# NEW SOLAR-INDUCED CHLOROPHYLL FLUORESCENCE RETRIEVAL ALGORITHM BASED ON TANSAT SATELLITE DATA

Yushi Zhou[1,2], Xiaoping Lu[1,*], Yongbo Huang[1,2], Zhong Gao[1,2], Yaqing Zheng[1,2]

[1] Key Laboratory of Spatio-temporal Information and Ecological Restoration of Mines of Natural Resources of the People's Republic of China, Henan Polytechnic University, Jiaozuo 454003, China – lxp@hpu.edu.cn
[2] School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China - (hpuzys, 211804020034, 111904010005, 211704020040) @home.hpu.edu.cn

**Commission III, WG III/4**

**KEY WORDS:** Chlorophyll Fluorescence, Random Sample Consensus, TanSat, KI Fraunhofer Line, Algorithm, Retrieval

**ABSTRACT:**

Solar-induced chlorophyll fluorescence (SIF) is an indicator of plant photosynthesis which could be detected by satellite. However, some existing algorithms are easily affected by the inaccuracy of satellite data which will causing deviation in the retrieval of SIF. To avoid "outliers" with inaccuracy affecting the retrieval results, a random sample consensus algorithm (RANSAC) was introduced to retrieve SIF in this paper. The results show that the chlorophyll fluorescence value obtained by this method is consistent with the OCO-2 SIF product ($R^2 = 0.81$), and also consistent with the MODIS vegetation index ($R^2 = 0.87$ with NDVI, $R^2 = 0.85$ with EVI). Compared with the existing SIF products (OCO-2 SIF), the SIF retrieved in this paper was better in spatial details and outlier distribution.

## 1. INTRODUCTION

Sun-induced chlorophyll fluorescence (SIF) is a type of light released by vegetation during photosynthesis and its wavelength is within 650-800 nm (Porcar-Castell et al., 2014). Studies have shown that SIF is closely related to Gross Primary Productivity (GPP) (Damm et al., 2010; Frankenberg et al., 2011) and the state of vegetation stress (Liu, Cheng, 2010; Liu et al., 2013). By detecting SIF, the light energy utilization rate of vegetation can also be perceived (Liu, Cheng, 2010). Therefore, sensing and detecting SIF are extremely significant for monitoring the growth of vegetation and investigating ecological conditions.

SIF is a kind of weak canopy emission information, and its energy value accounts for about 5% of the total reflected energy (Van der Tol et al., 2014; Zhang et al., 2018). Therefore, the main problem solved by satellite retrieval SIF is to extract SIF from the detected radiance. Many SIF retrieval algorithms have been proposed, such as FLD (Plascyk, Gabriel, 1975), 3FLD (Maier et al., 2003), iFLD (Alonso et al., 2008) and SFM (Mazzoni et al., 2008). However, the current SIF retrieval technology based on remote sensing products is not mature enough yet.

In 2013, Xinjie Liu and Liangyun Liu (Liu X, Liu L, 2013) used the weighted least squares method to search GOSAT satellites and obtained SIF products in China. Their paper provides a new idea for the SIF retrieval method based on the physical model. But the SIF retrieval method based on the weighted least squares method is still affected by outliers containing random inaccuracy.

Random sample consensus (RANSAC) was proposed by Fischler and Bolles in 1981 to eliminate the effects of "outliers" that are unrelated to real data (Fischler, Bolles, 1981). Therefore, in the process of fitting linear equations, RANSAC can effectively exclude "outliers", thereby effectively avoiding the effects of "outliers".

Based on the above theory, this paper uses the RANSAC algorithm to retrieve the global SIF values in March and August 2018 with the help of Chinese TanSat satellite ACGS (Atmospheric Carbon dioxide Grating Spectroradiometer) sensor data. Then, compared with other satellite products in terms of spatial distribution and quantitative analysis of data to verify the reliability of the algorithm.

## 2. MATERIALS AND PREPROCESSING

### 2.1 The TanSat Satellite Data

Launched in December 2016, the TanSat satellite is China's first satellite to monitor and detect carbon dioxide ($CO_2$). TanSat flies in a sun-synchronous, 700 km altitude orbit. TanSat has a spatial resolution of 2 km×2 km and a revisit period of 16 days (Yang et al., 2018).

