Data acquisition & processing overview

Jeremy Werdell, Jessie Turner
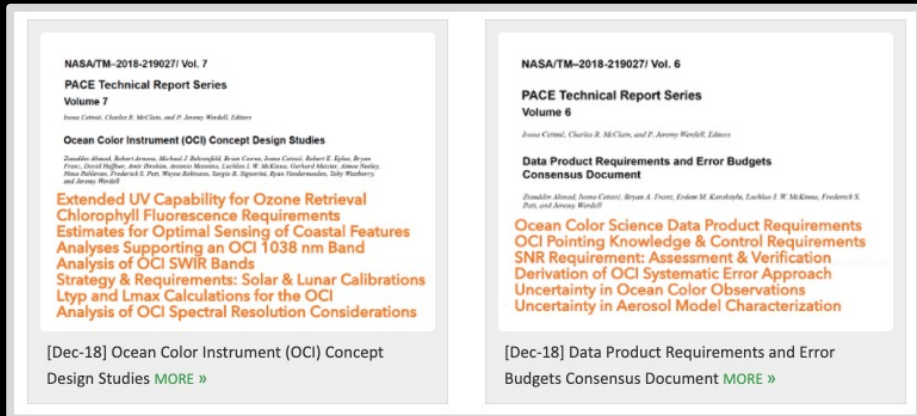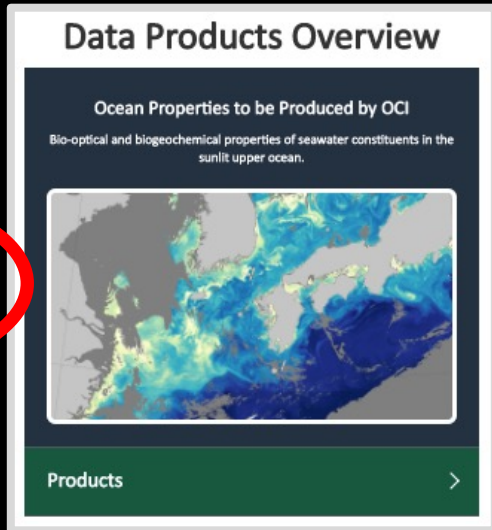
PACE class @ UMBC
1-5 Aug 2022

GODDARD
EARTH SCIENCES
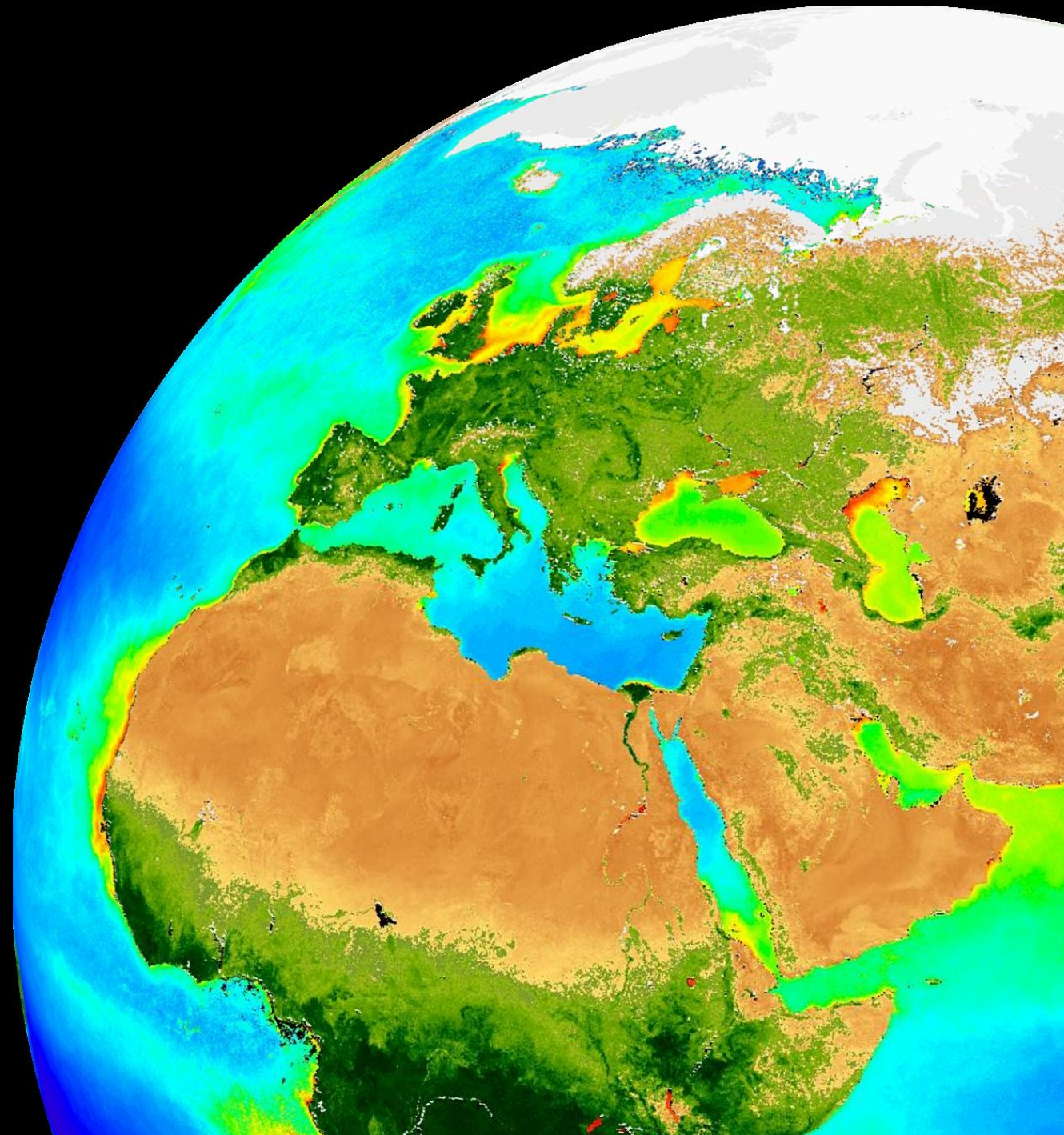
resources & useful info

data product descriptions + access to simulated data & characterizations

PACE technical memos & other documents

https://pace.gsfc.nasa.gov

@NASAOcean

data acquisition

data processing

OCI L1B simulated data (PyTOAST)

https://oceancolor.gsfc.nasa.gov

how to get the data (x3)
SIPS / SDS vs. DAACs

https://seadas.gsfc.nasa.gov

fussiness
EarthData forum

SeaDAS infrastructure

visualization GUI (graphical user interface) written in Java

wrapper scripts written in Python (modis_GEO.py, etc.)

