Project Title	**Blue-Cloud 2026: A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters**
Project Acronym	Blue-Cloud 2026
Project Number	101094227
Type of project	RIA – Research and Innovation Action
Topics	HORIZON-INFRA-2022-EOSC-01
Starting date of Project	01 January 2023
Duration of the project	42 months
Website	www.blue-cloud.org

D4.3 – VLabs User's Handbook

Work Package	**WP4 | Blue-Cloud Virtual Labs for demonstrating cross-domain web-based open science**
Task	T4.1 | WP4 coordination
Lead author	Cyrielle Delvenne (VLIZ)
Contributors	VLab1 : João Vitorino (IH), Vânia Lima(IH), Enrique Castrillo (SOCIB), Jay Pearlman (IEEE), Rene Garello (IEEE) VLab 2 : Alexander Barth (ULiège), Abel Dechenne (ULiège) VLab 3 : Steven Pint (VLIZ) VLab 4 : Francesco Palermo (CCMC), Catalina Reyes (OGS), Simona Simoncelli (INGV), Enrico Baglione (INGV) VLab 5 : Julien Barde (IRD), Yannis Marketis (FORTH)
Peer reviewers	Marine Vernet, IFREMER, Sara Pittonet, Trust-IT
Version	V1.0
Due Date	30/01/2025
Submission Date	31/01/2025

Dissemination Level

X	PU: Public
	CO: Confidential, only for members of the consortium (including the Commission)
	EU-RES. Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
	EU-CON. Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
	EU-SEC. Classified Information: SECRET UE (Commission Decision 2005/444/EC)

Version History

Funded by the European Union

Revision	Date	Editors	Comments
0.1	06/02/2024	Patricia Cabrera (VLIZ)	Table of Content
0.2	23/09/2024	Cyrielle Delvenne (VLIZ)	Executive summary
0.3	13/11/2024	Cyrielle Delvenne (VLIZ)	Structure update
0.4	01/12/2024	VLab leaders and partners (IH, SOCIB, ULiege, VLIZ, CMCC, OGS, INGV, IRD)	Content of each section
0.5	13/12/2024	Cyrielle Delvenne (VLIZ)	Content harmonization

Glossary of terms

Item	Description
AI	Artificial Intelligence
API	Application Programming Interface
CKAN	Comprehensive Knowledge Archive Network
CMEMS	Copernicus Marine Environment Monitoring Service
CNR	Italian National Research Council
CMS	Content Management System
CVS	Code Versioning System
CWP	Coordinating Working Party on Fishery Statistics
DD&AS	Data Discovery & Access Service
DIVAnd	Data-Interpolating Variational Analysis in n dimensions
D4Science	Data Infrastructure promoting Open Science (managed by CNR)
ECMWF	European Centre for Medium-Range Weather Forecasts
EMODnet	European Marine Observation and Data Network
EOSC	European Open Science Cloud
EOVs	Essential Ocean Variables
FAO	Food and Agriculture Organisation
FIRMS	Fisheries and Resources Monitoring System
GRSF	Global Record of Stocks and Fisheries
GUIS	Graphical User Interface
JSON	Javascript Object Notation
IDE	Integrated Development Environment
MFS	Mediterranean Forecasting System (MFS)
MHW	Marine HeatWave
OGC	Open Geospatial Consortium
RFMO	Regional Fisheries Management Organisations
RIA	Rich Internet Application
SSC	Sea Surface Currents
SPARQL	Standard query language and protocol for Linked Open Data on the web or for RDF triplestores
TBD	To Be Defined
TBA	To Be Arranged
VLab	Virtual Laboratory
VRE	Virtual Research Environment

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Keywords

EOSC; Virtual Labs; Big Data; Virtual Research Environment; Data infrastructures

Disclaimer

The Blue-Cloud 2026 project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101094227. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

EXECUTIVE SUMMARY

By working closely with long-term EU marine data services (i.e., EMODnet, Copernicus), and research data infrastructures (i.e., EuroArgo, SeaDataNet, Ecotaxa and others), the Blue-Cloud project is offering to the marine community a state-of-the-art cyber platform designed to implement an incubator of new research ideas, supporting scientists and researchers in designing, testing, and evaluating innovative computational analytical flows that extract valuable knowledge from diverse datasets, while accelerating open science and collaborative research in ocean sustainability.

The Blue-Cloud Virtual Research Environment (VRE) offers a comprehensive environment for sharing data, tools, and knowledge, fostering innovation and discovery. By integrating advanced analytical tools, robust computing resources, and diverse datasets from observations and models, the platform empowers researchers to tackle complex ocean challenges and enhances the scientific community's ability to understand and predict marine phenomena while fostering the development of cutting-edge technologies and methodologies.

The Blue-Cloud innovation potential is explored and unlocked by **dedicated demonstrators** as Virtual Labs (VLabs) co-designed with top-level marine researchers to demonstrate the power of the Blue-Cloud Open Science platform. The objective is to deploy a wide range of custom methods and technologies, such as softwares and algorithms, using cloud-based services in the Blue-Cloud VRE. The VLabs will also bring new data types, such as currents or carbon data, and will also be able to incorporate the Data Discovery and Access (DD&AS) Service to their workflows.

This deliverable provides guidelines on how to use each of the VLabs developed during the Blue-Cloud 2026 project. Users are encouraged to discover and understand the services and run them as proposed by the developers.

The document is structured by sections for each of the VLabs, from VLab one to five. Please note that this document will be improved throughout the project duration and with the input of users, so feel free to reach out if you have any comments or suggestions.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1. Introduction

Welcome to the Users Handbook for the Blue-Cloud 2026 Virtual Laboratories (VLabs). This handbook serves as a comprehensive guide, providing essential information and practical instructions for effectively utilizing the VLabs.

The document is organized into five main sections, each dedicated to a specific VLab. These sections highlight the unique capabilities and objectives of each laboratory, ensuring users can navigate and apply their features efficiently. Notably, Section 5 (VLab 4) is further divided into four subsections, each focusing on a distinct marine environmental indicator, as detailed below:

- **VLab #1**: Integration of Coastal Ocean Observations Along Europe (ICOOE)
- **Vlab #2**: Coastal Currents from Observation
- **VLab #3**: Carbon Plankton Dynamics
- **VLab #4**: Marine Environmental Indicators
    - Marine Heatwaves (MHW)
    - Ocean Heat Content
    - Eutrophication (TRIX)
    - Storm Severity Index Version 2 (SSI V2)
- **VLab #5**: Global Fisheries Atlas

Each section begins with an introduction to its respective VLab, outlining its purpose, scope, and **target users**. This is followed by a detailed explanation of the **state-of-the-art** methods employed and the **input data sources** used.

To guide users in maximizing the potential of each VLab, a **step-by-step user guide** is included, offering clear and practical instructions. For those working with alternative data sources, dedicated subsections provide guidance on adapting the outlined procedures.

Finally, each section concludes with information about the contributing authors and a comprehensive list of references, ensuring proper attribution and supporting further exploration.

This handbook is a **living document**, meaning it will evolve over time. Updates will be made throughout the project based on user feedback and advancements in the VLabs' capabilities. Your insights are invaluable in helping us refine and improve this resource.

Thank you for being part of the Blue-Cloud 2026 journey. We hope this handbook serves as an invaluable tool in your exploration and application of the VLabs.

2. Registration to the Blue Cloud 2026 VRE

To be able to access Blue-Cloud 2026 services, users first need to register to the Blue-Cloud 2026 platform by creating an account on the **Blue-Cloud 2026 Gateway** (Figure 1).

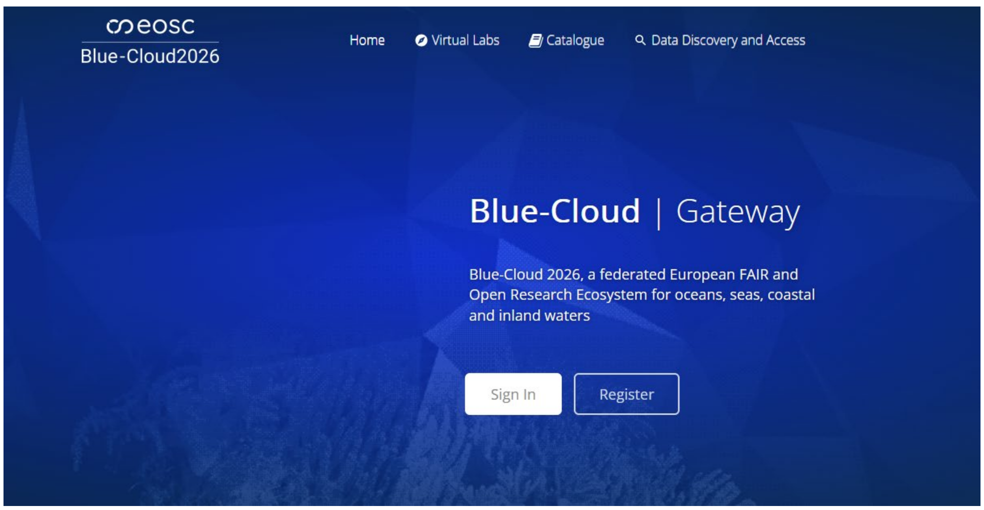


Figure 1. Blue-Cloud Gateway

The registration process ends with a pop-up advertising on “email verification” sent by the Blue Cloud management team. After the confirmation of registration, the user can start using the Blue Cloud platform. Video tutorial on how to register on Blue-Cloud is also available here: **Blue-Cloud Hackathon - Using the Blue-Cloud Virtual Research Environment** .

Once logged in, VLabs can be accessed by clicking on “Virtual Labs” in the top centre button (Figure 2).

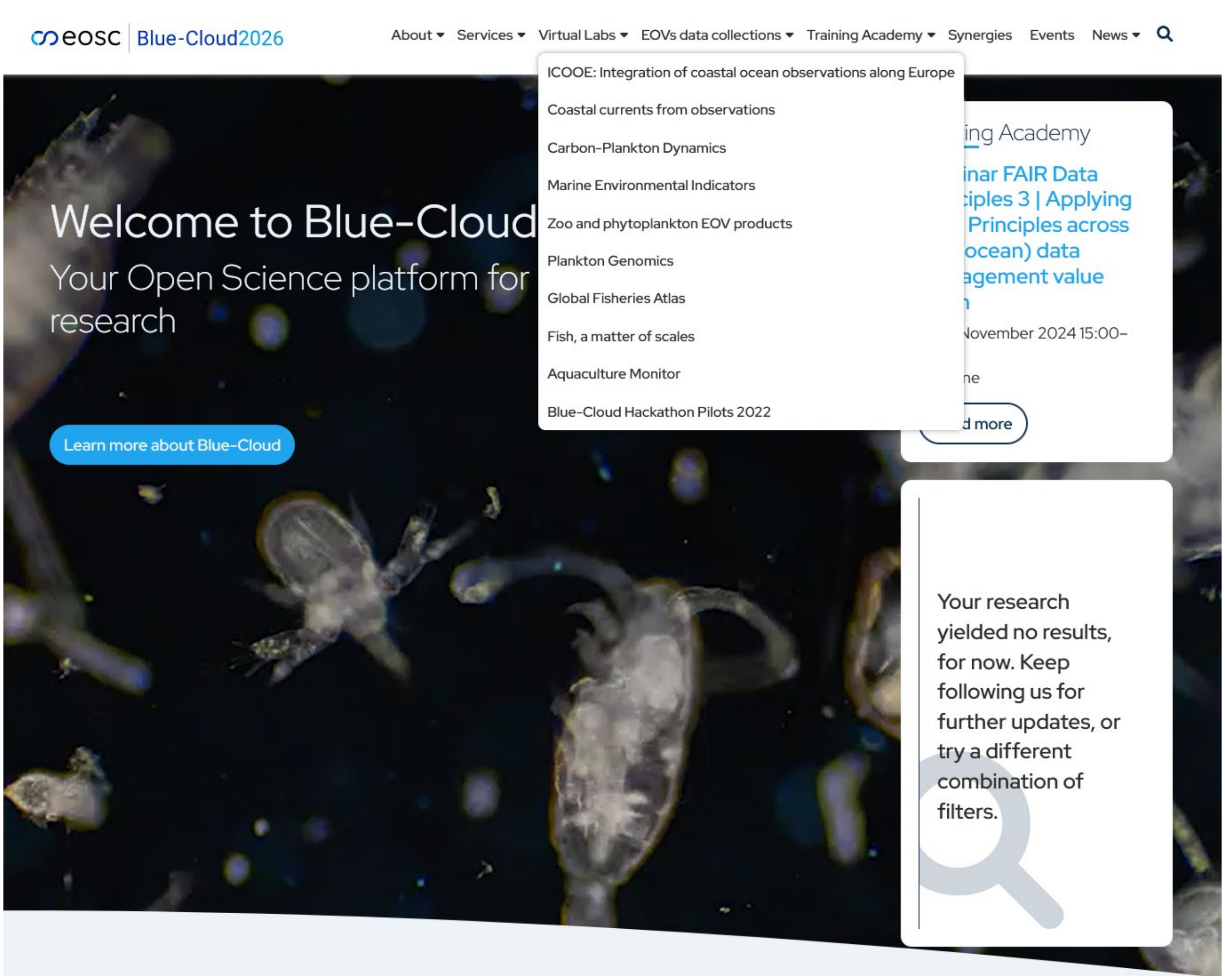


Figure 2. How to access VLabs once logged in

The credentials allow access to the **Virtual Research Environment (VRE)** as well as four BlueCloud demonstrators and five Thematic VLabs (Figure 3). BlueCloud demonstrators were developed during the first BlueCloud project while VLabs are currently being developed during the Blue-Cloud 2026 project.

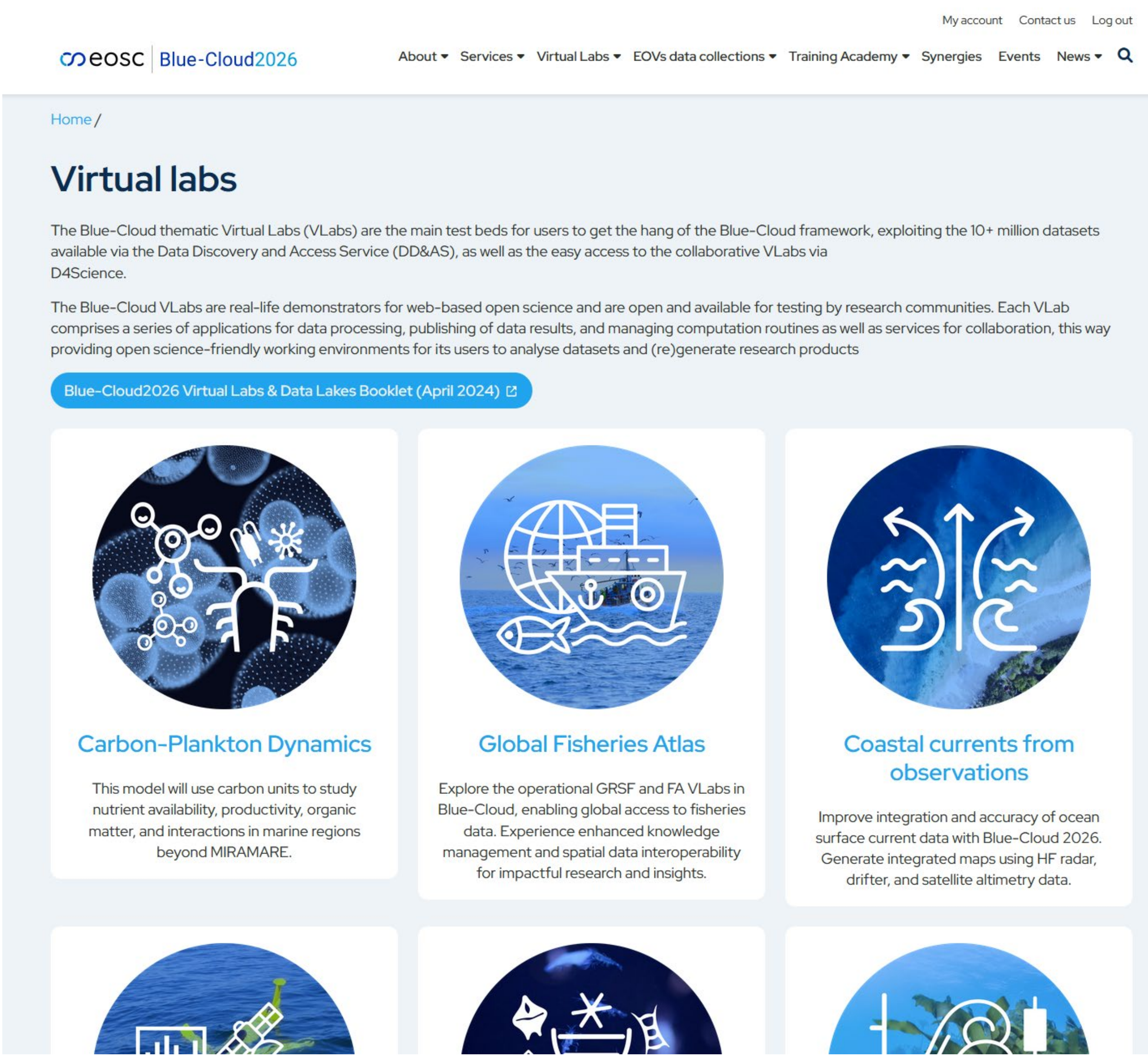


Figure 3. Virtual Labs web page

3. Introduction to VLab 1 “Integration of coastal ocean observations along Europe”

This VLab provides users with an environment that is specifically designed to support them in their explorations of the available coastal ocean information, developing and implementing tools to facilitate the access to different types of data, to data processing and quality control when needed, to exploitation of individual datasets and to integrated analysis of different types of data. It will bring together observations collected by partners of the Joint European Research Infrastructure for Coastal Observatories (JERICO-RI) with other available data from international data repositories.

VLab 1 will support the coastal ocean users through three Thematic Services:

- **Thematic Service 1,** addressing **“Transboundary processes and connectivity along the European margins”,** aims to support user interest in processes such as biological connectivity, contaminants spread, and impacts of river outflows along margins.
- **Thematic Service 2, “Extreme Events”** , focusing on coastal impacts of major storms.
- **Thematic Service 3, “Ocean glider”** will show the added value of repeated glider sections.

The VLab will provide:

- ➢ Easy **identification** and access to available **observations** and complementary data for the coastal ocean areas of Europe;
- ➢ **Easy** handling/management of the available datasets, implementing, when required, the **pre-processing** (including additional quality control of the data) and **reformatting** of the data in a transparent way for the user and following the accepted Best Practices.
- ➢ Access to a panoply of exploitation tools aimed to **extract** from the datasets valuable information;
- ➢ Provide specific tools for the integrated **analysis** of different datasets, covering complementary aspects of the coastal ocean environment;
- ➢ **Visualization** tools adapted to each Thematic Service specific objectives.

One central idea in the base of the implementation of this VLab is the recognition of the multiplicity of skills that are required for a user to progress in the use of the available observations of the coastal ocean environment to extract relevant information about the different processes that take place in these areas, of the impacts these processes promote. The simple identification of, and access to, the different available datasets is frequently not a guarantee that full extended use can be made of this data.

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union’s Horizon Europe programme under Grant Agreement no. 101094227

3.1. Target Users & Community

VLab 1 specifically aims to support the broad community of users that work on the coastal ocean environments of Europe. The Coastal Ocean Research community is clearly at the front of the high priority users of this VLab. The three thematic services proposed in this VLab bring to researchers the tools for processing, exploring, integrating and visualizing the different observations and associated complementary information that are available in coastal ocean areas, helping to remove the barriers to a comprehensive use of this datasets and boosting the understanding of the variety of processes operating in those areas. Users from the Blue Economy sectors or from Crisis Management entities can find in the Thematic Services offered by this VLab an easy way to access, explore and integrate the different observations or modelling results that can be relevant to support their operations. . Environmental managers and policy makers can use the Virtual Lab as a rapid and powerful vehicle to build an integrated vision of geographical areas of interest and to explore different levels of understanding about the relevant processes and impacts occurring on these areas, in this easy improving the knowledge base that supports the decision making process. Finally, students and the general public can use Vlab1 as a window to the coastal ocean, exploring available observations and complementary information using the different tools provided.

The VLab builds from the expertise in coastal ocean observation and research that is gathered in the JERICO-RI community, and a close interaction with this community is expected.

3.2. State of art

VLab 1 is based on methods that are commonly accepted by the scientific community and conform to the Best Practices recommendations and documents from OBPS and Aquadocs. In particular, the methods and tools implemented as part of Vlab1 follow the Best Practices developed by the Joint European Research Infrastructure for Coastal Observatories (JERICO) community, for observations in coastal ocean using, for example, HF radars or fixed platforms and Brest Practices commonly adopted by the glider community.

The Thematic Service #1 aims to provide a variety of tools that can support users in working with the observations and complimentary information that are available for a coastal ocean and in extracting from these the information about the transboundary transport and connectivity processes. The Thematic Service #1 is structured on the basis of two central ideas about the coastal ocean environment:

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

1) The coastal oceans are areas of confluence of multiple influences and forcing mechanisms, many of which originate from the domains that directly interact with the coastal ocean, such as the deep ocean environment, the river and estuarine environment or the atmosphere.
2) The coastal ocean environment contains specific mechanisms that promote the long distance propagation of perturbations. The bathymetric variations associated with the shelf and slope topography, in combination with the presence of the coastal boundary, provide the basic mechanism allowing the coastal ocean to behave as a waveguide. Disturbances caused in a given location (due, for example, to wind forcing variability or to the impact of located continental or open ocean influences) can efficiently be transmitted along the continental margin in the form of barotropic shelf waves (Robinson,1964) or, in the more general case of combined influence of bottom topography and stratification, as continental shelf waves (Allen, 1975; Brink, 1991). These processes allow that the disturbances can propagate sometimes by thousands of kilometers along the continental margin, impacting the coastal ocean conditions in areas far from the location of the disturbance. In the transition between the continental shelf and the open ocean, the large-scale density distribution (and associated geostrophic transport) interacts with the abrupt topographic change associated with the continental slope leading to the development of strong, slope intensified polewards flows through the JEBAR - Joint Effect of Baroclinicity and Relief - mechanism (Huthnance, 1984). These slope currents can transport open ocean influences (e.g. warm and high salinity waters or biological species) along the outer edges of the coastal ocean areas. They can also collect shelf and continental (e.g river discharges) influence and propagate these influences to remote areas far from their origin.

   In addition to these long distance along-margin propagation, important coastal ocean processes also act to promote across-margin propagation of disturbances. Particularly important examples of these processes are the development of large upwelling filaments extending one or two hundreds of kilometers offshore (areas exposed to upwelling favourable winds) and the important exchanges promoted by long submarine canyons. These processes all contribute to promote an important interaction between the coastal ocean and the complementary domains such as the open ocean environment or the continental domain.

   Thematic Service #1 focus specifically on the processes described above, providing a broad range of tools that users can apply to extract from the available observations (and other complementary data) the relevant information about how these processes are expressed in the area of interest and how they are contributing to promote the interaction between the different parts of the coastal ocean domain inside this area . Since the processes described above operate mainly at subinertial time scales (periods longer than the local inertial periods), which are dominated by the effects of Earth rotation, the Thematic Service #1 analytical tools are restricted to low-pass filtered data.

Thematic Service #2 will open to the users the access to a number of tools aimed to support the research on and understanding of extreme events that impact the coastal ocean regions of Europe. These include, for example, the identification of the periods affected by storm conditions in a given geographical area and for a given period of interest, the characterization of these storms from the available measurements of waves, winds and other oceanographic parameters or the potential impact of the storms on the bottom sedimentary cover.

Thematic Service #3 is a toolbox for the computation and visualization of physical/biogeochemical variables from gliders, as well as derived variables and indicators that are relevant for both science and society.

Taking as a base example The Ibiza Channel (IC) for being a well-known biodiversity hotspot and a choke point of the western Mediterranean Sea with complex topography and circulation. The significant variability of the meridional circulation at the scales of weeks and few kilometres has been explained through the variability of the water masses in the vertical, with very relevant implications on the marine ecosystem (e.g., Bluefin Tuna, Jellyfish). By monitoring the IC quasi-continuously through the deployment of gliders along an endurance line since January 2011, we have noticed that the repeated high-resolution glider sections allow us to describe and understand the ocean processes involved in the ocean circulation variability from daily/weekly to interannual scales. Semi-permanent glider sections are becoming available in different coastal to ocean areas (Boundary ocean Observing Network).

Tools such as this Thematic Service #3 will facilitate understanding the circulation in key ocean areas and its relation to water masses driving these changes that are essential to understand the role of the ocean in climate change. The developed metrics and diagnostics are: (1) vertical sections of temperature (T), salinity (S), density (ρ) and geostrophic velocity (GV), (2) T/S diagrams and water mass identification, and (3) geostrophic transports (total and per water mass). The tool also allows the use of biogeochemical (BGC) data from glider measurements (chlorophyll-a concentration, oxygen concentration and saturation, and turbidity). The data processing applied to the BGC data is the same as for the hydrographic variables (T, S) to be able to relate BGC values to the GV and associated water mass transports.

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

3.3. Input data sources

At the present stage of development of VLAB 1 users can access and work with the following variables/datasets:

VARIABLES	DATA SOURCES	DATA ACCESS
Surface Currents	Coastal high frequency (HF) radars retrieved from the European HFR node thredds catalog (data version 3.0)	https://thredds.hfrnode.eu:8443/thredds/catalog.html
Surface Currents	Nemo Model Fields retrieved from CMEMS	https://data.marine.copernicus.eu/product/IBI_MULTIYEAR_PHY_005_002/description and https://data.marine.copernicus.eu/product/IBI_ANALYSISFORECAST_PHY_005_001/services

Table 1. VLab 1 Input data sources

3.4. Step by step guideline to use the VLab

Users access the VLab from their own account in the Blue Cloud gateway (mandatory). Accessing the ICOOE VLab, the user can find two thematic services (TS) available, Thematic Service #1 (Transboundary Processes and Connectivity) and Thematic Service #3 (Ocean Glider).