The two main instruments onboard the TanSat satellite are the Atmospheric Carbon dioxide Grating Spectroradiometer (ACGS) and the Cloud and Aerosol Polarimetry Imager (CAPI). CAPI measurements are mainly used to provide cloud and aerosol optical properties, which can determine whether the pixel is affected by the atmosphere. ACGS has three sets of grating spectrometers with similar structures, which can realize hyperspectral detection of radiation information in three bands of 0.76 μm (O2-A band, 758–778 nm), 1.61 μm (Weak CO2 band, 1,594–1,624 nm) and 2.06 μm (Strong CO2 band, 2,042–2,082 nm). The spectral resolutions of ACGS are 0.044 nm in the O2-A band, 0.12 nm in the weak CO2 band, and 0.16 nm in the strong CO2 band. In this paper, the O2-A band was used to invert the SIF value. TanSat ACGS L1b data provides radiance data from Science product and irradiance data from Calibration product which they are both radiometrically accurate and spectrally calibrated. Their wavelength extraction formula is as follows:

---

[*] Corresponding author

$$\lambda = \sum_{i=0}^{5} C_i \cdot P^i \tag{1}$$

where $P$ = the pixel number
$C_i$ = the dispersion coefficients
$\lambda$ = wavelength

In this paper, spectral offset correction was performed on the radiance and irradiance spectra with offset before retrieval.
Figure 1 shows the example of the radiance spectrum and irradiance spectrum at O2-A band, and figure 2 shows the normalized spectral curve of radiance spectrum and irradiance spectrum between 769~771 nm. Within this range, the most obvious change is the KI Fraunhofer line (770.1 nm). Since the KI Fraunhofer spectral line has the least interference with oxygen absorption，so several radiance data covering around 770.1 nm were selected for SIF retrieval in this study.
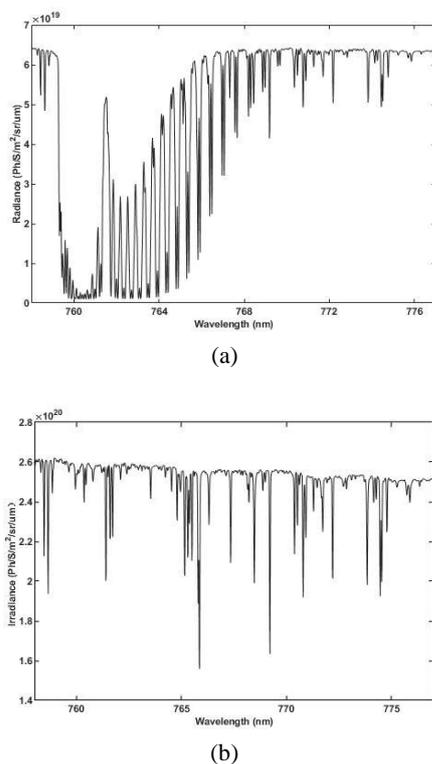


(a)



(b)

Figure 1. Example of TanSat-ACGS L1b radiance (a) and irradiance (b) spectral curve
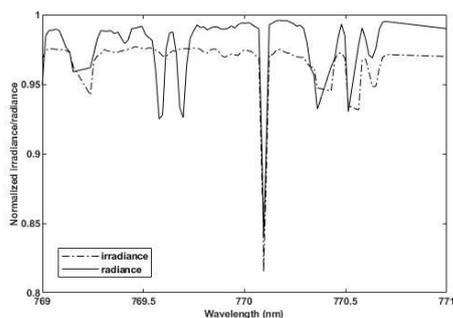


Figure 2. The normalized spectral curve of radiance spectrum and irradiance spectrum

## 2.2 Comparative Data

In order to evaluate the reliability of the results in this paper, comparisons were made by using other satellite-based datasets.

### 2.2.1 OCO-2 SIF Products

Same with TanSat, OCO-2 (Orbiting Carbon Observatory-2) is a NASA mission dedicated to studying atmospheric carbon dioxide, which has been launched in July 2014. OCO-2 can provide 757 nm and 771 nm SIF retrievals and some supporting geographic, cloud and meteorological information from September 2014 (DOI: 10.5067/AJMZO5O3TGUR). Those OCO-2 SIF products are retrieved based on the IMAP-DOAS method (Frankenberg et al., 2011, 2012 & 2014).

