D3.3 - Blue Cloud Demonstrator Users Handbook V1

Work Package	WP3, Blue Cloud Pilot Demonstrators
Lead Partner	IFREMER
Lead Author (Org)	NYS Cécile (OCEANSCOPE – IFREMER) MAUDIRE Gilbert (IFREMER)
Contributing Author(s)	AUGOT Jérémy (CLS), BARDE Julien (IRD), BLONDEL Emmanuel (FAO), CABRERA Patricia (VLIZ), COCHRANE Guy (EMBL), DEBELJAK Pavla (CNRS), DRUDI Massimiliano (CMCC), ELLENBROEK Anton (FAO), GENTILE Aureliano (FAO), LAVERGNE Emeric (CLS), LECCI Rita (CMCC), MARKETAKIS Yannis (FORTH), PESANT Stéphane (EMBL), SCHEPERS Lennert (VLIZ), TYBERGHEIN Lennert (VLIZ)
Reviewers	Pasquale Pagano (CNR); Dick M.A. Schaap (MARIS)
Due Date	30.11.2020, M14
Submission Date	17.12.2020
Version	0.21

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Blue-Cloud - Piloting Innovative services for Marine Research & the Blue Economy - has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement n. 862409.

DISCLAIMER

“Blue-Cloud, Piloting Innovative services for Marine Research & the Blue Economy” has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement n.862409.

This document contains information on Blue-Cloud core activities. Any reference to content in this document should clearly indicate the authors, source, organisation, and publication date.

The document has been produced with the funding of the European Commission. The content of this publication is the sole responsibility of the Blue-Cloud Consortium, and it cannot be considered to reflect the views of the European Commission. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

COPYRIGHT NOTICE

VERSIONING AND CONTRIBUTION HISTORY

Version	*Date*	*Authors*	*Notes*
0.1	16.10.2020	Cécile NYS (OceanScope/IFREMER)	Layout
0.2	20.10.2020	Cécile NYS (OceanScope / IFREMER) & Gilbert MAUDIRE (IFREMER)	Structure of the demonstrator part
0.3	29.10.2020	Anton Ellenbroek (FAO)	Added draft of demonstrator 4
0.4	16.11.2020	Patricia CABRERA (VLIZ)	Input of demonstrator 1
0.5	16.11.2020	Massimiliano DRUDI (CMCC)	Input of demonstrator 3
0.6	17.11.2020	Anton ELLENBROEK (FAO)	Added draft of demonstrator 5
0.7	17.11.2020	Stéphane PESANT (EMBL-EBI) & Pavla DEBELJAK (CNRS)	Input of demonstrator 2
0.8	17.11.2020	Cécile Nys (OceanScope/IFREMER)	Review, add legends, harmonize layout
0.9	18.11.2020	Cécile NYS (OceanScope / IFREMER) & Gilbert MAUDIRE (IFREMER)	Add executive summary, introduction and conclusion
0.10	18.11.2020	Cécile Nys (OceanScope/IFREMER)	Review and optimization text and structure, add legends, harmonize layout.
0.11	18.11.2020	Cécile Nys (OceanScope/IFREMER)	Major hierarchy change in the titles and organization of document
0.12	19.11.2020	Cécile Nys (OceanScope/IFREMER)	Add the internal reference in the document. Version submitted for review to WP3 partners and contributors.
0.13	23.11.2020	Massimiliano DRUDI (CMCC)	Revision by demonstrator 3 partners
0.14	23.11.2020	Anton ELLENBROEK (FAO)	Revision by partners of demonstrators 4 and 5.
0.15	23.11.2020	Patricia CABRERA (VLIZ)	Revision by demonstrator 1 partners
0.16	02.12.2020	Pasquale Pagano (CNR)	Project Review
0.17	07.12.2020	Anton Ellenbroek (FAO)	Additional input from demonstrators 4 and 5
0.18	09.12.2020	Dick M.A. SCHAAP (MARIS)	Project Review
0.19	10.12.2020	Massimiliano DRUDI (CMCC)	Additional input from demonstrator 3
0.20	14.12.2020	Patricia CABRERA (VLIZ)	Additional input from demonstrator 1
0.21	14.12.2020	Cécile NYS (OceanScope/IFREMER)	Concatenation & Review of corrections
0.22	15.12.2020	Pasquale Pagano (CNR)	Reviewer Approval
0.23	15.12.2020	Dick M.A. SCHAAP (MARIS)	Reviewer Approval
1.0	17.12.2020	Cécile Nys (OceanScope/IFREMER)	Final version

Contents

Table of illustrations

Executive summary

This deliverable, D3.3 “Blue Cloud Demonstrator Users Handbook V1” is a first version of the handbook and guidelines on how to use the Virtual Research Environments (VREs) of the five demonstrators. For each demonstrator, the following information will be described: the objective, the targeted users, a summary of their provided services and the guidelines of how to use their respective VREs.

As this version of the Handbook describes the β-version of the VREs, it is not fully complete and final, as the VREs are not fully finalised yet. It must be considered as an intermediate version to be used internally for integration and testing. This Handbook V1 is an ongoing and living document. It will be updated following advances of the demonstrators and their VREs.

The final version V2 of this handbook will be published in M27 of the Blue-Cloud project (December 2021) and will be open to users. Intermediate versions in between now and then will be written and possibly made accessible online, following progress of demonstrators and their VREs, deploying functionalities, achieving results, etc.

1 Introduction

The Blue-Cloud innovation potential will be explored and unlocked by developing five dedicated Demonstrators as Virtual Labs in a Virtual research Environment (VRE) together with excellent marine researchers. For that purpose, Blue-Cloud selected five varied and domain-coverage rich scientific demonstrators:

- **Demonstrator #1 – Zoo- and Phytoplankton EOV products.**
  This demonstrator aims to produce phytoplankton, zooplankton and nutrients EOV products. They will contribute to improve knowledge and vastly reduce uncertainty regarding the present state of the marine plankton ecosystems and their response to ongoing and future climate change.
- **Demonstrator #2 – Plankton Genomics.**
  This demonstrator aims to highlight a deep assessment of plankton distributions, dynamics and fine-grained diversity to molecular resolution by working across biomolecular, image and environmental data domains.
- **Demonstrator #3 – Marine Environmental Indicators.**
  This demonstrator aims to calculate and distribute online information and indicators on the environmental quality of the Mediterranean Sea, which will serve intermediate users such as environmental protection agencies. Tests will be conducted to extend the geographical area to the North-East Atlantic and demonstrators 1, 4 and 5, will use the results.
- **Demonstrator #4 – Fish, a matter of scales.**
  This demonstrator aims to expand data management and analytic capabilities for fisheries by:
  - Expanding the existing Virtual Lab for the FAO Tuna Atlas (tuna and billfish catch data) into the Fisheries Atlas. The Atlas should have features for data analysis (using indicators, interactive maps or dashboards, state-of-the-art analytical models);
  - Expanding the existing Global Record of Stocks and Fisheries (GRSF) Virtual Lab with new stocks and include and/or link to the results of approved status assessments of fisheries, including those from the Fisheries Atlas and other demonstrators.
- **Demonstrator #5 – Aquaculture Monitor.**
  This demonstrator aims at developing the remote sensing data capacity for the monitoring of aquaculture in marine cages and in coastal areas. The ambition is to deliver a tool to produce online national aquaculture sector overviews using OGC compliant data services also accessible through the Demonstrator #4 integrated Open FAIR Viewer.

The work to be undertaken by demonstrators will be conducted in four steps:

- Step 1: Identifying the requested technical requirements. **DONE**
  The Demonstrators will be analysed in detail with the involved researchers in order to describe their workflows and to use these for designing the Virtual Labs. This is done from the perspective of the Demonstrators.
- **Step 2: Implementing the first version of the demonstrators.**
  This first version will be based on pre-existing services, but implemented in a standardized common environment and relying, as far as possible, on common components to be set up mainly by WP2 and WP4. While developing, much interaction will take place between the researchers (WP3) and the technical developers (WP2 & WP4) in order to test and fine tune prototypes. ➔ Handbook V1
- Step 3: Second version of the demonstrators.
  Focusing on innovative development for the second versions of the Demonstrators, as more Blue-Cloud functionality will become available in WP2 and WP4. ➔ Handbook V2
- Step 4: Promoting and demonstrating the achieved demonstrators.
  The final phase leading up to the project completion will be dedicated to promote and to demonstrate the achieved demonstrators at different events which will be organized as part of WP5. Maintenance and support to the users will also be supported.

This document, D3.3 “Blue Cloud Demonstrator Users Handbook V1” reflects work and accomplishments of step 2. It summarises to the users the thematic services and Virtual Labs that have been set up by the demonstrators.

2 Demonstrator # 1 – Zoo- and Phytoplankton EOV products

2.1 Objectives of demonstrator

The zoo-phytoplankton Essential Ocean Variables (EOV) demonstrator aims to provide a methodology to generate:

1) **zooplankton products** based on in situ observations of abundance of different zooplankton species in a region encompassing the North East Atlantic;
2) **global ocean three-dimensional (3D) products of chlorophyll-a** (Chla) concentration, that is a proxy for total phytoplankton biomass, based on Argo vertical profiles matched up with satellite imagery;
3) a **mechanistic model** using near real-time data to quantify the relative contributions of the bottom-up and top-down drivers in phytoplankton dynamics.

2.2 Targeted users

The phyto-zooplankton EOV demonstrator provides a **description** of the **current state** of the **plankton communities** and **forecasts** their evolution, representing valuable information for the modelling, assessment and management of the marine ecosystem.

For example, ***fisheries advisory organisations*** can use these plankton products to study the availability of food resources for fish stocks and assess the effects on fish stocks. This knowledge will help the ***marine policy officers*** to address threats such as food insecurity, as foreseen under the EU Biodiversity Strategy for 2030. Moreover, the proposed EOV products are of interest for ***fundamental research*** (e.g. researchers and consultants from environmental agencies) contributing to the understanding of the environmental conditions and top-down factors at new scales of observations (e.g. regional/global, seasonal and time series).

2.3 Necessary data sources & Blue-Cloud (VRE) services used

The following table (Table 1) contains the necessary data sources for the demonstrator. At this stage, the Scientific Validation is the only part of the demonstrator that does not require to be pre-processed outside the VRE.

Table 1. Data sources needed for demonstrator 1 (Zoo- and phytoplankton EOV products).

	Variable	Source	URL
ZOOPLANKTON PRODUCT	Zooplankton abundances	EurOBIS	https://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=216
	Bathymetry	GEBCO	https://www.gebco.net/
	Distance to Nearest Coastline: 0.01-Degree Grid	GSFC, NASA	https://oceancolor.gsfc.nasa.gov/docs/distfromcoast/
PHYTOPLANKTON PRODUCT	Satellite-derived reflectance	CMEMS	OCEANCOLOUR_GLO_OPTICS_L3_REP_OBSERVATIONS_009_086
	Sea Level Anomaly	CMEMS	SEALEVEL_GLO_PHY_L4_REP_OBSERVATIONS_008_047 product
	Physical data: Global ARMOR 3D products	CMEMS	CMEMS MULTIOBS_GLO_PHY_REP_015_002
	BGC-Argo Float NetCDF files (S-files)	Argo GDAC (Coriolis center)	ftp.ifremer.fr/ifremer/argo/; http://www.argo.ucsd.edu
	Satellite-derived Photosynthetically Available Radiation	GlobColour	ftp://ftp.hermes.acri.fr
	Bathymetry	GEBCO	https://www.gebco.net/data_and_products/gridded_bathymetry_data/
SCIENTIFIC VALIDATION	Zooplankton abundances	EMODnet Biology	https://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=4687
	Phytoplankton abundances	EMODnet Biology	https://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=4688
	Abiotic data (nutrients, PAR and temperature)	LifeWatch	http://rshiny.lifewatch.be/station-data/

2.4 Summary of provided services

2.4.1 Zooplankton EOVs

The main service offered in this prototype is a complete workflow using the DIVAnd software tool (Data Interpolating Variational Analysis in n dimensions) to create interpolated maps of zooplankton abundances. DIVAnd has been designed to interpolate sparse, in situ measurements onto a regular grid in an optimal way, considering constraints such as the presence of obstacles (coastlines, islands) or currents**. The service is provided as a set of Jupyter notebooks** that describe the full procedure to create the final, gridded products:

1) data reading;
2) choice of analysis parameters;
3) spatial interpolation;
4) creation of plots;
5) writing of netCDF file storing the results.

The main dataset for this service is the abundance data, consisting of a set of positions (longitude, latitude) and the bathymetry and distance to nearest coastline datasets used for the interpolation.

2.4.2 Phytoplankton EOVs

The phytoplankton EOV products are global 3D products using machine learning-based methods, following the method developed by *Sauzède et al. (2016)*. This methodology extends surface bio-optical properties to depth to create a 3D product. An artificial neural network (Multi-Layer Perceptrons, MLPs) is trained to produce a vertical distribution of Chla concentrations.

