



D5.21 REPORT ON DEMO OF THE PROTOTYPE WEB APPLICATION (ACTIVITY 5.4; B14; INGV-BO)



Deliverable number:	D5.21
Work package:	WP5 – Marine Domain
Intermediate Objective:	IO5.X, Activity 5.4
Deliverable type:	Report
	<input type="checkbox"/> Websites, patent filings, videos, etc.
	<input type="checkbox"/> Other: please specify
Dissemination level:	Public
	<input type="checkbox"/> Restricted
Estimated delivery (bimester):	B14, postponed to B15 (01/04/2025-31/05/2025)
Actual delivery date:	31/05/2025
Author(s) (Partner-OU):	Simona Simoncelli, Paolo Frizzera, Claudia Fratianni, Damiano Delrosso (INGV-BO)
Reviewed by:	
Note:	

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006 (D.D. n. 130/2022)
 Funded by EU - Next Generation EU
 Mission 4 “Education and Research” - Component 2: “From research to business” -
 Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

1. INTRODUCTION.....	3
1. Purpose of the document.....	3
2. METHODOLOGY.....	4
1. SOURCE Software.....	5
2. Data	6
3. WORKFLOW	7
1. Data Access	7
2. Observation Module	7
3. MODEL MODULE.....	10
4. CAL/VAL MODULE.....	11
4. DATABASE	13
5. WEBGIS APPLICATION.....	13
6. SUMMARY AND FUTURE DEVELOPMENTS.....	15

Table of contents

1. INTRODUCTION

1. Purpose of the document

The aim of this document is to present the prototype web application developed by INGV-BO as one of the proposed tasks (A5.4.3) within activity 5.4.

ITINERIS WP5 integrates the Italian marine Research Infrastructures (RIs) and the marine component of other environmental RIs, for the establishment of a national integrated marine observing system of systems. Marine RIs are an essential component of the European contribution to the global ocean observing system since they share facilities, resources and services to monitor the ocean, providing scientific data as foundation of derived information and knowledge. The integration and harmonization of the national component of marine RIs represents a great advantage in terms of efficiency and sustainability. The implementation of a unique portal where all marine monitoring systems are displayed and the data visible and easily accessible represents a key step towards this integration and the long-term monitoring of essential ocean/climate and biological variables.

The design and implementation plan of the ITalian Integrated Ocean Observing System (IT-IOOS), has been presented in D5.6. Its aim is to harmonize data and products from the different RIs and to establish an integrated system of systems, capable of ensuring continuity of data and services and responding to user requirements. Data collected by the RIs from different measuring systems and platforms are integrated and made visible by a national Marine Data Portal (<https://it-ioos.eu/>), designed as a distributed system interfaced with existing RIs and thematic data centers to provide to the users a single data access point.

One of the main project’s objectives (OBJ4) is also to develop pilot services to tackle overarching marine issues and to respond to key stakeholder requirements. Our proposed pilot aims to show the

impact of improved data availability in developing a prototype¹ service for monitoring the marine environment. The prototype is a web application for the visualization of data products, such as **reprocessed and qualified time series of selected EOVs/ECVs** from moorings and fixed stations based on SOURCE (Sea Observations Utility for Reprocessing, Calibration and Evaluation) open software (Oliveri, and Simoncelli <https://doi.org/10.5281/zenodo.5106546>, Oliveri et al. 2022, <https://doi.org/10.3389/fmars.2021.750387>) and the corresponding synthetic time series extracted from open access model data.

Our prototype fits multiple purposes:

1. to access RIs in situ data of selected EOVs (i.e. Temperature and Salinity) from moorings and fixed stations;
2. to re-process the RIs in situ data aggregating fragmented time series in daily, monthly time series for visualization and analysis purposes;
3. to perform secondary Quality Control (QC) of the in situ data. Secondary QC means applying an additional QC level to the data with respect to the primary QC performed by the data provider.
4. to access and extract co-located time series from open access models;
5. the visual intercomparison of models and observations with the provision of quantitative model skill scores which indicate the model capability to reproduce the observations.

We believe that this application could serve multiple stakeholders, starting from the observing and the modeling communities that can visually monitor the performance and integrity of their data and products through the intercomparison of different type of data. The visualization of EOVs time series of temperature and salinity at fixed locations allows real time monitoring of the ocean state and its environmental conditions. In fact, this application sets the basis for future developments such as the computation of environmental monitoring indicators or the analysis of other physical or biogeochemical variables.

2. METHODOLOGY

SOURCE code has been totally rewritten in python by Oliveri et al. 2022 starting from the operational workflow that manages the <http://calval.bo.ingv.it/> (Fig.1) webpage, showing the evaluation of INGV models' performance versus independent in situ observations from Eulerian systems (moorings and fixed buoys) available through the Copernicus Marine Service.