Thematic Service#1: By selecting this TS, a Dashboard Virtual Research Environment is launched (Figure 2). The Dashboard VRE shows, on the right, a map of the European coastal ocean domain and, on the left, several user options.

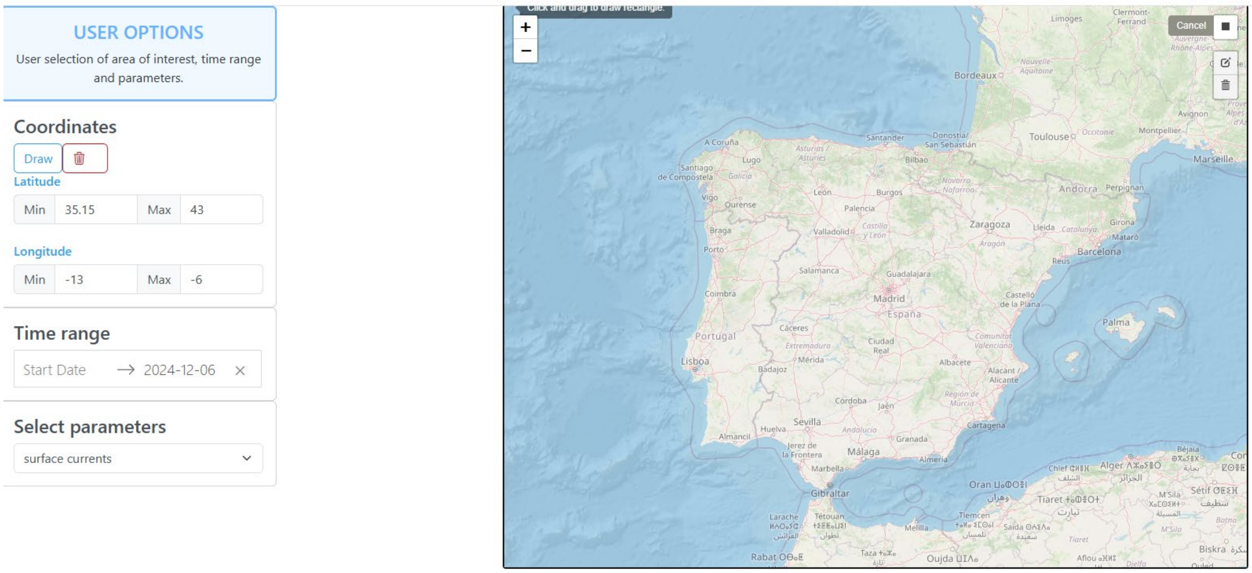


Figure 2. Dashboard Virtual Research Environment

Using the mouse/keyboard on the map, users can move around in the domain and zoom in/zoom out the geographical area (Figure 3). The geographical area of interest can be drawn directly on the map or can be input in the appropriate fields in the Coordinates box (*min* stands for minimum latitude/longitude coordinate, *max* stands for maximum latitude/longitude coordinate). To draw the area of interest directly in the map, click the *Draw* button from the Coordinates box and then draw the area of interest using the mouse. Clicking the garbage bin symbol (next to Draw) resets the geographical selection made with the *Draw* button. The information about the coordinates of the selected box is saved and presented in the Coordinates box on the left of the dashboard. The user can then proceed by selecting the time interval of interest, in the Time range box on the left side, to indicate, in a calendar, the year, month, day of the start and end of the time period of interest.

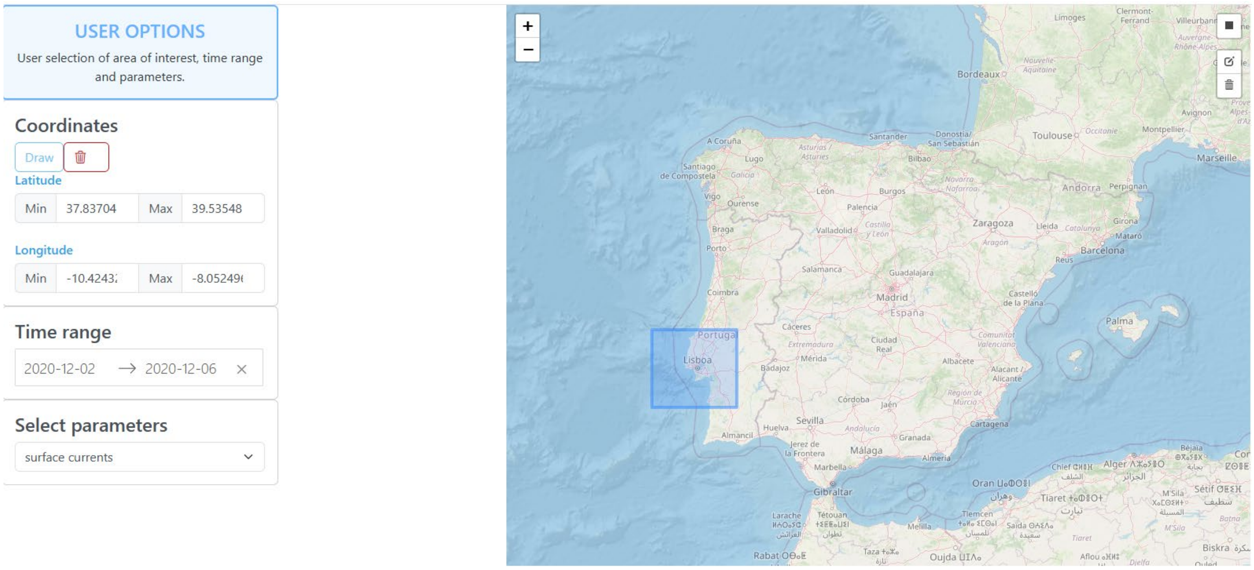


Figure 3. Area of interest

In the drop-down menu the user can select the physical parameters available, comprising surface currents, surface temperature, surface salinity, subsurface currents, subsurface temperature, subsurface salinity and sea surface height. At the present stage of development of this Thematic Service, the demonstrator only allows the selection of surface currents, which are retrieved for 2 sub-systems, concerning surface currents from high frequency radar (HFR) data and surface currents from the NEMO model. High frequency radar data is retrieved from the European HFR node for all the available networks and the NEMO model data is retrieved from Copernicus Marine Service.

After the user selection is finished, a check on the Overlap Area (OA) and Overlap Time (OT) is done, to make the user aware of the data availability from the selection made before. This way a new selection can be easily performed, if that is the case. The Available Data button becomes highlighted when all the mandatory selection fields are completed and the information of the available data is displayed afterwards, after clicking the Available data button (Figure 4).

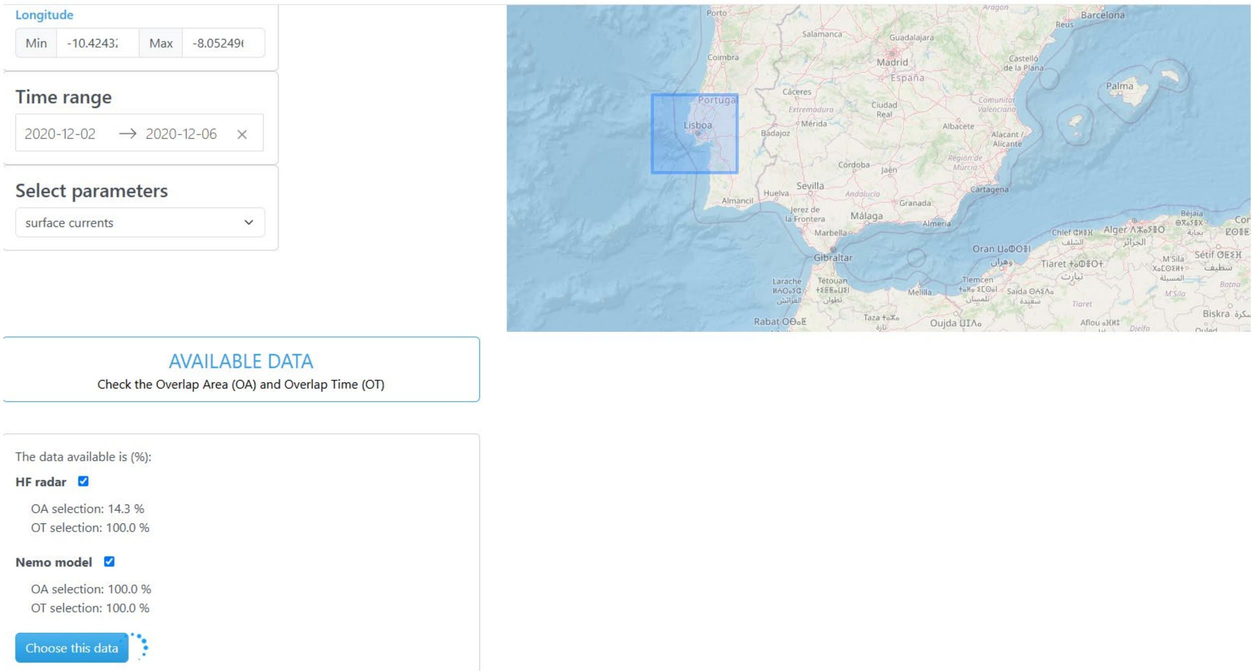


Figure 4. Available data button

Access to the data and download to the Dashboard VRE is done in a transparent way regarding the user and involves a process of connection to the data providers, with authentication in the case of the CMEMS data access. The data download for the European HFR node thredds catalog is done using the NetCDF Subset Service available and netCDF 4.1.3 data format are downloaded. The Nemo Model surface currents retrieved from CMEMS are accessed through the Copernicus Marine Toolbox (version 1.3.4) API in a Python integrated development environment. By using the Python Library (API) and the Subset function, it is possible to download a subset of the datasets of interest (by specifying the variables, geographical area, time range, and depth), as only specific parts of the data are required.

It should be noted that the data download feature is dependent on the successful connections established with the mentioned providers, which includes the login feature. Therefore, we emphasise the need to, at times, retry the data download dashboard feature, as the primary connections can sporadically fail at times due to the thredds server being down, authentication failures, and/or other updates related problems. The periodic updates of the Copernicus Marine Toolbox also require the verification of the correct work flow of the data download functionality.

The European HFR node data download always includes a broader time range than requested by the user, to accommodate data pre-processing requirements. To allow a more manageable handling of datasets comprising a long time period, the system segments the data to be downloaded in blocks of limited time length. When the data downloads start, a progress spinner appears right next to the Choose this Data button. When the data downloads are completed, the user is informed about the final status of the data download (including details about the data and datasets downloaded), by means of a message which shows up in the dashboard, right next to the Choose this Data button (Figure 5). There is a Show Logs menu, in which the user can access a session report on the tasks performed, including the ones involving the data download.

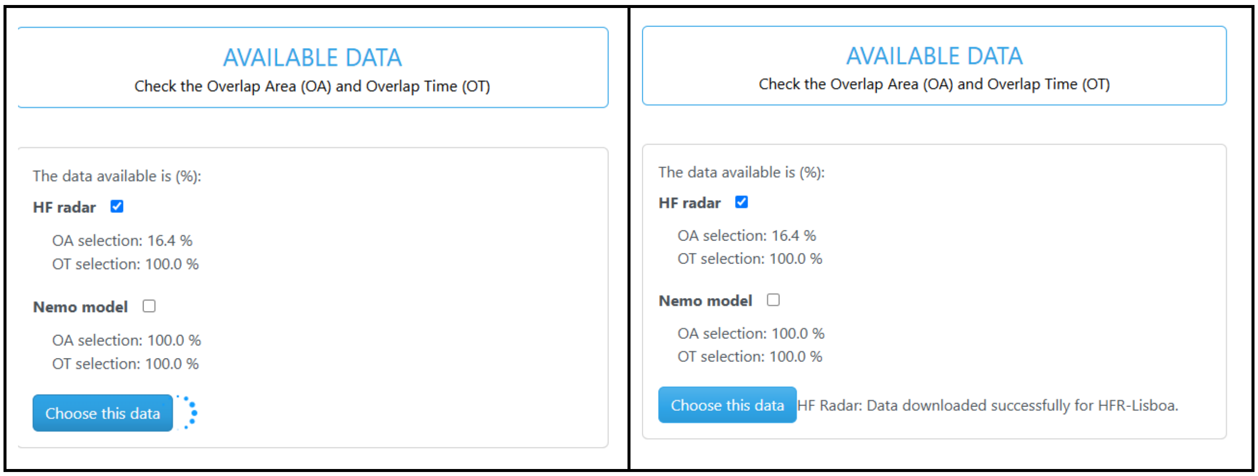


Figure 5. Status of data download

Following the downloading of the relevant datasets to be used, the Dashboard VRE proceeds with the pre-processing of these datasets. This stage, also transparent to the user, involves a number of steps that aim to transform the data in such a way that it becomes consistent with the specific requirements of the Thematic Service #1. For the HF radar data download from the European node, these steps involve the definition of a common regular grid comprising all the observed nodes, the introduction of an additional quality control criteria based on the data quality parameters provided, the identification of time gaps in the data for each node and the time interpolation of the small time gaps, the low-pass filtering of the hourly data using a running mean of 36 hours. For NEMO surface current fields, the pre-processing stage includes the detection of eventual time gaps and, in case the time period selected by the user extends from times covered by the low resolution grid product IBI_MULTIYEAR_PHY_005_002 to periods covered by the the high resolution grid product IBI_ANALYSISFORECAST_PHY_005_001 , it interpolates the low resolution data in the high resolution grid. This procedure aims to assure the compatibility of data retrieved from each one of the two products in the following exploitation/integration steps. NEMO fields correspond to daily means so no low-pass filtering is required.

After the preprocessing step is completed, the user is sent to a window presenting the different analysis that is possible to develop with the data sets selected. The analysis includes Exploration Tools aimed to explore each individual variable (e.g. Basic Statistics, Correlation Analysis, Variability Structure, others), Integration Tools to explore the interaction of different variables and different datasets (e.g. Combined Parameters, Integration Assessment or Cross-Correlation Analysis, among others) and Advanced Integration Tools (e.g. coastal ocean adjustment to external forcings).

The first demonstrator of the Thematic Service's #1 capacities shows the "Exploration Tools-Basic Statistics" option. When this option is selected, the Dashboard Virtual Environment shows a window with, on the left side, the option of the type variables to be analysed and, on the right side, three options of products that can be developed inside the "Basic Statistics" option.

By selecting one of the variables presented, the available datasets (exported previously) are presented to the user. One or multiple choices of these datasets can be selected by the user, the analysis being developed for each one of these choices.

The user can then select the type of product that can be developed from the Basic Statistics option. A predefined option, that will be always produced, corresponds to the basic statistics for the complete time extension of the dataset. Two other options are available to be added to the standard product. One corresponds to the basic statistics for predefined time periods. The user can choose either monthly statistics or weekly statistics or both. These options are only available if the length of the time series allows for the calculation of the statistics. The second option that the user can select corresponds to the basic statistics for a user-defined time period. In this case, the user is asked to provide the start time and the end time of the period to be analysed.

The products built from the Basic Statistics option are presented to the user both in table format as well as in graph format.

Thematic Service#2: This thematic service, presently in implementation, will open to users the access to a number of tools aimed to support the research on extreme events impacting the coastal ocean areas of Europe. The thematic service follows the same structure implemented for thematic service #1, described above. The users interface allows the users to select the geographical area of interest, the time period of interest and the variables to be analysed. This selection guides a search for available datasets in a number of pre-identified data aggregators. After confirmation by the users, the datasets are downloaded to the thematic service enviroment. A choice of tools for Data Exploitation, Data Integration and Advanced Analysis is then presented to the user. Following users selection of the type of analysis and the datasets on which these analyses are performed, the TS builds a number of results that can be presented both in table format as well as graphically.

Thematic Service#3: By selecting this Thematic Service (TS), users gain access to the JupyterHub utility, a cloud-based interactive computing environment preconfigured with the necessary libraries and packages to seamlessly run the Glider Toolbox. This robust toolbox enables the processing of user-stored data within the workspace, ultimately generating an advanced data product. This advanced product comprises a NetCDF file formatted in OG1.0 standards and an array of visual outputs relevant to the specific glider mission.

Upon accessing the JupyterHub interface, users are provided with a streamlined workflow to input the required information into the Glider Toolbox. The input parameters include, but are not limited to, the geographical bounding box coordinates encompassing the transects of interest, the start and end dates of the mission, and additional configurable variables such as the smoothing parameter (expressed in kilometers) for transport computation and the criterion for Water Intrusion and Withdrawal (WIW) detection.

Once the Glider Toolbox completes its computations, a dedicated folder is generated within the user's workspace. This folder houses the resulting NetCDF file, visual plots derived from the analysis, and supplementary log files that document the processing workflow. These outputs serve as comprehensive representations of the processed glider mission data.

For visualization and further analysis of the advanced data product, users have two options: they may either directly download the generated files from the workspace or leverage the Viewer utility. This utility can be accessed via a dedicated link in the upper menu bar of the Virtual Lab interface. Upon launch, the Viewer opens the Graphical Interface of Thematic Service #3, which automatically retrieves the data stored in the user's workspace. Within this interface, mission lines are displayed on an interactive map, providing an intuitive overview of the processed transects.

Furthermore, the Viewer offers users the ability to select and visualize any of the four key graphical outputs produced during the analysis: potential temperature, practical salinity, potential density, and geostrophic velocity. For each transect, users can choose the specific parameter they wish to examine in detail. Additionally, the Viewer includes a convenient download button located at the bottom of the interface, allowing users to retrieve the Advanced Data Product in its entirety for further offline analysis or archival purposes.

3.5. Using other data sources

The different data sources that are used by VLab 1 were detailed in Deliverable 4.1. Other additional data sources may include specific observations collected by members of the JERICO-RI community and are not yet available from data aggregators such as EMODNET or the European HF radar node. These are not, however, included at the present stage.

3.6. Authors

Instituto Hidrográfico : João Vitorino, Vânia Lima

SOCIB : Enrique Castrillo

3.7. References

Allen, J. S., 1975. Coastal Trapped Waves in a Stratified Ocean. Journal of Physical Oceanography, 5, 300-325. DOI:https://doi.org/10.1175/1520-0485(1975)005<0300:CTWIAS>2.0.CO;2

Brink. K.H., 1991. Coastal-Trapped Waves and Wind-Driven Currents Over the Continental Shelf. Annual Review of Fluid Mechanics, Vol. 23, 389-412. DOI:https://doi.org/10.1146/annurev.fl.23.010191.002133

Huthnance, J., 1984. Slope currents and "JEBAR". Journal of Physical Oceanography (1984) 14(4) 795-810. DOI: 10.1175/1520-0485(1984)014<0795:sca>2.0.co;2

Robinson, A. R., 1964: Continental shelf waves and the response of the sea level to weather systems. J. Geophys. Res., 69, 367-368.https://doi.org/10.1029/JZ069i002p00367

CMEMS:

Atlantic-Iberian Biscay Irish- Ocean Physics Analysis and Forecast. DOI (Product): https://doi.org/10.48670/moi-00027

Atlantic-Iberian Biscay Irish- Ocean Physics Reanalysis. DOI (Product): https://doi.org/10.48670/moi-00029

European High Frequency Radar node thredds catalog:

HFR-South: Instituto Hidrográfico & Puertos Del Estado. (2014). High Frequency Radar South network (HFR-South). European HFR-Node. https://doi.org/10.57762/RNCF-D423

HFR-Lisboa: Instituto Hidrográfico. (2012). High Frequency Radar Lisboa network (HFR-Lisboa). European HFR-Node. https://doi.org/10.57762/v31p-h205

HFR-Galicia: n/a

HFR-EusKOOS: HFR-EUSKOOS, AZTI, Euskalmet – Basque Meteorological Agency, & Basque Government Security Department. https://doi.org/10.57762/T4WH-DQ48

HFR-Ibiza: Tintoré, J., Lana, A., Marmain, J., Fernández, V., Casas, B., & Reyes, E. (2020). HF Radar Ibiza data from date 2012-06-01 (Version [x.y.z]) [Data set]. Balearic Islands Coastal Observing and Forecasting System, SOCIB. https://doi.org/10.25704/17GS-2B59

HFR-PLOCAN: n/a

4. Introduction to VLab 2 “Coastal Currents from Observations”

The VLab 2 called : “Coastal Currents from Observations” aims to show a multi-source approach framework to study and reconstruct surface currents through a variational inverse method: Data-Interpolating Variational Analysis in *n* dimensions (DIVAnd). It is presented as a Jupyter notebook and written in Julia language.
The current approach aims to reconstruct Mediterranean surface currents. The final objective is that the users can easily adapt the approach to a different region assuming that enough data is available .

Three types of data are considered for the approach:

- **Surface drifters** (essential for the cross validation framework)
- **Satellite altimetry** (dataset with a high spatial and temporal coverage used to derive geostrophic currents)
- **High Frequency (HF)** Radars

3.1 Target Users & Community

For the Coastal Currents VLab, target users are scientists in the field of oceanography. It is known that surface currents are involved in the upper layer dynamical processes and are thus important for several applications. The most likely community to use our products are modelers that need to force surface currents for their analysis or for validation of hydrodynamical models. As an example, the MEDSLIK-II model built by Euro-Mediterranean Centre on Climate Change (CMCC) researchers will take our product as a forcing variable to forecast the oil spill model.

These results will also be helpful to understand the seasonality and the interannual variability in surface currents for the Mediterranean Sea. Another use could be an educational use to teach students about reconstruction of data in oceanography and to showcase an “interpolation” technique.

3.2 State of art

Sea surface currents (SSC) are central for understanding ocean upper dynamics and interactions. Indeed, currents transport heat, mass, create conditions for up- or downwelling systems and drive the content of the primary variables such as oxygen and chlorophyll. Thus, SSC is an important factor that conditions primary production and therefore, the whole biological interaction of the trophic chain. Surface currents play also a central role in dispersion of pollutants, for example in the case of oil spill models. Due to this, SSC are part of the **Essentials Ocean Variables** (**EOVs)** (as defined by Global Ocean Observing System community), which confirms their importance to oceanography knowledge and research. As a consequence, analysis over large timescales and areas are precious for capturing variability and recurrent patterns of circulation.

However, direct SSC measurements are quite sparse at the scale of an ocean basin. Due to this issue, we need to find a way to get complete gridded files as they are used to force models or to predict dispersion of pollutants. To overcome this problem, researchers often use interpolation techniques to fill the gaps.

We propose here to bring a new way to combine indirect remote sensing data and direct in-situ observation to obtain a reconstructed gridded surface currents dataset.

The variational inverse method is called **Data-Interpolating Variational Analysis in *n* dimensions (DIVAnd)** and is based on a cost function which will penalize abrupt gradients in the analyses. The approach also considers dynamic constraints which are added to the cost function. We consider the presence of the coastline, the generally small horizontal divergence, Coriolis force, the surface pressure gradient, and temporal coherence of the system.

3.3 Input data sources

VARIABLES	DATA SOURCES	DATA ACCESS
Altimetry	CMEMS dataset: SEALEVEL_EUR_PHY_L3_MY_008_061	https://data.marine.copernicus.eu/product/SEALEVEL_EUR_PHY_L3_MY_008_061/description

Sea Surface Current (drifters)	CMEMS dataset: INSITU_GLO_PHY_UV_DISCRETE_MY_013_044	https://data.marine.copernicus.eu/product/INSITU_GLO_PHY_UV_DISCRETE_MY_013_044/description
High Frequency Radars	CMEMS dataset: INSITU_GLO_PHY_UV_DISCRETE_MY_013_044	https://data.marine.copernicus.eu/product/INSITU_GLO_PHY_UV_DISCRETE_MY_013_044/description

Table 2. VLab 2 Input data sources

3.4. Step by step guideline to use the VLab

The main code is presented as a Jupyter Notebook called "Surface_Currents_Main.ibynb". This is the main body of the script and where validation and results are shown. For facilities, we advise users to order your folders as mentioned below.

This notebook is stored in ~/CoastalCurrents/examples/, ("~", is the home directory).

To use it, first copy the notebook from the shared workspace to your home directory before running the notebook. The user cannot run the notebook from the shared workspace, as it is read-only.

The following explanation will help you to understand the different parts of the code.

1) Code structure and dependencies:

Surface_Currents_Main.ibynb is runned. For its run, it needs different functions:

Direct functions of the notebook are "Data_Preparation.jl" & "Parms_to_RMS.jl" (stored in ~/CoastalCurrents/examples/)

- Data_preparation.jl loads main variables from the dataset, removes missing values (encoded as NaNs) and selects the year.
- Parms_to_RMS.jl takes main parameters and performs the analysis (DIVAnd computation and error calculus based on independent drifter's).

Useful functions
(stored in ~/CoastalCurrents/src/)

- CoastalCurrents.jl, Altimetry.jl & hfradar.jl are used for their **loaddata()** functions which are used inside of "data_preparation.jl". It stores variables in vectors for computing facilities.
- Plotting.jl is useful for **CoastalCurrents.Plotting.plotmap(bathname)** which adds the mask on results maps.
- nc2leafletvelocities.jl is used to get dynamical outputs of the surface currents.

2) Dataset:

(stored in ~/tmp/BlueCloud2026/)

- Three folders are present there; Altimetry, Drifters and HF Radars in separate directories.
- In a separate file we find bathymetry and mask file "gebco_30sec_4.nc". This file is downloaded automatically if it is not done yet.

HFRadars files were all averaged by day. The script to do it is not given as the way of doing it is specific to the considered dataset.

