AGU100 ADVANCING EARTH AND SPACE SCIENCE

JGR Biogeosciences

RESEARCH ARTICLE
10.1029/2018JG004938

Key Points:
- The differences between three space-borne SIF products (one from OCO-2 and two from GOME-2) are quantified at the biome level
- While two SIF products show consistent magnitude and seasonal amplitude, the third product is twice as large for most biomes
- Using a toy assimilation framework we discuss the impact of these differences on the optimization of process-based parameters related to GPP

Correspondence to:
C. Bacour,
cedric.bacour@noveltis.fr

Citation:
Bacour, C., Maignan, F., Peylin, P., MacBean, N., Bastrikov, V., Joiner, J., et al. (2019). Differences between OCO-2 and GOME-2 SIF products from a model-data fusion perspective. *Journal of Geophysical Research: Biogeosciences*, *124*, 3143–3157. https://doi.org/10.1029/2018JG004938

Received 16 NOV 2018
Accepted 11 SEP 2019
Accepted article online 11 OCT 2019
Published online 29 OCT 2019

Author Contributions:
Conceptualization: C. Bacour
Formal analysis: C. Bacour, F. Maignan, P. Peylin, N. MacBean, J. Joiner, P. Köhler, L. Guanter
Investigation: C. Bacour
Methodology: C. Bacour, P. Peylin, N. MacBean
Software: C. Bacour, V. Bastrikov
Supervision: C. Bacour, F. Maignan
Validation: C. Bacour, F. Maignan, P. Peylin, J. Joiner
Visualization: C. Bacour
Writing - original draft: C. Bacour, F. Maignan, P. Peylin, N. MacBean
Writing – review & editing: C. Bacour, F. Maignan, P. Peylin, N. MacBean, J. Joiner, P. Köhler, L. Guanter, C. Frankenberg

Differences Between OCO-2 and GOME-2 SIF Products From a Model-Data Fusion Perspective

C. Bacour[1], F. Maignan[2], P. Peylin[2], N. MacBean[3], V. Bastrikov[2], J. Joiner[4], P. Köhler[5], L. Guanter[6], and C. Frankenberg[5,7]

[1]NOVELTIS, Labège, France, [2]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France, [3]Department of Geography, Indiana University, Bloomington, IN, USA, [4]National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC), Greenbelt, MD, USA, [5]Divison of Geological and Planetray Sciences, California Institute of Technology, Pasadena, CA, USA, [6]Centro de Tecnologías Físicas, , Universitat Politècnica de València, València, Spain, [7]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Abstract Space-borne retrievals of solar-induced chlorophyll fluorescence (SIF) over land surfaces have recently become a resource for studying and quantifying the broad scale dynamics of gross carbon uptake (gross primary productivity—GPP) across ecosystems. To prepare for the assimilation of SIF data in terrestrial biosphere models, we examine how differences between SIF products (due to differences in acquisition characteristics and processing chain) may affect the optimization of model parameters and the resultant GPP estimate. We compare recent daily mean SIF products (one from the Orbiting Carbon Observatory-2 [OCO-2] and two from the Global Ozone Monitoring Experiment–2 [GOME-2], GlobFluo [GF] and NASA-v28 [N28], missions), averaged at 0.5° × 0.5° spatial resolution and 16-day temporal resolution, at the biome level. Phase differences between these products are relatively small. A first-order correction of the difference in spectral sampling between the two instruments shows that OCO-2 and N28 are consistent in terms of magnitude and amplitude, while GF is twice as large as the others. Using a bias-blind toy data assimilation framework, we analyze how biases between SIF products, and between model and products, can be partially alleviated by optimizing the slope and intercept parameters of a linear GPP-SIF operator. As observation biases can transfer to biases in other optimized process-based parameters and to modeled carbon fluxes— thereby resulting in unidentified inaccurate parameter values—we argue that potential SIF biases should be treated cautiously in real-world experiments in order to achieve realistic and reliable future simulations.

1. Introduction

Large uncertainties remain in our modeling of the global photosynthesis flux, which is the main source of carbon uptake by land ecosystems (Le Quéré et al., 2017). Simulated fields of plant gross carbon uptake (gross primary productivity – GPP) by dynamic global vegetation models (DGVMs) vary significantly in magnitude and exhibit different spatial and temporal gradients, including different inter-annual variability (Anav et al., 2015). Bolstered by the evidenced linear relationship between remotely-sensed solar-induced chlorophyll fluorescence (SIF) and GPP at broad spatial and temporal scales (Guanter et al., 2014; Zhang et al., 2016), it is anticipated that satellite SIF products will provide a significant constraint (reduction in uncertainty) on global GPP estimates (Parazoo et al., 2014).

SIF and GPP are functionally related through the absorbed photosynthetically active radiation (APAR); as such, SIF is often considered as a proxy of GPP. Indeed, at large spatial and temporal scales, they exhibit a strong linear relationship, with scaling factors between the two depending on biome type and climate (Guanter et al., 2012; Sun et al., 2017). In addition, it has been shown that the seasonal cycle of GPP is better reproduced by space-borne SIF observations than by classical remotely sensed estimates of vegetation greenness such as the fraction of APAR, the normalized difference vegetation index or the enhanced vegetation index (Joiner et al., 2014; Jeong et al. 2017; Walther et al., 2016). Nevertheless, although SIF and GPP are both largely driven by APAR, they are also both separately modulated by other processes. For example, there are additional processes controlling the fluorescence emission at the leaf level due to the dissipation of energy in the photosystems. SIF is in principle only a proxy for the light reaction, whereas the electrons needed for carboxylation differ between the photosynthetic pathways which is the reason why we observe different slopes

between C3 and C4 species (Wood et al., 2017). At the canopy level also, the fluorescence flux is partly scattered and reabsorbed. These interactions make for a complex relationship between instantaneous SIF and photosynthesis. Accounting for and formalizing within a model that complex functional link between SIF and photosynthesis opens up challenging opportunities for improving our understanding of the dynamics of ecosystems' photosynthetic activity.

Since the very first remotely sensed estimates of SIF from the land surface with Medium-Resolution Imaging Spectrometer (Guanter et al., 2007), global SIF estimates are now routinely produced from observations of space-borne spectrometers that have sufficient spectral resolution and signal to noise ratio in the red and far-red regions of the solar spectrum to evaluate the in-filling of solar Fraunhofer lines by SIF. This is the case for the OCO-2 (Orbiting Carbon Observatory-2), GOME-2 (Global Ozone Monitoring Experiment–2), SCIAMACHY (SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY), GOSAT (Greenhouse Gases Observing Satellite) missions (Joiner et al., 2011; Frankenberg et al., 2011; Guanter et al., 2012; Joiner et al., 2013; Frankenberg et al., 2014; Köhler et al., 2015; Khosravi et al., 2015; Somkuti et al., 2015; Sanders et al., 2016), and now TROPOMI (TROPOspheric Monitoring Instrument) (Köhler, Frankenberg, et al., 2018). Differences between SIF products derived from different instruments are to be expected depending on three main factors: (1) the evaluated wavelengths (SIF emission has a spectral signature with maxima around 685 and 740 nm); (2) their own observation characteristics (time of satellite overpass, spatial resolution, revisit frequency, spectral resolution, observing geometry such as view angles, etc.); and (3) the whole processing chain for retrieving SIF estimates. For instance, a scaling factor of about 1.5 is expected between SIF products estimated at 757 nm (GOSAT and OCO-2) and those estimated at 740 nm (GOME-2, SCIAMACHY) (Joiner et al., 2013; Köhler, Frankenberg, et al., 2018), which is related to the spectral variation of fluorescence emission intensity. Nevertheless, as highlighted by Köhler et al. (2015) and recently by Parazoo et al. (2019), significant differences in absolute SIF values may arise for products derived from the same set of raw observations (GOME-2) due to the use of different retrieval algorithms.

The large spread in GPP simulated by DGVMs results from structural differences in both the process representation in the model and from the different fixed values assigned to the associated parameters (Anav et al., 2015). Several modeling groups are currently investigating the use of gridded space-borne SIF products for correcting their DGVM GPP predictions through statistical optimization (or data assimilation [DA]) approaches. The resulting optimized model parameters are expected to provide a better characterization of the spatial distribution of GPP as well as a better representation of GPP temporal phase. These studies have used a range of methods to link SIF and GPP from assuming a linear SIF-GPP relationship (MacBean et al., 2018; Parazoo et al., 2014) to more complex process-based SIF-GPP models (Koffi et al., 2015; Lee et al., 2015; Norton et al., 2018). Discrepancies between data sets are highly problematic in a DA system because biases will be aliased onto model parameters and will therefore impact the calibrated GPP. Exactly how biases will impact model parameters and GPP depends not only on the functional link between the SIF observation operator (that relates the model variables and parameters to observed SIF) and the GPP model but also on the characteristics of the assimilated SIF data (magnitude and seasonality). Therefore, we should attempt to thoroughly investigate and identify the cause and impact of these SIF data discrepancies on calibrated GPP estimates.