¹**Prototype:** the first early representation of the service which offers the expected functionality and performance of the final implementation.

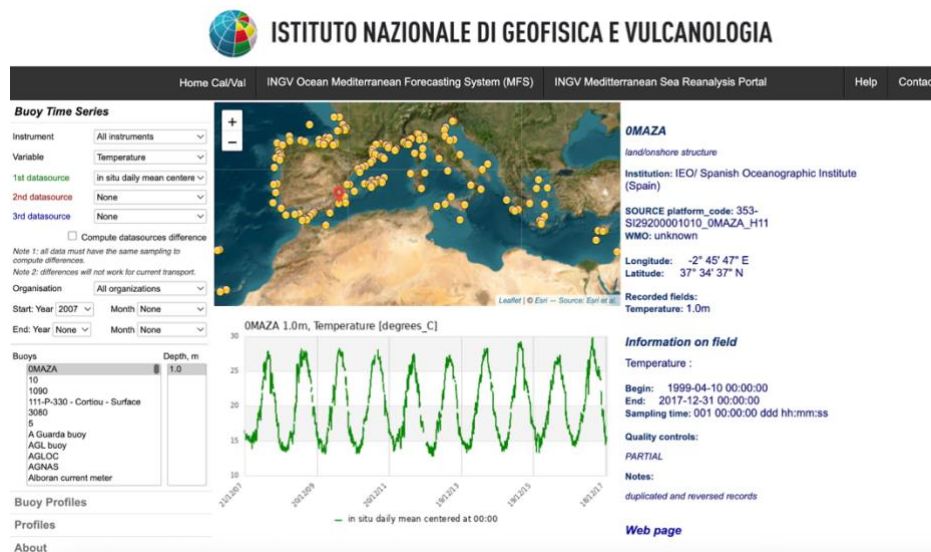


Figure 1. Screenshot of the INGV CalVal webpage (<http://calval.bo.ingv.it/>).

Our initial goals were to:

1. enhance SOURCE Technology Readiness Level (TRL) bringing it to a system prototype demonstration in operational environment under limited conditions (7); and
2. increase SOURCE FAIRness (Findability, Accessibility, Interoperability, Reusability), mainly improving its Interoperability with different data sources and access services and its Reusability in different environments and for slightly different scopes.

In specific, SOURCE has been developed with the double ambition of running it on an INGV server and integrating it into IT-IOOS service, once operational.

SOURCE has been adapted to use ITINERIS data and data products (validated time series), specifically from the European Multidisciplinary Seafloor and water column Observatory (EMSO), one of the RIs participating in the project.

EMSO observatories are high-technology platforms hosting sensors monitoring various parameters, such as temperature, pH, salinity, water circulation, and seabed movements. In our pilot we focus on temperature and salinity data from the Italian EMSO facilities.

During the project a considerable time has been dedicated to the participation in the review of the RIs data management practices, their access services, the adopted standards and the data quality control procedure (D5.6). This crucial information was fundamental to our activity.

1. SOURCE Software

The software has been first organized in Jupyter notebooks (<https://jupyter.org/>) and the developments are managed through the Github repository (<https://github.com/>). The Jupyter notebooks allow the execution of the code step by step and the integration of the code with documentation and plots. This approach also makes SOURCE suitable for VRE deployment.

The code is organized in several notebooks, shematized in Figure 1:

- **data access module:** it retrieves and organizes the in-situ data;
- **observational module:** it processes in situ data and metadata;
- **model module:** it downloads and processes model outputs, extracting model time series collocated with observations;
- **CalVal module:** it computes model skill scores comparing model versus in situ data.

The SOURCE output data and metadata are organized in a database which is then used for the WebGIS **visualization**.

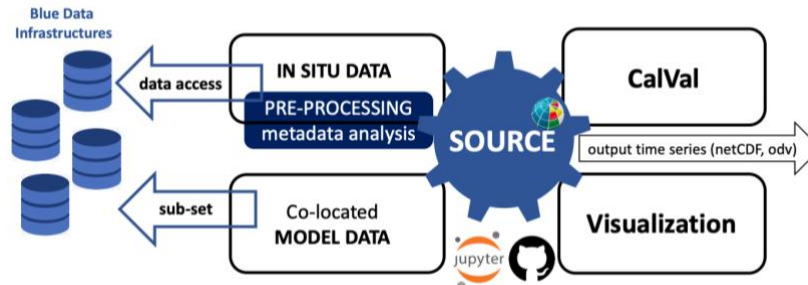


Figure 1 - Diagram of SOURCE analytics tool which is the foundation of the prototype web application.

2. Data

The IT-IOOS architecture is structured as a distribution system with a single access point for users and for the ITINERIS hub (D5.6). IT-IOOS consists of sub-systems (SS), one per Research Infrastructure participating in the ITINERIS Marine Domain. In addition, the Marine Data Store and the Marine Data Portal are key components of the IT-IOOS architecture, responsible for accessing and archiving data.