3) Resolution and boundaries:

Resolution and boundaries of the analysis are tuned in ~/CoastalCurrents/examples/common.jl (Figure 6). To do so:

- lonmax (maximum longitude), lonmin (minimum longitude), latmax (maximum latitude), latmin (minimum latitude) are set.
- The resolution is given by dlon and dlat (in arc degrees).
- For squared gridded output, dlon = dlat.

```
# Grid definition
lonmax = 37
lonmin = -7
latmax = 46
latmin = 30

# resolution
dlon = dlat = 0.25              # Resolution
lonr = lonmin:dlon:lonmax       # -7 Gibraltar to 37 which is BlackSea east
latr = latmin:dlat:latmax       # 30 on Lybia to 46 onto the north of the Black Sea

```

Figure 6. Resolution and boundaries composition

4) Mask:

To load the mask, “gebco_30sec_4.nc” is used. **It is downloaded automatically.**

In “common.jl”, we find:

```
# Base Directory for Dataset

basedir = expanduser("~/tmp/BlueCloud2026")
```

```
# Mask and bathymetry in basedir

bathname = joinpath(basedir,"gebco_30sec_4.nc")
bathisglobal = true
```

Figure 7. Load the mask

Which set “bathname” and “bathisglobal”. Main variables to load the mask (Figure 7).

Then, in “Surface_Currents_Main.ibynb” we find (Figure 8):

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union’s Horizon Europe programme under Grant Agreement no. 101094227

Define the mask and necessary variables for interpolation

```
[44]: # Mask loading
      # bathname & bathisgoal are defined in common.jl

      mask,(pm,pn),(xi,yi) = DIVAnd.domain(bathname,bathisglobal,lonr,latr)
      mask = DIVAnd.floodfill(mask) .== 1
      hx, hy, h = DIVAnd.load_bath(bathname, bathisglobal, lonr, latr);


      # Trick of sea removing to avoid mistakes
      mask[:,end] .= 0;
      mask[:,end-1] .= 0;
```

Figure 8. Define the mask

Which sets the mask from bathname, bathisgoal and (lonr,latr), defined in "**3) Resolution and boundaries**".
(pm,pn), (hx,hy,h) are derived from the mask and are used in **DIVAndrun_HFRadar()**.

5) DIVAnd run:

DIVAndrun_FHRadar() which is the function used for DIVAnd method for currents and is run in the file "parm_to_rms.jl" which runs for each month. This feature could change in the future since the temporal coherence (forcing correlation between 2 time steps) is still not tuned.

This function need the following parameters (Figure 9);

- mask (Computed with **DIVAnd.domain()**, which is a 2d matrix fulfilled by 0 and 1)
- h (Bathymetry, computed with **DIVAnd.load_bath()**) in meters
- pm,pn (pm pn are parameters computed from the mask) inverse of the resolution (m^{-1})
- xi,yi (interpolated grid in degrees)
- x_tot,y_tot (observations locations in degrees)
- robs (the radial velocity or a single velocity component in m/s)
- directionobs_tot (direction vector computed from u & v in degrees)
- len (correlation length, which tune how far an observation can be interpolated)
- epsilon2 (error variance of observations, which set the accuracy of the different dataset)

It returns:

- u and v velocities for the interpolated field (in m/s).
- Sea surface height of the region.

More information is provided at https://gher-uliege.github.io/DIVAnd_HFRadar.jl/dev.

```
####################################################################################################################################################
#                                                       DIVAnd run : HFRAdar
####################################################################################################################################################


    # Month per month : DIVAnd Interpolation

    for i in 1 : 12

uri[:,:,i],vri[:,:,i],ηi[:,:,i] = DIVAndrun_HFRadar(
    mask,h,(pm,pn),(xi,yi),(x_tot[i],y_tot[i]),robs_tot[i],directionobs_tot[i],len,epsilon2[i];
    eps2_boundary_constraint = eps2_boundary_constraint,
    eps2_div_constraint = eps2_div_constraint,
    # eps2_Coriolis_constraint = -1,
    # f = 0.001,
    # residual = residual,
    # g = g,
    # ratio = 100,
    # lenη = (000.0, 000.0, 24 * 60 * 60. * 10),
    # maxit = 100000,
    # tol = 1e-6,
);
        if i == 1
    println("$i er mois claculé")
        else
    println("$i eme mois claculé")
        end
    end

```

Figure 9. DIVAnd run

3.5. Using other data sources

Using other data sources is still not available on this VLab. On further updates, the code will be automated to consider any region as soon as the user has at least Drifter's and altimetry data. HF radar won't be mandatory but recommended.

3.6. Authors

Université de Liège: Abel Dechenne, Alexander Barth

3.7. References

For technical aspect of DIVAnd technique, please refer to: Barth, A., Troupin, C., Emma, R., Alvera Azcarate, A., Beckers, J.-M., & Joaquín, T. (2021). *Variational interpolation of high-frequency radar surface currents using DIVAnd*.

4. Introduction to VLab 3 "Carbon-Plankton Dynamics"

This VLab provides a workflow to run **mechanistic models** of the NPZD-type (Nutrient-Phytoplankton-Zooplankton-Detritus), using near real-time data to quantify the relative contributions of the bottom-up and top-down drivers in phytoplankton dynamics. The NPZD models are commonly used, and describe four ecosystem components, nutrients, phytoplankton, zooplankton, and detritus with varying levels of detail. Two models are implemented: (1) a NPZD model with fixed elemental ratios, adjusted from the NPZD model of *Soetaert and Herman (2009)* and *Otero et al. (2023),* hereafter termed basic NPZD model, and (2) a quota-type of NPZD model, based on the work by *Geider et al. (1998)*, extended with different organic matter pools after *Schartau et al. (2007)*. Hereafter termed extended NPZD model. Both models resolve parts of the carbon and nitrogen cycle. The basic model is used to demonstrate a workflow for assessing model uncertainty. The extended model is conceived as a demo version to look at parameter sensitivity in an interactive way.

In the basic model phytoplankton dynamics are simulated based on information from nutrient concentrations and zooplankton density. Based on these simulations using near real-time data, it is possible to calculate and visualize the relative contribution of each bottom-up or top-down driver, i.e. (1) nutrients, (2) Sea Surface Temperature (SST), (3) photosynthetically active radiation (PAR) and (4) zooplankton grazing, over time (Figure 10). Carbon dynamics are simulated based on the marine biological carbon pump (Figure 10). The carbon pump in marine ecosystems operates as a key mechanism for carbon dynamics. During photosynthesis, the phytoplankton captures carbon from the atmosphere, converting it into organic matter in the pelagic food web. This carbon is then transported downward as plankton and other organic material die and sink, forming detritus. As the detritus descends through the water column, it undergoes decomposition and remineralization, releasing carbon back into the water as Dissolved Inorganic Carbon (DIC). This process not only regulates the distribution of carbon in the ocean but also plays a critical role in the global carbon cycle.

Funded by the European Union
Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

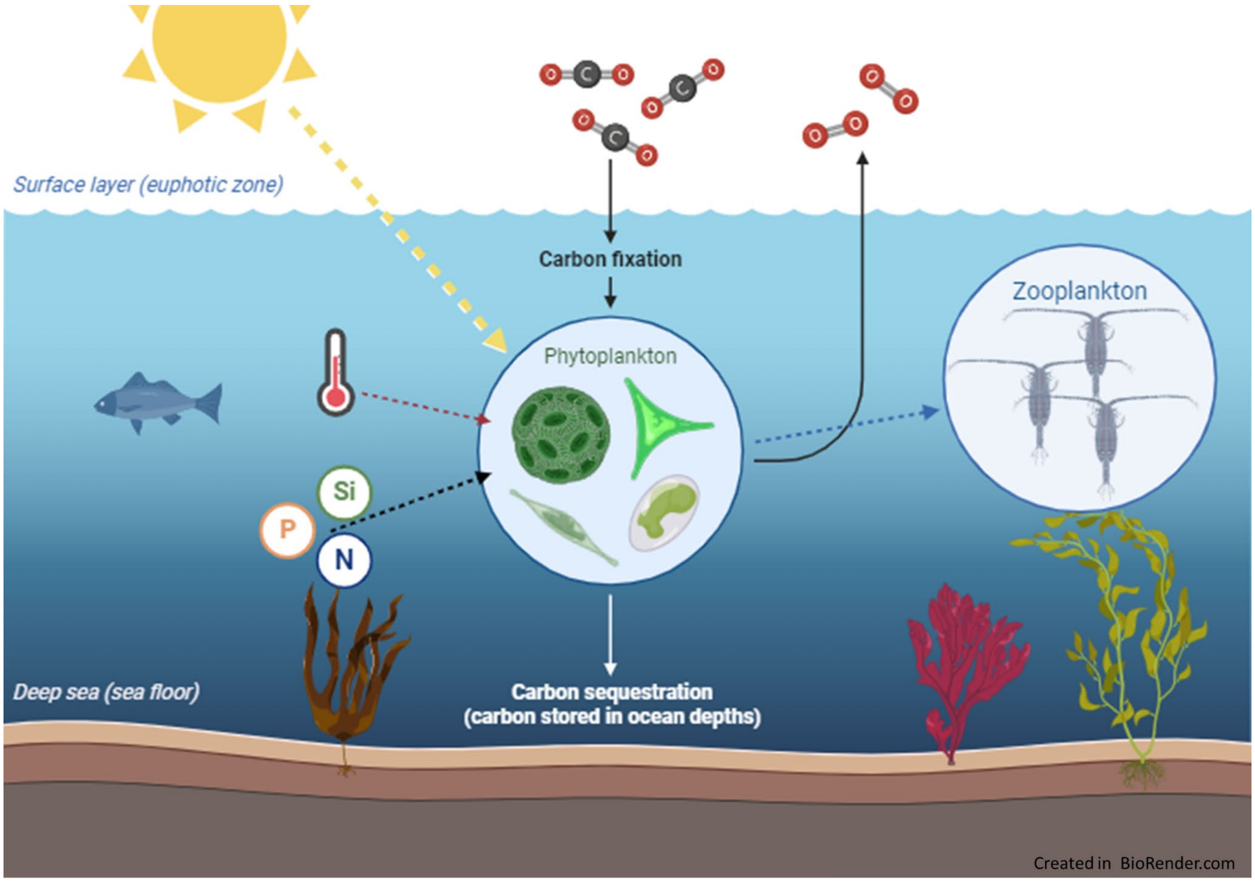


Figure 10. Visualization of the Nutrient-Phytoplankton-Zooplankton-Detritus (NPZD) model.

The validation of the model is performed by comparing the model predictions, e.g. phyto- and zooplankton biomass, with field observations. The Root Mean Square Error (RMSE) is calculated between prediction values and observational values. By doing so and by running the model for multiple parameterizations, we can select the best 10% simulations (lowest RMSE) to predict phyto- and zooplankton, and carbon dynamics and define confidence intervals around the model predictions. To estimate the relative contributions of the drivers, we use these best 10% simulations or select the best 5% simulations (based on the RMSE) to decrease computational effort.

The extended NPZD model resolves a water column and a rudimentary bottom in one vertical dimension (Figure 11). Two nutrient compartments exist for (1) reduced inorganic nitrogen (state variable: NH4) and (2) oxized inorganic nitrogen (state variable: NO3).

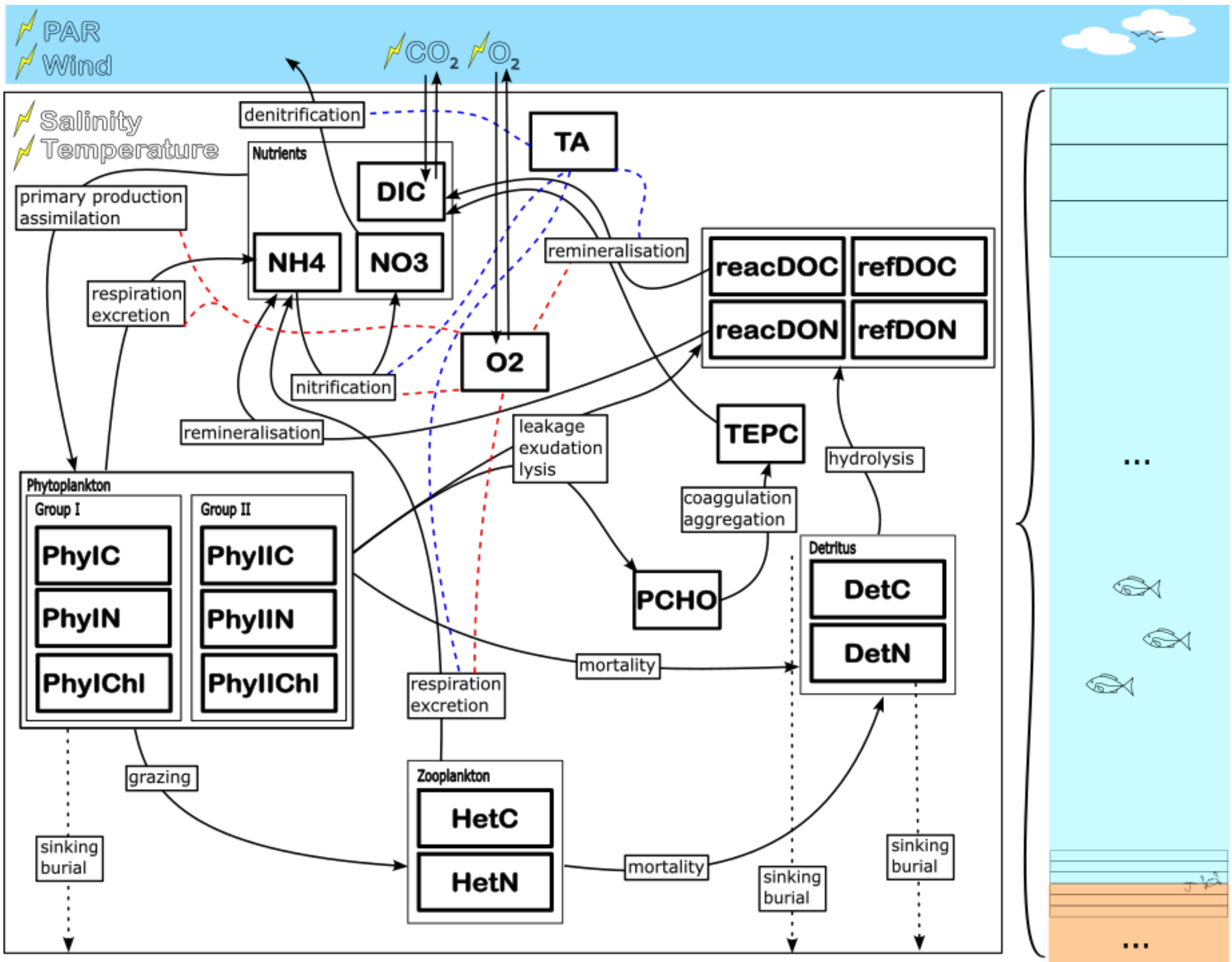


Figure 11. Graphical representation of the structure of the extended NPZD model.

Two phytoplankton groups take up nitrogen from these pools with variable C/N ratios. Both phytoplankton groups are grazed by a zooplankton compartment. In contrast to the phytoplankton, the zooplankton has a fixed C/N ratio. Phytoplankton also produces dissolved organic matter through leakage and active exudation. The dissolved organic carbon is in the model split up into a carbohydrate and a non-carbohydrate fraction (following *Schartau et al. 2007*), where the carbohydrate fraction is involved in the formation of marine snow (transparent exopolymeric particles, or TEP). This sinks to the bottom and is continuously remineralized. Phytoplankton and zooplankton that die enters a particulate detritus pool. This detritus nitrogen and carbon sinks to the bottom and is continuously broken down to form dissolved organic matter. Remineralization of this dissolved organic material occurs through oxic remineralization, anoxic remineralization, and denitrification. Since the presence of oxygen is a regulatory factor in these remineralization processes, it is also explicitly modelled as a state variable. The relationships between oxygen production and carbon fixation, and between oxygen consumption and all respiratory and remineralization processes are implemented with fixed elemental ratios. Oxygen and CO_2 are taken up by the sea from the atmosphere. Atmosphere-ocean gas exchange is modelled with gas transfer functions based on the work by Wanninkhof (*Wanninkhof 2014*). CO_2 that enters the model through the ocean surface enters a DIC pool and will modify the acid-base equilibrium. To calculate the acid-base equilibrium total alkalinity (TA) and DIC are both included as state variables. The pH at the surface is implemented as a derived variable. A full description of the model is available in the R-package that contains the model and is accessible from the R-Studio server (see below).

The model is implemented as a Fortran program in an add-on package for the R Statistical Software (R Core Team, 2024). The Fortran program is called from an R function that accepts a number input datasets. These are in detail described in the help files and vignettes of the package. A web interface is callable from within the R-package (cf. help files). This web application allows for 'on-the-fly' interaction with the model (Figure 12). A choice can be made for a simple 0D box model version or a 1D vertically resolved model version (as described above). Model output is plotted against a background of data from the Dutch national monitoring program of Rijkswaterstaat, for which the data can also be obtained through **their web services**. A choice can be made from a selection of stations. Model simulations always start from 2001 and run for a number of years, selectable in the web application.

4.1 Target Users & Community

The target users of this VLab are researchers who are at the forefront of coastal ecosystem studies. The models are particularly well-suited for researchers that want to extend their activities into the field of ecosystem modeling or for educational purposes (i.e. to educate students interested in marine ecology). In such contexts, the models can be used either as a starting point for model development or as a demo on how to work with ecosystem models of low to moderate complexity.

In addition, the models present a view on the role of various biological compartments in carbon dynamics of coastal ecosystems. Once calibrated and validated, the models can simulate different scenarios and assess the effect on the carbon system and the marine ecosystem. In this context, they may be useful as analysis tools for researchers to reflect on processes that they deem responsible for patterns in their own data. The models could be useful as components of a Digital Twin for the Ocean.

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

4.2 State of art

Achieving a comprehensive characterization of phytoplankton, the basis of the marine food webs, requires multidisciplinary data. These use cases demonstrate how to link and integrate different data types (biology, biogeochemistry, and physics), available from several Blue-Cloud data infrastructures, and made interoperable through a Blue-Cloud Vlab, to run a mechanistic model identifying the drivers of phytoplankton abundance, and a machine-learning algorithm to calculate carbon sequestration.

Data on long-term zooplankton and phytoplankton observations, accessible from EMODnet Biology through the Discovery and Data Access and Service (DDA&S) were used in combination with carbon data from ICOS and SOCAT, and environmental data such as sea surface temperature and nutrient regimes from EMODnet Chemistry and other available data resources. Such heterogeneous data, in their nature and acquisition mode, cannot be combined and exploited in a straightforward manner. The data was used to detect and explain anomalies in long-term trends and quantify the relative contribution of the environmental drivers for these plankton essential ocean variables (EOVs). Coupling these systems allows testing the operational situation and assessing gap-filling or potential capabilities. The models could be used in the future to simulate different scenarios and assess the effect on the carbonate system and the marine ecosystem.

4.3 Input data sources

VARIABLES	DATA SOURCES	DATA ACCESS
Zooplankton abundances	EMODnet Biology	Blue Cloud (VRE or DD&AS)
Phytoplankton proxy (Chla)	EMODnet Biology	Blue Cloud (VRE or DD&AS)
Abiotic data (nutrients, PAR, temperature, salinity, pH and carbon)	EMODnet chemistry SOCAT ICOS	Blue Cloud (VRE or DD&AS)
Biogeochemical data for the extended NPZD model	Rijkswaterstaat, Waterinfo	https://waterinfo.rws.nl Subset for the model demo webapp available in the NorthSeaCarbon R-package
PAR and wind data for the extended NPZD model	Royal Netherlands Meteorological Institute (KNMI)	https://www.knmi.nl/nederland-nu/klimatologie/gemeten-reeksen

VARIABLES	DATA SOURCES	DATA ACCESS
		Subset for the model demo webapp available in the NorthSeaCarbon R-package

Table 3. VLab 3 Input data sources

4.4 Step by step guideline to use the VLab

4.4.1 The basic NPZD model

Before starting the model, there are requirements that need to be fulfilled. First of all, input data, such as nutrients, temperature, salinity, pH, wind and carbon data, are needed covering at least three years. The three-year period is needed to obtain daily data using generalized additive modelling. The daily data is used to run the model on daily time steps. Further, you need the Jupyter Notebook and accompanying R scripts. The Jupyter Notebook and R scripts are provided for you. Additionally, input data for the Belgian part of the North Sea and Gulf of Trieste are provided as well as demo. It is advised to have a basic understanding of R to run the NPZD model in R.

The workflow to use the basic NPZD model is provided in a Jupyter Notebook and contains the required information to run the model. When loading a Jupyter Notebook in the Virtual Research Environment (VRE) with the required R packages pre-installed (Blue-Cloud Carbon-Plankton Server - 8 Cores / 64G RAM), this document will be available in your *Home* folder. This Jupyter Notebook document contains all necessary information and code to (1) (re)calibrate the NPZD model, (2) simulate phyto- and zooplankton dynamics, (3) validate modeling results with observational data, (4) calculate the relative contribution of the bottom-up and top-down drivers on phytoplankton dynamics and (5) visualize the modeling results. The NPZD model runs in terms of nitrogen, as well as in terms of carbon.

Steps to run the Jupyter Notebook:

1. Open Jupyter Notebook from your CarbonPlanktonDynamics VRE homepage (Figure 12) and select the Blue-Cloud Carbon-Plankton Server - 8 Cores / 64G RAM. This server has the required R packages pre-installed.

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Figure 12. Carbon Plankton Dynamics VRE Homepage

2. This will open a new tab in your browser where *JupyterLab* is launched (Figure 13). In the launcher, you now see two folders and a Jupyter Notebook in the column on the left of the screen. Select the *Jupyter Notebook* 'NPZD.ipynb'. This is the only document that you need to run the NPZD model with examples of the Belgian part of the North Sea (BPNS) or the northern Adriatic Sea. The *NPZD folder* contains the necessary files to run the NPZD model, such as .csv files and R scripts. The *workspace folder* is your personal folder where you can store your results. At the end of the Jupyter Notebook is a section where you can copy your results to your personal workspace.

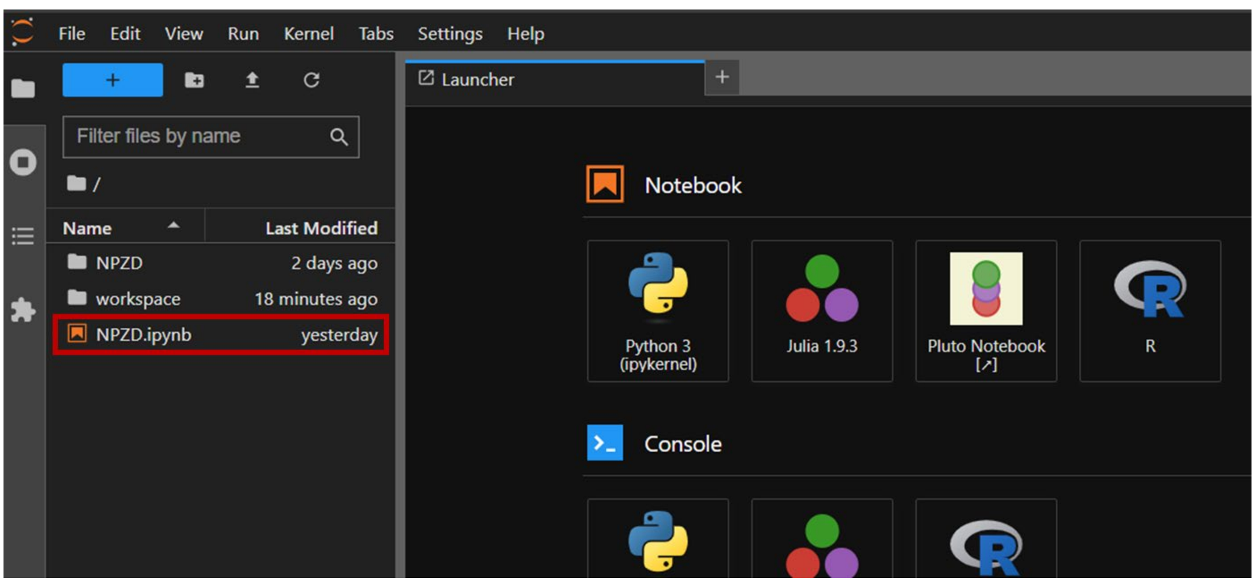


Figure 13. Jyputer notebook 'NPZD.ipynb'

3. Now, you can go through the Jupyter Notebook and run the model step by step following the instructions in the document (Figure 14).

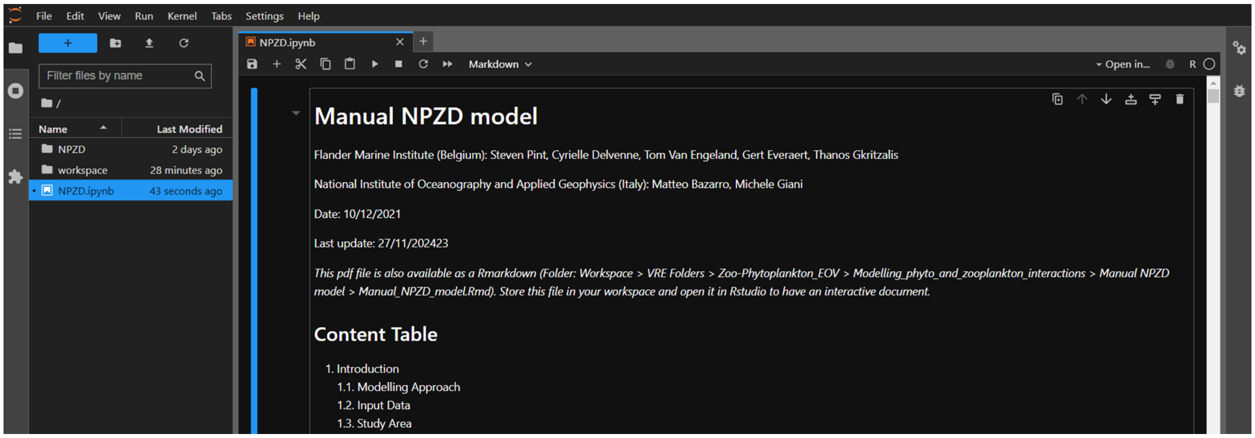


Figure 14. Model manual

Below we provide Figure 15 showing a potential output of the model for the nearshore region of the BPNS, where the relative contributions of each of the drivers in phytoplankton abundances is illustrated.

