unlike typical vegetation indices based on MODIS data (e.g., NDVI or EVI).

Findings from the new RSIF product emphasizes the large amount of information available in already existing visible and near-infrared observations to reproduce GPP. The results also open up some questions related to SIF measurements. In particular, an important question is whether SIF remote sensing observations can be used to detect changes in LUE which should vary across plant functional type, phenological stage, or environmental conditions (Heinsch et al., 2006). Given that the RSIF product improves the overall correlations, we may wonder whether RSIF is simply a smoothed version of SIF or whether additional information in SIF is lost in the processed data. We unfortunately do not have a direct way to answer this question; however, the very high correlations in midlatitudes (dominated by fPAR changes) and much lower in tropical rainforest point to the fact that remote sensing observations of far-red SIF may still carry important information on SIF yield or could highlight changes in *fesc*, related to changes in canopy structure (Saleska et al., 2016) or light (Morton & Cook, 2016).

The emission of SIF peaks in the near-red and far-red spectra at 683 and 736 nm, respectively. It has been shown that relative magnitude of the red and far-red fluorescence is sensitive to nitrogen uptake, chlorophyll content, and responses to stress conditions such as low temperature (Agati et al., 1995, 1996; Campbell, Carmichael, & Chai, 2008; Campbell, Middleton, et al., 2008; Campbell et al., 2007; Corp et al., 2003; Rossini et al., 2015). Most of the studies using SIF, including ours, to estimate terrestrial GPP have so far focused

on using the far-red SIF emission. However, the red SIF seems less correlated with $APAR_{Ch}$ (Du et al., 2017), and thus the combination of both red and far-red SIF may have the potential to better constrain the estimates of plants' photosynthetic activity. Recently, a new retrieval of red SIF from GOME-2 instrument was developed (Joiner et al., 2016), which is noisy but could yield important additional information on LUE. New fluorescence observations at higher spatial and temporal resolution and with higher signal-to-noise ratio, such as from OCO-2 (Frankenberg et al., 2014; Li et al., 2017; Schimel et al., 2015), Florescence Explorer (FLEX) (Kraft et al., 2012), TROPOspheric Monitoring Instrument (TROPOMI) (Guanter et al., 2015), or Geostationary Carbon Cycle Observatory (GeoCARB) (O'Brien et al., 2016), might help better retrieve changes in LUE, maybe in combination with this RSIF product. In addition, SIF observations at other spectral frequencies and at higher temporal resolution (diurnal) might yield critical information regarding variations in SIF yield and LUE. The approach presented here could be further improved using hyperspectral data over shorter duration when hyperspectral data are available.

Acknowledgments

This work used eddy covariance data from the FLUXNET 2015 Tier 1 data set acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The FLUXNET eddy covariance data processing and harmonization was carried out by the ICOS Ecosystem Thematic Center, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux, and AsiaFlux offices. The authors would like to thank and acknowledge Joanna Joiner for providing the processed GOME-2 SIF data and Youngryel Ryu for providing the BESS data. The MODIS MYD091A data product was retrieved from the online Data Pool, courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, https://lpdaac.usgs.gov/data_access/data_pool. P. Gentine acknowledges funding from NASA NNH16ZDA001N-AIST, Computational Technologies: "An Assessment of Hybrid Quantum Annealing Approaches for Inferring and Assimilating Satellite Surface Flux Data into Global Land Surface Models," as well as funding from the STR3S project supported by the Belgium Space Agency. The 0.5° RSIF data are available at: https://gentinelab.eee.columbia.edu/content/datasets. The source codes and neural network structure can be downloaded in MATLAB format from here: https://github.com/HamedAlemo/RSIF.

References

Agati, G., Mazzinghi, P., Fusi, F., & Ambrosini, I. (1995). The F685/F730 chlorophyll fluorescence ratio as a tool in plant physiology: Response to physiological and environmental factors. *Journal of Plant Physiology*, *145*(3), 228–238. https://doi.org/10.1016/S0176-1617(11)81882-1

Agati, G., Mazzinghi, P., Lipucci di Paola, M., Fusi, F., & Cecchi, G. (1996). The F685/F730 chlorophyll fluorescence ratio as indicator of chilling stress in plants. *Journal of Plant Physiology*, *148*(3–4), 384–390. https://doi.org/10.1016/S0176-1617(96)80270-7

Aires, F., Prigent, C., Rossow, W. B., & Rothstein, M. (2001). A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations. *Journal of Geophysical Research*, *106*, 14,887–14,907. https://doi.org/10.1029/2001JD900085