The Marine Data Store is the dynamic archive for ITINERIS marine data, metadata, and products through a federated approach connecting with IT-IOOS sub-systems. The data, products, and services generated by each sub-system are made available through an endpoint, typically ERDDAP.

The Marine Data Portal serves as a single access point for IT-IOOS and acts as the interface between users and the various IT-IOOS sub-systems, including the IT-IOOS Marine Data Store. Its main objective is to provide access to the data and services of IT-IOOS.

Every SS consist of one or more components depending on the sub-system. Each component can be organized into Production Units (PUs). PUs are responsible for producing a specific set of data or products corresponding to a data endpoint. The architecture of the EMSO SS has been schematized in Figure 2 and it consists of 3 components organized in 5 PUs, as detailed in D5.6. EMSO facilities are platforms equipped with multiple sensors, placed along the water column and on the seafloor, providing continuous time-series of measurements of different physical and biogeochemical variables.

EMSO provides data to its end-users through a single ERDDAP endpoint that federates a network of ERDDAP servers managed by each single facility. Thus, each EMSO facility manages its data repository where the latest data version is stored.

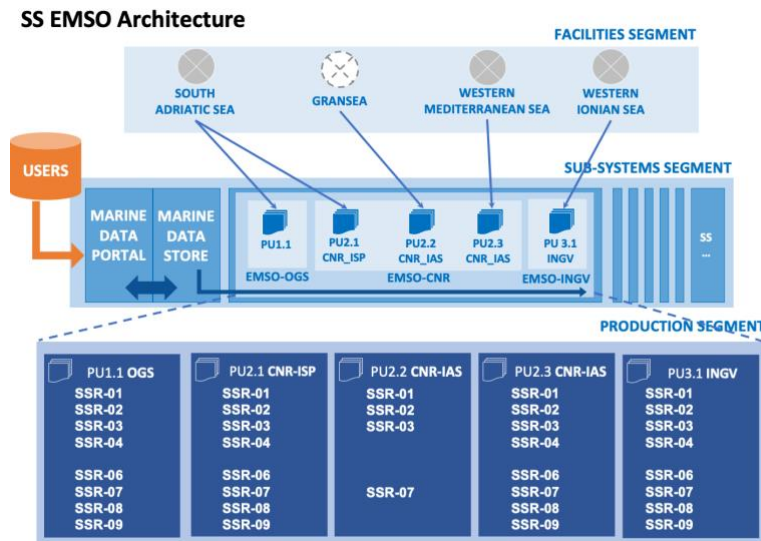


Figure 2 Architecture of the IT-IOOS Sub-System EMSO.

ITINERIS end-users should have access to the national EMSO SS datasets from the Marine Data Portal but during its implementation phase we used the EMSO PUs endpoints inserted in [Tabella DATASET D5.6 data dispoenibilita integrazione.xlsx](#). Once the Marine Data Portal data access service is operational, the data retrieval process will go through it.

3. WORKFLOW

1. Data Access

Data files have been accessed through the ERDDAP servers managed by each EMSO PU endpoints listed here [Tabella DATASET D5.6 data dispoenibilita integrazione.xlsx](#). Two ocean physical EOVs/ECVs, among the various monitored by EMSO facilities, have been considered for the first pilot implementation: the Subsurface Temperature and Subsurface Salinity. The ERDDAP queries included: latitude, longitude, time, depth, TEMP, TEMP_QC, PSAL, PSAL_QC, station_id. Data quality information from the data provider and contained in the quality flag associated to each temperature and salinity measurements is essential for data reusability in any application and it is the base of our data reprocessing procedure which aims at further checking them through a secondary QC, providing consistency of the application output.

EMSO provides interoperable data and metadata in a coherent and standardized manner, following the EMSO Data and Metadata Specifications. Since the EMSO metadata schema is going to be mapped to a common IT-IOOS metadata schema to ease the harmonization of its data access service, some SOURCE code adaptation and testing are required once the data query will go through it.

Some difficulties have been encountered in this phase due to various data FAIRness issues, which required some tailored manual solutions in order to start data processing. The ongoing metadata harmonization process and data access through the Marine Data Portal should enhance data interoperability avoiding human intervention. This is a crucial aspect for implementing any operational data application.

2. Observation Module

The SOURCE Observations module is designed to prepare observation-based data through a sequence of pre-processing and re-processing operations (flowchart in Figure 3). Initially, the data are pre-processed and re-formatted into SOURCE's internal netCDF format. During this step, spatial and temporal limits, instruments, variables and the desired Quality Flags (QFs) are selected for further analysis.