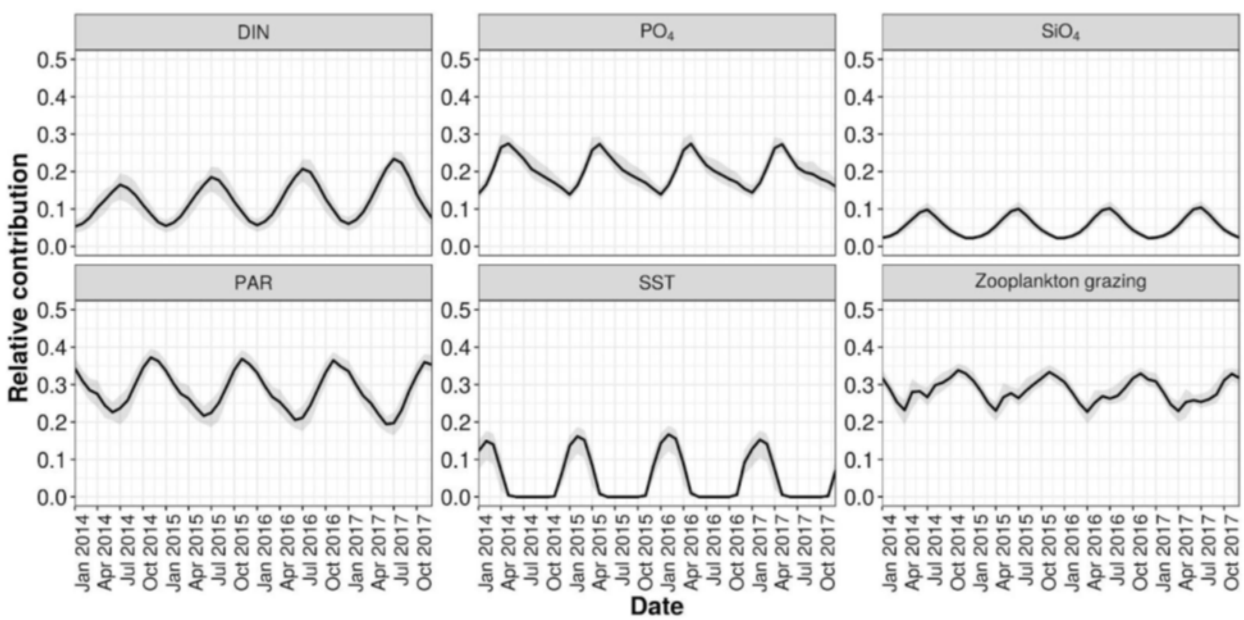


Figure 15. Average monthly relative contributions for each limitation factor in the growth of phytoplankton for the nearshore region.

4.4.2 The extended NPZD model

The extended NPZD model is implemented as an add-on package, called NorthSeaCarbon, for the R Statistical Software. When a R-Studio server session is started up, the package is available in the package library and can be loaded from the R console in the usual way by typing ‘require(NorthSeaCarbon)`. To call the help-file of the model, type “?NSC1D” (the name of the function that accepts all inputs to the model and calls the appropriate model integration routines). A general introductory help page can be loaded by typing “?`NorthSeaCarbon-package`”. The latter is a good starting point to explore the package. To start up the demo web application, type “PlayModel()” on the R console. A browser window will be opened (Figure 16).

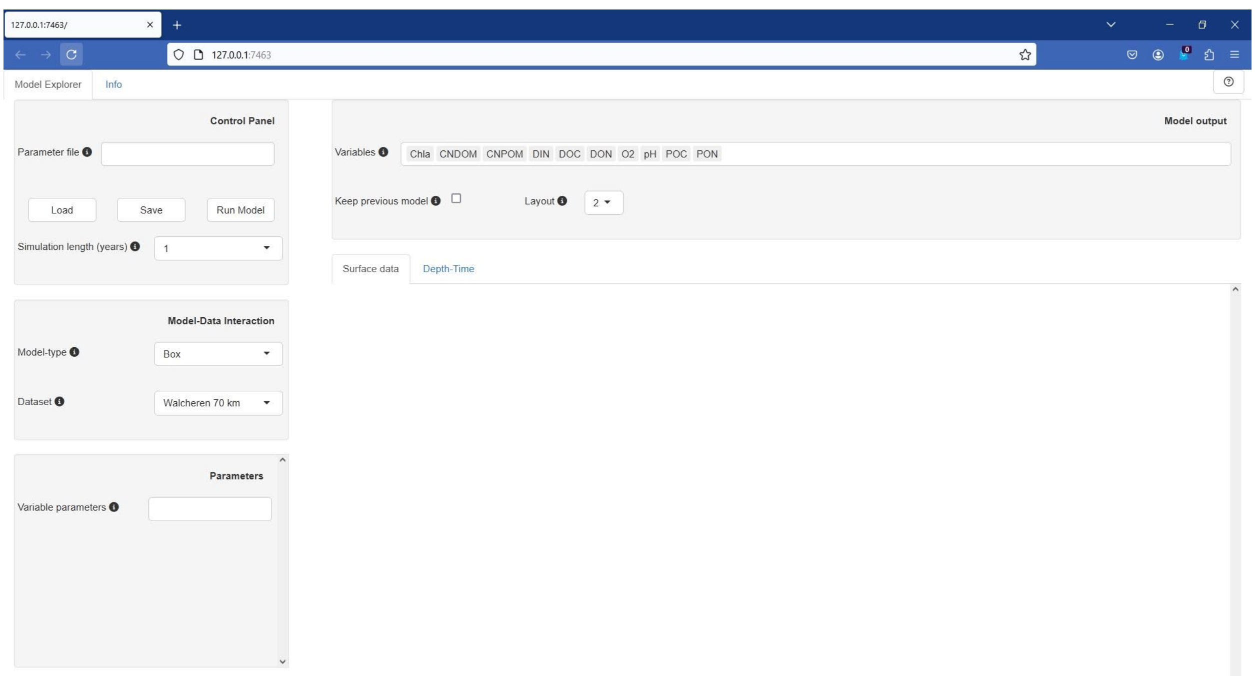


Figure 16. The initial browser window after starting up the web application with the model demo.

The application has two tabs. The ‘Model Explorer’ tab contains the actual graphical interface to control the model. The ‘Info’ tab gives a full description of the model. The content of this tab is also available as a package vignette in the NorthSeaCarbon package, which is called from the R-console by typing “`vignette(“NSCmodel”)`”. In the upper right-hand corner a button with a question mark gives access to a help page explaining how to use the interface.

4.5 Using other data sources

Some requirements as discussed in 4.4.1 need to be fulfilled to run the NPZD models with other data. For the basic NPZD model, the time series of the input data should cover at least three years. These three years are needed to create daily data using interpolation methods such as Generalized Additive Models (GAM) or Generalized Linear Models (GLM) to obtain a complete set of input data.There are scripts available that can help prepare daily input data for other regions

To use the model with data from regions other than the Belgian or Adriatic, it is recommended to obtain the necessary data from EMODnet, as the model is designed to work with this data source. Using data from other regions or sources may lead to errors, primarily due to differences in column names. To minimize potential issues, refer to the input data provided for the Adriatic region as a guide when working with EMODnet data from other regions.

For the extended NPZD model more elaborate preparations are required before the model can be run with other data. It is important to note that the web interface only works with the built-in data structures, and only allows for selecting between different built-in datasets and for changing parameters. If the model is required to run with different input data, this is only possible with scripts in the BlueCloud R-Studio Server). Broadly speaking, forcing functions should be supplied as 2-column matrices which enter the NSC-function as function parameters. Model parameters should be supplied as a named list or vector. Default model grids are available in the package but can be altered as function parameters to the NSC-function as well. We refer to the help files (“``?`NorthSeaCarbon-package` ``” and references therein) for complete definitions of the requirements for the input data structures.

4.6 Authors

Flander Marine Institute: Steven Pint, Cyrielle Delvenne, Tom Van Engeland, Gert Everaert, Thanos Gkritzalis

National Institute of Oceanography and Applied Geophysics: Matteo Bazarro, Michele Giani

4.7 References

Geider, R.J., MacIntyre H.L., Kana T.D. (1998). A dynamic regulatory model of phytoplankton acclimation to light, nutrients, and temperature. Limnology and Oceanography 43(4): 679-694

Otero, V.; Pint, S.; Deneudt, K.; De Rijcke, M.; Mortelmans, J.; Schepers, L.; Martin-Cabrera, P.; Sabbe, K.; Vyverman,W.; Vandegehuchte, M.; et al. (2023) Pronounced Seasonal and Spatial Variability in Determinants of Phytoplankton Biomass Dynamics along a Near–Offshore Gradient in the Southern North Sea. J. Mar. Sci. Eng. 11, 1510. https://doi.org/10.3390/jmse11081510

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Soetaert, K. and Herman, P.M.J. (2009). A practical guide to ecological modelling. Using R as a Simulation Platform. Springer-Verlag, New York, US, p. 372.

Schartau, M., Engel, A., Schröter, J., Thoms, S., Völker, C., and Wolf-Gladrow, D. (2007) Modelling carbon overconsumption and the formation of extracellular particulate organic carbon. Biogeosciences, 4, 433–454, https://doi.org/10.5194/bg-4-433-2007.

Wanninkhof, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited. *Limnology and Oceanography: Methods*, *12*(6), 351-362.

5. Introduction to VLab 4 "Marine Environmental Indicators"

The Marine Environmental Indicators (MEI) VLab enables users to monitor and evaluate the environmental conditions of marine areas, providing essential support for decision-making in ocean management. By integrating multiple data sources, the platform offers a unified data analysis service, facilitating online computation of indicators through Jupyter Notebooks or a dedicated web application. This application leverages the VRE Analytics Engine services to generate indicator outputs. This VLab 4 users handbook will be divided into three sections representing each indicator (OHC, TRIX, MHW). The indicator SSI V2 will be described in the next version of this handbook. A **video tutorial** on how to access the VLab and its indicators can be found on the **Blue-Cloud Youtube Channel**.

5.1 Target Users & Community

VLab 4 is designed for a wide range of users involved in marine research and environmental management. Key user groups include:

- **Environmental Scientists & Researchers**: Those focusing on marine ecosystems, oceanography, environmental and climate monitoring.
- **Policy Makers & Marine Resource Managers**: Users seeking insights for informed decision-making in marine and coastal resource management.
- **NGOs & Environmental Advocacy Groups**: Organizations working on marine conservation and sustainability initiatives.
- **Academic Institutions & Universities**: Research groups focusing on marine science and environmental protection.
- **General Public & Educational Institutions**: Those interested in understanding marine environmental issues through accessible data.

5.2 Input data sources

INDICATOR	VARIABLES	DATA SOURCES	DATA ACCESS
Ocean Heat Content (OHC)	In situ Temperature (°C)	SeaDataNet World Ocean Database CMEMS CORA EuroArgo Blue Cloud EOV dataset (WP3 output) Currently from SeaDataCloud Mediterranean Sea - Ocean Heat Content 1955-2018 https://doi.org/10.12770/504EA4EE-ABEE-4EBF-AB89-C4200E1CDAD4	Data is not accessible via the Blue Cloud Data Access service. It has been uploaded to Vlab.

Trophic Index (TRIX)	-DIN, water body dissolved inorganic nitrogen. -CHL, water body chlorophyll-a. -TP, water body total phosphorus. -DO, water body dissolved oxygen saturation.	DIVAnd interpolated climatologies* calculated for the stated parameters (Chl-a, DIN, TP & DO%).	Data is not accessible via the Blue Cloud Data Access service. It has been uploaded to Vlab.
Marine Heat Wave (MHW)	Sea Surface Temperature (°C)	CMEMS Mediterranean Model Data and Satellite Observation	BlueCloud VRE, CMEMS

Table 4. VLab 4 Input data sources

5.3. General instructions for using the Jupyter Notebooks in the Vlab

Jupyter Notebooks are available for generating the different environmental indicators. Each notebook is self-explanatory and guides the user through its execution and customization to meet specific needs. Users can modify parameters such as the time range or the geographical region to be analyzed.

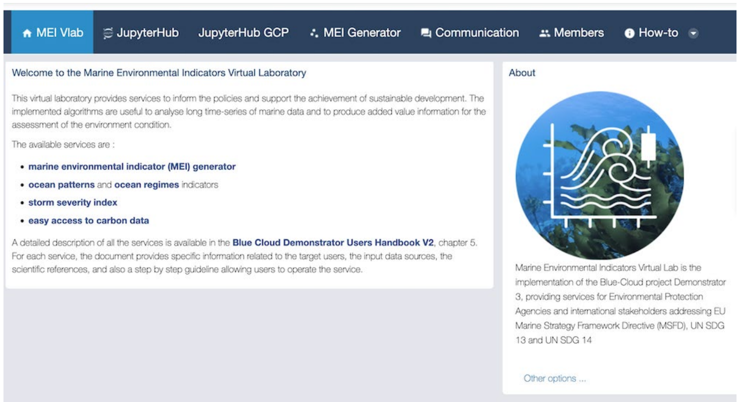
To access one of these notebooks, follow the instructions below.

These steps are tailored for the MHW notebook but can be easily adapted to others using the information provided in the table below:

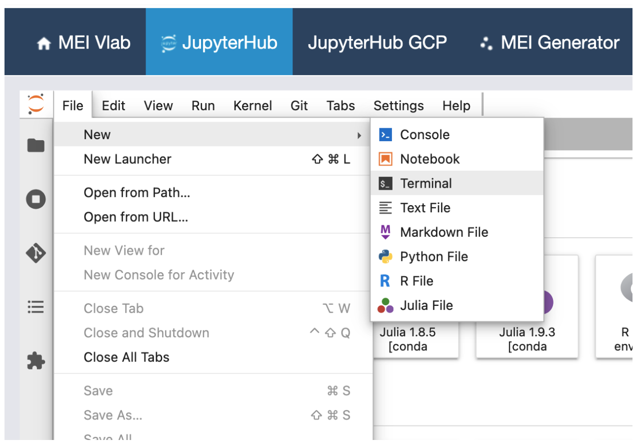
1. Open the vlab at the address:

   https://blue-cloud.d4science.org/group/marineenvironmentalindicators

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227



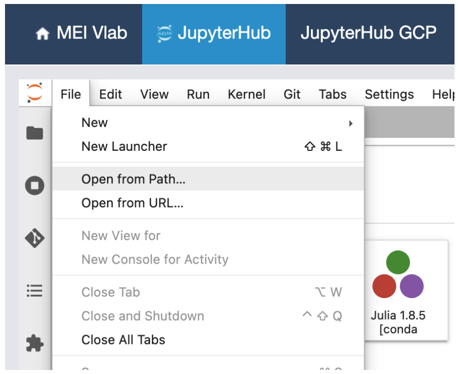
2. Open a **JupyterLab** instance by selecting the **JupyterHub tab** inside the VRE VLab.
3. Open a terminal in JupyterHub by navigating to **File > New > Terminal**.



4. In the terminal, **copy the archive** in your own space on JupyterHub and **unzip it** using the following commands:

```
cp /workspace/VREFolders/MarineEnvironmentalIndicators/notebooks/MarineHeatWaves/MHW.zip .
unzip MHW.zip
```

5. Open the desired notebook by selecting **Open from Path** in the menu and entering the name of the folder where the zip archive was unzipped, as specified in the table below.



Main path for compressed notebook archives in the workspace:
/workspace/VREFolders/MarineEnvironmentalIndicators/notebooks/

INDICATOR	Subpath in workspace	Archive name	Folder in JupyterLab	Notebook filenames
Ocean Heat Content (OHC)	OceanHeatContent	OHC.zip	OHC	OHC.ipynb
Trophic Index (TRIX)	EutropicathionIndicator	TRIX.zip	TRIX	TRIX.ipynb
Marine Heat Waves (MHW)	MarineHeatWaves	MHW.zip	MHW	MHW-Timeseries.ipynb, MHW-Maps.ipynb

Table 5. VLab 4 notebook archives in the workspace

The output produced by the notebooks is the user's personal workspace in: **/workspace/MEI**

5.4. Marine Heat Waves (MHWs)

5.4.1. State of art

Marine Heatwaves (MHWs) are prolonged periods of anomalously high sea surface temperatures (SSTs) that significantly impact marine ecosystems, biodiversity, and socio-economic activities.

Over recent decades, global ocean warming, particularly in the Mediterranean Sea, has led to increased intensity, frequency, and duration of these events. MHWs are defined as periods when SSTs exceed the

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

90th percentile of a baseline period for at least five consecutive days. This threshold-based approach considers both the duration and intensity of warming, making it a widely accepted standard for identifying MHWs (Hobday et al., 2016). The Mediterranean Sea is a hotspot for MHWs due to its semi-enclosed nature, high regional SST variability, and sensitivity to atmospheric and oceanic changes.

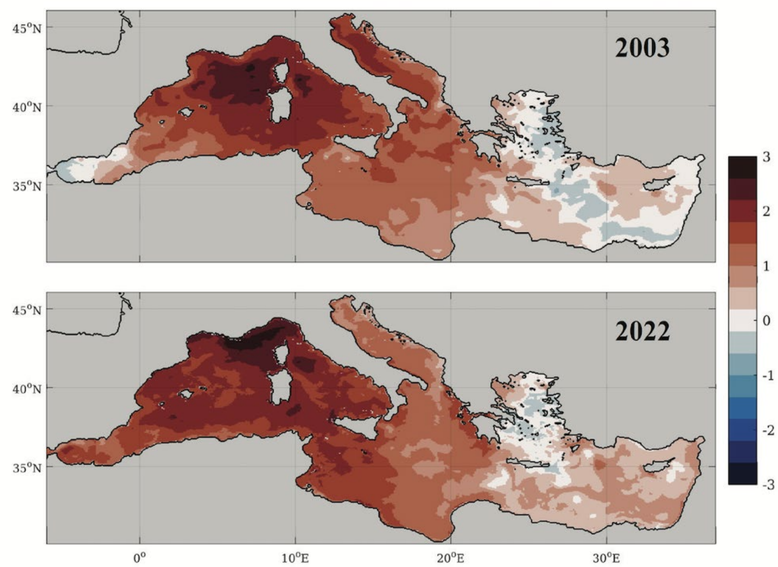


Figure 17. Comparison of summer SST anomalies during June - August of 2003 (top) and 2022 (bottom) MHWs in the Mediterranean Sea, relative to the corresponding 1993-2022 period.

The Mediterranean Sea has shown an accelerated warming trend, with SST increases ranging from 0.034 to 0.048 °C/year, surpassing the global average (Darmaraki et al., 2024). MHWs have been reported across various Mediterranean sub-basins, with notable events in 2003, 2012, 2017, and 2022. These events were characterized by SST anomalies exceeding 4°C and durations extending up to several months. For instance, the 2003 event affected 46-70% of the Mediterranean, with SST anomalies reaching up to 7°C, extensively studied for its ecological and economic impacts (Darmaraki et al., 2024 - Fig.17). In contrast, the 2022 MHW set records for its duration and geographic spread, covering up to 70% of the basin (McAdam et al. - Fig. 18).

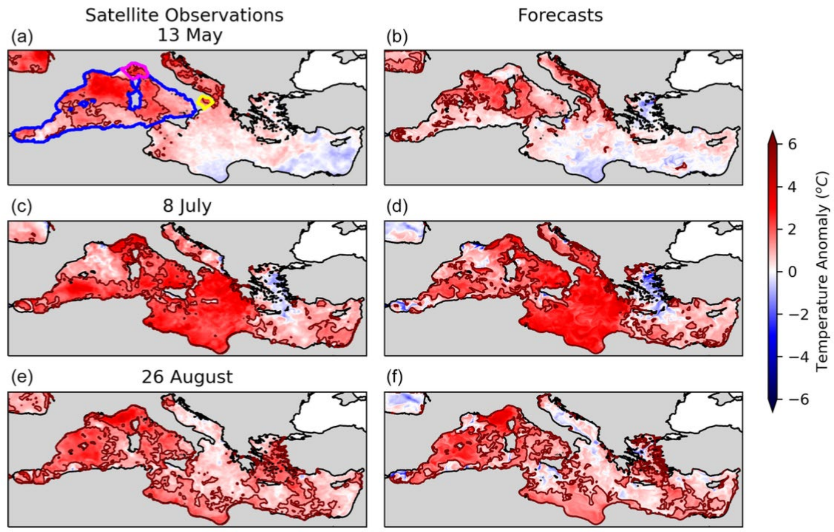


Figure 18. Snapshots of SST anomalies and MHW occurrence during the different stages of the 2022 MHW. (a, c, e) - Reprocessed satellite observations. (b, d, f) - Forecasts with a lead time of 4 d.

The occurrence of MHWs results from complex interactions between atmospheric and oceanic processes. Key drivers include atmospheric forcing such as increased solar radiation, reduced cloud cover, and persistent anticyclonic conditions. Ocean dynamics also play a critical role, with reduced vertical mixing, strong stratification, and advection of warm water masses amplifying their effects. Additionally, large-scale climate modes such as the East Atlantic Pattern and North Atlantic Oscillation significantly modulate the characteristics of these events (Darmaraki et al., 2024).

MHWs have profound impacts on both ecological and socio-economic systems. Ecologically, they cause mass mortality of marine species, disrupt ecosystems, and drive shifts in species distributions (Darmaraki et al., 2024). Economically, they result in substantial losses for fisheries and aquaculture, while also affecting industries reliant on marine biodiversity. These impacts underscore the need for robust monitoring and forecasting systems. Recent advancements, such as the Copernicus Marine Service, have significantly enhanced our ability to detect and predict MHWs. Short-term forecasting systems, including the Mediterranean Forecasting System (MedFS), have demonstrated reliability in predicting the onset, intensity, and duration of these events, aiding in mitigation efforts (McAdam et al).

Despite significant progress, several challenges remain. Variations in MHW definitions and detection methodologies hinder cross-regional comparisons. Additionally, the limited understanding of subsurface MHWs and their vertical extent presents a research gap. Uncertainty in predicting long-term trends under

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

various emission scenarios and inadequate data on socio-economic impacts also complicate adaptive measures.

To address these gaps, future research should prioritize enhanced monitoring through the integration of satellite observations with in situ measurements. Developing coupled ocean-atmosphere models will better simulate MHW dynamics. Quantifying socio-economic impacts and identifying vulnerable regions will aid in crafting targeted mitigation strategies. Furthermore, leveraging scientific insights to inform climate adaptation policies will enhance resilience against these extreme events.

In conclusion, MHWs serve as a critical indicator of ocean health and the broader impacts of climate change. Advancements in detection, monitoring, and forecasting are essential to mitigate their adverse effects and ensure sustainable management of marine resources. The Mediterranean Sea, as a prominent MHW hotspot, offers valuable insights into the evolving dynamics of these phenomena and their far-reaching consequences.

5.4.2. Input data sources

VARIABLES	DATA SOURCES	DATA ACCESS
Sea Surface Temperature (℃)	CMEMS Mediterranean Sea High Resolution and Ultra High Resolution Sea Surface Temperature Analysis https://doi.org/10.48670/moi-00172 CMEMS Mediterranean Sea - High Resolution L4 Sea Surface Temperature Reprocessed https://doi.org/10.48670/moi-00173 CMEMS Mediterranean Sea Physics Analysis and Forecast https://doi.org/10.25423/CMCC/MEDSEA_ANALYSISFORECAST_PHY_006_013_EAS8	BlueCloud VRE, CMEMS

Table 6. VLab 4 MHW Input data sources

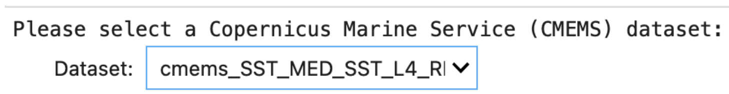
5.4.3. Step by step guideline to use the VLab

The notebook MHW-Timeseries purpose is detecting **Marine Heatwaves (MHWs)** using time series data from **Sea Surface Temperature (SST)** observations and climatological datasets.

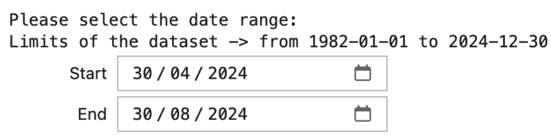
The following steps will occur, executing the notebook's cells:

1. **Initial Cells**

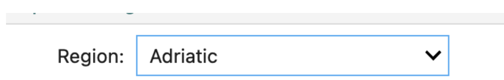
   After loading the needed Python packages, the user can select the input dataset of interest:



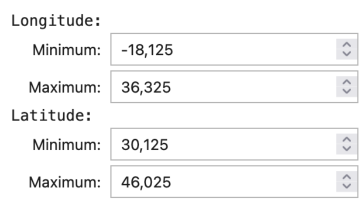
   Then the user can choose the time range to configure the analysis. The user can select the start date and end date.



