Part 4: Questions & Answers Sessions

Please type your questions in the Question Box. We will try our best to get to all your questions. If we don't, feel free to email Erika Podest (erika.podest@jpl.nasa.gov).

Question 1: What are the differences between physical and data-driven retrieval methods?

Response 1: In principle these are two different retrieval mechanisms. With "physical" we mean that we model and fit fluorescence based on first principles. It means that we assume a theoretical solar spectrum and we have a well defined instrument line shape that we use to convolve the solar spectrum with and thereby generate synthetic data. We call this process an instrument model. All we need then is some sort of assumption about the spectral shape of SIF and then all of this can be constructed into a simple forward model with which we can generate synthetic measurements used for the retrieval process.

In the data driven approach we train the instrument model from the data itself. We use data from on-orbit satellite measurements, typically over fluorescence-free scenes to train the instrument models to include everything going into the detector, most importantly the spectral line shape that determines how each detector sees a Fraunhofer line. This method is typically done when we use larger spectral ranges as subtle instrument features can affect the retrievals. We often find that we can better capture the instrument models by training the model using actual data.

Question 2: Can we use Python or R to process OCO-2 data? What is your suggestion to process the data in those platforms?

Response 2: You should use the tools that you are most familiar with, which is often the easiest way. If you are an expert in one code, we do not want you to switch your main processing platform. You can access netCDF data in any of these scripting languages as long as you have a netCDF or HDF library installed. There are examples on the github link below with R and Python scripts.
https://github.com/cfranken/SIF_tools

Question 3: How important is it to evaluate the performance of SIF against the eddy covariance flux tower GPP measurements?

Response 3: Comparing SIF against flux tower observations is the best possible way to fully understand how SIF works and how it relates to GPP at the larger canopy scale. It helps understand under what conditions SIF will work well and under what conditions SIF-GPP relationships break down.

Towards that goal, these measurements are crucial. In some aspects it might work at the leaf level even better because you are actually measuring the same thing. At the flux tower site we have the opposite problem from the satellites. The fluorescence footprint measured from a flux tower might capture a much smaller area - in the vicinity of a couple of meters around the tower than the actual flux tower footprint. A satellite covers a much larger area than the tower footprint. These are some of the caveats to keep in mind, especially if there is heterogeneous vegetation within the eddy covariance flux tower footprints, which can be hundreds of meters.

Question 4: Why do you prefer to use Julia and what is the advantage of using Julia? Can't we use Python instead? What are the advantages of Pluto over Jupyter or RMarkdown? What are the advantages of Julia over R or Python?

Response 4: Good question, you can use the language of your choice, the basic principles of data processing remain the same. We have switched over to Julia for its speed and simplicity, allowing production speed (like C/Fortran) in a modern scripting language. However, at times this can be a personal choice and you have to decide for yourself which language is most "efficient" for you. The advantage of Pluto over Jupyter is that it is more interactive, i.e. content will be updated throughout all cells if variables change. If you'd like to learn more, the following link contains a good discussion:
https://towardsdatascience.com/r-vs-python-vs-julia-90456a2bcbab
Notebooks are nice for fundamentals, but not for working on pure script, so keep that in mind. Single SIF data is very noisy and needs to be averaged in either space or time, hence accessing it via GEE may not be feasible yet.

Question 5: Is it possible to map dormant volcanoes' lithology by the fluorescence of some minerals? Or only lithological mapping could be inferred somehow from active volcanoes' emissions?

Response 5: Volcanoes are typically non-fluorescent, but you are right, some minerals can luminesce. Typically the luminescence signal is even weaker than vegetation fluorescence and for most minerals the luminescence does not take part in this 750 nm range but more on the visible spectral range. In general the signal is much weaker than vegetation fluorescence and you would need many more measurements to average this

out in order to reveal areas of luminescence (due to the relatively high measurement uncertainties).

Question 6: In the last session the SIF data were referred to as retrievals, in this session they are being referred to as soundings, is there a difference?

Response 6: It is a terminology that we switch back and forth. In the framework here, a sounding and retrieval are almost synonymous, in both cases they are related to a single measurement from the satellite. Strictly speaking, a retrieval is the process from which we compute fluorescence from the spectra for each sounding. For sounding we mean a single measurement from the satellite that might be associated with the retrieved fluorescence value for that sounding or the spectrum from that sounding. It is referring to a single measurement point.