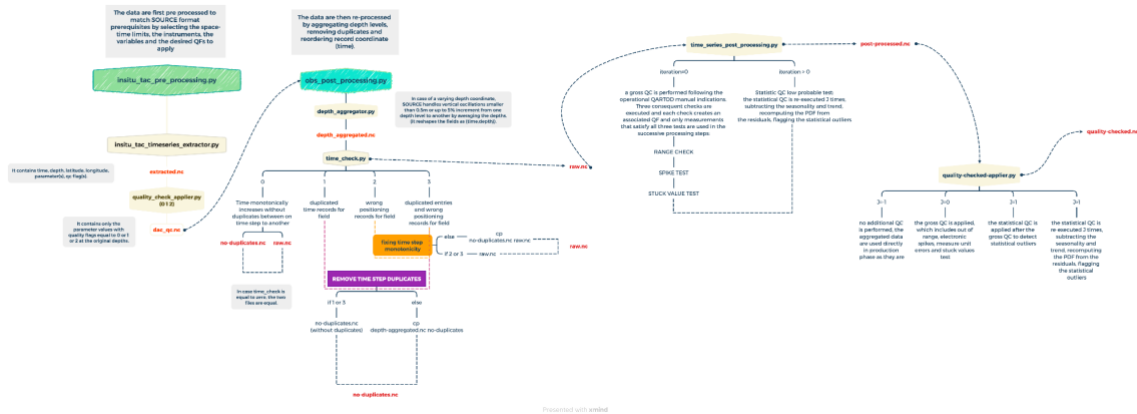


Figure 3: Flow chart of SOURCE observations module.

Next, the data are grouped by parameters, and the QFs are applied. The re-processing phase involves aggregating depth levels, removing duplicate records, and reordering data along the time coordinate (Figure 4).

Afterwards, the data undergoes a secondary quality control (QC) procedure, preserving the original sampling and applying different QC tests depending on the value of an input parameter J:

- $J = 0$: A gross QC is applied, following the operational guidelines from the QARTOD manual. Three sequential tests - range check, spike test and stuck value test - are executed. Each test generates a corresponding QF, and only measurements that satisfy all three tests are retained for the next steps;
- $J > 0$: A statistical QC is also performed, after the gross QC, and repeated J times. This process involves subtracting the seasonal and trend components (Figure 5), recomputing the probability density function (PDF) from the residuals, and flagging statistical outliers.

Finally, the re-processed and quality-controlled data are saved and temporally averaged based on the original sampling interval (Figure 6).

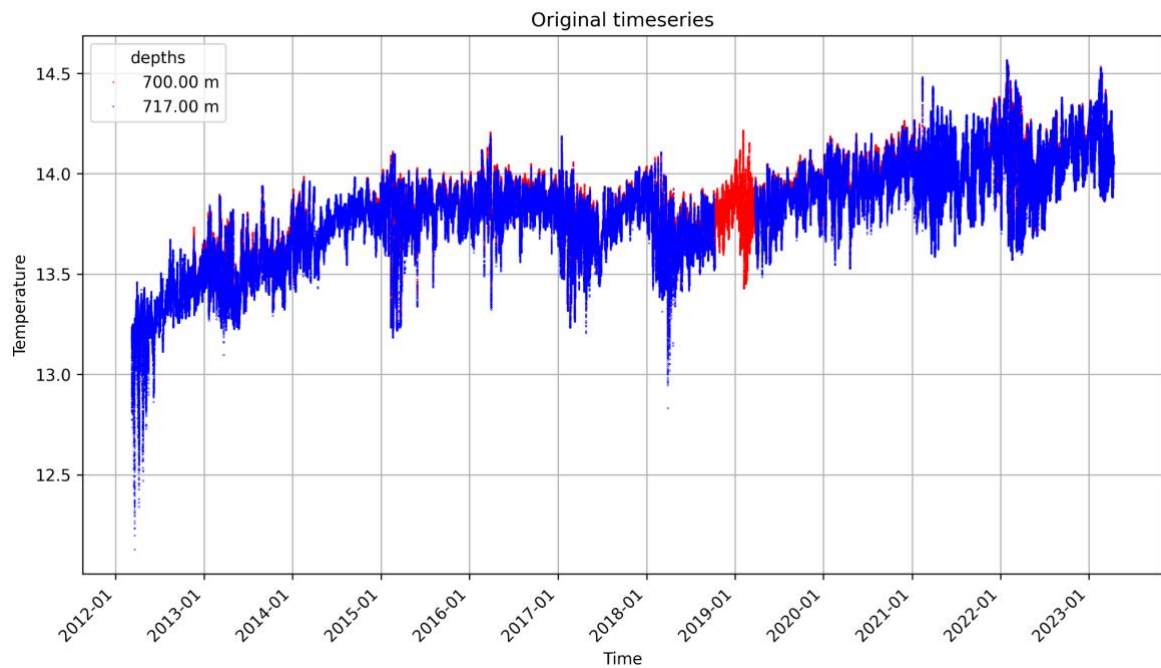


Figure 4 - Example of an input time series pre-processed and re-formatted.