The MLPs consists of several layers: one input layer, one output layer and one or several hidden layers. Each layer is composed of neurons, which are elementary transfer functions that provide outputs when inputs are applied. The MLP retrieves the Chla associated with the total phytoplankton biomass and it is trained using Chla outputs from the Biogeochemical-Argo database (available from Coriolis data centre). The MLPs input layer is composed of three main components:

1) satellite-based inputs from CMEMS and GlobColour, such as the ocean colour remote sensing reflectance (Rrs) at five wavelengths, the photosynthetically available radiation (PAR) and the sea level anomaly (SLA);
2) depth-resolved physical properties from CMEMS ARMOR3D products, such as the mixed layer depth and components derived from the principal component analysis (PCA) of the vertical profiles of hydrological properties (Temperature and Salinity);
3) time (day of the year transformed in cycles) and geographical coordinates for the ocean colour and hydrological data.

2.4.3 Scientific validation

This service provides a workflow to run a mechanistic model analysis, using near real-time data to quantify the relative contributions of the bottom-up and top-down drivers in phytoplankton dynamics. The Nutrient, Phytoplankton, Zooplankton and Detritus (NPZD) model used in this demonstrator has been created by *Soetaert and Herman (2009).* **The workflow is provided in R markdown documents and uses the parallel computing on the VRE (Figure 1)**. It consists of

1) applying Generalized Additive Models (GAM) on available observed data (e.g. temperature, dissolved inorganic nitrogen, phosphate or silicate) to complete daily time series for the input data of the NPZD model;
2) run the NPZD Model to calculate and visualize phyto- and zooplankton abundances;
3) validate the NPZD model based on observed Chlorophyll-a and zooplankton abundances;
4) calculate and visualize the drivers that limit phytoplankton abundances (i.e. relative contributions), such as temperature, light and nutrients.

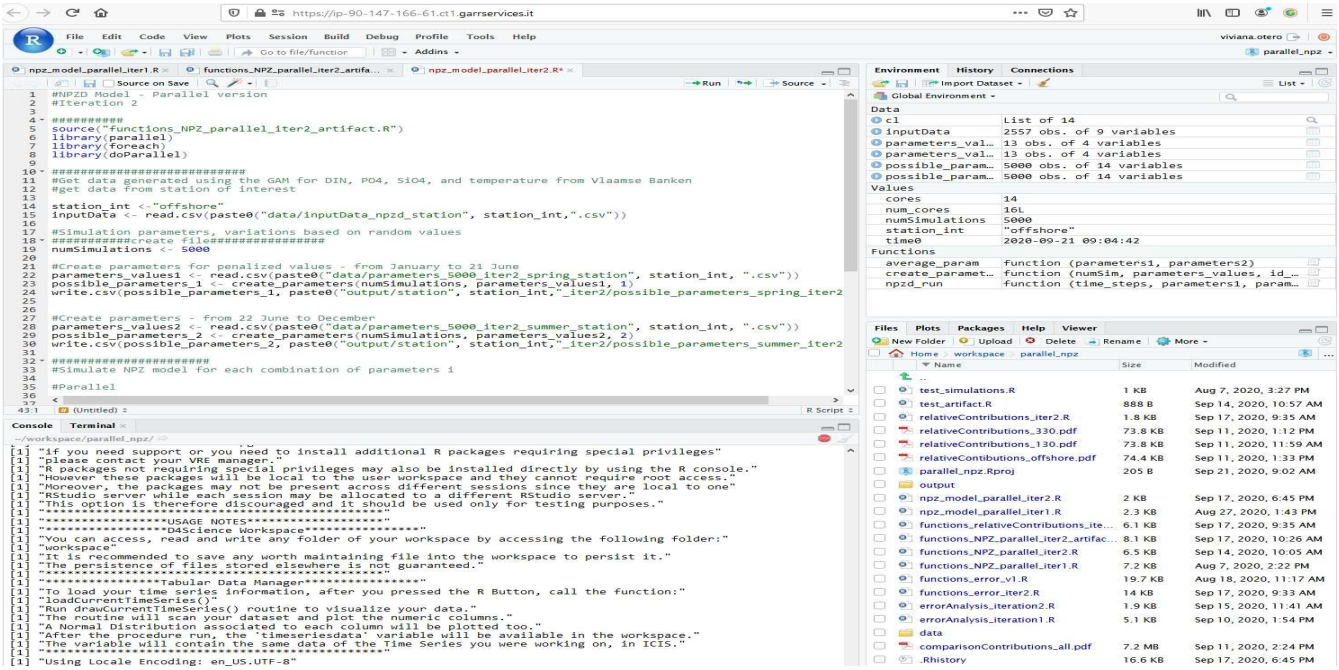


Figure 1. Parallel computing on the VRE to run the NPZD model.

2.5 Guidelines to use the services

This section includes the steps to use the services explained on the previous section and considerations to take into account. Users need to visit and register, to be able to run these services, on the following link:

https://blue-cloud.d4science.org/web/zoo-phytoplankton_eov .

2.5.1 Zooplankton EOVs

The process to run the tool DIVAnd using the "Docker Image Executor", in the Blue-Cloud Virtual Research Environment (hosted on the D4Science e-infrastructure) uses the Analytics Engine.

The steps to run the Docker Image Executor are explained in the documentation section “Zooplankton EOV docs” on the Zoo-Phytoplankton EOV Vlab. https://blue-cloud.d4science.org/group/zoo-phytoplankton_eov/zooplankton-eov-docs . The "Docker Image Executor" looks for the docker containers on the docker-hub (https://hub.docker.com/repository/docker/abarth/divand-bluecloud). Input files and results are transferred using the workspace.

2.5.2 Phytoplankton EOVs

All necessary files to calculate the Chla product are located in the "Zoo-Phytoplankton_EOV VRE"→ "VRE folders" → "Chla_Product", and consist of:

- "Inputs" folder: contains the input data to derive Chla for each month of a particular year (monthly ".nc" files).
- "Programs" folder: contains the scripts (Jupyter notebook) and models (trained MLP and PCA models).
- "Outputs" folder: contains the output files that will be generated as ".nc" files.
- "Plots" folder: contains the visualization of the outputs as ".png" or ".svg" files.

This workflow can be executed in the Jupyter Lab of the VRE, running the script on the Jupyter notebook. To reproduce this workflow with different data, users must have their inputs data in the same format as the data provided in the "Inputs" folder and change the paths on the Jupyter notebook script to read this data accordingly. Outputs and plots will be generated in the corresponding folders.

2.5.3 Scientific validation

At the moment, this is not publicly available in the VRE yet. The NPZD model has to be parametrized for 13 different parameters. The optimized parameters for the Belgian part of the North Sea (BPNS) are provided from 2014 to 2017. Nevertheless, if regional seas are considered with another biogeochemical cycling, consider that the model might need to be recalibrated. The code for this re-parametrization will be available in the R Markdown document. The input data (e.g. temperature, nutrients) are provided for the BPNS as daily time series. If other sites are studied where the temporal resolution of the abiotic conditions is low (not on a daily basis), it is necessary to use interpolation methods (e.g. GLM or GAM) to obtain a complete set of input data.

2.6 References

Sauzède, R., Claustre, H., Uitz, J., Jamet, C., Dall'Olmo, G., d'Ortenzio, F., Gentili, B., Poteau, A. and Schmechtig, C. (2016*). A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth: Retrieval of the particulate backscattering coefficient*. Journal of Geophysical Research: Oceans, 121(4), pp.2552-2571.

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.

3 Demonstrator # 2 – Plankton Genomics

The following information of demonstrator 2, concerns, at this moment, only the first Notebook of Plankton Genomics.

3.1 Objectives of demonstrator

The aim of the plankton genomic demonstrator is to showcase a deep assessment of plankton distributions by mining data across biomolecular, imaging and environmental domains. It will draw on the outputs of initiatives such as Tara Oceans[1] and will focus on two key objectives:

- ***Notebook 1 - Species and functions discovery*** – The demonstrator will enable the discovery of, as yet, not described biodiversity from genetic and morphological signals, and the characterization of their geographical distributions, co-occurrences/exclusions and correlation with environmental variables.
- ***Notebook 2 - Biodiversity and ecology*** – The demonstrator will enable the exploration of genetic and morphological markers of plankton diversity, in particular the ones discovered above, and predict their spatio-temporal distribution.

3.2 Targeted users

The initial users of the plankton genomics demonstrator are, primarily, scientific researchers, including taxonomists, computational ecologists and bioinformaticians with extensive knowledge of the data collected during the Tara Oceans Expedition. In the short term, we expect an important uptake of the demonstrator by European initiatives such as the H2020 Blue Growth project AtlantECO, the Ocean Sampling Day initiative, and the Marine Genomic Observatories in close collaboration with EMBRC and ASSEMBLE Plus.

The end-users include a broad base of scientists in quest of the identification of unknown sequences in the oceanic environment, and also interested, for example in plankton biogeography, marine biogeochemistry, ecosystem health, and climate science.

3.3 Necessary data sources & Blue-Cloud (VRE) services used

3.3.1 Data discovery & access

The demonstrator will enable scientific exploration of large datasets, including raw images and sequences, as well as taxonomic and functional annotations derived from images and sequences. This will require access to federated data infrastructures such as Elixir, EMODnet and Copernicus, as well as interoperability solutions such as standard provenance metadata and programming interfaces. The β-version will use data provided directly by the Blue Cloud partners.

[1] https://oceans.taraexpeditions.org/en/

3.3.2 Computing

The demonstrator will use computational methods to correlate unannotated, "unknown" raw imaging and molecular data with known taxonomies, morphologies and functions. It will also model the biogeography and co-occurrence of knowns and unknowns with environmental conditions. The first part will require access to considerable computational power at the European Bioinformatics Institute (EMBL-EBI), whereas the second part will be coded directly in the Blue Cloud VRE.

3.4 Summary of provided services

The plankton genomics demonstrator will consist of two Jupyter Notebook (Figure 2 & Figure 3) with R packages that allow users to

1) obtain lists of unknown taxonomies and functions,
2) correlate these unknowns with environmental parameters,
3) model the biogeography of unknowns using environmental climatologies from Copernicus, and
4) visualize these biogeographies on maps.

