Marine Data 4 Ocean Health

Training Workshop
29 & 31 January 2025

PROGRAMME OF THE EUROPEAN UNION

Copernicus
Europe's eyes on Earth

Copernicus Marine Service

Implemented by
MERCATOR OCEAN INTERNATIONAL

MERCATOR OCEAN
INTERNATIONAL
2 avenue de l'Aérodrome de Montaudran, 31400 Toulouse, FRANCE
Tél : +33 5 61 39 38 02 - Fax : +33 5 61 39 38 99
Société civile de droit français au capital de 2 000 000 € - 522 911 577 RCS Toulouse - SIRET 522 911 577 00024

marine.copernicus.eu
mercator-ocean.eu

SOMMAIRE

Introduction

Here you will find the unanswered questions of the training workshop "Marine Data 4 Ocean Health" held the 29th and 31st of January 2025. More than 1450 people took part of the event, hence this post-event report. You can also have a look at the replays on our YouTube Channel here.

Please keep in mind that some questions asked were very similar, as a result we've summarized the different topics you covered in the Q&A 😊

We thank you again for your presence and hope to see you again!

1) Session #1 – Products & Use Cases

General questions about the Copernicus Marine Service

Some useful links :

- Home | CMEMS
- Products catalogue : Copernicus Marine Data Store | Copernicus Marine Service
- Past use cases : Use Cases | CMEMS
- Everything you need as a user : User Corner | CMEMS
- Tutorials from past events : E-learning materials | CMEMS

- Is the data freely available or do we need to pay? Can it be use for commercial purposes or to inform policy makers ?

The vizualisation, the access and the download of data are totally free: Access data | CMEMS.

Once it is downloaded, you can do whatever you want with the data, you just need to cite the Copernicus Marine Service in your work (the citations are pre-made on each product to facilitate the task).

- Is it possible to use a format other than netcdf ?

Yes, it is possible to use a data format other than .nc for Copernicus Marine products. You can specify the format of the downloaded data to be either NetCDF or Zarr. To do this, you can use the following options:

- For the Command Line Interface (CLI), use --file-format <format> where <format> can be [netcdf | zarr].
- For the Python Library (API), use file_format=<format> where <format> can also be [netcdf | zarr].

By default, the data is delivered in NetCDF format, but if you specify Zarr format, you can download the data in that format as well. Additionally, the majority of Copernicus Marine products are produced in NetCDF format, but they can also be downloaded

in Zarr format via the Copernicus Marine Toolbox, and some datasets provide GeoTIFF and Shapefile data.

Here is a link to an article to look into it in detail:
What are the formats of Copernicus Marine products? | Copernicus Marine Help Center

- How can you be sure about Copernicus data accuracy ?

Copernicus data come from several sources : satellite, in-situ and model data. All of them are tested, updated, scientifically-driven data. Useful information, such as error levels of observations or indicators on the reliability of forecasts is provided consistently and effectively.

MyOceanPro Viewer Demonstration : MyOcean Viewer | CMEMS

- Can we save our progress or history to continue our work later or to make some corrections or edits?

Yes, in MyOcean Pro Viewer, you can save your work progress. The tool allows you to save graphs in your session, which means you can continue your work later or make corrections and edits as needed. Additionally, you can minimize and expand graph windows, which can be useful for managing your workspace and focusing on specific data or analyses.

- This website looks very familiar to Giovanni/GESDISC of NASA. Is it from the same data structure?

This is not the same information system. CMEMS information, including Marine Data Store and Frontend tools are specially developed by and for CMEMS.
Similar technologies may exist, but we don't know enough about GDISC/NASA to be able to provide any further details.

- What is the programming language used to analyze the ocean data using Myocean Viewer? Are the results justifiable?

Most of MyOcean is in JavaScript/TypeScript, including the data analysis the user refers to. There is a small amount of Python in the viewer, which is the part that creates the netCDF subset files.
The only data analysis done are statistics : averages (geographic and temporal), minimum, maximum, very basic calculation.
We'd like to see examples of calculations that could be questioned.

- Please, can we find this data and work in cloud with Google earth engine?

We don't think so, but that's more up to Google. The user should communicate to Google that they should give access to this CMEMS data.

- The CMEMS presents results of numerical simulation of the vast field dataset and satellite data? or we can find raw field data and raw satellite data?

CMEMS provide numerical model output and satellite observations. The observations are provided in different level of processing (Level 3 & Level 4).

- I have 20 sites I visit once within the 20 years. Can I scrape the sea surface temperature for each site and the specific year visited? Possible?

The Marine Point Coordinates Downloader can do this if you specify the date in the table to be imported. The points must be space-time points, which seems to be the case here.

- Can Python Google Collab can easily take the link for the analysis as it does with Google Earth Engine in Google Collab

If the question is whether it's possible to view a MyOcean dataset in a Google Colab notebook, as can be done with Google Earth Engine, then the answer is yes with the WMTS link. If this isn't what you meant with this question, please contact us directly in the Help Center chat.