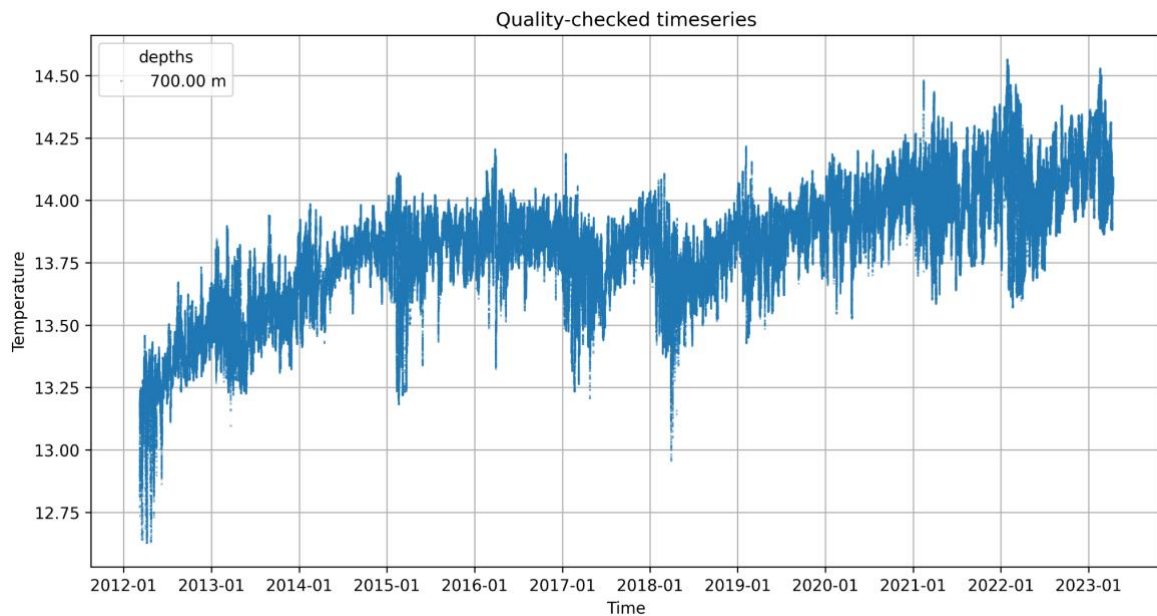


Figure 5 - The example time series after going through the secondary QC (statistical QF applied).

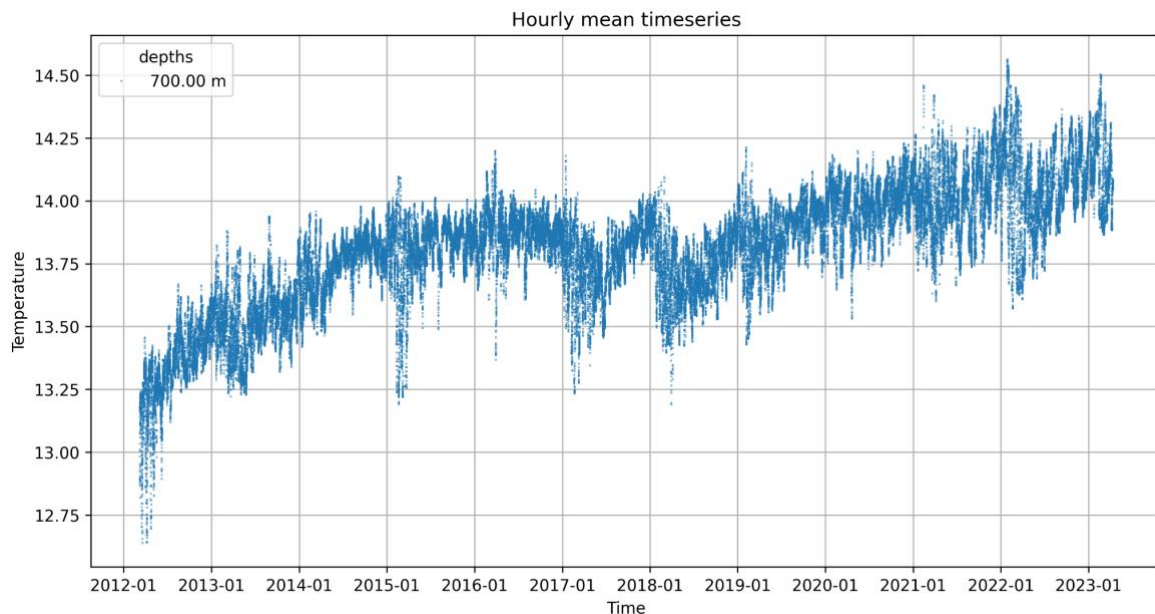


Figure 6 - Hourly averages computed at the end of SOURCE observations module.