### 2.2.2 MODIS Vegetation Index Products

MODIS (Moderate Resolution Imaging Spectroradiometer) L3 geographic Climate Modelling Grid (CMG) product MOD13C2 contains Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) data at global scale (0.05° grid) (doi: https://doi. org//10.5067/MODIS/MOD13C2.006) (Huete et al., 2002).

All comparative data were converted to GeoTIFF format using Matlab R2019a software.

## 3. SIF RETRIEVAL ALGORITHM

### 3.1 Basic equations

Assuming that the surface reflection conforms to Lambertian-equivalent reflectivity surface model and that the chlorophyll fluorescence intensity $F$ and surface reflectance $R$ do not change with the wavelength in the studied radiance spectrum (769-771 nm), the radiance can be expressed as (Joiner et al., 2011):

$$I(\lambda) = I_0(\lambda) + \frac{R T_\uparrow(\lambda) T_\downarrow(\lambda)}{1 - RS(\lambda)} \cdot \frac{E(\lambda)\cos\theta}{\pi} + \frac{F T_\uparrow(\lambda)}{1 - RS(\lambda)} \tag{2}$$

Where $I(\lambda)$ = observed radiance
$I_0(\lambda)$ = atmospheric path radiance
$T_\uparrow(\lambda)$ = up-welling transmittances
$T_\downarrow(\lambda)$ = down-welling transmittances
$E(\lambda)$ = solar irradiance
$S(\lambda)$ = atmospheric spherical albedo
$\theta$ = solar zenith angle

In the 769 ~ 771nm radiation spectral range, the atmosphere has significant permeability, so the effects of atmospheric scattering and atmospheric absorption can be ignored. Thus, we set atmospheric path radiance and atmospheric spherical albedo value as 0, up-welling and down-welling transmittances as 1, then equation 2 could be simplified as:

$$I(\lambda) = E(\lambda) \cdot \frac{R\cos\theta}{\pi} + F \tag{3}$$

Let constant $\frac{R\cos\theta}{\pi} = k$, the equation 2 can be written as:

$$I(\lambda) = k \cdot E(\lambda) + F \tag{4}$$

We can use radiance and irradiance data provided by TanSat to retrieve the SIF.

## 3.2 Random Sample Consensus Algorithm

Random sample consensus (RANSAC) is an algorithm that can estimate mathematical model parameters from a set of observations containing outliers (Fischler, Bolles, 1981). A basic assumption is that the data consists of "inliers" and is also affected by "outliers" that contain inaccuracies. Therefore, RANSAC assumes that a process can be used to estimate model parameters that can explain or fit the data, eliminating the influence of "outliers" (Figure 3).
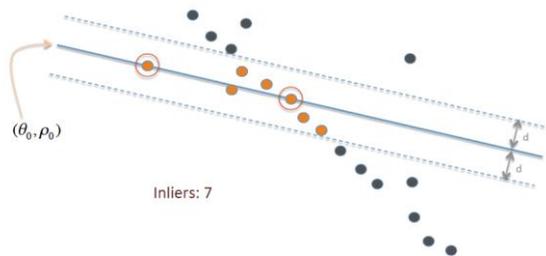


Figure 3. RANSAC: Inliers and Outliers (URL: https://upload.wikimedia.org/wikipedia/commons/b/b7/RANSAC_Inliers_and_Outliers.png)

RANSAC algorithm needs to input a set of observation data, a parametric model that can be interpreted or adapted to observation data, and some credible parameters. Then, the algorithm achieves its goal by repeatedly selecting a random subset of the data. The selected subset is assumed to be the "inliers" and verified using by the following method:

**Step 1:** There is a model that is adapted to the hypothetical "inliers", that is, all unknown parameters can be calculated from the hypothetical "inliers".
**Step 2:** Use the model obtained in step 1 to test all other data. If a point is suitable for the estimated model, it is considered to be an "inlier".
**Step 3:** If enough points are classified as "inliers", the estimated model is reasonable enough.
**Step 4:** Re-estimate the model with all hypothetical "inliers".
**Step 5:** Finally, the model is evaluated by estimating "inliers" and the model's error rate.

