MASTER THESIS

Spaceborne Observations of Sun-Induced Vegetation Fluorescence: A 2007-2015 global time series from GOME-2A

Author:
Maurits L. Kooreman

Supervisors:
dr. Abram F. J. Sanders
dr. ir. Willem W. Verstraeten

September 25, 2015

Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

Universiteit Utrecht

Abstract

The increasing CO_2 levels in the atmosphere over the past two centuries give rise to a global greenhouse effect and so interest in measuring, modeling and predicting the global carbon cycle is growing. We have a good understanding of the global carbon cycle, but large uncertainties still remain, particularly concerning the role of the terrestrial biosphere. Fixation of carbon through photosynthesis is the largest terrestrial sink for atmospheric CO_2. One way of gaining further insights into carbon fixation is to quantify the biosphere photosynthetic activity on a global scale. During photosynthesis, solar light is absorbed by chlorophyll and this provides the energy for photosynthesis. Any energy that is not used for photosynthesis is dissipated as heat or re-emitted as fluorescence photons at longer wavelengths. This means that photosynthetic activity in turn could be quantified by measuring sun-induced fluorescence (SiF). Although these SiF emissions are small (about 1% of the total absorbed radiation energy [*Porcar-Castell et al.*, 2014]), they can can be measured from space.

This study uses the GOME-2 spectrometer (on board the MetOp-A platform) to retrieve the SiF signal from radiance spectra. We adopt a retrieval methodology from *Joiner et al.* [2013] and *Guanter et al.* [2014], to implement a retrieval setup that enables near real-time processing. A data set is constructed providing a time series of SiF on a global scale spanning the entirety of the GOME-2A mission (January 2007 to present).

Furthermore, the study investigates the relation between in situ (flux tower) derived GPP measurements and the remotely sensed SiF retrieval. This study focuses on the United States Cornbelt which provides an excellent region of research, as its homogeneous distribution of vegetation enables good comparison with the coarse GOME-2A footprint. In this area, the correlation between Cornbelt-averaged SiF and flux tower derived GPP is high (R=0.88) and this new SiF retrieval method shows promises for better future understanding of carbon dynamics in the terrestrial biosphere.

Contents

1 Introduction

Due to rapidly increasing CO_2 levels over the last five decennia giving rise to a global greenhouse effect, interest is growing in measuring, modeling and predicting the global atmospheric and terrestrial carbon budget. Terrestrial vegetation accounts for approximately 54% of the global carbon sink [*Field et al.*, 1998] and therefore, knowledge on its dynamics is of great importance for understanding the global carbon budget. This study aims to improve the uncertainties on one particular terrestrial component of the global carbon budget: Gross Primary Productivity (GPP).

This GPP is the amount of carbon uptake in plants and is expressed in grams of carbon per square meter per second (gC m^{-2} s^{-1}). It is the amount of atmospheric carbon assimilated by plants during photosynthesis. Figure 1 shows that the global GPP flux is approximately 123 PgC yr^{-1}.
Plants also put carbon back into the atmosphere. The atmospheric carbon source is partly caused by autotrophic respiration, it is the plants equivalent to exhaling and partly by heterotrophic respiration which happens when dead biomass is decomposing. Fig. 1 shows a total respiration flux of approximately 118.7 PgC yr^{-1}. When including river export (1.7 PgC yr^{-1}) and weathering of soil (0.4 PgC yr^{-1}) as sinks of carbon in the biosphere, terrestrial vegetation accounts for a net biosphere carbon uptake of approximately 3.0 PgC yr^{-1}, called Net Ecosystem Production (NEP). However, significant uncertainties still remain in the NEP flux as small uncertainties in GPP and respiration can lead to large uncertainties in the NEP [*Ciais et al.*, 2013].

Plants assimilate CO_2 from the atmosphere by means of photosynthesis. For this reaction sunlight is required. The sunlight is absorbed by leaf chlorophyll pigments and converted into usable chemical energy. Usually, plenty of energy is available, and excess energy has to be removed. Part of this excess energy is emitted as light at longer wavelengths. This emission is known as sun-induced fluorescence (SiF; *Zarco-Tejada et al.* [2000]). SiF occurs in the red and near-infrared regions of the electromagnetic spectrum and shows a distinct spectral structure. The radiance spectrum therefore holds information about SiF at canopy level.

Recent developments in remote sensing enable measuring this canopy SiF signal from space using grating spectrometers. Mounting such a spectrometer on a polar-orbiting satellite allows it to scan the entire Earths surface in the time span of one and a half day.
The GOME-2A grating spectrometer on board the MetOp-A satellite has been collecting spectral data since January 2007. With its spatial resolution of 40 km x 80 km and a spectral resolution of 0.5 nm (FWHM) it is sufficiently accurate for this SiF study. With these resolutions, spatial and temporal patterns can be distinguished. Seasonal variability can be observed, as well as distinct vegetation patterns such as for example the Nile delta and the more fertile coasts of Australia.

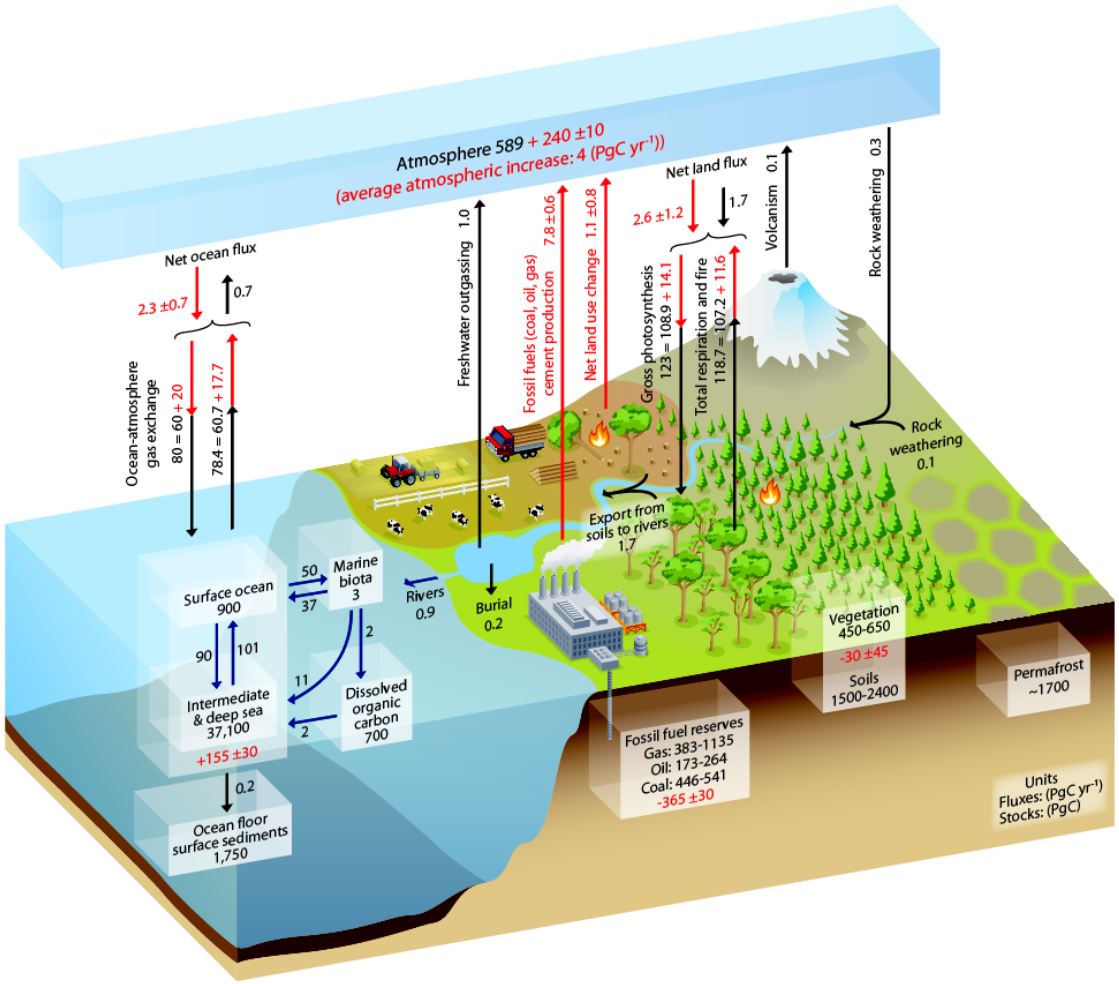

Figure 1: A schematic view on the global carbon cycle. Numbers indicate reservoir mass estimates in PgC and black arrows indicate carbon fluxes in PgC yr^{-1} before the industrial age (1750). Red arrows and values indicate the current antropogenic perturbation on the (pre-industrial) steady state. Source: The fifth IPCC report [*Ciais et al.*, 2013].

1.1 Chlorophyll *a* Fluorescence

The process of carbon uptake by vegetation is photosynthesis. Under the influence of light, CO_2 and H_2O are converted into sugars and oxygen following net reaction 1.

$$6CO_2 + 6H_2O \xrightarrow{light} C_6H_{12}O_6 + 6O_2. \quad (1)$$

Photosynthesis occurs in leaf cells, where the main pigment responsible for the absorption of solar light is chlorophyll *a* (leaf-green), supplemented by chlorophyll *b*, carotene and other pigments. The absorbed energy is made available for photo-chemical reactions. These reactions occur in a photo system, which is the larger light harvesting protein complex.

The Fraction of Absorbed Photosynthetically Active Radiation (fAPAR) is regulated by the concentration of chlorophyll in the leaf, which can change within a period of days in response to changes in the environment such as drought, rainfall, heat etc. [*Porcar-Castell et al.*, 2014]. One visible example is the turning yellow of leaves after an extended drought period where chlorophyll pigment concentrations diminish. Fig. 3 shows an overview of environmental stress factors that can cause changes in leaf chlorophyll content.

Evolved plants have two photo-systems which work in series to harvest light energy: PSI and PSII. They are the main engine in the process of light energy harvesting.

Energy collected by a photo-system is used in three different ways: It can be used for photochemistry to assimilate CO_2, it can be dissipated non-radiatively as heat, or it can be re-emitted as an infrared fluorescence photon (e.g. *Porcar-Castell et al.* [2014]). The fluorescence emission process comprises only about 1% of the total available solar energy. PSII is responsible for the majority of the SiF emissions. The direct link to CO_2 assimilation by means of photosynthesis and chlorophyll fluorescence is one of the building blocks of this study.

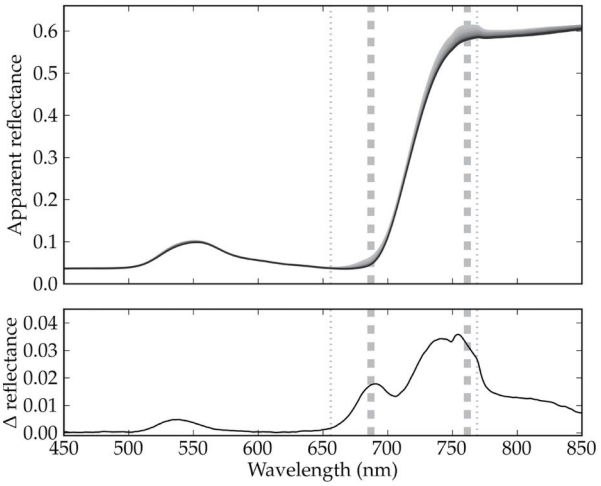

Figure 2: The gray area in the upper panel shows the change in apparent reflectance[1] after a dark adapted maple leaf is exposed to full sunlight, the steady state situation is indicated by the solid line. The bottom panel shows the difference between both spectra and it reveals the spectral structure of the fluorescence emission. The thick gray lines indicate the location of the O_2-A and B absorption bands. Reproduced from *Porcar-Castell et al.* [2014].

At dusk or dawn, when there is only few light energy available for plants, almost all energy is efficiently used for photosynthesis. This results in little heat loss and a low fluorescence

[1] In apparent reflectance, effects like shadows and directional effects are not compensated for.

signal. During the day light intensity increases and more excess energy is released through fluorescence because not all absorbed solar energy can be used for photosynthesis. This results in a fluorescence peak in the early afternoon.
In low light situations (e.g. mid-latitude winter months) chlorophyll fluorescence can be so small that it cannot be seen from space. Such situations might prove useful as a way of calibrating the SiF retrieval, as we can assume conditions without SiF to yield a value of zero.

The spectral structure of fluorescence emission is shown in Fig. 2. Two distinct maxima are visible. One fluorescence peak near 685-695 nm (F_{690}) and a larger peak around 730-740 nm (F_{735}). *Subhash and Mohanan* [1997] show that the re-absorption within the leaf affects SiF spectral line shape and the ratio of these two peaks F_{690}/F_{735}. Absorption takes mostly influences the F_{690} peak. Because of this we choose to look only at the F_{735} peak in this study.
As explained earlier, chlorophyll content changes when plants undergo stress. The ratio F_{690}/F_{735} decreases with increasing chlorophyll content [*Buschmann*, 2007]. As the ratio F_{690}/F_{735} can be measured, it can be used as a proxy for chlorophyll content and leaf stress.
A problem is the upscaling from leaf-level to canopy level fluorescence which is needed in this research. At the leaf level, up to 90% of the fluorescence can be reabsorbed, either within the leaf itself, or within the canopy, so canopy scale fluorescence depends much on e.g. vegetation structure and geometry, which is highly variable at small spatial scales.

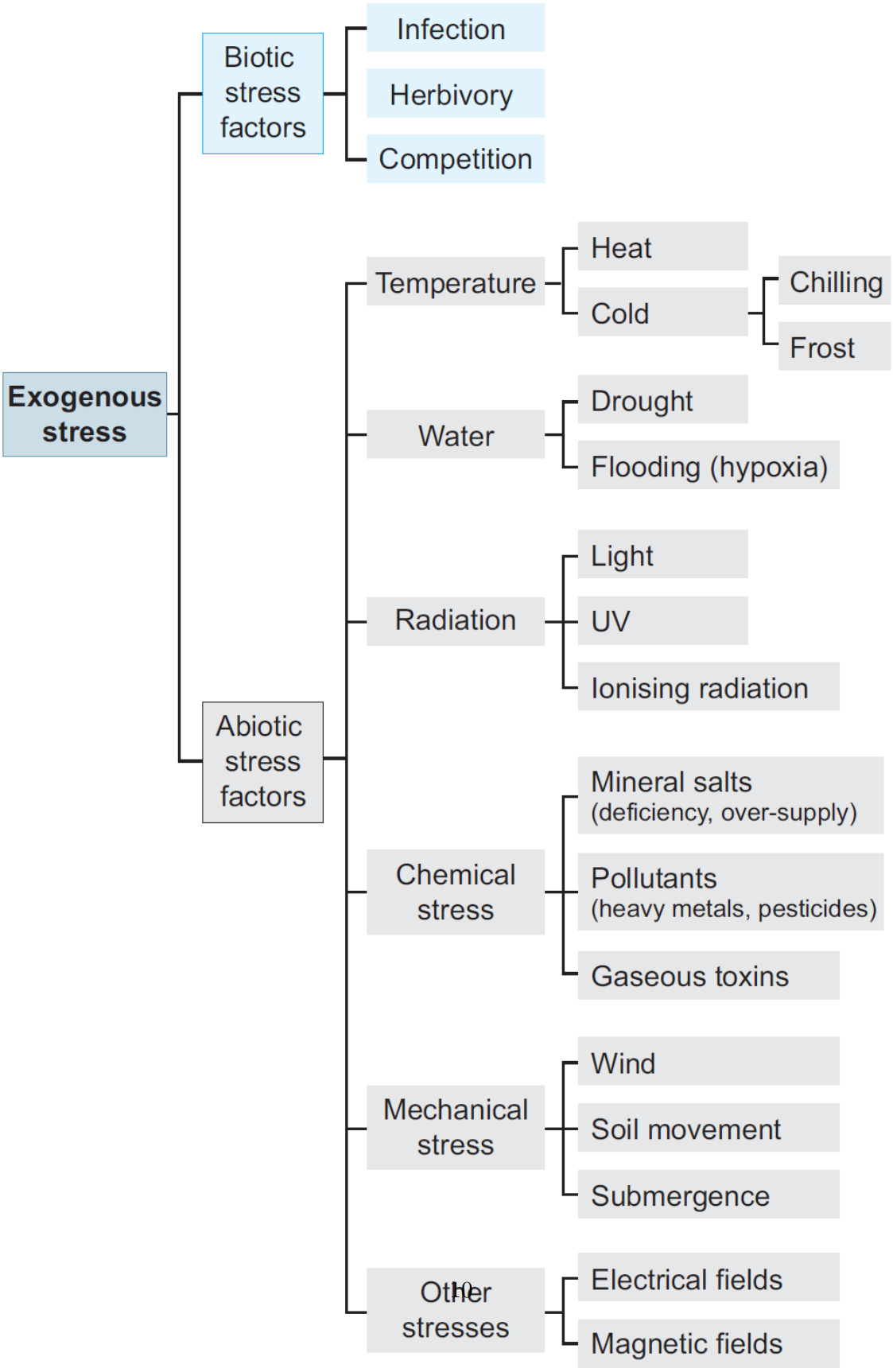

Figure 3: The multitude of stress factors which affect LUE. Source: *Schulze et al.* [2005].

1.2 Gross Primary Productivity

It is not possible to measure GPP directly through remote sensing. A proxy for GPP which can be measured by remote sensing is the amount of photosynthetically active radiation (PAR). A linear relation between GPP and PAR can be established [*Monteith*, 1972]:

$$\mathrm{GPP} = \mathrm{PAR} \times \mathrm{fAPAR} \times \mathrm{LUE} \tag{2}$$

PAR is defined as the amount of incoming solar radiation between 400 nm and 700 nm. The fAPAR is a coefficient determining the fraction of absorbed PAR. It is determined by the leaf geometry and vegetation type. The light use efficiency (LUE) is a coefficient determining the efficiency in converting the energy absorbed by the leaf to photosynthesis.

The efficiency of photosynthesis is directly related to the leaf stress. Leaf stress is influenced by many factors as illustrated by Fig. 3. This figure shows that stress can be categorized into two categories. Biotic stress factors, which are inherent to the plant itself include competition for sunlight and nutrients, diseases and grazing. The abiotic stresses are caused by the plant environment and include a whole range of factors. Many of these factors cannot be measured or are not yet completely understood. This suggests that finding a proxy for LUE could be helpful in determining the relationship between these stress factors and GPP.
In this study we compare SiF retrieved from GOME-2A with flux tower derived GPP data.

1.3 Flux towers

The CO_2 flux can be measured in situ by gas analyzers. From this, the net conversion of CO_2 into biomass by photosynthesis (GPP) can be determined using the following linear relation:

$$\mathrm{GPP} = \mathrm{ER} - \mathrm{NEE}, \tag{3}$$

Where ER is the ecosystem respiration and NEE is the Net Ecosystem Exchange (NEE > 0 denotes ecosystem uptake), see section 1.

Flux tower derived GPP data is provided by AmeriFlux[2]. A summary of the towers used in this study is shown in Table 1.
The footprint of GOME-2A is very large compared to the footprint of a flux tower which usually ranges from several hundred meters to a few kilometers.
This implies that a GOME-2A pixel includes more different types of vegetation than a flux tower footprint. So in order to make a proper comparison, we need to find an area with very homogeneously distributed vegetation such as the United States Cornbelt.

The Cornbelt in North America is defined as a rectangle (35°N-49°N, 76°W-106°W), similar to *Guanter et al.* [2014]. It spans from New Mexico to Québec covering roughly 5 million km^2. According to the SiBCASA biomes map there are only two dominant vegetation types in the Cornbelt, shown in Fig. 5. The left half of the area mostly consists of agricultural lands, while the rest of the area is covered by tall broad leaf-deciduous trees and small patches of needle leaf and evergreen trees. A small stretch of agricultural lands can also be found around 42° latitude among the forests. In this area, 13 flux towers matched the criteria of data availability and location:

- Flux tower location is within the defined Cornbelt rectangle.
- Derived level 2 GPP data is available.
- Data available for at least a year within the range 01-01-2007 to 31-12-2015.

Fig. 4 shows the flux tower locations. A fairly good spread over the area is accomplished. Flux towers are indicated by their ID.

[2]http://ameriflux.ornl.gov/

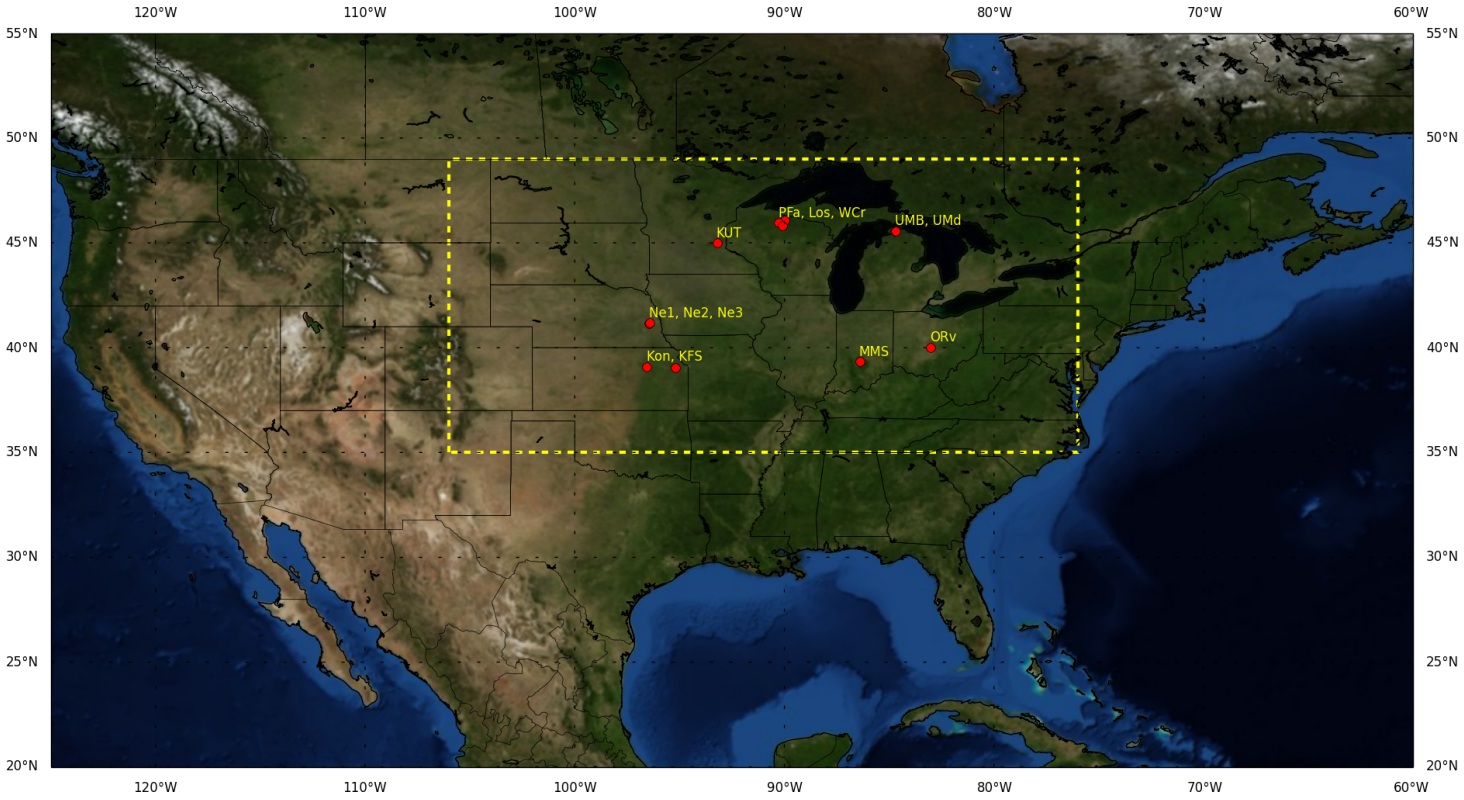

Figure 4: United States Cornbelt (35°N-49°N, 76°W-106°W). U.S. Cornbelt is indicated by the dashed yellow line. Flux tower locations are indicated by the red dots and labeled with their corresponding code. Image source: NASA, Blue Marble.