   The user can select a predefined sub-region of the Mediterranean Sea:



   Or enter the longitude and latitude coordinates of a custom sub-region:



Using datasets from **Mediterranean Forecasting System** (**MFS**), Satellite-based SST observations and Climatology for Historical SST averages and percentiles (1987–2021), the notebook loads files dynamically based on selected time periods and regions.

2. **Time Series Extraction**: to extract area-averaged time series of SST values. Processes datasets with spatial masks for specific regions and calculates metrics like 90th percentile thresholds.
3. **Marine Heatwave Detection**: Uses the wavesnspikes package to detect MHWs by identifying SST values exceeding the 90th percentile of climatology. And outputs metrics for observed and forecasted conditions.
4. **Visualization**: plots SST time series and highlights marine heatwave events with comparisons to climatology. Save figures to the specified directory.

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Figure 19 is the example result for the selected time period and region:

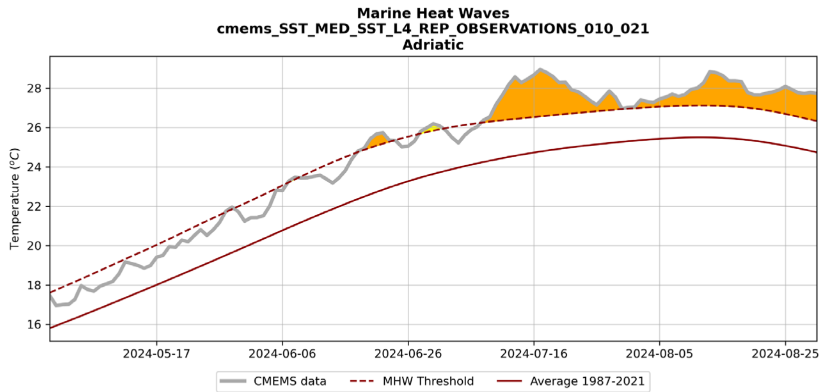


Figure 19. Example result for the selected time period and region

The other notebook MHW-Maps focuses on generating spatial maps for marine heatwave (MHW) analysis using Sea Surface Temperature (SST) datasets and climatological baselines.

The following steps will occur, executing the notebook's cells:

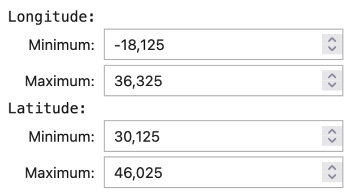
1. **Dataset selection:** after loading the needed Python packages, the user can select the input dataset of interest:

```
Please select a Copernicus Marine Service (CMEMS) dataset:
Dataset: cmems_SST_MED_SST_L4_RI
```

2. **Date Selection**: Allows users to specify a target date for analysis.

```
Please select the date of interest:
Dataset limits -> from 1982-01-12 to 2024-12-30
Date 30 / 12 / 2024
```

3. **Region of interest:** the user can select a region of interest introducing the longitude and latitude coordinates:



4. **Intensity Calculation**: for Temperature anomaly, difference between observed SST and climatological mean SST.
5. **Data Loading**: loads datasets dynamically based on the selected date and type: Satellite Observations (real-time or historical data) and Climatology
6. **Visualization**: Generates spatial plots to visualize temperature anomalies.

Figure 20 shows the example result for the selected time period:

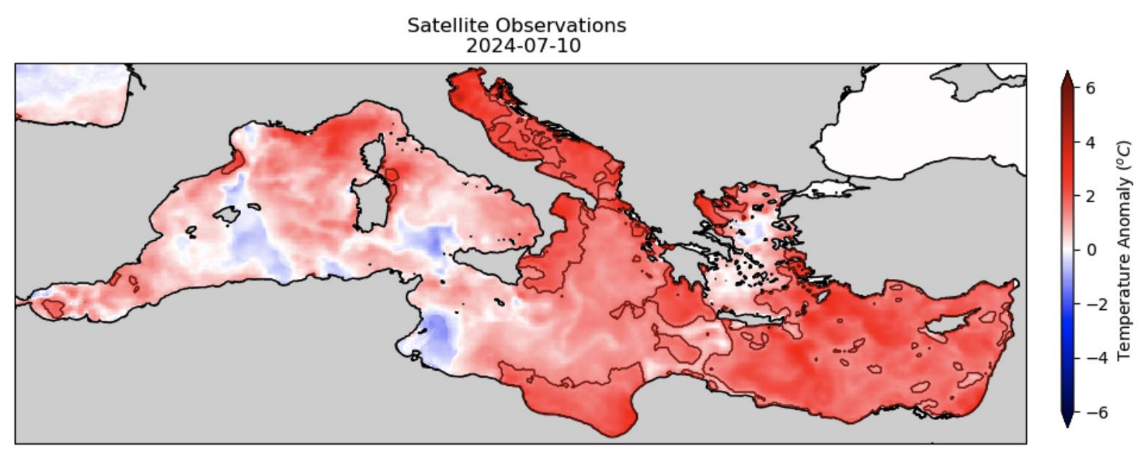


Figure 20. Example result for the selected time period

5.4.4. Authors

CMCC: Francesco Palermo, Giovanni Coppini, Emanuela Clementi, Ronan J. McAdam, Megi Hoxhaj, Fabrizio Antonio, Rafael Gomes de Menezes

5.4.5. References

Coppini, G., Marra, P., Lecci, R., Pinardi, N., Cretì, S., Scalas, M., Tedesco, L., D'Anca, A., Fazioli, L., Olita, A., Turrisi, G., Palazzo, C., Aloisio, G., Fiore, S., Bonaduce, A., Kumkar, Y. V., Ciliberti, S. A., Federico, I., Mannarini, G., . . . Negro, G. (2017). SeaConditions: a web and mobile service for safer professional and recreational activities in the Mediterranean Sea. *Natural Hazards and Earth System Sciences*, *17*(4), 533–547. https://doi.org/10.5194/nhess-17-533-2017

D'Anca, A., Conte, L., Nassisi, P., Palazzo, C., Lecci, R., Cretì, S., Mancini, M., Nuzzo, A., Mirto, M., Mannarini, G., Coppini, G., Fiore, S., & Aloisio, G. (2017). A multi-service data management platform for scientific oceanographic products. *Natural Hazards and Earth System Sciences*, *17*(2), 171–184. https://doi.org/10.5194/nhess-17-171-2017

Darmaraki, S., Denaxa, D., Theodorou, I., Livanou, E., Rigatou, D., E, D. R., Stavrakidis-Zachou, O., Dimarchopoulou, D., Bonino, G., McADAM, R., Organelli, E., Pitsouni, A., & Parasyris, A. (2024). Marine Heatwaves in the Mediterranean Sea: A literature review. *Mediterranean Marine Science*, *25*(3), 586–620. https://doi.org/10.12681/mms.38392

Drago, F., Cabrera, P., Irisson, J., Bittner, L., Schickele, A., Drudi, M., Balem, K., Noteboom, J. W., Castaño-Primo, R., Jones, S., Taconet, M., Ellenbroek, A., Vallejo, B. R., Haberle, I., Hackenberger, D. K., Djerdj, T., Hackenberger, B. K., Ćaleta, B., Purgar, M., . . . Zavala-Romero, O. (2023). Blue-Cloud Virtual Labs in support of Sustainable Development Goals. In *Zenodo (CERN European Organization for Nuclear Research)*. https://zenodo.org/records/7663960

Drudi, M., Palermo, F., Mariani, A., Lecci, R., Andrea, G. J., Balem, K., Maze, G., Bachelot, L., Noteboom, J. W., Pfeil, B., Castaño-Primo, R., Paul, J., Dussurget, R., & Arnaud, A. (2022). Test the Blue-Cloud Virtual Labs: Marine Environmental Indicators. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.6628701

McAdam, R., Bonino, G., Clementi, E., & Masina, S. (2024). Forecasting the Mediterranean Sea marine heatwave of summer 2022. *State of the Planet*, *4-osr8*, 1–10. https://doi.org/10.5194/sp-4-osr8-13-2024Ocean Decade. (2024, September 16). *10 Challenges - Ocean Decade*. https://www.oceandecade.org/challenges/

Simoncelli, S., Oliveri, P., & Mattia, G. (2020). SeaDataCloud Mediterranean Sea - Ocean Heat Content 1955-2018 (Version 1.1) [Data set]. IFREMER / IDM/SISMER. https://doi.org/10.12770/504EA4EE-ABEE-4EBF-AB89-C4200E1CDAD4

5.4 Ocean Heat Content

5.4.1. State of art

The Ocean Heat Content (OHC) is considered an important Ocean Monitoring Indicator of ocean warming due to climate change. With about 90% of the excess heat accumulated in the Earth system deposited in the world's ocean, the Earth Energy Imbalance causes rising ocean temperatures and increasing ocean heat content (OHC). Cheng et al. (2021; 2022; 2023; 2024) provide every year an updated estimate of OHC for the global ocean and its regional basins starting from the World Ocean Database data. Two different

OHC products are used for the annual OHC report assessment (Fig. 21): (1) the Institute of Atmospheric Physics (IAP) at the Chinese Academy of Sciences (CAS); (2) National Centers for Environmental Information (NCEI) at the National Oceanic and Atmospheric Administration (NOAA). The Mediterranean Sea is the ocean region that shows the highest warming (Fig. 22), thus it is extremely important to have a rapid and efficient OHC assessment. The Copernicus Marine Service also provides in its Ocean Monitoring Indicators catalogue the Mediterranean OHC Anomaly (0-700m) time series (Fig. 23) and trend from Reanalysis & Multi-Observations Reprocessing, but without a systematic yearly update.

The Blue Cloud 2026 proposal of OHC indicator would provide an operational workflow that would allow the rapid Mediterranean OHC estimation from multiple data sources. Moreover the use of the workbench 1 EOVs dataset, which will integrate data from the four main BDIs, would allow the maximization of data spatial and temporal coverage, reducing the associated product uncertainty.

OHC has been defined within the Copernicus Marine Service (von Schuckmann et al. 2016) as the deviation from a reference period and it is closely proportional to the average temperature change in a specific ocean layer, usually from the surface to 700 m or 2000m depth:

$$OHC = \rho_0 C_p \int_{z1}^{z2} (T_i - \underline{T})dz$$

with a reference density (ρ_0) of 1030 kg m^{-3} and a specific heat capacity (C_p) of 3980 J/kg°C.

Within the SeaDataCloud project a Mediterranean OHC product was delivered (Simoncelli et al., 2020; 2021). Input data from SeaDataNet and CMS-CORA (version 5.2) were integrated and provided to DIVAnd tool to obtain temperature gridded fields (sliding decades) between the sea surface and 2000 m of depth, from which to compute OHC anomaly time series and trends in the layers 0-700 m (Fig. 24) and 0-2000 m. The workflow implemented in the framework of the SeaDataCloud project has been deployed in the MEI VLab and developments are ongoing to use the Beacon monolithic instances as input data, extending the time coverage closest to the present.

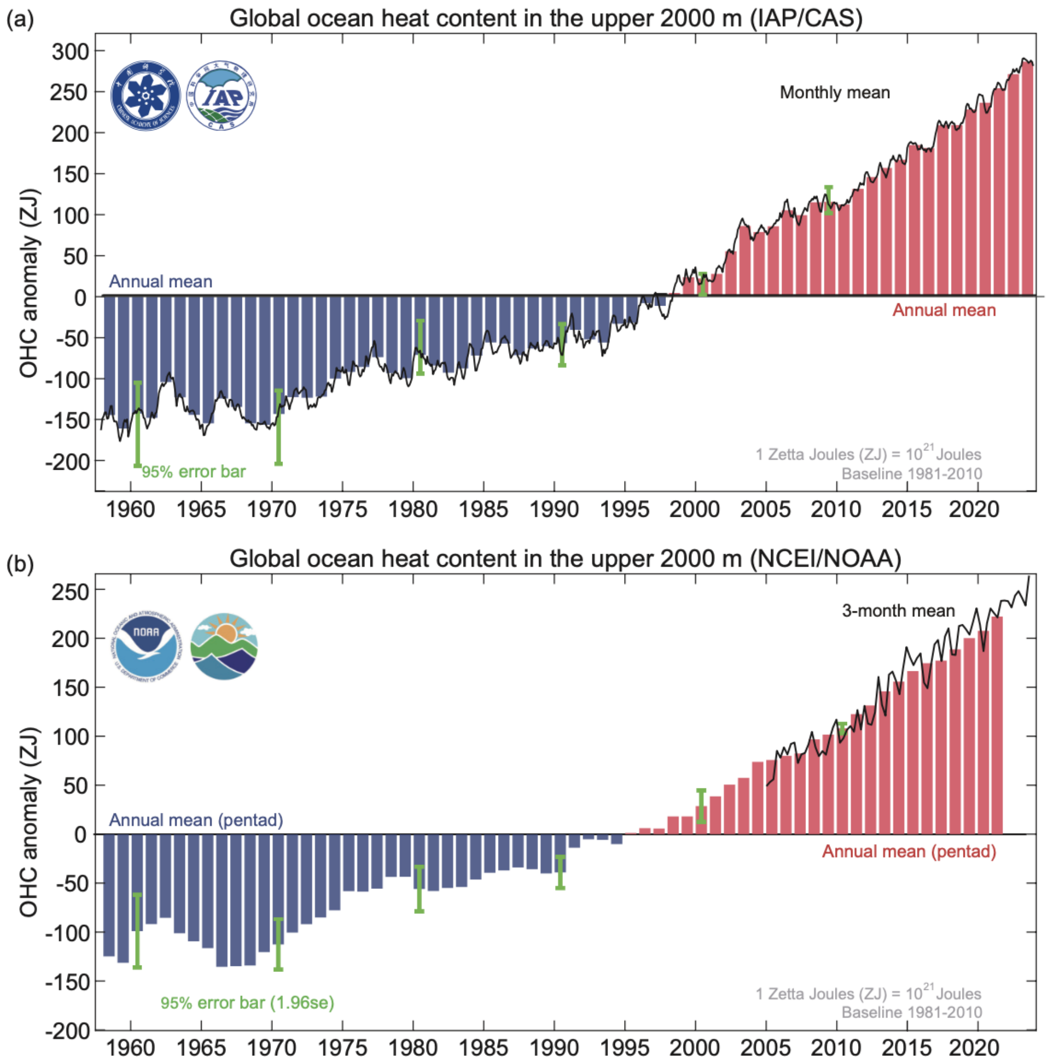


Figure 21. Global upper 2000m OHC from1958 through 2023 according to (a)IAP/CAS and (b) NCEI/NOAA (1 ZJ = 1021 J). The line shows (a) monthly and (b) seasonal values, and the histogram presents (a) annual and (b) pentad anomalies relative to a 1981–2010 baseline. (Fig 2. from Cheng et al. 2024)

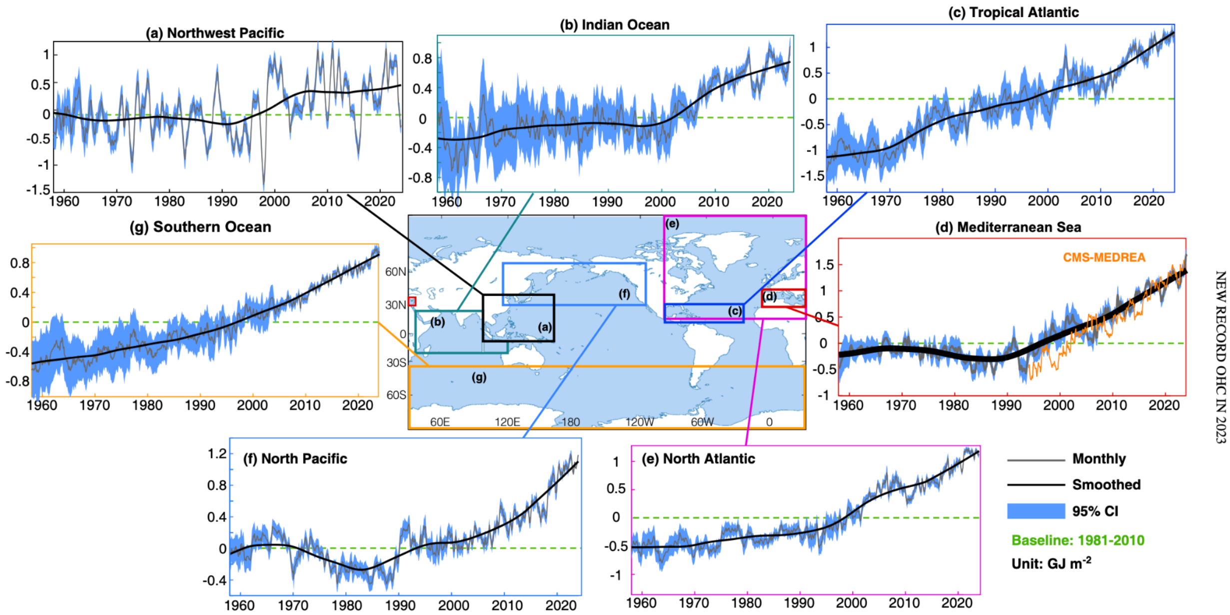


Figure 22. Regional observed upper 2000 m OHC change from 1958 through 2023 relative to a 1981-2010 baseline. (Fig 8. from Cheng et al. 2024)

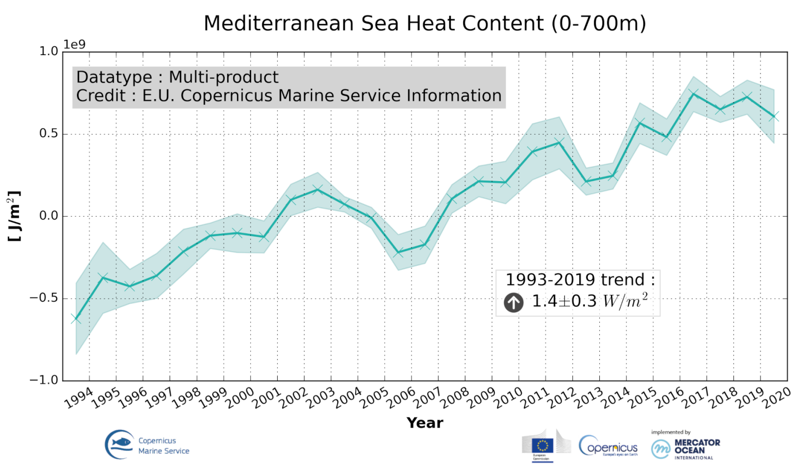


Figure 23. Time series of annual mean area averaged ocean heat content in the Mediterranean Sea (basin wide), and integrated over the 0-700m depth layer during 1993-2019 (European Union-Copernicus Marine Service, 2019): ensemble mean and ensemble spread (shaded area). The ensemble mean is based on different data products, i.e. Mediterranean Sea Reanalysis, global ocean reanalysis GLORYS, C-GLORS, ORAS5, FOAM; global observational based products CORA, ARMOR3D.

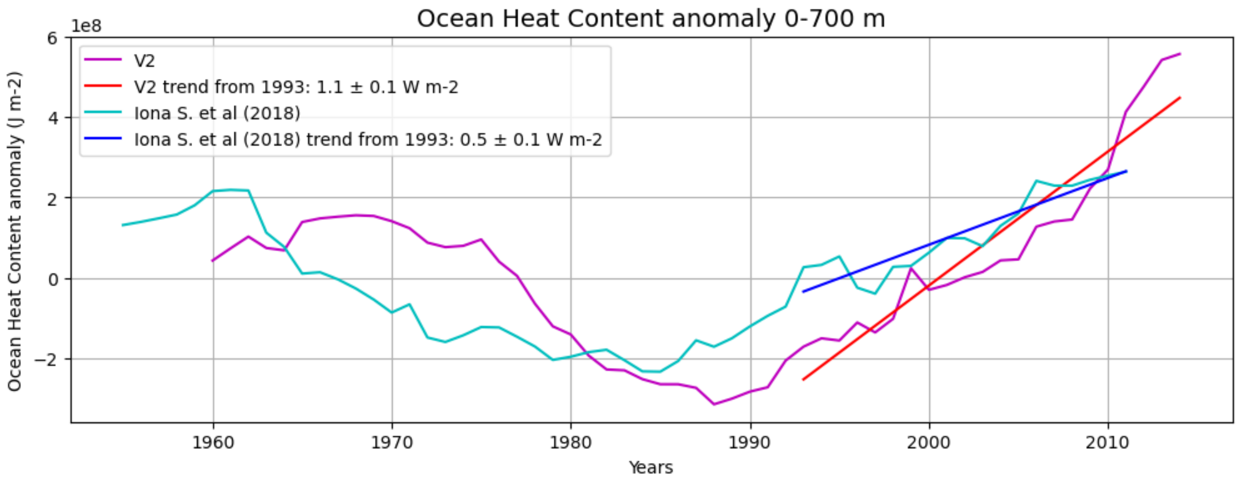


Figure 24. Annual mean area averaged ocean heat content anomaly in the Mediterranean Sea (basin wide), and integrated over the 0-700m depth layer during 1960-2014 (Simoncelli et al. 2020; 2021).

5.4.2. Input data sources

The aim is to provide to the users an operational workflow to rapidly estimate the OHC indicator in the Mediterranean Sea from multiple historical datasets using the DIVAnd mapping tool for gridding temperature in situ data.

The evaluation of OHC indicator in the Mediterranean Sea domain will be possible from several input data sources listed in Table 7. The in situ temperature data spanning the time period after 1950 from CMEMS CORA, EuroArgo, World Ocean Database and SeaDataNet will be gathered querying the beacon monolithic instances available to the Blue-Cloud VLabs. The data will then be an input to the DIVAnd tool for the generation of sliding decadal gridded fields to be used for the computation of the OHC from various data sources (Fig. 25). The present implementation of the OHC indicator uses temperature gridded fields computed by Simoncelli et al. (2020; 2021), but new temperature gridded fields from the above mentioned sources are in development. The

OHC notebook will be updated accordingly. Moreover, once the Blue Cloud EOV dataset (WP3 output) will be produced, the relative OHC estimation will be integrated in the workflow.

The users can compute the OHC in sub-regions of the Mediterranean Sea, in different layers of the water column between the sea surface and 2000 m depth and for selected time periods.

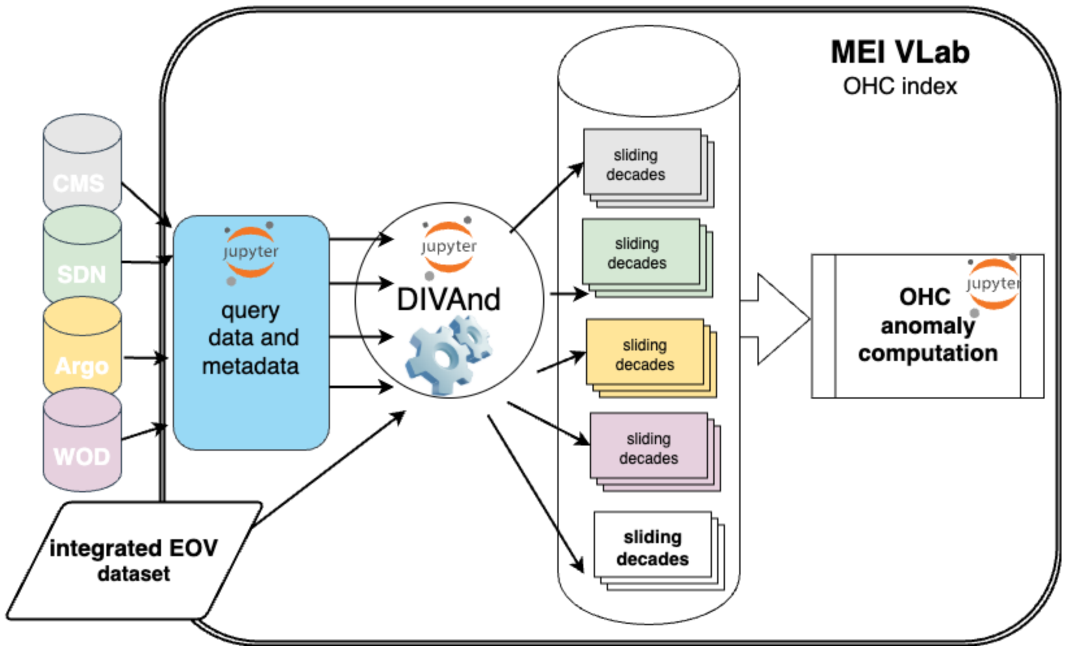


Figure 25. Schematic representation of the OHC workflow.

VARIABLES	DATA SOURCES	DATA ACCESS
In situ Temperature (℃)	CMEMS CORA	Data will be accessed using beacon monolithic instances to generate gridded Temperature fields with DIVAnd, from which to derive the OHC.
In situ Temperature (℃)	Argo	Data will be accessed using beacon monolithic instances to generate gridded Temperature fields with DIVAnd, from which to derive the OHC.