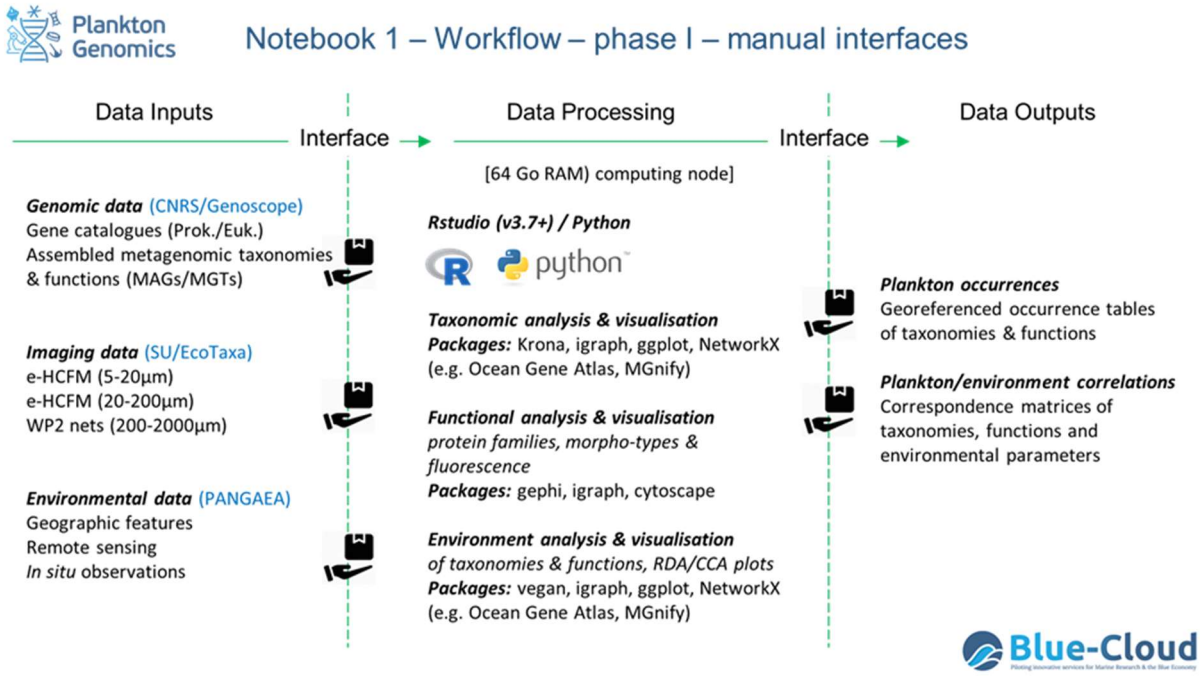


Figure 2. Notebook 1 - Species and functions discovery

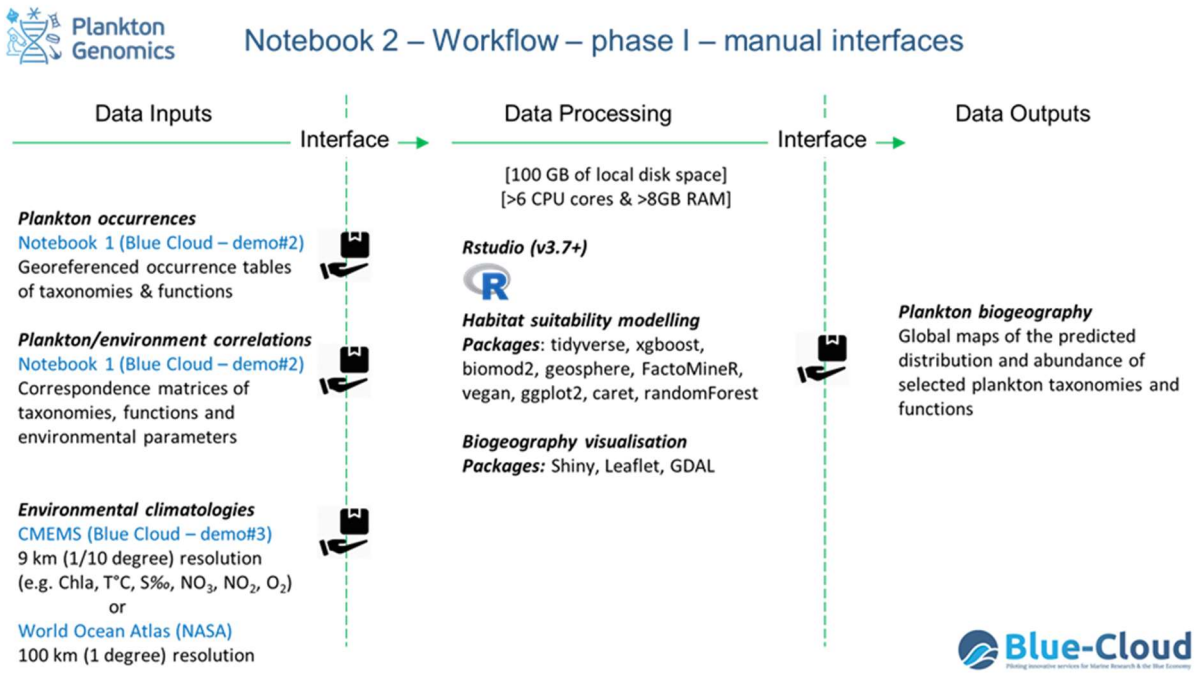


Figure 3. Notebook 2 - Biodiversity and ecology

3.5 Guidelines to use the services

The β-version of the plankton genomics demonstrator is under development and will be available in February 2021. No guidelines are available yet.

4 Demonstrator # 3 – Marine Environmental Indicators

4.1 Objectives of demonstrator

The objectives of demonstrator 3 – Marine Environmental Indicators – are to:

- Calculate and distribute online information and indicators on the environmental quality of the marine area.
- Obtain new added-value data applying big data analysis and machine learning methods on the multi-source data sets.
- Enable users to perform on line and on the fly operations such as selecting portion of a dataset, to perform statistical analysis or display the data.

4.2 Targeted users

The target audience of the service are ***intermediate users*** such as environmental protection agencies, and international stakeholders in the MSFD, in the UN SDG 14 and in the Blue Economy.

The service offers to them a flexible capacity to perform statistical analyses of the quality and characteristics of the marine environment for the Mediterranean Sea region, with possibility to scale in the next version up to the Global.

Attention is also dedicated to scientific users, providing to them a tool to facilitate the discovery of new climatic indicators based on machine learning, and a simplified way to analyse oceanographic data.

4.3 Necessary data sources & Blue-Cloud (VRE) services used

A subset of product MEDSEA_REANALYSIS_PHYS_006_004 in CMEMS catalogue is made available as a sample input dataset inside the VRE that the user can select from the web interface. The sample input dataset inside the VRE has the same format of the external data source (CMEMS).

For the notebooks, two examples of input datasets are available: a selection of monthly mean fields of GLOBAL_REANALYSIS_PHY_001_030 CMEMS product covering the Mediterranean Sea in 2018 and the same selection in 2017. Input dataset is downloaded from CMEMS using a CMEMS Motu client and saved in the user workspace. The code for downloading data is included in the notebooks. Some pre-trained models are also provided and can be used in the notebooks by the user.

ERA5 data from the Copernicus C3S Climate Data Store is used to retrieve hourly reanalysis 10 m wind data above sea. Using this source data, VRE processes the data to obtain daily and monthly SSI distributions for the ERA5 spatial area (from 1979 to present).

4.4 Summary of provided services

This version of the service provides the following features:

- a prototype MEI Generator app;
- notebooks for specific investigation related to the Ocean Patterns;
- Storm Severity Index (SSI) dataset.

4.4.1 Prototype MEI Generator app

The prototype MEI Generator app provides an interface that allows the user to generate new added-value data and to display data already available. To generate new data, the user can specify the desired type and field, then specify the additional parameters depending on the selected type of output. In this prototype version, the working domain is the Mediterranean Sea and the available input data are inside the time range 1987-1989. After the selection of all the expected input parameters, the user can submit the job. At its termination, the new output will be available for the display.

4.4.2 Ocean patterns indicator: workflow & notebooks

For the 'Ocean patterns indicator', the workflow (Figure 4) is structured in two notebooks: a model development notebook and a prediction notebook. In the model development notebook, the user will download a training dataset, parameter, optimize and train the model, and then save it in a NetCDF file. In the second notebook (prediction notebook), the user will upload the model generated in the first notebook, download the dataset to be predicted and plot the results. The figures and the dataset including the computed variables can be saved in user workspace, in a NetCDF file.

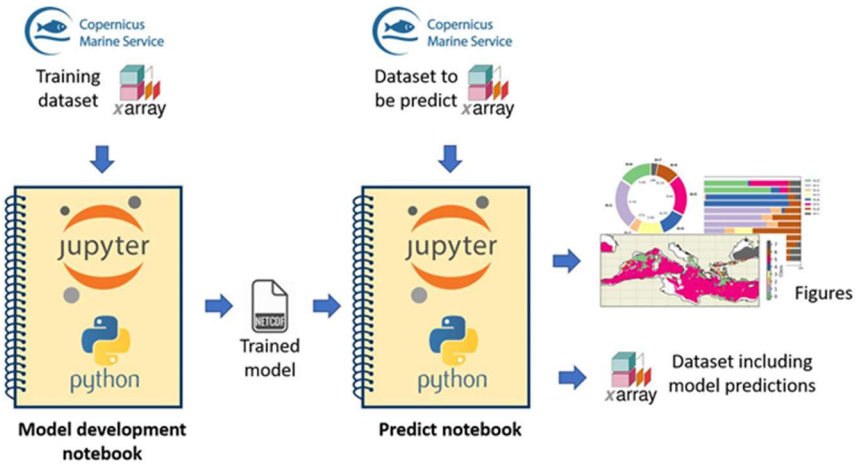


Figure 4. 'Ocean patterns indicator' workflow

4.4.3 Storm Severity Index

The Storm Severity Index (SSI) dataset provides users insights on atmospheric wind/storm circumstances that impact the circulation of seas such as the Mediterranean Sea. The user can combine this information with other marine environmental indicators for correlations.

Calculated SSI can also be related to individual storms or seasonal distribution across a spatial sea area for a longer period of time (i.e. the storm season for the Aegean Sea). Series of calculated SSI distributions over a period like 30 years can provide insight in changes of the storm climate of a sea region hence the sea circulation impact.

4.5 Guidelines to use the services

The internet link to reach the VLab of this demonstrator is

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

From the above link, new users can follow the registration procedure, while registered users can access the VRE and use the available services, according to the guideline provided in this paragraph.

4.5.1 Prototype MEI Generator app - Guidelines

The Web User Interfaces (UI) allows the user an easy and transparent way for accessing the processing methods available on the Data Miner service of the Blue-Cloud VRE. In fact, in order to obtain new data, the UI presents to the user all the available options which are needed for the customization of the processing.

The functionalities currently provided by the UI are summarized as follow:

1) Generating new data starting from the existing ones.
2) Access and visualize the data previously calculated by the user.

4.5.1.A Generating new data starting from the existing ones

The user can select and choose the following relevant info:

a) **Data source** (Figure 5 a)
   N.B.: At this moment of time, this is currently not modifiable yet and set, by default, with the product 'MEDSEA_REANALYSIS_PHYS_006_004, Mediterranean Sea Reanalysis Data of Physics' (a CMEMS product).
b) The **Type of data to generate** (Figure 5 b)
   i.e. monthly mean time series, annual climatologic map, etc.
c) The **Environmental field** (Figure 5 c)
   i.e. temperature, salinity, density, etc.
d) The **time range** (Figure 5 d)
   N.B.: This information is related to the type of data to be generated.
e) The **area** (of interest) in terms of boundaries (Figure 5 e)
   N.B. The points of minimum latitude and longitude and the points of maximum latitude and longitude build the area of interest. The chosen area (blue rectangle on map in Figure 5) is then visible on the map that is centred on the geographic domain of the selected data source.

f) The **depth** range (Figure 5 f).

After the user made his/her selection, he/she can click on the "Execute process" button (Figure 5 g) and the UI packages all the selections made by the user and send them to the Data Miner via a WPS (Web Processing Service) request.

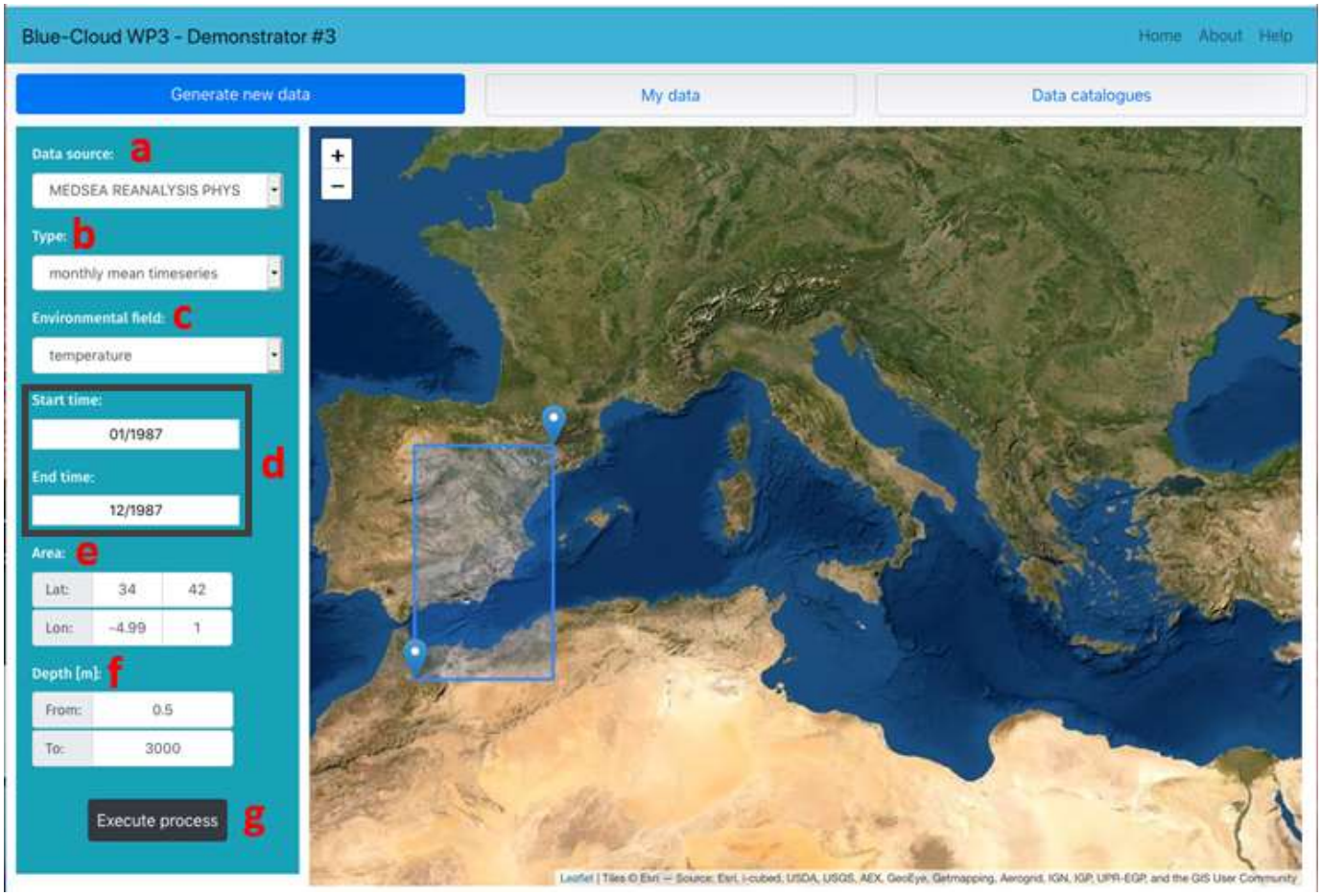


Figure 5. 'Generate new data' – User Interface (UI).