In this context, this study aims to assess the potential impacts of the differences between several remotely sensed SIF data sets from a model-data assimilation perspective. The assessment is performed at the biome scale because DGVMs classically group plants species with a similar eco-physiological behavior into broad plant functional types [PFTs] (Cramer, 1997; Prentice et al., 1992); hence, data assimilation experiments aim to optimize PFT-dependent parameters. First, we quantify the similarities and discrepancies between some of the most commonly used gridded satellite SIF products that have sufficient temporal and spatial coverage at $0.5° \times 0.5°$ (which is a typical resolution of a global DGVM simulation). These are: (1) OCO-2 products (Sun et al., 2018); (2) GOME-2 GlobFluo (Köhler et al., 2015); and (3) GOME-2 NASA (Joiner et al., 2013) retrievals. We evaluate both the absolute value and seasonality of daily mean SIF. The analysis focuses on year 2015, and the temporal evolution of the differences between products is assessed over the 2015–2018 period. Second, a *bias-blind* assimilation toy experiment is performed to evaluate the potential impacts of the differences between SIF products on calibrated GPP estimates. Bias-blind (Dee, 2005) means here that no attempt has been made to correct any systematic error in either the observations, the model, or both. The assimilation relies on a linear regression between SIF and GPP simulated by the Organizing Carbon and

Hydrology In Dynamic Ecosystems [ORCHIDEE] DGVM (as in MacBean et al., 2018, but without optimization of the ORCHIDEE parameters). The specific questions that are addressed are thus:

1. What are the main differences in terms of phase, amplitude, and magnitude of the seasonal cycle between the selected gridded SIF products?
2. How do these differences vary between PFTs?
3. What is the potential impact of these differences on the parameters and on the optimized GPP within an optimization scheme based on a linear GPP-SIF relationship?

2. Data sets

2.1. OCO-2

The OCO-2 instrument has been providing high spatial resolution SIF estimates (1.3×2.25 km^2) since its launch in July 2014. The estimation relies on the evaluation of the in-filling of solar Fraunhofer lines at 757 and 770.1 nm (Frankenberg et al., 2014; Sun et al., 2018) surrounding the O_2 A-band. The swath width of 10 km results in a revisit frequency of 16 days but also limits a global coverage (Köhler, Guanter, et al., 2018; Sun et al., 2018). We use the OCO-2 B8100 SIF data available at ftp://fluo.gps.caltech.edu/data/OCO2/sif_lite_B8100/ which has shown a very good consistency with airborne SIF measurements (Sun et al., 2017). The theoretical systematic error is below 0.05 W·m^{-2}·sr^{-1}·μm^{-1} (Frankenberg et al., 2014). The overpass time of OCO-2 is at 13:30 local solar time. In the following, we use the raw, non-gridded, SIF products estimated at 757 nm and acquired at nadir.

2.2. GOME-2

The estimation of terrestrial SIF using the GOME-2 instrument is possible owing to the very high signal to noise ratio and broad spectral coverage of its fourth channel (593–790 nm). This compensates for the lower spectral resolution with respect to OCO-2. Compared to OCO-2, the spatial resolution of GOME-2 is much broader (80×40 km^2/40×40 km^2 before/after July 2013). The large swath width of 1,920/960 km, permits a high revisit frequency (1.5/3 days) and spatial contiguous measurements as opposed to OCO-2. Two widely used GOME-2 SIF data sets have been chosen for this study: We compare the GlobFluo products (Köhler et al., 2015; referred to as GF in the following) and the release Number 28 of the NASA SIF products derived using the Joiner et al. (2013) approach (referred to as N28). They are respectively available at ftp://ftp.gfz-potsdam.de/home/mefe/GlobFluo/GOME2/ and at https://avdc.gsfc.nasa.gov/pub/data/satellite/MetOp/GOME_F/v28/. The two approaches enable estimating SIF at 740 nm. The equator overpass time of the instrument is 9:30 local solar time.

2.2.1. GOME-2-N28

The retrieval approach of GOME-2-N28 uses a data driven principal component [PC] analysis (Joiner et al., 2013) to model atmospheric absorption (mainly due to water vapor) within the spectral fitting window of 734–758 nm and reports a value at 740 nm (Joiner et al., 2016). The PCs are derived from fluorescence-free observations over ocean (both clear and cloudy scenes) and the Sahara desert. The retrieval of SIF relies on the filling-in of solar Fraunhofer features from SIF emission. A main difference between the two GOME-2 retrievals is that GOME-2-N28 makes use of a narrower fitting window that eliminates larger H_2O absorption features, as documented in Joiner et al. (2016). Another difference is the way in which cloudy scenes are filtered. In GOME-2-N28, an effective cloud fraction is determined from the derived surface reflectivity and climatological 16-day black-sky land surface albedo Aqua MODIS products at 656 nm (MOD43B3; Lucht et al., 2000). GOME-2-N28 also employs quality control to remove erroneous radiances and retrievals such as spikes that occur within the South Atlantic anomaly region over South America and a bias correction scheme (Joiner et al., 2016) to account for systematic errors identified by Köhler et al. (2015).

2.2.2. GOME-2-GF

The GOME-2-GF retrieval described by Köhler et al. (2015) uses a similar principal component analysis approach as GOME-2-N28. Differences lie in 1) the fitting window wavelength range used, 2) the derivation of PCs, 3) the estimation of the atmospheric transmittance and its representation in the forward model, and 4) the automated selection of included PCs. It essentially models the radiance spectrum in the 720–758 nm interval and provides a value for the fluorescence emission spectrum at 740 nm with a single-measurement random error about 0.5 W·m^{-2}·sr^{-1}·μm^{-1}. Modeled spectra are compared to the original measurements to exclude corrupted retrievals. This is done by a radiance residual check (residual sum of squares is limited

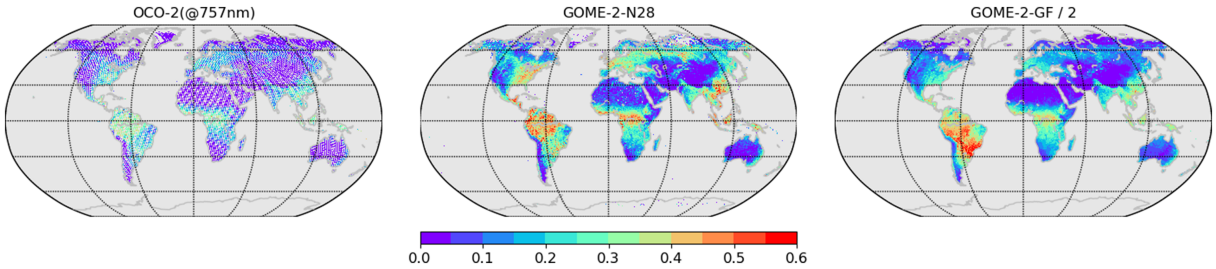

Figure 1. Global maps of the yearly average of the daily mean SIF products (in $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$) for year 2015 derived from OCO-2(@757 nm), GOME-2-N28, and GOME-2-GF. Note that GOME-2-GF solar-induced chlorophyll fluorescence values are divided by two (in order to use the same color scale for the three maps).

to 2 ($W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$)^2). Effective cloud fractions from the Fast Retrieval Scheme for Clouds from the O_2 A-band [FRESCO] (Wang et al., 2008) are used for cloud screening. Furthermore, the soundings were filtered for observation times between 8 a.m. and 2 p.m. mean solar time.

2.3. Raw Data Processing

In order to account for GOME-2 observations close to OCO-2 nadir observation geometry, only observations with a view zenith angle lower than 40° were considered. For the two GOME-2 data sets, we used only the pixels with a cloud fraction below 20%. For the GOME-2-N28, only the data with the best quality flag (2) were considered. For the three SIF products, we used the raw, non-gridded, data for years 2015 to 2018. They were binned every 16 days on a common grid at a 0.5° × 0.5° spatial resolution, keeping only grid cells with at least three individual observations for each aggregation period. In order to correct for the impacts of the differences in the overpass time of the two instruments (Köhler et al., 2015), we normalized the SIF products by a daily correction factor that provides an approximation of the daily mean value (Frankenberg et al., 2011; Sun et al., 2018). The daily correction factor was determined using the same approach for all three processed SIF products—hence ensuring a consistent cross-comparison—and differs only by 5% in average from the one provided in OCO-2 monthly data. Given the differences in spectral resolution between OCO-2 and GOME-2, a first-order spectral normalization of the OCO-2 products was performed to facilitate the intercomparison: The 16-day gridded OCO-2 SIF data were scaled to the 740 nm reference wavelength by applying a multiplicative factor of 1.5 (similar to Parazoo et al., 2019). The two OCO-2 estimates are further referred to as OCO-2(@757 nm) and OCO-2(@740 nm).

The global maps of the average daily mean SIF products (yearly average values for 2015) derived from the three data sets already reveal differences between them (Figure 1). While the mean SIF levels of OCO-2 and GOME-2-N28 are comparable, in spite of their differences in terms of spectral resolutions and fitting windows, GOME-2-GF provides about twice higher values. Differences in SIF spatial gradients are also noticeable, more particularly in South America. GOME-GF shows high SIF values spreading over a large area east of the Andes and south of the Amazon basin, covering Bolivia, Paraguay, and most of Brazil (except the Northeast Region, which is more arid), possibly an artifact from the South Atlantic Anomaly (Köhler et al., 2015); OCO-2 and GOME-2-N28 show a more pronounced latitudinal discontinuity with high SIF values mainly within the Amazon forest.