Alemohammad, S. H., Fang, B., Konings, A. G., Aires, F., Green, J. K., Kolassa, J., et al. (2017). 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*(18), 4101–4124. https://doi.org/10.5194/bg-14-4101-2017

Anber, U., Gentine, P., Wang, S., & Sobel, A. H. (2015). Fog and rain in the Amazon. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(37), 11,473–11,477. https://doi.org/10.1073/pnas.1505077112

Badgley, G., Field, C. B., & Berry, J. A. (2017). Canopy near-infrared reflectance and terrestrial photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *3*(3), e1602244. https://doi.org/10.1126/sciadv.1602244

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., et al. (2010). Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science*, *329*(5993), 834–838. https://doi.org/10.1126/science.1184984

Campbell, J. E., Carmichael, G. R., & Chai, T. (2008). Photosynthetic control of atmospheric carbonyl sulfide during the growing season. *Science*, *322*(5904), 1081–1085. https://doi.org/10.1126/science.1164498

Campbell, P. K. E., Middleton, E. M., Corp, L. A., & Kim, M. S. (2008). Contribution of chlorophyll fluorescence to the apparent vegetation reflectance. *Science of the Total Environment*, *404*(2–3), 433–439. https://doi.org/10.1016/j.scitotenv.2007.11.004

Campbell, P. K. E., Middleton, E. M., McMurtrey, J. E., Corp, L. A., & Chappelle, E. W. (2007). Assessment of vegetation stress using reflectance or fluorescence measurements. *Journal of Environmental Quality*, *36*(3), 832–814. https://doi.org/10.2134/jeq2005.0396

Corp, L. A., McMurtrey, J. E., & Middleton, E. M. (2003). Fluorescence sensing systems: In vivo detection of biophysical variations in field corn due to nitrogen supply. *Remote Sensing of Environment*, *86*(4), 470–479. https://doi.org/10.1016/S0034-4257(03)00125-1

Du, S., Liu, L., Liu, X., & Hu, J. (2017). Response of canopy solar-induced chlorophyll fluorescence to the absorbed photosynthetically active radiation absorbed by chlorophyll. *Remote Sensing*, *9*(12), 911–919. https://doi.org/10.3390/rs9090911

Duveiller, G., & Cescatti, A. (2016). Spatially downscaling Sun-induced chlorophyll fluorescence leads to an improved temporal correlation with gross primary productivity. *Remote Sensing of Environment*, *182*, 72–89. https://doi.org/10.1016/j.rse.2016.04.027

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., et al. (2011). New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters*, *38*, L17706. https://doi.org/10.1029/2011GL048738

Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Köhler, P., et al. (2014). Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2. *Remote Sensing of Environment*, *147*(C), 1–12. https://doi.org/10.1016/j.rse.2014.02.007

Friend, A. D., Arneth, A., Kiang, N. Y., Lomas, M., Ogée, J., Rödenbeck, C., et al. (2007). FLUXNET and modelling the global carbon cycle. *Global Change Biology*, *13*(3), 610–633. https://doi.org/10.1111/j.1365-2486.2006.01223.x

Gitelson, A. A., Buschmann, C., & Lichtenthaler, H. K. (1999). The chlorophyll fluorescence ratio F735/F700 as an accurate measure of the chlorophyll content in plants. *Remote Sensing of Environment*, *69*(3), 296–302. https://doi.org/10.1016/s0034-4257(99)00023-1

Green, J. K., Konings, A. G., Alemohammad, S. H., Berry, J., Entekhabi, D., Kolassa, J., et al. (2017). Regionally strong feedbacks between the atmosphere and terrestrial biosphere. *Nature Geoscience*, *10*(6), 410–414. https://doi.org/10.1038/ngeo2957

Guan, K., Berry, J. A., Zhang, Y., & Joiner, J. (2016). Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. *Global Change Biology*, *22*(2), 716–726. https://doi.org/10.1111/gcb.13136

Guan, K., Pan, M., Li, H., Wolf, A., Wu, J., Medvigy, D., et al. (2015). Photosynthetic seasonality of global tropical forests constrained by hydroclimate. *Nature Geoscience*, *8*(4), 284–289. https://doi.org/10.1038/ngeo2382

Guanter, L., Aben, I., Tol, P., Krijger, J. M., Hollstein, A., Köhler, P., et al. (2015). Potential of the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence. *Atmospheric Measurement Techniques*, *8*(3), 1337–1352. https://doi.org/10.5194/amt-8-1337-2015

