WORK IN PROGRESS: SIF Retrieval Using Neural Networks: A Satellite-Based Approach

Vrije Universeit Amsterdam

Supervisor: Gerbrand Koren and Sandjai Buhlai

Carmen Oliver Huidobro

From February 1, 2025 to March 11, 2025

Introduction

Solar Induced Fluorescence (SIF) is the faint glow emitted by plants during photosynthesis. Since SIF is directly linked to the photosynthetic process, it serves as a promising proxy for monitoring plant health, carbon uptake, and overall ecosystem productivity. With Earth's rainforests, especially the Amazon, playing a crucial role in regulating the global climate, remote sensing of SIF allows us to study these ecosystems in a non-destructive way. Moreover, while dense canopies like those in the Amazon pose a challenge for *in situ* measurements, satellite data provide extensive spectral, spatial, and temporal coverage that can help bridge this gap.

Satellite Overview and SIF Equation

Our study uses data from the TROPOspheric Monitoring Instrument (TROPOMI), a key sensor aboard the Sentinel-5 Precursor satellite. TROPOMI is designed to measure various atmospheric constituents with high spectral resolution, providing critical data for air quality, climate research, and increasingly, vegetation studies. Its multispectral capabilities allow us to capture both the reflected sunlight and the subtle SIF signals emerging in the far-red spectral region (typically 730–780 nm). We use Level-2 (L2) data, which is preprocessed and of higher quality than Level-1 (L1) data. Although this processing drops information that does not meet quality criteria, its enhanced data quality comes at the expense of spatial and temporal coverage compared to L1.

To retrieve SIF from the satellite data, we use the following radiative transfer equation:

$$R(\lambda, \mu, \mu_0) = a_s(\lambda)\, T^{\downarrow}(\lambda, \mu_0)\, T^{\uparrow}(\lambda, \mu) + \frac{\pi\, I_{SIF}(\lambda)}{E_0(\lambda)\, \mu_0}\, T^{\uparrow}(\lambda, \mu)$$

- $\mathbf{R}(\lambda, \mu, \mu_0)$ is the top-of-atmosphere reflectance at wavelength λ, observed at a viewing angle with cosine μ and solar zenith angle with cosine μ_0.
- $\mathbf{a_s}(\lambda)$ is the surface albedo (the fraction of incident light reflected by the surface) at wavelength λ.
- $\mathbf{T}^{\downarrow}(\lambda, \mu_0)$ and $\mathbf{T}^{\uparrow}(\lambda, \mu)$ represent the atmospheric transmittances for the downward and upward paths, respectively. These quantify the fraction of solar radiation that reaches the surface and the fraction of surface-emitted radiation that reaches the sensor after traversing the atmosphere.
- $\mathbf{I_{SIF}}(\lambda)$ is the solar-induced fluorescence intensity emerging from the vegetation.
- $\mathbf{E_0}(\lambda)$ is the extraterrestrial solar irradiance, which normalizes the observed signal.

In the following Figure 1, one can observe the reflectance over the Sahara captured by TROPOMI, which serves as the basis for our study:

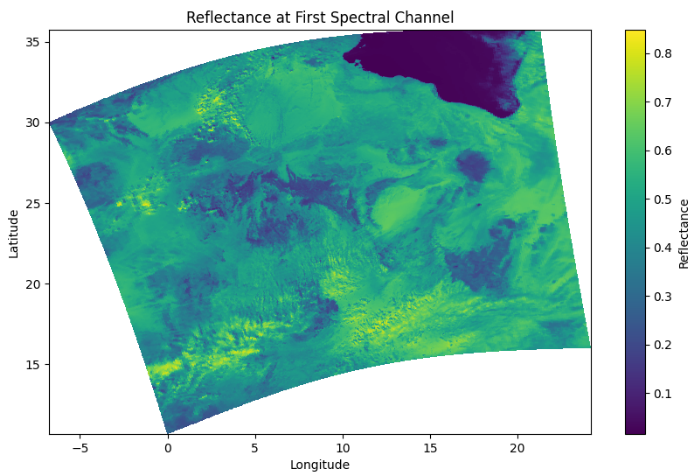


Figure 1: Reflectance data from TROPOMI over the Sahara

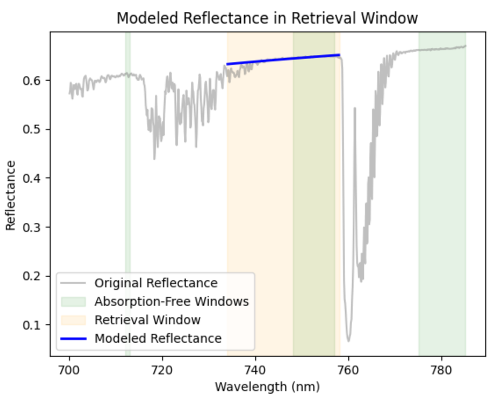
In our workflow, we first assume conditions where the vegetation contribution (i.e., SIF) is negligible (as in desertic areas) to validate the model. By setting $I_{SIF}(\lambda) = 0$, the equation above simplifies and we can then obtain transmittance values—a key variable in our subsequent SIF retrieval efforts.

Methodology

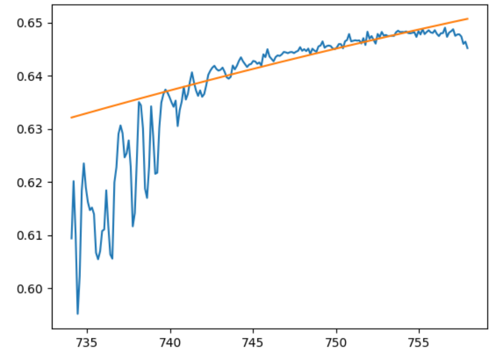
Modeling Surface Albedo from Reflectance

It is common practice to model surface albedo using reflectance data measured in spectral non-absorbing window regions, wavelength intervals where absorption by atmospheric gases (e.g., water vapor, CO_2) is minimal. In these windows, although scattering by molecules and aerosols remains significant, the negligible absorption simplifies the radiative transfer, making the observed reflectance predominantly a function of the surface properties and scattering effects. This approach enables us to reliably derive the surface albedo $a_s(\lambda)$ directly from the reflectance measurements.

In our approach, we construct a model for the albedo using the reflectance in three non-absorbing windows. The model then describes the albedo for the retrieval window of interest (730–758 nm, divided into 195 spectral channels). For each of the 406 scanlines, we form a matrix (406 rows $\times$ 195 columns) of modeled albedo values.



(a) Modeled albedo with the TROPOMI relfectance and the different spectral windows



(b) Close up of modeled albedo and reflectance in retrieval window

Atmospheric Transmittance in Desertic Conditions

To retrieve the atmospheric transmittance, we start with data acquired over desertic regions (e.g., the Sahara) where vegetation is sparse and the SIF contribution is negligible. Under these conditions, the retrieval equation simplifies (with $I_{SIF}(\lambda) = 0$):

$$R(\lambda, \mu, \mu_0) = a_s(\lambda)\, T^{\downarrow}(\lambda, \mu_0)\, T^{\uparrow}(\lambda, \mu).$$

This relation allows us to derive the atmospheric transmittance.

For modeling purposes, we approximate the transmittance using an expression inspired by the Beer–Lambert law:

$$T^{\uparrow}\, T^{\downarrow} = e^{-\tau(\lambda)\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)},$$
$$T^{\uparrow} = e^{-\frac{\tau(\lambda)}{\mu}},$$

where $\tau(\lambda)$ is the optical depth. Although the Beer–Lambert law is classically used to describe absorption in a homogeneous medium, it provides a reasonable approximation in our desertic scenario where scattering and absorption are relatively simple and multiple scattering effects are minimal. Consequently, we retrieve the optical depth via:

$$\tau(\lambda) = -\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right) \log\left(\frac{\text{reflectance}}{\text{surface albedo}}\right).$$

Figure 3 displays the optical depth computed for all 406 scanlines.