These large-scale differences are partly attributed to differences in instrumental characteristics between OCO-2 and GOME-2 (Table 1), which result in contrasted temporal and spatial coverages (Figure 2). The average number of individual observations per 16 days and per 0.5° × 0.5° grid cell between 60°S and 60°N latitudes is about 23 for OCO-2, 5 for GOME-2-GF, and 3 for GOME-2-N28 (Figure 2a). The higher observation density obtained with OCO-2 within a 16-day window is due to its higher spatial resolution. An opposite pattern is obtained when comparing the yearly average number of 16-day observations per grid cell (Figure 2b): It is larger for GOME-2-GF (16) than for GOME-2-N28 (12) and OCO-2 (5). Also, while the GOME-2 products have an almost global coverage, the higher spatial resolution of OCO-2 combined with the narrow swath width and the alternation between nadir and glint modes lead to a sparser coverage of the globe at 0.5°: About one pixel in two in longitude is not observed by the instrument in the nadir products. The different spatio-temporal coverages between the two GOME-2 products (over South America in particular) is a consequence of different cloud screening approaches, which is less selective for GOME-2-GF.

Table 1
Main Instrumental Characteristics of OCO-2 and GOME-2 SIF Observations for Year 2015

	OCO-2	GOME-2
Spatial resolution	1.3 × 2.25 km^2	40 × 40 km^2
Revisit frequency	16 days	3 days
Overpass time (Local Solar Time)	13:30	9:30
Orbital period	98.8 min	101 min
Fitting window	Micro window at 757 nm	720/734–758 nm centered at 740 nm
Spectral resolution	~0.05 nm	~0.5 nm

For each data set, we assessed a gross estimation of the retrieval error (assumed purely random) for the 16-day and 0.5° × 0.5° binned products depending on their theoretical retrieval errors. We accounted for the decrease of SIF (and hence the corresponding error) with the daily normalization factor: The reduction of SIF is on average of 0.32 for OCO-2 and 0.4 for GOME-2. The OCO-2 theoretical retrieval error is between 0.3 and 0.5 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$ (Frankenberg et al., 2014). For purely random errors, the theoretical retrieval error decreases as 1/sqrt(N) with averaging, with N observations in 16-day and 0.5° × 0.5° bins. Considering values of N corresponding to the first and last quartiles of the distributions derived from the processing performed in this study, the corresponding random error for the normalized OCO-2(@757nm) SIF products in the 16-day 0.5° bins lies approximately between 0.016 and 0.045 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$. Regarding the two GOME-2 products, the random error for monthly products at 0.5° varies between 0.1 and 0.4 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$ for NASA N26 (Joiner et al., 2014) and is comprised typically between 0.05 and 0.2 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$ for GlobFluo (Köhler et al., 2015). For our 16-day processed products, considering that there are about twice less individual observations within a 16-day composite than at a monthly resolution, these values roughly translate (by applying a correction factor of sqrt(2)) into random errors between 0.06 and 0.23 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$ for GOME-2-N28 (assuming similar error statistics between the N26 and N28 versions) and between 0.03 and 0.11 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$ for GOME-2-GF.

3. Methods

3.1. Pixel Selection

As previous studies showed that the SIF-GPP relationship is dependent upon biome and climate (Guanter et al., 2012), we analyzed the coherence between OCO-2 and GOME-2 SIF products using a vegetation map that discretizes the Earth surface into 14 vegetated PFTs and bare soil. The comparison focused on the 14 natural and agricultural PFTs. The vegetation map was derived from the ESA CCI Land Cover products according to the Poulter et al. (2015) methodology and is currently used with the ORCHIDEE DGVM.

The analysis was not performed globally but on selected 0.5° × 0.5° pixels for each PFT. The selected pixels satisfy prescribed selection criteria, detailed hereafter and determined following a trial and error

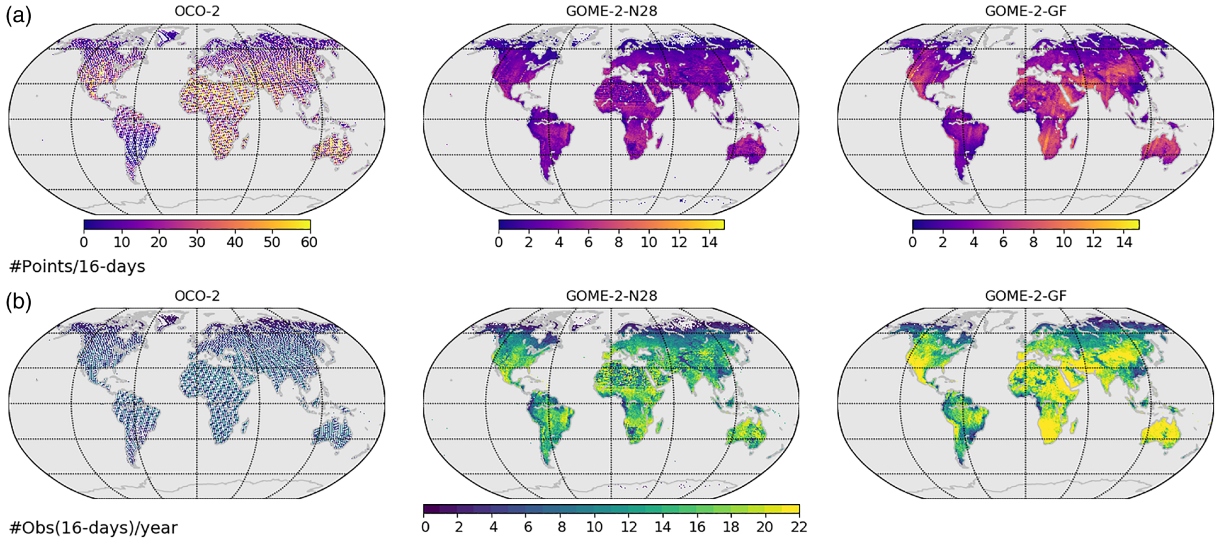

Figure 2. For the three SIF data sets and year 2015, global maps (land only) of (a) the average number of individual observations within a 16-day period; (b) the total number of 16-day products at 0.5° × 0.5°.

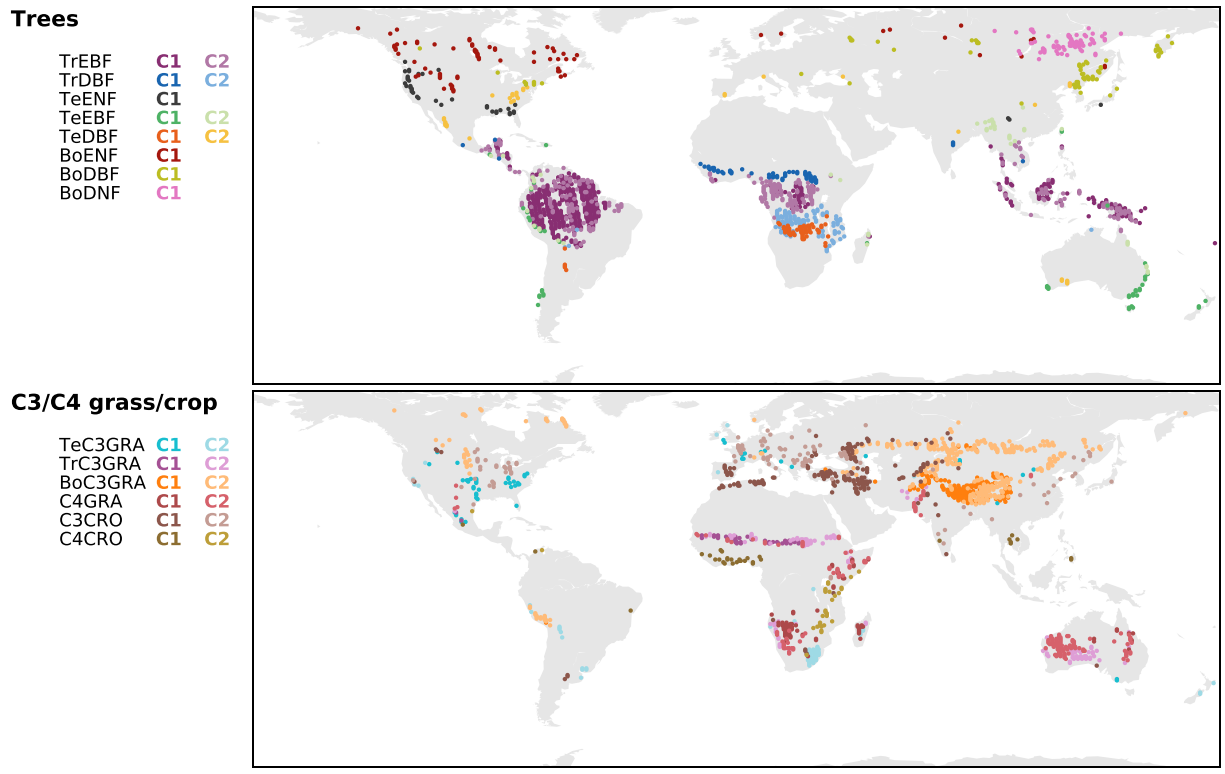

Figure 3. Location of the considered pixels, in relation to their respective plant functional type cluster (a maximum of two clusters per PFT is considered).

process, so as to ensure a sufficient PFT homogeneity of the pixels and a sufficient temporal sampling. Only pixels with (1) a fraction of the dominant PFT greater than 50% and (2) at least six 16-day observations for year 2015, were retained. Larger values of the PFT fraction would have strongly reduced the number of pixels selected for analysis for some PFTs (in particular for temperate and boreal PFTs that are rather mixed with other vegetation types at 0.5°). The number of selected pixels strongly varies with PFT (as seen in Figure 5).