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J. A., et al. (2014). Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(14), E1327–E1333. https://doi.org/10.1073/pnas.1320008111

Heinsch, F. A., Maosheng Zhao, Running, S. W., Kimball, J. S., Nemani, R. R., Davis, K. J., et al. (2006). Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations. *IEEE Transactions on Geoscience and Remote*, *44*(7), 1908–1925. https://doi.org/10.1109/TGRS.2005.853936

Hilton, T. W., Whelan, M. E., Zumkehr, A., Kulkarni, S., Berry, J. A., Baker, I. T., et al. (2017). Peak growing season gross uptake of carbon in North America is largest in the Midwest USA. *Nature Climate Change*, *7*(6), 450–454. https://doi.org/10.1038/nclimate3272

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Huete, A. R., Didan, K., Shimabukuro, Y. E., Ratana, P., Saleska, S. R., Hutyra, L. R., et al. (2006). Amazon rainforests green-up with sunlight in dry season. *Geophysical Research Letters*, *33*, L06405. https://doi.org/10.1029/2005GL025583

Jeong, S.-J., Schimel, D., Frankenberg, C., Drewry, D. T., Fisher, J. B., Verma, M., et al. (2017). Application of satellite solar-induced chlorophyll fluorescence to understanding large-scale variations in vegetation phenology and function over northern high latitude forests. *Remote Sensing of Environment*, *190*, 178–187. https://doi.org/10.1016/j.rse.2016.11.021

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, *112*(10), 3833–3845. https://doi.org/10.1016/j.rse.2008.06.006

Jimenez, C., Clark, D. B., Kolassa, J., Aires, F., & Prigent, C. (2013). A joint analysis of modeled soil moisture fields and satellite observations. *Journal of Geophysical Research: Atmospheres*, *118*, 6771–6782. https://doi.org/10.1002/jgrd.50430

Jimenez, C., Prigent, C., & Aires, F. (2009). Toward an estimation of global land surface heat fluxes from multisatellite observations. *Journal of Geophysical Research*, *114*, D06305. https://doi.org/10.1029/2008JD011392

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., et al. (2013). Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: Methodology, simulations, and application to GOME-2. *Atmospheric Measurement Techniques*, *6*(10), 2803–2823. https://doi.org/10.5194/amt-6-2803-2013

Joiner, J., Vasilkov, A. P., Yoshida, Y., Corp, L. A., & Middleton, E. M. (2011). First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences*, *8*(3), 637–651. https://doi.org/10.5194/bg-8-637-2011

Joiner, J., Yoshida, Y., Guanter, L., & Middleton, E. M. (2016). New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: Simulations and application to GOME-2 and SCIAMACHY. *Atmospheric Measurement Techniques*, *9*(8), 3939–3967. https://doi.org/10.5194/amt-9-3939-2016

Joiner, J., Yoshida, Y., Vasilkov, A. P., Schaefer, K., Jung, M., Guanter, L., et al. (2014). The seasonal cycle of satellite chlorophyll fluorescence observations and its relationship to vegetation phenology and ecosystem atmosphere carbon exchange. *Remote Sensing of Environment*, *152*, 375–391. https://doi.org/10.1016/j.rse.2014.06.022

Jung, M., Reichstein, M., & Bondeau, A. (2009). Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, *6*(10), 2001–2013. https://doi.org/10.5194/bg-6-2001-2009

Köhler, P., Guanter, L., & Joiner, J. (2014). A linear method for the retrieval of Sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data. *Atmospheric Measurement Techniques Discussions*, *7*(12), 12,173–12,217. https://doi.org/10.5194/amtd-7-12173-2014

Koirala, S., Jung, M., Reichstein, M., de Graaf, I. E. M., Camps-Valls, G., Ichii, K., et al. (2017). Global distribution of groundwater-vegetation spatial covariation. *Geophysical Research Letters*, *44*, 4134–4142. https://doi.org/10.1002/2017GL072885

Kolassa, J., Aires, F., Polcher, J., Prigent, C., Jimenez, C., & Pereira, J. M. (2013). Soil moisture retrieval from multi-instrument observations: Information content analysis and retrieval methodology. *Journal of Geophysical Research: Atmospheres*, *118*, 4847–4859. https://doi.org/10.1029/2012JD018150

Kolassa, J., Gentine, P., Prigent, C., & Aires, F. (2016). Soil moisture retrieval from AMSR-E and ASCAT microwave observation synergy. Part 1: Satellite data analysis. *Remote Sensing of Environment*, *173*, 1–14. https://doi.org/10.1016/j.rse.2015.11.011