The user can select the data source (a), the type of data to be generated (b), the environmental field (c), the time range (d), the area of interest (e) and the depth (f). After selection, the user 'Execute process' (g) and the UI packages all the selections.

4.5.1.B Access and visualize the data previously calculated by the user

By selecting the "My data" tab (Figure 6) the user can see his/her last launched process together with all his/her previous requests. There the user can see and follow:

- the creation time for each process (Figure 6 a);
- the current status of their execution on DataMiner (Figure 6 b & c);
- the data source (Figure 6 d);
- the related parameters of the process (Figure 6 e);
- the area of interest (Figure 6 f);
- the depth range (Figure 6 g) and
- the time range (Figure 6 h).

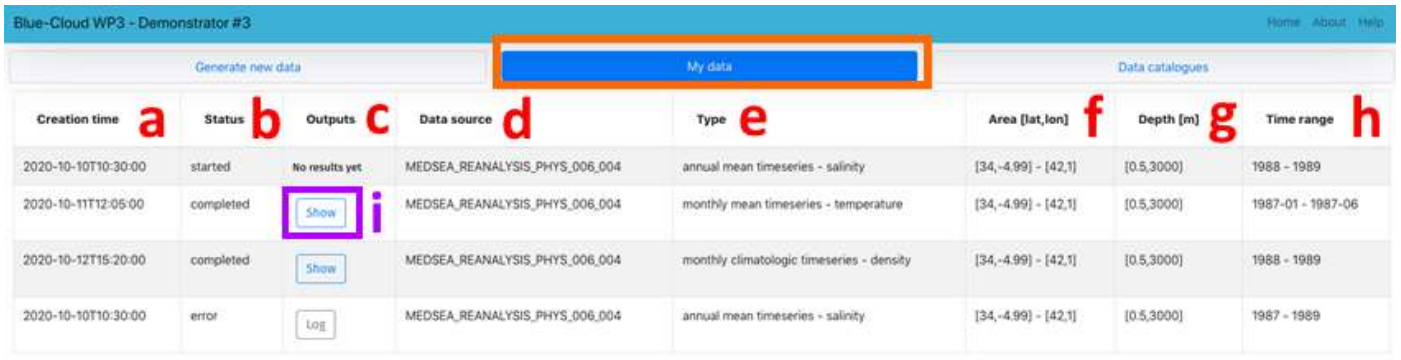


Figure 6. 'My data' – User Interface (UI).

On the 'My data' tab (box in orange) of the UI, the user can see all his/her requests (past, present and ongoing) with their detailed info : creation time (a), status of process (b & c), original data source (d), parameters used in process (e), area of interest (f), depth range (g) and time range (h). Once a process is finished (status = completed), the user can visualize the end-product by clicking on the 'Show button' (i).

When the status of the process is presented as "completed", then the user can visualize the data produced by the process by clicking on the "Show" button (Figure 6 i). The user then obtains the screen as showed below (Figure 7). The user can elect to download the image (Figure 7 a - Download Image), the NetCDF file of the data (Figure 7 b – Download Data) and eventually the execution log (Figure 7 c – Download Log) of the process.

Figure 7. 'Show result' in the 'My data' UI.

The user can download the illustration of the results (Download Image – a), download the produced data in a NetCDF file (Download Data – b) and/or download the execution logs of the process (Download Log – c).

4.5.2 Ocean patterns indicator – Guidelines

To use the Ocean Patterns notebooks, the user has to follow the workflow of the two following notebooks:

1) Development notebook
2) Predict and Plot notebook.

In order to access one of the notebooks, please follow the instructions:

1) Open a JupyterLab instance using the JupyterHub tab inside the VREVLab
2) Open a Terminal inside the JupyterLab - from the menu > File > New > Terminal
3) Inside the terminal, copy the archive, change the permission and unzip, with the commands:
    - cp workspace/VREFolders/MarineEnvironmentalIndicators/notebooks/OceanPatternsIndicator/OceanPatternsIndicator.zip .
    - chmod 777 OceanPatternsIndicator.zip
    - unzip OceanPatternsIndicator.zip
4) Open the desired notebook - from the menu > Open from Path, using the following path:
    - For the development notebook: OceanPatternsIndicator/Develop_PCM_model.ipynb
    - For the predict and plot notebook: OceanPatternsIndicator/predict_PCMlabels_and_plot.ipynb

4.5.2.A Development notebook

The user should carefully read the descriptions and follow the instructions as presented in the 'Development' notebook. This notebook is structured as follow:

a) Model parameters
User should provide the following parameters to design the model: the number of classes (Figure 8 a) and the name of the chosen variable (Figure 8 b).

b) Load training dataset
User should provide CMEMS account user and password (Figure 9 a). The training dataset is downloaded from CMEMS and saved in a NetCDF file in datasets/ folder in the workspace. (Figure 9 b). Finally, the training dataset is loaded locally using xarray library. (Figure 9 c).

c) Create and train model
User can create (Figure 10 a) and train (fit) the model (Figure 10 b) following the instruction in the notebook.

d) Development plots
The user can produce some plots that will be used as a guide to choose the best model parameters. (Figure 11)

e) Refit and save model
Model is trained again with the correct number of classes and saved as a NetCDF file in the models/ folder (Figure 12). It will be used as an input in the next notebook (See 4.5.2.B Predict and plot notebook).

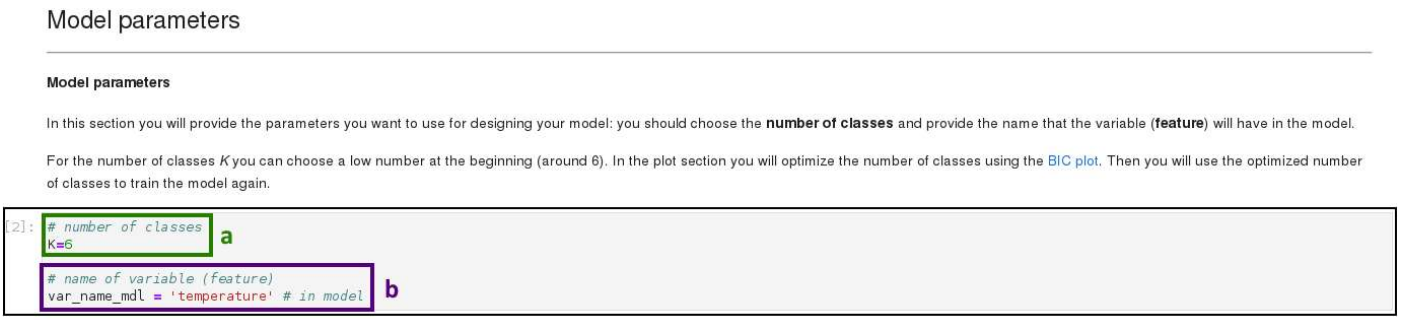


Figure 8. Development notebook – Model parameters.

In the parameters, the user can introduce the number of classes (a) to train the model in at first and the chosen variable (b) in which he wants to work.

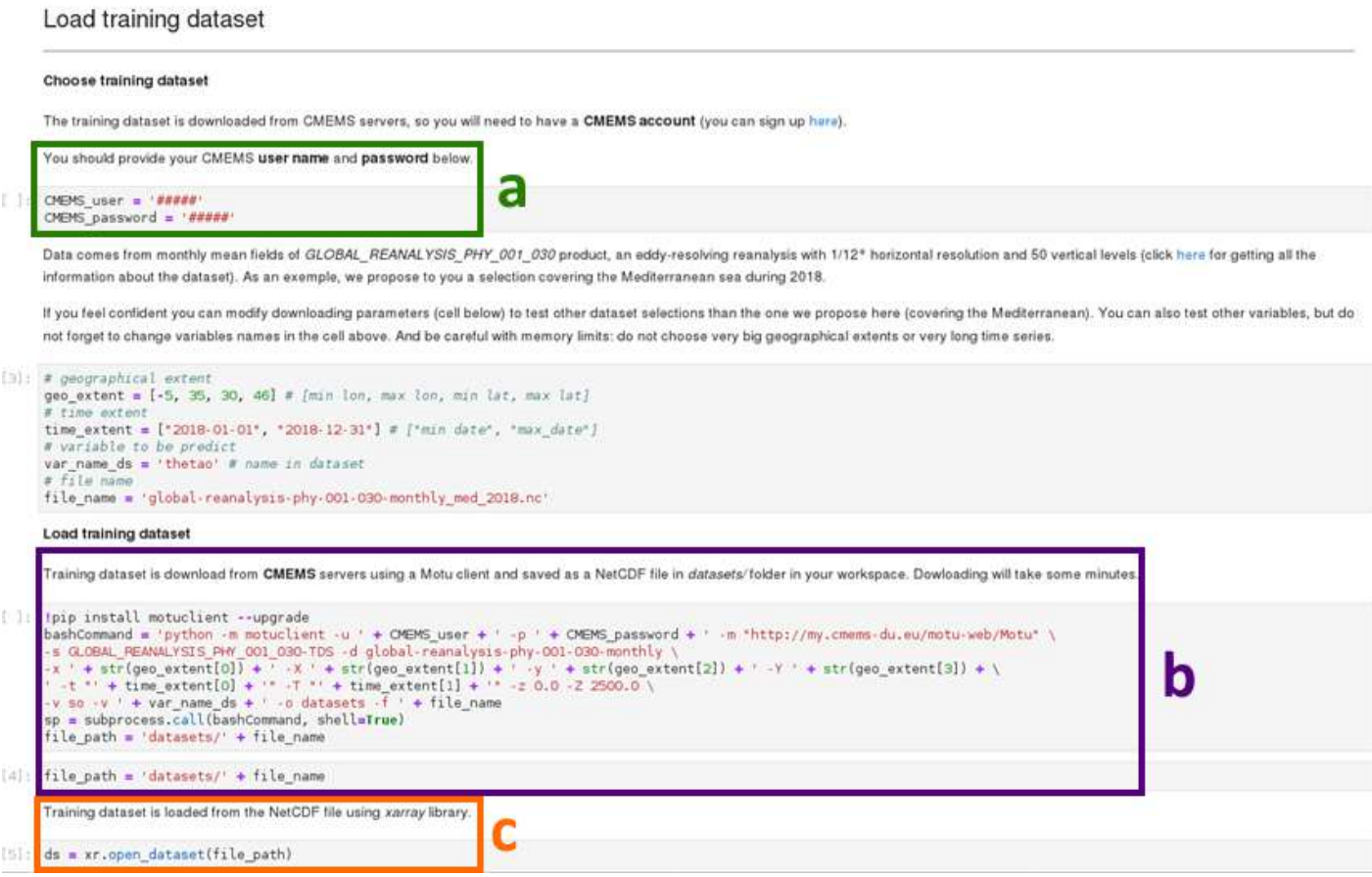


Figure 9. Development notebook – Load training dataset.

User provides his/her CMEMS account user and password (a). User can then download the training dataset from CMEMS in a NetCDF file (b) on the local workspace. Finally the training dataset is loaded locally using the xarray library (c).

