

HANDS-ON POST-CALIBRATION OF IN VIVO FLUORESCENCE USING OPEN ACCESS DATA

A GUIDED JOURNEY FROM FLUORESCENCE TO PHYTOPLANKTON BIOMASS

By Pierre Marrec, Amanda Herbst, Stace E. Beaulieu, and Susanne Menden-Deuer

<https://doi.org/10.5670/oceanog.2025.314>

Lab Instructions for Students

BACKGROUND

We are in an era of big data, where high-resolution sensors gather and share information at an unprecedented rate, particularly in the field of oceanography. Oceanographic data come from a wide variety of sources, including sensors on remote observing platforms, including satellites, and sensors deployed during research cruises or from samples analyzed in the laboratory. To make these data FAIR—findable, accessible, interoperable, and reusable (Wilkinson et al., 2016)—they need to be processed in ways that make them accessible to a wide range of users and that adhere to strict scientific standards. In this lab you will be introduced to data handling as well as sharing practices; open-access data are now a mandate for those who receive funding from many government agencies. Many publishers also require transparent documentation of data and analysis procedures. Lastly, it is immensely useful to the scientific community to have open-access data for meta-analyses and for investigations that the generator might not have thought of.

Oceanic physical parameters (temperature, salinity, depth) and biogeochemical parameters (dissolved oxygen, bio-optical properties, nitrate, and carbonate system chemistry components) are routinely measured by various sensors. The data generated by these sensors are substantial and often made available as open access through dedicated data portals. However, before these data can be effectively utilized for research purposes, additional steps are often necessary. The raw values obtained by sensors often require quality control, calibration, and validation to ensure accuracy, as using them directly from data portals can lead to erroneous absolute values and misinterpretations. Therefore, developing and implementing robust data processing procedures is crucial to produce science-ready data from oceanographic sensors.

The goals of this lab are to:

1. Introduce you to ocean observing data and the principles of FAIR data management.
2. Execute the necessary steps for processing data from oceanographic sensors before using them to analyze relevant oceanographic processes.
3. Enable you to interpret oceanographic data effectively, properly and reliably.

DATA SOURCE

You will be working with high-frequency in vivo chlorophyll-*a* (Chl-*a*) fluorescence data collected by fluorometers installed on R/V *Endeavor*. These fluorometers continuously record Chl-*a* fluorescence in the surface waters along the ship's track. At the same time, discrete Chl-*a* concentration data are collected during these cruises, extracted from phytoplankton cells that were filtered from underway water samples.

The specific data you'll be analyzing are openly accessible in a .csv (comma separated values) file and originate from six cruises conducted within the Northeast US Shelf Long-Term Ecological Research (NES-LTER) program (<https://nes-lter.who.edu/>).

THE NORTHEAST US SHELF

The Northeast US Shelf (NES) stretches from Cape Hatteras in North Carolina to the Gulf of Maine. It is a rich and important marine ecosystem, both ecologically and economically (Hoagland et al., 2005). This area is known for being highly productive and for its strong seasonal changes (Fontaine et al., 2024), with water temperatures differing by more than 15°C between winter and summer.

The NES has a mix of very different water types. Cold, less salty water from the Arctic flows south along the coast, while farther offshore, warmer, saltier water from the Gulf Stream moves north (Chapman and Beardsley, 1989). These two types of water meet along the edge of the continental shelf, creating a sharp boundary called the shelf-break front.

In winter and early spring, the shelf waters are full of nutrients, thanks to mixing, river runoffs, and coastal currents. These nutrients help large phytoplankton (microscopic algae) grow to high Chl-*a* concentrations, forming what we call "blooms" (Marrec et al., 2021). In summer, the surface waters become layered and more stable, which limits the supply of nutrients. As a result, smaller phytoplankton at lower Chl-*a* concentrations are common during this season (Marrec et al., 2021).

These seasonal changes in phytoplankton affect the entire food web—from tiny animals like zooplankton to fish and larger predators—because they influence the availability of food.

The NES is also changing quickly due to climate change. Surface waters here are warming about three times faster than the global ocean average (Saba et al., 2016). This warming is causing many marine species to shift northward (Friedland et al., 2023) and is changing when and where different organisms appear, from the smallest plankton (Hunter-Cevera et al., 2020) to large fish and marine mammals (Lucey and Nye, 2010).