Finally, in order to account for the diversity within each PFT in terms of climate, physiological, and structural characteristics, which would induce different SIF responses, each PFT has been discretized into a given number of clusters. The number of clusters varies for each PFT. This number has been set empirically, based on the analysis of the dendrograms inferred from both the temporal series of SIF retrieved by OCO-2 for year 2015 and the determination of cut-off thresholds (adapted depending on the number of pixels selected for each PFT), leading to group pixels with similar temporal profiles. Each PFT can be divided between one and up to four clusters, which results in a total number of 32 clusters. The clustering was performed with the Python library scipy.cluster.hierarchy (https://docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html), using the Ward variance minimization algorithm. The principle of the clustering approach is to minimize the variance within each cluster while maximizing the variance between clusters (for more details, see tutorial https://joernhees.de/blog/2015/08/26/scipy-hierarchical-clustering-and-dendrogram-tutorial/). Each selected pixel was associated to one of the clusters, and the first analysis of the SIF differences was performed at the scale of the PFT clusters. The spatial distribution of these clusters is shown in Figure 3 (for clarity only up to two clusters per PFT are represented).

3.2. Quantification of the Differences Between SIF Products

In order to quantify the differences between the SIF products in terms of bias, variance, and phasing, we determined the mean square deviation (MSD) between each pair of them and decomposed it based on the approach of Kobayashi and Salam (2000):

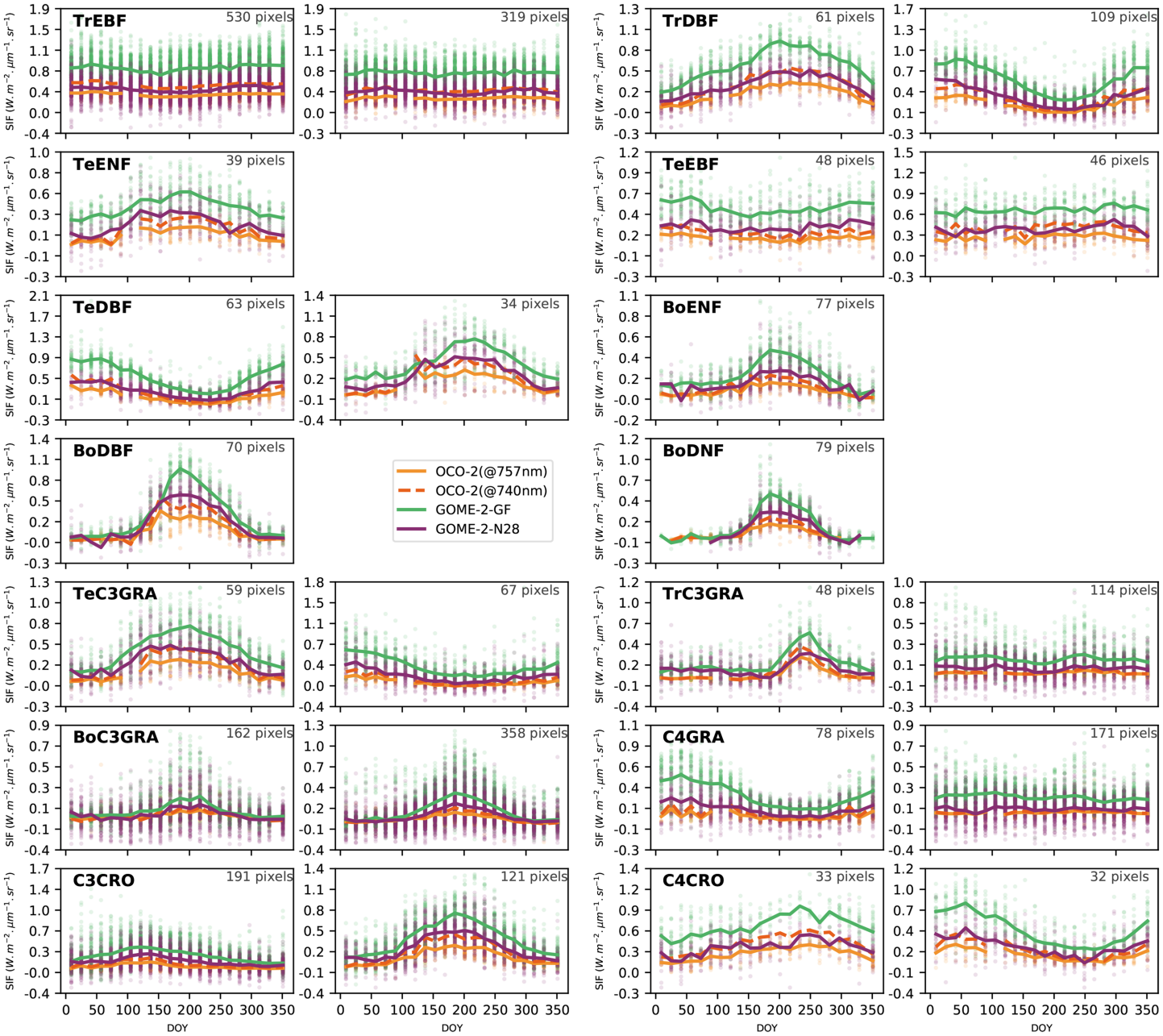

Figure 4. Examples of temporal variations of the SIF products for two clusters of each plant functional type (year 2015). The lines correspond to the means of the corresponding cluster and the dots to the actual products, with a different color for each product. For OCO-2, the time series are shown for the raw (@757 nm) and spectrally scaled (@740 nm) data. The number of pixels considered in each cluster is indicated. The PFTs are tropical evergreen broadleaf (TrEBF), tropical deciduous broadleaf (TrDBF), temperate evergreen needleleaf (TeENF), temperate evergreen broadleaf (TeEBF), temperate deciduous broadleaf (TeDBF), boreal evergreen needleleaf (BorENF), boreal deciduous broadleaf (BorDBF), boreal deciduous needleleaf (BorDNF), forests, C3 grasslands (temperate—TeC3GRA, tropical—TroC3GRA, and boreal—BorC3GRA) and C4 grasslands (C4GRA), and C3 (C3CRO) and C4 (C4CRO) crops.

$$MSD = \left(\overline{SIF_x} - \overline{SIF_y}\right)^2 + \left(\sigma_{SIF_x} - \sigma_{SIF_y}\right)^2 + 2\left(\sigma_{SIF_x}.\sigma_{SIF_y}\right)(1-R). \quad (1)$$

The first term characterizes the bias between two SIF data sets; the second term gauges their differences in terms of variability (i.e., the magnitude of the observed fluctuations); the third one mostly expresses the lack of correlation (R) between them, which can be considered as an estimator of their disagreement in terms of phasing (temporal variations; although not fully independent from the second term; Gauch et al., 2003).

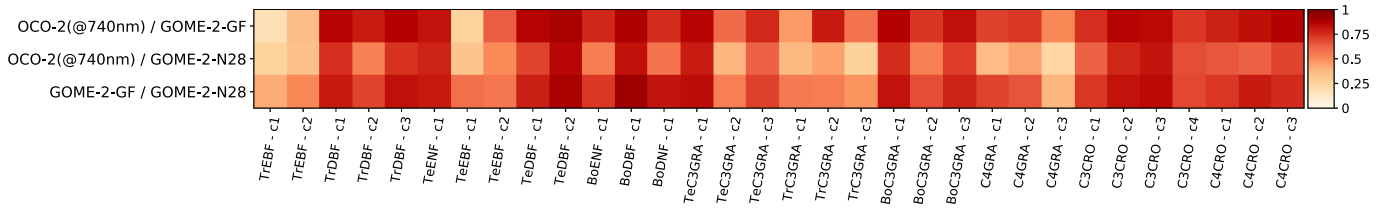

Figure 5. Correlation coefficients between each pair of products at 740 nm estimated for each cluster (year 2015).