Table 1: Different towers used along with their location and their vegetation types.

ID	Name	Vegetation type	Lat. [°N]	Lon. [°E]	Elev. [m]
KFS	Kansas Field Station	Grasslands	39.0561	-95.1907	333
Kon	Konza Prairie	Grasslands	39.0824	-96.5603	443
KUT	KUOM (University of Minnesota) Turfgrass Fields	Grasslands	44.9950	-93.1863	301
Los	Lost Creek	Shrublands	46.0827	-89.9792	480
MMS	Morgan Monroe State	Deciduous Broad leaf Forest	39.3231	-86.4131	275
Ne1	Mead Irrigated	Croplands	41.1650	-96.4766	361
Ne2	Mead Irrigated Rotational	Croplands	41.1649	-96.4701	362
Ne3	Mead Rainfed	Croplands	41.1797	-96.4396	363
ORv	Olentangy River Research Park	Permanent Wetlands	40.0201	-83.0183	221
PFa	Park Falls	Mixed Forest	45.9459	-90.2723	470
UMB	University of Michigan Biological Station	Deciduous Broad leaf Forest	45.5598	-84.7138	234
UMd	University of Michigan Biological Station - disturbance	Deciduous Broad leaf Forest	45.5625	-84.6975	239
WCr	Willow Creek	Deciduous Broad leaf Forest	45.8060	-90.0798	515

For analysis, vegetation types have been classified into four different categories. *Grass*, containing all types of grasslands. *Forest*, including mixed forests and deciduous broad leaf forests. *Agriculture*, including all sorts of agricultural products. *Other* is used to describe the shrub land and wetlands sites in this study.
Flux tower GPP data is provided with a temporal resolution of 30 minutes. When data is missing, it is excluded from analysis, rather than interpolated.

1.4 SiBCASA model

The SiBCASA model combines photosynthesis and biophysical processes from a simple biosphere (SiB) model with carbon biogeochemical processes from the Carnegie-Ames-Stanford Approach (CASA) model. It calculates a range of meteorological variables as well as GPP on a 1° by 1° grid. The model temporal resolution is 10 minutes although this study uses data at a resolution of 3 hours to be able to cover a time span of 6 years. The SiBCASA model calculates the carbon, energy and water fluxes. The (latent and sensible) heat fluxes are parameterized in such a way that they account for effects of snow-cover, turbulence and rainfall. A leaf-level photosynthesis model is included. In order to upscale to canopy level, the model is driven by the Normalized Difference Vegetation Index (NDVI) data from AVHRR (Advanced Very High Resolution Radiometer) to calculate the absorbed fraction of photosynthetically active radiation (fPAR) [*Velde et al.*, 2013, 2014].

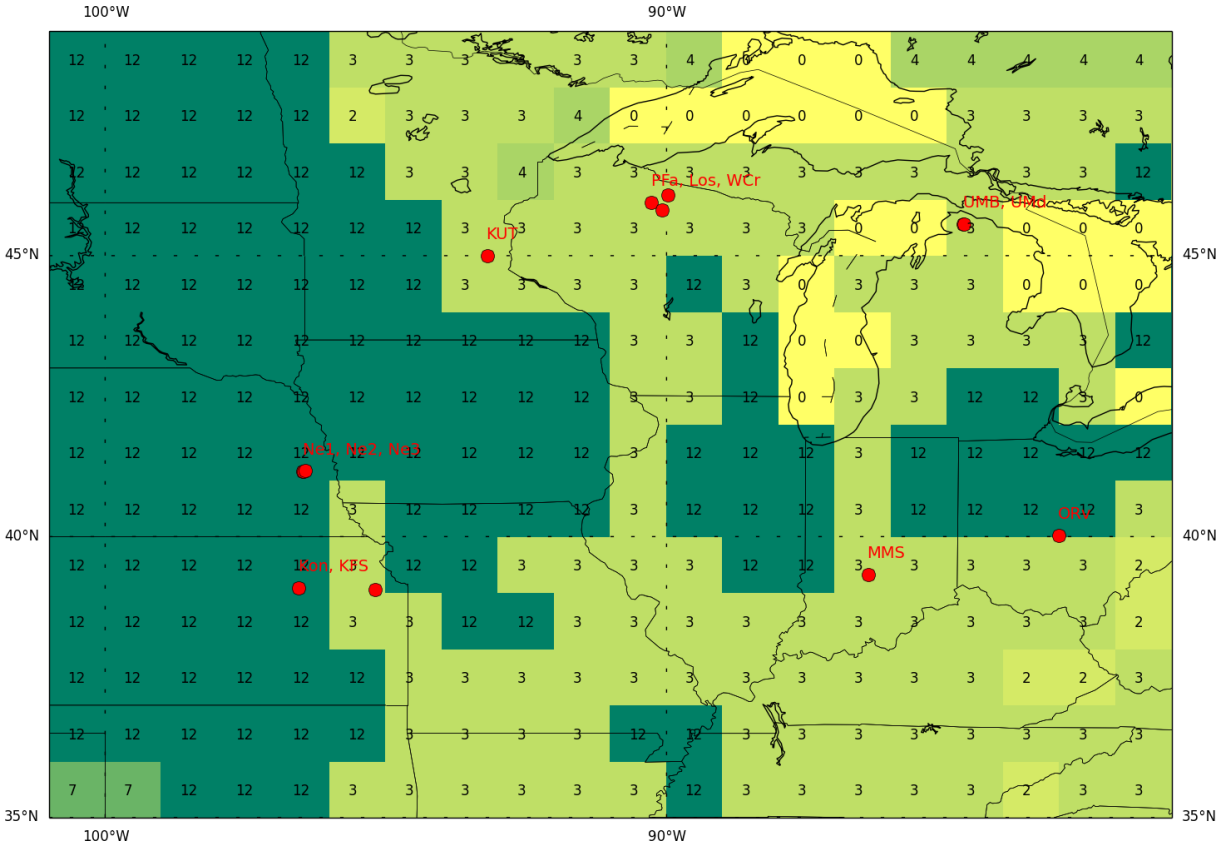

Figure 5: SiBCASA vegetation types on a 1° by 1° grid for the U.S. Cornbelt. Flux tower locations and names are indicated in red. For an explanation of biome types, see Table 2.

In order to differentiate between different vegetation types, SiBCASA uses a biome map. This map includes 12 different vegetation types: Fig. 5 shows the vegetation types for the U.S. Cornbelt. This area is comprised of two main vegetation types; agriculture (12) and tall broad- and needle leaf trees (3). All towers are within either one of those types.

Table 2: List of vegetation types included in the SiBCASA model.

Biome	Vegetation type
0	No biome specified
1	Tall broad leaf-evergreen trees
2	Tall broad leaf-deciduous trees
3	Tall broad leaf and needle leaf trees
4	Tall needle leaf trees
5	Tall needle leaf-deciduous trees
6	Short vegetation (savanna)
7	Short vegetation (grasslands)
8	Short vegetation (none)
9	Short broad leaf shrubs with bare soil
10	Short ground cover (tundra)
11	No vegetation (low latitude desert)
12	Agriculture

1.5 Research outline

The first objective of this study is to apply and to improve the retrieval method for SiF developed by *Joiner et al.* [2013] and *Guanter et al.* [2014]. It is based on satellite data to retrieve a SiF signal through a statistical comparison between reflectance spectra from vegetated and non-vegetated areas (e.g. oceans, ice caps and deserts). This methodology is used and improved in this study. We apply this methodology and retrieve the SiF product. The main difference is our choice of building a reference set of non-vegetated areas. *Joiner et al.* [2013] uses a multitude of areas and a reference time period of a single day whereas this study uses only one area and builds its reference set from a year of reflectance spectra. A second large difference is the choice of spectral window. This study includes the O_2-A band and water vapor absorption bands whereas those are excluded in the current version of the *Joiner et al.* [2013] study.
From GOME-2A data, a global monthly mean SiF data set has been constructed covering the period January 2007 - June 2015, in analogy with the entire GOME-2A mission so far. Complementing this first objective, we will compare the SiF retrieval using this method with the retrieval by *Joiner et al.* [2013].

A second objective is to evaluate the accuracy of the SiF-retrieval by comparing the satellite SiF data with ground based GPP observations derived from flux towers and model GPP output. To do this, the study focuses on a large area of relatively homogeneous vegetation spanning parts of the United States of America and Canada called the *corn belt*. GPP data derived from 13 flux towers spread over this area are used for comparison. The SiF retrieval is also compared with GPP estimates from a vegetation model (SiBCASA, details in section 1.4).

A third objective is to investigate temporal SiF anomalies in the corn belt. How do these anomalies correspond with SiBCASA and flux tower GPP anomalies of the same type and can we use this method to see the effect of climatological influences like droughts on SiF levels.

After this introduction, the retrieval implementation and GPP data sources are discussed in Section 2. Results of this SiF retrieval is presented in Section 3 along with the GPP data from flux towers and SiBCASA to provide a basis for the conclusions in Section 5.

2 SiF retrieval method

This section describes the forward model which is used to determine the top of atmosphere reflectance as function of atmospheric parameters. GOME-2A data is fed to this function and the forward model is solved iteratively to calculate SiF. Details on this process and additional necessities are presented in this section.

2.1 Forward model

The retrieval in this study is based on a method developed by *Joiner et al.* [2013] and *Guanter et al.* [2014]. The primary assumption in this model is that the ground surface reflects isotropically (Lambertian surface). This means it assumes that a particular surface element has the same apparent radiance no matter what angle it is viewed from.
The total reflectance ρ_{tot} at the top of the atmosphere (TOA) in the spectral region where SiF occurs can be decomposed into three components. The atmospheric reflectance, represented by the first term of Equation 4, the reflection due to surface reflections (second term) and a "reflectance" due to fluorescence emissions.

$$\rho_{\mathrm{tot}} = \rho_{\mathrm{a}} + \frac{\rho_{\mathrm{S}}(\lambda)T(\lambda)\bar{T}(\lambda)}{1-\rho_{\mathrm{S}}(\lambda)\bar{\rho}(\lambda)} + \frac{\pi F_{\mathrm{S}}(\lambda)\bar{T}(\lambda)}{[1-\rho_{\mathrm{S}}(\lambda)\bar{\rho}(\lambda)]E(\lambda)\cos(\theta_0)}, \tag{4}$$

here, ρ_{S} is the surface reflectance, ρ_{a} is the reflectance of the atmosphere (i.e. the reflectance when the ground surface is black), T is the total downward irradiance transmittance, $\bar{T}$ is the spherical transmittance of the atmosphere for illumination from below and $\bar{\rho}$ is the spherical reflectance of the atmosphere back to the surface. Furthermore, θ_0 is the solar zenith angle (SZA), $E(\lambda)$ is the extraterrestrial irradiance and F_{S} is the fluorescence radiance at the surface. This fluorescence radiance is assumed to be isotropic [*Joiner et al.*, 2013].

For the wavelength region of interest (700 nm - 800 nm), the molecular scattering optical thickness can be calculated with Equation 5 which considers light passing through an atmospheric column with mass M_{atm}:

$$\tau = \frac{M_{\mathrm{atm}}}{M_{\mathrm{molec}}} A_{\mathrm{V}} \sigma_{\mathrm{S}}. \tag{5}$$

Here M_{molec} is the average molecular mass of air and A_{V} is Avogrado's number.
Keeping this in mind, we follow *Joiner et al.* [2013] and neglect atmospheric scattering effects in the wavelength region of interest and hence we assume that $\rho_{\mathrm{a}}, \rho_s \simeq 0$. This assumption is reasonable because the molecular scattering optical thickness in the near-infrared is about 0.02. Using Equation 5 by calculating the amount of molecules within a light path and multiplying this with the atmospheric scattering cross section σ_S. This results in an atmospheric transmittance of 98%. Assuming negligible atmospheric scattering effects reduces Equation 4 to

$$\rho_{\mathrm{tot}} = \rho_{\mathrm{S}}(\lambda)T(\lambda)\bar{T}(\lambda) + \frac{\pi F_{\mathrm{S}}(\lambda)\bar{T}(\lambda)}{E(\lambda)\cos(\theta_0)}. \tag{6}$$

The model is greatly simplified, as now only two atmospheric paths are taken into account. The direct path of solar photons, passing through the atmosphere, reflecting at the surface and passing through the atmosphere again towards the sensor, and the path of fluorescence emissions going only from the surface to the top of the atmosphere.
The concept of total transmittance is illustrated in Fig. 6. It has a direct and a diffuse component:

$$T_{\text{total}} = \tau_{\text{dir}} + \tau_{\text{diff}}. \tag{7}$$

When light passes through the atmosphere, as illustrated in Fig. 6, three things can happen: (1) Light can be scattered out of the beam causing a decrease in τ_{dir}, (2), light can be absorbed, also leading to a decrease in τ_{dir} and (3), light can be scattered out and back into the original beam, causing an increase in τ_{diff}.

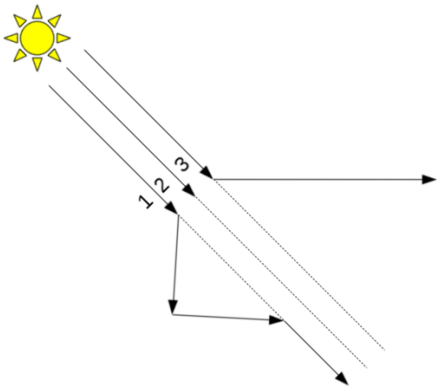

Figure 6: Illustration of the different components of atmospheric transmittance.

This means, that when neglecting diffuse effects, the transmittance now only includes atmospheric absorption as (1) and (3) from Fig. 6 are both gone. The two-way atmospheric transmittance $T\bar{T}$ can thus be expressed in terms of the atmospheric optical thickness $\tau(\lambda)$.

$$T\bar{T} = e^{-\tau(\lambda)\left[\cos(\theta)^{-1}+\cos(\theta_0)^{-1}\right]} = e^{-\tau^{\uparrow\downarrow}}, \tag{8}$$

where $\tau^{\uparrow\downarrow}$ is the two-way slant optical thickness, θ is the viewing zenith angle and θ_0 is the solar zenith angle, reducing equation (4) to:

$$\rho_{\text{tot}} = \rho_{\text{S}}(\lambda)e^{-\tau^{\uparrow\downarrow}(\lambda)} + \frac{\pi F_{\text{S}}(\lambda)}{E(\lambda)\cos(\theta_0)}e^{-\tau^{\uparrow\downarrow}(\lambda)\Phi} \qquad \Phi = \frac{\cos^{-1}(\theta)}{\cos^{-1}(\theta)+\cos^{-1}(\theta_0)}. \tag{9}$$

Now we have constructed an equation for the TOA reflectance in terms of this two-way slant optical thickness. Here, Φ is a multiplication constant due to the geometric path the light travels through the atmosphere. Equation 9 needs to be solved for $\rho_{\text{S}}(\lambda)$, $F_{\text{S}}(\lambda)$ and $\tau(\lambda)$, they are now all a function of only wavelength. Note that θ and θ_0 are essentially known constants, as they are measured by the satellite for every ground pixel. This paragraph will continue with the parameterization of these unknowns.

As illustrated in Fig. 2, the spectral distribution of fluorescence can approximated by a double Gaussian distribution [*Zarco-Tejada et al.*, 2000].

$$F_{\text{S}}(\lambda) = F_{\text{S},0}\left(\gamma\exp\left[\frac{-(\lambda-\mu_l)^2}{2\sigma_l^2}\right] + \exp\left[\frac{-(\lambda-\mu_h)^2}{2\sigma_h^2}\right]\right). \tag{10}$$

Here, μ is the peak wavelength, σ is the standard deviation and γ is a factor determining the ratio between the peak heights. Subscripts l and h respectively indicate the near-infrared and far-red peaks.
In this study only the near infrared peak is considered. Thus:

$$F_{\mathrm{S}}(\lambda) = F_{\mathrm{S},0} \exp\left[\frac{-(\lambda - \mu_h)^2}{2\sigma_h^2}\right], \tag{11}$$

with $\mu_h = 737$ nm and $\sigma_h = 33.9$ nm [*Subhash and Mohanan*, 1997][3]. $\mathrm{F}_{\mathrm{S},0}$ will be retrieved.

Surface reflectance ρ_s is assumed to be spectrally smooth and is parameterized by a fourth order polynomial for vegetated areas. For non-vegetated areas a second order polynomial proves to be sufficient [*Koelemeijer et al.*, 2003].

$$\rho_{\mathrm{S}} = \sum_{n=0}^{2/4} c_n x^n. \tag{12}$$

The spectral structure of light passing through the atmosphere is non-linearly dependent on albedo, surface pressure and other atmospheric variables. This leads to the last ingredient for solving Equation 9: finding a representation of $\tau^{\uparrow\downarrow}$.

We determine the spectral structure of the atmospheric slant optical thickness $\tau^{\uparrow\downarrow}$ from non-vegetated pixels and then apply a crucial assumption. This assumption states that this $\tau^{\uparrow\downarrow}$ is equal to $\tau^{\uparrow\downarrow}$ over vegetated pixels. *Leth* [2014] shows the dependence of the SiF retrieval to the choice of this reference set.
Of course, we know that this assumption is not true. By definition, the atmospheric column over a non-vegetated pixel is not the same as the atmospheric column over any other pixel. But trying to determine the atmospheric optical thickness over a vegetated area would cause the inclusion of SiF. This leads to a problem as $\tau^{\uparrow\downarrow}$ and F_{S} cannot be determined independently.
This topic remains a challenging part of modeling the atmospheric slant optical thickness.
In this study, we use a whole set of non-vegetated spectra in order to determine this $\tau^{\uparrow\downarrow}$. Many spectra are needed in order to encompass the possible variability in atmospheric optical thickness on different locations and at different points in time. A set of spectra is including non-vegetated pixels over the Sahara. 365 days of these spectra are collected and combined into a *reference set*.

[3]In *Joiner et al.* [2013] a σ_h is used of 21.2 nm. Recalculating cited literature however, reveals a possible error in the calculation or the writing of this value.

2.2 Determining PCs

To find a representation of $\tau^{\uparrow\downarrow}$ in the reference set, we follow *Joiner et al.* [2013] and write $\tau^{\uparrow\downarrow}$ as a linear combination of principle components (PCs):

$$\tau^{\uparrow\downarrow}(\lambda) = \sum_{n=0}^{k} c_n x_n(\lambda). \tag{13}$$

Here c_n are the coefficients of the empirically determined orthogonal functions (EOFs) x_n.

Principle Component Analysis (PCA) is frequently applied as empirical method in the field of remote sensing fluorescence retrieval [*Shlens*, 2014]. Using traditional methods such a Singular Value Decomposition or direct $X^T X$ diagonalisation is computationally demanding when dealing with large numbers of spectra.
In order to efficiently determine the the EOFs of a reference set (see section 2.4.1), this study uses the Non-linear Iterative Projections by Alternating Least-Squares (NIPALS) algorithm, developed by Herman Wold in 1966 [*Esbensen et al.*, 2001]. This algorithm calculates the PCs one at a time in an computationally efficient manner.

NIPALS is explained following *Esbensen et al.* [2001], it contains the following six steps:

- To start the process, a multidimensional matrix $\mathbf{X}$ (in our case an array of reference spectra with in each row a single measurement of the spectrum and in each column a single wavelength bin) is centered and normalized.
  Initially, for iteration parameter f=1; $\mathbf{X}_\mathrm{f} = \mathbf{X}$
- For $\mathbf{t}_\mathrm{f}$ choose any column in X to be the initialization t vector.
- $\mathbf{p}_\mathrm{f} = \frac{\mathbf{X}^T \mathbf{t}_\mathrm{f}}{|\mathbf{X}^T \mathbf{t}_\mathrm{f}|}$, is calculated for each iteration f.
- $\mathbf{t}_\mathrm{f} = \mathbf{X}\mathbf{p}_\mathrm{f}$
- Check convergence: if $|\mathbf{t}_{\mathrm{f,new}} - \mathbf{t}_{\mathrm{f,old}}| <$ some criterion, stop; else go to step 2. Convergence usually occurs within 20-40 iterations.
- $\mathbf{X}_{\mathrm{f}+1} = \mathbf{X}_\mathrm{f} - \mathbf{t}_\mathrm{f}\mathbf{p}_\mathrm{f}^T$. As long as the maximum amount of PCs is not reached, update the matrix and proceed to the next iteration.

The principle components are calculated in descending order of the variance they explain. Following *Leth* [2014], who empirically determined the amount of PCs resulting in the best SiF retrieval, we set the amount of PCs for this study to f=35. It must be noted that determining the amounts of PCs is a difficult topic which is still under investigation (e.g. *Köhler et al.* [2015]). Using 35 PCs means that to solve the forward model (Equation 9), a total of 41 fit parameters is required: 35 will be determined by PCs and describe the two-way atmospheric transmittance, 5 parameters are used to describe the surface albedo and one to describe the fluorescence emission.

2.3 GOME-2A characteristics

This study uses the data produced by the Global Ozone Monitoring Experiment-2 (GOME-2A) spectrometer flying aboard the MetOp-A satellite (launched on 19 October 2006) which is part of the European Meteorological Satellite series.
MetOp-A flies in a sun-synchronous polar orbit at a height of 817 km with a local equator overpass time of 09h30 in its descending node. GOME-2A has four main channels that measure radiance between 240 and 790 nm at a resolution of 0.2-0.4 nm. For our purpose, only the band-4 data is used, ranging from 593 nm to 790 nm.
A scan mirror enables across-track scanning in nadir with a swath width of 1920 km. This allows the instrument to achieve global coverage in 1.5 days. It takes 6 sec. for the mirror to go forwards (east to west, 4.5 sec.) and backwards (west to east, 1.5 sec.). As it goes back three times as fast, the ground pixel size is much larger and so we only use data from the forward scan. Pixel size depends on the satellite velocity and the scan integration time. With an integration time of 0.1875 sec., the along-track pixel size is 40 km and the cross-track pixel size is 80 km, the spatial resolution is thus 40 km x 80 km [*Callies et al.*, 2000].
This has been the nominal operation mode for GOME-2A until 15 July 2013. After that, the swath width has been reduced to 960 km resulting a pixel size of 40 km x 40 km. Global coverage is now achieved in 3 days.