APPROACH

The laboratory session will be based on a brief introduction of how to measure phytoplankton biomass using a fluorometer during the lecture prior to the lab and is divided in two sections:

1. Accessing underway fluorescence and discrete Chl-*a* data from oceanographic cruises
2. Post-calibrating the fluorescence data and interpreting the results

All the fundamental steps of this laboratory session can be completed using basic computer resources (e.g., Open Office Calc, Microsoft Excel); you do not require specific programming skills. However, it also offers you an opportunity to enhance your programming skills (e.g., using R, MATLAB, or Python) by automating and streamlining these processes. The answers to the questions posed below will be part of a group discussion and report out at the end.

To facilitate the activities, open-access templates (OpenDocument Spreadsheet, .ods) are included in the online supplementary materials. Additionally, .ods files containing the expected results for each activity are provided to ensure all tasks can be completed, even when facing challenges with specific steps.

Activity

SECTION 1. ACCESSING AND EXPLORING SENSOR-BASED FLUORESCENCE AND DISCRETE CHL-A DATA (1 hour of homework, 1.5 hours in class)

Part 1. Sensor-Based Fluorescence Chl-a Data

GOAL. Access and work with authentic raw underway fluorescence data, followed by a preliminary interpretation of these data.

1. Download the underway csv data files for the cruises EN644, EN649, EN655, EN657, EN661, and EN668 from the supplemental materials.
2. This step is optional, as it can be challenging to navigate through these data. If instead you wish to find the original source data, all these data are publicly available at the UNOLS R2R repository (<https://www.rvdata.us/>). Try to find the underway data from these cruises. For this first “Browse Vessels”, select R/V *Endeavor*, and find your assigned cruise. Navigate through the underway datasets available and click on the DOI of “TSG – SeaBird SBE-21.” Download the raw data and try to have a look at them (first extract compressed files, then navigate through the .raw files by opening them with a text editor).
3. For each cruise, copy and paste these columns into the “Data Exploration – underway fluorescence” sheet of the fluoro-activity-template.ods spreadsheet:
 - date
 - gps_furono_latitude (in latitude, °N)
 - gps_furono_longitude (in longitude, °E)
 - tsg1_fluorometer1_v0 (in raw_voltage_wetstar, raw fluorescence in V)
 - tsg1_fluorometer2_v1 (in raw_voltage_ecofl, raw fluorescence in V)
 - tsg1_fluorescence_wetstar (in raw_chla_wetstar, manufacturer-calibrated fluorescence in mg m^{-3})
 - tsg1_fluorescence_ecofl (in raw_chla_ecofl, manufacturer-calibrated fluorescence in mg m^{-3})
4. Convert the date into readable date formats. Indeed, the date is in ISO8601 format and cannot be read by spreadsheet editor software as a date as it is. For this select the date column, go to “find and select>replace,” and replace “+00:00” by “ ” (no character) and replace all. Time is expressed in UTC, to get access to local time $\text{LT} = \text{UTC} - 5\text{h}$ (in winter) and $\text{LT} = \text{UTC} - 4\text{h}$ (in summer).
5. Delete all the rows where fluorescence values are NaN (“not a number”). Most of the NaN values are at the beginning and at the end of, but use “sort & filter” to find NaNs in these columns.

6. Compare the different fluorometer values by making two figures:
 - a. Raw fluorescence (raw_voltage, in V) vs. time for both fluorometers on the same plot
 - b. Manufacturer calibrated values (raw_chla, in mg m^{-3}) vs. time for both fluorometers on the same plot.
7. Calculate the difference between WetStar and ECO-FL manufacturer-calibrated fluorescence values in the wetstar_ecofl_diff column.
8. Plot the different values thus obtained vs. time.

Council 1

We will break in class and have everyone describe their observations from their assigned cruise (same for all other Councils). For now, just think about these questions and note your thoughts.

- What are key differences between the Chl-a fluorescence from the UNOLS R2R repository vs the one available in the supplemental materials csv files?
- What do you observe in terms of general patterns of raw fluorescence (V) and manufacturer-calibrated fluorescence (mg m^{-3})?
- What do you observe when looking at the fluorescence difference between both sensors?

Part 2. Discrete Data of Extracted Chl-a

GOAL. Access and analyze authentic discrete chlorophyll-a (Chl-a) data, followed by a preliminary interpretation. Gain familiarity with the dataset required for Section 2 of the lab activity.