This process is repeated to perform a specified number of times. Each time the new model is generated, there are two situations: either it is abandoned by few "inliers", or it is selected cause it is better than the existing model.
Based on this, we designed the process of retrieval SIF via RANSAC. All calculations are performed by Matlab R2019a. The main steps of the algorithm are:

---

**Algorithm**: SIF Retrieval via RANSAC

1: **Input:** data [x = I ($\lambda$), y = E($\lambda$)], model[ $y = k \cdot x + F$ ], n = 2, k $= C_{\psi}^{2}$ , Number of radiation spectral: $\psi$ , t $= 0.6 \cdot D_{max}$ , d = 2, Maximum distance between data points: $D_{max}$

2: **Initialization** iterations = 0, best_model = null, best_consensus_set = null, best_num = 0
3: **while** (iterations < k)
    maybe_inliers = Random n points from the data set
    maybe_model = fit [maybe_inliers]

    consensus_set = maybe_inliers
4:     **For** ( $\forall$ [x, y] in the data that do not belong to [maybe_inliers])
5:     **if** ([x, y] fit [maybe_model] & [t])
6:       **do** add [x, y] to [consensus_set]
7:     **if** (The number of elements in [consensus_set] > d)
8:       **do** better_model = [k, F] suitable for all points in the [consensus_set]
        this_num = The number of elements
9:     **if** (this_num > best_num)
10:       **do** best_model = better_model
        best_num = this_num
        best_consensus_set = consensus_set
11:     **iteration++**
12: **End For**
13: **Output** best_model, best_consensus_set, best_num

---

## 4. RESULTS

### 4.1 Retrieved Result of SIF

We compare the retrieved SIF value with OCO-2 SIF and MODIS vegetation indices to evaluate the reliability of the algorithm. Figure 4 shows the global distribution of the retrieved TanSat SIF, OCO-2 SIF, NDVI, and EVI. All products were resampled to a uniform resolution under ArcGIS 10.5 software.
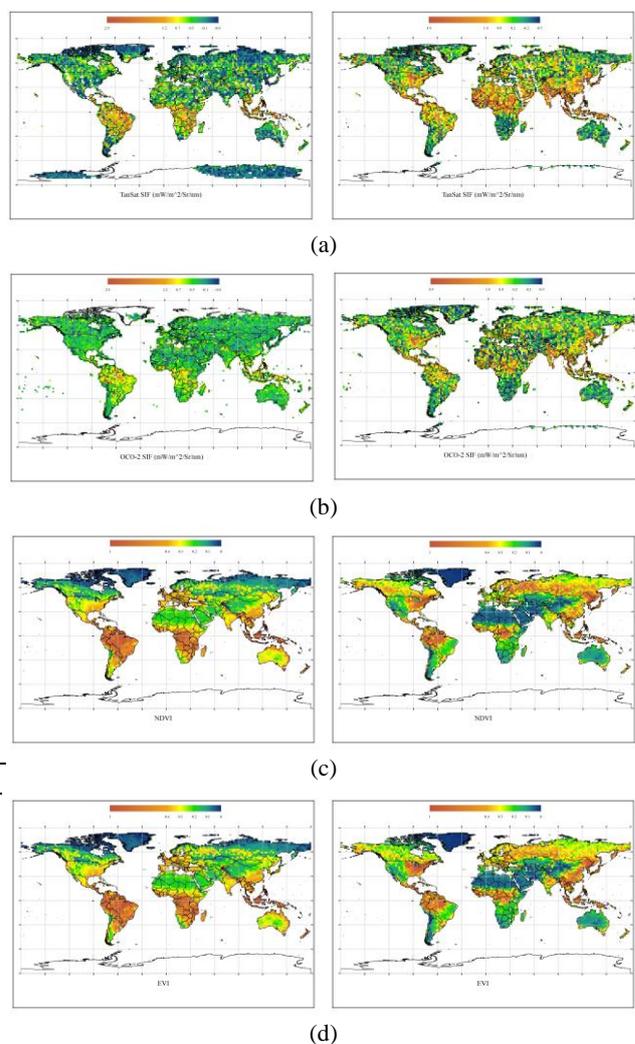


(a)



(b)



(c)



(d)

Figure 4. Global maps of retrieved TanSat SIF (a), OCO-2 SIF (b) products and vegetation indices including NDVI (c) and EVI (d). The date are March (left) and August (right) 2018.