- Is it possible to download data in CSV or JSON format?

It is possible to download data from objects drawn directly from the Viewer (point, line, area) in CSV format by clicking on the "..." icon in the graphics window > Export to CSV

You can download data in their NetCDF native format and convert them into CSV as explained in this article.

It's not possible to download data as JSON, except when downloading the metadata of the Copernicus Marine catalogue via the describe function of the Toolbox (cf. self-versioned documentation to see JSON as output dictionary).

- Does CMEMS have downscaled projections data?

Copernicus Marine funds research projects and participates in projects on climate projections (example: Towards downscaled regional climate change projections for marine environments, May 2021). But for the moment, we do not provide climate projections in our data catalog.

- Tried to find out data for Bangladesh territorial sea for 2019 for chlorophyll but it showed no data available. So what is the data viewing limitation?

When you don't see any expected data, please check the product and dataset are providing data for the area and date of interest. There are NRT products for present/forecast and 2 years in the past data, and MY products for decades in the past. Also, check the date picked from the bottom timeline in the Viewer.

Data assimilation & optimization – Quentin Hyvernat

- What is the length of the in-situ data ?

The in-situ data from the BGC-Argo floats used in my method lasts for one seasonal cycle, from January 2025 to December 2025. In general, the BGC-Argo products sample a vertical profile from 1,000 metres to the surface every 5 to 7 days with a resolution of the order of a metre. The lifetime of observations from a single float, and therefore the number of profiles seen by a single float, varies enormously. The vast majority of BGC-Argo floats manage to produce observations for several years.

Attached is a reference paper with more information : https://doi.org/10.1146/annurev-marine-010419-010956

- Does the satellite data need to be corrected or the data is ready to use for the user ? Or can it be used to filter as we do it sometimes for the optical data ?

This question goes beyond my area of expertise. I can only answer as a user. In my case, the satellite data was used as distributed, without any corrections on my part. The CMEMS datasets that I used enabled me to estimate the uncertainties on these products. I used this data to estimate the uncertainties I could have on my results rather than correcting these products directly. In addition, whenever possible I used several sets of observations in order to calculate a new uncertainty on the observation that I would then use.

For example, some BGC-Argo and satellite observations can be quite far apart. In this case, I would recommend taking the strong point of each observation as the truth. In this case, we kept the temporal variation of the Chla observed by the BGC-Argo float, because its temporal resolution was superior to the L3 satellite data along the float used in my study. And we could very well imagine using the L3 satellite observations to estimate a new uncertainty on the Chla observations from the BGC-Argo floats.

User testimony #1 – Giuseppe Aulicino (Parthenope University of Naples

Link to Giuseppe's paper: https://doi.org/10.1016/j.ecolmodel.2021.109619

We answered all of the questions during his Q&A session, you can have a look at the replay if you wish.

User testimony #2 – Georgios Sylaios (Democritus University of Thrace)

Link to Georgios' paper: https://doi.org/10.3390/jmse10030411

- I would like to generate seasonal variation of forel-ule ? Can you please suggest the best tool ? And where can I find data ?

This is a scale of ocean color classification mostly used in citizen science based mostly on Secchi disk observations. However, remote sensing data (e.g., from Sentinel 3 OLCI, Meris, Modis, etc) could be used to obtain indirectly the FU index. Please read the following publication

https://link.springer.com/article/10.1007/s11356-021-18083-0

- What is the usual standard duration considered in the oil spill simulations ?

A typical run (2 days hindcast + 5 days forecast) on a typical laptop with currents and waves from CMEMS and winds from NOAA-GFS takes about 20 mins.

User testimony #3 – Diego Panzeri (National Institute of Oceanography and Applied Geophysics)

Link to Diego's paper: https://doi.org/10.3389/fmars.2024.1308325

- Are these types of models useful also for other species, like sexile species as octocorals ?

Yes, this type of model is and has also been used for sexile species, see for example the work in the Strait of Sicily by Lauria et al. 2017 (here: https://www.nature.com/articles/s41598-017-08386-z), in the journal Scientific Report. In this case, I recommend models or an ensemble of models such as GAM, GLM and machine learning to assess any differences.

- Will it be possible to do modelling of shore-dwelling mussel species' movements? What kind of data collection method would you recommend for small mussel species?

I'm not a shellfish expert, but I think it might be possible to model this way too, depending on the data you use. I think a sample survey (e.g. with a dredge net along the coast) could be informative. Any georeferenced and standardized data (as a survey protocol) could be modeled with this type of approach. I'm not sure if this is the best solution, but it can be tried.

- Why do you choose these 9 species? Are they ecologically or economically important to the study site? or any major threats to these species? Did you try also marine birds for example ?

