Applied Remote Sensing Training (ARSET) Program

Solar Induced Fluorescence (SIF) Observations for Assessing Vegetation Changes Related to Floods, Drought, and Fire Impacts

Homework Questions

Question 1

Solar-Induced Fluorescence or SIF is a red to far-red signal emitted by excited chlorophyll molecules in vegetation.

Answers: (bold correct)

a. **True**
b. False

Feedback:
This statement is true. Plants harvest light and carbon dioxide to produce energy through photosynthesis, a reaction that occurs in the chloroplasts of plant cells. The main two reaction centers in the chloroplast are called Photosystem I (PSI) and Photosystem II (PSII). It is through a process called "photochemical quenching" in PSII that causes the plant to emit SIF photons, and this fluorescence has a peak in the red end of the visible spectrum.

Question 2

What were some of the different science applications for using SIF?

Answers: (bold correct)

a. Climate anomalies
b. Crop yield predictions

c. Drought impacts
d. All of the above

Feedback:
All the examples include studying the effects on vegetation. Excessive heat and insufficient water are key factors in affecting vegetative photosynthesis and the SIF signals. SIF provides the plants' response to various known external conditions such as heat waves and farming practices.

Question 3

Which of the following statements describe advantages to using SIF for observing plant health compared to other measures?

A. SIF is a direct emission with mechanistic ties to photosynthesis
B. It can detect changes in vegetation stress before changes appear in reflectance indices (e.g., NDVI)
C. It has low sensitivity to deep clouds
D. It is correlated with carbon fixation rather than just greenness

Answers: (bold correct)

a. Statement A only
b. Statements A, B and C
c. Statements A, B and D
d. Statements A, B, C and D

Feedback:
While SIF can be used to detect vegetation health with lower hysteresis than reflectance indices, it is still obscured by optically thick clouds when observed from space. The quality flag that we use in the code exercises helps us filter data points with heavy cloud cover.

Question 4

Space-based SIF measurements are available only from NASA.

Answers: (bold correct)

a. True

b. **False**

Feedback:
This statement is false. NASA (OCO-2, OCO-3), the European Space Agency/ESA (GOME, GOME-2, SCIAMACHY, TROPOMI), the Japan Aerospace Exploration Agency/JAXA (GOSAT, GOSAT-2, GOSAT-GW) all offer space-based SIF data for monitoring and studying global vegetation and health. Planned ESA missions including CO2M and FLEX will continue this important measurement.

Question 5

In the lessons, we learned about two key challenges of using Level 2 SIF data from OCO-2 and OCO-3: single sounding retrievals may have negative values (which is not possible in a physical sense) and the samples cover narrow strips of area in each orbit. What are some of the ways we can mitigate these limitations?

Answers: (bold correct)

a. Gap-filling with other data sets
b. Spatial gridding and averaging
c. Rejecting negative values
d. a & b
e. a & c

Feedback:
Gap-filled Level 3 datasets, such as GOSIF, which uses MODIS reflectance to estimate SIF based on relationships trained on OCO-2 data, can help fill gaps where Level 2 data may be lacking. Over longer time periods or when working with SAM data, it is useful to create gridded rasters or spatio-temporal averages with reduced noise. Negative SIF values are still considered valid especially for analyzing spatial and temporal variability, and arise primarily due to retrieval noise.

Question 6

Which of the following is an advantage of using Snapshot Area Map (SAM) mode observations, like the ones used in the second Jupyter Notebook, , 2_oco3_sam.ipynb, over other types of SIF measurements from OCO-3?

Answers: (bold correct)

a. SAM samples are more accurate than other OCO-3 data since they are all taken over a short time window.
b. SAM observations build up a record of a small region through repeated overflights, allowing for comparison with ground-based measurements.
c. SAM data are spatially contiguous so they can be used in place of gap-filled products.
d. All of the above

Feedback:
SAM observations are often taken in carefully chosen regions that align with ground-based stations to allow for validation and detailed analysis of vegetation in specific biomes. Individual SAM data points have the same or even greater uncertainty than other OCO-3 data. Although SAMs have greater spatial coverage than other OCO-3 data, they do not serve the same purpose as gap-filled data.