VARIABLES	DATA SOURCES	DATA ACCESS
In situ Temperature (°C)	World Ocean Database	Data will be accessed using beacon monolithic instances to generate gridded Temperature fields with DIVAnd, from which to derive the OHC.
In situ Temperature (°C)	SeaDataNet	Data will be accessed using beacon monolithic instances to generate gridded Temperature fields with DIVAnd, from which to derive the OHC.
In situ Temperature (°C)	Blue Cloud EOV dataset (WP3 output)	Blue Cloud catalog

Table 7. VLab 4 OHC Input data sources

5.4.3. Step by step guideline to use the VLab

1. Where is The Jupyter Notebook

1) Once in https://blue-cloud.d4science.org/group/marineenvironmentalindicators follow the instructions at paragraph 5.3 for installing the notebook

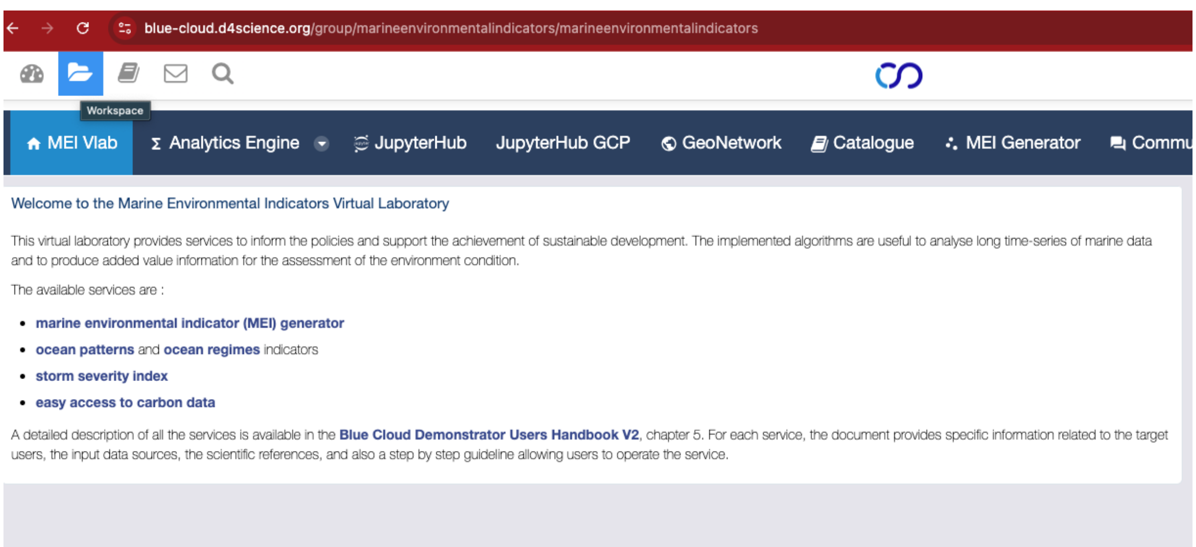


Figure 26. Open the VLab

2) Access to the folder OHC (Figure 27).

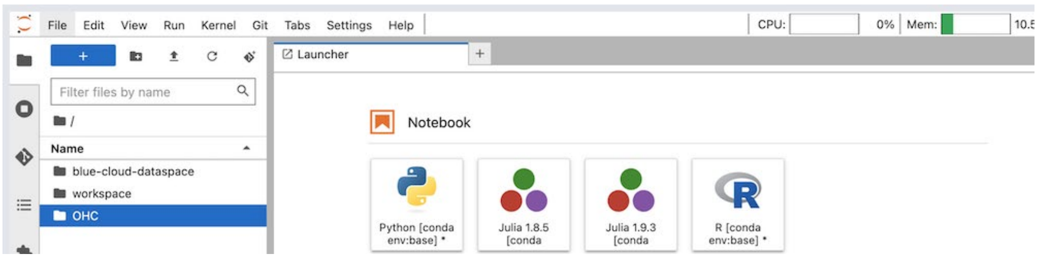


Figure 27. Access the folder OHC

3) Inside the folder there is the OHC.ipynb notebook (Figure 28).

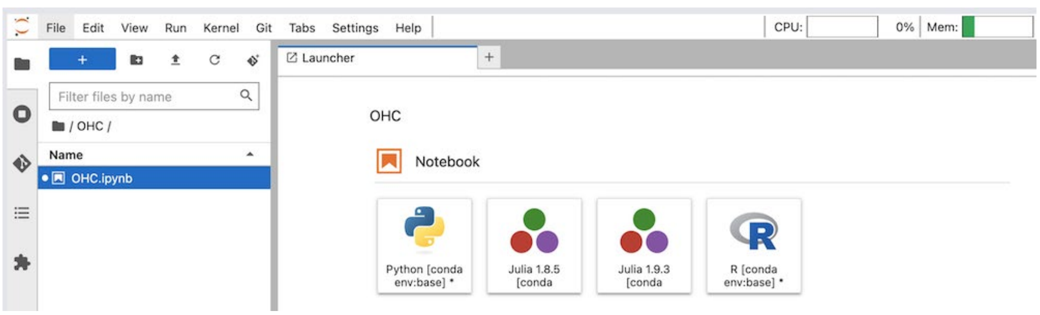


Figure 28. Open "OHC" folder

Now you can double click on the "OHC.ipynb" notebook and it will open in a new tab

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

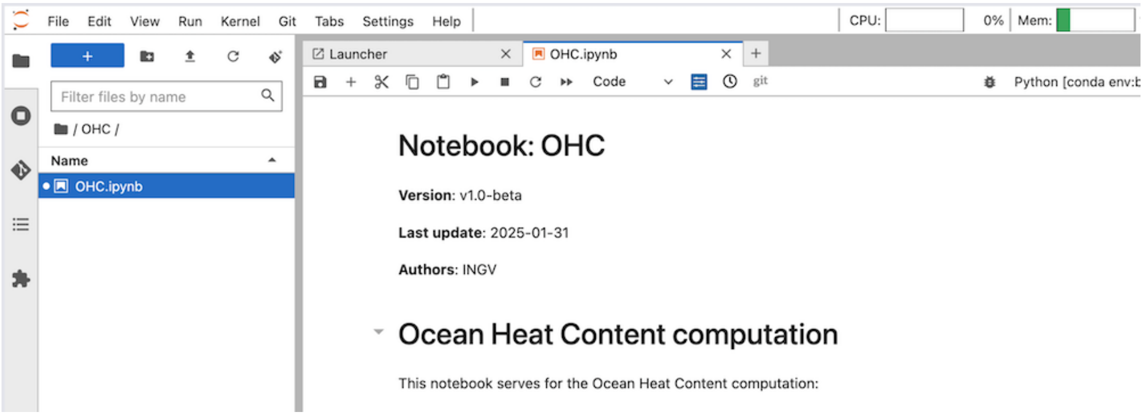


Figure 29. The notebook opened

2. Working with the right Server Option

Be sure your Jupyterhub Server Option is settled on the Blue-Cloud Julia 1.9.3 - 8 Cores/32G RAM (Figure 30):

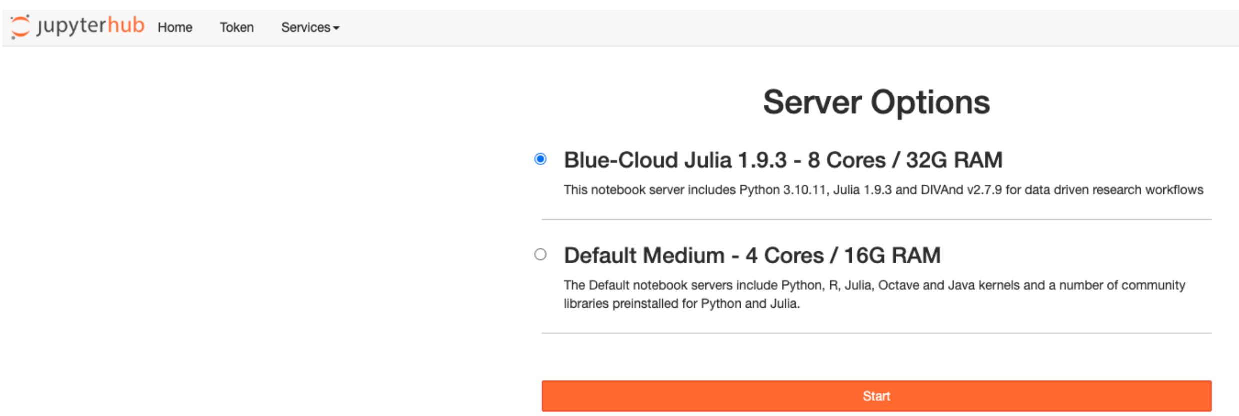


Figure 30. Blue-Cloud Julia 1.9.3 - 8 Cores/32G RAM

3. Reading and Running the Jupyter Notebook

Be sure that the “Python[conda env:base]” kernel is correctly selected in the top-right panel of the jupyter notebook.

The notebook is divided into cells that can be executed sequentially, giving the possibility to the user to interact manually on some parameters.

The notebook serves for the OHC computation. The actual configuration:

- horizontal resolution = 0.125 degrees;
- The largest domain coinciding with a “Mediterranean” box: lon = [-5.625 36.5] degrees E, lat = [30 46] degrees N;
- Input 1955-2017 temperature sliding decades annual analysis from Simoncelli et al. (2020; 2021);
- Mask “gebco_2019_mask_1_8_edited_final.nc” edited opening and closing numerous zones and with removed less than 3 points connected components (WOA18 depth levels versions);
- OHC plots 0-2000m, computation and save in netCDF format for publication.

1. The Initial cells import the libraries and packages you need and define the Metadata to be correctly saved within the final results.
2. The “Data Reading” section defines the path to be followed to find/store the input/output data. These INPUT/OUTPUT folders are currently directed in the “blue-cloud-dataspace/MEI/INGV/” directory and in “/workspace/MEI/OceanHeatContent/”
3. The “Grid Selection” section defines some fixed parameters used for the calculation and the resolution of the gridded domain (Figure 31). It allows through the use of widgets the selection of latitude and longitude interval window as like as the depth edges and the years window to which refer the following analysis. Executing the cells with the within-the-notebook panel , you load the domain and the time window selected.

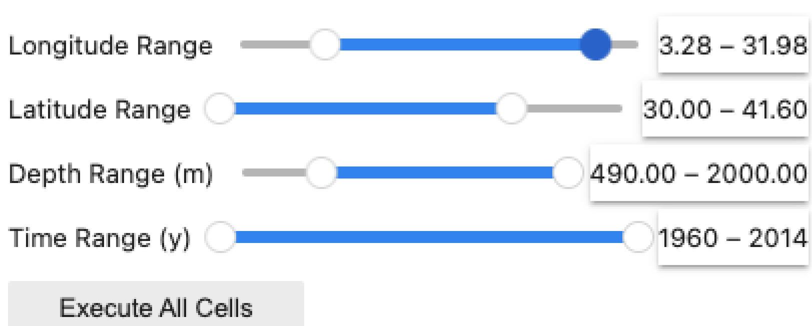


Figure 31. Grid selection window

4. The following cells operate some mask instructions to select the desired domain and create the baseline for the temperature fields and the related plots.
5. In the “Load Temperature file” section the necessary temperature fields are loaded and some example plots of those loaded fields are shown over the selected domain (Figure 32), together with the Hoev-Moeller plots for Temperature variable (Figure 33) and its anomaly for the time period (Figure 34).

```
plt.show()
```

```
-5.625 30.0
```

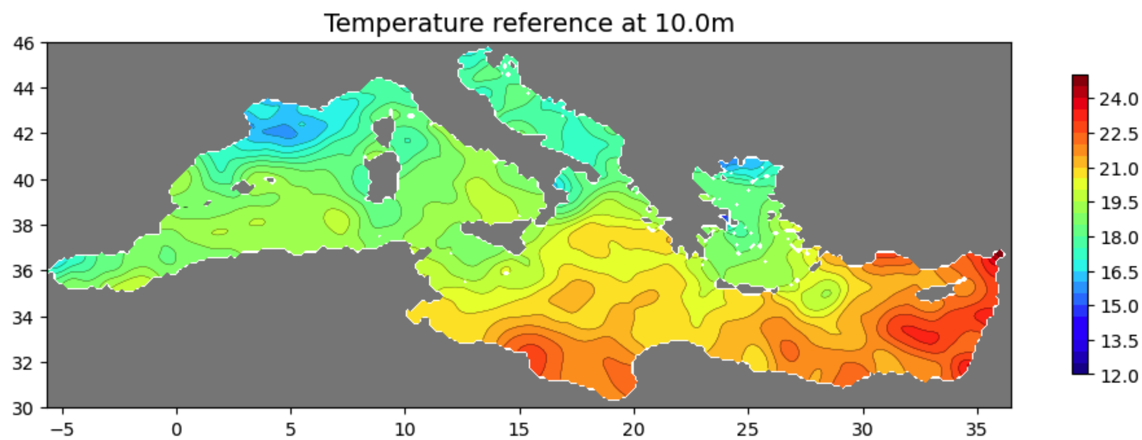


Figure 32. Example plot of temperature

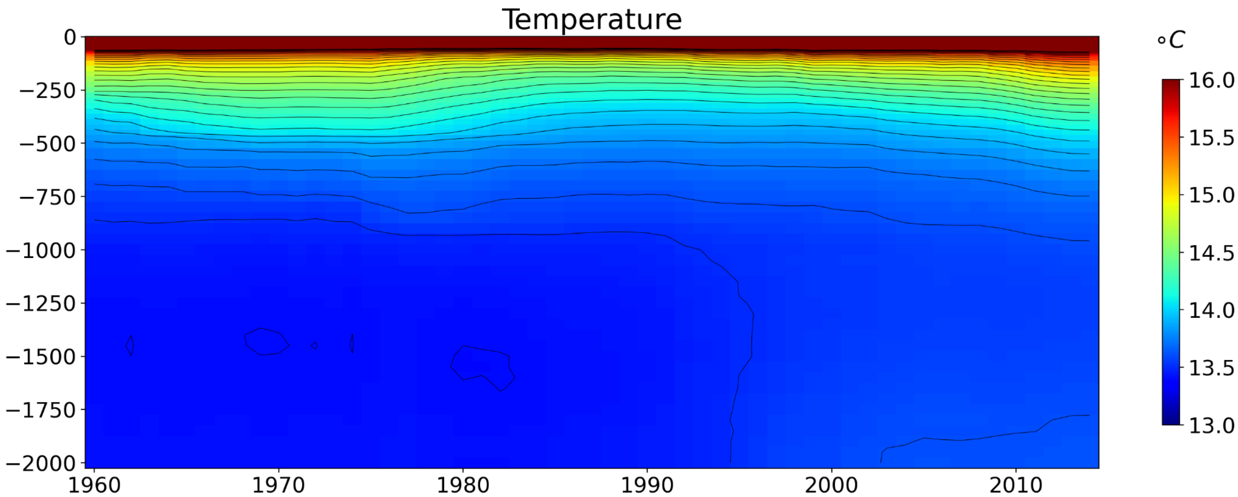


Figure 33. Example of Hoev-Moeller plot for temperature variable

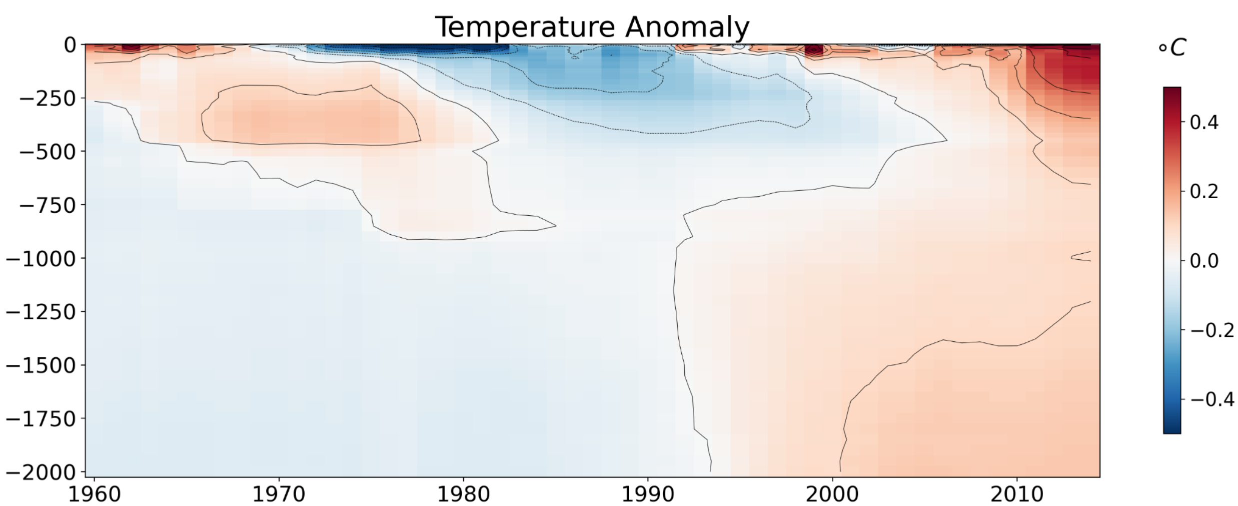


Figure 34. Example plot for temperature anomaly

6. The “Computation” section executes the OHC calculation for the loaded temperature fields over the selected domain, plotting the results within the Notebook, saving some plots in the OUTPUT directory and storing the numerical results, supplied by metadata, in a NetCDF file (Figure 35, 36 & 37).

```
plt.tight_layout()
# Save the plot as an image file with specified settings (e.g., resolution)
plt.savefig(OUTPUT_PATH+'TEMP_ANOMALY_TREND_'+suffix+'.png', bbox_inches='tight', dpi=200)

# Show the plot
plt.show()
```

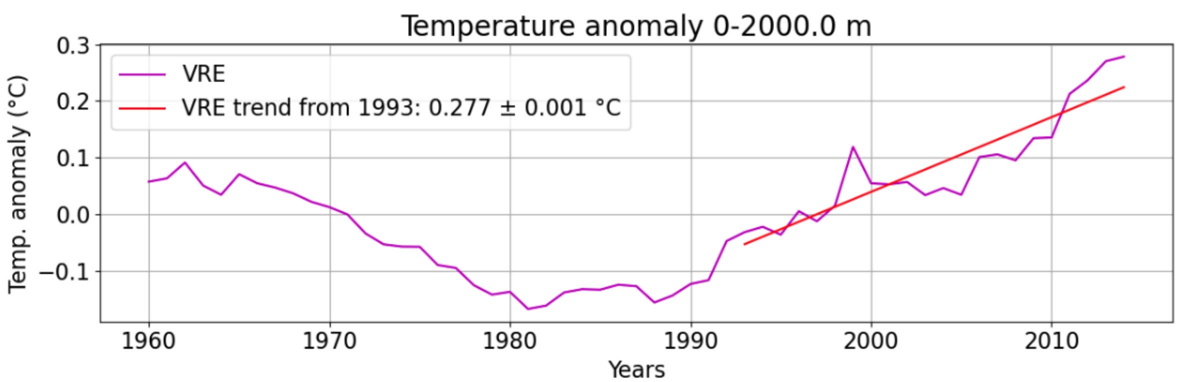


Figure 35. Temperature anomaly output

```
plt.tight_layout()
# Save the plot as an image file with specified settings (e.g., resolution)
plt.savefig(OUTPUT_PATH+'OHC_700_ANOMALY_'+suffix+'.png', bbox_inches='tight', dpi=200)

# Show the plot
plt.show()
```

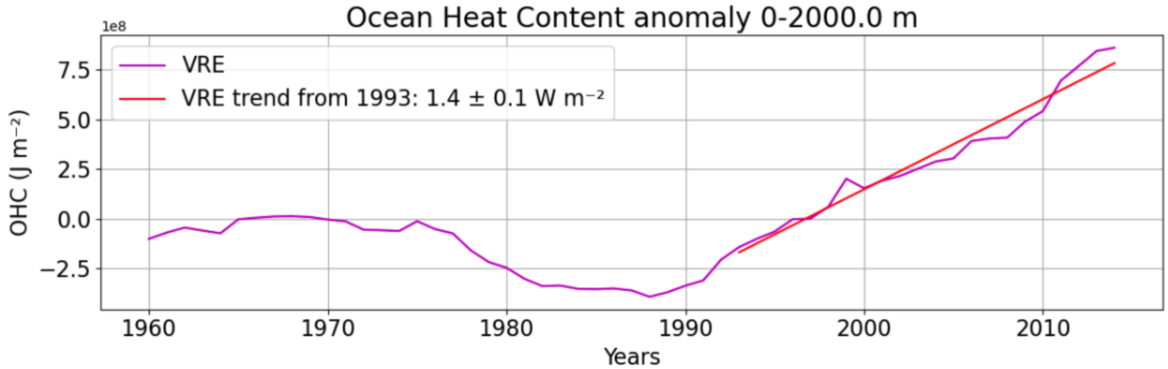


Figure 36. Ocean Heat Content anomaly output

Write output in netCDF dataset

```
[34]: # Handling netCDF file creation
out_file = OUTPUT_PATH+'Ocean_Heat_Content_WP4_'+suffix+'.nc'

# Delete the previous dataset if it exists
if os.path.isfile(out_file):
    os.remove(out_file)
    print(f"Removing file {out_file}")

# Create the output dataset
out_data = Dataset(out_file, "w", format="NETCDF4")

# Global attributes
out_data.setncatts({
    "Conventions": "CF-1.6"
})

# Define dimensions
depth_dim = out_data.createDimension("depth", 1)
time_dim = out_data.createDimension("time", len(years))
nv_dim = out_data.createDimension("nv", climatology_bounds.shape[0])

# Define variables
nctime = out_data.createVariable("time", "f8", ("time",))
nctime.units = "days since 1900-01-01 00:00:00"
nctime.standard_name = "time"
nctime.long_name = "time"
nctime.calendar = "standard"
nctime.climatology = "climatology_bounds"

ncdepth_bounds = out_data.createVariable("depth_bounds", "f8", ("nv", "depth"))
ncdepth_bounds.units = "m"

ncclimatology_bounds = out_data.createVariable("climatology_bounds", "f8", ("nv", "time"))
ncclimatology_bounds.units = "days since 1900-01-01 00:00:00"

# Variables for ocean heat content and temperature anomaly, using ("time", "depth") to switch dimensions
ncVariable1 = out_data.createVariable("Ocean_Heat_Content", "f4", ("time", "depth"), fill_value=9.96921e36, zlib=True)
ncVariable1.units = "J"
ncVariable1.standard_name = "ocean_heat_content"
ncVariable1.long_name = "Ocean Heat Content Anomaly"
ncVariable1.cell_methods = "time: mean within years time: mean over years"
```

Figure 37. NetCDF file output

All the output files (plots and netCDF) are saved with a suffix uniquely linked to the selection made by the user in the “Grid Selection” section.

5.4.4. Using other data sources

The proposed workflow is considering as input the data from the four main Blue Data Infrastructures and its added value will be the use of the Blue-Cloud EOV dataset produced within workbench 1, which will integrate them. Since the OHC is the main climate change indicator, its estimation should be based on the maximum number of validated data. However the user might want to integrate its own additional temperature data in a target region for a customized estimate, thus we will provide an example.

5.4.5. Authors

INGV team: Enrico Baglione, Paolo Oliveri and Simona Simoncelli

5.4.6. References

Cheng, L., Abraham, J., Trenberth, K. E., Boyer, T., Mann, M. E., Zhu, J., Wang, F., Yu, F., Locarnini, R., Fasullo, J., Zheng, F., Li, Y., Zhang, B., Wan, L., Chen, X., Wang, D., Feng, L., Song, X., Liu, Y., . . . Lu, Y. (2024). New record ocean temperatures and related climate indicators in 2023. *Advances in Atmospheric Sciences*, *41*(6), 1068–1082. https://doi.org/10.1007/s00376-024-3378-5

Cheng, L., Abraham, J., Trenberth, K. E., Fasullo, J., Boyer, T., Mann, M. E., Zhu, J., Wang, F., Locarnini, R., Li, Y., Zhang, B., Yu, F., Wan, L., Chen, X., Feng, L., Song, X., Liu, Y., Reseghetti, F., Simoncelli, S., . . . Li, G. (2023). Another year of record heat for the oceans. *Advances in Atmospheric Sciences*, *40*(6), 963–974. https://doi.org/10.1007/s00376-023-2385-2

Cheng, L., Abraham, J., Trenberth, K. E., Fasullo, J., Boyer, T., Mann, M. E., Zhu, J., Wang, F., Locarnini, R., Li, Y., Zhang, B., Tan, Z., Yu, F., Wan, L., Chen, X., Song, X., Liu, Y., Reseghetti, F., Simoncelli, S., . . . Reagan, J. (2022). Another Record: Ocean Warming Continues through 2021 despite La Niña Conditions. *Advances in Atmospheric Sciences*, *39*(3), 373–385. https://doi.org/10.1007/s00376-022-1461-3

Cheng, L., Abraham, J., Trenberth, K. E., Fasullo, J., Boyer, T., Locarnini, R., Zhang, B., Yu, F., Wan, L., Chen, X., Song, X., Liu, Y., Mann, M. E., Reseghetti, F., Simoncelli, S., Gouretski, V., Chen, G., Mishonov, A., Reagan, J., & Zhu, J. (2021). Upper Ocean temperatures hit record high in 2020. *Advances in Atmospheric Sciences*, *38*(4), 523–530. https://doi.org/10.1007/s00376-021-0447-x

European Union-Copernicus Marine Service. (2019). Mediterranean Ocean Heat Content Anomaly (0-700m) time series and trend from Reanalysis & Multi-Observations Reprocessing [Data set]. Mercator Ocean International. https://doi.org/10.48670/MOI-00261

Simoncelli, S., & Oliveri, P. (2021). SeaDataCloud Mediterranean Ocean Heat Content Product Information Document (Version 3). SeaDataCloud. https://doi.org/10.13155/79146

Simoncelli, S., Oliveri, P., & Mattia, G. (2020). SeaDataCloud Mediterranean Sea - Ocean Heat Content 1955-2018 (Version 1.1) [Data set]. IFREMER / IDM/SISMER. https://doi.org/10.12770/504EA4EE-ABEE-4EBF-AB89-C4200E1CDAD4

5.5 Eutrophication Indicator - TRIX

5.5.1 State of art

The assessment of the risks and impacts of eutrophication in estuarine and coastal waters is one of the key issues in marine environmental management (Painting, S.J., et al 2005). The Trophic index (TRIX), introduced by Volleinweider et al. (1998) to characterise the trophic conditions of seawater, is a linear combination of the logarithms of four state variables (Chl-a, DIN, TP and the absolute percentage deviation from oxygen saturation, aD%O). Therefore, this composite index aggregates pressure (nutrients), biological response (Chl-a, a proxy for biomass) and environmental disturbance in the water quality (oxygen) (Ærtebjerg, Gunni et al. 2001).