source code (l2gen, l3bin, etc.) written in C & Fortran
same code used in production at GSFC

long series of data processing exercises
sent via email yesterday

uses ocean color as a case study, but
agnostic to data type overall

feel free to ask questions in the future

PACE / SeaDAS data processing & analysis lab

Following this lab, students will be able to:
- Differentiate between Levels 1, 2, and 3
- Process a PACE-like data file from Level 1 to Level 3 in both the SeaDAS GUI and from the command line
- Identify how key processing choices affect results for derived geophysical variables

0. Requirements.
This should be completed before the start of the lab.

SeaDAS version 8.2
python (see https://seadas.gsfc.nasa.gov/downloads/ for details)
OCSSW data processing for MODIS-Aqua (at a minimum)

1**. The command line environment**
All SeaDAS data processing functions are available through a GUI and a command line executable. Using the latter requires setting up the proper command line environment.

Open a terminal window. Locate your OCSSW software installation. By default, this should be a subdirectory within your SeaDAS installation. For example: /Applications/seadas/occsw .

Define an environmental variable called OCSSWROOT that points to your OCSSW software installation. The language for doing so depends on your shell. On Macs, the default is bash:

export OCSSWROOT=[path to SeaDAS]
e.g., export OCSSWROOT=/Applications/seadas/ocssw

Confirm the installation via the command "printenv OCSSWROOT". This should print the [path to SeaDAS] that you entered to your terminal screen.

Finish setting up the command line environment by sourcing the data processing environmental variable file. This file defines the locations of all files required by OCSSW. In bash:

source $OCSSWROOT/OCSSW_bash.env

Confirm the environment has been set via the command "which l2gen". This should print the full path to the file l2gen to your terminal screen (e.g., /Applications/seadas/ocssw/bin/l2gen).

2. Data acquisition.
Next, let's locate, download, and identify the file type of Level-1A and -2 satellite data. We will use these files over the course of this lab.

Acquire the following MODIS-Aqua files from the OceanColorWeb Level-1/2 browser. Or, use Data > File Search on Ocean Color Web. They encompass 3 successive days in the Gulf of Maine - 29 Sep 2018 to 1 Oct 2018:
- A2018272172500.L1A_LAC
- AQUA_MODIS.20180929T172500.L2.OC.nc
- A2018273180500.L1A_LAC
- AQUA_MODIS.20180930T180501.L2.OC.nc
- A2018274171000.L1A_LAC
- AQUA_MODIS.20181001T171001.L2.OC.nc

The above provides everything you need to know to find these files.

The files come in different file formats. What are they?

It's time to open SeaDAS.

3. Level-1 and -2 data extraction.
Full granules can be large. Images with lots of water can take a significant amount of time to process from Level-1 to -2. Let's extract our region of interest, the Gulf of Maine, defined here as **40° to 46°N and -75° to -60°W**. You will need to open the full files in SeaDAS first.

Use the OCSSW extractor (GUI) to create subsetted Level-1A and -2 files. Do the Level-2 files first, using the above latitude and longitude boundaries to define the extract. You'll notice that the start and end pixel and line number fields get populated after you've added the latitude and longitude boundaries. Write these down. You will need them to run the Level-1A extracts.

Why can you use latitude and longitude to define extractions for Level-2, but not for Level-1A?

Once you've created the six extracts, you can close the full files for the remainder of this lab.

4. Geolocation

MODIS-Aqua Level-1A files need to be geolocated before Level-2 files can be generated. Note, this is not true for all ocean color satellites – SeaWiFS, for example, does not require this step. This doesn't mean that SeaWiFS files aren't geolocated, only that this process is consolidated elsewhere in the SeaWiFS data processing architecture.

Use the OCSSW MODIS GEO processor to geolocate 1 or 2 of the 3 Level-1A extracts. What are the pixel numbers for the file(s) you chose to geolocate?

File	spixl	epixl	sline	eline

5. Command line processing

All SeaDAS data processing GUIs list the name of their corresponding command line executable on the top bar of the GUI window.

What is the name of the executable for the MODIS GEO processor?

Where is this script located?

What language is it written in?

Run the MODIS GEO script from your terminal window without any inputs. Help and usage information will scroll in your terminal window. All SeaDAS command line executables do this when you run the command without input variables. If you've set up the terminal environment correctly, you do not need to provide the full paths to the executables to run them.

Use the command line to geolocate the remaining 1 or 2 L1A extracts.

6. Level-1B processing

MODIS-Aqua Level-1A files need to be processed to Level-1B before Level-2 files can be generated. Again, not all ocean color satellites require this independent processing step.

Use the OCSSW MODIS L1B processor to generate Level-1B files. Consider using both the GUI and the command line. Do not provide any inputs other than the Level-1A and GEO file names.

How many files were created per run?

How do these files differ? (Hint: open them in SeaDAS and have a look.)

We will not be using the HKM or QKM files the remainder of this lab.

7. Standard Level-2 processing

The "l2gen" executable in SeaDAS provides Level-2 processing capabilities for all ocean color instruments under the auspices of the GSFC Ocean Biology Processing Group. When you run l2gen without any inputs (other than the input file names and ancillary data file names), you will exactly recreate the Level-2 files distributed on the OceanColorWeb.

Open the OCSSW l2gen GUI.