Question 7

SIF and Gross Primary Production (GPP) are known to be directly correlated with one another, but in our Jupyter notebook exercise on SAMs, 2_oco3_sam.ipynb, we saw a good correlation at the Michigan site and poor correlation at the Oregon site. What are some reasons for the poor correlation at the Oregon site?

Answers: (bold correct)

a. Changes in land cover type around the Oregon flux tower site over the course of the study period.
b. The data is noisy and more outlier points need to be discarded.
c. The Oregon tower site experienced a wildfire during the study period.
d. All of the above
e. a & c

Feedback:
While there are optimizations that can be made by rejecting SAM points with land cover that doesn't match the land around the tower site, the main cause of the poor SIF-GPP correlation at the Oregon site is that there was a wildfire in that region in 2020 and subsequent logging around the tower. Because we spatially average SAM points to reduce their uncertainty, there is scale mismatch between aggregated SAM observations and the small grassland area (1.5 mi x 0.5 mi or 2.5km x 0.85 km) around the Oregon tower.

Question 8

Gap-filled SIF products like GOSIF are a form of generative AI (genAI) that synthesize values by being fine-tuned from remote sensing foundation models.

Answers: (bold correct)

a. True
b. False

Feedback:
This statement is false. GOSIF and essentially all other gap-filling techniques in this area of remote sensing use data-driven or traditional machine learning (ML) approaches that produce estimates by developing correlations between carefully chosen predictor variables. Unlike generative deep learning models, these models have explainable results; i.e., the influence of each predictor can be quantified in the final estimate.

Question 9

In the third code exercise on GOSIF, we looked at the 2019 Midwestern floods and how they impacted the corn belt in the midwestern US. One conclusion we made was that SIF was overall lower in 2019 (the flood year) compared to 2018, our control year. Review the time series plot we made at the end of the notebook. Why did we see a peak in SIF late in the 2019 season when SIF was already declining over the same period in 2018?

Answers: (bold correct)

a. Planting was delayed in the 2019 season, resulting in a corresponding later harvest.
b. Late-season heat waves in 2018 caused crops to dry out too early.
c. More water in the soil due to the flood allowed the corn and soy to "catch up" later in the season.
d. None of the above

Feedback:
Sometimes the SIF behavior we see can be explained by additional context about human behavior or the environment. In this case study, the Yin et al. paper we discussed mentioned that planting was delayed by as much as 4 weeks in 2019 because fields were flooded. While the GRACE-FO mission observed higher Terrestrial Water Storage (TWS) in 2019, this resulted in higher SIF-GPP activity primarily in wildland regions rather than over cropland.

Question 10

Take a moment to review the code in the Jupyter Notebook from exercise 1, `01_exploration.ipynb`. What format does NASA use to store OCO-2 and OCO-3 data and what Python modules can we use to open it?

Answers: (bold correct)

a. HDF4 format and the pyhdf module
b. NetCDF format and the netcdf4 module or xarray
c. GeoTIFF format and the gdal Python bindings
d. Zarr format and the zarr-python module

Feedback:
GES DISC archives OCO-2 and OCO-3 granules in the netcdf format, which can be opened and manipulated using the netcdf4 module or xarray. A separate Data Access Protocol (DAP) server is offered to allow remote access to the data without needing to download whole files, and this can be implemented in Python with the pydap module.

Question 11

Which of the following is a reason to use a gap-filled product (like GOSIF) for studying vegetation change as opposed to a gridded raster of instrument data (like the ones we made in Part 1)?

Answers: (bold correct)

a. Gap-filled products have reduced uncertainty compared to gridded rasters.
b. Gap-filled products account for biome type, and therefore ecosystem-level SIF behavior is easier to observe in the data compared to instrument data.
c. Gap-filled products have higher availability over time and space than instrument data, allowing for studies of regions and events that might not have been covered by the source instrument.
d. Gap-filled products take up less disk space than gridded rasters.