1. Download the discrete underway Chl-a data (chl-transect-underway-discrete.csv) from the Environmental Data Initiative EDI data portal: Menden-Deuer, S., P. Marrec, and A. Herbst. 2022. Underway discrete chlorophyll and post-calibrated underway fluorometer data during NES-LTER Transect cruises, ongoing since 2019 ver 1. Environmental Data Initiative. <https://doi.org/10.6073/pasta/16c8e5937a860c882b524fda73408baf> (Accessed 2024-08-16).

2. For each cruise, copy and paste the cruise Id, date/time, latitude (°N), longitude (°E), replicate, chl (discrete Chl-a concentration, mg m^{-3}) and the iode_quality_flag columns into the “Data Exploration – discrete chl-a” sheet of the fluoro-activity-template.ods spreadsheet.
3. Convert the date into readable date format. For this select the date column, go to find and select>replace, and replace “T” by “ ” (space) and replace all. Do the same with “Z.”
4. Delete all the discrete Chl-a concentrations with a iode_quality_flag = 3 or 9. IODE quality flags are used to assess the quality of the discrete Chl-a data (1 = good, 3 = questionable, 9 = missing). More information about the quality check procedure can be found online in the detailed meta-data of the EDI data package.
5. For each sampling point (replicate a, b, and c sampled at the same time), calculate the mean, standard deviation, and coefficient of variation ($\%CV = \text{StdDev}/\text{Mean} * 100$) values for each triplicate discrete Chl-a values. In some instances, none or only 1 or 2 values on the 3 triplicate values will be available.

Council 2

- What are the min/max discrete Chl-a values for your cruise?
- What about the %CV, which stands for coefficient of variation? What does this value mean?
- How do you explain the variability amongst triplicates?

SECTION 2. USING DISCRETE CHL-A TO POST-CALIBRATE SENSOR-BASED FLUORESCENCE (1 hour of homework, 1.5 hours in class)

Part 1. Plotting Sensor-Based Chl-a Fluorescence vs. Extracted Chl-a Concentrations

GOALS: Establish connections between sensor-based underway chlorophyll-a (Chl-a) fluorescence and discrete Chl-a data. Introduce methods required for post-calibrating sensor-based Chl-a fluorescence data.

1. For a given cruise, copy and paste the mean and the standard deviation of each discrete Chl-a sample into the “Post-calibration – linear regression” sheet of the fluoro-activity-template.ods spreadsheet.
2. For each discrete Chl-a point, find the corresponding underway manufacturer-calibrated fluorescence value of each fluorometer in the underway file (sheet “Data Exploration – underway fluorescence”) at the same time (same minute) as the sampling time of discrete Chl-a. Make sure the date/time are expressed in the same format (UTC or local time).
3. Then plot for each sensor the manufacturer-calibrated fluorescence (x-axis) vs. the average discrete Chl-a (y-axis), with the standard deviation of each discrete Chl-a sample represented by error bars (y-axis).
4. Perform a linear regression. As an optional step, instead of using the average values, assign the matching fluorescence to each triplicate value (with IODE quality flag =1). The linear regression model I will then consider the error associated with the discrete Chl-a measurements. This introduction to linear regression can be expanded to consider different regression models (e.g., model II, by considering the uncertainty relative to the underway fluorescence when considering different matching intervals).
5. Identify the slope and the intercept of the linear regressions and get the R^2 (coefficient of determination) values.

Council 3

- What do the slope and the intercept represent?
- What does the R^2 represent?
- Can you think of reasons for the behavior of some values?
- What could be the sources of error of discrete Chl-a?
- Which fluorometer is the most reliable (for a given cruise)? Why?

Part 2. Post-Calibration to Estimate Chl-a Concentration from In Vivo Fluorescence

GOALS: Post-calibrate the underway fluorescence data by applying the regression relationships established in Section 2, Part 1, between the raw fluorescence measurements and the discrete Chl-a concentration data. Compare the raw fluorescence values with the post-calibrated data collected during the three summer cruises and interpret the resulting figures.