Select the extracted **AQUA_MODIS.20180929T172500*** L1B and GEO files as ifile and geofile. If you have a Level-1B file highlighted in the File Manager window, these fields will self-populate.

Rename the ofile – otherwise, you may overwrite the Level-2 extract you previously created.

Acquire the ancillary data associated with this data file (hint: there is a GUI button).

Run l2gen.

Open the Level-2 extract you created previously and the Level-2 file that you just produced.

View the chlor_a product for each. How do they compare?

Repeat for the other 2 Level-1B extracts. Continue until you are convinced that the Level-2 files you generate are identical to the Level-2 extracts you created previously. This is also a good way to verify that you ran the extractor properly.

8. l2gen parameterizations and command line usage

Without exaggeration, l2gen can produce hundreds of geophysical data products and be configured in thousands of different combinations.

Open the l2gen GUI and explore the tabs.

Note several highlights: The "Processing Options" and "Ancillary Inputs" tabs should bring back memories of the atmospheric correction lecture. The "IOP Options" tab should bring back memories of the ocean color data products lecture.

To run l2gen from the command line:
l2gen par=[parameter file]
where the parameter file includes a list of keyword=value pairs (one per line).

Note that you can also run l2gen from the command lines as:
l2gen header1=keyword1 header2=keyword2 header3=keyword3 … (and so on)

Notes on the parameter files:

- For a complete GUI-based list of these keyword=value pairs, select the "Show Defaults" button under the "Main" tab. This will list all of the keyword=value pairs with the default values used in standard processing. For a complete terminal-based list, run l2gen from the command line without any inputs.

- l2gen does not require providing a keyword=value pair to use a default configuration, which is why they are commented out in this "Show Defaults" display (the "#" at the beginning of the line tells l2gen to ignore that line). To use an alternate l2gen setting in the GUI, you can either make the change in one of the tabs or you can edit the keyword values within the "Main" tab display by removing the "#" and changing the value.

- A parameter file only needs to include those values that diverge from their defaults. This is why we didn't need to change anything in the previous exercise to exactly replicate the standard Level-2 files distributed online. The "Main Tab" always shows those keyword=value pairs that diverge from the default configuration.

- You can download a parameter file developed within the GUI using the "Save Parameters" button. Likewise, you can also load one using "Load Parameters".

9. Example non-standard l2gen runs
Let's explore how a geophysical data product changes when we diverge from the default configuration of l2gen. Use the **AQUA_MODIS.20180929T172500*** L1B and GEO files for this exercise.

We will rerun l2gen 3 times. Make sure to **rename the ofile for each instance** or you will overwrite the Level-2 file(s) that you previously generated.

Generate a Level-2 file using climatological ancillary data. This is exactly the same as the exercise in Section 7, but without hitting the "Get Ancillary" button.

Quantify the differences in chlor_a between your standard run of l2gen (from Section 7) and this run. Do this by using: (a) SeaDAS to generate a difference map (perhaps using Math Band), (b) one or more of SeaDAS's plotting or graphing functions; or (c) all of the above.

Revert to using the correct ancillary data for the remainder of this exercise.

Generate a Level-2 file with the vicarious gains all set to 1.0 .

Quantify the differences in Rrs_443 between your standard run of l2gen and this run.

Revert to using the correct vicarious gains for the remainder of this exercise.

Generate a Level-2 file using SWIR bands instead of NIR bands in the atmospheric correction (set aer_wave_short=1240 and aer_wave_long=2130)
.
Quantify the differences in aot_869 and angstrom between your standard run and this run.

Revert to using the Level-2 files you created in Section 7 for the remainder of this lab.

10. Spatially compositing data
Someday, you may find yourself wanting or needing to re-grid (re-project) your Level-2 data onto a standard map grid. Here are examples of two common grids:
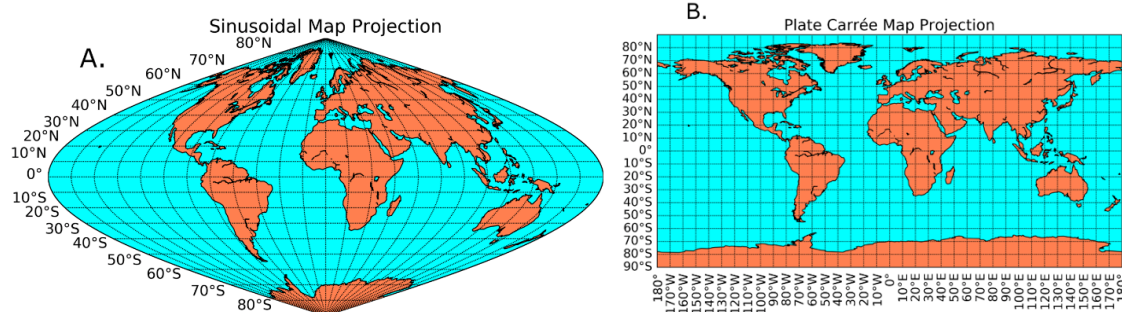
Note the differences in bin size as a function of latitude. Examine, e.g., Greenland.

One quick way to visualize the transformation from the satellite swath coordinate system to a standard map grid is to use OCCSSW l2mapgen. This utility reprojects a Level-2 file onto a Plate Carree grid (panel B above) and generates a PNG or PNM image.

Use l2mapgen to generate a PNG image of **AQUA_MODIS.20180929T172500** chlor_a. You will need to assign the "prod", the "outmode", and select "apply_pal".

The NASA processing cadence for producing Level-3 maps is: (1) generate a Level-3 bin file, which is always in a sinusoidal map projection (panel A above) that retains equal distance in latitude and longitude space from north to south; then, (2) generate the Level-3 map file in the projection of your choice (standard is Plate Carree). SeaDAS cannot open Level-3 bin files.

Let's next explore the effects of spatially re-gridding Level-2 data on geophysical values.