Feedback:
The primary reason to use gap-filled products like GOSIF is due to their high spatiotemporal coverage. By predicting SIF from MODIS predictor data, GOSIF provides 5 km/pixel global estimates of SIF over the long observational record of the Terra and Aqua missions, the first of which was launched in 1999.

In comparison, the OCO-2 data that was used to train the GOSIF model suffers from spatial gaps due to the narrow swath width of the instrument, and the observational record begins in 2014. While creating a gridded raster can mitigate spatial gaps, better coverage can only be achieved by averaging over a longer time window.

Optional Hands On Activities Between Parts

Please write instructions for each activity

PART 1

Instructions for Participants with links to resources needed:

Now that you have learned how to create and plot gridded rasters from OCO-2 and OCO-3 SIF data, try creating a dataset to compare the values obtained from the two instruments over the year 2021. Using the downloading and gridding code from the notebook, create a set of gridded rasters from OCO-2 data over at least 3 separate months in 2021, then do the same thing for OCO-3. These rasters can have global extent or can be subsetted to cover a region of interest to you. Next, average the SIF values from each raster and create a time series plot with one line representing the OCO-2 monthly average SIF, and the other line representing OCO-3 monthly average SIF. Create a screenshot of the time series plot when you are finished. How well do measurements from the two instruments agree with each other? Are there any notable differences?

PART 2

Instructions for Participants with links to resources needed:
The two sites we discussed in the Jupyter notebook, US-UMB and US-Me2, represent two different types of forest biomes found in the United States, broadleaf deciduous forest and evergreen needleaf forest, respectively. For this homework exercise, we will explore the phenological characteristics of a tropical forest using the same analysis technique that we learned in the notebook. First, take a look at the Ameriflux page for La Selva Biological Station in Costa Rica: https://ameriflux.lbl.gov/sites/siteinfo/CR-Lse. You can find the corresponding SIF_High site on the OCO-3 SAM webpage by searching for "sif_La_Selva_Costa_Rica" under the Site Name: https://ocov3.jpl.nasa.gov/sams/index.php

Download the Ameriflux data for this site and note down the coordinates of the tower. Run the 4_appendix.ipynb notebook with the information you gathered to create a SIF dataset over the tower site from OCO-3 observations near this location. You may wish to filter the data to only include "evergreen broadleaf forest" biome pixels, which have an IGBP index of 2. Additionally, since the biological preserve extends primarily to the south of the research station, you may want to adjust your spatial aggregation in the notebook accordingly. Once you have completed the data preparation, run the cells in the notebook again with your new data. What do you notice about the phenological pattern in the data for this biome? How well do SIF and GPP correlate with each other at this site?

PART 3

Instructions for Participants with links to resources needed:
In our exercise using GOSIF, we studied the corn belt region of the Midwestern US. Due to climate, some crops like cane sugar are primarily grown outside of the US. In particular, most cane sugar imported into the US comes from Brazil and Paraguay. As a result of its strategic and economic importance, the USDA maintains reports on major crops produced in foreign countries through its Foreign Agricultural Service (FAS), and you can find the report on Brazilian sugar production here: https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=Sugar%20Annual_Brasilia_Brazil_BR2025-0011.pdf

Pay special attention to the note on wildfires on pages 18-20. The report notes that there was a significant drought intensified by the El Niño weather pattern that year and subsequent wildfires. Create a comparison of the GOSIF data from June - December 2024 (the drought/fire year) over São Paulo state compared to the same time period in 2023 (our control). Use the same animation and time series plotting code from the notebook, adjusting the time range of files to be downloaded and the geographic bounds of the data during the PNG preparation step as needed. What do you notice about the SIF behavior over the cropland in the state during the wildfires as compared to a normal year?

Hint: Use geographic bounds of 19°S - 25°S, 55°W - 45°W. There is a geojson file called "sao_paulo.geojson" in the "notebooks/" directory that you can use in place of the "corn_belt.geojson file from the main exercise.