Kolassa, J., Gentine, P., Prigent, C., Aires, F., & Alemohammad, S. H. (2017). Soil moisture retrieval from AMSR-E and ASCAT microwave observation synergy. Part 2: Product evaluation. *Remote Sensing of Environment*, *195*, 202–217. https://doi.org/10.1016/j.rse.2017.04.020

Kolassa, J., Reichle, R. H., & Draper, C. S. (2017). Merging active and passive microwave observations in soil moisture data assimilation. *Remote Sensing of Environment*, *191*, 117–130. https://doi.org/10.1016/j.rse.2017.01.015

Konings, A. G., Williams, A. P., & Gentine, P. (2017). Sensitivity of grassland productivity to aridity controlled by stomatal and xylem regulation. *Nature Geoscience*, *10*(4), 284–288. https://doi.org/10.1038/ngeo2903

Kraft, S., Del Bello, U., Bouvet, M., Drusch, M., & Moreno, J. (2012). "FLEX: ESA's Earth Explorer 8 candidate mission," IEEE International Geoscience and Remote Sensing Symposium (pp. 7125–7128). Munich. https://doi.org/10.1109/IGARSS.2012.6352020.

Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., et al. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, *2*(12), 831–836. https://doi.org/10.1038/ngeo689

Lee, J.-E., Berry, J. A., van der Tol, C., Yang, X., Guanter, L., Damm, A., et al. (2015). Simulations of chlorophyll fluorescence incorporated into the Community Land Model version 4. *Global Change Biology*, *21*(9), 3469–3477. https://doi.org/10.1111/gcb.12948

Lee, J.-E., Frankenberg, C., van der Tol, C., Berry, J. A., Guanter, L., Boyce, C. K., et al. (2013). Forest productivity and water stress in Amazonia: Observations from GOSAT chlorophyll fluorescence. *Proceedings of the Biological Sciences*, *280*(1761), 20130171. https://doi.org/10.1098/rspb.2013.0171

Li, X., Xiao, J., & He, B. (2017). Chlorophyll fluorescence observed by OCO-2 is strongly related to gross primary productivity estimated from flux towers in temperate forests. *Remote Sensing of Environment*, *204*, 659–671. https://doi.org/10.1016/j.rse.2017.09.034

Liu, L., Guan, L., & Liu, X. (2017). Directly estimating diurnal changes in GPP for C3 and C4 crops using far-red Sun-induced chlorophyll fluorescence. *Agricultural and Forest Meteorology*, *232*, 1–9. https://doi.org/10.1016/j.agrformet.2016.06.014

Long, S. P., & Bernacchi, C. J. (2003). Gas exchange measurements, what can they tell us about the underlying limitations to photosynthesis? Procedures and sources of error. *Journal of Experimental Botany*, *54*(392), 2393–2401. https://doi.org/10.1093/jxb/erg262

McDowell, N. G., Coops, N. C., Beck, P. S. A., Chambers, J. Q., Gangodagamage, C., Hicke, J. A., et al. (2015). Global satellite monitoring of climate-induced vegetation disturbances. *Trends in Plant Science*, *20*(2), 114–123. https://doi.org/10.1016/j.tplants.2014.10.008

Meiyappan, P., & Jain, A. K. (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, *6*(2), 122–139. https://doi.org/10.1007/s11707-012-0314-2

Monteith, J. (1972). Solar radiation and productivity in tropical ecosystems. *The Journal of Applied Ecology*, *9*(3), 747–721. https://doi.org/10.2307/2401901

Monteith, J. (1977). Climate and the efficiency of crop production in Britain. *Philosophical Transactions of the Royal Society of London, Series B*, *281*, 277–294. https://doi.org/10.1098/rstb.1977.0140

Morton, D. C., & Cook, B. D. (2016). Amazon forest structure generates diurnal and seasonal variability in light utilization. *Biogeosciences*, *13*(7), 2195–2206. https://doi.org/10.5194/bg-13-2195-2016

O'Brien, D. M., Polonsky, I. N., Utembe, S. R., & Rayner, P. J. (2016). Potential of a geostationary geoCARB mission to estimate surface emissions of CO_2, CH_4 and CO in a polluted urban environment: Case study Shanghai. *Atmospheric Measurement Techniques*, *9*(9), 4633–4654. https://doi.org/10.5194/amt-9-4633-2016