Use OCSSW l2bin to generate a series of chlor_a Level-3 bin files for **AQUA_MODIS.20180929T172500**. Always set "l3bprod" to chlor_a. Always set "prodtype" to regional. Run 4 times with "resolution" set to 1, 2, 4, and 9 km . Rename the ofile each time to avoid overwriting existing files.

Next, use OCSSW l3mapgen to convert these Level-3 bin files into Level-3 map files. Manually type "chlor_a" as the "product", leave the projection at "platecarree", and set the resolution to 1km, 2km, 4km, and 9km based on the resolution of the Level-3 bin file you are mapping.

This is a good time to revisit the help/usage information provided when running an OCSSW program without any input values. It's recommended that you do this for any program you plan to use to best understand the operation and options of that program. Type "l3mapgen" in a terminal window. Focus on usage for the "product" input. Note that you can modify an input product to report alternate statistics, such as standard deviations and the number of observations that filled each Level-3 grid-bin. Such information on capabilities is not always intuitive within a GUI environment. Best practices involve studying command line options.

Open the original Level-2 and the 4 Level-3 mapped chlor_a products.

To explore how geophysical values change with re-gridding, let's focus on a specific high chlorophyll area to the east of Cape Cod:
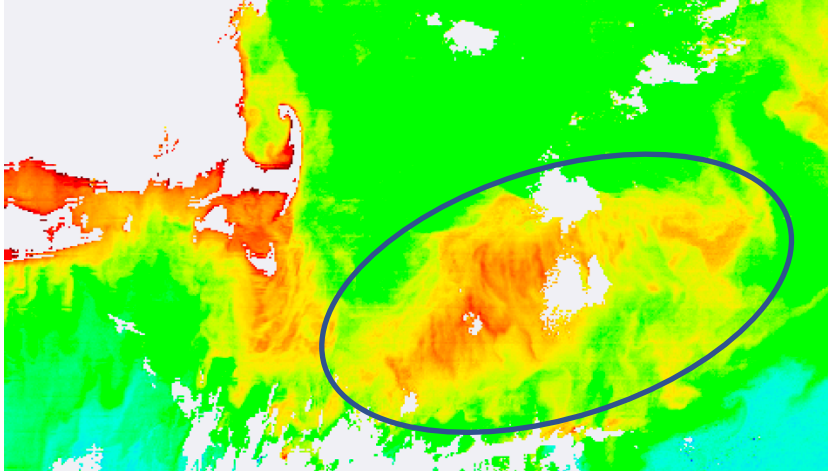
Quantify if/how this feature distorts relative to the Level-2 data when re-gridded at different spatial resolutions. You can do this through analysis of samples sizes, histograms or cumulative distributions, metrics of central tendency, or by other means of your choice. SeaDAS provides many options for such image analysis.

What happens to the fidelity of data off the coast of Cape Cod with increasing pixel size?

Now imagine that you're studying a small or filamentous feature, say a thin bloom of Trichodesmium offshore of Australia or Noctiluca in the Gulf of Oman. How do you think your choice in spatial representation of the feature will influence your results?

10. Temporally compositing data
Someday, you may find yourself wanting or needing to temporally composite your data into a longer term average (say, weekly, monthly or seasonally).

Use OCSSW l2bin to create regional 2-km chlor_a Level-3 bin files for the other two Level-2 files you generated in Section 7 (**AQUA_MODIS.20180930T180501** and **AQUA_MODIS.20181001T171001**). When done, you should have 3 Level-3 2-km chlor_a bin files (one for each day).

We now want to create a single, temporal composite of these 3 files. We'll use OCSSW l3bin to do this, which will involve a different process than that for l2bin.

First, create a plain text file that lists the full name of all three Level-3 bin files to be composited, with one file name per line.

Select this text file as the ifile in the l3bin GUI. Next, select chlor_a as the prod and change the latitude and longitude boundaries to those we've been using in this lab (otherwise, you will create a global map … worth trying, though!).

After your 3-day Level-3 bin file has been created, use l3mapgen again to create a 2-km Level-3 map, as you did in Section 10.

Open the Level-3 2-km map from the first day in the series and the 3-day Level-3 2-km map that you just created.

Quantify how the spatial coverage changed.

What are the advantages of using temporally composited data?

Quantify how the chlor_a retrievals changed.

What are the disadvantages of using temporally composited data?

Thinking back to the filamentous features the "Why aren't all satellites hyperspectral with meter-scale resolution?" lecture … how will these and other transient or fragile features be treated in temporal composites?

Currently, a major goal of ocean color remote sensing is the identification of phytoplankton community structure. How do you think consideration of Level-2 versus spatially re-gridded Level-3 data versus temporally composited Level-3 data influences the interpretation of remotely-sensed community structure?

12. Level-3 data processing
Perhaps not widely known, OCSSW software within SeaDAS allows generation of geophysical data products from Level-3 Rrs bin files.

To explore this, we need to generate a Level-3 bin file that includes Rrs. Run l2bin on **AQUA_MODIS.20180929T172500**, but this time do not set assign a "l3bprod" in the GUI. Leaving this field blank will bin all products in the Level-2 file (if you ran l2bin from the command line without any inputs you should already know this). Do set prodtype to "regional" and resolution to "2".

Open the OCSSW l3gen GUI. Assign the Level-3 bin file you just created as ifile.

Note that it looks and feels like the l2gen GUI, but without the ability to specify ancillary files. Why do you suppose this is?

Let's produce IOP products this time. Add adg_443_giop, aph_443_giop, and bbp_443_giop to the list of l2prods, either using the GUI buttons or by manually changing the field in the "Main" tab. Feel free to remove the rest of the l2prods. As always, rename ofile as necessary to avoid overwriting existing files.

Run l3gen. This will create a new Level-3 bin file with IOP products derived from Level-3 Rrs.

Run l3mapgen to convert this Level-3 bin file into a Level-3 map. Leaving the "product" field blank will create a map file containing all 3 IOP data products (again, you know this already if you ran l3mapgen from the command line to see its usage).

Do you think that IOP data products produced at Level-2 and Level-3 will be the same? What information contained at Level-1/-2 is lost in Level-3 data? Is this important for IOP retrievals?