3.3. Toy Data Assimilation Experiment

We performed a bias-blind (Dee, 2005) data assimilation to assess how differences in SIF products could impact the optimized GPP that would be obtain from a data assimilation experiment with the ORCHIDEE DGVM (Krinner et al., 2005). As in most data assimilation systems, the errors in either the observations or the model are assumed to be purely random and hence unbiased. Our approach should be seen as a "toy model" experiment to provide qualitative information and specific recommendations for data assimilation experiment. Following the assimilation framework of MacBean et al. (2018), we assumed a linear relationship between SIF and GPP based on the studies of Yang et al. (2015), Guanter et al. (2014), Zhang et al. (2016), and Sun et al. (2017). Assuming such linear SIF-GPP relationship and taking the GPP from the ORCHIDEE model, we estimated the slope and offset parameters for each PFT, a_{PFT} and b_{PFT} (see equation below). Their estimation is performed for the three global scale SIF products (considering all grid cells common to the three data sets) as in MacBean et al. (2018), but without optimizing model parameters that control GPP (e.g., related to photosynthesis and phenology). The linear regression between the SIF observations and the SIF ORCHIDEE simulations was determined by the following equation:

$$SIF = \sum_{PFT} a_{PFT}.GPP_{PFT} + b_{PFT}. \quad (2)$$

The prior values for a and b (0.25 for *both*) and their range of variation ([−0.5; 1] for a and [−1.5; 2.5] for b) were set the same for all PFTs. The prior modeled GPP was determined following a classical simulation setup (spin-up simulation followed by a transient simulation up to the observation period, accounting for changing climate, and the yearly increase of the global atmospheric CO_2 concentration) since 1901. We used six hourly CRU-NCEP v7 meteorological forcing (available at https://vesg.ipsl.upmc.fr/thredds/catalog/store/p529viov/cruncep/V7_1901_2015/catalog.html). In this toy data assimilation experiment, only the slope a and offset b parameters of the SIF-GPP relationship are optimized in order to evaluate how (1) their estimates and (2) the corresponding optimized SIF simulations, differ depending on the "assimilated" SIF products. We then discuss the potential impact of SIF product differences in the context of real data assimilation systems (i.e., optimizing parameter controlling GPP) using the result of the "toy model" experiment.

4. Results

4.1. Seasonal Cycle Comparison Between SIF Products

The analysis of the differences between the SIF products focuses mostly on year 2015 and relies on the following figures: Figure 4 shows the SIF seasonal cycles for one or two clusters of each PFT. The clusters were chosen to illustrate the similarities or discrepancies between SIF data; for some PFTs only one cluster has been used because too few pixels were available that met the selection criteria (see Figure 3). Figure 5 shows the correlation coefficients calculated between each pair of products to evaluate the coherence between the 16-day mean temporal variations. The mean square deviation decomposition performed between each pair of products over the various PFTs is presented in Figure 6 to identify the main causes of the discrepancies between the SIF data sets.

4.1.1. Mean Bias

Figure 4 highlights large differences relative to the mean magnitude of the SIF products for all PFTs/clusters. GOME-2-GF exhibits a higher SIF magnitude than GOME-2-N28 and OCO-2 (regarding both the "raw" data at 757 nm and the rescaled ones at 740 nm), which are close to each other. Note that, as expected, the spectral normalization of OCO-2 SIF reduces the mismatch with the GOME-2 products. The "overestimation" of

Figure 6. Mean square deviation decomposition into bias (purple), variance (green), and phase (orange) between each pair of products by plant functional type (GoGF = GOME-2-GF; GoN28 = GOME-2-N28) for year 2015. The square of the bias and of the variance are displayed. In the analyses involving OCO-2, the quantities are calculated for the "raw" (@757 nm, plain bars) and the spectrally normalized (@740 nm, light bars) OCO-2 SIF products.

GOME-2-GF is more pronounced for tropical (TrEBF, TrDBF) and temperate (TeENF, TeEBF, TeDBF) forests and the C3/C4 crops. Indeed, the bias (or mean difference following the definition of Kobayashi and Salem, 2000) in magnitude is the dominant cause of discrepancy between GOME-2-GF and the two other products for these ecosystems. Considering the SIF products at 740 nm, the bias between GF and N28 or OCO-2(@740nm) varies typically between 0.1 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$ (between GF and N28 for C3/C4 crops) and 0.37 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$ (between GF and N28 for TrEBF). Apart for BoC3GRA that exhibits lower biases (below 0.1 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$), the biases for the other PFTs are also significant, varying between 0.1 and 0.2 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$. Given that the two GOME-2 products are derived from the same instrument, we expected them to be rather similar, although this disagreement has already been pointed out (Köhler et al., 2015; Parazoo et al., 2019). The mean ratio between GOME-2-GF and GOME-2-N28 varies typically between 1.5 and 2 depending on the PFT. The bias in magnitude is one order of magnitude higher than the theoretical accuracy of OCO-2 (0.05 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$) and is comparable to the systematic error associated to GOME-2-N28 (about 0.2 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$). Furthermore, the bias is also higher or about the same order of magnitude than the random error of GOME-2 SIF products at comparable spatial and temporal resolutions (0.06–0.23 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$ for GOME-2-N28 and 0.03–0.11 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$ for GOME-2-GF). For boreal needleleaved forests and grass PFTs, the bias between products is less prominent; indeed, it is not the main cause of the differences between GOME-2-N28 and the other two products for those PFTs.

The GOME-2-N28 and OCO-2 SIF products have more comparable annual mean values (Figure 4). The mean bias between these two SIF products is often negligible for all PFTs (Figure 6 Panel 2), varying between 0.02/0.05 and 0.06/0.2 $W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$ (considering OCO-2@740 nm/OCO-2@757 nm respectively) depending on PFT. A higher mismatch is found for TrEBF and BoDBF.

4.1.2. Seasonal Amplitude (Variance Term)

The largest SIF seasonal amplitude is obtained for GOME-2-GF, followed by GOME-2-N28 and OCO-2(@740 nm) and finally OCO-2(@757 nm). This is seen particularly for deciduous forests or grass PFTs (Figure 4). The mismatch in terms of fluctuation magnitude (referred to as the "variance term" of the MSD decomposition) scales with the mean seasonal amplitude. It is also related to the noise level of each products, which actually scales with the square root of the signal level (highly reflective surfaces are associated

with higher retrieval errors). The variance term is generally the second cause of discrepancy between OCO-2(@757 nm) and the GOME-2-GF products as seen in Figure 6. The spectral normalization of OCO-2 reduces that discrepancy. The variance mismatches between GOME-2-N28 and either GOME-2-GF or OCO-2(@740 nm) are small (being the more consistent with the latter): They are the third driver of the global difference between GOME-2-N28 and the other two products.

4.1.3. Seasonal Phase (Correlation Term)

The illustrative clusters of Figure 4 show similar SIF seasonalities recorded by the three space-borne products. Although the difference in phase is the first most important driver of the differences between GOME-2-N28 versus OCO-2(@740 nm) and the second driver between GOME-2-GF and both GOME-2-N28 and OCO-2(@740 nm), its relative contribution to the total error (i.e., to the MSD) between products is most often small compared to the square bias (the mean contribution of the square bias is higher by a factor of about 2 between GOME-2-GF and GOME-2-N28 and of about 3 between OCO-2@740 nm and GOME-2-GF, depending on the PFT).

The lower correlations (typically between 0.25 and 0.5) between each pair of SIF products are found for the PFTs/clusters exhibiting weak seasonal cycles (TrEBF, TeEBF-c1, TeC3GRA-c2, TrC3GRA-c1 and c3, and C4GRA-c3). However, this is less significant when the seasonal cycle is weak. For the other PFTs/clusters the values are mostly above 0.7 (40th percentile) which reveals rather consistent SIF seasonalities between the three data sets. One may however observe in Figure 4 some phasing discrepancies for some clusters: for TeDBF (cluster 2) and BoDBF, where both OCO-2 and GOME-2-N28 tend to detect an earlier start of the growing season than GOME-2-GF.

4.1.4. Temporal Evolution of the Differences Between Products

We performed the same MSD decomposition between SIF data sets over the 2016–2018 period (using the same pixels/clusters), in order to investigate a potential evolution of the SIF product differences. We mostly observed a decreasing trend of the bias and the variance (in a lower extent however) between OCO-2 (@757 nm) and the two GOME-2 SIF data. Using year 2015 as the reference, the median bias (over all PFTs) between OCO-2 and GF decreases by 4.4% in 2016, by 10.2% in 2017, and by 9.6% in 2018; the decrease is of 9.1%, 11.9%, and 12.2% between OCO-2 and N28. On another hand, the median bias between GOME-2 GF and N28 is more stable over time (it increases by 1% in 2016, and decreases by 3.4% in 2017 and 4% and 2018). The magnitude of the change in bias varies with PFT, and we observed that the decrease was more pronounced for the boreal pixels whatever the pair of SIF products considered; the decrease (between OCO-2 and the two GOME-2 products) is more important in 2016 and 2017 while an increase (relative to 2017) is observed in 2018. The changes in variances (median over all PFTs) show similar features (although to a lower extent) than the changes in biases, with decreasing differences between OCO-2 and the two GOME-2 products. These changes are between OCO-2 and GF, −3.2% in 2016, −8.5% in 2017, and −9.6% in 2018 (relative to 2015); between OCO-2 and N28, −4%, −3.4%, and −9.3%; between GOME-2 GF and N28, +10.9%, −0.8%, and +2.2%. The variations in phase are on average much lower (considering also that the absolute values for 2015 are relatively low). The median values are 0% between OCO-2 and both GOME-2 products for 2016 and 2017, while in 2018 they increase by 2.5% for GF and decreases by 3.3% for N28. The variations are about −7% to −9% between GOME-2 GF and N28 over 2016–2018.