Space borne instruments like GOME-2A suffer from radiometric degradation. To adjust for this, a solar irradiance spectrum is measured by GOME-2A every day which is then used as the $E(\lambda)$ term in Equation 6 for each retrieval that day. In this way, $E(\lambda)$, and hence $F_s(\lambda)$ both degrade in the same way which compensates for most of the radiometric degradation.

2.4 Retrieval setup

2.4.1 Selecting reference pixels

As explained in section 2.1, PCA has to be applied to a set of non-fluorescent reference spectra in order to determine the PCs describing $\tau^{\uparrow\downarrow}$ (Equation 13).
This study selects the Sahara desert over ocean or ice areas as a reference area as this area yields the best retrieval results [*Leth*, 2014]. The Sahara is a large barren area, which is important to minimize residual fluorescence signals in the large GOME-2A footprint. Moreover, a year of reference spectra is selected to include the strong seasonal variability in the H_2O vapor column over the Sahara, since water vapor absorption is included in our spectral window.
A rectangle is defined (16°N-30°N, 8°W-29°E) to represent the Sahara area in the retrieval algorithm. Any patches of vegetation in this area are filtered out using the United States Geological Survey global land cover characteristics database v2.0[4].

Scenes with abundant cloud cover cannot be used for the SiF retrieval. Cloud data is supplied by the Fast Retrieval Scheme for Clouds from the Oxygen A band (FRESCO), which is an operational product from the same GOME-2A instrument. This scheme combines cloud optical thickness and cloud cover fraction into a single parameter: *effective cloud fraction* (F_c) [*Koelemeijer et al.*, 2001]. Following *Joiner et al.* [2013], we discard data if $F_c \geq 0.4$.

Data that passes the cloud selection criteria is collected for each day, building a data set from 2007 until present. From this set, an algorithm collects a reference set comprised of 365 days in such a way that the last day of the reference set is aligned with the last day of the retrieval month. In essence, the reference set *walks* with the retrieval moving forwards. For example, a retrieval of July 2009 yields a reference set ranging from 01-08-2008 until 31-07-2009. This way of reference set selection is needed for near real-time SiF retrieval. Fig. 7 illustrates the 'walking' motion of the reference set along with the retrieval time.
In practice, some days are skipped due to missing data. When this happens, the missing days are excluded and the reference set time range extends back until 365 days are collected. Collecting such a large number of reference spectra allows the PCA to account for much of the variance caused by seasonal effects (e.g. solar zenith angle, H_2O vapor column, etc.).

For building a reference set, also other non-vegetated areas could have been used such as data over highly clouded oceans and snow- and ice-covered surfaces can be used [*Joiner et al.*, 2013]. Moreover, *Joiner et al.* [2013] builds a reference set using a single day.
Snow or ice covered surfaces are avoided as well. These surfaces tend to be located close to the poles and are therefore geographically already very different from vegetated areas. Moreover, SZAs on ice-covered surfaces are high on average, whilst the SZA over fluorescent areas is much lower on average.

[4] http://edc2.usgs.gov/glcc/tablambert_af.php

2.4.2 Selecting target pixels

This study uses a spectral window ranging from 712 nm to 780 nm. This relatively broad window covers spectral features like the O_2-A band (∼760 nm) and part of the H_2O absorption band (approximately 715 nm - 740 nm). The next step is to select vegetated pixels. Selection criteria are loosely based on *Joiner et al.* [2013].

Pre-retrieval filters:

- Pixels with a FRESCO determined effective cloud fraction smaller than 0.4 are excluded.
- Pixels flagged by FRESCO for snow or ice scenes are excluded. FRESCO detects snow/ice scenes if the surface albedo is ≥ 0.8 at 772 nm or ≥ 0.2 at 360 nm [*Wang et al.*, 2010].
- If there is a mismatch between the pixel time index of FRESCO and the level 1b product, the filter is applied.
- Pixels within the sun glint cone (half of top corner: 18°) are excluded.

Post-retrieval filters:

- Pixels with a SZA of more than 70° are excluded.
- Pixels with a spectral residual RMS of more than 1% are excluded as they are assumed to be noisy. This criterion filters out a lot of the affected areas. However, extremely noisy areas like the South Atlantic Anomaly [*Casadio and Arino*, 2011] still show up in the results.
- Pixels for which the retrieval takes a very high number of iterations (more than 4200) are excluded. Convergence usually occurs within five iterations.

2.4.3 Baseline settings

Table 3: Summary of the baseline retrieval parameters

Parameter	Value
Cloud fraction threshold	≤ 0.4
Ice	Absent
FRESCO data	Available
Sun glint cone threshold	$\geq 18°$
SZA	$< 70°$
Spectral window	712 - 783 nm
Length of reference time frame	365 days
Number of PCs	35
Surface reflectivity (vegetated)	4^{th} order polynomial
Surface reflectivity (barren)	2^{nd} order polynomial
SiF Gaussian fit	$\mu_h = 737$ nm, $\sigma_h = 33.9$ nm
Spectral residual RMS	< 0.01
Retrieval convergence threshold	4200 iterations

2.5 Retrieval method for GOME-2A

The retrieval method can be divided into three categories; Reference set selection, retrieval and post processing, resulting in a level 3 SiF product. A summary is presented based on *Leth* [2014].

Reference set selection:
From the GOME-2A instrument, level 1b data is used as input. From this data, barren classified spectra from the Sahara desert are selected and aggregated towards a barren spectra database. From this database, a year of fluorescence spectra is selected depending on which time period is requested for retrieval. From the selected data, PCs are determined by the NIPALS algorithm and the 35 most significant PCs are selected to describe the atmospheric transmittance in the forward model (Equation 9).

Retrieval:
With the atmospheric transmittance component in place, pixels that contain fluorescence are filtered for e.g. clouds and SZA. For each pixel, a cost function is built from the reflectance, the instrument error and the forward model. The function is minimized using the Levenberg-Marquardt scheme. From the minimization procedure, a set of fit parameters is obtained to describe the fluorescence peak, the coefficients of the surface albedo polynomial and the coefficients of the atmospheric transmittance PCs [*Leth*, 2014].

Post processing:
The post retrieval filters are applied and the fluorescence peak is determined for the final SiF retrieval. Center coordinates of GOME-2A pixels are regridded to a regular 0.5° x 0.5° grid box. Fluorescence values are then averaged over each pixel within this grid cell and daily and monthly values are calculated resulting in the final level 3 SiF product.

2.6 Significant events for GOME-2A

Using satellite data over such a long time span is challenging. Continuous work is being done to ensure continuity and consistency of the data. However, on some occasions this is not possible and we have to manually adapt the retrieval process. The most important events for this retrieval are described below. One ought to keep this in mind when interpreting the data.

07-09-2009 through 11-09-2009:
Throughout its lifetime, an absorbing layer is built up on the optical components causing extra radiometric degradation. To solve this issue, the instrument was heated to evaporate this absorbing layer. During this five day de-icing procedure no data was recorded. After de-icing, radiometric degradation was slowed down as one of its causes was eliminated. However, the procedure caused a jump in radiance levels. A detailed description on this process can be found in *Dikty and Richter* [2011].
To compensate for this effect, the retrieval algorithm has started a new reference set right after this event happened (see Fig. 7).

15-07-2013:
The launch of MetOp-B causes MetOp-A to no longer be the primary satellite. The swath width of the MetOp-A GOME-2A instrument has been reduced to 960 km resulting in a cross-track pixel size of 40 km, the spatial resolution becomes 40 km x 40 km.

17-06-2014:
The level 1B data product provided by EUMETSAT is processed from the raw satellite data (level 0) by a so called processor. The processor version used for this study is 5.3 for which the data ranges from 17 January 2007 until 17 June 2014. We continued with data from version 6.0 because EUMETSAT states that no updates to the calibration have been made[5].
This change affects both the level 1b data as well as the FRESCO data, but the SiF retrieval remains the same. Consequently, we can proceed with the new 6.0 processor.

23-02-2015 10:00:
Fresco data is no longer available for processor version 6.0-0.06-1.28. The rest of the day is classified as faulty data. From 24-02-2015 onward, version 6.0-0.07-1.32 is used.

[5]http://www.eumetsat.int/website/home/TechnicalBulletins/GOME2/DAT_2217717.html under 'Release of version 6 of the level 0 to 1B processor'

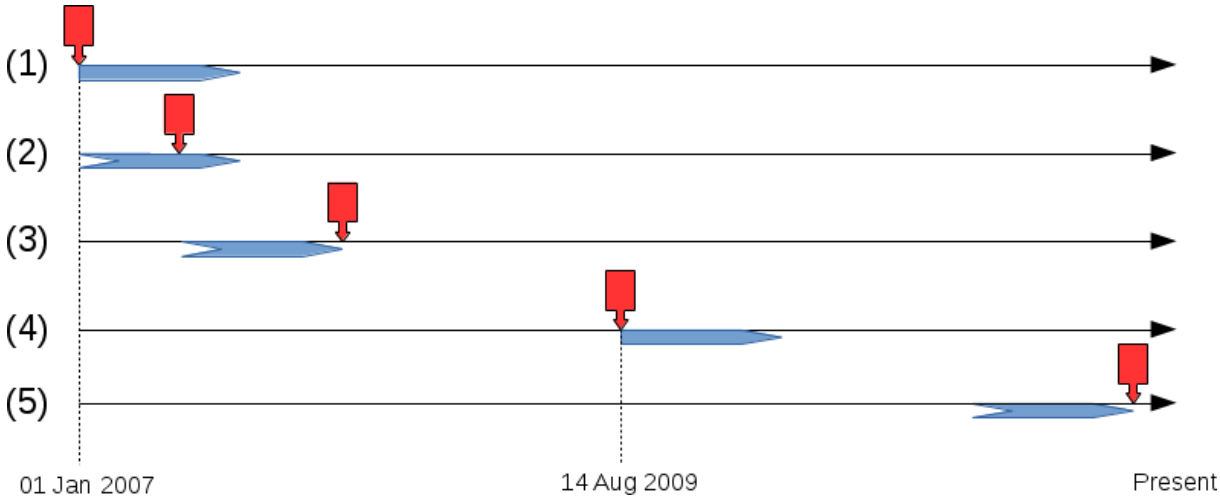

Figure 7: Time line showing the reference set (blue) selection as a function of the retrieval month (red). (1), no data prior to retrieval month is available, the reference extends one year into the future. (2) and (3): A year prior to the retrieval date is available and will be selected. (4) The de-icing maneuver caused an inconsistency in the reference set, a new starting point has been selected at August 14^{th} 2009. (5): The retrieval continues to run using the past year as a reference set.

3 Retrieval results

Using the baseline parameters as listed in Table 3, the retrieval algorithm is applied to GOME-2A data covering the period January 2007 to June 2015, producing level 2 data. The reference set is composed of 365 days of Sahara spectra, and trails behind the retrieval month as illustrated in Fig. 7. On the level 2 data, the post-retrieval filters are applied, the results are binned to a 0.5° x 0.5° latitude-longitude grid and daily and monthly averages are computed to produce the final level 3 SiF maps (e.g. Fig. 8 & 9).

3.1 Baseline GOME-2A SiF retrieval

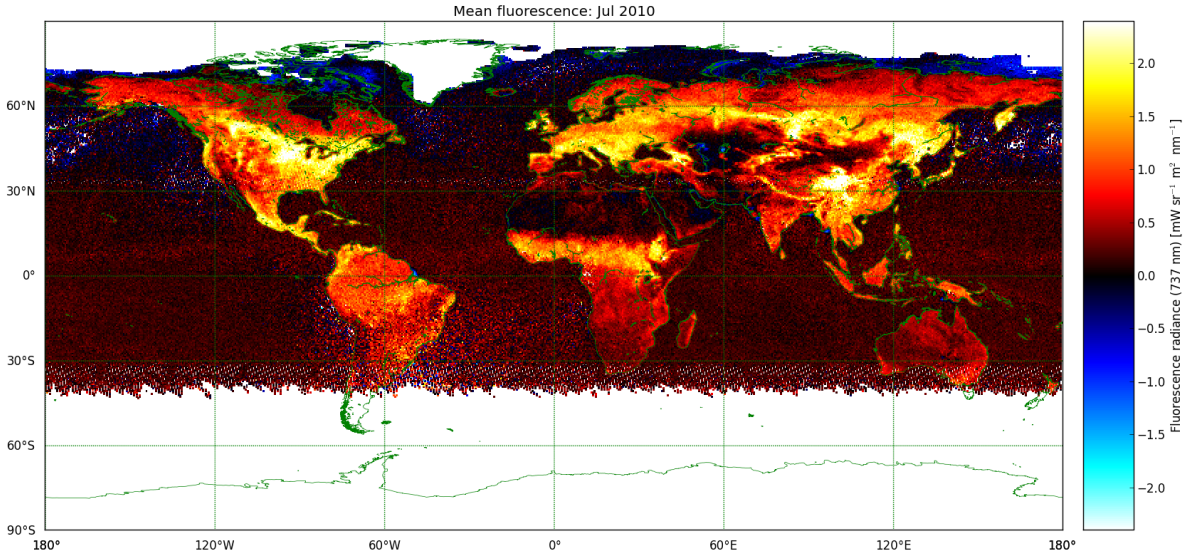

Figure 8: SiF retrieval at 737 nm in mW sr^{-1} m^{-2} nm^{-1} using the baseline retrieval parameters. For July 2010, summer in the Northern Hemisphere (NH) is clearly visible in yellow. Bins for which no pixel passed the data selection are shown in white.

Fig. 8 shows a monthly averaged SiF retrieval for July 2010. Retrieved SiF ranges -3.7 to 5.0 mW sr^{-1} m^{-2} nm^{-1}. However, in non-vegetated areas (SiF $<$ 0.35 mW sr^{-1} m^{-2} nm^{-1}, the mean of all retrieved values), the retrieval yields a mean value μ_{nv} of 0.32 with a standard deviation σ_{nv} of 0.4 mW sr^{-1} m^{-2} nm^{-1}.
Over vegetated areas, $\mu_{\mathrm{v}} = 1.0$ with $\sigma_{\mathrm{v}} = 0.4$ mW sr^{-1} m^{-2} nm^{-1}.
In July, the sun is located above the Northern Tropics and no retrieval results are shown at high latitudes due to the selection criterion on the SZA (see Table 3).
Other white areas (e.g. in Canada and above India) can be attributed to snow or ice cover. On some occasions, missing data can be found over the ocean. This is usually caused by extensive cloud cover where certain pixels are shielded by clouds for a whole month. Such particular areas can be seen over the North Pacific Ocean in Fig. 8.

3.1.1 Spatial patterns

In the NH summer (Fig. 8), the highest SiF values can be found in the southern parts of China and throughout the Russian forests, as well as in the eastern United States and Canada.

Now we turn to the Southern Hemisphere summer. Fig. 9 shows the SiF retrieval for January 2010. The Northern Hemisphere is mostly dark as SZAs are high, and vegetation is dormant. High SiF values can be found in Southern Africa and the South American subtopics (note that the green color indicates coastlines).
Distinct SiF patterns are visible in southern Chile, along the coast of Australia and in New Zealand.

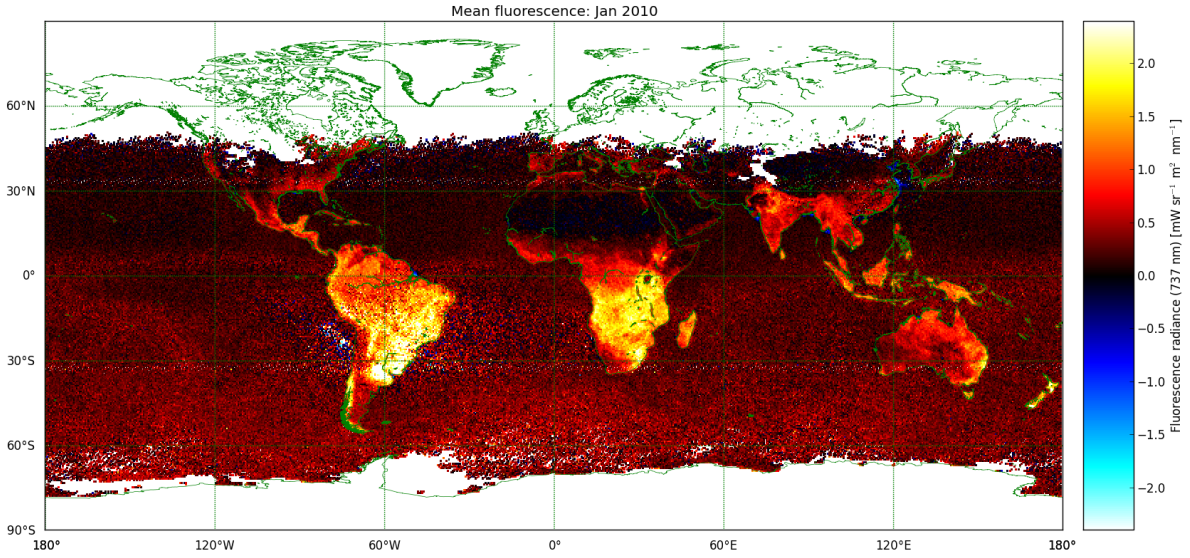

Figure 9: January 2010 shows typical SiF spatial patterns for the NH winter.

Over the ocean, the SiF values are expected to be around zero, except for some random statistical noise. Indeed, hardly any structure is visible. Also for this month, values over the ocean are of the same order as in Fig. 8. Similarly low values can be found on other non-vegetated areas such as the Sahara Desert, the Saudi Arabian Peninsula, the Middle East, Kazakhstan, the Taklamakan Desert (China), the Rocky Mountains, the Australian Outback and the South American Andes mountain range in Chile and Argentina. Although the latter seems to be partially obstructed by noise.

In the South American region, the Earths Inner Radiation Belt reaches deep into the atmosphere. Highly energetic particles can therefore penetrate into low earth orbit altitudes. These particles can collide with the GOME-2A spectrometer and cause instrument noise. This phenomenon is known as the South Atlantic Anomaly (SAA) [*Casadio and Arino*, 2011] and is visible in Fig. 8 as an oval shaped region above South America, centered around 25° South and 50° West. The SAAs interference causes random noise to increase in the retrieval. During

summer in the SH, this phenomenon is less clearly distinguishable. This is likely caused by a higher number of photons hitting the detector, reducing statistical noise.

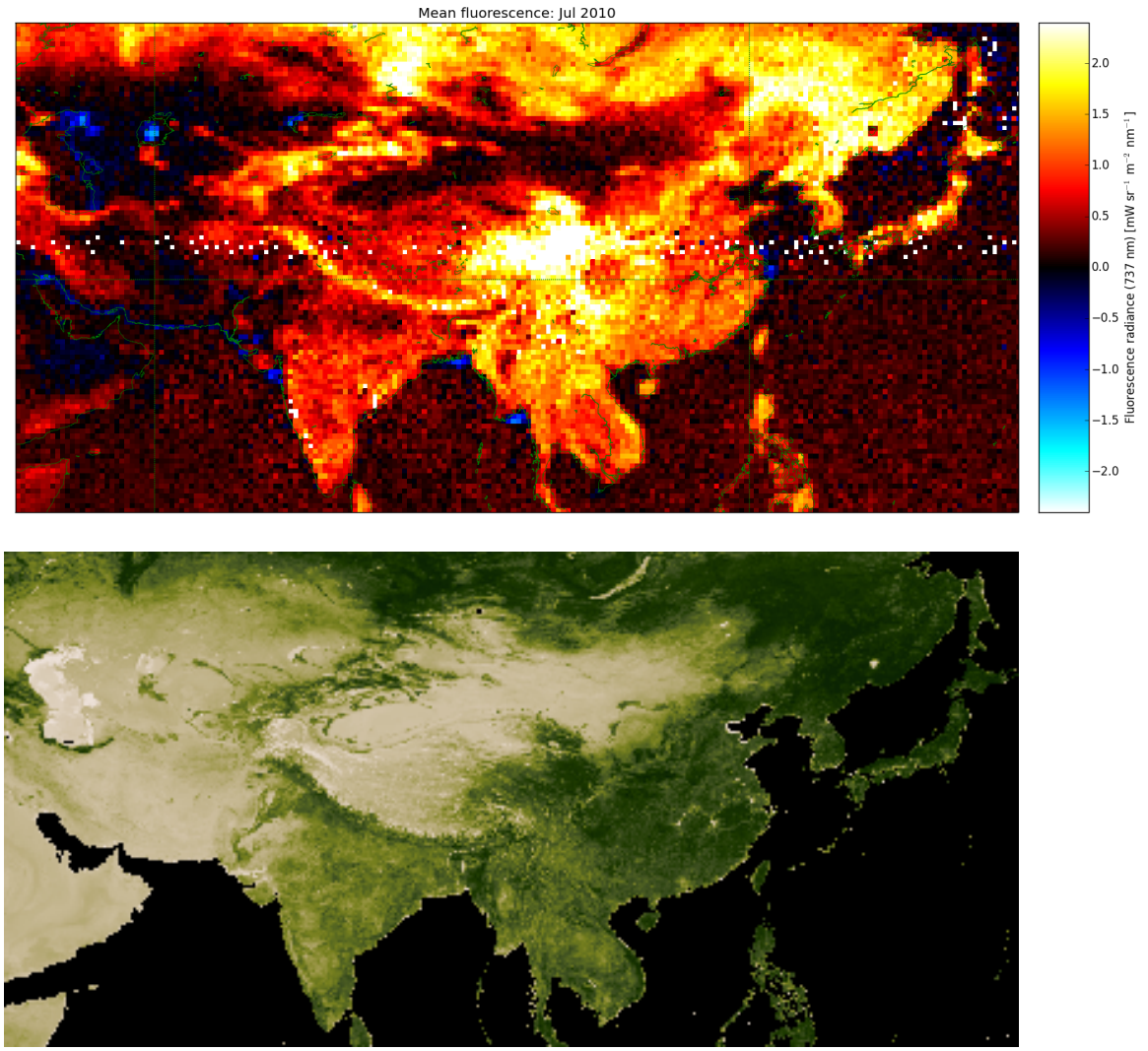

Figure 10: SiF retrieval in Asia compared with the NASA Blue Marble NDVI image for the same region. For SiF, white pixels indicate no data, whereas black pixels indicate no NDVI data. The band of white pixels around 30° North is due to a daily switch to calibration mode for a short period of time.

Zooming in on the Asian continent, more spatial patterns becomes visible as shown in Fig. 10. The SiF map is compared with the Blue Marble: Next Generation NDVI image (BMNG; *Stöckli et al.* [2005]) for the Asian region. The BMNG-NDVI image series is created using NASA Terra MODerate resolution Imaging Spectroradiometer (MODIS) data for 2004. Similarities are abundant when comparing the spatial patterns of the SiF retrieval with the green patches of vegetation on the NDVI images.