This is a good place to explore how different configurations of an IOP model influence the derived adg, aph, and bbp. Repeat this exercise, but use the "IOP Options" tab in the l3gen GUI to reparametrize GIOP (similar to Lab 10 on Friday).

End of laboratory. Nice work.

PACE Course August 4, 2022

Where to find simulated PACE data

https://oceancolor.gsfc.nasa.gov/

Data tab

Direct Data Access

Download one Level 2 file and one Level 3 Mapped file

Optional: Level 1B > PACE_OCI

Optional: Level 1C > PACE_SPEXONE

Level 2 > PACE_OCI, choose year 2019. Find one with a large file size (more data) with "OC" (Ocean Color) in the filename

Level 3 Mapped > PACE_OCI, choose any day and product you like.

Plotting simulated PACE data in Panoply

Optional:

Open Level 1B file in Panoply

Navigate to "observation data"

What are the groups of bands? How many bands are in each group?

Why can't you plot a map of colored Lt ("rhot") with longitude and latitude as your x and y axes?

Visualizing Level 2 and Level 3:

Open Level 2 file in Panoply.

Find and click on Rrs and click "Create Plot."

Window > Arrays

How many bands are available?

Click the small up and down arrows next to the band number to advance through the bands. Or, can enter a number in the box and hit Enter. New plots will be created in the plot window.

Open level 3 mapped file in Panoply.

Click on your product and click "Create Plot."

Window > Scale

Play with the max value, hit Enter to update the plot.

Units: Scalar or Log10 depending on your product of interest.

Is this different from what you could get from SeaWiFS or MODIS?

PACE / SeaDAS data processing & analysis lab answer key

2. Data acquisition

Level-1A download as bunzip2 files, which extract to an hdf4 file
Level-2 files are netCDF / hdf5

3. Level-1 and -2 data extraction.

Level-2 files have been geolocated and, thus, include latitude and longitude information.

Level-1A files have not been geolocated and, thus, latitudes and longitudes are not yet assigned to each pixel.

File	spixl	epixl	sline	eline
A2018272172500.L1A_LAC	270	1278	1	774
A2018273180500.L1A_LAC	1	675	1109	1890
A2018274171000.L1A_LAC	450	1345	976	1924

5. Command line processing I

modis_GEO.py
$OCSSWROOT/scripts/
Python

6. Level-1B processing

3
LAC file has the 1 km data products
HKM files have the 500-m bands (469, 555, 1240, 1640, 2130 nm)
QKM files have the 250-m bands (645, 859 nm)

7. Standard Level-2 processing

There may be an offset in the geographic boundaries of the imagery, but the geophysical data and its spatial patterns should be identical.

8. l2gen parameterizations and command line usage

List of three processing options of interest.

9. Example non-standard l2gen runs
Chl differences from change in ancillary data:

chl diff *(%)*

−20   −10   0   10   20
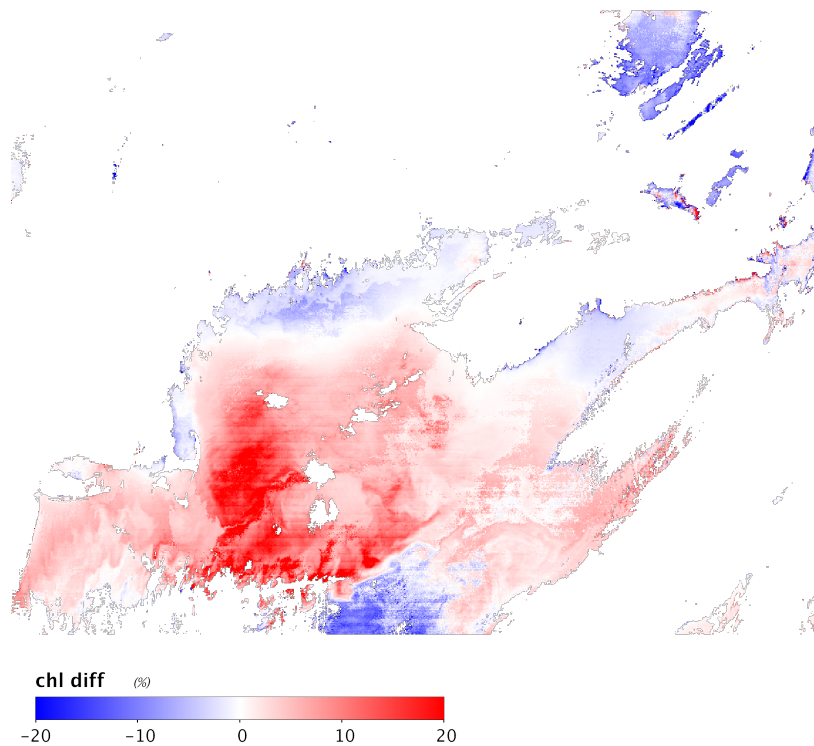
Rrs_443 differences from change in vicarious calibration:

rrs_diff *(%)*

-40   -20   0   20   40
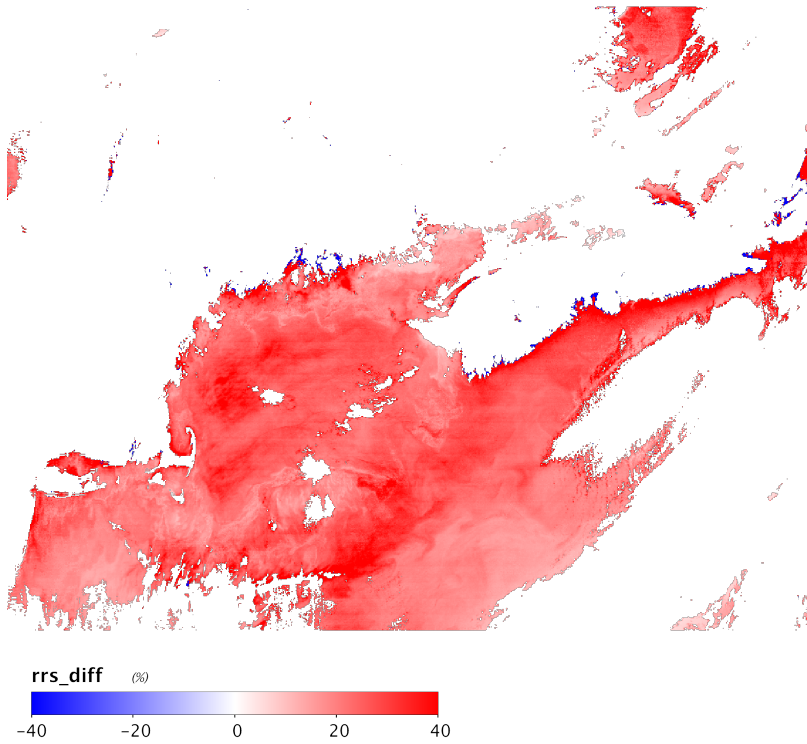
AOT_869 and angstrom differences from change in A/C bands:

aot_diff *(%)*

-20   -10   0   10   20

ang_diff *(%)*

−100   −50   0   50   100
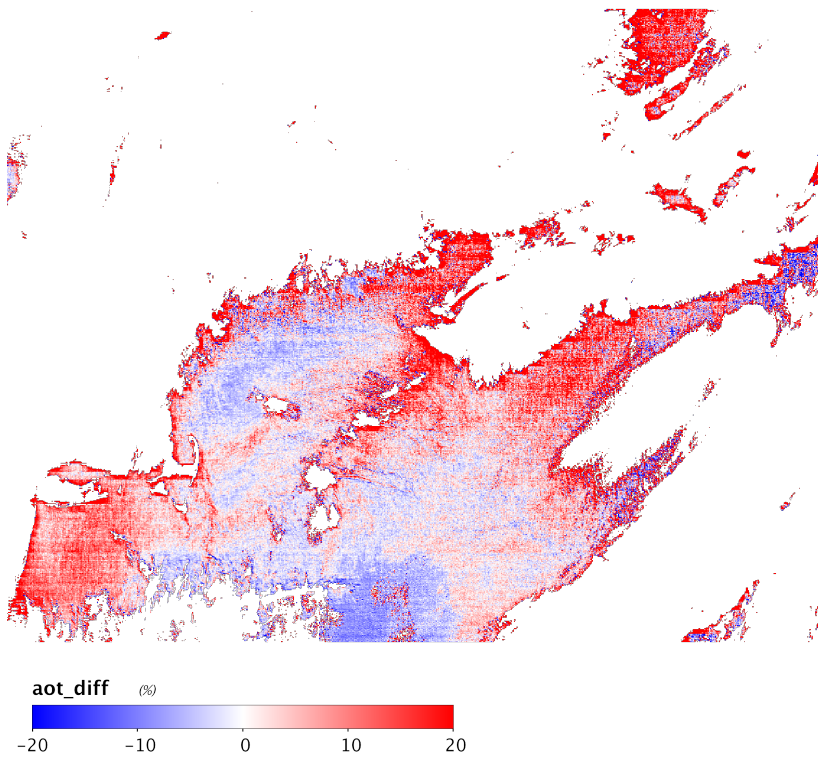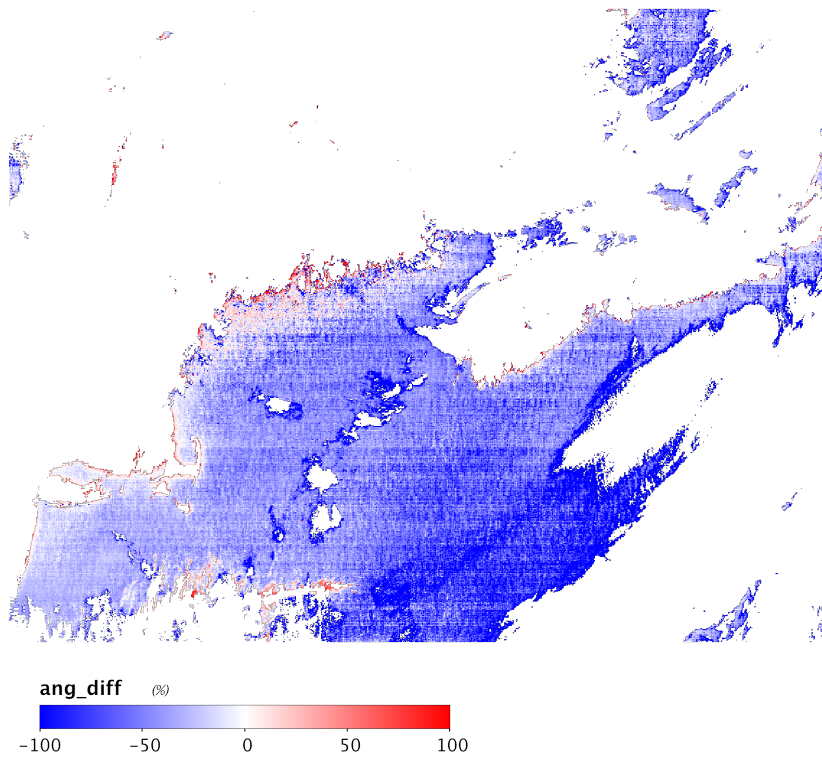
10. Spatially compositing data