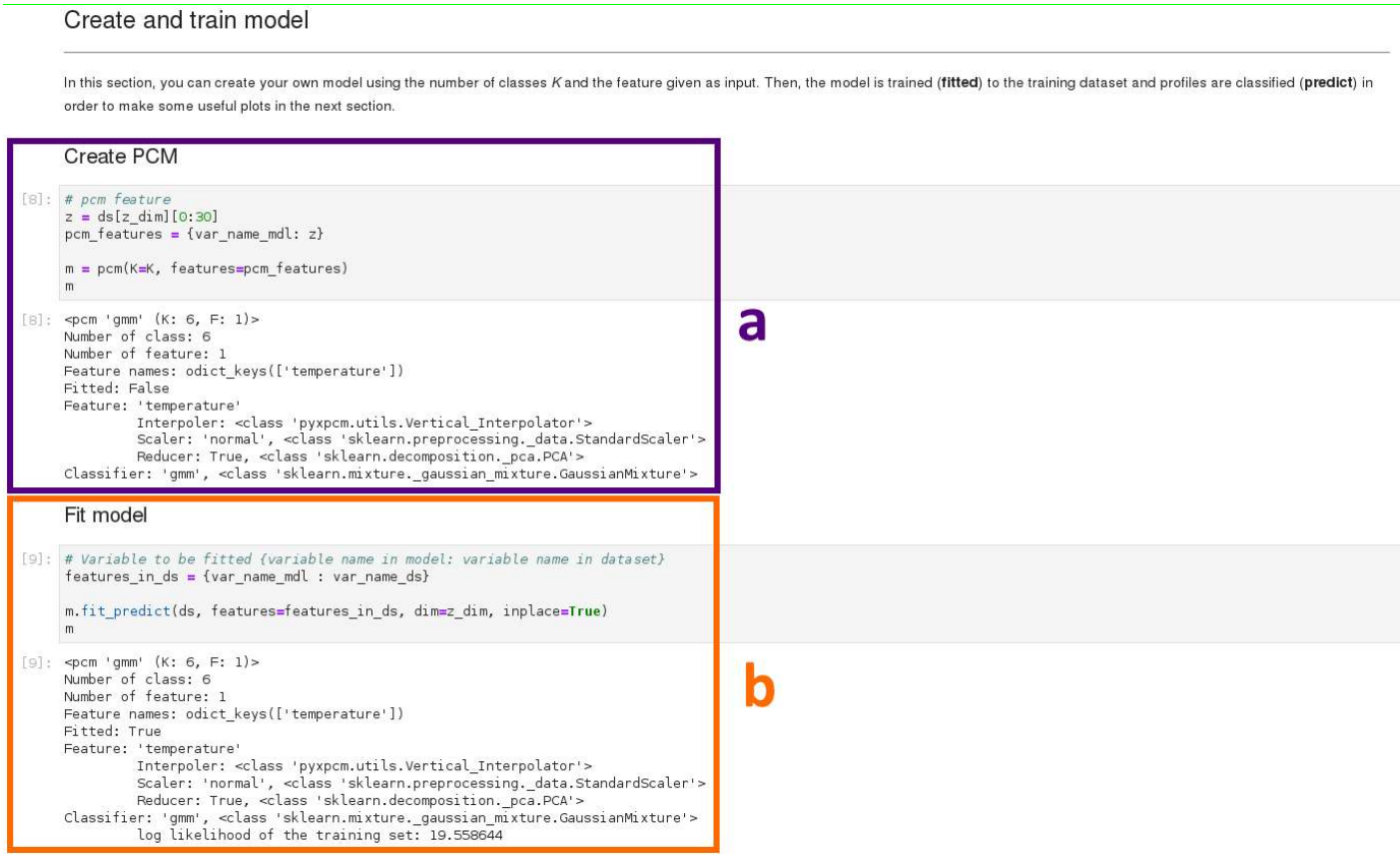


Figure 10. Development notebook – Create and train model.

User can create (a) and train/fit the model (b) following the instruction in the notebook (text in green preceded by a hashtag '#').

Development plots

The plots in this section will help you to **optimize** the model parameters (specially the number of classes) and to take a look on how the model is working.

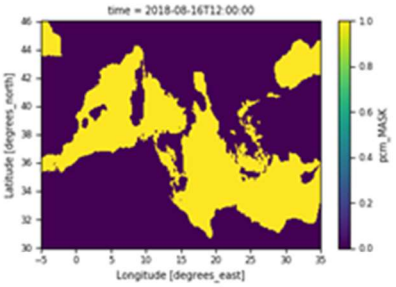
1. Mask

When fitting the model, the pyxpcm sorfware **preprocessed** the data to use profiles without NaN values. NaNs can be found when the profile is not depth enough, for example. Profiles with NaN values are masked and they are not used for fitting the model.

You can plot the mask below to know how is the dataset which is actually used to fit the model. As we are working with a time series, you should choose the time slide to be plotted.

```
[10]: mask = ds.isel(time=7).pyxpcm.mask(m, features=features_in_ds, dim=z_dim)
      mask.plot()
```

```
[10]: <matplotlib.collections.QuadMesh at 0x7f943edec310>
```



2. BIC

The BIC (Bayesian Information Criteria) is used to **optimize the number of classes** in the model, trying not to over-fit or under-fit the data. For calculating this index, the model is fitted to the training datas for a range of K values from 0 to 20, doing 10 runs each time to calculate the standard deviation. The **minimum** in the BIC curve will give to you the best number of classes to be used.

For each run, a sub-dataset of the training dataset is used, as profiles should not be correlated neither spatially nor temporally to get a minimum. Spatial and temporal correlation in the training dataset are user inputs that should be given below. For our example in the Mediterranean sea, values have been optimized. If you want to try another geographical selection or another variable you may change this numbers.

You can also choose the number of runs and the maximum number of classes, taking in to account that increasing this numbers will increase the computation time. Calculation is parallelized but it still take some minutes.

Figure 11. Development notebook – Development plots.

User can produce some plots that will be a guide to choose the best model parameters.

Refit and save model

As you know know which is the **best number of classes** to be used with your training dataset, you can train (fit) the model again with the good number of classes.

```
[ ]: # goog number of classes
     K = 8

     m = pcm(K=K, features=pcm_features)
     m.fit_predict(ds, features=features_in_ds, dim=z_dim, inplace=True)
```

b a

If you are happy with you model, you can save it in the *models/* folder and use it in the *predict_PCMlabels_and_plot.IPYNB* notebook to classify (predict) a dataset and plot the results corresponding to **ocean patterns indicators**.

```
[ ]: m.to_netcdf('models/test_model_mediterranean_temp_2018.nc')
```

c

Figure 12. Development notebook – Refit and save model.

Model is trained again (a) with the correct number of classes (b) and saved as a NetCDF file in the folder (c). It can be used as an input in the next notebook

4.5.2.B Predict and plot notebook

The user should carefully read the descriptions and follow the instructions as presented in the 'Predict and plot' notebook. This notebook is structured as follow:

a) Load model and dataset
The user should provide the CMEMS account user and password (Figure 9 a) and the path to the trained model generated in the first notebook (Figure 12 c).
User can also use one of the already trained models available in the folder (Figure 13 a). The input dataset is downloaded from CMEMS (a CMEMS user account is need – Figure 13 c) and the model is load from the NetCDF file (Figure 13 b).

b) Predict labels
Classes' prediction is done using pyxpcm library. User should follow notebook instructions (Figure 14).

c) Plot results
Results are plotted in different ways and figures are saved in the figures' folder on the user workspace (Figure 15).

d) Save data
User can also save the dataset including the new computed variables in a NetCDF file (Figure 16).

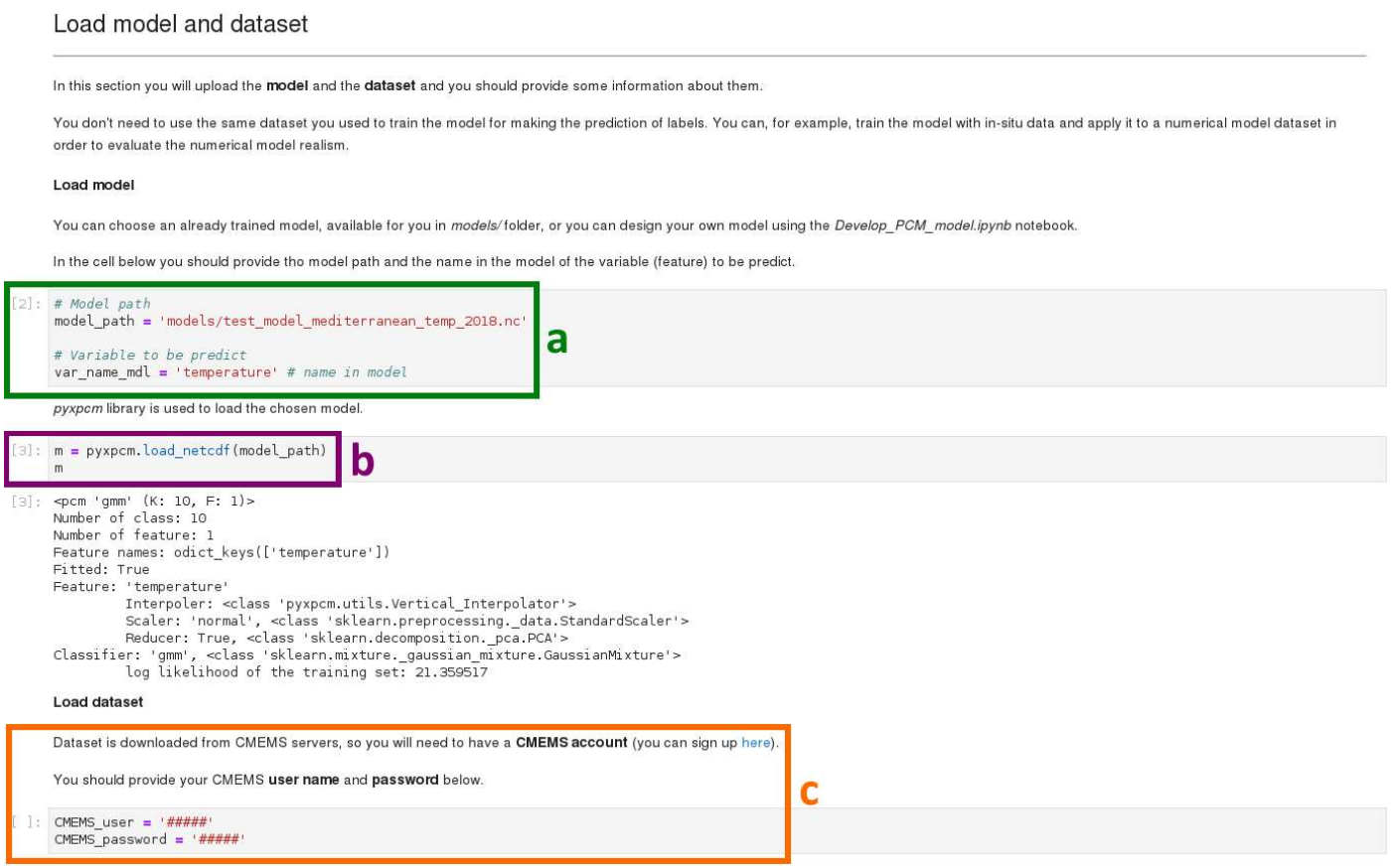


Figure 13. Predict & Plot notebook – Load model and dataset.

User has to provide the CMEMS account user and password (c). User can use one of the already trained models available in the folder (a). The input dataset is downloaded from CMEMS (CMEMS account needed) and the model is load from the NetCDF file (b).

Predict labels

Classes labels and some statistics are computed using *pyxpcm* library. New variables with the results are added to the dataset (`inplace=True` option).

Predict class labels

Taking into account the characteristics of the classes already determine in the trained model, each profile in the dataset is classified (**predicted**) into one of the classes. A new variable *PCM_LABELS* is created, including one class label for each profile.

```
[8]: features_in_ds = {var_name_mdl: var_name_ds}
m.predict(ds, features=features_in_ds, dim=z_dim, inplace=True);
```

Probability of a profile to be in a class (Posteriors)

As *pyxpcm* software is using a GMM (Gaussian Mixture Model) to determine clusters, it is possible to calculate the probability of a profile to belong to a class, also call **posterior**. This is the first step to determine the robustness of the model, that will be calculated below. A new variable *PCM_POST* is created.

```
[9]: m.predict_proba(ds, features=features_in_ds, dim=z_dim, inplace=True);
```

Classes quantiles

Class vertical structure can be represented using the quantiles of all profiles corresponding to a class. We advise to you to calculate at least the **median profile** and the 5% and 95% quantiles (`q=[0.05, 0.5, 0.95]`) to have a good representation of the classes, but feel free to add other quantiles if you want. A new variable `outname=var_name_ds + '_Q'` is added to the dataset.

```
[10]: ds = ds.pyxpcm.quantile(m, q=[0.05, 0.5, 0.95], of=var_name_ds, outname=var_name_ds + '_Q', keep_attrs=True, inplace=True)
```

Robustness

Robustness represents the **probability** of a profile to belong to a class, as posteriors, but the value range is more appropriated for graphic representation. Two new variables are added to the dataset: *PCM_ROBUSTNESS* and *PCM_ROBUSTNESS_CAT*.

```
[11]: ds.pyxpcm.robustness(m, inplace=True)
ds.pyxpcm.robustness_digit(m, inplace=True)
```

```
[11]: xarray.Dataset
```

▸ Dimensions: (**depth**: 41, **latitude**: 193, **longitude**: 481, **pcm_class**: 8, **quantile**: 3, **time**: 12)

▾ Coordinates:

pcm_class	(pcm_class)	int64	0 1 2 3 4 5 6 7
depth	(depth)	float32	-0.494025 -1.541375 ...
latitude	(latitude)	float64	30.0 30.08 30.17 ... 4...
time	(time)	datetime64[ns]	2017-01-16T12:00:00...

Figure 14. Predict & Plot notebook – Predict labels.

Plot results

Plots are created using the *Plotter* class, which is instantiate below. Plots include the vertical structure and the spatial and the temporal distribution of classes. These plots would allow you to determine if classes show a spacial or temporal coherence: the **ocean patterns indicators**.

save_BlueCloud function save the figure and add dataset information and logos below.