The north side of the Elbrus Mountains, which form the border between the Caspian Sea and the Iranian Plateau, shows a strip of moderate SiF. A more prominent feature in the SiF retrieval is the southern slope of the Himalayas. It can be clearly distinguished as an arc on the border between India and Tibet. On the eastern side, it ends in an area with SiF of over 2.0 mW sr^{-1} m^{-2} nm^{-1}. This is the location of the Sichuan Bamboo Forest, which is a fast growing plant species with high photosynthetic efficiency [*Kumar et al.*, 2002]. The southeast of Russia shows similarly high SiF in vegetated areas as well as the mountainous region of Kyrgyzstan.

Sparse vegetation is found in the rocky desert of Central Asia. Only the Amu and Syr rivers provide enough water for vegetation and can be seen respectively to the south and to the east of the Aral Sea as small strips of SiF, although these regions are hard to distinguish on the NDVI picture.
The eastern part of China is a mostly rocky arid area and yields hardly any vegetation. This low SiF region stretches all the way into Mongolia. The oval shape of the Taklamakan sand desert is clearly visible in the south-western part of this area.

A number of blue spots break the overall red glow of the picture. Apparently some regions show negative SiF values. The north of the Caspian Sea, the Aral Sea, Lake Balkhash in east Mongolia, the mouth of the Padma River in northeast India, the Gulf of Martaban and the Hangzhou Bay in Shanghai. These areas show up on the BMNG-RGB map as cyan regions outside the country borders. The effect indicates that the assumed model cannot fit the observed radiative spectrum in these locations well. The scenes probably contain characteristics that show spectral features inversely proportional to the assumed fluorescence spectral shape. This is possibly caused by absorption and reflection characteristics of shallow water or suspended sediment [*Leth*, 2014].

3.1.2 Temporal patterns

Apart from spatial patterns in SiF, the long operational period of GOME-2A makes it possible to analyze its temporal patterns in more detail, presented in Fig. 11. This figure shows SiF maps for every month of 2010. An overview of the entire GOME-2A mission retrieval can be found in Appendix A (Fig. 23 - Fig. 31).

Little to no SiF is present in the Northern Hemisphere during the winter months (November through February), only a background signal remains.
In February however, the African coast along the Mediterranean Sea starts producing the first SiF. During the following months, the SiF signal propagates into the European and Asian. The same phenomenon can be seen on the American continent. SiF signals grow all the way up to 60° North in August.

In August, SiF increases in the coastal regions of Australia which has strong SiF during the NH winter. Also, New Zealand and southern Chile, as well as a bit of Madagascar show this pattern.

Tropical regions show high variability throughout the year. Maximum SiF values range from roughly 20° south between December and February up to 15° north between August and October. This pattern follows the overhead sun.

The SAA is persistent throughout the year. This phenomenon reduces the accuracy of retrieval above the South American region, in particular the Amazon rain forest.

An interesting temporal phenomenon that does not have to do with vegetation but rather with a possible artifact of the retrieval method is shown in Fig. 12. Four real color images are shown of the Aral Sea (left) and SiF retrieval for the same area (right). The original size of the lake is indicated by a thin line. Over the past decades, this Sea has been suffering from severe desiccation because of extensive irrigation and other human activities since 1960. This has caused the water level of the Aral Sea to drop by 23 m and its area has shrunk by 74% so far [*Micklin*, 2007].
The SiF retrieval responds by reducing the negativity of the retrieval. It disappears over the period 2008 through 2014. In 2014, no negative SiF is left in the central basin. Notably, this negative retrieval only occurs for the right part of the basin, where the water is shallow. The water in the left basin is now strongly controlled by flood gates and dams and is therefore much deeper. No negative SiF values are found in the left basin, indicating that this phenomenon does not occur over deeper water.
The small patch of vegetation towards the south of the Aral Sea reduces its SiF signal as water becomes less and less available.

When looking at Appendix A, the retrieval for September 2009 shows significantly less SiF than for the next month. Also, the 2010 retrieval shows higher average values. The possible cause for this result is discussed in Section 5.

4 Retrieval Evaluation

4.1 Comparison: *Joiner et al.* [2013] SiF retrieval

As this study is based on a method by *Joiner et al.* [2013], a logical step is to compare the results with this previous work.

SiF retrieval data were made available for download on the NASA website[6]. For comparison, we use version 26 data of *Joiner et al.* [2013] (Jv26), which shows monthly averaged SiF on a 0.5° x 0.5° grid. For convenience, the data is represented in the same way as in this study.

4.1.1 Jv26 product description

The Jv26 SiF product characteristics:

- GOME-2A data.
- Fitting window of 734 nm - 758 nm.
- The same cloud and SZA filters are applied as in this study.
- The PCA is performed using daily reference spectra over highly cloudy ocean and snow- and ice-covered surfaces (ρ at 670 nm >0.7), as well as over the Sahara [*Joiner et al.*, 2013].
- Ocean data is included for monitoring the biases.
- Attempts are made to correct for a small zero-level offset problem in the previous versions.

The most important retrieval parameters are summarized in Table 4.

Table 4: A summary of the differences in retrieval settings between Jv26 and this study.

Variable	Jv26	This study
SiF peak	736.8 nm	737 nm
Width of SiF peak	21.2 nm	33.9 nm
Number of PCAs used	12	35
Fitting window	734 nm - 758 nm	712 nm - 783 nm
Reference spectra	1 day (Ice, Ocean and Sahara)	365 days (Sahara only)

[6]http://avdc.gsfc.nasa.gov/index.php?site=296832442&id=66&go=list&path=/GOME_F

Compared to this study, Jv26 uses a much smaller spectral fitting window. It does not include the O_2-A band and most of the H_2O absorption structures, just the spectral area in which hardly any molecular atmospheric absorption takes place (the *continuum*) [*Joiner et al.*, 2013, 2014, *Köhler et al.*, 2015].
Leth [2014] suggests that 25 PCs is the minimum amount for getting a valid SiF retrieval, but in the Jv26 retrieval, 12 PCs are used to describe the atmospheric optical thickness [*Köhler et al.*, 2015]. This difference is caused by the amount of PCs needed for an accurate description of the atmospheric optical thickness as function of the spectral fitting window. When O_2-A and H_2O bands are not included, the spectral structure within the fitting window is less complex and fewer PCs will suffice in describing it.
The choice of reference spectra is a completely different one. Accumulating spectra from an entire year gives a lot of variability in the reference state. However, a higher amount of reference spectra may not be representative for the retrieval state. Jv26 chooses to select only reference spectra from the retrieval day. This is thus a representative time period of the retrieval, however, in order to aggregate enough reference spectra, spectra are collected from an abundance of different scenes. This results in a different type of variety which could also hinder the retrieval.

4.1.2 Comparison with Jv26: spatial patterns

Fig. 13 shows a comparison between the two SiF retrieval methods for a single month, July 2010, to analyze the difference in spatial patterns.
At first glance, the structure looks very similar. In absolute value, the SiF signal from this study is slightly higher. Taking a closer look however, reveals subtle differences. Fig. 13c visualizes these differences.
The first major difference is seen in high altitude regions such as the Himalayan Mountain Range and its high altitude desert and forest lands to the north, the Rocky Mountains and the Andes Mountain Range. This study shows a much higher SiF signal in these high altitude regions compared with Jv26. Especially the Sichuan Bamboo Forest, also shown in Fig. 10 shows much higher SiF in this study. This spatial pattern is in accordance with high topography visible in the topography map of Fig. 14.

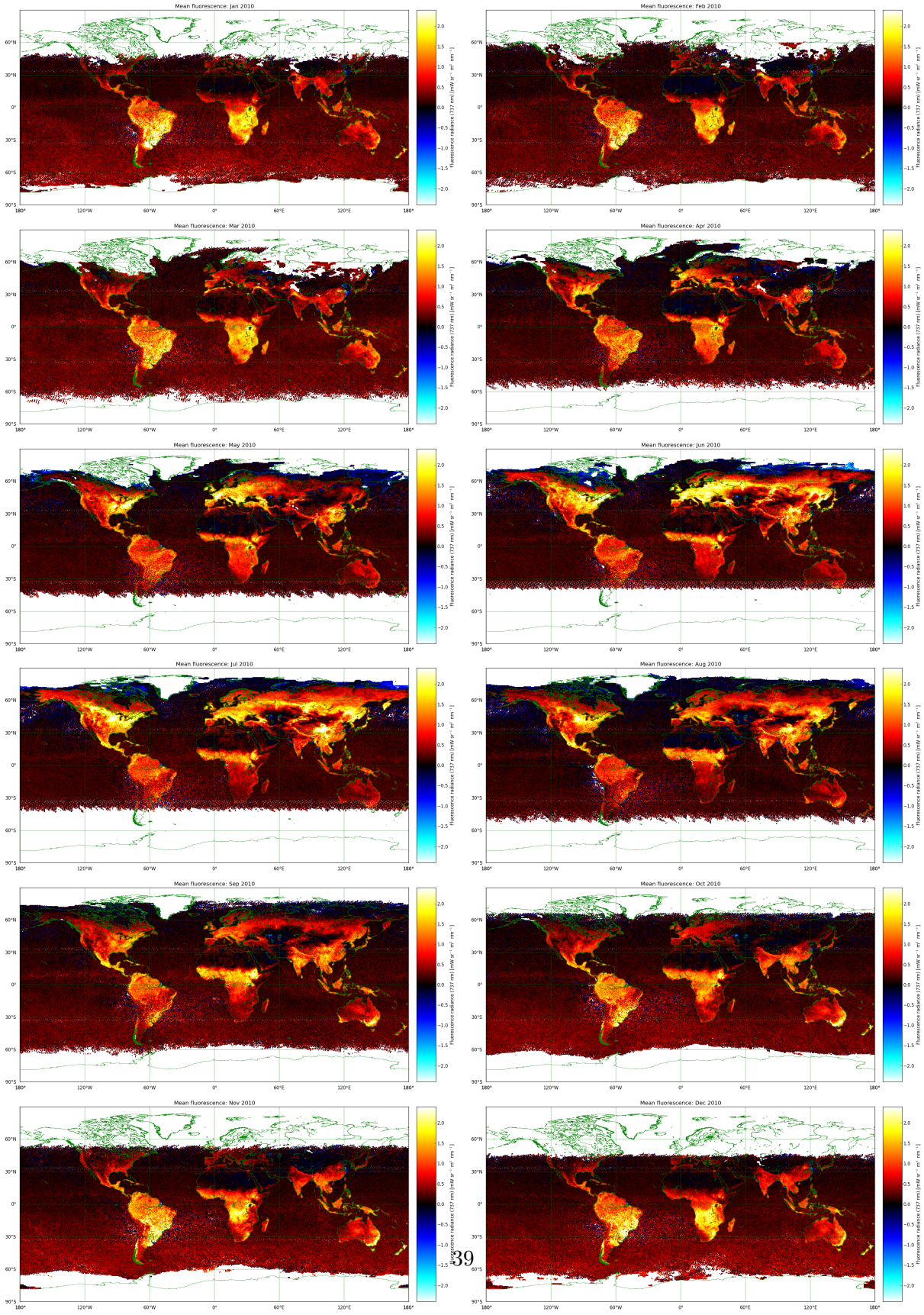

Figure 11: Monthly mean SiF retrieval for 2010.

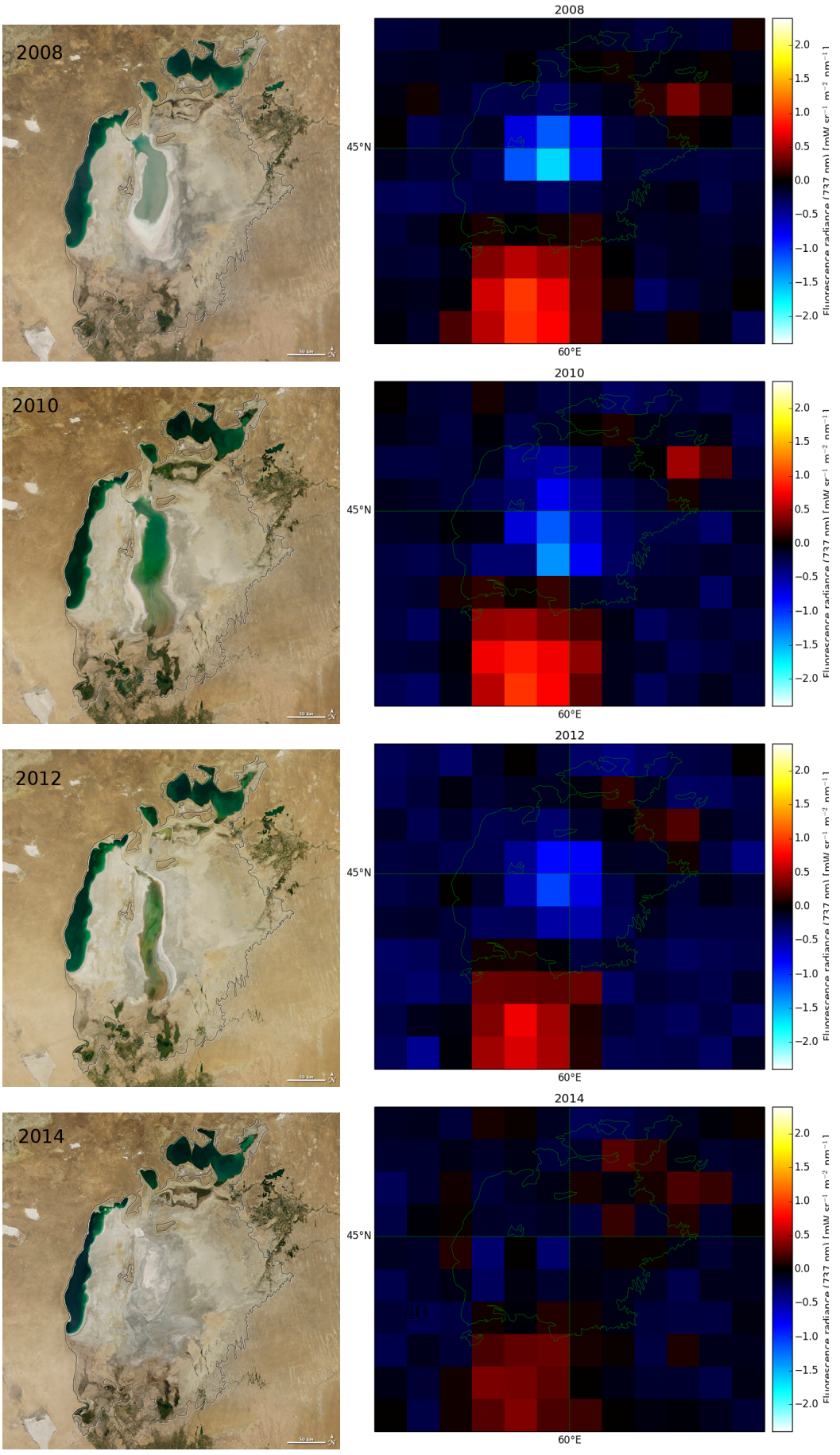

Figure 12: Real color image of the Aral Sea (left) and SiF retrieval (right). Credit: Rebecca Lindsey (NASA Earth Observatory)

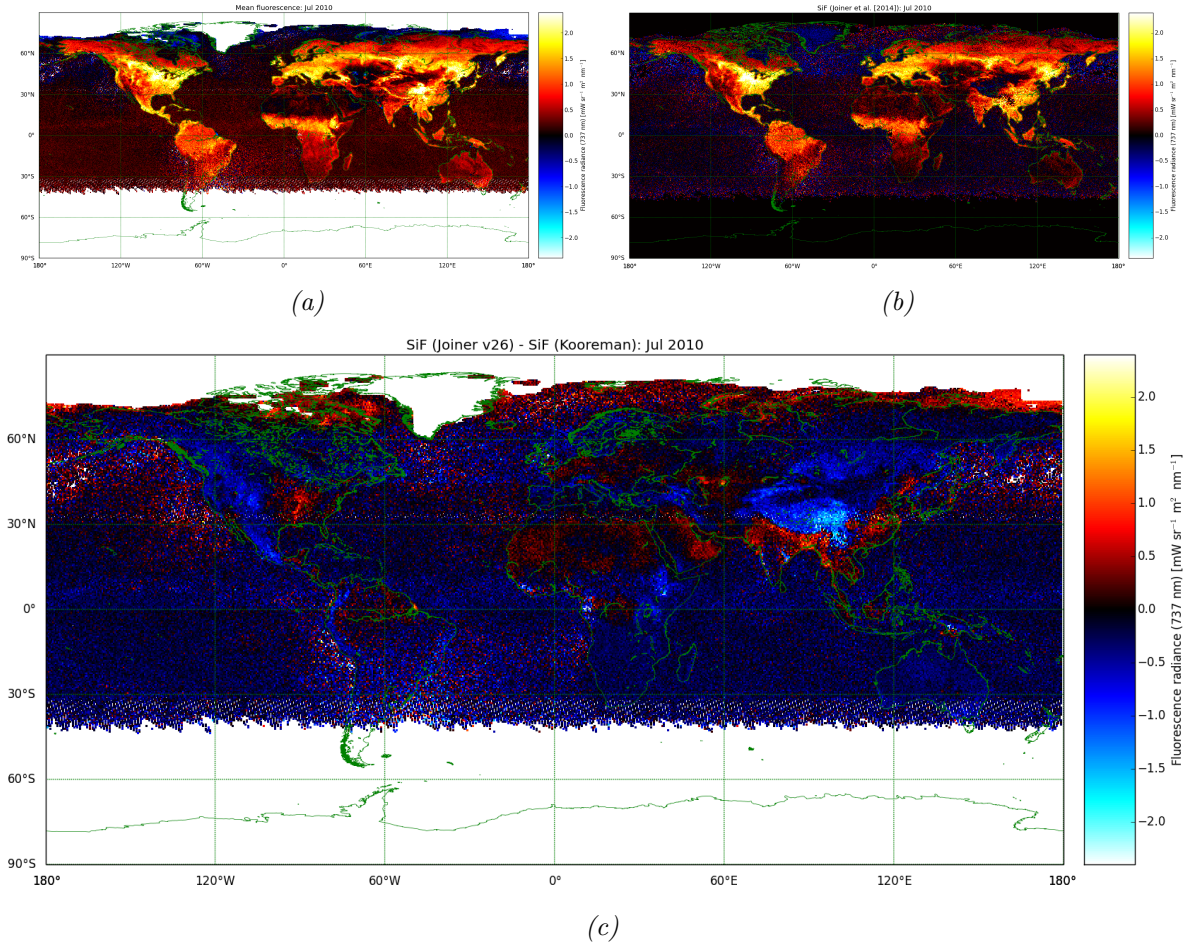

Figure 13: Comparison between a SiF retrieval from this study (a) and from Jv26 (b). The bottom figure shows a grid by grid difference between the retrieval of July 2010 by Jv26 and this study. Blue values indicate a higher value in this retrieval, red values indicate a higher value in Jv26.

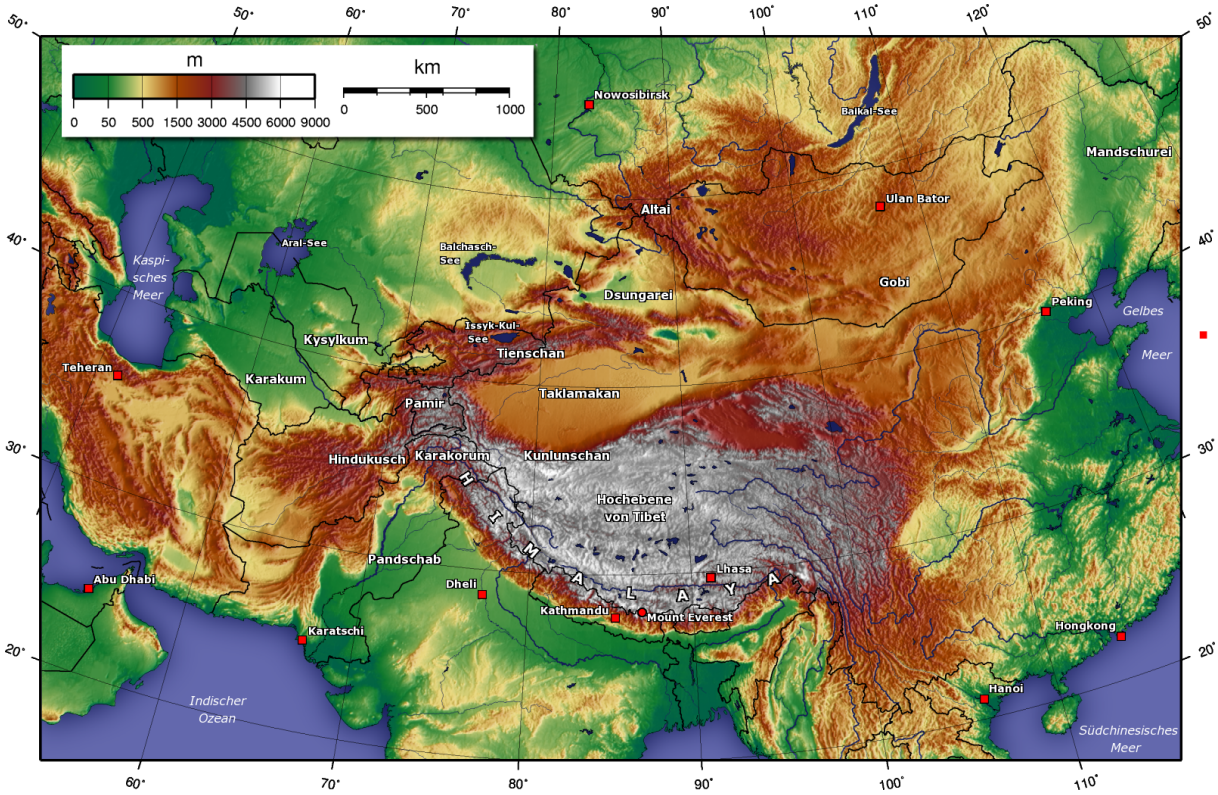

Figure 14: Topography map of Central Asia. Source: Worldofmaps.net

4.1.3 Comparison with Jv26: temporal patterns

Fig. 15 and Fig. 16 show a comparison between the SiF retrieval in this study and in Jv26. Monthly mean maps of the year 2014 are presented.

The temporal pattern looks very similar overall. Both years clearly show the seasonal cycle of vegetation. Fig. 15 & 16 suggest that in the Jv26 retrieval, SiF on the NH seems to set in earlier than in this study. The SiF signal starts intensifying over the European continent in April and May, whereas in this study, SiF is still mostly absent in April and reaches the same intensity roughly a month later. The same pattern is visible in the North American Continent. When looking at the end of the summer and the fall, we see a similar pattern. The Jv26 SiF retrieval shows a SiF signal in Europe and North America all the way through October. The SiF in this study is already faded back to the background level by October.