It is obvious that the global change of SIF is closely related to the season. In March, the direct solar point was in the southern hemisphere, so higher SIF values appeared in South America, southern Africa, and near the equator. Meanwhile, the global vegetation index (NDVI and EVI) distribution shows the same trend. In August, due to the enhanced photosynthesis of vegetation in the northern hemisphere, higher SIF values appeared in the eastern United States, eastern Asia, and near the equator, Europe and Siberia also show higher vegetation activity.

It can be seen that the global distribution of several products has highly consistent. However, retrieved TanSat SIF can provide richer details than OCO-2 SIF, such as the SIF distribution in North America in March 2018 (Figure 5). But it is worth noting that TanSat has more strip loss than OCO-2 (Red circle position in Figure 5). The preliminary verification shows the consistency between the retrieved TanSat SIF and the actual situation.
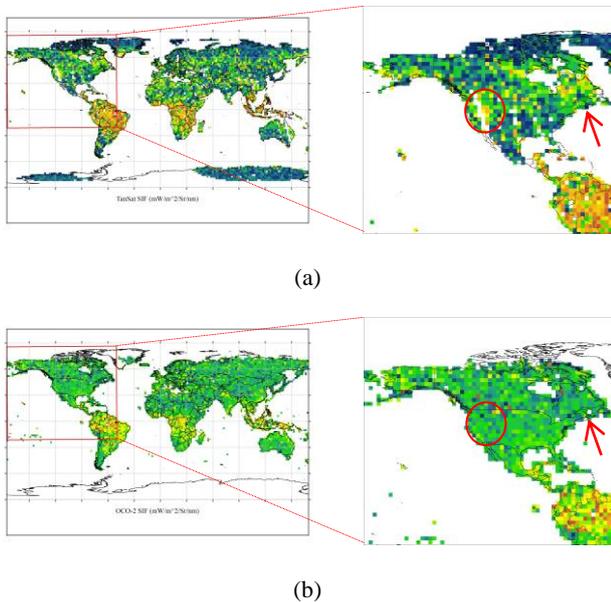


(a)



(b)

Figure 5. Retrieved TanSat SIF (a) and OCO-2 SIF (b) product distribution in North America in March 2018. (The difference in details is more obvious where the arrows point.)

### 4.2 Quantitative comparison

To quantitatively investigate the reliability of the retrieved TanSat SIF against other remote sensing datasets, scatter plot and Box-plot were drawn in this paper, as shown in Figure 6 and Figure 7. It should be noted that several locations scattered on the land which have avoided the effects of default values and cloud occlusion were randomly selected for statistics to generate a scatter plot in this study.

In order to evaluate the reliability of SIF products retrieved from TanSat data, OCO-2 SIF Products and other satellite-based vegetation datasets (MODIS Vegetation Index Products) were used for comparison to overcome the problems caused by the lack of direct observation data of SIFs in global regions which is a common practice in related research
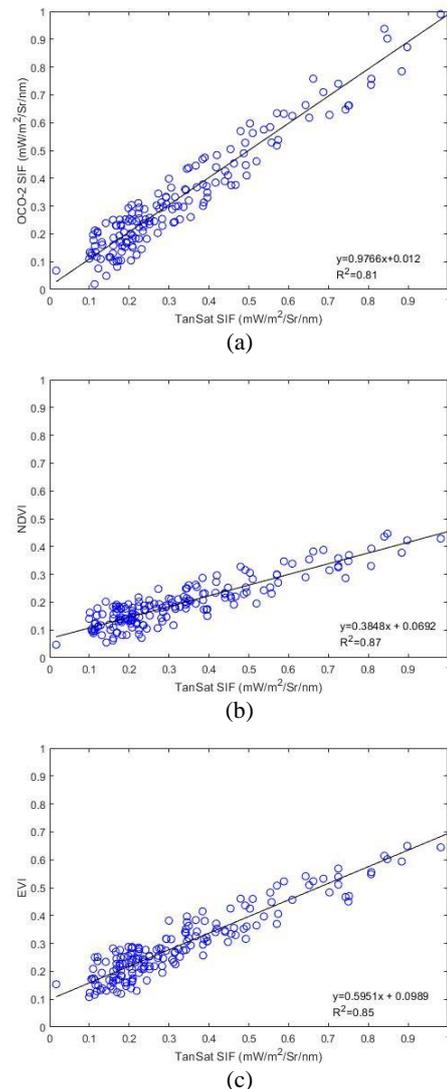