Please, feel free to change plot options if you need it.

```
[12]: P = Plotter(ds, m)
```

1. Vertical structure of classes

The graphic representation of quantile profiles reveals the vertical structure of each class and how clusters are created, as the clustering method is base on finding similarities in the vertical structure of the feature (profiles). The median profiles will give you an idea of the **typical profile** representing each class and the rest of quantiles, the **variability** of the profiles within a class.

```
[13]: P.vertical_structure(q_variable = var_name_ds + '_Q', sharey=True, xlabel='Temperature (°C)')
      P.save_BlueCloud('figures/vertical_struc_EX.png')
```

```
Figure saved in figures/vertical_struc_EX.png
```

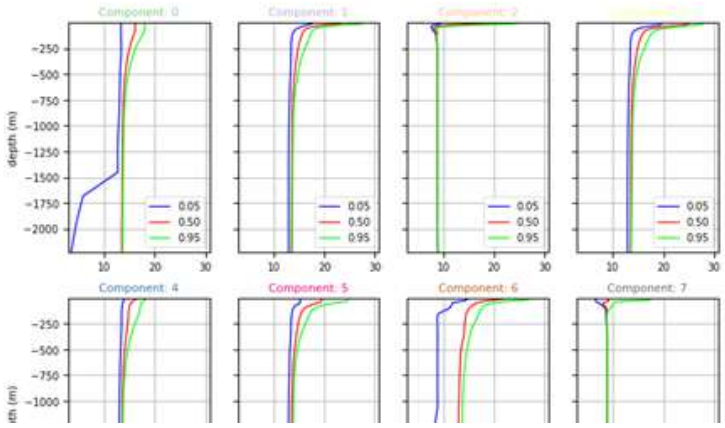


Figure 15. Predict & Plot notebook – Plot results.

Save data

If you are happy with the results and you want to work on the data by your own, you can save the dataset including the new PCM variables (PCM labels, robustness, ...) in the cell below

```
[21]: ds.to_netcdf('datasets/tests_predicted_dataset.nc')
```

Figure 16. Predict & Plot notebook – Save data.

4.5.3 Storm Severity Index - Guidelines

The Storm Severity Index (SSI) calculation needs either daily or monthly SSI grid data as input for the calculation of the SSI map (i.e. grid data in NetCDF). This can be an individual storm, or a seasonal SSI map for a given area and time period or even a SSI climatology map that covers a period of 30 years (e.g. 1980-2010).

The user can access daily and monthly SSI grid data, which are provided (NetCDF format) into the VRE workspace (folder: *input_dataset>SSI*) and are initially available for a few years only (e.g. 1987 until 1989) and for the Mediterranean Sea area only.

This daily or monthly SSI grid data have been derived from hourly Copernicus C3S ERA5 reanalysis data (10 m wind data above sea) using the following calculation (1):

$$\mathrm{SSI}_{\mathrm{grid}} = \sum_{t=1}^{T} [(\max(0, \frac{v_t}{v_{threshold}} - 1))^3] \qquad (1)$$

$\mathrm{SSI}_{\mathrm{grid}}$: 31 x 31 km grid cells (ERA5)
T : # hours in a day or a month
$V_{\mathrm{threshold}}$: no impact below this wind speed

More years and bigger areas of daily and monthly SSI grid data will become gradually available (in folder *input_dataset>SSI)* because the C3S ERA5 hourly reanalysis data covers more than 40 years (1979 until present) of data worldwide.

For the SSI service, a processing method for the SSI computation will be available in next version of the MEI Generator app, providing output data in NetCDF format. In addition, will also be available a viewing service (OGC WMS) underpinned by ADAGUC server. To start the computation of a SSI distribution for given area and time period, the following user inputs on MEI Generator app are required:

- Selection of the data source (daily or monthly SSI grid data) **-** Figure 5 **a**
- Selection of the SSI type (only one method) - Figure 5 **b**.
- Selection of a time period - Figure 5 **d**
- Selection of an area of interest - Figure 5 **e**

The computation of the SSI distribution is performed interactively. For each of the K grid cells in the area of interest for time-period T, the SSI is calculated using daily or monthly SSI grid data (Figure 17). The selection of the time-period (Figure 5 d) should however align with the selected data source (Figure 5 a). Users should not select daily SSI source data to calculate of a long period of time (e.g. years).

Besides the SSI distribution map, the total SSI score for the area of interest is calculated by summating the SSI indices of the K grid cells that are included in the area of interest.

$$\mathrm{SSI}_{\mathrm{total}} = \sum_{k=1}^{K} \sum_{t=1}^{T} [(\max(0, \frac{v_{t,k}}{v_{threshold}} - 1))^3] \qquad (2)$$

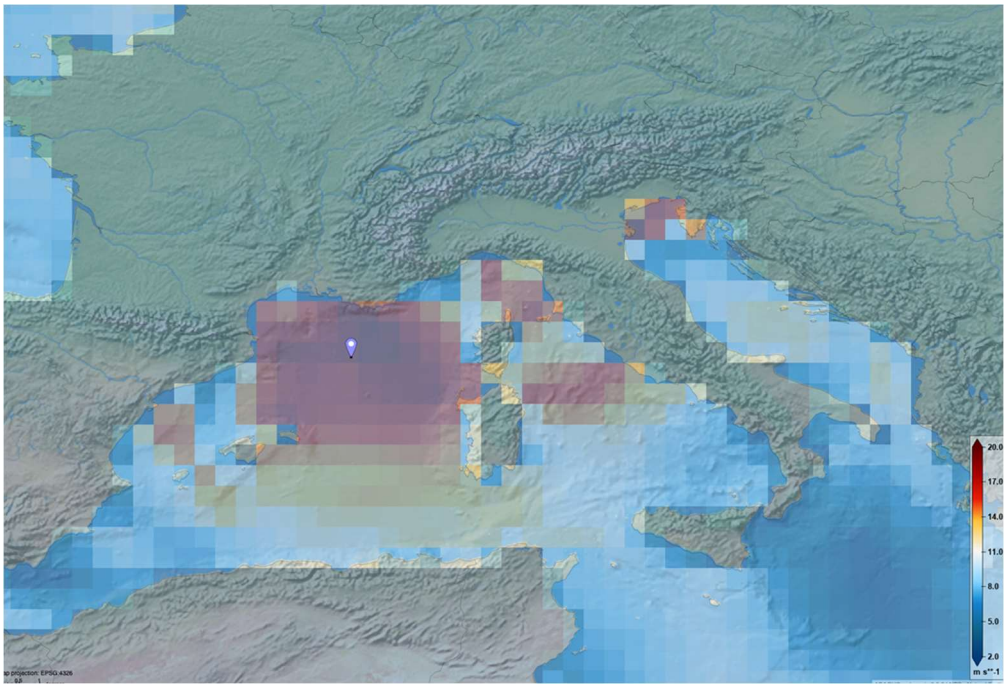


Figure 17. Illustration of the SSI distribution of a storm in November 2017

5 Demonstrator # 4 – Fish, a matter of scales

5.1 Objectives of demonstrator

The objective is to deliver a scalable and robust open data portal for fisheries data in EU waters and beyond, with a focus on a Global Tuna Atlas and a Global Record of Stocks and Fisheries (GRSF). Hence, the demonstrator will expand the existing VREs of FAO Tuna Atlas and GRSF, with more features for data analysis using indicators, interactive maps, etc., for the former, and new as well as expanded information for approved status assessments of fisheries, including those from other sources and demonstrators for the latter.

5.2 Targeted users

5.2.1 Targeted in Blue Cloud

End-users: general public with an interest in stocks and fisheries, fish provenance, fisheries distribution, fisheries and SDG2 and SDG 14, accessed through e.g. web-portals, atlases, API's or QR codes.
Advanced users: regional fisheries data analysts that need to show how fisheries in their area of interest develop over time using fisheries analytical models and in relation to environmental variables and other ancillary data.
Developers: system developers in need of a 'boilerplate' solution for the management of fisheries time-series on catch and effort that brings collated statistical data into a data harmonization and QA process.
Fisheries Managers: regional fisheries managers that require access to overviews of fisheries to inform their management decision making processes. (Figure 18)

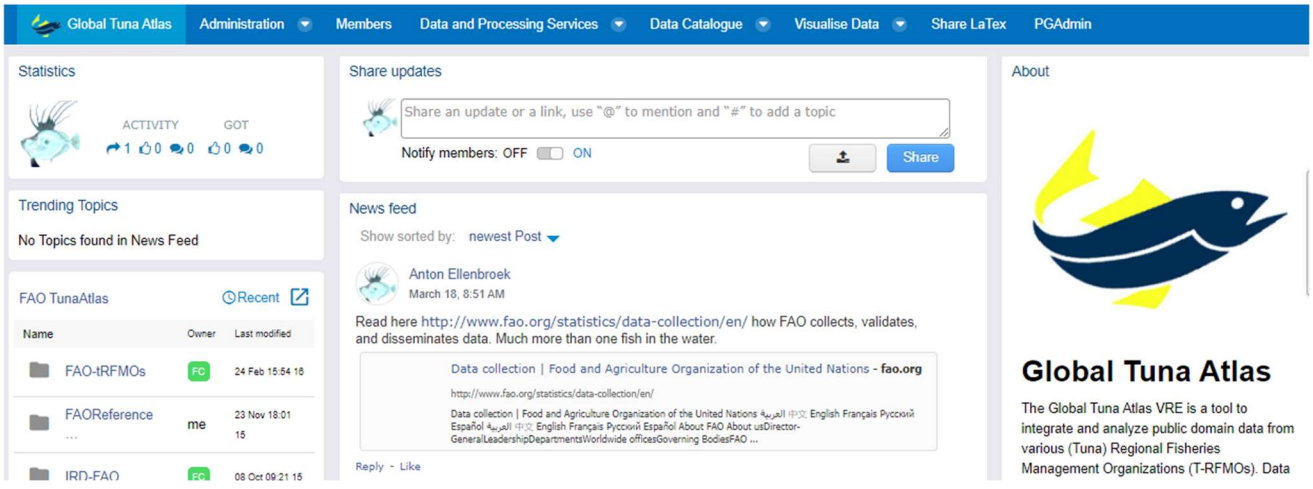


Figure 18. User portal serving fisheries data management communities.

5.2.2 Future potential user communities

All communities with a need to have a fully FAIR compliant data management solution spanning statistical and geospatial data workflow in the Fisheries, Aquaculture and related aquatic and land-based domains.

5.3 Necessary data sources & Blue-Cloud (VRE) services used

"In theory" data: this VRE can use in a transparent fashion any FAIR compliant ISO-OGC dataset as a native source of data.

"In Virtue" data: the VRE is designed to ingest, harmonize and standardize fisheries collated data on catch and effort. Any compliant time-series can be included. Global core information on Global Stocks and Fisheries is incorporated relying on semantic web technologies.

"In Blue Cloud" data: the VRE is capable to access and query ISO/OGC data through WMS/WFS and analyse and display results. This data can be combined with fisheries specific data for further analysis in the WPS based Blue Cloud DataMiner. Furthermore, VRE Cataloguing services (publish, discover and access) are exploited for exposing the demonstrator information through a VRE operated from BlueCloud.

External data: access and display Ancillary information from FNS Cloud on Fish composition and several bespoke datasets are ingested, using a variety of ad-hoc services. Data include Global Effort Maps of fisheries, FAO global project data, and fish tagging data.

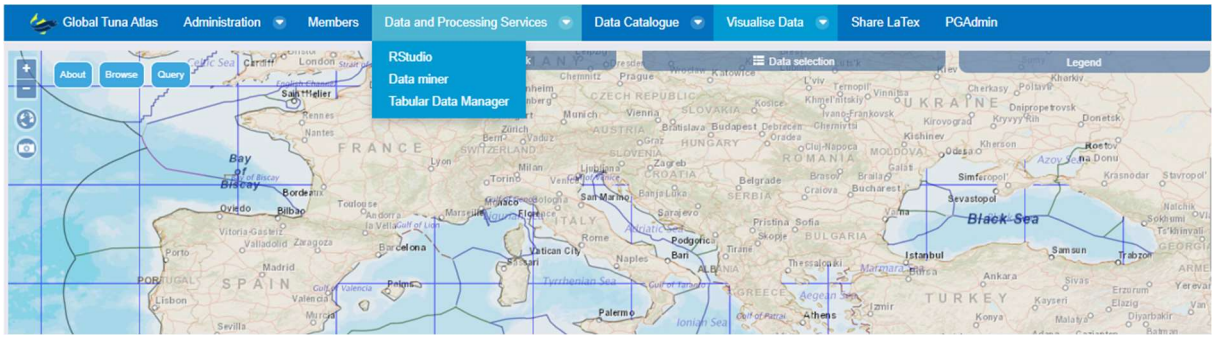


Figure 19. Preliminary Dashboard to Find, Access, Interoperate with and Replicate fisheries data analytics.