The Southern Hemisphere shows a different pattern. In the Jv26 data, South America hardly shows any change in SiF. Only late in the year it seems to become brighter in SiF emissions. This study however, shows a SiF growth starting in August, clearly engulfing the Southern Hemisphere only to weaken again from December onward. We elaborate this effect in section 4.3

There is a noticeable difference in retrieval pixels, especially in the winter season. The retrieval of this study shows little areas of missing data to the north of the Himalayas, in Central Asia and to the west of China. Also off the coast of Chile and the southern half of the African continent this missing data is visible. Jv26 data does not show any missing data within the retrieval area.

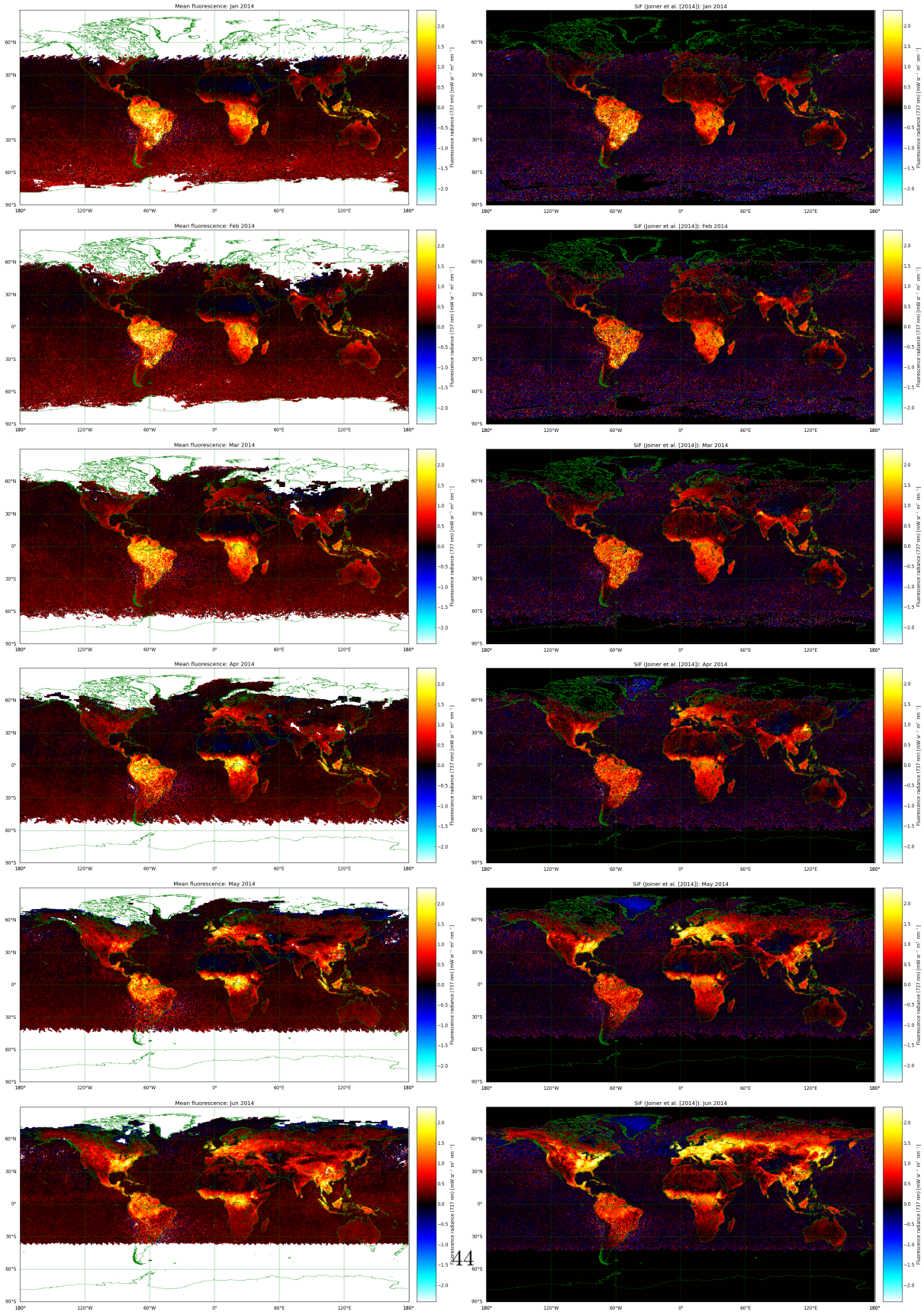

Figure 15: SiF retrieval from January through June 2014. This study (left) and Jv26 (right).

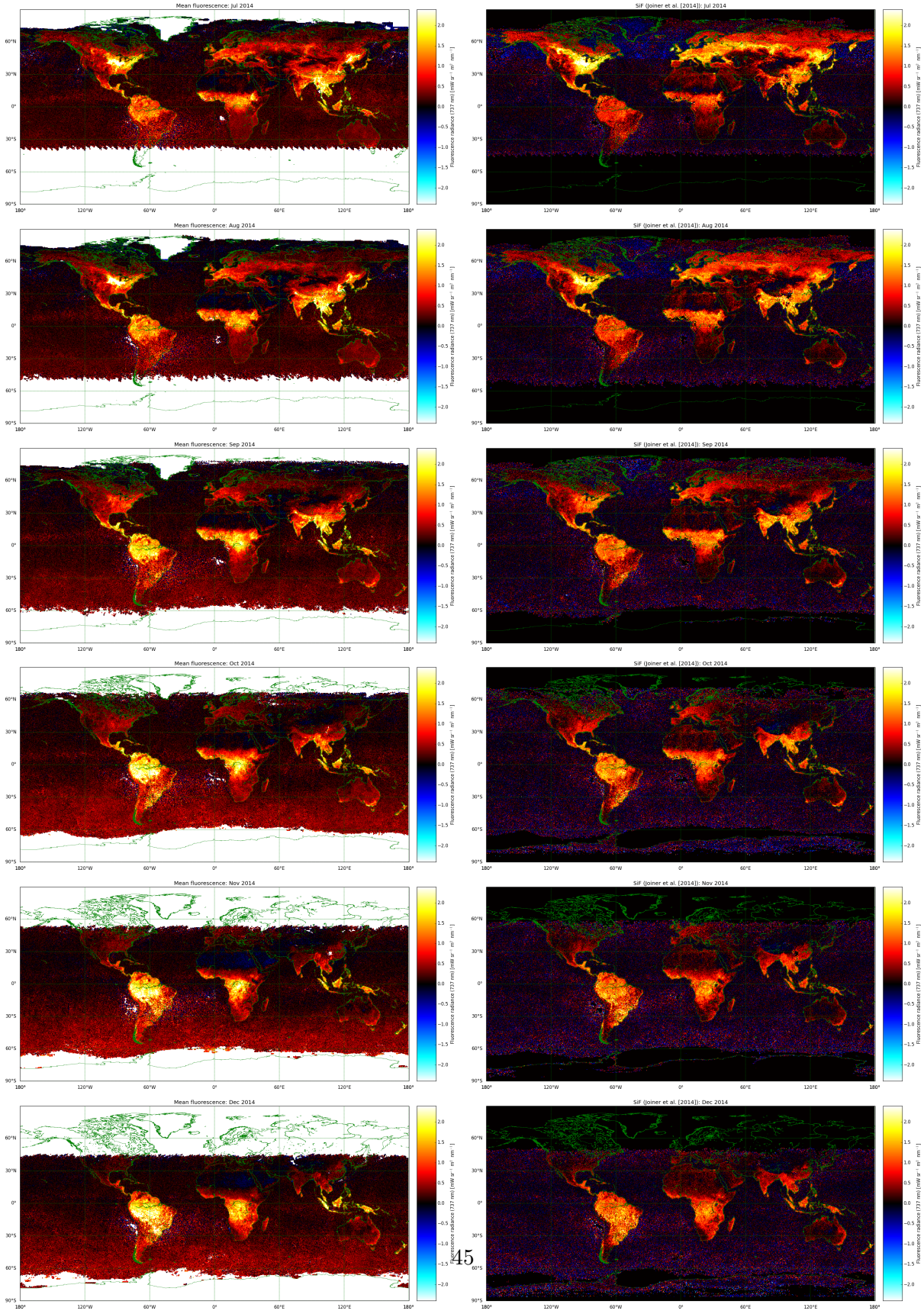

Figure 16: SiF retrieval from July through December 2014. This study (left) and Jv26 (right).

4.2 Correlating SiF to GPP

In order to evaluate the SiF retrieval, this study compares the GOME-2A SiF with ground based derived GPP observations from flux towers and SiBCASA model GPP data. For the entire duration of the GOME-2A mission so far, monthly mean SiF and GPP are shown in Fig. 17 for the United States Cornbelt (Fig. 4) area. The U.S. Cornbelt area is particularly useful for comparison with satellite data as the vegetation is very homogeneous. This reduces the representativeness errors between the localized flux tower measurements and the coarse GOME-2A footprint.

As explained in section 1 flux tower GPP is derived from CO_2 measurements. Flux towers are rather vulnerable to weather circumstances. Because of this, data is often incomplete or missing in the winter as instruments are frozen over or snow cover obstructs the flux measurements. SiBCASA is a computer model and will therefore always yield results although data was available from January 2007 through August 2012, so far.
To further enhance the data for comparison, the U.S. Cornbelt area daily and monthly average is calculated to further reduce the representativeness errors by taking U.S. Cornbelt averaged GOME-2A SiF and flux tower derived GPP. Results are plotted for the entire GOME-2A time series in Fig. 17.

Analyzing Fig. 17 in detail shows interesting aspects. First of all it should be noted that the correlation coefficients $R_{\text{SiF-tower}}$ and $R_{\text{SiF-SiBCASA}}$ are calculated based on all data points shown in the figures. This means that the seasonal variability present in all three data sets accounts for a part of this correlation coefficient. In order to determine whether this data set really includes inter-annual variability, the set would have to be deseasonalized. This is done in Section 4.5
It should also be noted that whereas the GOME-2A and SiBCASA data is provided on a continuous basis, the flux tower data density varies throughout the years. As all flux tower data are aggregated and some flux towers do not yield data for certain years, some years are covered by more flux towers than other years.

Keeping this in mind, we take a closer look at its temporal dynamics. First of all the GOME-2A data. U.S. Cornbelt averaged SiF values reached an average of 1.2 mW sr^{-1} m^{-2} nm^{-1} in the summer and 0.2 mW sr^{-1} m^{-2} nm^{-1} in the winter. The years 2010, 2013 and 2014 show periods of more than 1.5 mW sr^{-1} m^{-2} nm^{-1}. Also 2015 tends towards having a summer with high SiF values. The flux tower data does not show these dynamics. The mean GPP is approximately 12 μmol m^{-2} s^{-1} with notable low values in the summers of in 2008 and 2013. In the daily mean time series some outliers are observed in SiF values, including some negative values in the winter months of 2008, 2009, 2011, 2012 and 2015. Also in the summers of 2008 and 2010 some extremely high SiF values are shown.

The monthly mean time series gives us insights of the onset and offset period of the SiF. Notably, the periodic SiF signal is wrapped around the signal from the flux towers. This indicates that SiF emissions start before the onset and are continuing after the offset has already taken place.

This effect is consistent throughout the years.
When comparing this effect with the SiBCASA model data, we see that the onset season coincides much better, as the model predicts an earlier onset. The SiF offset is structurally behind by as much as a month.
This is probably also a reason why the SiF retrieval correlates better with the model data (R=0.93) than with the flux tower data (R=0.88).

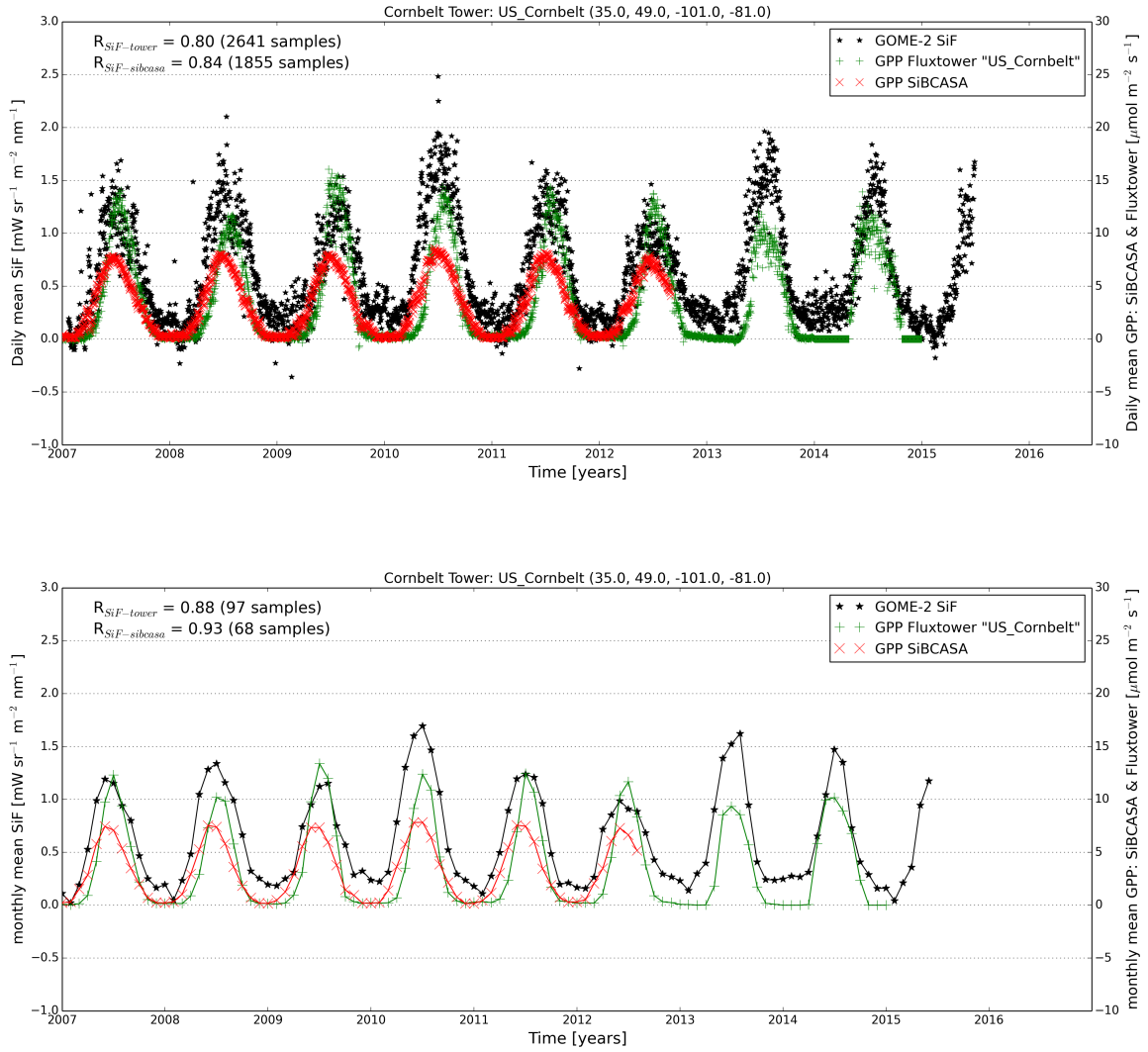

Figure 17: Time series of the U.S. Cornbelt showing area averages of the GOME-2A SiF retrieval (black), the flux tower derived GPP data (green) and SiBCASA GPP data (red). The top window displays daily averages, the bottom window displays monthly averages.

4.3 Jv26 time series comparison

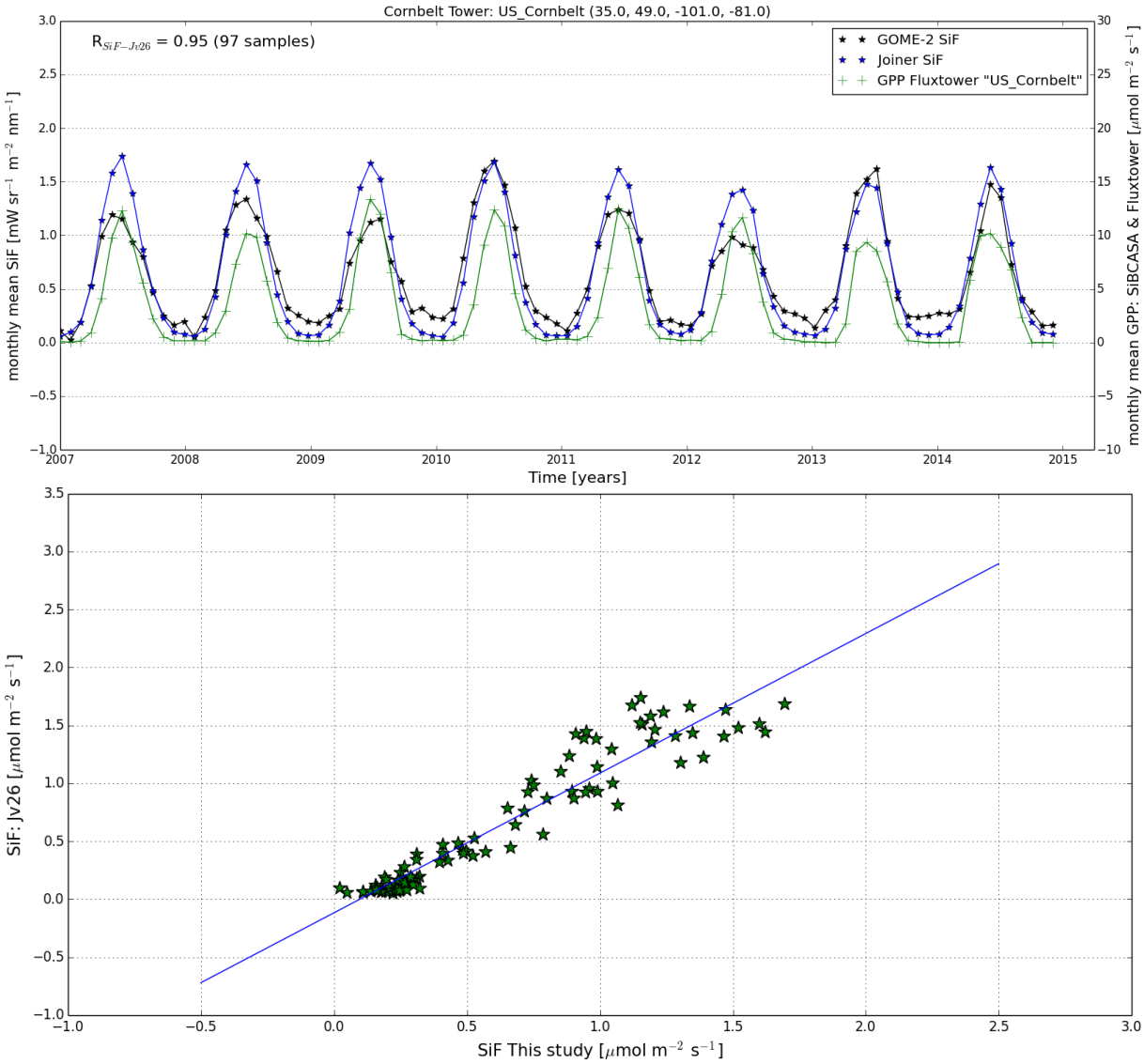

Figure 18: The top panel shows a time series from 2007-2015 for the Jv26 SiF retrieval (blue), the SiF retrieval of this study and flux tower GPP (green) in the U.S. Cornbelt. The bottom panel indicates the correlation between the two SiF retrievals.

As an additional comparison, Fig. 18 presents the two U.S. Cornbelt averaged SiF retrievals together with the flux tower GPP measurements. This allows for a more detailed analysis of the onset and offset periods. In section 4.1.3 we suggest based on the global SiF maps that there is a difference in onset and offset of approximately a month. However, Fig. 18 shows us that this is not the case at all. Yes, the absolute value in a particular month is often higher in the Jv26 retrieval but that does not mean that the onset and offset are different. In fact, when we

define the onset to be the first month in which the SiF value reaches above the winter noise value of 0.32 mW sr^{-1} m^{-2} nm^{-1} (section 3.1), these months are always identical, possibly with exception of the 2011 onset. We can take the same definition for the offset (first month where SiF $\leq$ 0.32) which also shows an identical offset in the two retrievals. Winter months show higher SiF values in the retrieval of this study compared to Jv26 which is possibly due to a bias correction in the Jv26 data.
We have to note that compared to the flux tower GPP time series, the SiF retrieval still show an 'envelope' with earlier onsets and later offsets.

The correlation between the two SiF time series, shown in the bottom panel of Fig. 18, is very high ($R = 0.95$). A linear regression leads to $\text{GPP}_{\text{jv26}} = 1.21 \times \text{GPP} - 0.12$.

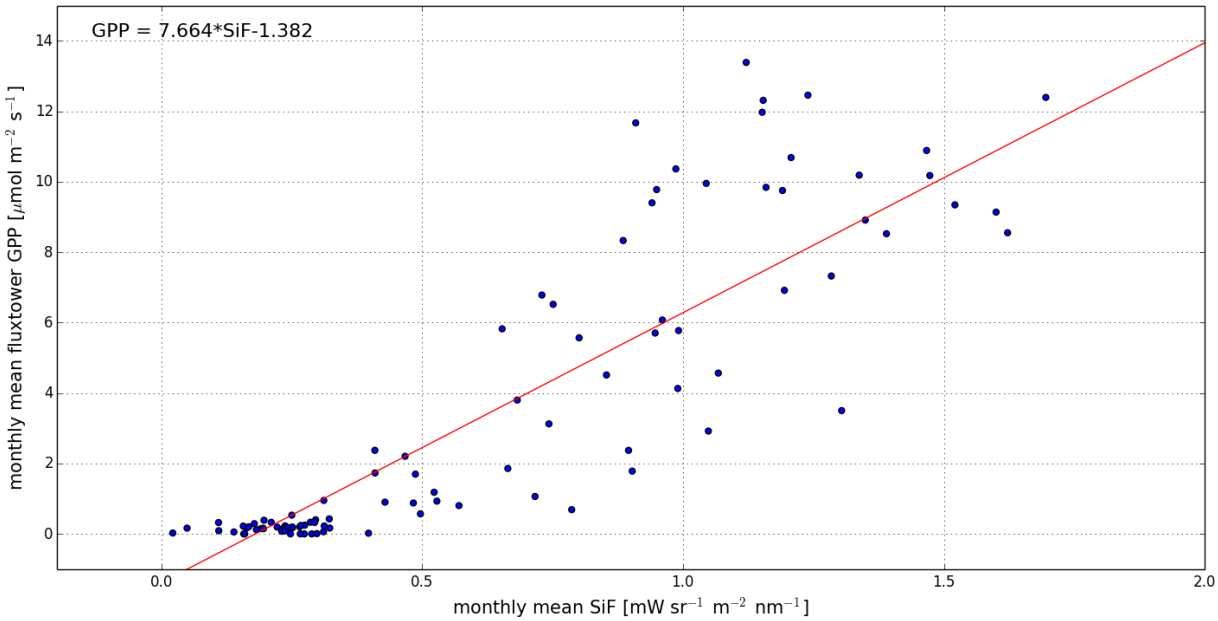

Figure 19: Scatter plot showing the relation between flux tower derived GPP and GOME-2A SiF in the U.S. Cornbelt for the entire operational period.