(a)



(b)



(c)

Figure 6. Scatter plots of retrieved TanSat SIF against OCO-2 SIF (a), MODIS vegetation indices (NDVI: (b) and EVI: (c))

As shown in Figure 6, retrieved TanSat SIF is consistent with other remote sensing products. Further comparison with OCO-2 SIF shows that it has a linear relationship with OCO-2 SIF ($R^2$ = 0.81), and most of the points are located near the $x = y$ line, which indicates that the high-level consistency between retrieved TanSat SIF and OCO-2 SIF. A comparison with the vegetation index (NDVI and EVI) found that retrieved TanSat SIF also had a linear relationship with NDVI ($R^2$ = 0.87) and EVI ($R^2$ = 0.85).
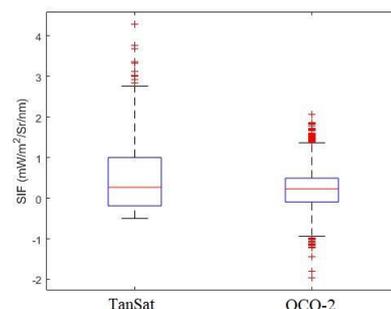


Figure 7. Box-plot of retrieved TanSat SIF against OCO-2 SIF

Figure 7 shows the difference between the data distribution of retrieved TanSat SIF and OCO-2 SIF. The two products are consistent medians, but the data distribution of retrieved TanSat SIF is relatively more scattered, especially the part larger than the median. The difference between the two groups of values is relatively large, which may be due to random errors generated during satellite observations and changes brought by the time difference between satellites covering the same area. In contrast, the OCO-2 SIF has more outliers, which may be due to the filtering of abnormal results by the RANSAC algorithm.

In summary, through a series of quantitative studies, it has been found that retrieved TanSat SIF and related products showed consistency and extra advantages, which proves that the SIF value retrieval based on the RANSAC algorithm is feasible.

## 5. CONCLUSION

Chlorophyll fluorescence is closely related to plant photosynthesis, and detecting chlorophyll fluorescence by remote sensing has important application value. At present, limited by remote sensing detectors and retrieval methods, chlorophyll fluorescence cannot be accurately retrieved.

Based on the random sample consensus algorithm, this paper used TanSat satellite's ACGS sensor data to retrieve the global SIF values in March and August 2018. It was found that both in the global spatial distribution and data consistency, retrieved TanSat SIF showed correlations with other satellite products (OCO-SIF, NDVI, and EVI). Compared with OCO-SIF, the SIF retrieval product in this paper has advantages in spatial detail performance and outlier data distribution. This article has proved that SIF products retrieved using random sample consensus algorithm have certain reliability.

However, the method in this paper still needs to be improved. The basic retrieve equations used in this paper ignore the effects of the atmosphere (such as temperature, pressure, aerosol, etc.), which will bring uncertainty to the result. Besides, the resolution of the satellite has an inestimable impact on the ground SIF retrieval, and the difference in scale also brings considerable difficulties to the SIF verification. Next, it is positive to apply the RANSAC algorithm to OCO-2 data and compare it with its SIF products.

Fortunately, the FLuorescence EXplorer (FLEX) mission, which will be launched by ESA (European Space Agency) in 2022, provides expectations for solving these problems.

## ACKNOWLEDGEMENTS

## REFERENCES

Alonso, L., Gómez-Chova, L., Vila-Francés, J., Amorós-López, J., Guanter, L., Calpe, J., Moreno, J., 2008. Improved Fraunhofer Line Discrimination method for vegetation fluorescence quantification. IEEE Geoscience and Remote Sensing Letters, 5(4), 620-624.