The trophic state depends on the availability of nitrogen and phosphorus for primary production, which in terms determines the phytoplankton biomass and oxygen saturation. In TRIX the nutrients are represented ideally by total nitrogen and total phosphorus; chlorophyll-a is a substitute parameter for phytoplankton biomass, as production is not routinely measured; and the deviation of oxygen saturation from 100% (aD%O) in the productive layer indicates the production intensity of the system (Ærtebjerg, Gunni et al. 2001). This simple index permits one to synthesise key eutrophication variables into a simple numeric expression to make information comparable over a wide range of trophic situations, while avoiding the subjectivity in the usage of traditional trophic terminology.

5.5.2. Input data sources

VARIABLES	DATA SOURCES	DATA ACCESS
Water body Dissolved Inorganic Nitrogen (DIN) **Water body chlorophyll-a (CHL)** **Water body Total Phosphorus (TP)** **Water body Dissolved Oxygen saturation (DO)**	EMODnet Chemistry 2024 Eutrophication MediterraneanSea: - Eutrophication Med profiles 2024 unrestricted - Eutrophication Med time series 2024 unrestricted DIVAnd interpolated climatologies* calculated for the stated parameters (Chl-a, DIN, TP & DO%).	Data is not accessible via the Blue Cloud Data Access service. It has been uploaded to Vlab. The data is available through the bluecloud dataspace at: **/blue-cloud-dataspace/MEI/Eutrophication_indicator/data/** The datasets are not yet public since the toolbox is still in beta testing or until the 1st release of the workbench dataset is available.

VARIABLES	DATA SOURCES	DATA ACCESS
	*The climatologies have been calculated by merging time series and profiles collections from EMODnet chemistry.	

Table 8. VLab 4 TRIX Input data sources

5.5.3. Step by step guideline to run TRIX Jupyter Notebook

1) From the D4Science gateway https://blue-cloud.d4science.org/group/bluecloud-gateway enter the **Marine environmental Indicators (MEI) VLab**. Once there choose JupyterHub (Figure 38).

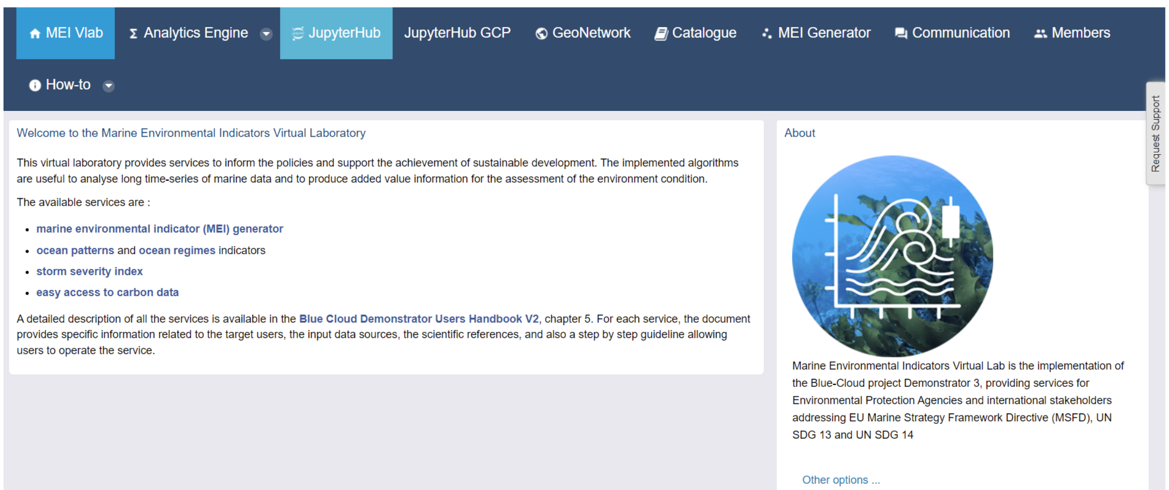


Figure 38. Marine environmental Indicators VLab

2) In your Jupyterhub the following options will be granted (Figure 39).

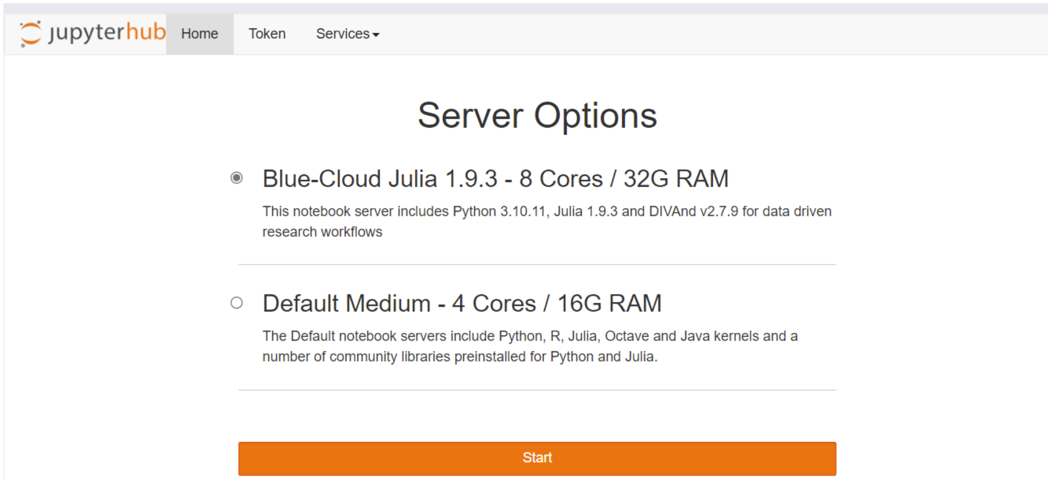


Figure 39. Server options

3) Once the server has been chosen the user will be asked to pre-build the dash to create the interactive dashboard. Go on with "*build*" and "*save and reload*" (Figure 40).

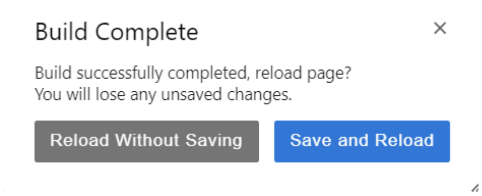


Figure 40. Window to be selected to build the interactive dashboard

4) After following the instructions in paragraph 5.3 for unzipping the notebook, the user will now see some folders and a Jupyter Notebook in the column on the left of the screen. Select the *Jupyter Notebook TRIX.ipynb* (Figure 41).

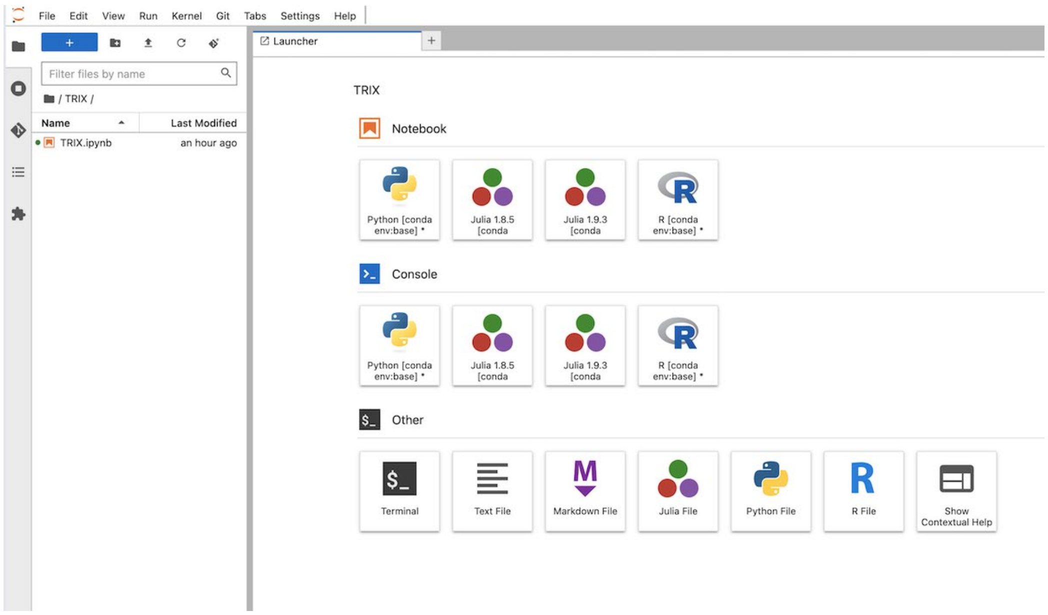


Figure 41. Select the Jupyter Notebook TRIX.ipynb

The notebook contains all the necessary steps to

a) Read the DIVAnd analysis outputs stored on the dataspace (Figure 42). DIVAnd climatologies have been calculated from in-situ data from EMODnet chemistry data collections for the Adriatic sea representative of a time period of 10 years (2013-2023).

```
[4]: data_path="~/blue-cloud-dataspace/MEI/Eutrophication_indicator/data"

[5]: ds_DIN = xr.open_dataset(f'{data_path}/Water_body_Dissolved_inorganic_nitrogen_Adriatic.4Danl.nc',decode_cf=True)
     ds_CHL = xr.open_dataset(f'{data_path}/Water_body_Chlorophyll_a_Adriatic.4Danl.nc',decode_cf=True)
     ds_TP = xr.open_dataset(f'{data_path}/Water_body_Total_phosphorus_Adriatic.4Danl.nc',decode_cf=True)
     ds_DO = xr.open_dataset(f'{data_path}/Water_body_Dissolved_oxygen_saturation_Adriatic_24072024.4Danl.nc',decode_cf=True)

[6]: # examine the structer of the parameter
     ds_DO.to_dataframe().describe()
```

[6]:

	climatology_bounds	Dissolved_oxygen_saturation	Dissolved_oxygen_saturation_L1	Dissolved_oxygen_saturation_L2	Dissolved_oxygen_s
count	16915392	4.869144e+06	184576.000000	289692.000000	
mean	2008-06-30 16:59:59.999996928	9.950395e+01	94.438965	95.371582	
min	2003-01-01 00:00:00	5.681399e+01	56.813988	56.813988	
25%	2003-06-23 12:00:00	9.225182e+01	88.919846	90.371361	
50%	2008-07-01 12:00:00	1.017738e+02	94.972382	95.172478	
75%	2013-07-07 18:00:00	1.068053e+02	102.577202	102.785599	
max	2013-12-31 00:00:00	1.414151e+02	141.415115	141.415115	
std	NaN	8.224477e+00	11.406313	10.545814	

Figure 42. Read the DIVAnd analysis outputs stored on the dataspace

b) filter outliers using the ranges for the specific area (Figure 43).

- **General Filtering Process**:
  - Values that fall outside the specified range are replaced with `NaN` (Not a Number), effectively filtering them out.
- **Variables and Ranges**:
  - **Chlorophyll_a (`CHLF`)**:
    - Filters values to retain only those between 0 and 4.
  - **Dissolved Inorganic Nitrogen (`DINF`)**:
    - Filters values to retain only those between 0 and 30.
  - **Total Phosphorus (`TPF`)**:
    - Filters values to retain only those between 0 and 1.
  - **Dissolved Oxygen Saruration (`DOF`)**:
    - Filters values to retain only those between 0and 120.

This filtering process is essential for focusing on the range of interest for each variable and discarding outliers or irrelevant data points.

```
[7]: CHLF = ds_CHL.Chlorophyll_a.where((ds_CHL.Chlorophyll_a >=0) & (ds_CHL.Chlorophyll_a <=4))
     DINF = ds_DIN.Dissolved_inorganic_nitrogen.where((ds_DIN.Dissolved_inorganic_nitrogen >= 0) & (ds_DIN.Dissolved_inorganic_nitrogen <=3
     TPF = ds_TP.Total_phosphorus.where((ds_TP.Total_phosphorus >=0) & (ds_TP.Total_phosphorus <=1))
     DOF = ds_DO.Dissolved_oxygen_saturation.where((ds_DO.Dissolved_oxygen_saturation >= 0) & (ds_DO.Dissolved_oxygen_saturation <=120))
```

Figure 43. Filter outliers using ranges for specific area

c) calculate the TRIX indicator by first checking the units and if necessary convert them (Figure 44). Absolute percentage deviation from oxygen saturation needs to be calculated in this step as well.

Each state variable is scaled by the highest (U_i) and the lowest (L_i) values in the data time series, and TRIX is defined as:

$$\mathrm{TRIX} = \frac{k}{n}\sum_{i=1}^{n}\frac{(\log M_i - \log L_i)}{(\log U_i - \log L_i)}$$

where $k = 10$ is another scaling factor; $n = 4$ is the number of state variables considered; and M_i are the observed Chl, DO, DIN, and TP values.

Vollenweider et al. (1998) further simplified the TRIX formula by assuming (on the basis of the data used) that the difference $(\log U_i - \log L_i)$ was equal to 3 for all state variables. Therefore, considering $k = 10$, $n = 4$, and the specific $\log L_i$ values, the TRIX formula was rewritten as follows:

$$\mathrm{TRIX} = \frac{10}{12}\left[(\log M_{\mathrm{Chl}} + 0.5) + (\log M_{\mathrm{DO}} + 1) + (\log M_{\mathrm{DIN}} - 0.5) + (\log M_{\mathrm{TP}} + 0.5)\right]$$

Or

$$\mathrm{TRIX} = \frac{1}{1.2}\left[(\log(M_{\mathrm{Chl}}\,M_{\mathrm{DO}}\,M_{\mathrm{DIN}}\,M_{\mathrm{TP}}) + 1.5\right]$$

The latest equation gives the TRIX index currently used by ARPAE and adopted by the Italian national legislation (D.L 260/2010).

[15]:
```
TRIX = abs((np.log10(CHL*DIN*TP*DO)+1.5)/1.2)

print(np.nanmax(TRIX))
print(np.nanmin(TRIX))
```
```
6.0775857
0.070939064
```

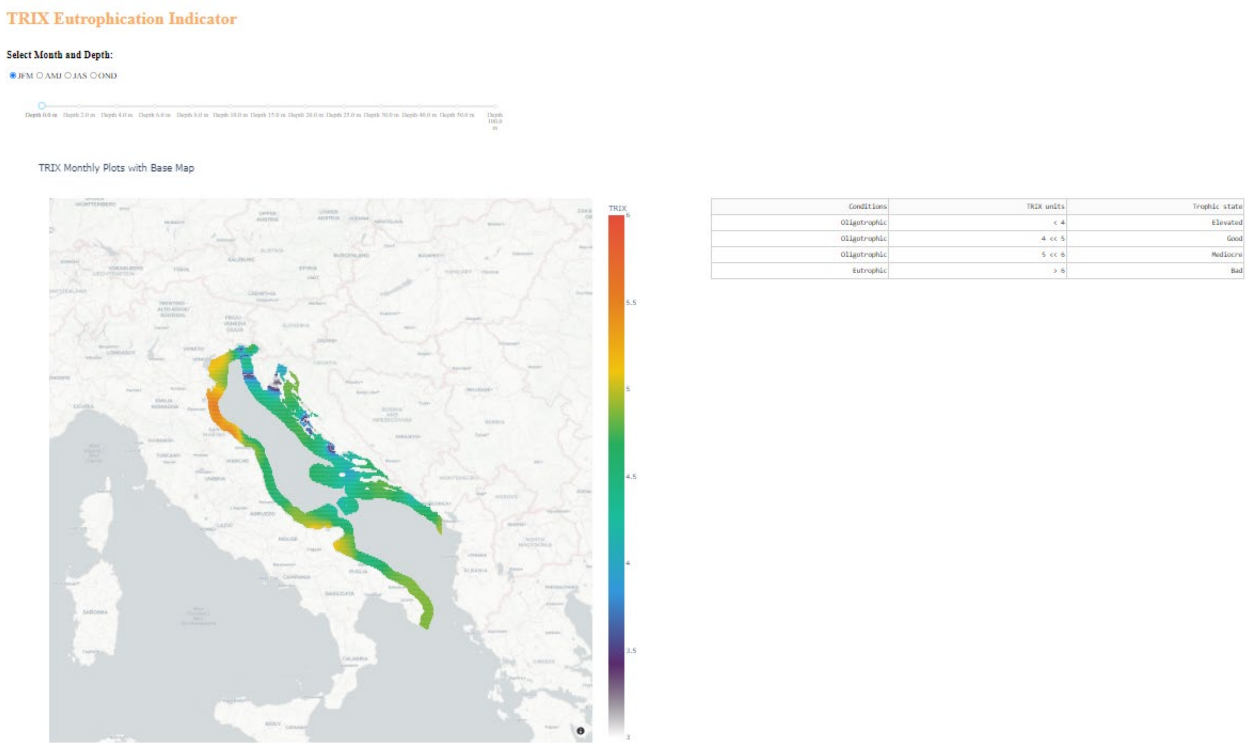
[27]:
```
TRIX
```

[27]: xarray.DataArray (**time**: 12, **depth**: 12, **lat**: 234, **lon**: 251)

Figure 44. Calculate the TRIX indicator

The output TRIX is an array which contains spatial (lon, lat and depth) and temporal (monthly and seasonal) information which we will use to create the dashboard.

5) An interactive map is created from the output where the user can check the eutrophication state of an area of interest by changing depth and season (Figure 45).



Conditions	TRIX units	Trophic state
Oligotrophic	< 4	Elevated
Oligotrophic	4 << 5	Good
Oligotrophic	5 << 6	Mediocre
Eutrophic	> 6	Bad

Figure 45. Interactive map where the user can check the eutrophication state of an area of interest by changing depth and season

5.5.4. Using other data sources

The Beta version of the indicator will be calculated based on EMODnet chemistry data while at the same time an alpha version with the Eutrophication workbench output dataset will be in development.

In order to calculate the index with other data sources the user needs the following variables: Depth, water temperature, water body salinity, water body dissolved oxygen concentration, water body dissolved oxygen saturation, water body total phosphorus, water body chlorophyll-a and water body dissolved inorganic nitrogen. Then the following the following preprocessing steps should be applied:

1. Take the 0–100 m depth layer to consider only the eutrophication signal.

2. Take QFs 1, 2, 6, and Q for all parameters except depth and time (exclude QF 4).

3. The dissolved oxygen solubility concentration needs to be calculated for data compiling with application domain T = 0–40 °C and S = 0–40 (Benson and Krause, 1984).

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

4. Calculate oxygen saturation from Benson and Krause, 1984.

Once this has been done, the code provides a way to discard outliers and to check units which is important for the final calculation of the TRIX index.

5.5.5. Authors

OGS: Nydia Catalina Reyes Suarez, Marina Lipizer , Oussema Fersi, Megan Anne French,Alessandra Giorgetti

5.5.6. References

Adolf Stips, Diego Macias, Elisa Garcia-Gorriz, Svetla Miladinova; Alternative assessments of large scale Eutrophication using ecosystem simulations: hind-casting and scenario modelling; EUR 27904; doi:10.2788/156650

Ærtebjerg, G., Carstensen, J., Dahl, K., Hansen, J.L., Nygaard, K., Rygg, B., Sørensen, K., & Severinsen, G. (2001). Eutrophication in Europe's coastal waters.

Benson, B. B., & Krause Jr, D. (1984). The concentration and isotopic fractionation of oxygen dissolved in freshwater and seawater in equilibrium with the atmosphere 1. Limnology and oceanography, 29(3), 620-632.

Fiori, E., Zavatarelli, M., Pinardi, N., Mazziotti, C., and Ferrari, C. R.: Observed and simulated trophic index (TRIX) values for the Adriatic Sea basin, Nat. Hazards Earth Syst. Sci., 16, 2043–2054, https://doi.org/10.5194/nhess-16-2043-2016, 2016.

Flanders Marine Institute (2023). Maritime Boundaries Geodatabase: Territorial Seas (12NM), version 4. Available online at https://www.marineregions.org/. https://doi.org/10.14284/633

OSPAR Commission, 2003. The OSPAR Integrated Report 2003 on the Eutrophication Status of the OSPAR Maritime Area based upon the first application of the Comprehensive Procedure. Includes ‘‘baseline/assessment levels used by the Contracting Parties and monitoring data (MMC 2003/2/4; OSPAR Publication 2003: ISBN: 1-904426-25-5).

Painting, S.J., Devlin, M.J., Rogers, S.I., Mills, D.K., Parker, E.R., Rees, H.L., 2005. Assessing the suitability of OSPAR EcoQOs for eutrophication vs ICES criteria for England and Wales. Marine Pollution Bulletin 50, 1569–1584.

Pettine, M., Casentini, B., Fazi, S., Giovanardi, F., & Pagnotta, R. (2007). A revisitation of TRIX for trophic status assessment in the light of the European Water Framework Directive: application to Italian coastal waters. Marine Pollution Bulletin, 54, 1413–1426.

Volleinweider, R.A., Giovanardi, F., Montanari, G., Rinaldi, A., 1998. Characterization of the trophic conditions of marine coastal waters, with special reference to the NW Adriatic Sea: proposal for a trophic scale, turbidity and generalized water quality index. Environmetrics 9, 329–357.

6. Introduction to VLab 5 “Global Fisheries Atlas”

The Global Fisheries Atlas VLab acts as an umbrella built on top of authoritative data or knowledge sources describing fisheries activities (Table 9).

This VLab aims to better describe fisheries activities at a global scale either by disseminating knowledge or data. To do so, more than ten sources have been harmonised and are accessible from the Global Fisheries Atlas. The VLab content is made accessible by various entry points either through **graphic user interface (GUIS)** or programmatic access depending on users needs and profiles. Users can access or use web mapping products to browse and explore the **Global Record of Stocks and Fisheries (GRSF)** knowledge base as well as datasets describing Tuna Atlas fisheries at a global scale. Beyond data already integrated, this VLab also provides methods and tools to reproduce, update or customize this work to deal with other use cases.

For what regards **GRSF**:

- Programmatic access
  - GRSF API
  - GRSF SPARQL endpoint
- GUIs
  - GRSF CKAN catalogue

For what regards **Global Tuna Atlas**:

- Programmatic access
  - Zenodo GUIs and API to access the main versions of data or code releases
  - R code in RStudio projects
- GUIs
  - Zenodo DOIS for both data and code

- Shiny apps
- Dynamic reports: R markdown

6.1 Target Users & Community

Among possible users of the data products and code made available by the Global Fisheries Atlas VLab, we identified the following profiles:

- **Scientists**
  - Fisheries Scientists
    - Updating the current time series of catch and fishing effort remains the main goal of the Global Tuna Atlas. This task is meant to be achieved by the early adopters of the VLab consolidated group (**Fisheries and Resources Monitoring System (FIRMS)** members: scientists and **Regional Fisheries Management Organisations (RFMOs)** data managers). Indeed, this work is already used and advertised by the FIRMS network (**Food and Agriculture Organisation (FAO)** and RFMOs), datasets have also been assigned DOIs on Zenodo.
    - Accessing updated time series of catch and fishing effort data to e.g. calibrate models, drive stock assessment and population dynamics models.
    - Understand, reproduce or customize the work to tailor new Global Fisheries Atlas for other species (e.g. by executing the work from the VLab RStudio server).
  - Marine Biologists
    - Finding data describing species distribution and abundance that are essential biological variables for ecological research and ocean observatories also called **Essential Ocean Variables (EOVs).**
  - Climate Change Analysts
    - To study the impact of climate on fisheries by running cross domain analysis with physical and chemical parameters (e.g. ocean temperature data, salinity levels, and other climate-related factors).
  - Data scientists can use available fisheries data e.g. to train **AI** models and cross-domain analysis with environmental parameters such as those provided by Copernicus.
- **Policy Makers, Commercial Fisheries, and Conservation Organizations**
  - Conservation Organizations and NGOs

        - Will benefit from detailed fisheries data for habitat protection efforts, biodiversity conservation strategies, and environmental impact assessments.
    - Policy Makers and Regulatory Bodies
        - RFMOs and other policymakers can utilize the GFA for helping craft legislation and marine spatial planning
    - Commercial Fisheries and Industry Stakeholders
        - Operators, fisheries managers, and industry representatives can leverage the GFA for strategic planning, sustainable practice adoption.

- **Data managers in the marine domain and beyond :** interested in implementing similar data management plans for other kinds of data can reuse (execute and customize) the code directly in the VLab. Developers will find "ready to go" solutions for the FAIR management of fisheries data (e.g. time-series on catch and effort that brings collated statistical data into a data harmonization and QC process with dynamic R markdown reports).

- **End-users**: general public with an interest in stocks and fisheries, fish provenance, fisheries distribution, fisheries and SDG2 and SDG 14, are meant to access VLab products through e.g. web-portals, atlases, API's or QR codes.

6.2 State of art

High-resolution data in the fisheries domain (meaning at the fishing operation level) are often highly confidential, primarily due to economic and strategic interests. However, there is an increasing demand from both scientists and the public for fisheries data, information, and knowledge to be made accessible in order to address emerging scientific questions and support informed decision-making. Despite this, there is currently no global policy or EU directive enforcing the FAIR principles for fisheries data, nor is there a standardized fisheries data management framework comparable to the e.g. INSPIRE or Water framework directives, which mandate the FAIRness of spatial data by implementing OGC standards.

As a result, we used pre-harmonized public domain datasets that are reported by countries to the RFMOs at a lower spatio-temporal resolution (e.g. using grids at a 1° or 5° per month). These datasets are made freely available on RFMOs Websites but have not been assigned DOIs and unfortunately don't share the same structure. Indeed, unlike other domains, the fisheries sector lacks widely adopted *de facto* or international *standards* for data formats, which underscores the need for our VLab to harmonize the heterogeneous data structures provided by RFMOs. This is the reason why we promote the implementation of the CWP standards for fisheries domain data and the reason why we have to focus on specific use cases like tuna fisheries as we can't harmonize all fisheries datasets worldwide in the framework of a single project.