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., et al. (2017). Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: Algorithms and uncertainty estimation. *Biogeosciences*, *3*(4), 571–583. https://doi.org/10.5194/bg-3-571-2006

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., et al. (2014). Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. *Nature*, *509*(7502), 600–603. https://doi.org/10.1038/nature13376

Rossini, M.d., Nedbal, L., Guanter, L., Ač, A., Alonso, L., Burkart, A., et al. (2015). Red and far red Sun-induced chlorophyll fluorescence as a measure of plant photosynthesis. *Geophysical Research Letters*, *42*, 1632–1639. https://doi.org/10.1002/2014GL062943

Running, S. W. (2004). A continuous satellite-derived measure of global terrestrial primary production. *Bioscience*, *54*(6), 547–515. https://doi.org/10.1641/0006-3568(2004)054%5B0547:acsmog%5D2.0.co;2

Ryu, Y., Jiang, C., Kobayashi, H., & Detto, M. (2017). MODIS-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000. *Remote Sensing of Environment*, *204*, 812–825. https://doi.org/10.1016/j.rse.2017.09.021

Saleska, S. R., Wu, J., Guan, K., Araujo, A. C., & Huete, A. (2016). Dry-season greening of Amazon forests. *Nature*, *531*(7594), E4–E5. https://doi.org/10.1038/nature16457

Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., et al. (2015). Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology*, *21*(5), 1762–1776. https://doi.org/10.1111/gcb.12822

Sellers, P. J. (2007). Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, *6*(8), 1335–1372. https://doi.org/10.1080/01431168508948283

Song, C., Dannenberg, M. P., & Hwang, T. (2013). Optical remote sensing of terrestrial ecosystem primary productivity. *Progress in Physical Geography*, *37*(6), 834–854. https://doi.org/10.1177/0309133313507944

Sun, Y., Fu, R., Dickinson, R., Joiner, J., Frankenberg, C., Gu, L., et al. (2015). Drought onset mechanisms revealed by satellite solar-induced chlorophyll fluorescence: Insights from two contrasting extreme events. *Journal of Geophysical Research: Biogeosciences*, *120*, 2427–2440. https://doi.org/10.1002/2015JG003150

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

van der Tol, C., Verhoef, W., & Rosema, A. (2009). A model for chlorophyll fluorescence and photosynthesis at leaf scale. *Agricultural and Forest Meteorology*, *149*(1), 96–105. https://doi.org/10.1016/j.agrformet.2008.07.007

Wan, L., Zeiler, M., Zhang, S., Le Cun, Y., & Fergus, R. (2013). Regularization of neural networks using dropconnect. In *International Conference on Machine Learning* (pp. 1058–1066).

Xu, L., Saatchi, S. S., Yang, Y., Myneni, R. B., Frankenberg, C., Chowdhury, D., & Bi, J. (2015). Satellite observation of tropical forest seasonality: Spatial patterns of carbon exchange in Amazonia. *Environmental Research Letters*, *10*(8), 1–10. https://doi.org/10.1088/1748-9326/10/8/084005

Yang, X., Tang, J., Mustard, J. F., Lee, J.-E., Rossini, M., Joiner, J., et al. (2015). Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest. *Geophysical Research Letters*, *42*, 2977–2987. https://doi.org/10.1002/2015GL063201

Zhang, Y., Xiao, X., Guanter, L., Zhou, S., Ciais, P., Joiner, J., et al. (2016). Precipitation and carbon-water coupling jointly control the interannual variability of global land gross primary production. *Scientific Reports*, *6*(1), 39748. https://doi.org/10.1038/srep39748

Zhang, Y., Xiao, X., Jin, C., Dong, J., Zhou, S., Wagle, P., et al. (2016). Consistency between Sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America. *Remote Sensing of Environment*, *183*, 154–169. https://doi.org/10.1016/j.rse.2016.05.015

Zhao, M., Heinsch, F. A., Nemani, R. R., & Running, S. W. (2005). Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sensing of Environment*, *95*(2), 164–176. https://doi.org/10.1016/j.rse.2004.12.011

Zhao, M., Running, S. W., & Nemani, R. R. (2006). Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses. *Journal of Geophysical Research*, *111*, G01002. https://doi.org/10.1029/2004JG000004

Zhou, S., Yu, B., Huang, Y., & Wang, G. (2015). Daily underlying water use efficiency for AmeriFlux sites. *Journal of Geophysical Research: Biogeosciences*, *120*, 887–902. https://doi.org/10.1002/2015JG002947