1. For a given cruise, copy and paste the date/time, latitude, longitude, raw fluorescence values from the selected fluorometer (manufacturer-calibrated fluorescence values expressed in mg m^{-3}) into the “Post-calibration” sheet of the fluoro-activity-template.ods spreadsheet.
2. Calculate the post-calibrated fluorescence values using the slope and intercept (i.e., coefficients) obtained from the linear regression in the “post-calibrated_fluorescence” column.
3. Compare the raw manufacturer-calibrated fluorescence values and the post-calibrated values by plotting both values (1) vs. time and (2) vs. latitude.
4. Now gather the underway post-calibrated fluorescence data obtained during the three summer cruises (EN644, EN655, EN668). Since each student worked on a different cruise,

data should be shared among the groups, so that everyone has access to all calibrated data from each cruise. Post-calibrated data for all cruise can be find in chl-transect-underway-post-cal.csv in the supplementary materials if needed.

5. Plot on a single figure raw manufacturer-calibrated fluorescence values together vs. latitude for all three summer NES-LTER cruises.
6. Make a similar figure from the three summer (EN644, EN655, EN668) NES-LTER cruises post-calibrated fluorescence values.

Council 4

- *What do you observe when comparing manufacturer-calibrated fluorescence (in mg m^{-3}) with the post-calibrated fluorescence for your cruise?*
- *What do you observe when comparing the raw manufacturer-calibrated fluorescence values from the three summer NES-LTER cruises?*
- *What differs when comparing the post-calibrated fluorescence values together from the three summer NES-LTER cruises?*
- *Why is the post-calibration of the fluorescence sensor essential?*

References

- Chapman, D.C., and R.C. Beardsley. 1989. On the origin of shelf water in the Middle Atlantic Bight. *Journal of Physical Oceanography* 19:384–391, [https://doi.org/10.1175/1520-0485\(1989\)019<0384:OTOOSW>2.0.CO;2](https://doi.org/10.1175/1520-0485(1989)019<0384:OTOOSW>2.0.CO;2).
- Fontaine D.N., P. Marrec, S. Menden-Deuer, H.M. Sosik, and T.A. Ryneerson. 2024. Time series of phytoplankton net primary production reveals intense interannual variability and size-dependent chlorophyll-specific productivity on a continental shelf. *Limnology and Oceanography* 70(1):203–216, <https://doi.org/10.1002/lno.12749>.
- Friedland, K.D., K.R. Tanaka, S. Smoliński, Y. Wang, C. Hodgdon, M. Mazur, J. Wiedenmann, C. Goetsch, and D.E. Pendelton. 2023. Trends in area of occurrence and biomass of fish and macro-invertebrates on the Northeast US Shelf Ecosystem. *Marine and Coastal Fisheries* 15:e10235, <https://doi.org/10.1002/mcf2.10235>.
- Hoagland, P., D. Jin, E. Thunberg, and S. Steinback. 2005. Economic activity associated with the northeast shelf large marine ecosystem: Application of an input-output approach. Pp. 157–179 in *Large Marine Ecosystems*, vol. 13, T.M. Hennessey, J.G. Sutinen, eds, Elsevier, [https://doi.org/10.1016/S1570-0461\(05\)80031-X](https://doi.org/10.1016/S1570-0461(05)80031-X).
- Hunter-Cevera, K.R., M.G. Neubert, R.J. Olson, A. Shalapyonok, A.R. Solow, and H.M. Sosik. 2020. Seasons of Syn. *Limnology and Oceanography* 65:1,085–1,102. <https://doi.org/10.1002/lno.11374>.
- Lucey, S., and J. Nye. 2010. Shifting species assemblages in the Northeast US Continental Shelf Large Marine Ecosystem. *Marine Ecology Progress Series* 415:23–33, <https://doi.org/10.3354/meps08743>.
- Marrec, P., H. McNair, G. Franzè, F. Morison, J.P. Strock, and S. Menden-Deuer. 2021. Seasonal variability in planktonic food web structure and function of the Northeast US shelf. *Limnology and Oceanography* 66:1,440–1,458, <https://doi.org/10.1002/lno.11696>.
- Marrec, P., and S. Menden-Deuer. 2024. Changes in phytoplankton size-structure alter trophic-transfer in a temperate, coastal planktonic food-web. *Limnology and Oceanography Letters* 9:624–633, <https://doi.org/10.1002/lol2.10410>.
- Saba, V.S., S.M. Griffies, W.G. Anderson, M. Winton, M.A. Alexander, T.L. Delworth, J.A. Hare, M.J. Harrison, A. Rosati, G.A. Vecchi, and R. Zhang. 2016. Enhanced warming of the Northwest Atlantic Ocean under climate change. *Journal of Geophysical Research: Oceans* 121(1):118–132, <https://doi.org/10.1002/2015JC011346>.