Part 3 Questions & Answers Session

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 Amita Mehta (amita.v.mehta@nasa.gov).

Question 1: How many granules can we download at a time?

Answer 1: When a dataset has a significant amount of granules, you are limited to downloading 1,000,000 granules at a time due to processing time required. When this is true, the 'Download All' button will be grayed out until you select fewer granules to download. Granules will be downloaded from Earthdata Search individually (or follow the instructions while downloading to install the Earthdata Download tool). If using the earthaccess library in Python, the files are opened or downloaded individually.
Example:

In this link, I cannot download 'all' since the dataset contains a total of 1,520,754 granules:
https://search.earthdata.nasa.gov/search/granules?p=C1214470533-ASF&pg[0][v]=f&pg[0][gsk]=-start_date&tl=1728503767.239!3!!

If I narrow down my search by inputting parameters in the 'Filter Granules' panel to the left, the 'Download' button becomes active again, allowing me to retrieve a smaller number of granules to complete my transaction. In this case, I chose to limit my search by choosing temporal parameters, restricting my search to return granules from 1 June 2024 - 30 June 2024:
https://search.earthdata.nasa.gov/search/granules?p=C1214470533-ASF&pg[0][v]=f&pg[0][qt]=2024-06-01T00%3A00%3A00.000Z%2C2024-06-30T23%3A59%3A59.999Z&pg[0][gsk]=-start_date&tl=1728503767.239!3!!

Question 2: Is there any cloud contamination threshold bar to filter the cloudy images?

Answer 2: Yes, you can set a cloud cover threshold to Level 2 data using the earthaccess Python library. An example is shown in the tutorial. The Earthdata Search

GUI will include Cloud Cover in the “Filter Granule” box if the cloud cover information is available for the selected collection.

Question 3: Is it possible to use Google Colab as a Jupyter notebook?
Answer 3: You can upload the jupyter notebook into Google Colab and work with the data from there. Note that Google Colab is not in the AWS cloud, so the data will have to transfer from AWS to Google Colab over the internet.

Question 4: If we're looking for specific bands and need them in GeoTIFF format, can earthaccess handle the conversion so that only those specific bands in GeoTIFF format are downloaded?
Answer 4: The earthaccess library does not provide file format conversions. If you use “earthaccess.open” and “xarray.open_dataset”, then only the bands you request are streamed to you (but, in some cases this may be slower than downloading the whole NetCDF file).

Question 5: Why do PACE global composited data from the Ocean Color website have 4 km resolution? Is there a place to access global composited data that have 1 km resolution?
Answer 5: All global files are necessarily composites of multiple scenes at the native instrument resolution, and compositing works better at 4 km resolution. You can use the Level-2 granules for higher resolution, but would have to generate your own global composite from those granules (about 150 needed for OCI). The OB.DAAC is considering (eventually) providing the global data at 1 km resolution. However, the data are not ALL at 1 km. It is only **at nadir** that the sensor has a 1 km resolution. At the swath edges, the samping between pixels grows to as much as 7 km for OCI, so when binning at higher resolution, an area weighting needs to be applied as a nearest neighbor approach will result in gaps. There is also an issue of data volume. At 4 km resolution, the OCI RRS L3 bin file is on the order of 3GB for a single daily file. A 1km file would be 16X larger (~48GB). BTW, a monthly 4 km RRS bin file is about 16GB, or 256GB at 1 km.

Question 6: Can I use the dataset for the great lakes or inland water bodies?
Answer 6: PACE-OCI does collect data over inland water bodies. All PACE-OCI ocean color products provide data for inland water bodies IF they are of sufficient size to be

identified as water and are not blocked by cloud cover. The Great Lakes are large enough to be analyzed using OCI data, but not all inland water bodies are. The spatial resolution of PACE is 1.2km x 1.2km, so the smallest possible water body it can work with should be about 3 to 4 times that spatial resolution. For example, a 7 km^2 water body could not be studied by PACE-OCI, but a 25 km^2 could be studied. This is because mixed pixels that contain both land and water, or both ice and water, or that are within a couple of pixels of land or ice do not allow for accurate water surface reflectance retrievals. In contrast, a 25 km^2 water body (if roughly 5 x 5 km dimensions) would have roughly up to 9 central pixels with potentially good retrievals.