3. MODEL MODULE

This module aims to download model time series at the observed space-time moorings' locations (defined by the Observation Module) in order to compare model outputs with observations (in the CAL/VAL Module). The model product selected to test the workflow is the Mediterranean Sea Physics Analysis and Forecast (MEDSEA_ANALYSISFORECAST_PHY_006_013, https://doi.org/10.25423/CMCC/MEDSEA_ANALYSISFORECAST_PHY_006_013_EAS8), which is produced by the physical component of the CMEMS Mediterranean Forecasting System (Med-Physics). The model horizontal grid resolution is $1/24^\circ$ (4 km, approximately) and it is resolved over 141 unevenly spaced vertical levels. The product covers the whole Mediterranean Sea.

The first section of the module downloads the model mesh mask from the CMEMS portal which will be used later to determine the closest sea grid point to the mooring location. Once identified the nearest sea grid point, there is a check on the distance between mooring location and the point itself: if the distance exceeds the threshold defined by the parameter `Dist_max`, currently equal to 9 km but freely definable by the user, model data will not be downloaded for the location of the mooring, and the procedure moves on to the next mooring.

After determining the closest model grid point to the mooring location, the procedure uses this point's indices to download the potential temperature and salinity time series, for the entire user-defined period and for the whole water column, via the Copernicus Marine Toolbox (<https://toolbox-docs.marine.copernicus.eu/en/stable/>).

In the subsequent section, the model sub-set undergoes post-processing through some SOURCE utilities:

- computation of in situ temperature from model potential temperature and salinity to properly compare the model with observations
- model fields are vertically linearly interpolated onto the observed depths. If the observed depths fall outside the model's vertical bounds, the value is extrapolated from the nearest model depth, both at the surface and at the bottom.

4. CAL/VAL MODULE

The aim of the CAL/VAL module is to visualize the temperature and salinity time series (both in situ and from model outputs) downloaded and processed through the previous modules, as well as to compute the relative model skill scores (RMSE and bias).

Temperature and salinity variables are processed in separate cells of the module, so the user can choose to display only the information they are interested in.

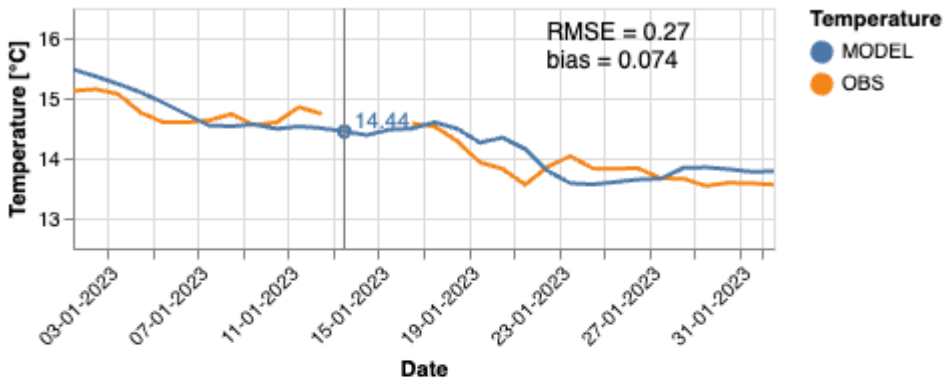
Once a Pandas DataFrame has been created for in situ data and model data (for the desired variable), the processed moorings are displayed on an interactive map, each represented by a buoy icon labeled with the platform name (see an example of visualization contained in the Jupyter notebook in Figure 6). By clicking on the icon of the desired mooring, the temperature or salinity time series for each mooring depth is visualized along with the corresponding model RMSE and bias values. Examples of temperature and salinity visualization are shown in Figures 7 and 8, respectively.



Figure 6 - Locations of the moorings whose data (temperature) were downloaded and processed for the January 2023 test period.