Damm, A., Elbers, J. A. N., Erler, A., Gioli, B., Hamdi, K., Hutjes, R., Kosvancova, M., Meroni, M., Miglietta, F., Moersch, A., Moreno, J., Schickling, A., Sonnenschein, R., Udelhoven, T., van der Linden, D., Hostert, P., Rascher,U., 2010. Remote sensing of sun‐induced fluorescence to improve modeling of diurnal courses of gross primary production (GPP). Global Change Biology, 16(1), 171-186.

Fischler, Martin A., Bolles, Robert C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the Acm, 24(6), 381-395.

Frankenberg, C., Butz, A., Toon, G. C., 2011. Disentangling chlorophyll fluorescence from atmospheric scattering effects in $O_2$ A‐band spectra of reflected sun‐light. Geophysical Research Letters, 38(3).

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J. E., Toon, G. C., Butz, A., Jung, M., Kuze, A., Yokota, T., 2011. New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. Geophysical Research Letters, 38(17).

Frankenberg, C., O'Dell, C., Guanter, L., McDuffie, J., 2012. Remote sensing of near-infrared chlorophyll fluorescence from space in scattering atmospheres: implications for its retrieval and interferences with atmospheric $CO_2$ retrievals. Atmospheric Measurement Techniques, 5(8), 2081-2094.

Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Köhler, P., Pollock, R., Taylor, T. E., 2014. Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2. Remote Sensing of Environment, 147, 1-12.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., Ferreira, L. G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote sensing of environment, 83(1-2), 195-213.

J. Joiner, Y. Yoshida, A. P. Vasilkov, Y. Yoshida, L. A. Corp., 2010. First observations of global and seasonal terrestrial chlorophyll fluorescence from space. Biogeosciences Discussions, 8(3), 637-651.

Liu, L., Cheng, Z.,2010. Detection of vegetation light-use efficiency based on solar-induced chlorophyll fluorescence separated from canopy radiance spectrum. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3(3), 306-312.

Liu, L., Zhao, J., Guan, L., 2013. Tracking photosynthetic injury of Paraquat-treated crop using chlorophyll fluorescence from hyperspectral data. European Journal of Remote Sensing, 46(1), 459-473.

Liu, X., Liu, L., 2013. Retrieval of chlorophyll fluorescence from GOSAT-FTS data based on weighted least square fitting. J. Remote Sens, 17, 1518-1532.

Maier, S. W., Günther, K. P., Stellmes, M., 2003. Sun-induced fluorescence: A new tool for precision farming. Digital imaging and spectral techniques: Applications to precision agriculture and crop physiology. Digital imaging and spectral techniques, 209-222.

Mazzoni, M., Falorni, P., Del Bianco, S., 2008. Sun-induced leaf fluorescence retrieval in the O 2-B atmospheric absorption band. Optics express, 16(10), 7014-7022.

Plascyk, J. A., Gabriel, F. C., 1975. The Fraunhofer line discriminator MKII-an airborne instrument for precise and standardized ecological luminescence measurement. IEEE Transactions on Instrumentation and measurement, 24(4), 306-313.

Porcar-Castell, A., Tyystjärvi, E., Atherton, J., Van der Tol, C., Flexas, J., Pfündel, E. E., Moreno, J., Frankenberg, C., Berry, J. A., 2014. Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges. Journal of experimental botany, 65(15), 4065-4095.

Van der Tol, C., Berry, J. A., Campbell, P. K. E., Rascher, U., 2014. Models of fluorescence and photosynthesis for interpreting measurements of solar - induced chlorophyll fluorescence. Journal of Geophysical Research: Biogeosciences, 119(12), 2312-2327.

Yang, Z., Zhen, Y., Yin, Z., Lin, C., Bi, Y., Liu, W., Wang, Q., Wang L., Gu, S., Tian, L., 2018. Prelaunch radiometric calibration of the tansat atmospheric carbon dioxide grating spectrometer. IEEE Transactions on Geoscience and Remote Sensing, 56(7), 4225-4233.

Zhang, L., Wang, S., Huang, C., 2018. Top-of-atmosphere hyperspectral remote sensing of solar-induced chlorophyll fluorescence: A review of methods. Remote Sens, 22, 1-12.