Question 7: How is the index of the dataset classified? Is it done with a classification algorithm reference or something else?
Answer 7: Can the meaning of “index of the dataset” be clarified? We will look into this further with clarification.

Question 8: How can one get the remote sensing reflectances (Rrs) of a particular point such as a bloom event using PACE data and create an algorithm based on that Rrs to map the bloom extent?
Answer 8: The second Jupyter Notebook tutorial in this training demonstrates how to plot an Rrs spectrum of single pixels. You can get spectra from SeaDAS as well.

Question 9: Are all the flags listed (ATMFAIL,LAND,HILT,HISATZEN,STRAYLIGHT,CLDICE,COCCOLITH,LOWLW,CHLWARN,CHLFAIL,NAVWARN,MAXAERITER,ATMWARN,HISOLZEN,NAVFAIL,FILTER,HIGLINT) automatically applied to the L3 data?
Answer 9: That looks like the “l2_flag_names” global attribute for L3m CHL granule, so that is correct. This site also explains what flags are applied to L3 data: https://www.earthdata.nasa.gov/apt/documents/rrs/v1.1.

Question 10: When you name a data package "results," does it override files of the same name from earlier code?
Answer 10: Yes, it does. Use different variable names to prevent that from happening.

Question 11: Can you please tell me how can I work with monthly and weekly binned data? Where can I change the code?

Answer 11: The “granule_names” argument is used in the second notebook from today showing an example of filtering for monthly files. For weekly (actually 8-day composites), use the filter “*.8D.*” instead of “*.MO.*”.

Question 12: What’s the difference between Rrs and Surface Reflectance? Can we use surface reflectance for terrestrial surface water monitoring?
Answer 12: Rrs is a radiance reflectance (as opposed to an irradiance reflectance) It is the ratio of upwelled radiance to downwelled irradiance. If the upwelled radiance field is Lambertian, Rrs would be a factor of PI different from surface reflectance… but yes, you can use surface reflectance to monitor terrestrial surface waters. For instance, CyAN monitoring uses surface reflectance.

Question 13: Can we download L3 data for a region of interest? Or do we have to download the global file and crop using xarray?
Answer 13: You can define a region of interest using the OB.DAAC Data Dashboard and set a subscription for data over that region. Navigate to the Ocean Color website @ https://oceancolor.gsfc.nasa.gov and go to “Quick Links” on the right of the navigation bar. Choose “Data Dashboard” from the dropdown. Here is the direct link: https://oceandata.sci.gsfc.nasa.gov/data_dashboard. **NOTE - Cloud data are accessible via Earthdata Search, which does not currently support subscriptions. That means subscriptions are not yet available for PACE data, but work is being done to allow this soon.** You can also choose to extract a region when you order data. Additionally, you can use the earthaccess Python library to access granules and open them with xarray. When using earthaccess and xarray in the AWS cloud, then you can stream only the data used without downloading the whole global file.

Question 14: Where can we find more information about the atmospheric correction and flags algorithms used in the product PACE OCI AOP Rrs?
Answer 14: You can read about how we calculate Rrs here: https://www.earthdata.nasa.gov/apt/documents/rrs/v1.1. It lists what flags are applied to L3 data.

Question 15: Is the available data from the cloud land-sea masked or do we do it in a different step? If so, are there online sources teaching how to do so?

Answer 15: You can apply a cloud mask using the SeaDAS analysis and visualization software. The website for the software is in the description of the video. Here is a tutorial: https://www.youtube.com/watch?v=kI6x3sDjFQQ.

Question 16: Which PACE product shows the color of the ocean from green to blue?
Answer 16: If you are referring to the L3 composite shown in the first Jupyter Notebook tutorial, that was plotting chlorophyll a concentration using a blue-green color palette.