5.4 Summary of provided services

Tuna / Fisheries Atlas; an online overview of harmonized time-series of catch and effort accessible through a map Viewer, ISO/OGC metadata and data services, analytical and reporting tools, and R Shiny, Jupyter and Markdown reporting services. (Figure 20)

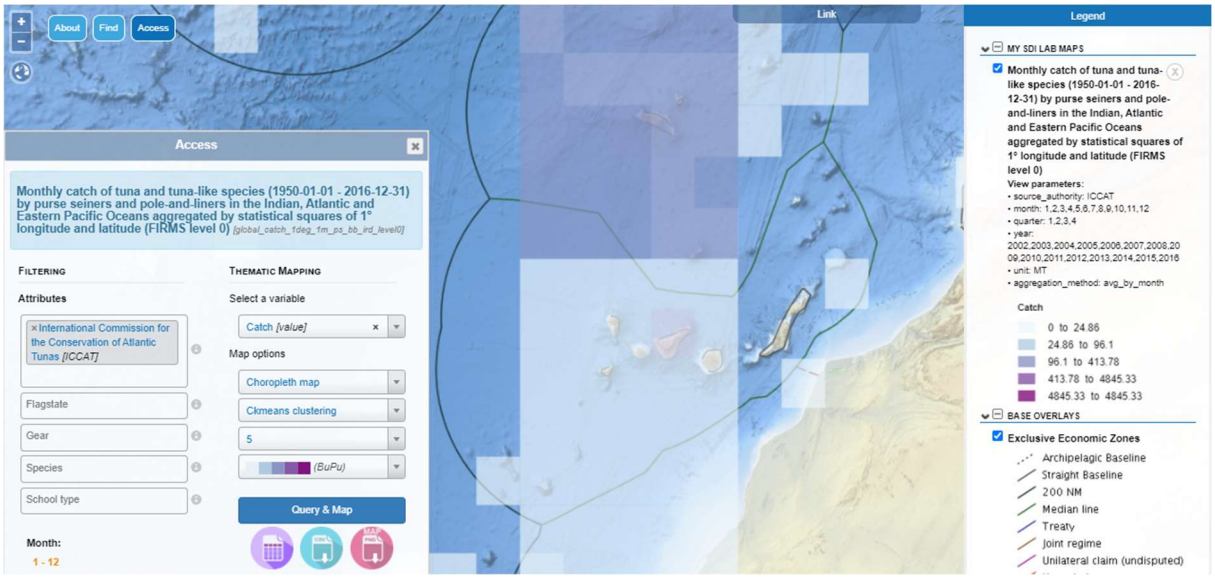


Figure 20. Find and Access Global Tuna capture data.

Global record of Stocks and Fisheries: the global reference repository for stocks and fisheries accessible through:

a) a Blue-Cloud VRE-operated catalogue that enables the hierarchical organization of those resources, with respect to several groupings (e.g. their corresponding types, exploiting resources, provenance information, etc.), which allows users discovering and accessing them, including QR codes for the visual identification and ease of sharing across different users and platforms,

b) a set of APIs that allow retrieving particular information for stocks and fisheries in a programmatic manner,

c) a set of competency queries able to answer complex (and in many cases common) questions that are impossible to answer from the original data sources of GRSF. Each competency question, is assigned a small description, to describe its purpose, and under the hood a SPARQL query is formulated and submitted to the GRSF Knowledge Base, so that it can answer and present the result to the user in a user-friendly manner (Figure 21).

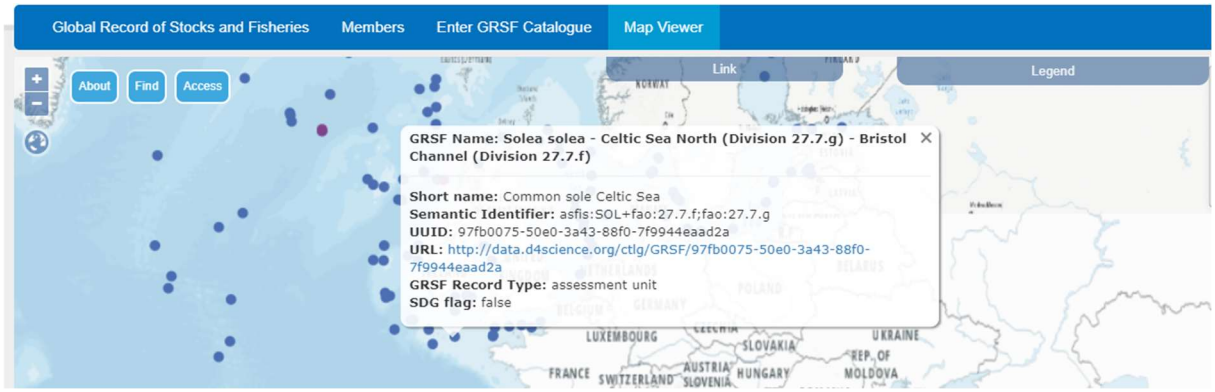


Figure 21. GRSF Public Map interface.

SDG14.4.1 Services: (2021) A suite of analytical services to serve a community of fisheries data analysist with models and methods based on widely published software.

API's: A GRSF API is under development.

5.5 Guidelines to use the services

The demonstrator is a rich environment of services that interact at various levels to connect different collaboration teams that need clearly separate workflows driven by underlying services. These VREs include the Fisheries Atlas with a focus on Statistical Data Management and the GRSF with a focus on stocks and fisheries Information management. These VREs are available on the following links:

a) Fisheries Atlas: https://blue-cloud.d4science.org/web/fisheriesatlas
b) GRSF PRE: https://blue-cloud.d4science.org/web/grsf_pre.

The Fisheries Atlas maps, through the public user interface, will show thematic maps of the Earth's fisheries, but also their production and trade in pop-up stats. The design aims to show trends in Fisheries indicators towards achieving the SDG's. It contains a map viewer for relevant layers and spatial information such as:

- FAO and IRD produced Tuna Atlas layers;
- FAO Productions and trade national statistics;
- Selected Regional Fisheries Organizations data;
- Selected EMODNet layers served by WP2);
- Selected CMEMS products served by WP and generated through WP4 data processing (in progress)
- Selected species distribution maps (From D4Science)
- Selected FAO and VLIZ layers (From D4Science / WP2)

This list will grow with the underlying Blue-Cloud Spatial Data Infrastructure; the VRE manager will decide which dataflows are mature enough and ready for demonstrator display. It can inherit other features from specialized fisheries dataflows later on.

The GRSF shares many services with the Fisheries Atlas, but has a focus more on information management and sharing of data at the level of stocks and fisheries. The GRSF PRE VLAB is used to validate new content in the GRSF Knowledge Base and thus is not a public service. In 2020, the GRSF team used this environment to validate data harvests from 3 global GRSF sources that are now published into the GRSF Admin and GRSF VREs. All GRSF environments are connected to the Blue Cloud Spatial Data Infrastructure (SDI) and all geospatial-referenced content can be viewed in context of other data accessible within the Blue-Cloud (e.g. Bathymetry layers from EMODnet bathymetry).

The SDI behind the Fisheries Atlas and the GRSF Map Viewer (OpenFairViewer) is continuously enriched with FAIR compliant data. Dependent on their metadata, they are Interoperable across upgraded R and R Shiny based visualization and analytical tools. In addition, GRSF APIs and Competency Queries are made available for data access.

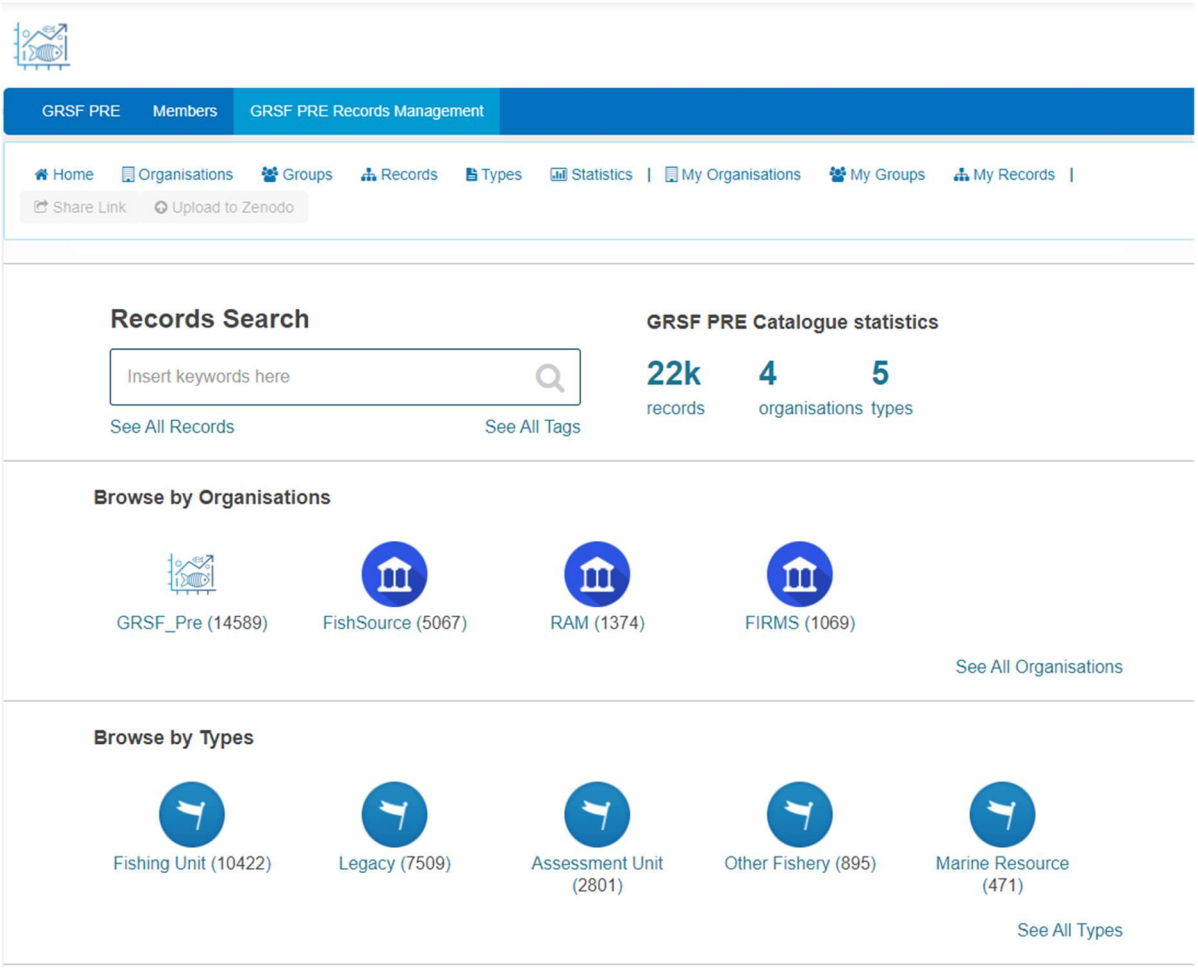


Figure 22. The GRSF VRE UI for the Global Record of Stocks and Fisheries (GRSF).

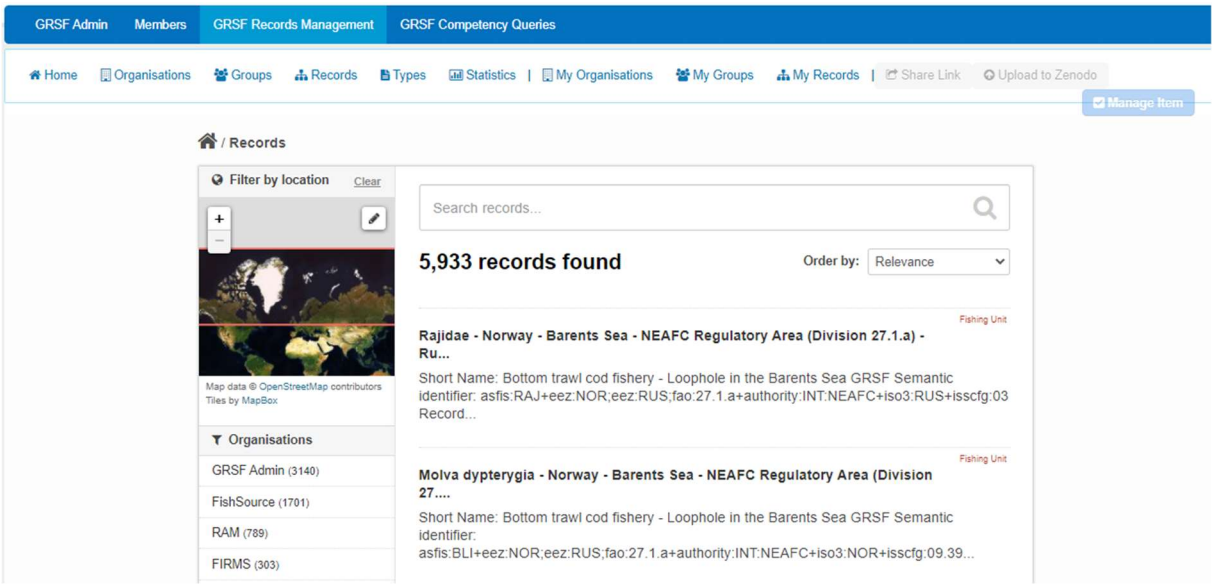


Figure 23. GRSF Record editing environment.

6 Demonstrator # 5 – Aquaculture monitor

6.1 Objectives of demonstrator

The objective is to deliver a scalable and robust open data portal for aquaculture cage detection and monitoring and coastal pond and land-type classification.