Overall, the continuous decrease in bias (in absolute values) over 2015–2017 between OCO-2(@757 nm) and the two GOME-2 SIF products, while the bias between the GOME-2 data remains stable, is consistent with the observed temporal degradation of the GOME-2 instrument and its impact on SIF estimates (Zhang, Joiner, et al., 2018).

4.2. SIF-GPP Linear Regression

In the toy assimilation experiment, we considered the three SIF products at 740 nm. As expected, given that the *a priori* (before assimilation) a and b coefficients of the SIF-GPP linear relationship are somewhat arbitray (although we use plausible values, which are identical for all PFTs), Figure 7a shows large misfits between prior ORCHIDEE SIF simulations and each of the three SIF data sets for the three MSD components: The mean biais is relatively large, but it varies between the different SIF products at the global scale (Figure 7a). More importantly, the phase differences between the prior ORCHIDEE-SIF and the two

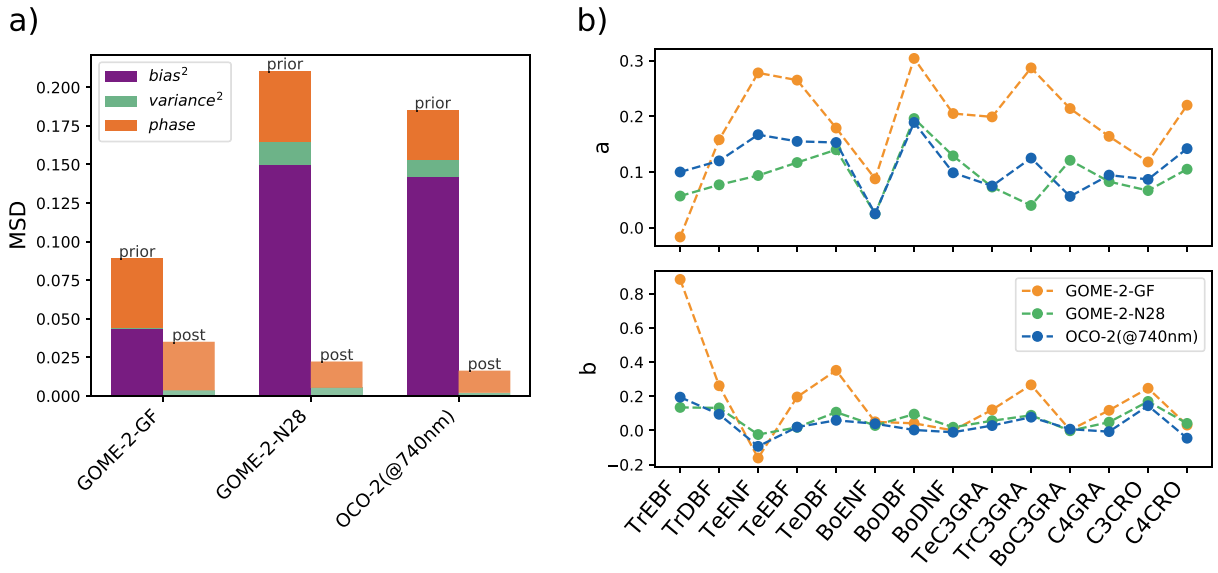

Figure 7. (a) MSD (mean square deviation) decomposition between ORCHIDEE SIF simulations prior and posterior to the optimization of the scaling *a* and offset *b* parameters, and the 3 SIF products at 740 nm (year 2015) over all plant functional types (bias [purple], variance [green], phase [orange]); (b) Estimates of *a* and *b* by plant functional type using three SIF products.

GOME-2 products are important (as important as the squared bias with GOME-2-GF) and are of the same order of magnitude for both GOME-2 products and slightly lower with OCO-2.

As expected, the optimization of the slope and offset parameters of the SIF-GPP linear relationship leads to a strong reduction of the MSD between the model and each SIF products. Following the optimization of the *a* and *b* parameters, the residual bias between all three space-borne SIF data and the optimized SIF model becomes negligible (Figure 7a), whatever the bias between the prior model and the SIF data. The main cause of the posterior residual misfit between ORCHIDEE SIF and the SIF data sets is linked to phase differences. Although the phase mismatch has been slightly reduced after assimilation (in relation to the tuning of the slope parameter *a* of equation (2)), a residual phase difference between the model and the data was expected given that no parameter acting on the GPP itself is optimized in this toy experiment. The differences in the optimized values of *a* and *b* for each PFT illustrate the amount of the correction required for cancelling out the initial mismatch in bias (and variance and phase, to a lower extent) between ORCHIDEE SIF and each SIF product (Figure 7b). The values of *a* and *b* estimated from OCO-2(@740 nm) and GOME-2-N28 are closer to each other than to those retrieved from GOME-2-GF. This is to be expected given that OCO-2 and GOME-2-N28 were in close agreement (see section 4.1). Differences in *a* and *b* between PFTs can be significant. The offset parameter *b* varies between PFTs for GOME-2 GF while it remains relatively similar for the other two SIF products. The slope parameter *a* varies between PFTs for the three SIF products, with however no clear pattern between boreal, temperate, and tropical PFTs or between forests and grasses/crops.

Real-world DGVM assimilations using SIF data would optimize GPP-related parameters associated to photosynthesis and phenology at the same time as optimizing parameters related to SIF observation operator (i.e., the *a* and *b* parameters of our toy model, as in MacBean et al., 2018). Tuning parameters acting like *a* and *b*, that is, parameters that have no direct control on GPP, will thus correct for most potential biases between modeled and observed SIF data. Nevertheless, any systematic error on the SIF data would also transfer to the estimation of the GPP related parameters (Dee, 2005). Although the aliasing of bias in the observations onto the retrieved parameters might result in improved SIF and GPP estimates during the assimilation window, it could nevertheless likely result in inaccurate future SIF and GPP simulations under different climate/surface conditions. Put differently, the assimilation would get the right GPP estimates for the wrong reasons (due to incorrect parameter values).

5. Discussion and Conclusion on the Potential Impact of SIF Difference on GPP Optimization

Overall, the differences between the three SIF products are not negligible. The disagreements in SIF annual mean values and seasonal amplitude that we see between GOME-2-GF and OCO-2/GOME-2-N28 are significant enough to require further investigation so we can better understand the causes.

In particular, we observed an overestimation of GOME-2-GF SIF magnitude as compared to GOME-2-N28 and OCO-2. The mean bias between GOME-2-GF and GOME-2-N28/OCO-2(@740 nm) is 0.18/0.2 $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$ (averaged value over all PFTs considered and for year 2015) when it is only of 0.04 between GOME-2-N28 and OCO-2(@740 nm). On another hand, the three SIF products show relatively consistent seasonal variations across PFTs (also found by Parazoo et al., 2019). Likely, the differences in annual mean value and seasonal amplitude originate from multiple sources. They are expected depending on the spectral bands that are sampled. Therefore, we expect a difference between OCO-2(@757 nm) and GOME-2 (740 nm), but not between the two GOME-2 products. Indeed, a first-order correction of the difference in spectral sampling between OCO-2 and GOME-2 results in a closer agreement between OCO-2(@740 nm) and GOME-2-N28 SIF amplitudes than between the two GOME-2 data sets. The discrepancy in annual mean values between the GOME-2 products is likely caused by differences in the spectral fitting windows that are used and the processing chains with the wider GOME-2-GF window associated to higher SIF estimates (Parazoo et al., 2019). GF uses a larger retrieval window (720–758 nm) than N28 (734–758 nm) that is more impacted by H_2O absorption features and reabsorption of the fluorescence signal at the shorter wavelengths (between 720 and 734 nm).

Differences in data selection are another potential source of discrepancy. For instance, GF and N28 use different cloud filtering approaches, which results in different amounts of data contamination and spatial sampling. These contrasting temporal sampling protocols, after aggregation on the 0.5° × 0.5° grid are visible regionally (Figure 2). The GF product seems globally less selective than N28 which generally results in a higher number of available individual observations for the 16-day / 0.5° aggregated data. This result however varies spatially. For example GF may provide more than three times more individual measurements than N28 over large regions in North America, Europe, Asia, and South Africa; on anoter hand, N28 may provide about as twice more data than GF in South America (as for instance around the Tropic of Capricorn on the eastern coast) or in eastern China. Joiner et al. (2014) identified the contamination of SIF data by clouds as a source of systematic error and therefore another possible explanation for the differences between GF and N28. The definition of the cloud fraction products used in the two GOME-2 processing chains likely differs therefore leading to a possible different acceptance or rejection outcome for a given observation. As highlighted by Köhler et al. (2015), the cloud fraction threshold has a direct impact on the estimated SIF value (which decreases as the cloud fraction threshold increases). In addition, it also impacts both the number of individual observations used for the 16-day and 0.5° synthesis and its associated error (defined as the inverse of the square root of the number of observations).