4.4 Relating SiF to GPP

Following *Guanter et al.* [2014], we determine an empirical linear relationship between the flux tower derived GPP and the GOME-2A SiF within the U.S. Cornbelt. This relation is shown in Fig. 19 and roughly corresponds to the literature which yields GPP$= 3.55 \times \text{SiF} - 0.88$ with a correlation coefficient of R=0.96. It must be noted that *Guanter et al.* [2014] only uses flux tower GPP for cropland sites, whereas this study averages over the entire U.S. Cornbelt. This results in a wider range of vegetation types covered within the data set and this leads to a lower correlation.

To take a closer look at the relation between SiF and GPP, results are shown for two particular flux towers.

The Morgan Monroe State Forest (MMS) tower, which is situated in a deciduous broad leaf forest and the Konza Prairie (Kon) tower located in an agricultural region. To compare with satellite and model observations, an area of 2° by 2° has been selected centered at the flux tower location. Mean values have been calculated for this small region and are presented in Fig. 20 and Fig. 21.

4.4.1 Forest

Morgan Monroe State Forest

The long sample time of the MMS flux tower gives a large sample size for calculating the correlation coefficients between SiF and the flux tower and SiBCASA GPP values as shown in Fig. 20. For monthly averaged values, correlations are significant (R $\sim$ 0.9). The spread in the daily measurements causes a lower correlation, which is then due to daily variation and noise.

The flux tower measurements (green) reveal the absence of vegetation activity in the winter, as GPP values are zero. The SiF retrieval however, does not yield values of completely zero. In March a strong increase in GPP can be seen. This increase is also shown by the SiF retrieval. However, the onset season is approximately one to two months early. The offset starts approximately in the same month grows to a steady value of about 10 μmol m^{-2} s^{-1} each year.

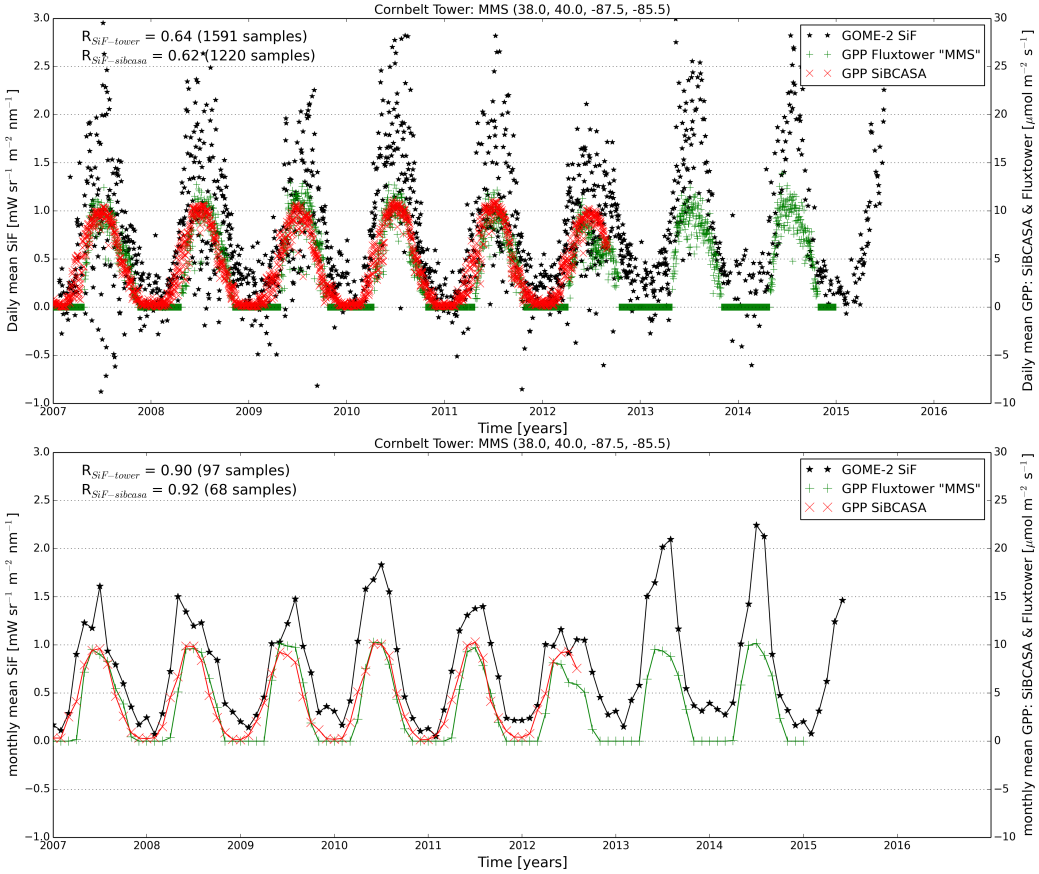

Figure 20: GPP and SiF time series for flux tower MMS. Daily means (upper panel) and monthly means (lower panel) for 2007-2015.

4.4.2 Agriculture

Konza Prairie

For the prairie tower in the agricultural area, the on- and offset seasons differ slightly. The onset of GPP starts in April and then increases within two months to its peak value (Fig. 21). This peak is more variable than for the forest area case with peak values ranging from 6 to 13 μmol m^{-2} s^{-1}.
The SiF retrieval shows a more regular pattern. The strong increase in the onset season is visible and again approximately a month behind the onset in the flux tower data.

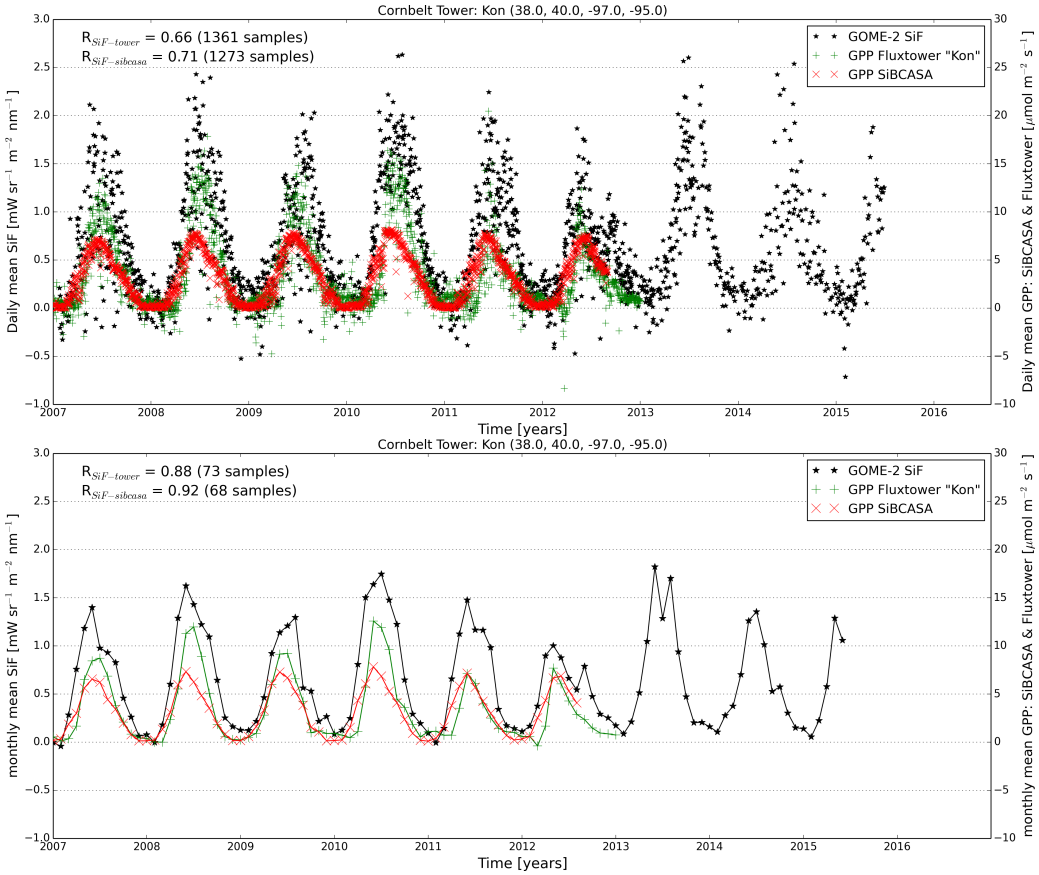

Figure 21: GPP and SiF time series for flux tower Kon. Daily means (upper panel) and monthly means (lower panel) for 2007-2015.

For both of the areas, SiBCASA seems to follow the pattern in a similar way as is the case in the entire U.S. Cornbelt region (Fig. 17)

4.5 Monthly SiF anomaly

In order to investigate the third objective of this study, relative anomalies have been constructed for each month during the GOME-2A mission. This is done by calculating the average of each month over the entire mission. Subsequently, a particular month is divided by this average and normalized, following Equation 14:

$$A_{\mathrm{X}} = \frac{\mathrm{SiF_X} - <\mathrm{SiF_X}>}{<\mathrm{SiF_X}>} * 100\% \tag{14}$$

Where A is the anomaly, x represents a certain month and the <> notation indicates the mean over all years of month x. This process is repeated for GOME-2A SiF, flux tower GPP and SiBCASA GPP. The results are aggregated in Fig. 22.

SiF retrieval relative anomalies are indicated by the yellow bins in Fig. 22. With the exception of January 2014, they do not show extreme values and range on average between -50% and +50%. A clearly positive signal can be observed in 2010, where all months show a positive SiF anomaly.

The SiBCASA GPP (red) shows even smaller relative anomalies ranging from -25% to +50%. Only 2012 shows relatively high anomalies in the first half of the year. This can be expected as model outputs are generally low in random fluctuations.

The green bins representing anomalies in the GPP derivation of the flux tower CO_2 measurements in Fig. 22 are highly variable. Anomalies of over 100% can be seen in the early months of 2011 through 2015. This indicates that GPP values are more than twice as high as the average value for that month.
Observation shows that these extreme values occur mostly in the winter and early spring (January to April). In this time of the year, GPP is very low in deciduous vegetation enlarging the anomaly profile.
For an overview of each particular month, including error analysis, the results have been summarized in Appendix B (32). For the individual months, error bars indicated the spread of daily values within that particular month. Note the different scaling on the y-axis. Error bars and anomalies are generally higher in the winter. This is to be expected as noise becomes more dominant in times of low vegetation activity.

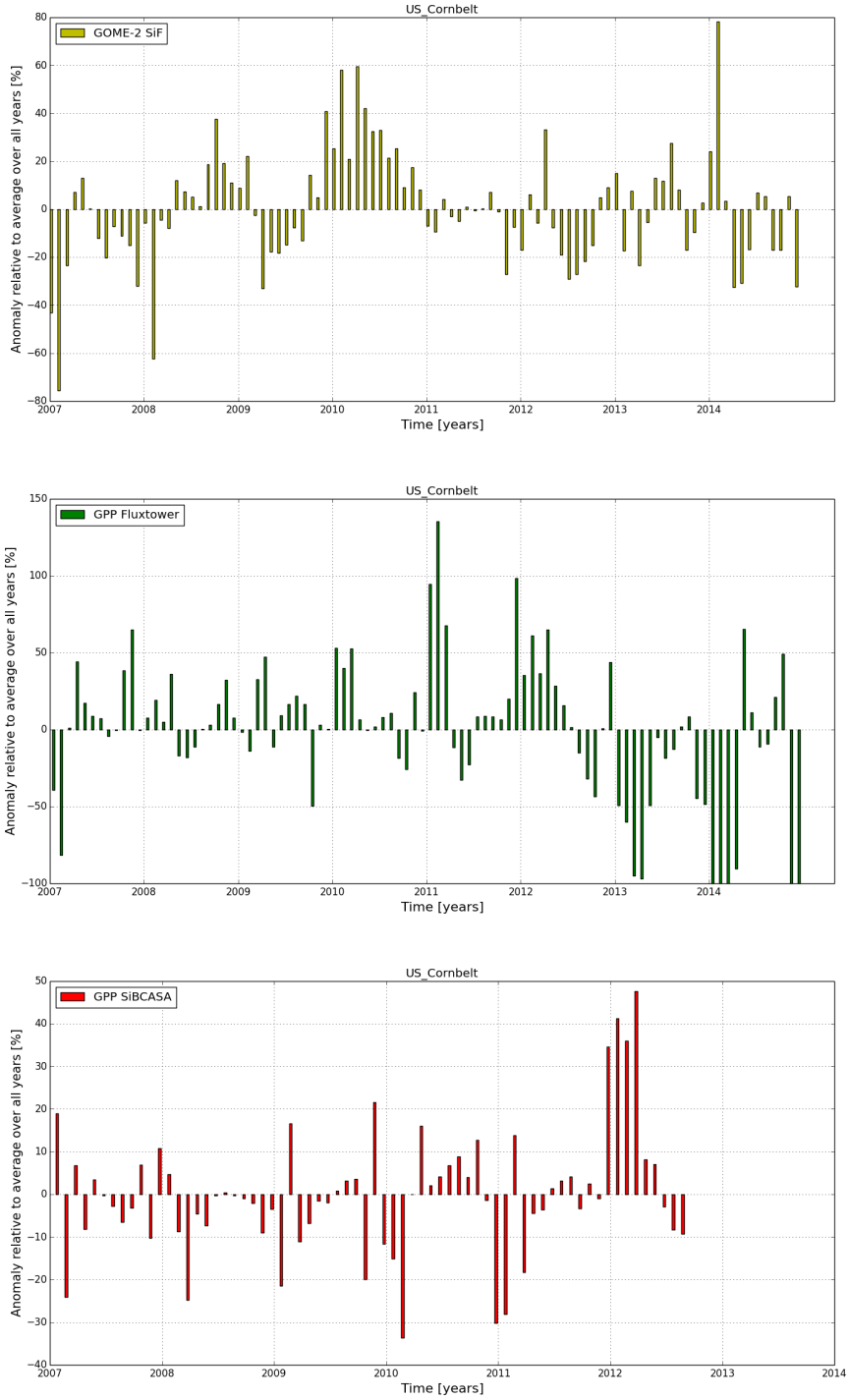

Figure 22: Anomalies relative to overall mean shown for every month during the GOME-2A mission for the SiF retrieval (yellow, top), flux tower GPP (green, middle) and SiBCASA GPP (red, bottom).

5 Discussion & Conclusion

First objective: Apply and improve the retrieval method for SiF developed by Joiner et al. [2013] and Guanter et al. [2014].

Using a new reference set selection method, this study successfully produced a global SiF retrieval. The retrieval encompasses the entirety of the GOME-2A mission from January 2007 through June 2015. With the new implementation of this algorithm it is possible to perform near real-time retrieval from the GOME-2A data on a monthly basis. Furthermore, this research lays another foundation for investigating the relation between upscaling leaf-level SiF to canopy level.

5.1 Baseline Retrieval

The baseline retrieval results show strong similarities with the BMNG-NVDI vegetation map. Spatial patterns show remarkable resemblance and also the seasonal vegetation cycle is clearly represented. However, in ocean, ice and other regions with little to no vegetation, the retrieval shows significant noise. It seems that the standard deviation on this background noise is at least 10% of the maximum retrieval signal. This causes an unreliable retrieval in regions with little vegetation. SiF values in areas such as the Sahara, the deserts in central Asia, Australia and China should therefore be treated with this in mind.
On top of this noise, the SSA shows a large area for which the retrieval is extremely noisy in the South American region. A large part of the Amazon Rain Forest is obstructed by this anomaly and it hampers time series analysis of for example deforestation of the Amazon Forest.

The negative SiF in areas such as the Aral Sea, river mouths and other areas is probably an artifact produced when a retrieval is attempted over shallow water. The results show that the negative values disappear when the Aral Sea is dried up. Moreover, sediment rich areas such as river mouths (e.g. Amazon, Ganges) and estuaria also show negative SiF retrieval. These results suggest that this effect is caused by the reflectance properties of suspended sediment, the bottom or other reflectance effects of shallow water, but further investigation is needed to determine the exact cause of these negative SiF values.

A strong seasonal cycle of SiF can be seen around the tropics. This is rather unexpected, as variations in the SZA are very limited in that area ($\pm 23.5°$) causing a less distinct growing season compared to higher latitudes. The presence of the Inter Tropical Convergence Zone (ITCZ) in this region shows a strong seasonal cycle and brings with it a large water vapor column which influences the SiF retrieval. This could be the cause of this relatively strong SiF seasonality in the tropics. Moreover, the ITCZ is a great source of water needed for plant growth which also influences the SiF seasonality.
The Savanna region, especially in Africa, shows a strong seasonal cycle with SiF following the overhead sun. This suggests that the grassy vegetation exhibits strong SiF and is therefore the cause of the seasonal variation visible in the tropics and subtropics.

5.2 Aerosols

Although we neglect any atmospheric scattering effects in the forward model (Equation 9), aerosol absorption does play a small role in the absorption and reflectance of light on its way from the canopy to the top of the atmosphere. Aerosol absorption is much larger [*Eck et al.*, 1999] than the molecular absorption (section 2.1). This means that our assumption on negligible atmospheric scattering does not hold for areas with high aerosol content. This would thus yield an incorrect SiF retrieval. We hypothesize that aerosol absorption is partly described by the PCs which determine the atmospheric optical thickness. This means that when the reference set contains pixels with aerosols, retrieval pixels with high aerosol content could still yield a valid SiF retrieval. Hence, for now, the possibly invalid assumption of negligible atmospheric scattering due to aerosol influences is no matter of concern. However, we recommend keeping this in mind when attempting SiF retrieval over an area with significantly high aerosol content compared to the reference area.

5.3 *Joiner et al.* [2013] v26 comparison

The comparison with the Jv26 data set shows an overall higher absolute SiF magnitude in this study. Moreover, these differences in SiF can be seen particularly in areas with high surface altitude. The retrieval by Jv26 shows significantly lower values in these regions. We hypothesize that this might be due to the different spectral window used for the retrieval since Jv26 only uses the continuum, and does not include the O_2-A and H_2O absorption bands as used in this retrieval. The atmospheric transmittance is significantly affected when attempting a retrieval at higher altitude as the atmospheric column of oxygen and water vapor is much smaller over high altitude terrain, simply because the atmospheric column is shorter and lower temperatures at high altitude allow for less atmospheric water vapor content. This results in a smaller absorption by O_2-A and H_2O which influences the PCs directly. Looking back at Equation 9, this results in a $\tau^{\uparrow\downarrow}$ which is possibly not well described for these regions. We hypothesize that this in turn leads to an overestimation of the fluorescence term F_S.

The new implementation of the reference spectra selection provides a different bases for the SiF retrieval. We can see a set of differences in the two retrievals as the correlation between them is of the order of R=0.75. A small positive bias is found in the non-vegetated areas where no SiF is expected, whereas in Jv26, this bias is slightly negative.

The time series in Fig. 18 gives more insight in these differences. The two satellite retrieval methods shows a high correlation with each other (R=0.95). We compare this result with a study done by *Köhler et al.* [2015]. This study shows that the correlation between different SiF retrievals (Joiner v25 and a retrieval using a different spectrometer (SCIAMACHY[7])) is high for the entire globe. Correlations between the SiF retrieval in this study and Jv26 are of the same order of magnitude as in *Köhler et al.* [2015], however, only derived for U.S. Cornbelt retrievals. This suggests that, although a different area is used, our SiF retrieval is comparable with the other state of the art fluorescence retrievals of *Joiner et al.* [2013] and *Guanter et al.* [2014].

The time series of the retrieval in this study also clearly shows a positive bias in the SiF retrieval during winter months. A possible way to correct for this bias would be to average over all a large number of non-vegetated pixels (e.g. the ocean) and simply subtract this value from all retrieved SiF values. Before this can be done however, we would first need to investigate whether the bias over the ocean is also representative for land pixels.

[7]SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY.

5.4 Flux towers and SiBCASA

Second objective: Evaluate the accuracy of the SiF-retrieval by comparing the satellite SiF data with ground based GPP observations derived from flux towers and model GPP output.

In the U.S. Cornbelt, variability in SiF is compared with variability in GPP measured by flux towers and GPP calculated by the SiBCASA model. Results show that the onset of SiF are approximately a month earlier than the flux tower GPP onset. Also the offset is on average a month later.
Comparing the results with the SiBCASA model data, we see an onset season which coincides with the SiF retrieval, but this onset also comes a month before the flux tower GPP onset. The offset season for SiBCASA is even earlier than the flux tower GPP offset, resulting in a difference of 1 to 2 months between SiBCASA GPP and SiF in the offset season.

A variability of about 20% is observed between the maximum SiF values of different years. This same variance is visible in the flux tower GPP data. However, high SiF years do not always coincide with high GPP years. In general, SiF and GPP show a strong correlation with the flux tower data (daily: $R \sim 0.80$, monthly: $R \sim 0.88$) and an even stronger correlation with the SiBCASA model GPP (daily: $R \sim 0.84$, monthly: $R \sim 0.93$) over the U.S. Cornbelt region. Although a part of this correlation is due to the mutual seasonal variability, this result strongly indicates a link between SiF and GPP.

The onset and offset seasons are nearly identical in the two satellite retrievals (this study and Jv26). However, the SiF and GPP onset and offset do not coincide that well. This effect is also found in *Joiner et al.* [2014] which focuses on the relation between satellite SiF retrieval and the in situ derived GPP measurements in the U.S. Cornbelt. In that study, especially the onset of SiF is earlier than the GPP measurements.

We hypothesize that the differences in the onset and offset between the SiF retrievals and the GPP is caused by the upscaling from the flux tower footprint to the GOME-2A footprint. Flux tower derived GPP can show a distinct onset and offset season for a single type of vegetation (e.g. corn) whereas the GOME-2A pixel will always include multiple types of vegetation. For example: In the case that corn starts growing relatively late compared to other vegetation, the GOME-2A SiF retrieval will automatically mismatch with the flux tower onset. One way to test this hypothesis would be to look at flux towers in fields of very specific vegetation. If for example a certain type of vegetation would be known to have a late onset and a late offset, we would expect to see this in the time series analysis comparison. In that case, the flux tower GPP would show an offset season later than the SiF offset season.

5.5 Relative monthly anomalies

Third objective: Investigate temporal SiF anomalies in the U.S. Cornbelt and compare them with anomalies in flux tower derived GPP measurements and modeled GPP

When working with anomalies, one must keep in mind that calculating anomalies involves operations with small numbers, especially in the winter when SiF and GPP is close to zero. This results in an amplification of the relative anomaly value meaning that anomalies for dormant vegetation months should be handled with care.