Question 17: When using global mapped data, do we need to reproject the data if we want to observe an area of interest with the local projection, or keep it as a global projection?
Answer 17: You can define a region of interest using the OB.DAAC Data Dashboard and set a subscription for data over that region. **NOTE - Cloud data are accessible via Earthdata Search, which does not currently support subscriptions. That means subscriptions are not yet available for PACE data, but work is being done to allow this soon.** You can also choose to extract a region when you order data. Navigate to the Ocean Color website @ https://oceancolor.gsfc.nasa.gov and go to "Quick Links" on the right of the navigation bar. Choose "Data Dashboard" from the dropdown. Here is the direct link: https://oceandata.sci.gsfc.nasa.gov/data_dashboard.

Question 18: Where in the PACE documentation can we find the tabular listing of spectral frequencies (bands) to avoid (invalid data)?
Answer 18: The diagram that Carina showed can be found here: https://pace.oceansciences.org/about_pace_data.htm.

Question 19: Is PACE data only available for oceans? What about land and inland water bodies?
Answer 19: PACE-OCI collects hyperspectral data over land as well. You can view the 10 land data products that are currently available here: https://pace.oceansciences.org/data_table.htm. PACE- OCI does also cover inland waterbodies if they are large enough – please see the answer to question number 6 above for more explanation.

Question 20: Are the SWIR bands available on EarthData?

Answer 20: Yes, all bands, including SWIR bands, are included in the PACE-OCI L1B data found in Earthdata.

Question 21: I tried to run earthdata library through Google Colab, tried to install the dependencies, yet it did not recognize its presence. Do you know what is happening?
Answer 21: On Colab, you need to first install the package itself with “%pip install earthaccess” in a cell by itself. After that, “import earthaccess” should work. For more assistance, you may want to post on the Earthdata Forum so that you can share more detail, including screenshots if necessary, about your issue:
https://forum.earthdata.nasa.gov.

Question 22: Is the data downloaded from the sites you presented during the three-day training course intended for reliable scientific study, or is it just used to give an overview of ocean quality?
Answer 22: PACE data is intended for scientific studies. However, we just launched earlier this year, some data is still pending validation. That means we have provisional and test data products for the moment. You can look at the status of the data products in the data products table. They will eventually get to the standard level when they are validated.
https://pace.oceansciences.org/data_table.htm

Question 23: Are there speed-ups to using EC2 in other AWS regions, e.g., Northern Virginia, compared to locally hosted notebooks or other clouds?
Answer 23: If we are only talking about performance improvements having to do with data transfer over a network, then using a compute instance (e.g. EC2) located in AWS region “us-west-2” is the only way to avoid transferring data over the **inter**net. Within AWS region us-west-2, all data transfer is on a fast, local network. Because the data is in S3 buckets within AWS us-west-2, any transfer to another AWS region, another cloud provider, or your local machine, goes over the **inter**net. This improvement matters when you are processing a lot of data and need each granule once. When you are still developing your analysis, and running code on the same granule many times, you’ll want a local copy of the data.

Question 24: When running the Jupyter Notebooks, I'm receiving an ImportError to import open_datatree from xarray. I have checked and I have xarray already installed and updated to version 2023.1.0, could you help me to solve it? Do you have an idea of what it could be?
Answer 24: If you follow the instructions we provided, you will get xarray version 2024.3.0 or better (as required by the environment.yml file). This version brought in the "open_datatree" backend, and future changes will provide a fully-documented file opener like "xarray.open_dataset".

Question 25: How can we get access to a .csv file of Rrs for a large scale area?
Answer 25: You can export Rrs data as a .csv using SeaDAS. Or you save the Rrs variable in an xarray dataset and use the Python library Pandas to export it as a .csv.

Question 26: Can we do atmospheric correction for OCM3 by replacing the PACE wavelengths in SeaDAS by resampling?
Answer 26: No. The l2gen code (the program that implements the atmospheric correction for PACE) does not support OCM3. The atmospheric correction uses sensor-specific look-up tables for the aerosol and Rayleigh components.