Several other possible sources of differences between GOME-2 and OCO-2 may be linked to differences in the acquisition time: (1) The daily correction might not work well for biomes with distinct daily cycles in SIF/GPP; (2) the instantaneous cloud cover can have an impact on the daily average; and (3) the early afternoon overpass time of OCO-2 (13:30 compared to 9:30 for GOME-2) may result in SIF observations that are more sensitive to hydric stress for some biomes. Additionally, the following may also play a direct role: (1) absolute radiometric calibration errors; (2) instrumental artifacts (drift and changes in overpass time); (3) change of the instrument response function over an orbit (related to temperature variations); (4) directional effects impacting SIF estimates for observations associated with different view angles (Zhang, Zhang, et al., 2018); and (5) representativeness error due to inhomogeneous sampling of the grid cells (depending on the footprint and revisit time of an instrument).

If not well understood nor accounted for, biases in the SIF data can induce misinterpretation of the modeled processes driving photosynthesis (and possibly also phenology). A recent publication of Yang et al. (2018) interpreted the observed decrease in GOME-2-SIF signal (NASA SIF product) in the Amazon forest during the 2015–2016 El Niño event as a reduction of photosynthetic capacity during the drought given that both the incoming solar radiation and the enhanced vegetation index were showing an opposite trend. The findings have since been questioned by Zhang, Joiner, et al. (2018), who attributed at least part of the negative

trend in SIF to an artifact in the data set caused by a degradation of the instrument. The temporal degradation of GOME-2 was also pointed out in our study where we showed that the mean bias between OCO-2 and the two GOME-2 products mostly decreases over the 2015–2018 period, while it remains more stable between the two GOME-2 products. Furthermore, cross-comparison of the temporal variations of SIF products derived from different missions are generally performed at broad spatial scales (Jeong et al., 2017; Köhler, Guanter, et al., 2018; Sun et al., 2018). Such a broad comparison may smooth discrepancies between products that could occur at a higher spatial resolution (such as the PFT level of current DGVMs) therefore hiding possible disagreements in terms of seasonality. Similarly, artifacts (if any) in temporal variations of SIF may result in an erroneous analysis of leaf phenology and photosynthesis temporal dynamics and the associated bioclimatic drivers.

In the context of data assimilation, MacBean et al. (2016) already showed how biases in observations can result in different parameter estimates using a toy model DA experiments. From the perspective of assimilating remotely sensed SIF products to constrain DGVM GPP simulations, the theoretical impact of data biases can be illustrated in the "ideal" case of a linear model H with Jacobian matrix $\mathbf{H}$. Then, formally, the observation bias $\boldsymbol{\beta}_\mathbf{y}$ transfers linearly to a bias $\boldsymbol{\beta}_\mathbf{x}$ in parameters following the equation $\boldsymbol{\beta}_\mathbf{x} = (\mathbf{H}^\mathbf{t}\mathbf{R}^{-\mathbf{1}}\mathbf{H} + \mathbf{B}^{-\mathbf{1}})^{-1}\mathbf{H}^\mathbf{t}\mathbf{R}^{-\mathbf{1}}\boldsymbol{\beta}_\mathbf{y}$ (see Supporting Information S1 in Bousquet et al., 2018, for demonstration). Although not directly applicable to the toy model optimization presented in this study, this equation however highlights that the biases in estimated parameters and optimized GPP (i.e. $H(\boldsymbol{\beta}_\mathbf{x})$) are proportional to the observation bias and depend on (1) the model sensitivity with respect to each parameter ($\mathbf{H}$) and on (2) the error covariance matrices ($\mathbf{R}$ and $\mathbf{B}$). The $\mathbf{R}$ and $\mathbf{H}$ matrices may differ depending on the SIF data set that is assimilated and hence the resulting impact of observation bias. Overall, this means that although we may reduce the misfit between the data and the model, the optimized parameters may be less optimal than the prior estimates—with a subsequent detrimental impact on GPP predictions. To illustrate this, let us consider the case where GPP is modeled using a simple light use efficiency (LUE) model (GPP = LUE × APAR) and where the a and b parameters were previously calibrated against a reference data set (*in situ* measurements or satellite product); a and b will now be fixed during the assimilation, and the *LUE* coefficients will be optimized per PFT. Our toy model experiment showed that the optimized value of the scaling coefficient of GPP (a) can double depending on the assimilated SIF data (i.e., varying typically from 0.1 to 0.2 when either one of the GOME-2 product is used, depending on the PFT; see Figure 7b). The optimized LUE coefficients will change similarly whether one or another SIF product is assimilated, leading to a doubling of the optimized GPP. Optimising both a and LUE will however decrease the impact on the GPP, depending on the ratio of the *a priori* errors associated to both parameters in the $\mathbf{B}$ matrix. Further quantification of SIF biases on GPP estimates would require real data assimilation experiment, in particular the characterization of the $\mathbf{H}$ matrix.

The impact of data biases on the optimized GPP hence depends on the implemented functional link (observation operator) between SIF and GPP. In the absence of tuning parameters in the SIF observation operator capable of compensating for model-data biases, systematic errors in the observations will more strongly impact the estimation of parameters controlling the magnitude of the GPP, which in turn will likely lead to decrease the model skill. If a linear SIF-GPP empirical relationship is assumed, the differences in bias and variance between the modeled GPP and any space-borne SIF products will partially be cancelled out by the scaling factor and offset parameters (e.g., a and b in this study) optimized for each PFT. The use of a mechanistic SIF-GPP model implies that the observation operator is able to represent SIF values in the observed waveband, at the time of the satellite overpass, or daily integrated values depending on the assimilated quantity. This may therefore account for expected differences between OCO-2 and GOME-2. However, it is still the case that a remaining bias or systematic error in any data set will still negatively impact the posterior mean GPP values. For evergreen ecosystems where SIF mean magnitude in particular is the major observational constraint that can be made use of (as opposed to phenology), this issue is critical.

In a real data assimilation experiment, seasonal phase differences between SIF products will also have a direct impact on phenology-related parameters that control GPP seasonality. Although the seasonal phase differences between space-borne SIF products appeared to be generally small in our analysis, we observed some discrepancies (like earlier vegetation onsets for TeDBF and BoDBF recorded by OCO-2 and GOME-2-N28 for instance), which are also depicted by the different phase mismatch with ORCHIDEE in Figure 7

a. Depending on the choice of the assimilated SIF data set, we may get a different calibrated model and different responses of the timing of the growing season to climate change.

Based on the results of this study, we argue for more detailed cross-platform comparisons (Lu et al., 2018; Sun et al., 2018) and for more validation of space-borne SIF products using independent measurements (Frankenberg et al., 2018; Sun et al., 2017; Yang et al., 2015). The recent launch of TROPOMI (Guanter et al., 2015; Köhler, Frankenberg, et al., 2018) and the forthcoming OCO-3 (Stavros et al., 2017) and FLEX (FLuorescence EXplorer) missions (Drusch et al., 2017) will bring complementary SIF data and will therefore contribute to addressing differences between SIF products. In addition, the scientific community will also benefit from the preparatory studies as well as from the ground validation network supporting the FLEX mission (Damm et al., 2015; Verhoef et al., 2018). Understanding and addressing the issue of biases and artifacts in the space-borne SIF data sets is thus of primary importance; otherwise, their potential benefit for improving our knowledge on the fate of the carbon cycle evolution under a changing climate may be more limited than expected.

Acknowledgments
The authors are very grateful to LSCE's computer team for the computing support and resources. This study was supported by CNES-TOSCA under the FluOR and ECOFLUO projects, and the Copernicus Atmosphere Monitoring Service (CAMS41 project) implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF) on behalf of the European Commission. The OCO-2 SIF product is available at ftp://fluo.gps.caltech.edu/data/OCO2/sif_lite_B8100/; GOME-2-GF is available at ftp://ftp.gfz-potsdam.de/home/mefe/GlobFluo/GOME2/; GOME-2-N28 can be accessed at https://avdc.gsfc.nasa.gov/pub/data/satellite/MetOp/GOME_F/v28/.

References

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., et al. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, *53*, 785–818. https://doi.org/10.1002/2015RG000483

Bousquet, P., Pierangelo, C., Bacour, C., Marshall, J., Peylin, P., Ayar, P. V., et al. (2018). Error budget of the MEthane Remote LIdar missioN and its impact on the uncertainties of the global methane budget. *Journal of Geophysical Research: Atmospheres*, *123*, 11–766. https://doi.org/10.1029/2018JD028907

Cramer, W. (1997). Using plant functional types in a global vegetation model. In T. M. Smith, H. H. Shugart, & F. I. Woodward (Eds.), *Plant Functional Types: Their Relevance to Ecosystem Properties and Global Change*, *IGBP Book Series* (Vol. 1, pp. 271–288). Cambridge: Cambridge University Press.

Damm, A., Guanter, L., Paul-Limoges, E., Van der Tol, C., Hueni, A., Buchmann, N., et al. (2015). Far-red sun-induced chlorophyll fluorescence shows ecosystem-specific relationships to gross primary production: An assessment based on observational and modeling approaches. *Remote Sensing of Environment*, *166*, 91–105. https://doi.org/10.1016/j.rse.2015.06.004

Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3323–3343.