The results show that anomalies in the SiF retrieval show a positive SiF signal throughout 2010 which does not compare to any of the anomalies in flux tower or model GPP. We hypothesize that this is caused by the de-icing of the GOME-2A spectrometer in September 2009. After this procedure, a higher number of photons are hitting the spectrometer causing an apparent increase in TOA radiance. However, further research would be necessary to pinpoint the exact cause of this anomalous feature, since we are still convinced that radiometric degradation of the TOA reflectance spectrum is accounted for within the principle component parameterization of the atmospheric optical thickness.
From Fig. 22 we cannot conclude that temporal SiF anomalies in the U.S. Cornbelt show the same structure as the flux tower derived GPP measurements and the modeled GPP from SiB-CASA. A comparison with anomalies calculated with the Jv26 data set could however lead to new insights into for example the positive anomalies seen in the SiF retrieval of 2010.

5.6 Future Research

GOME-2A on board of the MetOp-A satellite is no longer in its original operational mode. It has been replaced by a similar spectrometer on board of MetOp-B which is a more modern satellite in a similar orbit. The retrieval algorithm can be easily modified to also accept GOME-2B data from MetOp-B and hence further the retrieval time line.
Another big opportunity for this SiF retrieval is usage of the new TROPOMI Sentinel 5 precursor. This instrument operates on similar wavelengths and with a much higher spectral and spatial resolution.

The link between SiF and GPP could be further investigated using the results of this research. One way of doing so would be implementing a leaf level to canopy radiative upscaling model in the retrieval of vegetated pixels. Research is on this subject [*Rossini et al.*, 2015] looks promising.

When analyzing the retrieval differences with Jv26, one might consider running the retrieval with a smaller spectral window. This way, variation in the spectral reflectance structure is mitigated by removing contributions from the O_2-A and H_2O absorption bands. Retrieving SiF in this way might increase accuracy at higher surface altitudes. One might also choose to test the choice of reference area selection. Including reference areas which have a high surface altitude (e.g. snowy mountain areas, high altitude lakes or ice caps and glaciers) could possibly help the retrieval of high altitude fluorescence pixels.

The implementation of the reference set selection in this study is necessary for near real-time retrieval. However, results show that changes in GOME-2A calibration or sudden spectral degradation (e.g. de-icing) can cause a problem in the SiF retrieval. An event as such causes a small discontinuity in the retrieval. Further investigation of these effects might be necessary before making this SiF retrieval operational.

6 References

Buschmann, C., Variability and application of the chlorophyll fluorescence emission ratio red/far-red of leaves, *Photosynthesis Research*, *92*, 261–271, 2007.

Callies, J., E. Corpaccioli, M. Eisinger, A. Hahne, and A. Lefebvre, GOME-2 - MetOp's Second-Generation Sensor for Operational Ozone Monitoring, *Bulletin*, ESA, 2000.

Casadio, S., and O. Arino, Monitoring the South Atlantic Anomaly using ATSR instrument series, *Advances in Space Research*, *48*, 2011.

Ciais, P., et al., *Carbon and Other Biochemical Cycles. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univsersity Press, Cambridge, UK and New York, NY, USA, 2013.

Dikty, S., and A. Richter, GOME-2 on MetOp-A Support for Analysis of GOME-2 In-Orbit Degradation and Impacts on Level 2 Data Products., *Final report*, University of Bremen, Institute of Remote Sensing, 2011.

Eck, T., B. Holben, J. Reid, O. Dubovik, A. Smirnov, N. o'Neill, I. Slutsker, and S. Kinne, Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, *Journal of Geophysical Research*, *104-D24*, 31,333 – 31,349, 1999.

Esbensen, K., D. Guyot, F. Westad, and L. Houmøller, *Multivariate Data Analysis - in practice: An introduction to multivariate data analysis and experimental design*, 5 ed., CAMO ASA, Oslo, 2001.

Field, C., M. Behrenfeld, J. Randerson, and P. Falkowski, Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components, *Science*, *281*, 237–240, 1998.

Guanter, L., et al., Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence, *Proceedings of the Natural Academy of Sciences*, *111-14*, 1327–1333, 2014.

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

Joiner, J., et al., 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, 2014.

Koelemeijer, R., P. Stammes, J. Hovenier, and J. Haan, A fast method for retrieval of cloud parameters using oxygen A band measurements from the Global Ozone Monitoring Experiment, *Journal of Geophysical Research*, *106*, 3475–3490, 2001.

Koelemeijer, R., J. Haan, and P. Stammes, A database of spectral surface reflectivity in the range 335-772 nm from 5.5 years of GOME observations, *Journal of Geophysical Research*, *108-D2*, 2003.

Köhler, P., L. Guanter, and J. Joiner, A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data, *Atmospheric Measurement Techniques*, *8*, 2589–2608, doi:10.5194/amt-8-2589-2015, 2015.

Kumar, R., M. Pal, and U. Teotia, Diurnal changes in chlorophyll fluorescence in four species of bamboo, *Journal for Bamboo and Rattan*, *1-4*, 341–349, 2002.

Leth, T. v., Remote Sensing of Sun-Induced Chlorophyll Fluorescence, Master's thesis, Utrecht University, the Netherlands, 2014.

Micklin, P., The Aral Sea Disaster, *Annual Review of Earth and Planetary Sciences*, *35*, 47–72, 2007.

Monteith, J., Solar Radiation and Productivity in Tropical Ecosystems, *Journal of Applied Ecology*, *9-3*, 747–766, 1972.

Porcar-Castell, A., E. Tyystjärvi, J. Atherton, C. van der Tol, J. Flexas, E. Pfündel, J. Moreno, C. Frankenberg, and J. Berry, Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges, *Journal of Experimental Botany*, *65*, 4065–4095, 2014.

Rossini, M., et al., Red and far red Sun-induced chlorophyll fluorescence as a measure of plant photosynthesis, *Geophysical Research Letters*, *42*, 1632–1639, doi:10.1002/2014GL062943, 2015.

Schulze, E.-D., E. Beck, and K. Müller-Hohenstein, *Plant Ecology*, Springer, Heidelberg, Berlin, 2005.

Shlens, J., A Tutorial on Principle Component Analysis, *Tech. rep.*, Google Research, 2014.

Stöckli, R., E. Vermote, N. Saleous, R. Simmon, and D. Herring, The Blue Marble Next Generation - A true color earth dataset including seasonal dynamics from MODIS, *Tech. rep.*, NASA: Earth Observatory, 2005.

Subhash, N., and C. Mohanan, Curve-Fit Analysis of Chlorophyll Fluorescence Spectra: Application to Nutrient Stress Detection in Sunflower, *Remote Sensing of Environment*, *60*, 347–356, 1997.

Velde, I. v. d., J. Miller, K. Schaefer, K. Masarie, S. Denning, J. White, P. Tans, M. Krol, and W. Peters, Biosphere model simulations of interannual variability in terrestrial $^{13}C/^{12}C$ exchange, *Global Biogeochemical Cycles*, *27*, 1–13, 2013.

Velde, I. v. d., J. Miller, K. Schaefer, G. van der Werf, M. Krol, and W. Peters, Terrestrial cycling of $^{13}CO_2$ by photosynthesis, respiration, and biomass burning in SiBCASA, *Biogeoscience*, *11*, 2014.

Wang, P., O. Tuinder, and P. Stammes, Cloud retrieval algorithm for GOME-2: FRESCO+, *Tech. rep.*, Royal Dutch Meteorological Institute, 2010.

Zarco-Tejada, P., J. Miller, G. Mohammed, and T. Noland, Chlorophyll Fluorescence Effects on Vegetation Apparent Reflectance: I. Leaf-Level Measurements and Model Simulation, *Remote Sensing of Environment*, *74*, 582–595, 2000.

7 Acknowledgments

I would like to thank the KNMI for facilitating this research internship. Special thanks go out to my supervisors, Bram Sanders (KNMI) and Willem Verstraeten (KNMI, WUR) for always supporting my progress throughout this internship and Olaf Tuinder (KNMI) for his computational help. I would also like to thank Sander Houweling (SRON) and Thomas Röckmann (IMAU) for being my external supervisors from my masters program. Furthermore I would like to acknowledge the following AmeriFlux sites for their data record: KFS, Kon, KUT, Los, MMS, Ne1, Ne2, Ne3, ORv, PFa, UMB, UMd and WCr. (Funding, U.S. Department of Energy's Office of Science). And Prof. W. Peters at Wageningen University for providing me with the SiBCASA data.

8 Appendix

Appendix A: 2007-2015 SiF maps

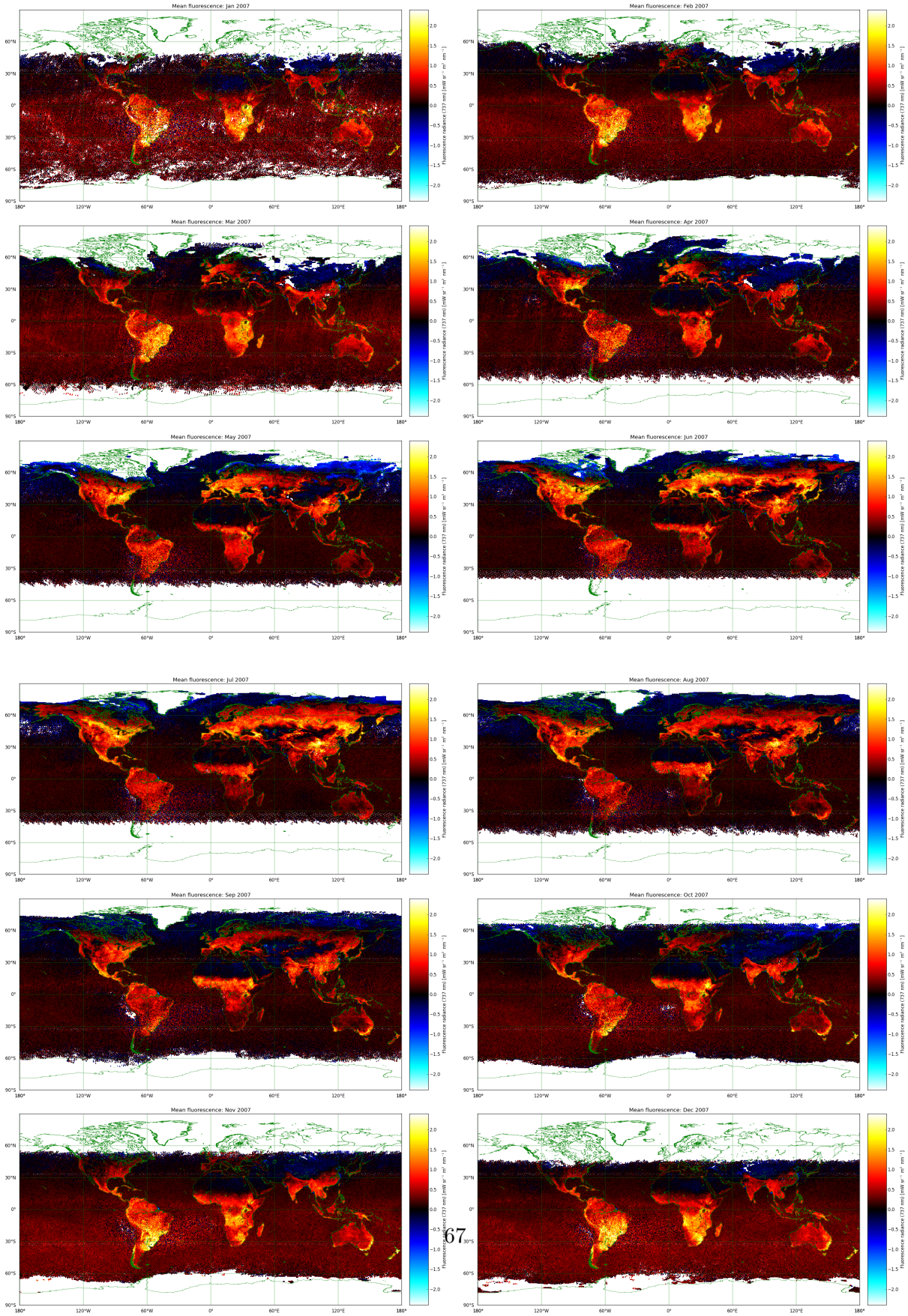

Figure 23: SiF retrieval for 2007.

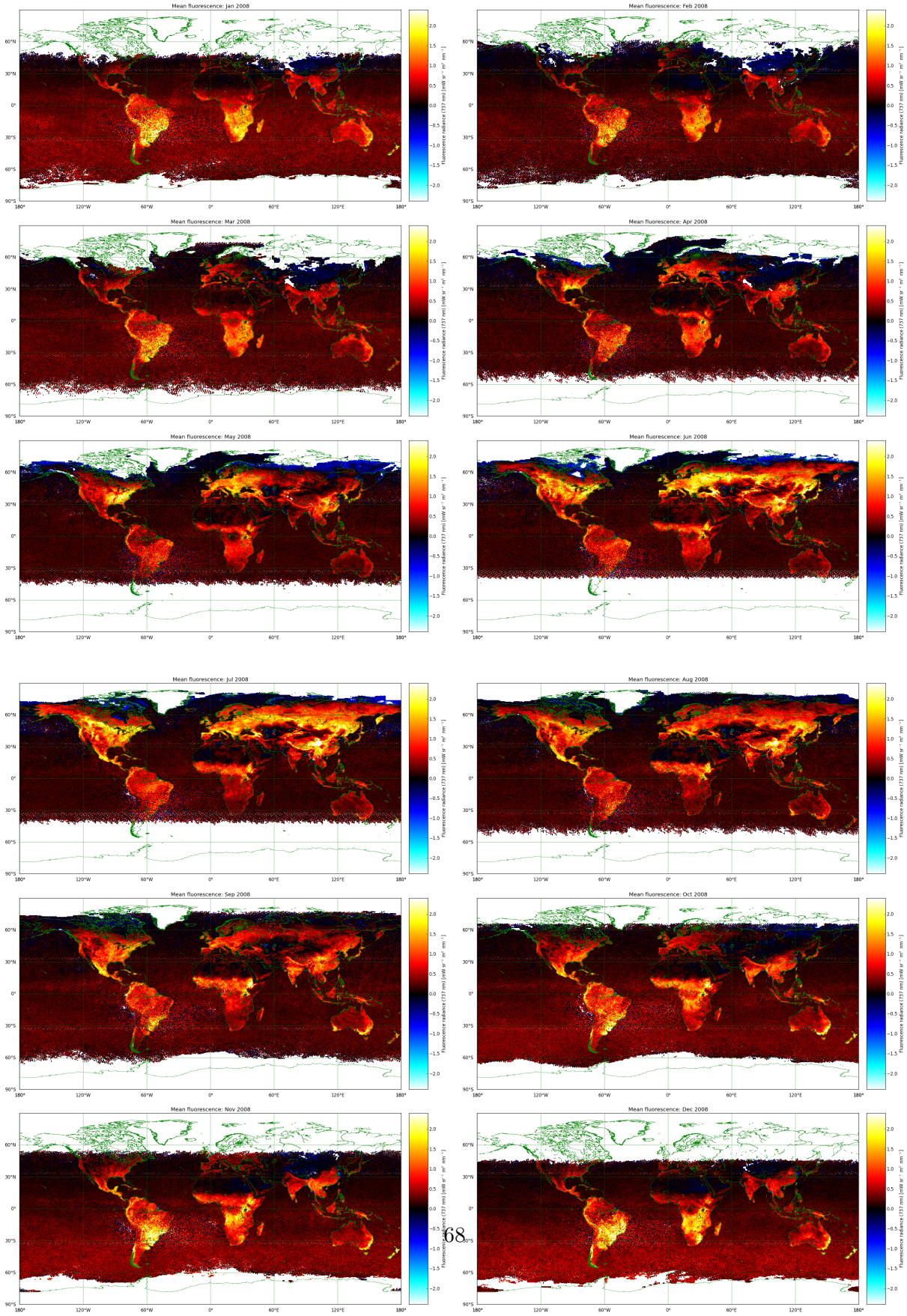

Figure 24: SiF retrieval for 2008.

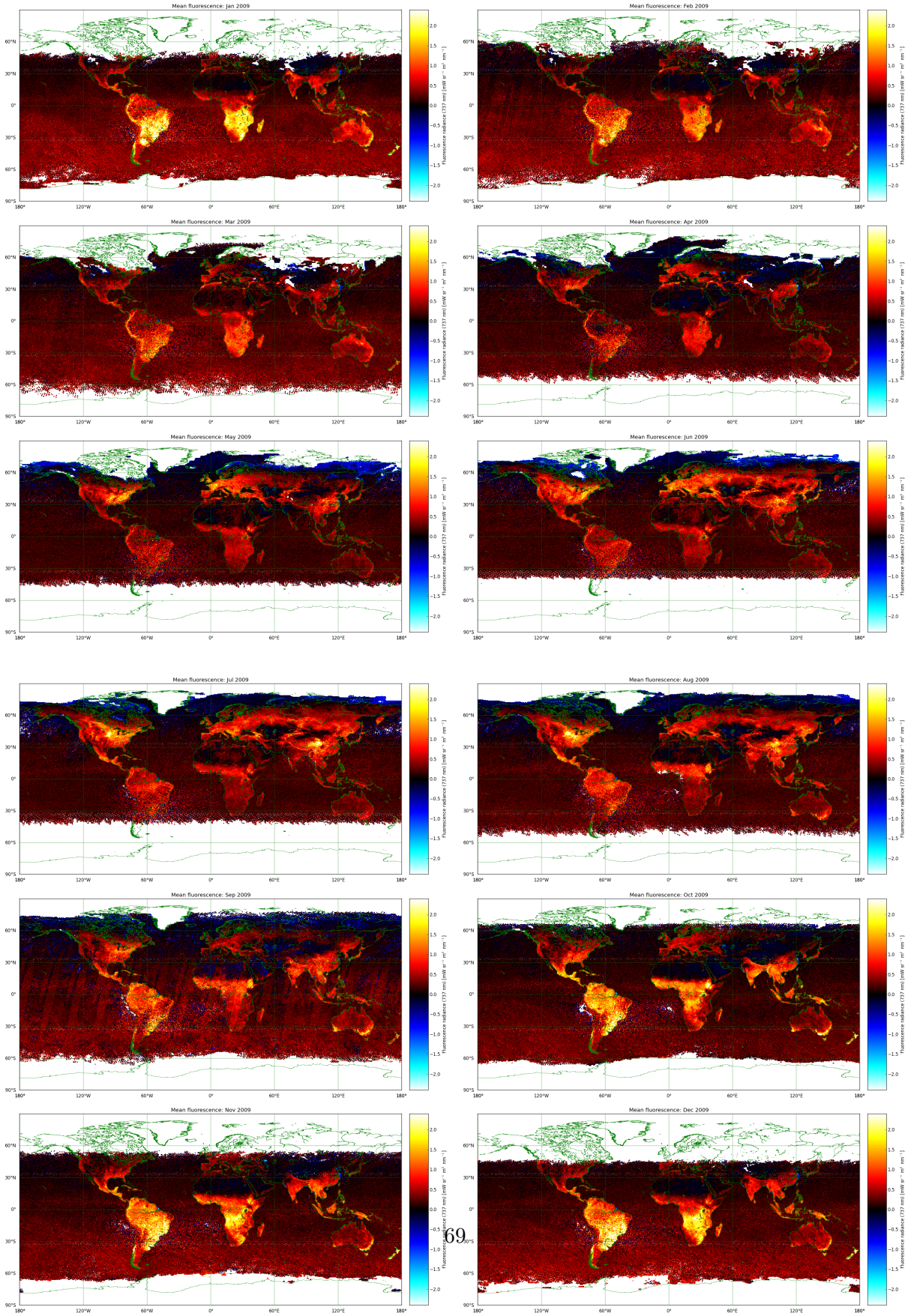

Figure 25: SiF retrieval for 2009.

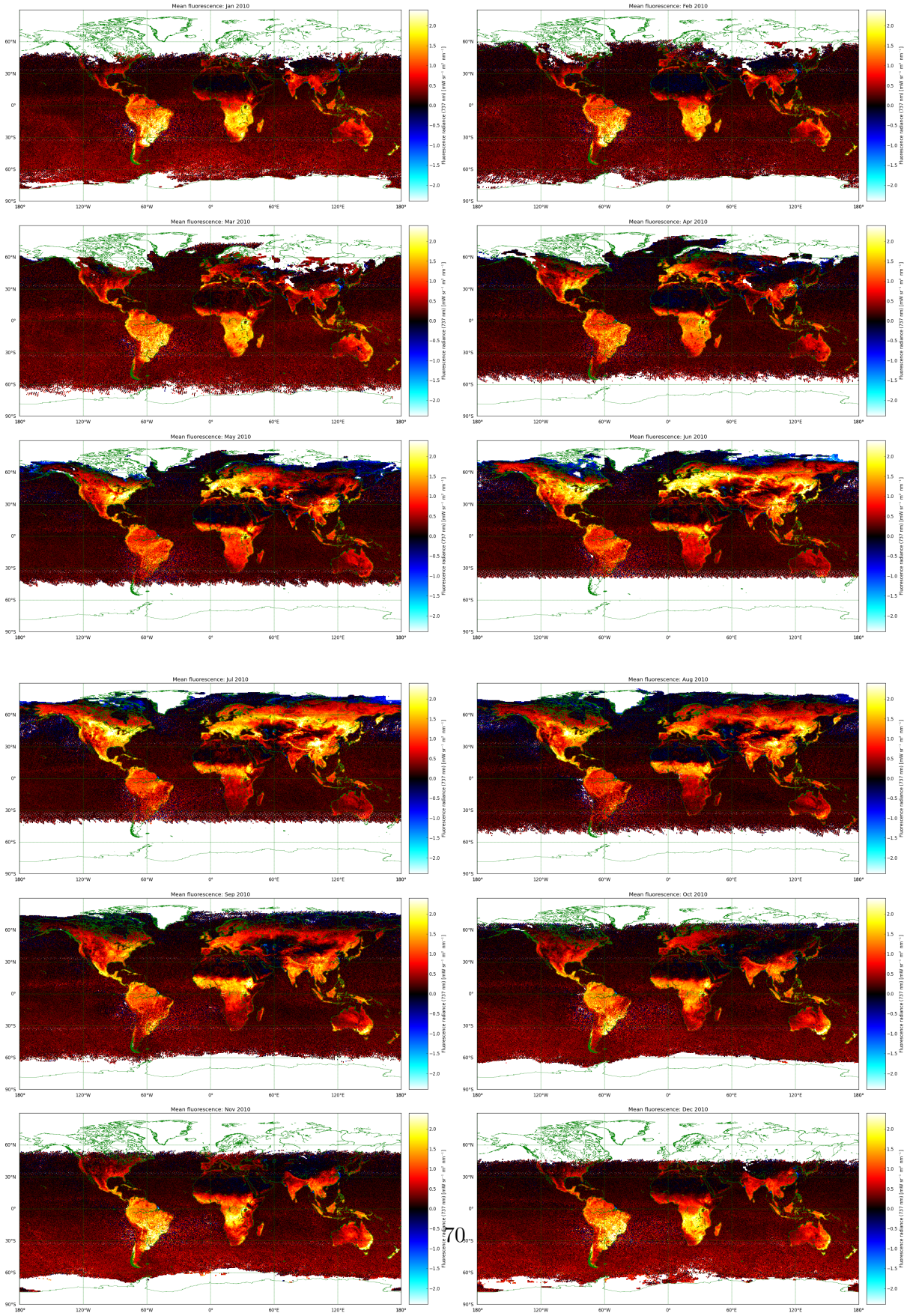

Figure 26: SiF retrieval for 2010.