6.2 Targeted users

6.2.1 Targeted in Blue Cloud

End-users: general public with an interest in aquaculture locations, production, and tracking, accessed through a public web-portal. (Figure 24)

Advanced users: regional aquaculture data analysts that need to monitor how aquaculture in their area of interest develops over time using Sentinel and other satellite data, and bring that in relation to environmental variables and other ancillary data such as site-inventories.

Developers: system developers in need of a 'template' solution for the management of sentinel and other satellites data access and processing in WeKEO. The demonstrator example workflow on cage monitoring can be adapted to other data and analytical WeKEO processes.

Regional Aquaculture Managers: regional aquaculture managers that require access to overviews of aquaculture sites and area estimates to inform their management decision making processes.

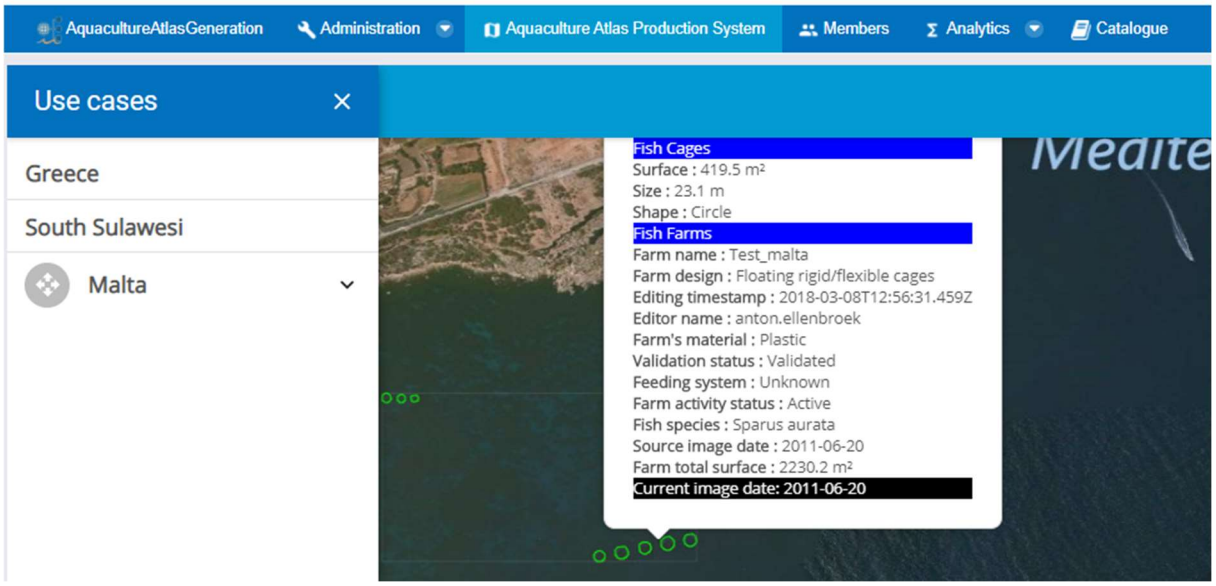


Figure 24. Aquaculture cages user portal.

6.2.2 Future potential user communities

All communities with a need to have a fully FAIR compliant data management solution driven by geospatial data workflows in the Fisheries, Aquaculture and Land-based agricultural domain.

6.3 Necessary data sources & Blue-Cloud (VRE) services used

“In theory” data: this VRE can use in a transparent way any FAIR compliant ISO-OGC dataset as a native source of data.

“In Virtue” data: the VRE is designed to access Sentinel and other satellite products, and perform a WeKEO based cage detection algorithm.

“In Blue Cloud” data: the VRE is capable to access and query ISO/OGC data through WMS/WFS and analyse and display results. This data can be combined with aquaculture site-specific information for further editing in an on-line map-based Data Editor. Data from other Blue Cloud demonstrators, such as on ocean variables and species, can be displayed as well.

External data: access and display ancillary information from high-resolution satellites and several bespoke datasets are ingested in a geoserver, using a variety of ad-hoc services and used to enrich the satellite-derived maps. Data include farm characteristics and statistics, and geographic areas such as Exclusive Economic Zone (EEZ).

6.4 Summary of provided services

Aquaculture Cage Atlas: an online overview of satellite data derived maps of cages and cage clusters. The products are delivered through an ISO-OGC compliant map viewer, and registered users can edit features of the detected cages and cage clusters. A proximity service can be called to automatically map across feature sets to enrich maps. In another process, estimates of cage activity over a production season can be made if there is a large enough sample available. (Figure 25)

Aquaculture Ponds Atlas: (2021) the versatility of the Blue cloud infrastructure and the re-usability of its components will be demonstrated in a test-service that uses some of the same data sources and a similar analytical data process to the Cage Atlas. The results will be a coastal land-use classification map, fully based on Copernicus data for its remote sensing component.

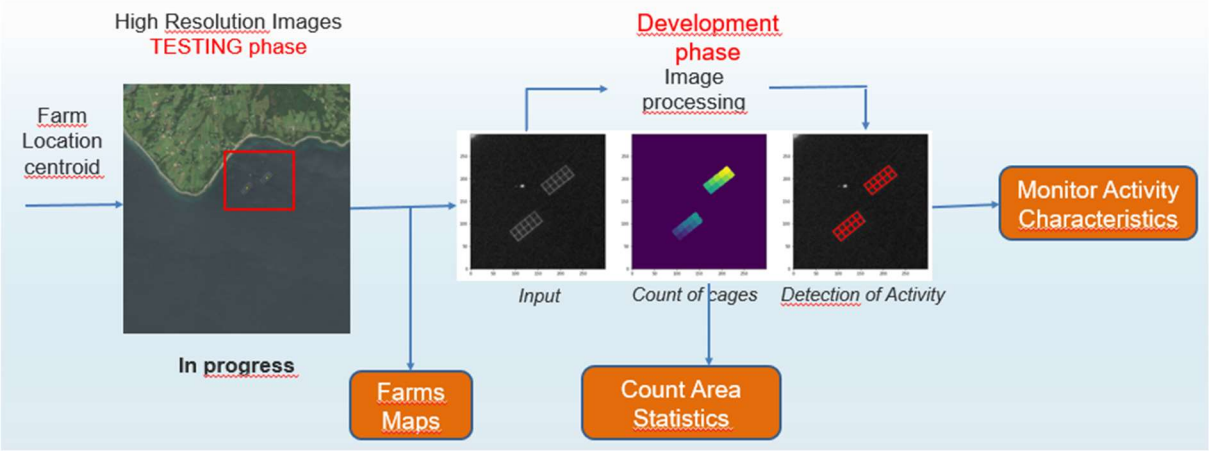


Figure 25. Summary workflow for cage monitoring.

6.5 Guidelines to use the services

The cage identification and monitoring services will be deployed on the Blue Cloud infrastructure in 2021, and are currently exclusively accessible through the partners' infrastructure.

The user oriented VRE facets are already embedded in the VRE and registered end-users can browse the analytical products. These are accessible through a GeoNetwork (Figure 26) and a Web-Portal (Figure 27) on the demonstrator's public page at the following link: Aquaculture Atlas Generation[2].

The framework for the Aquaculture Atlas Generation system shares significant approaches and services with other VREs. In Blue-Cloud, the planned improvement of Earth Observation applications for European seas and Indonesia will use Copernicus and related data sources provided Blue-Cloud. It supports the generation and the review of aquaculture maps for offshore fish farms and coastal ponds. For the detection of aquaculture fish farms, high-resolution optical satellite images analyzed in new semi-supervised algorithms that, after review and approval, will be embedded into the Blue-Cloud system across WP2 and WP4 resources. For the detection of coastal ponds (and rice paddy fields), a dedicated algorithm is still being developed. It will use Copernicus Sentinel-1 Synthetic Aperture Radar (SAR) and Sentinel-2 multispectral data, amongst others.

The output maps are visualized in an integrated map-viewer, similar to the technology of Demonstrator 4. For Super User (i.e. thematic expert), this VRE enables the interactive editing of attributes, the validation and publication of OGC-compliant data and catalog entries. The VLab relies on Blue-Cloud data services to manage users and content.

[2] https://blue-cloud.d4science.org/web/aquacultureatlasgeneration

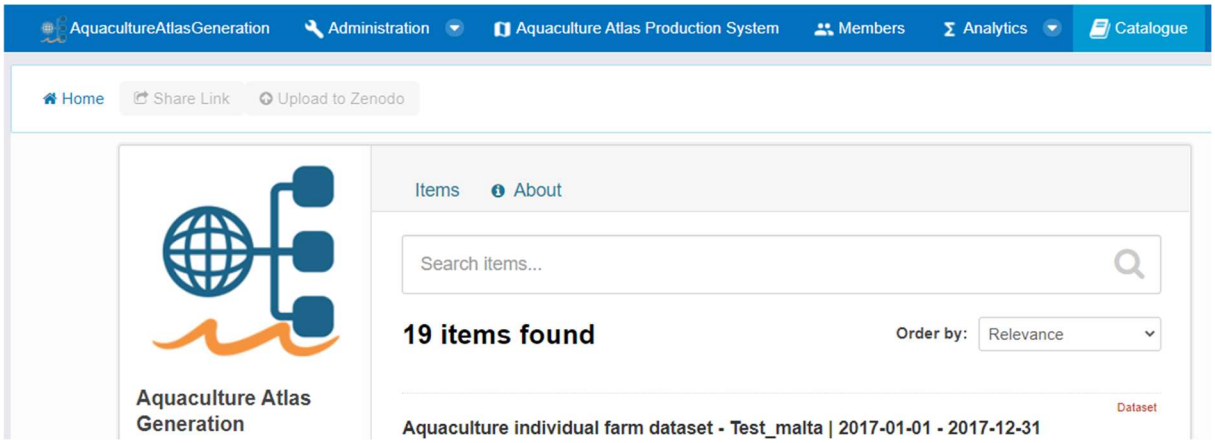


Figure 26. GeoNetwork portal for aquaculture map products.

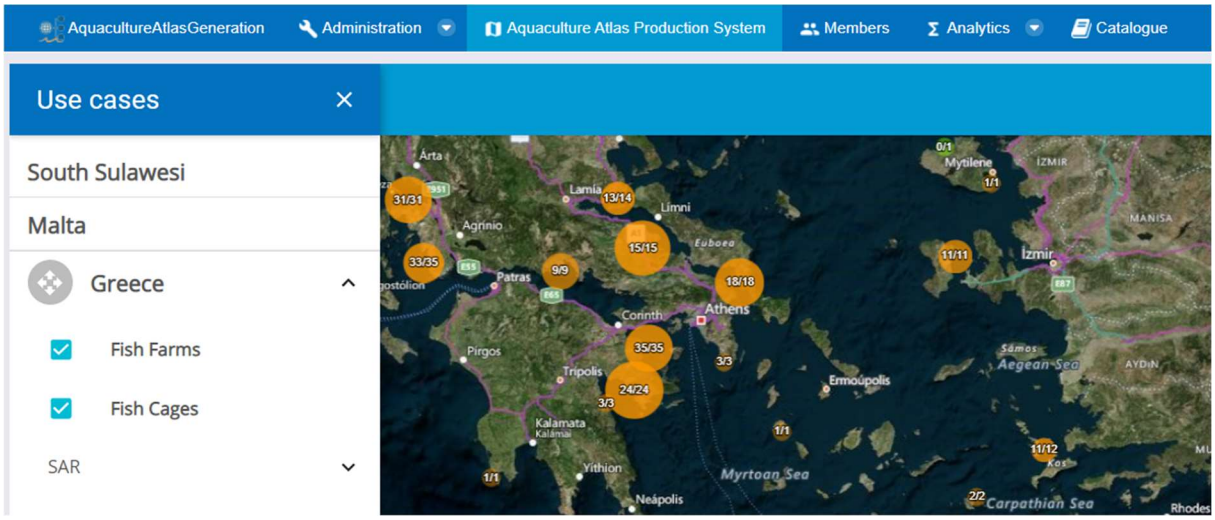


Figure 27. Portal for registered users to find and access farm details.

Advanced users will be supported in 2021 with facilities to run analytical services using Copernicus and other data, while FAO will provide a mapping service to connect to local statistical datasets.

7 Conclusions

At this stage of the implementation of the Blue-Cloud, this is only a temporary and ongoing/living document handbook. It gives the present state of the development of the demonstrators and, where already achieved, their implementation as Virtual Labs on the Blue-Cloud VRE.

This Handbook V1 describes the β-version of the demonstrators and their deployment as Virtual Labs. This version of the handbook is mostly for internal usage and updates. Work is still ongoing for deploying all demonstrators on the Blue-Cloud VRE and then to move from a β-version of Virtual Labs to their final operational versions, which is expected in M27 (December 2021) of the Blue-Cloud project.

This V1 of the Handbook will be further elaborated and expanded between now and the second (or final) release of the Virtual Labs. The V2 of the Handbook is expected at the same time as the final version of the demonstrator Virtual Labs in M27 of the Blue-Cloud project (December 2021).