l2mapgen chlor_a:
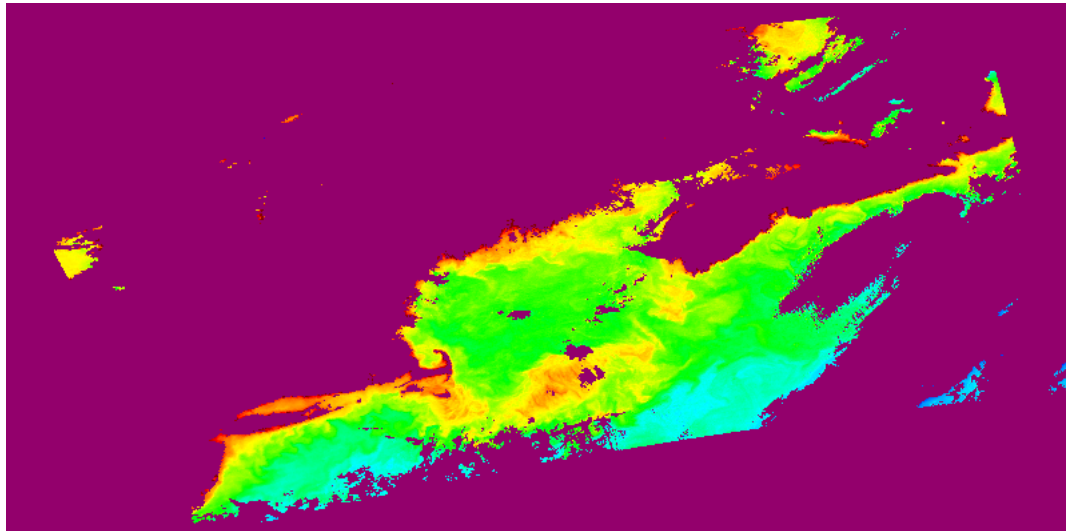
Level-2 vs. Level-3 maps at 1, 2, 4, and 9 km:
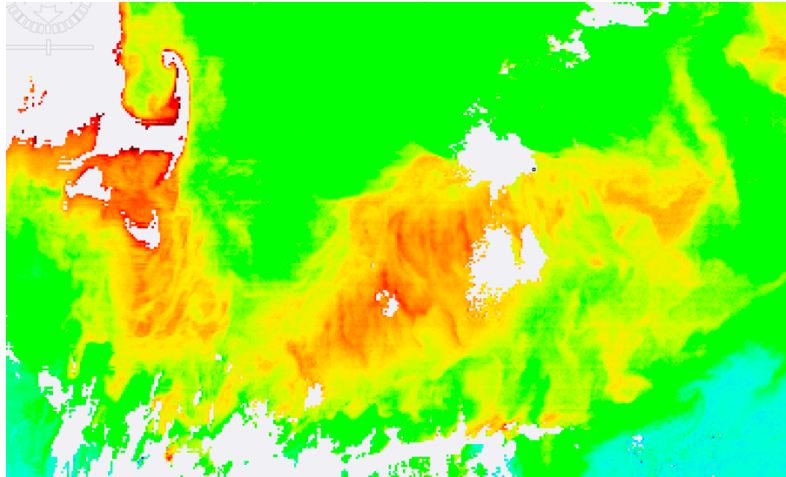
Level-2
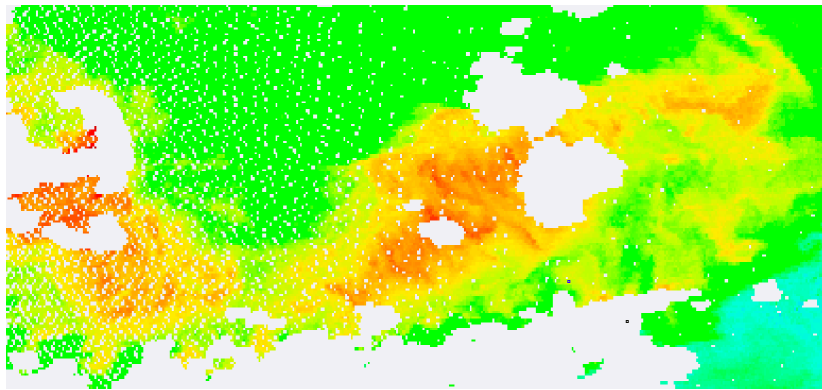
Level-3, 1 km
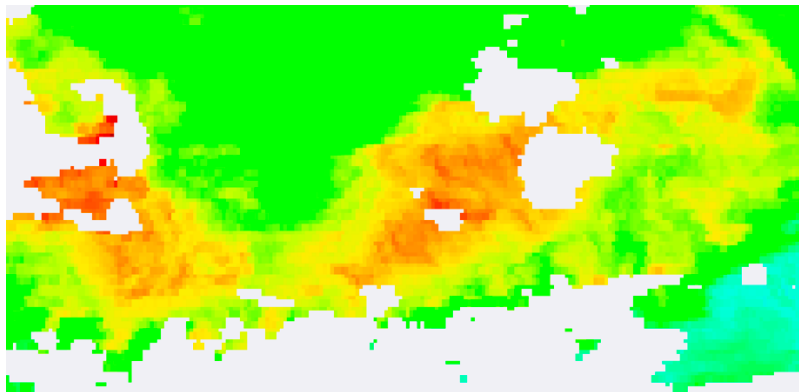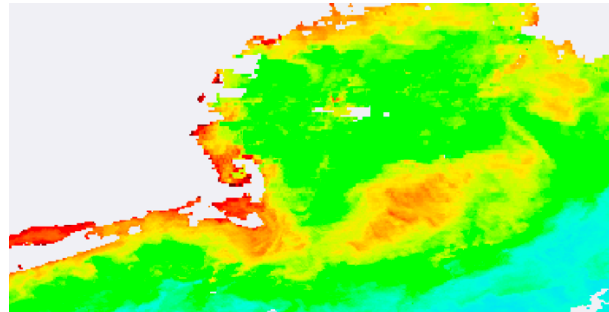
Level-3, 2 km
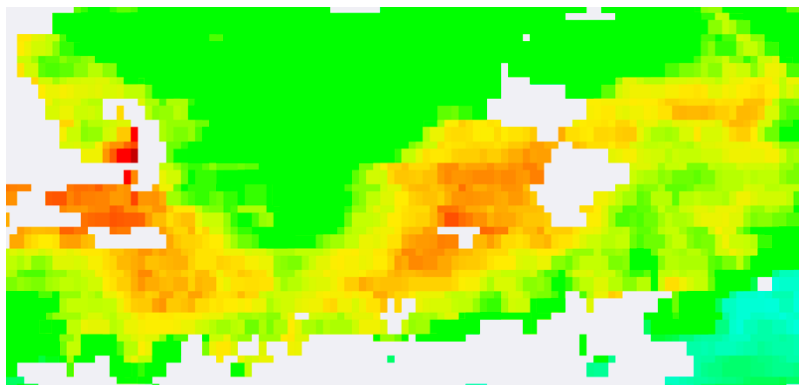
Level-3, 4 km
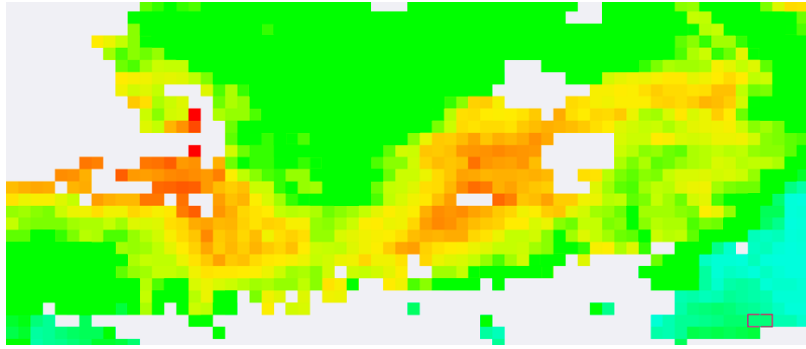
Level-3, 9 km