6.3 Input data sources

This VLab uses four different sources to build the **Global Record of Stocks and Fisheries (GRSF)** knowledge base and some of the public domain datasets provided by the five tuna **Regional Fisheries Management Organisations (RFMOs)**. The main characteristics of the data sources are described in the table below.

VARIABLES	DATA SOURCES	DATA ACCESS
GRSF	RAM legacy database	GRSF Knowledge Base
GRSF	FAO SDG 14.4.1 Questionnaire	GRSF Knowledge Base
GRSF	Fishsource	GRSF Knowledge Base
GRSF	FIRMS	GRSF Knowledge Base
Catch and effort	IOTC	Global Tuna Atlas: Zenodo
Catch and effort	ICCAT	Global Tuna Atlas: Zenodo
Catch and effort	IATTC	Global Tuna Atlas: Zenodo
Catch and effort	WCPFC Last data updated up to 2023: https://www.wcpfc.int/doc/annual-catch-estimates-2022-data-files	Global Tuna Atlas: Zenodo
Catch and effort	CCSBT- Southern bluefin tuna (SBF) ● annual catch data by year, flag or gear covering 1952-2023: https://www.ccsbt.org/userfiles/file/data/GlobalCatch_Flag_Gear.xlsx ● annual catch data by	Global Tuna Atlas: Zenodo

VARIABLES	DATA SOURCES	DATA ACCESS
	year, flag, and ocean covering 1965-2023: https://www.ccsbt.org/userfiles/file/data/CatchByOYF.xlsx ● However, for the GTA, we do need the annual catch data by year, flag, gear, and ocean covering 1952-2023 which was directly sent by the CCSBT colleague (dataset covering 1965-2021) last year ● geo-referenced catch and effort data for fisheries catching SBF:https://www.ccsbt.org/en/content/sbt-data.	

Table 9. VLab 5 Input data sources

6.4 Step by step guideline to use the VLab

Once logged in, VLab members can access the VLab content in different ways depending on their needs and profiles:

- **Advanced users**
  - can launch **RStudio server** where they will find ready to go R projects directly cloned from the underlying GitHub organization and related repositories. This is a key asset to foster the reproducibility or customization of the work. R projects include Shiny apps are easy to run and the data generation workflow can be executed from a single file.
  - can use the **GRSF API** or **SPARQL endpoint** to browse or extract specific content (Figure 51, 52). GRSF API provides a formal way to access GRSF data (i.e. stocks and fishery data, as well as supplementary resources such as marine species, water areas, fishing gears, etc.). The API provides a collection of RESTful methods supporting the retrieval of

resources using different combinations of parameters. Moreover, results can be delivered in various formats, such as JSON, CSV, XML, etc.
    - can query the content of the **Global Tuna Atlas relational database** (SQL, PostgreSQL / Postgis).
- The **general public** / wider audience can browse the content of the VLab knowledge base and related datasets by using menu items, e.g.:
    - **"CKAN catalogue" menu item:** to discover and access GRSF content with CKAN catalogue sheets. By leveraging CKAN facilities, GRSF resources are displayed as distinct datasets, grouped based on their distinctive attributes (e.g. grouped per FAO major area, per species, etc., Figure 50).
    - **"Shiny apps" menu item:** provides a list of Shiny apps built on top of the Global Tuna Atlas (Figure 46,47). Current apps demonstrates how atlas can easily be set up from FAIR fisheries data:
        - high resolution fisheries data complying with *Darwin Core* data format
        - lower resolution (gridded) fisheries data complying with ***Coordinating Working Party on Fishery Statistics* (CWP)** data format
    - **"viewer" menu item:** gives access to alternate Map viewers built on top of the Global Fisheries Atlas data or knowledge base. These web mapping apps demonstrate how other atlas can be set up from FAIR **Open Geospatial Consortium (OGC)** compliant fisheries metadata and data:
        - Fisheries Atlas map viewer: to spatially query, browse and display Global Tuna Atlas data catches (Figure 48). The same application is also in production out of the VLab on the **FIRMS Website** .

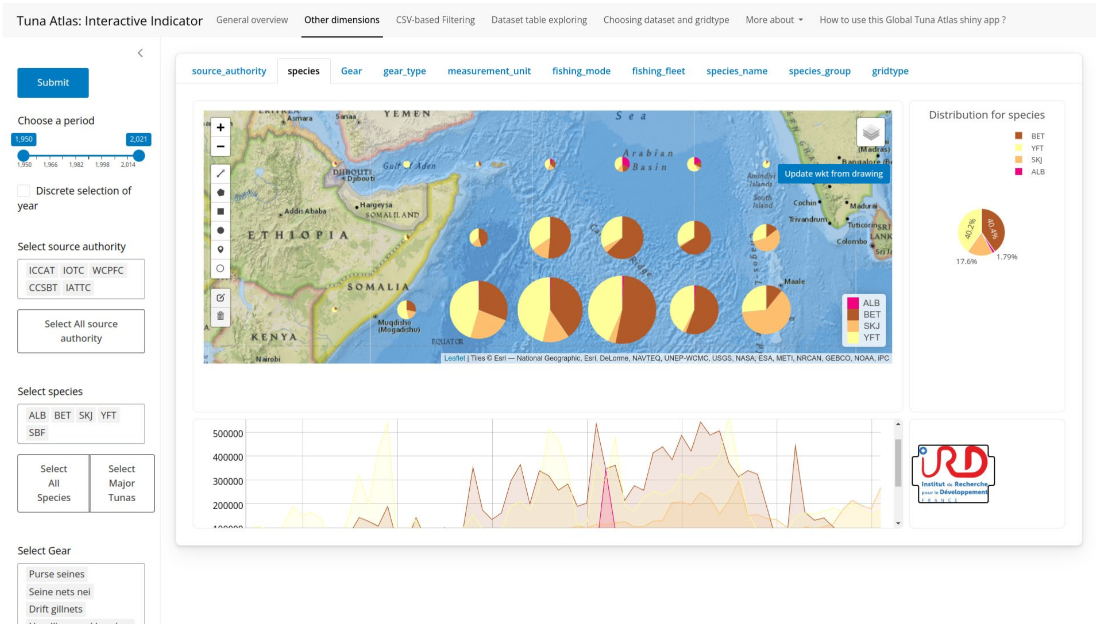


Figure 46. Fisheries Atlas shiny app for global gridded data, the example of Tuna Fisheries.

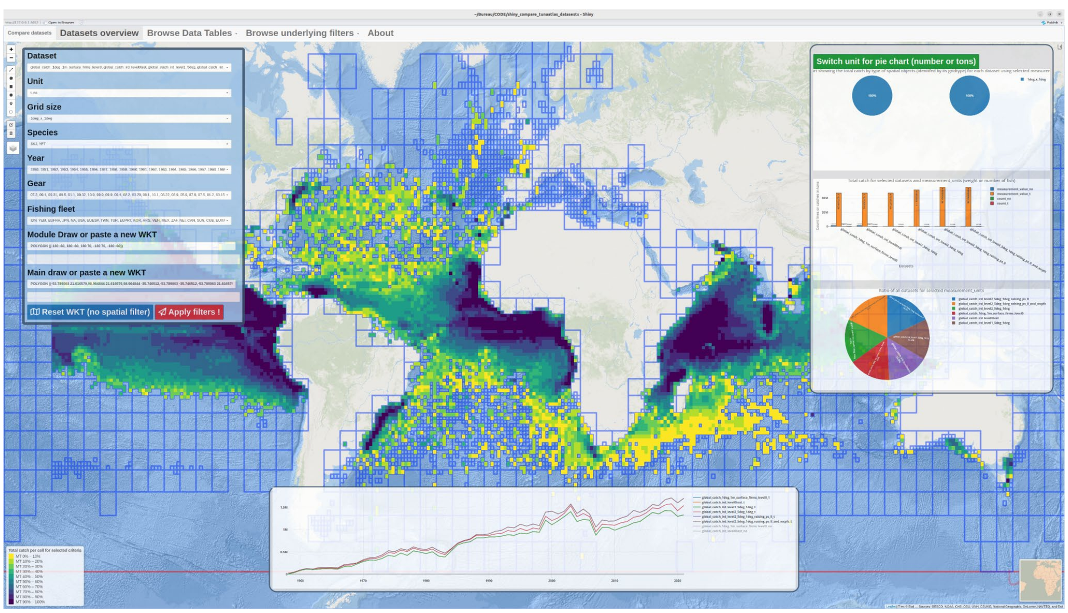


Figure 47. Fisheries Atlas shiny app to compare global gridded fisheries datasets with different levels of processing

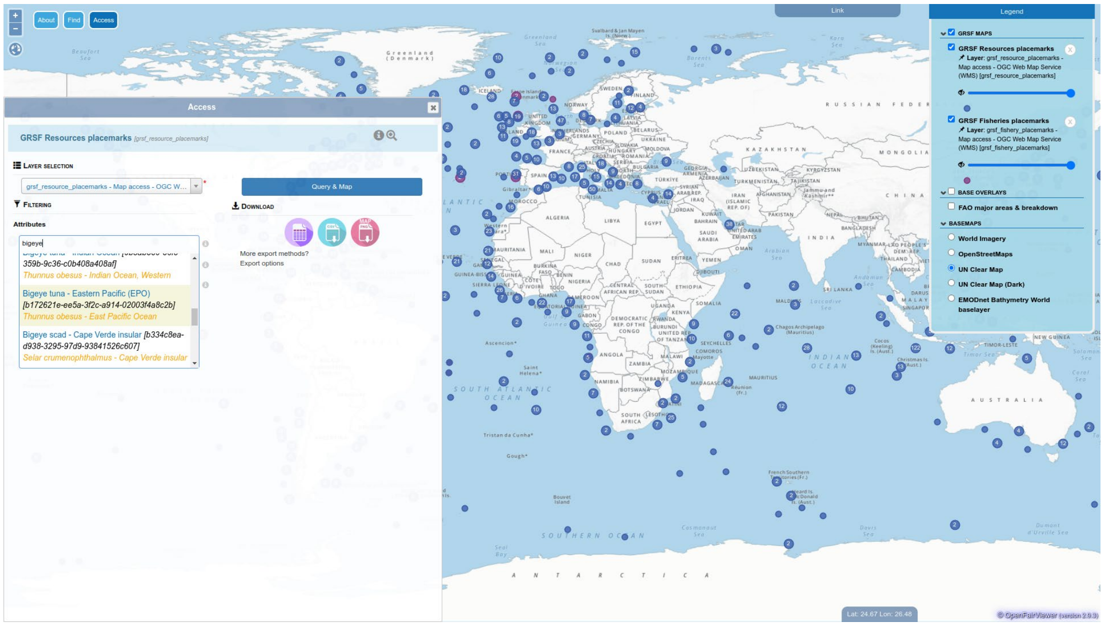


Figure 48. GRSF map viewer to discover, query and display Global records of stocks and Fisheries.

6.5 Using other data sources

The R workflow which generates the Global Tuna Atlas can easily be reproduced or customized by using new input datasets as long as they can be harmonized to comply with similar data structures (implementation of the CWP data format for fisheries data). This harmonisation process can be managed by reusing and adapting the code used to transform the various data structures of the five tuna RFMOs datasets. Once CWP compliant, the data generation workflow can easily use the new data by adding a new corresponding line in the CSV file of the geoflow R package which orchestrates the different steps and sequence of actions executed by the workflow.

6.6 Authors

The list of authors involved in this VLab is the following:

- **IRD**: Julien Barde, Bastien Grasset
- **FORTH**: Yannis Marketakis
- **CNR**
- **FIRMS**: FAO and tuna RFMOs data managers (IOTC, ICCAT, IATTC, WCPFC, CCSBT).

6.7 References

Part of the FAIR data and code generated by the Global Fisheries Atlas can also be found out of the Vlab in widely used repositories.

It is worth mentioning here some documents describing the first versions of Global (Tuna) Fisheries Atlas delivered first by FAO and IRD:

- IRD: Fonteneau Alain, Joseph J. (pref.). (1997). Atlas des pêcheries thonières tropicales : captures mondiales et environnement = Atlas of tropical tuna fisheries : world catches and environment. Paris : ORSTOM, 191 p. ISBN 2-7099-1370-4. https://www.documentation.ird.fr/hor/fdi:010012425
- FAO : see bibliography https://www.fao.org/fishery/en/collection/firms-tuna-atlas
- Building a Global Aquatic Resources Knowledge Base for Fisheries, Y. Marketakis, Y. Tzitzikas, A. Gentile, A. Ellenbroek, M. Taconet, 11th International Conference on ICT in Agriculture, Food & Environment Conference (HAICTA 2024), October 17-20, 2024, Samos, Greece, 2024

Datasets on Zenodo (Figure 49):

- All FIRMS Global Tuna Atlas products (Level 0):
  - Level 0 Nominal catch: Group, F. G. T. A. T. W. Global annual catches from tuna fisheries (1918 - 2021) (FIRMS level 0), 10.5281/zenodo.11410529 (2024).
  - Level 0 spatial catch: Group, F. G. T. A. T. W. Global monthly catches from tuna fisheries by 1° and 5° grids (1950-2021) (FIRMS level 0), 10.5281/zenodo.11460074 (2024).
- Upper levels of processing
  - Bastien Grasset, Julien Barde, & Paul Taconet. (2024). Global monthly catch of tuna, tuna-like and shark species (1950-2021) by 1° or 5° squares (IRD level 2) [Data set]. Zenodo 10.5281/zenodo.1164127 (2024).
  - Others to come: efforts, Level 1..

Code on Zenodo (DOIs being assigned to the main releases of repositories of the FIRMS Global Tuna Atlas GitHub page):

- Data generation workflow : https://doi.org/10.5281/zenodo.11563961
- Shiny apps:
  - Fisheries Atlas for CWP gridded : https://doi.org/10.5281/zenodo.12516879
  - Pending:
    - Fisheries Atlas for high resolution fisheries data (Darwin Core data format)

- Posters:
    - AI : Barde, J., Meneses, R., & Pittonet, S. (2024). Fisheries data to train Artificial Intelligence Models in predicting global fisheries distribution. Zenodo. https://doi.org/10.5281/zenodo.13305866
- Conference:
    - International Conference on Marine Data and Information Systems - Proceedings Volume. Miscellanea INGV, 80. https://doi.org/10.13127/MISC/80:
        - Extended abstract: Barde Julien, Blondel Emmanuel, Grasset Bastien, Workflows for marine metadata and data management, https://doi.org/10.13127/misc/80/7
        - Oral presentation: https://zenodo.org/records/14198690
- IOTC Working Party on Data Collection and Statistics (WPDCS19)

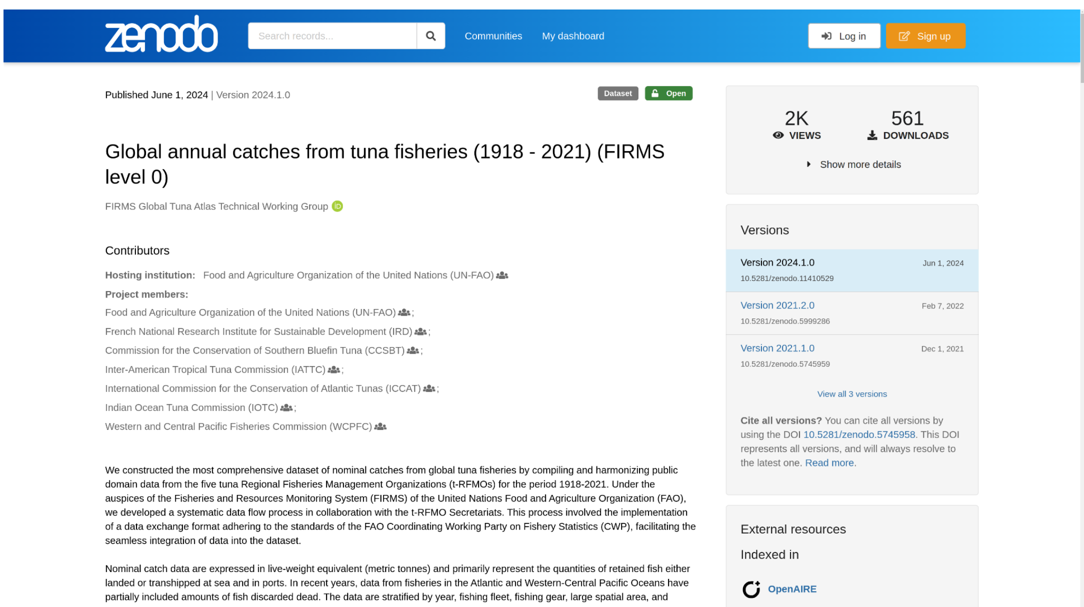


Figure 49. Example of Zenodo DOI where users can download a Global Fisheries Atlas data product.

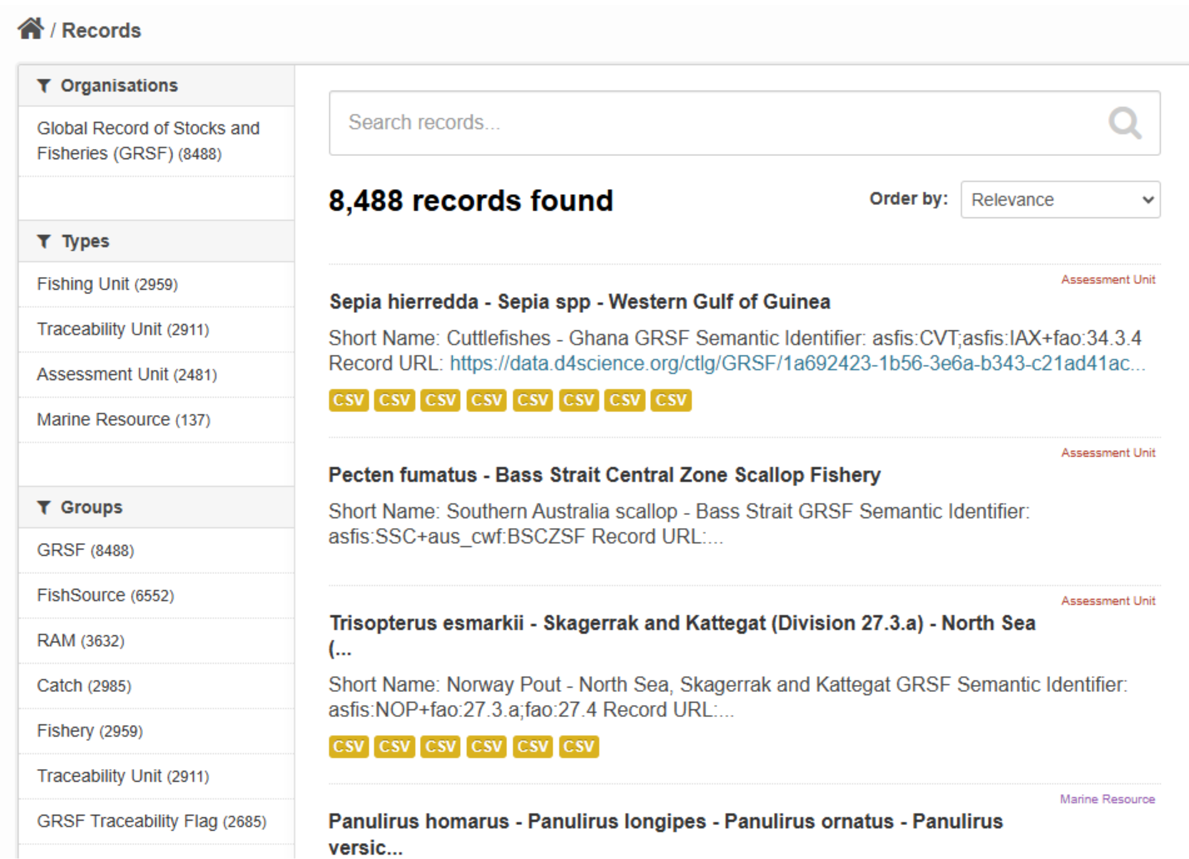


Figure 50. The Global Record of Stocks and Fisheries (GRSF) knowledge base is made accessible with UI generated by a CKAN catalog.

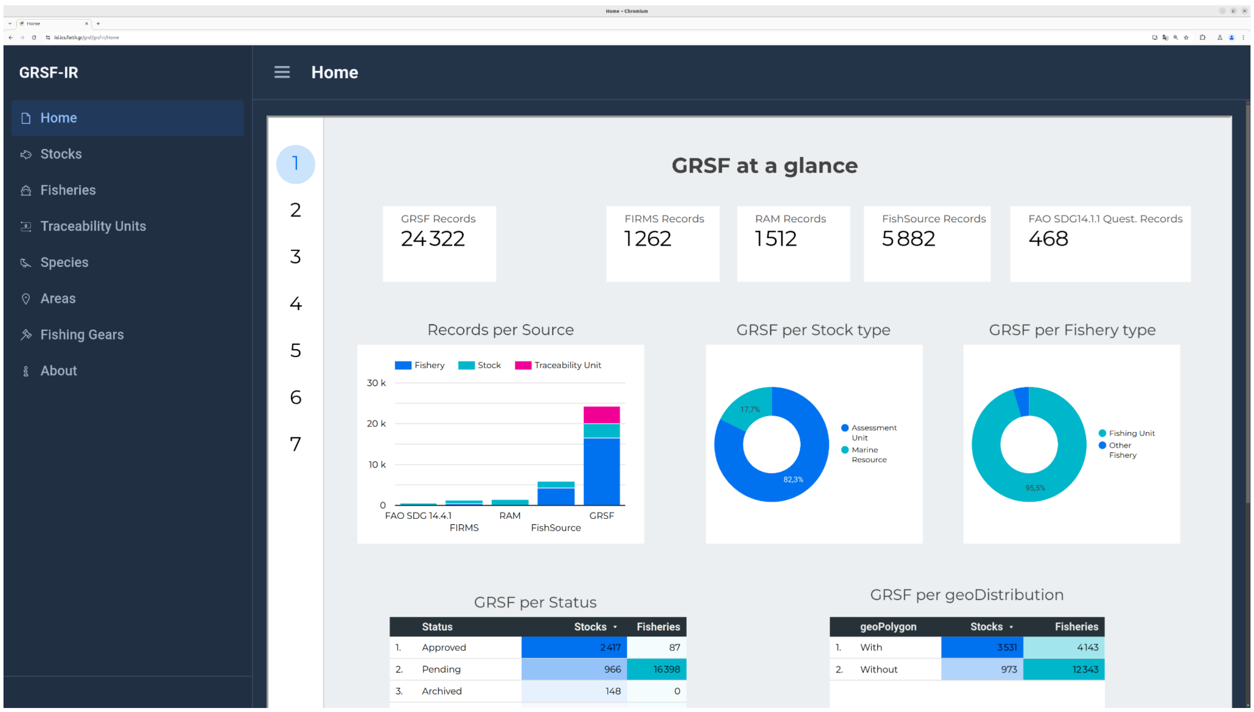


Figure 51. The Global Record of Stocks and Fisheries (GRSF) knowledge base is made accessible with a Dashboard

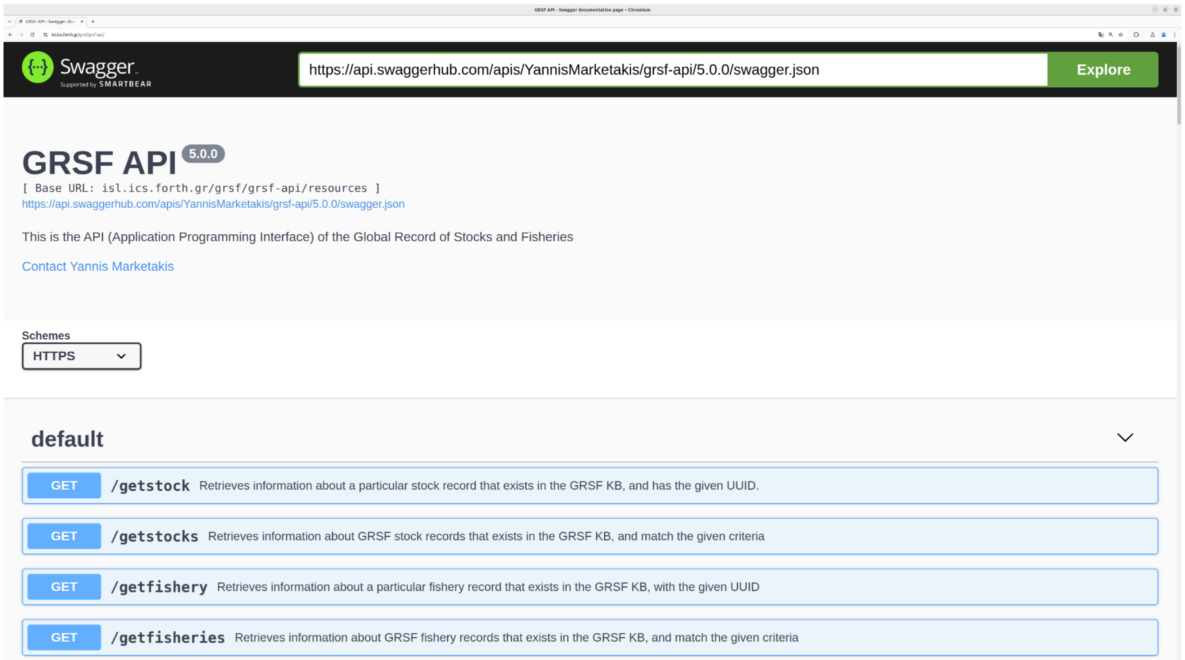


Figure 52. The Global Record of Stocks and Fisheries (GRSF) knowledge base is made accessible with a dedicated API.