OBSEA - OBSEA

Temperature [°C] at 1.0 m



OBSEA - OBSEA

Temperature [°C] at 20.0 m

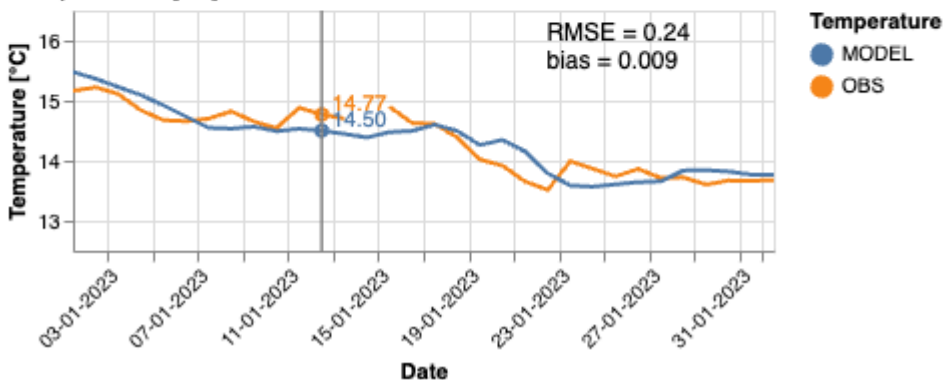


Figure 7 - Temperature time series from OBSEA platform and the corresponding MED-MFC model time series over the January 2023 test period. The model's RMSE and bias values with respect to in situ data are displayed in the upper-right corner of the plots.

OBSEA - OBSEA

Salinity [PSU] at 20.0 m

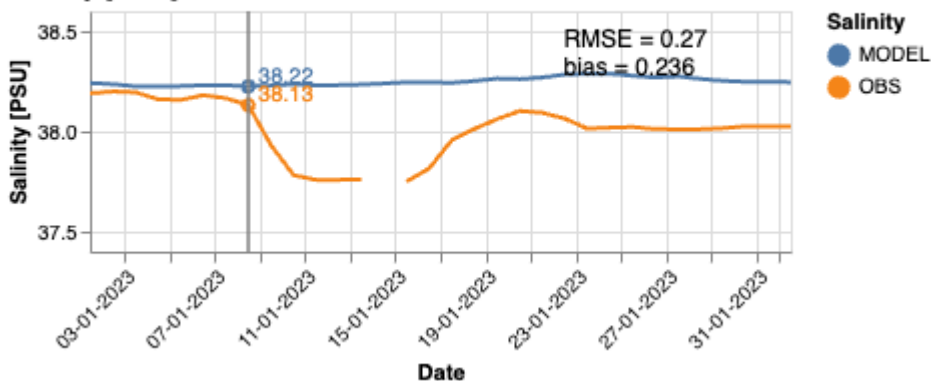


Figure 8 - Salinity time series from OBSEA platform and the closest MED-MFC model grid point time series in January 2023 test period. The model's RMSE and bias with respect to in situ data are displayed in the upper-right corner of the plots.

4. DATABASE

It was decided to use a database for storing the processed data, providing flexibility and faster system response, which is a key aspect for web applications. In selecting the best database option, we prioritized an open source solution that offers robustness, reliability, and security.

The choice of the PostgreSQL database was made due to the high flexibility of the product. It is a relational database that can be extended thanks to the ability to create custom data types, functions, and extensions. It supports standard SQL, is ACID compliant (Atomic, Consistent, Isolated, and Durable), and ensures security through various authentication and encryption methods. It also supports synchronous or asynchronous replication and offers an advanced indexing system with different types of indexes.

Figure 9 show the structure of the database where SOURCE intermediate and final information is organized for further usage by the webGIS.