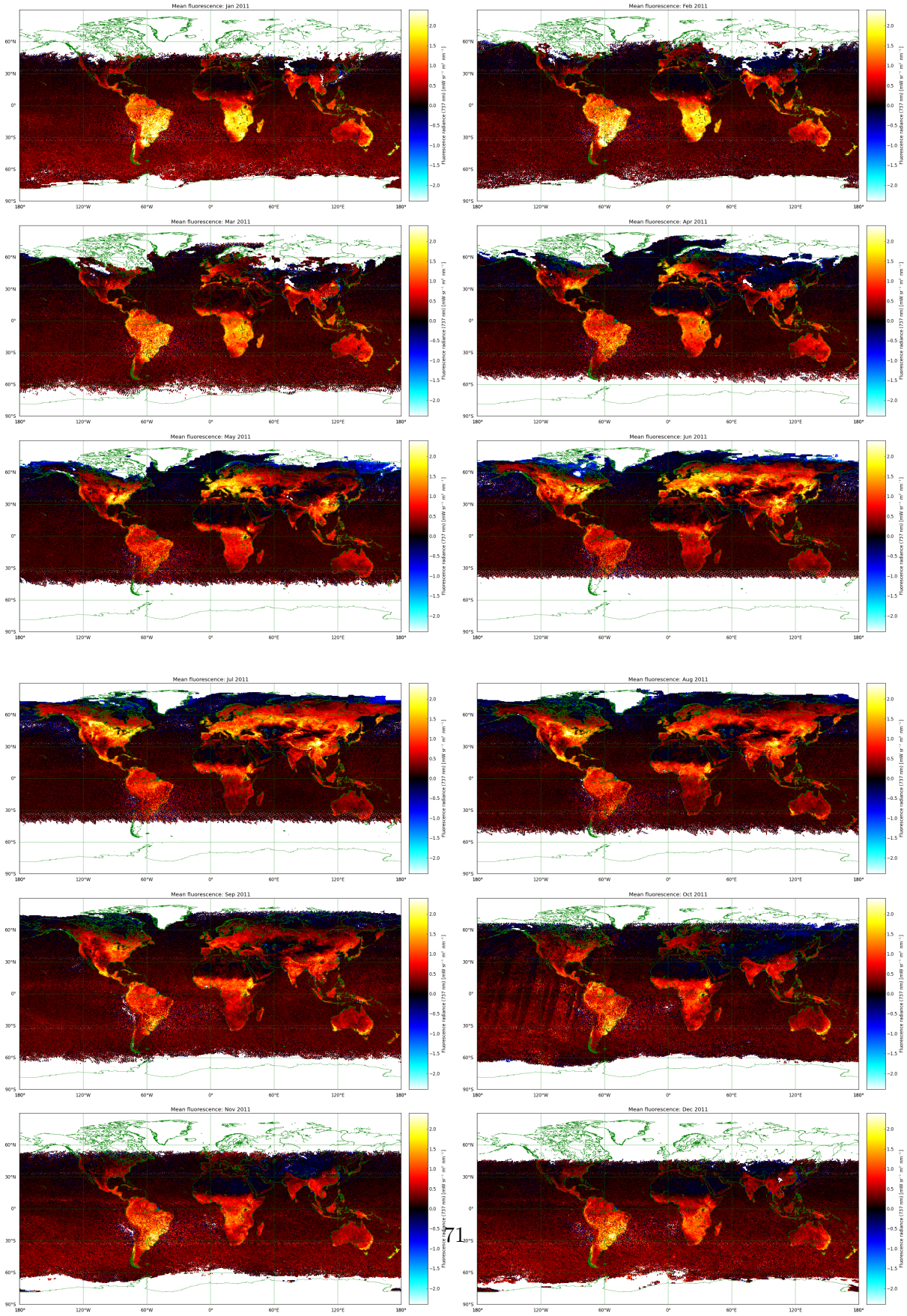

Figure 27: SiF retrieval for 2011.

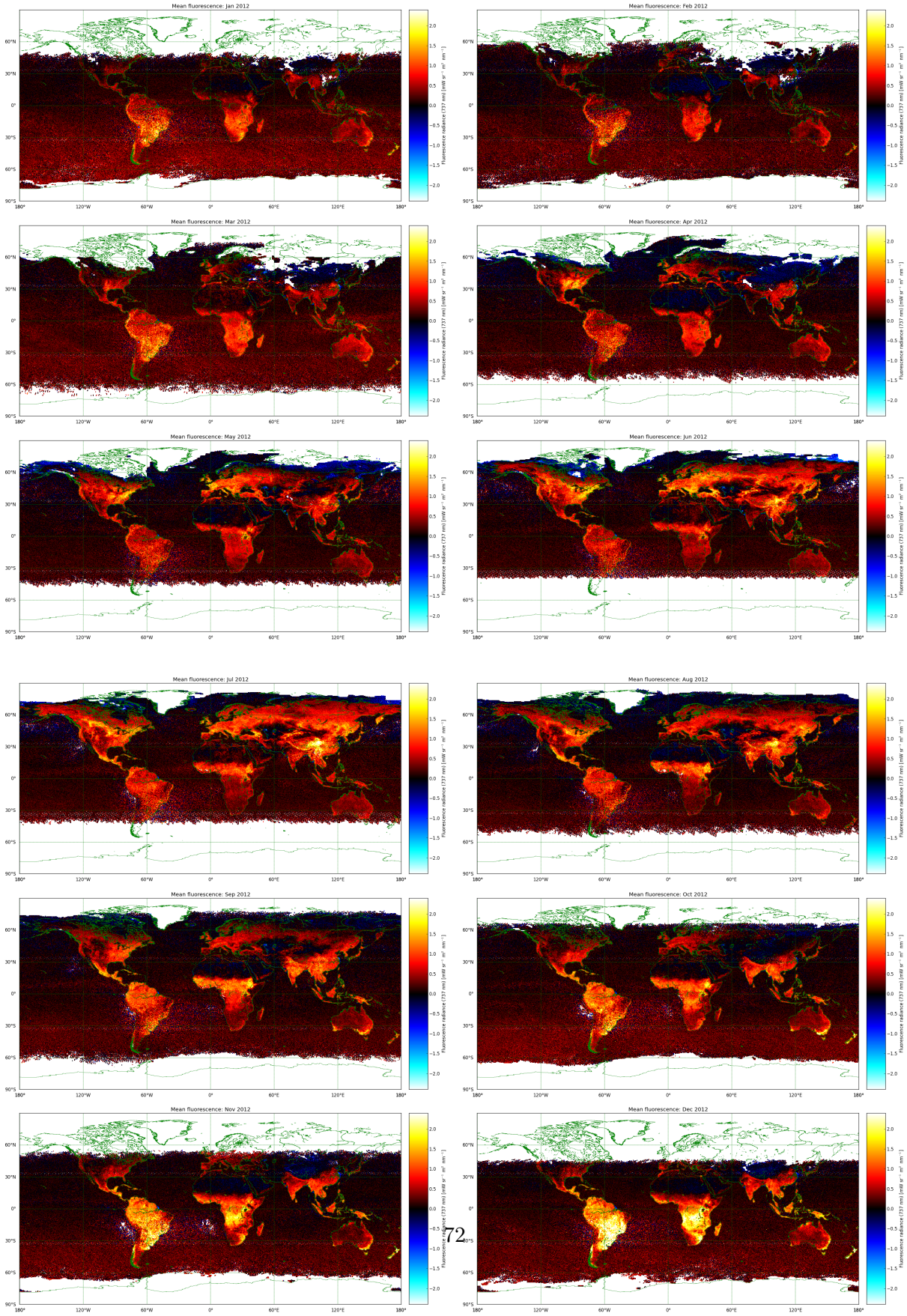

Figure 28: SiF retrieval for 2012.

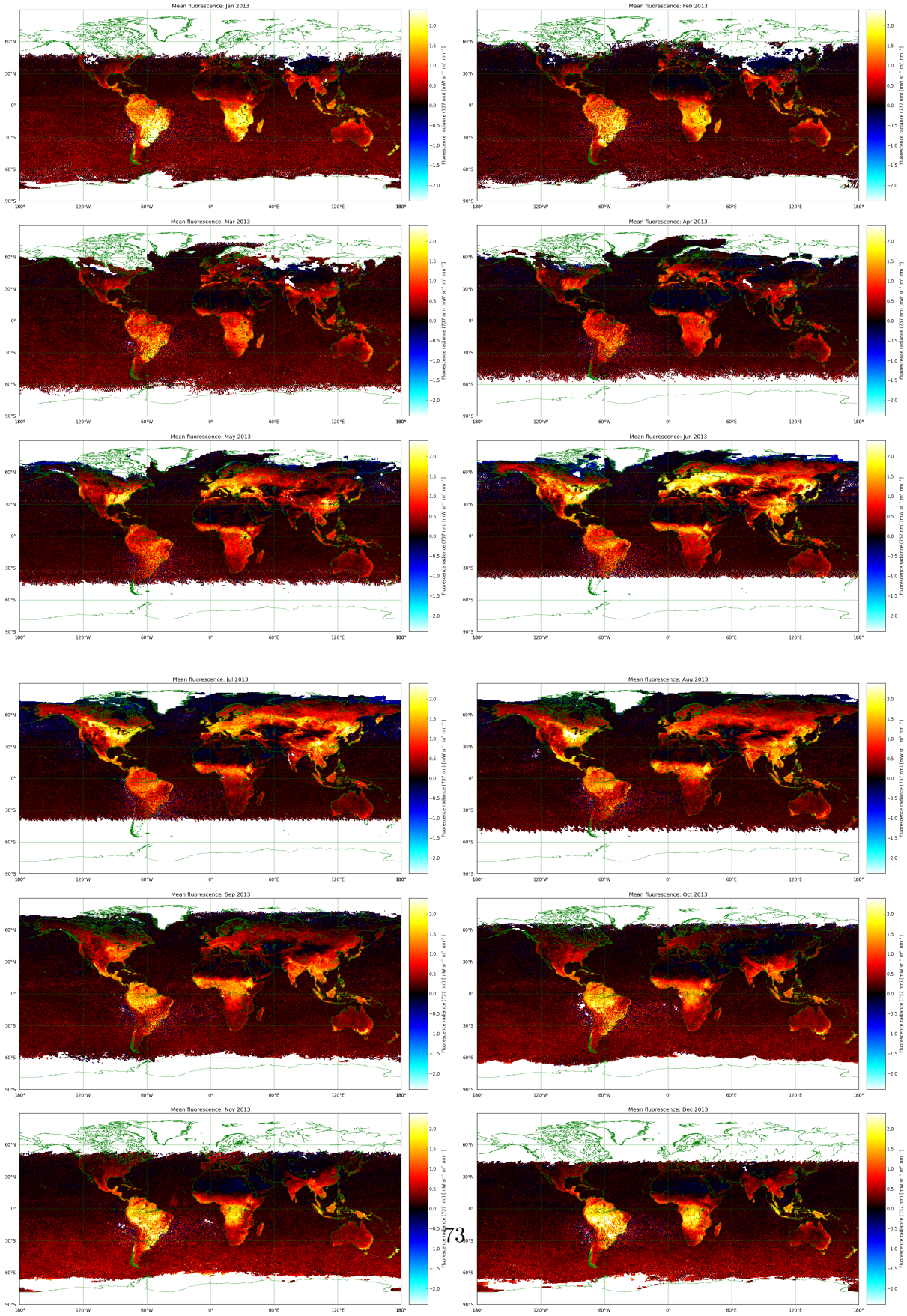

Figure 29: SiF retrieval for 2013.

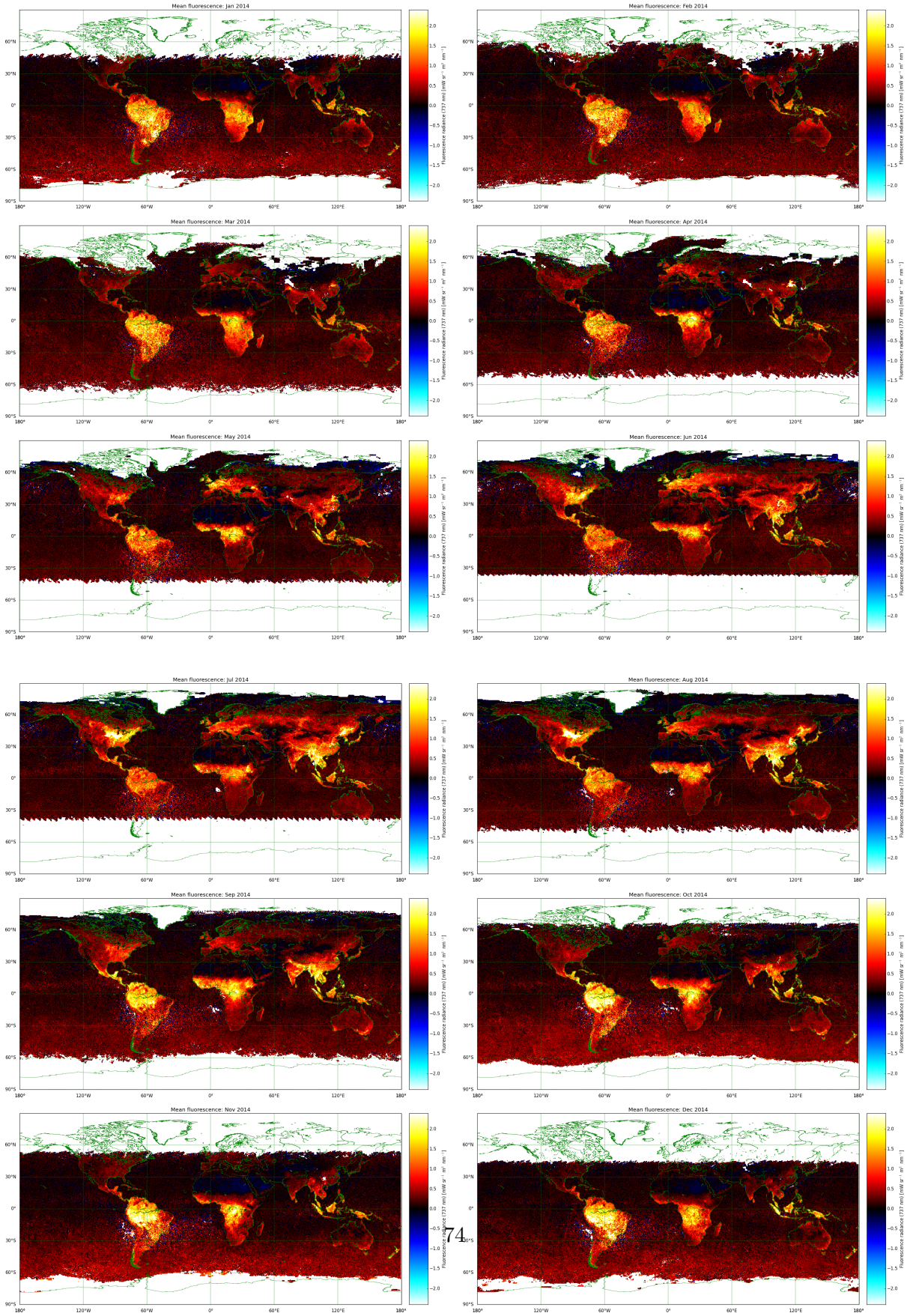

Figure 30: SiF retrieval for 2014.

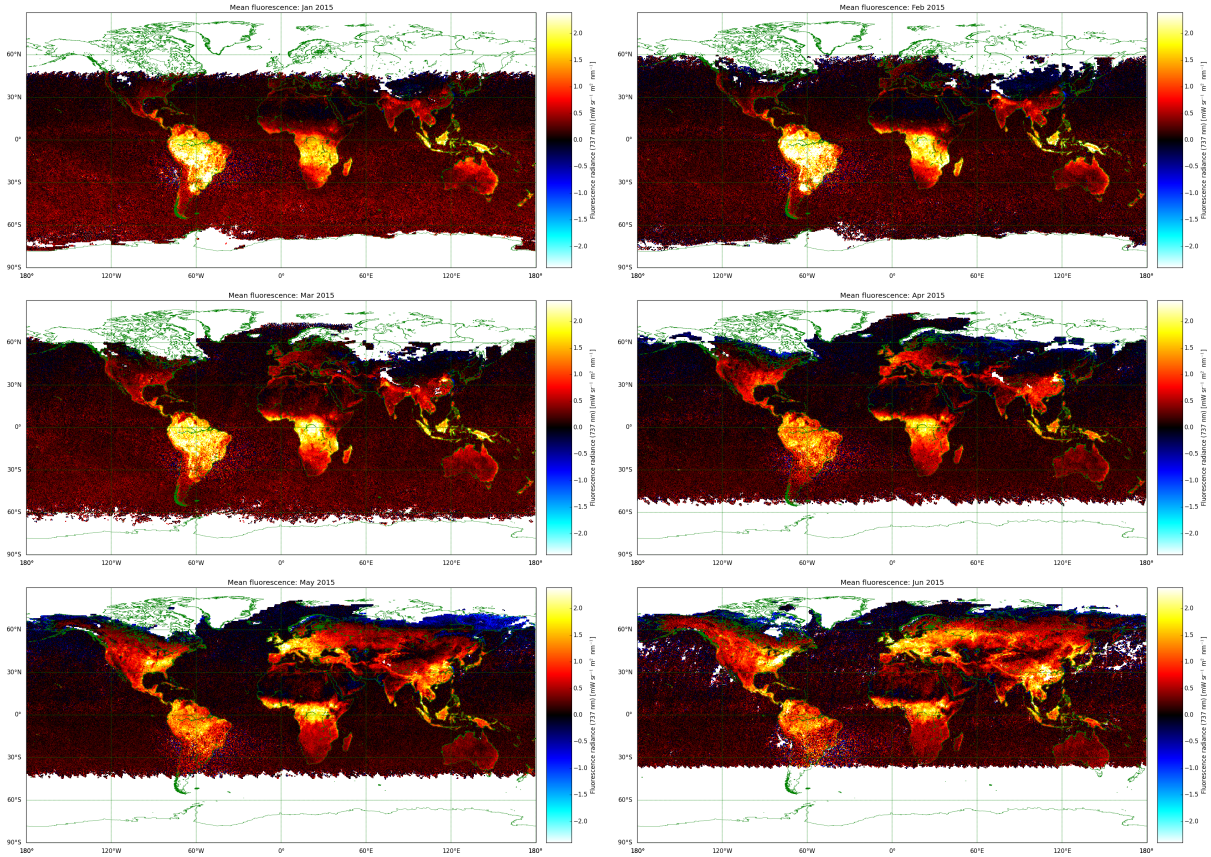

Figure 31: SiF retrieval for 2015.

Appendix B: Monthly anomalies

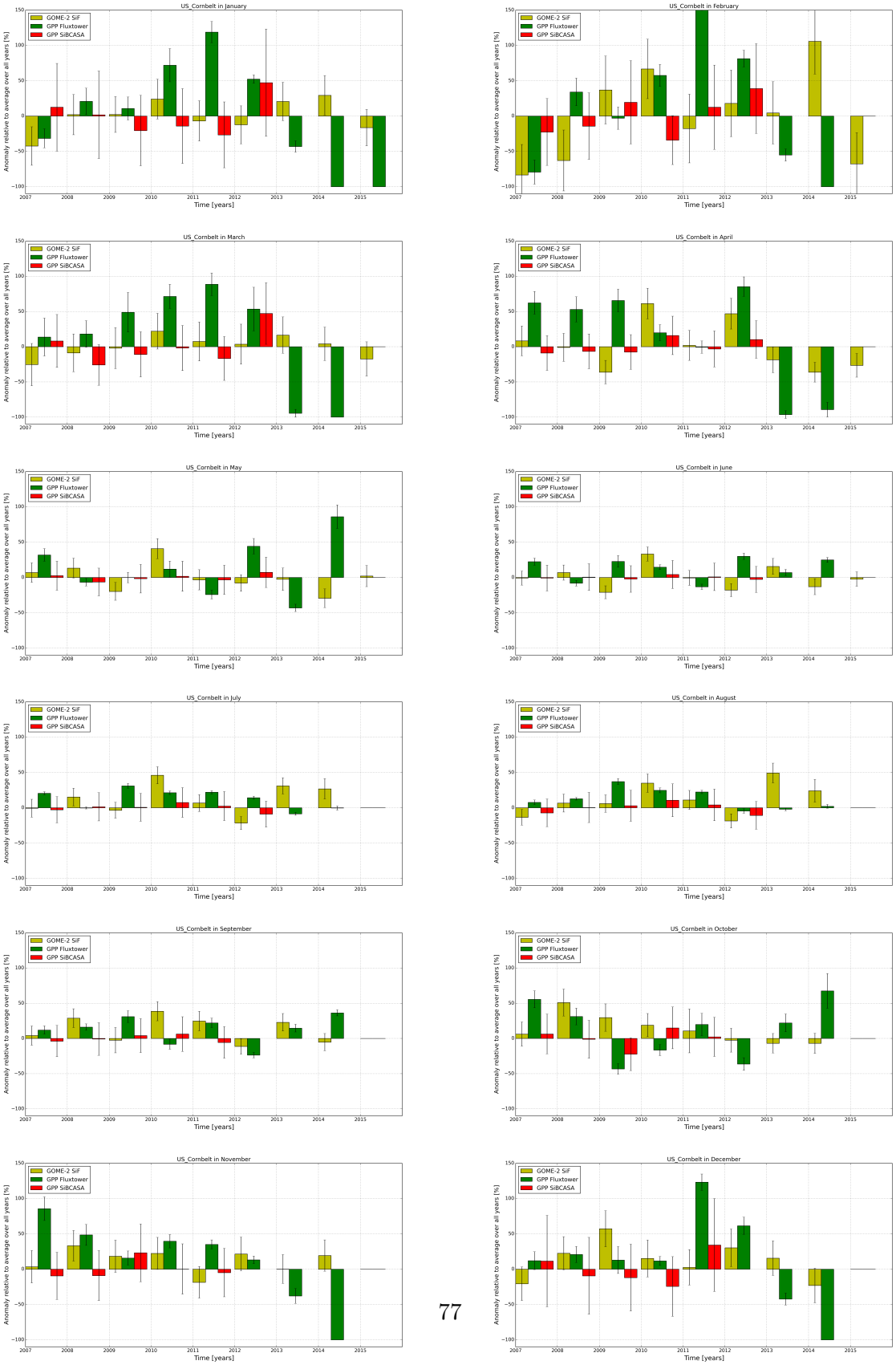

Figure 32: Anomaly with respect to the mission mean for each individual month.