I chose these species because they are important in the area from a commercial point of view and because the ultimate aim was to identify the best area for juveniles and spawners (under past, present and future conditions) that could be affected by climate change and have an indirect impact on fisheries in the Adriatic and Ionian Seas. In addition, the survey has a good catchability for this species and could provide good information on the status of these resources. Some species are overexploited, suche as European hake and Norway lobster. Regarding marine birds I've never tried, but here (https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.8272) you can find something.

2) Session #2 – Practical session

General questions

- How do I download CMEMS data directly using R?

As I mentioned in the presentation, it is possible to download CMEMS data using the toolbox. You can refer to this article to look into this in detail! https://help.marine.copernicus.eu/en/articles/8638253-how-to-download-data-via-the-copernicus-marine-toolbox-in-r

- How can a user navigate all of this for different data/analysis alone, without any prepared code ?

You can find on our website articles about everything : technical issues, how to use our tools, how to download the data and others here : User Corner | CMEMS
If you don't find the information you want, you can also contact our user support 😊

- I have general question: is there a possibility to compare data with sea flor topology map ? Is there in CopernicusMarineService sea floor topology map ?

A lot of Copernicus Marine products provide a bathymetry dataset (i.e. sea floor topology). You can download this dataset to do the analysis.

The Copernicus Marine Toolbox

Some useful links :

- Copernicus Marine Toolbox - Introduction | Copernicus Marine Help Center
- Copernicus Marine Toolbox - Installation | Copernicus Marine Help Center

- If using aggregation functions e.g. mean, max, etc. on dataset returned by open_dataset, where aggregation takes place (on server or user computer)

Open_dataset lets you remotely access the dataset online. However, all calculations on the dataset take place on the user's computer. You'll need to remember to save your work locally.

- Is it possible to connect MATLAB data to Jupyter Notebook? If I have image processing data in MATLAB, can I visualize it within Jupyter? If so, how can I do?

Although User Support does not guarantee MATLAB support, it is perfectly possible to import and visualize MATLAB data in a Jupyter Notebook, provided you export it in a Python-compatible format (e.g. CSV, MAT via the scipy.io library, etc.). Once the data has been exported, you can load it into Jupyter using Python libraries adapted to image processing and visualization.

- In my Jupyterlab, cmems v4 kernel is not present. How can İ incorporate this kernel?

It is not possible to add a Kernel to JupyterHub CMEMS online.
Please check that you are logged in at the following address with your CMEMS credentials: https://jupyterhub-cmems.mercator-ocean.fr/
Access to the JupyterHub is limited to 2 weeks from the last training session you attended. We recommend that all users create their own Python environment to work locally.

- Not a question strictly related to this presentation: It would be very interesting and useful if copernicus marine service releases a docker image

If ever the question concerns the Copernicus Marine Toolbox: there is a docker installation of the Toolbox (see article on installation).

- Some mediterranean sea products were not available for access through R CopernicusMarine package a few months ago. Has this been updated and fixed?

There is no difference of access to Copernicus Marine data between Toolbox Python API and R, but it could happen data is not accessible. In this case, don't hesitate to check our Notification User Service (UNS) and to contact us by chat.

- Is there a way to find the dataset id straight from the toolbox without going to the homepage?

No, it's not possible. To get the dataset_id, you will have to retrieve it from the datastore when selecting your product. You can also obtain it by using the 'Add a Layer' option in the MyOceanPro tool, which will take you to the datastore where you can select your dataset_id from your product page.

- Open_dataframe function works too when you have data in EASE Grid ?

In the context of working with EASE Grid data, the read_dataframe function can indeed be useful for reading data, provided the grid data is in a format that can be structured into a dataframe. However, some preprocessing might be required if the EASE Grid format differs significantly from typical dataframe-compatible formats (e.g., CSV, Parquet).

If the data is in EASE Grid format, you might need a compatible library or conversion step (such as using GDAL or another geospatial library) to correctly interpret the grid structure before loading it into a dataframe. Once converted or adapted, you can use read_dataframe to handle it within your workflow, especially for data analysis or transformation tasks.

Jupyter Notebooks – Alexandre Homerin (Noveltis)

All of the questions were answered during the event. You can watch the replay if you wish or download the tutorial if you didn't yet 😊 : Jupyter Notebook tutorial

QGIS – Daria Andrievskaia (Noveltis)

You can watch the replay if you wish or download the tutorial if you didn't yet 😊 : QGIS tutorial

- The Hurricane Michael being counterclock wise and being showed clockwise during the tutorial ?

The hurricane is definitely counter-clockwise. The arrows shown clockwise were those of currents, not hurricane winds.

R Studio – Simon Millet (Noveltis)

All of the questions were answered during the event. You can watch the replay if you wish or download the tutorial if you didn't yet 😊 : R Studio tutorial

Thanks again for your participation 😊