Question 7: I am not very comfortable using RS/GIS softwares. Can you please show/share in GitHub repository a demo on processing the NetCDF data to extract a time series over a region of interest?

Response 7: The code for this tutorial can be accessed at:

https://github.com/philag/TROPOMI-OCO-2_SIF_DEMO

We have further examples here:

https://github.com/cfranken/SIF_tools
https://github.com/cfranken/gridding

Most often if you do not want to use a scripting language you will have to work with the gridded Level 3 datasets and not the individual datasets.

Question 8: Is there a way to link SIF with LAI?

Response 8: Yes, you would need to download an LAI data set and co-locate both data sets. There are several strategies, for example using gridded data with the same spatial and temporal resolution.

Question 9: Can SIF data be used to study vegetation phenology?

Answer 9: Yes, mostly because phenology strongly impacts the amount of absorbed radiation by chlorophyll, which is perfectly captured in the fluorescence seasonal cycle. Crop areas probably have the strongest vegetation phenology because things are being harvested and completely cut down over the winter period (though sometimes there are winter crops in some areas). The phenology is very narrow in areas where crops grow incredibly fast in the growing season but are then harvested shortly thereafter. This very distinct highly peaked seasonal cycle and phenology can be captured very well with

fluorescence. It is mostly driven by changes in the amount of absorbed radiation by chlorophyll.

Question 10: Can we use a shapefile directly instead of a KML file to define the area of interest?

Response 10: Yes, you can use the Pluto notebook if your shapefile is a geoJSON file. Otherwise, there are other packages you may use to read shapefiles in a different format or convert to a geoJSON first. GIS softwares have an "In polygon" function where you can check if a footprint is within your shapefile, which might be helpful.

Question 11: Can the OCO-2 SIF Lite product be extracted more easily? In documentation i.e. Panoply is recommended, but I cannot get i.e. Daily_SIF_757 nm. Also, I once downloaded a subset over a specific area, but it is just a non georeferenced stripe, and it is not every 16 days, but daily, so I don't understand the temporal resolution.

Response 11: OCO-2 SIF Lite is a level 2 product. This means that it provides a measurement point for each and every single sounding from OCO-2. The SIF Lite product does not split up the data into satellite orbits, which is what is typically done with the raw Level 2 data. SIF Lite combines all the different orbits per day into a single daily file. The files are not gridded, they just contain the orbital overpasses over a single day, which is why they look like non-georeferenced stripes.
We use the gridding routine below to ingest these OCO-2 SIF Lite products and then specify the spatial and temporal resolution desired for your final product.
https://github.com/cfranken/gridding
In our ftp server you will be able to find some Level 3 gridded products.
The demo data used in this tutorial was Level 2 from Tropomi and OCO-2.

Question 12: Can you also explain the Level 3 products?

Response 12: Levels refer to processing stages (visit
https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels *for a detailed explanation). Level 3 means that the data is mapped on uniform space-time grid scales. There is no official SIF Level 3 product, because there are several ways to filter the data before gridding. Also, the temporal resolution (gridding the data for 1week/month,etc.) depends on the users' needs. Once you run the script presented in the tutorial in Part 4, you get a gridded product.*
For more examples, you can download several gridded (Level 3) data sets here:
ftp://fluo.gps.caltech.edu/data/tropomi/gridded/SIF740/

Question 13: How does GOME-2 SIF retrieval perform in comparison with OCO-2 and TROPOMI?

Response 13: In principle, GOME-2 SIF measurements are very similar to TROPOMI. GOME-2 data however, should be evaluated with care (especially with respect to trends) since there are known sensor degradation issues, which have not yet been resolved.

GOME-2 has slightly worse detector characteristics (in terms of signal-to-noise ratio, spatial footprint size, revisit time, and spectral resolution) since it is an older instrument. The TROPOMI retrieval uses a much narrower spectral range (thanks to the improved spatial resolution), which is a key advantage. Overall, a single sounding from TROPOMI is more reliable than GOME-2. The biggest disadvantage of GOME-2 is that its footprint is very large, on the order of 80km whereas for TROPOMI it is on the order of 5km. There is much more data for TROPOMI and better accuracy for an individual sounding, allowing for a much finer spatial and temporal resolution.

OCO-2 is the most accurate from a retrieval point of view because it has by far the highest spectral resolution but it comes at the expense of not being a real mapper.