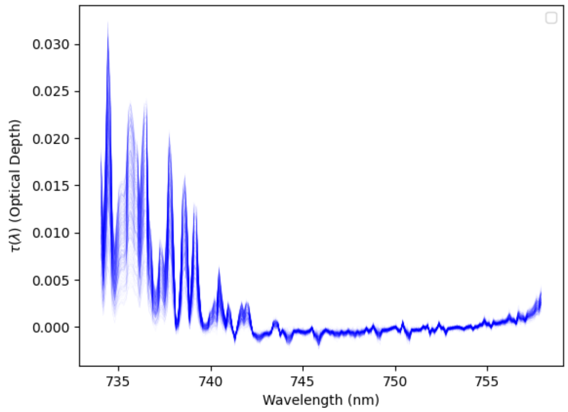


Figure 3: Optical Depth (τ) for 406 Scanlines

Neural Network Modeling of Atmospheric Transmittance

To capture the complex behavior of atmospheric transmittance—and by extension, the optical depth τ—we implement neural network models. Inspired by biological neurons each artificial neuron computes a weighted sum of its inputs, adds a bias, and applies a non-linear activation function:

$$y = g\left(\sum_i W_i x_i + b_i\right).$$

We experiment with differernt architectures of Neural Networks where we try to fit the data of a single time step, 406 different scanlines over 195 spectral channels that determine our retrieval window. where this data described the atmospheric optical density over the Sahara. A linear network fit resulted in a modest R^2 score of 0.1579. Incorporating non-linear activation functions such as `tanh` improved performance significantly, achieving an R^2 score of 0.6689. More advanced architectures like convolutional neural networks (CNNs) have yielded near-perfect predictions with an R^2 of 0.9987 on desertic data.

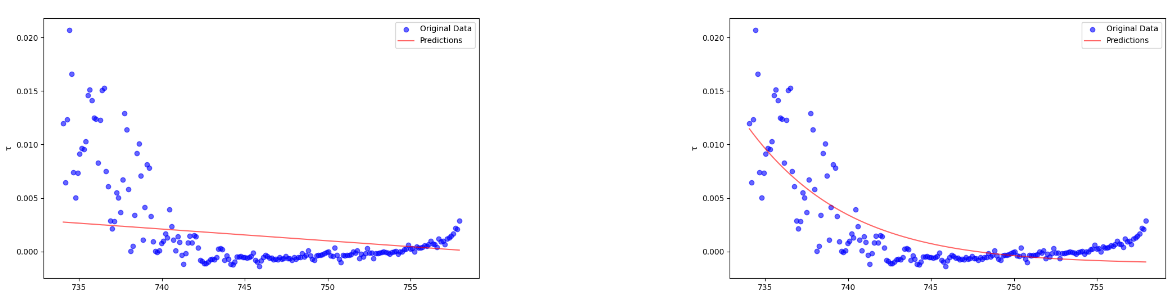


Figure 4: Neural Network fits for τ: (Left) Linear fit; (Right) `tanh` activation

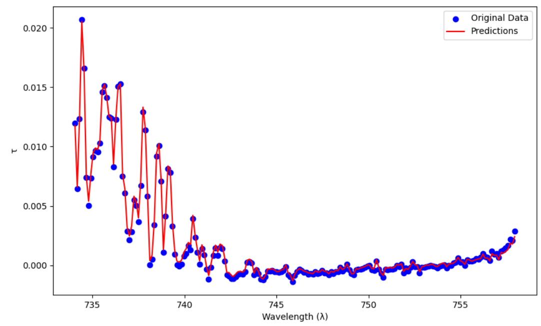


Figure 5: CNN fit to τ over 406 scanlines

I get an almost perfect fit with a CNN, but how is this giving me anything usefull? Now it is basically giving me a similar data set as the median of all the transmittance levels at different scanlines. We need a model that is more general and adaptive: Capture different modes of variability in transmittance (e.g., different aerosol levels, water vapor conditions). We need it to adapt dynamically to different input conditions instead of just smoothing everything. Instead of predicting just a single transmittance curve, we could train a probabilistic model: using a Variational Autoencoder (VAE) or a Mixture Density Network (MDN) to learn a distribution of possible transmittance curves, rather than just regressing to a single function. Instead of just using MSE, we could train the CNN to predict principal components of transmittance and reconstruct transmittance curves from those components. This would force the model to learn dominant modes of variation rather than just averaging.