11. Temporally compositing data
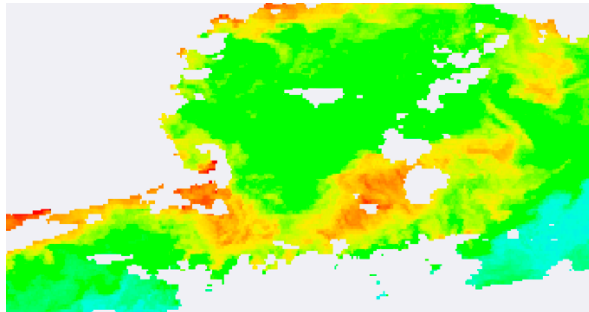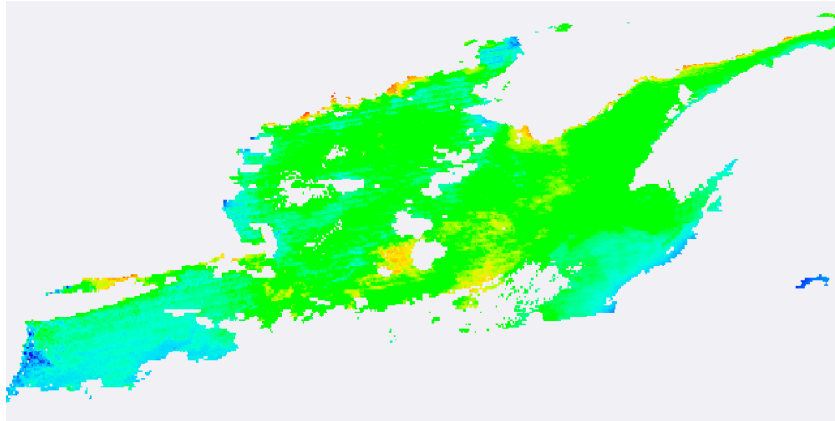
Single day: 3-day:
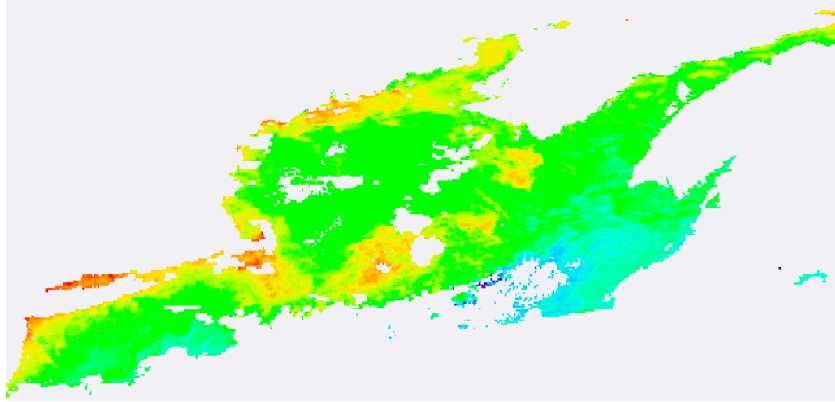

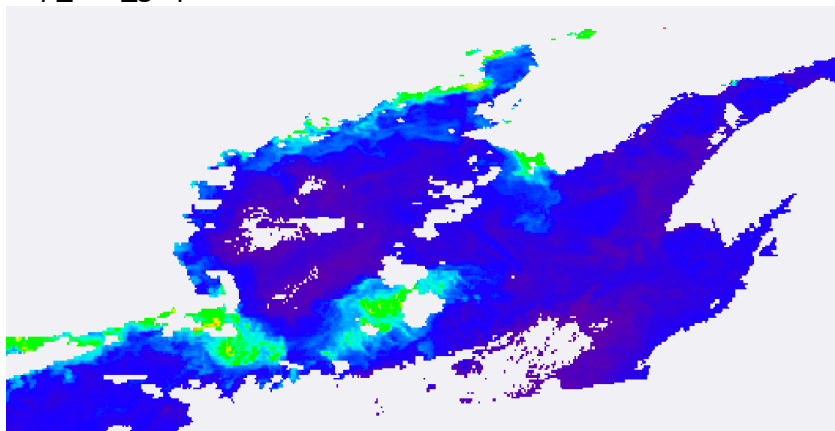## 12. Level-3 data processing

adg_443_giop

aph_443_giop

bbp_443_giop

Level-3 data products do not contain information on geometry or time at the moment of measurement – all useful information for deriving some geophysical data products.