Drusch, M., Moreno, J., Del Bello, U., Franco, R., Goulas, Y., Huth, A., et al. (2017). The FLuorescence EXplorer Mission Concept—ESA's Earth Explorer 8. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(3), 1273–1284. https://doi.org/10.1109/TGRS.2016.2621820

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., Köhler, P., Magney, T. S., Geier, S., Lawson, P., Schwochert, M., et al. (2018). The Chlorophyll Fluorescence Imaging Spectrometer (CFIS), mapping far red fluorescence from aircraft. *Remote Sensing of Environment*, *217*, 523–536. https://doi.org/10.1016/j.rse.2018.08.032

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*, 1–12. https://doi.org/10.1016/j.rse.2014.02.007

Gauch, H. G., Hwang, J. T., & Fick, G. W. (2003). Model evaluation by comparison of model-based predictions and measured values. *Agronomy Journal*, *95*(6), 1442–1446.

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., Alonso, L., Gómez-Chova, L., Amorós-López, J., Vila, J., & Moreno, J. (2007). Estimation of solar-induced vegetation fluorescence from space measurements. *Geophysical Research Letters*, *34*, L08401. https://doi.org/10.1029/2007GL029289

Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P. E., Gómez-Dans, J., Kuze, A., et al. (2012). Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. *Remote Sensing of Environment*, *121*, 236–251. https://doi.org/10.1016/j.rse.2012.02.006

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*, 201320008.

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

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., 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., & Middleton, E. M. (2011). First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences*, *8*(3), 637–651.

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

Khosravi, N., Vountas, M., Rozanov, V. V., Bracher, A., Wolanin, A., & Burrows, J. P. (2015). Retrieval of terrestrial plant fluorescence based on the in-filling of far-red Fraunhofer lines using SCIAMACHY observations. *Frontiers in Environmental Science*, *3*, 78.

Kobayashi, K., & Salam, M. U. (2000). Comparing simulated and measured values using mean squared deviation and its components. *Agronomy Journal*, *92*(2), 345–352.

Köhler, P., Frankenberg, C., Magney, T. S., Guanter, L., Joiner, J., & Landgraf, J. (2018). Global retrievals of solar induced chlorophyll fluorescence with TROPOMI: First results and inter-sensor comparison to OCO-2. *Geophysical Research Letters*, *45*, 10,456–10,463. https://doi.org/10.1029/2018GL079031

Köhler, P., Guanter, L., Kobayashi, H., Walther, S., & Yang, W. (2018). Assessing the potential of sun-induced fluorescence and the canopy scattering coefficient to track large-scale vegetation dynamics in Amazon forests. *Remote Sensing of Environment*, *204*, 769–785.

Koffi, E. N., Rayner, P. J., Norton, A. J., Frankenberg, C., & Scholze, M. (2015). Investigating the usefulness of satellite-derived fluorescence data in inferring gross primary productivity within the carbon cycle data assimilation system. *Biogeosciences*, *12*(13), 4067–4084.

Köhler, P., Guanter, L., & Joiner, J. (2015). A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data. *Atmospheric Measurement Techniques*, *8*(6), 2589–2608.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., et al. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, *19*, GB1015. https://doi.org/10.1029/2003GB002199

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., et al. (2017). Global carbon budget 2017. *Earth System Science Data Discussions*, 1–79. https://doi.org/10.5194/essd-2017-123

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 C ommunity L and M odel version 4. *Global Change Biology*, *21*(9), 3469–3477. https://doi.org/10.1111/gcb.12948

Lu, X., Cheng, X., Li, X., Chen, J., Sun, M., Ji, M., et al. (2018). Seasonal patterns of canopy photosynthesis captured by remotely sensed sun-induced fluorescence and vegetation indexes in mid-to-high latitude forests: A cross-platform comparison. *Science of the Total Environment*, *644*, 439–451. https://doi.org/10.1016/j.scitotenv.2018.06.269

MacBean, N., Maignan, F., Bacour, C., Lewis, P., Peylin, P., Guanter, L., et al. (2018). Strong constraint on modelled global carbon uptake using solar-induced chlorophyll fluorescence data. *Scientific Reports*, *8*(1), 1973. https://doi.org/10.1038/s41598-018-20024-w

MacBean, N., Peylin, P., Chevallier, F., Scholze, M., & Schürmann, G. (2016). Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. *Geoscientific Model Development*, *9*(10), 3569–3588.

Norton, A. J., Rayner, P. J., Koffi, E. N., & Scholze, M. (2018). Assimilating solar-induced chlorophyll fluorescence into the terrestrial biosphere model BETHY-SCOPE v1. 0: Model description and information content. *Geoscientific Model Development*, *11*(4), 1517–1536.

Parazoo, N. C., Bowman, K., Fisher, J. B., Frankenberg, C., Jones, D. B., Cescatti, A., et al. (2014). Terrestrial gross primary production inferred from satellite fluorescence and vegetation models. *Global Change Biology*, *20*(10), 3103–3121. https://doi.org/10.1111/gcb.12652

Parazoo, N. C., Frankenberg, C., Köhler, P., Joiner, J., Yoshida, Y., Magney, T., et al. (2019). Towards a harmonized long-term spaceborne record of far-red solar induced fluorescence. *Journal of Geophysical Research: Biogeosciences*, *124*, 2518–2539. https://doi.org/10.1029/2019JG005289

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., et al. (2015). Plant functional type classification for earth system models: Results from the European Space Agency's Land Cover Climate Change Initiative. *Geoscientific Model Development*, *8*(7), 2315–2328. https://doi.org/10.5194/gmd-8-2315-2015

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., & Solomon, A. M. (1992). Special paper: A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, *19*(2), 117–134. https://doi.org/10.2307/2845499

Sanders, A. F., Verstraeten, W. W., Kooreman, M. L., Van Leth, T. C., Beringer, J., & Joiner, J. (2016). Spaceborne sun-induced vegetation fluorescence time series from 2007 to 2015 evaluated with australian flux tower measurements. *Remote Sensing*, *8*(11), 895.

Somkuti, P., Boesch, H., & Parker, R. (2015). Remote sensing of chlorophyll fluorescence with GOSAT. In *EGU General Assembly Conference Abstracts* (Vol. 17).

Stavros, E. N., Schimel, D., Pavlick, R., Serbin, S., Swann, A., Duncanson, L., et al. (2017). ISS observations offer insights into plant function [Comments and Opinion]. https://doi.org/10.1038/s41559-017-0194

Sun, Y., Frankenberg, C., Jung, M., Joiner, J., Guanter, L., Köhler, P., & Magney, T. (2018). Overview of Solar-Induced chlorophyll Fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP. *Remote Sensing of Environment*, *209*, 808–823.

Sun, Y., Frankenberg, C., Wood, J. D., Schimel, D. S., Jung, M., Guanter, L., et al. (2017). OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science*, *358*(6360), eaam5747. https://doi.org/10.1126/science.aam5747

Verhoef, W., van der Tol, C., & Middleton, E. M. (2018). Hyperspectral radiative transfer modeling to explore the combined retrieval of biophysical parameters and canopy fluorescence from FLEX–Sentinel-3 tandem mission multi-sensor data. *Remote Sensing of Environment*, *204*, 942–963.

Walther, S., Voigt, M., Thum, T., Gonsamo, A., Zhang, Y., Köhler, P., et al. (2016). Satellite chlorophyll fluorescence measurements reveal large-scale decoupling of photosynthesis and greenness dynamics in boreal evergreen forests. *Global Change Biology*, *22*(9), 2979–2996. https://doi.org/10.1111/gcb.13200

Wang, P., Stammes, P., Pinardi, G., & van Roozendael, M. (2008). FRESCO+: An improved O 2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals. *Atmospheric Chemistry and Physics*, *8*(21), 6565–6576.

Wood, J. D., Griffis, T. J., Baker, J. M., Frankenberg, C., Verma, M., & Yuen, K. (2017). Multiscale analyses of solar-induced florescence and gross primary production. *Geophysical Research Letters*, *44*, 533–541. https://doi.org/10.1002/2016GL070775

Yang, J., Tian, H., Pan, S., Chen, G., Zhang, B., & Dangal, S. (2018). Amazon drought and forest response: Largely reduced forest photosynthesis but slightly increased canopy greenness during the extreme drought of 2015/2016. *Global Change Biology*, *24*(5), 1919–1934.

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., Joiner, J., Gentine, P., & Zhou, S. (2018). Reduced solar-induced chlorophyll fluorescence from GOME-2 during Amazon drought caused by dataset artifacts. *Global Change Biology*, *24*(6), 2229–2230.

Zhang, Y., Guanter, L., Berry, J. A., van der Tol, C., Yang, X., Tang, J., & Zhang, F. (2016). Model-based analysis of the relationship between sun-induced chlorophyll fluorescence and gross primary production for remote sensing applications. *Remote Sensing of Environment*, *187*, 145–155.

Zhang, Z., Zhang, Y., Joiner, J., & Migliavacca, M. (2018). Angle matters: Bidirectional effects impact the slope of relationship between gross primary productivity and sun-induced chlorophyll fluorescence from Orbiting Carbon Observatory-2 across biomes. *Global Change Biology*, *24*(11), 5017–5020. https://doi.org/10.1111/gcb.14427