Retrieval of SIF Data

Having modeled with a high R^2 value we decide to start the SIF retrieval to understand what could go wrong and right and therefore make smarter decisions earlier on. In non-vegetated regions, our method successfully yields near-zero SIF values with only minor residual errors (see Figure 6).

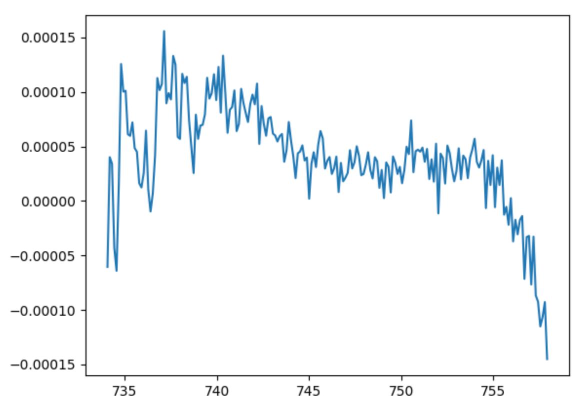
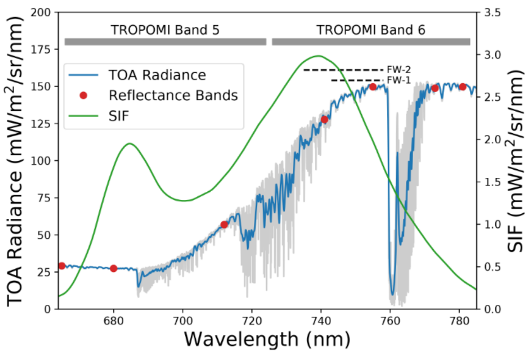
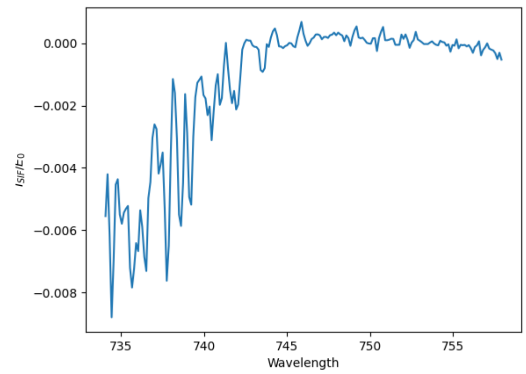


Figure 6: Retrieved SIF in a Sahara region from the modeled transmittance.

This result confirms that the retrieval process correctly identifies the absence of fluorescence up to three decimals, there can still be some improvement. Moreover, a comparison with SIF measurements from previous studies [Gua+21] in vegetated areas (Figure 7a) shows a close spectral shape, lending further credibility to our methodology. While many further refinements are necessary, these promising preliminary results demonstrate the potential of our approach to accurately extract SIF data from satellite observations.



(a) SIF over vegetated areas from previous work [Gua+21].



(b) Retrieved SIF over the Amazonas with modeled transmitance

Next Steps: SIF Retrieval in the Amazon

- Refining Neural Network Architectures: Exploring advanced CNNs and hybrid models to better capture spatial and spectral correlations inherent in heterogeneous environments.
- Validation with Ground-Based Measurements: Comparing retrieved SIF values with *in situ* observations to enhance accuracy.
- Extending Spectral-Spatial-Temporal Analysis: Integrating additional spectral channels, spatial pixels, and time samples to more comprehensively account for atmospheric scattering and absorption.
- Benchmarking Against Classical Methods: Comparing our neural network approach with traditional statistical analyses developed by the research group of Gerbrand Koren.
- Analyzing Extreme Events and SIF Ratios: Leveraging SIF data to detect and characterize extreme events (e.g., droughts, fires, and floodings) while investigating the red:far red SIF ratio as a potential indicator of vegetated area performance in the Amazon.

By adapting and extending our methodology validated in desertic regions, we aim to reliably retrieve SIF in the Amazon, providing crucial insights into the health and productivity of one of Earth's most vital ecosystems. Furthermore, assessing the performance of neural networks in this context will help determine their advantages over conventional methods for SIF retrieval.

References

[Gua+21] Luis Guanter et al. “The TROPOSIF global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission”. In: *Earth System Science Data* 13.11 (2021), pp. 5423–5440.