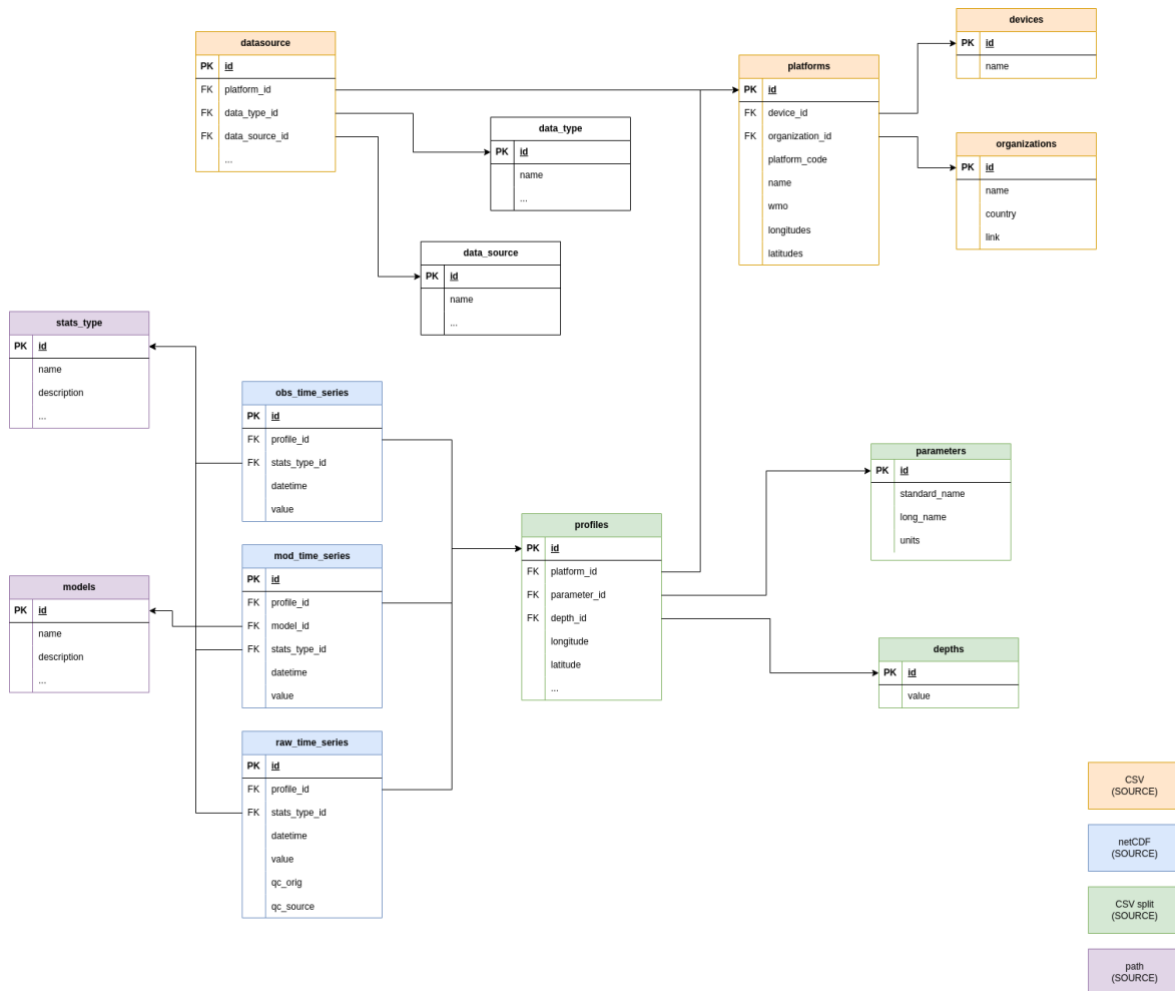


Figure 9- Diagram of the SOURCE database structure.

5. WEBGIS APPLICATION

The Django framework (<https://www.djangoproject.com/>) was chosen for the web application. Django is a high-level Python web framework designed to accelerate development by leveraging Python scripts. It is a comprehensive full-stack framework, providing built-in support for database connectivity, URL routing, authentication, an administrative interface, and more. Django follows the MTV (Model-

Template-View) architectural pattern, effectively decoupling data models, business logic, and presentation layers. This decision facilitated the reuse of the SOURCE Python code and the testing via Python Jupyter notebooks. Figure 10 presents the logical flow of service deployment.

The front-end is served by NGINX (<https://nginx.org/>), which functions as a reverse proxy, routing incoming requests to the back-end, specifically targeting the WSGI (Web Server Gateway Interface) application server.

Gunicorn (<https://gunicorn.org/>) acts as the WSGI server, enabling efficient and scalable execution of Python web applications by serving as the interface between the web server and the Python application.

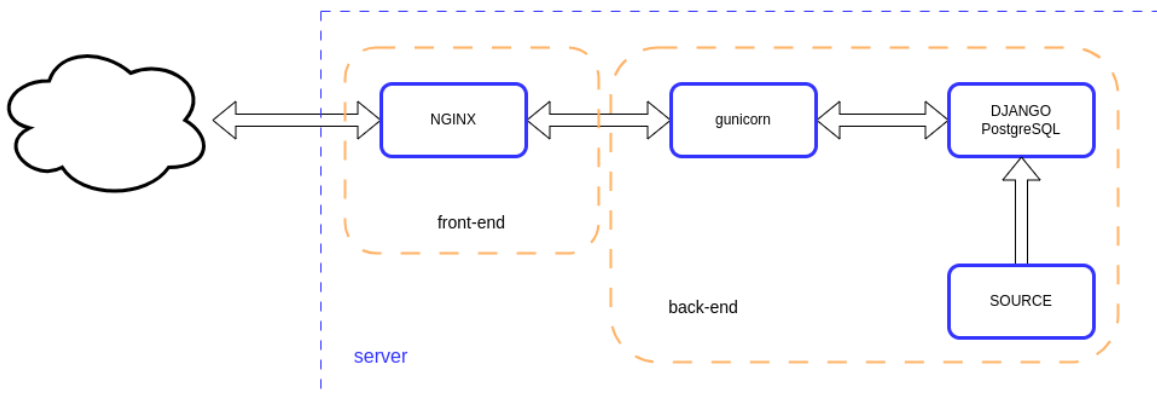


Figure 10 - Logical flow of the prototype web application.

Figure 11 is the implementation scheme on the dedicated INGV physical server which hosts the prototype web application. The server has been acquired with the project's funds, and it is dedicated to the operational implementation of the prototype web application.

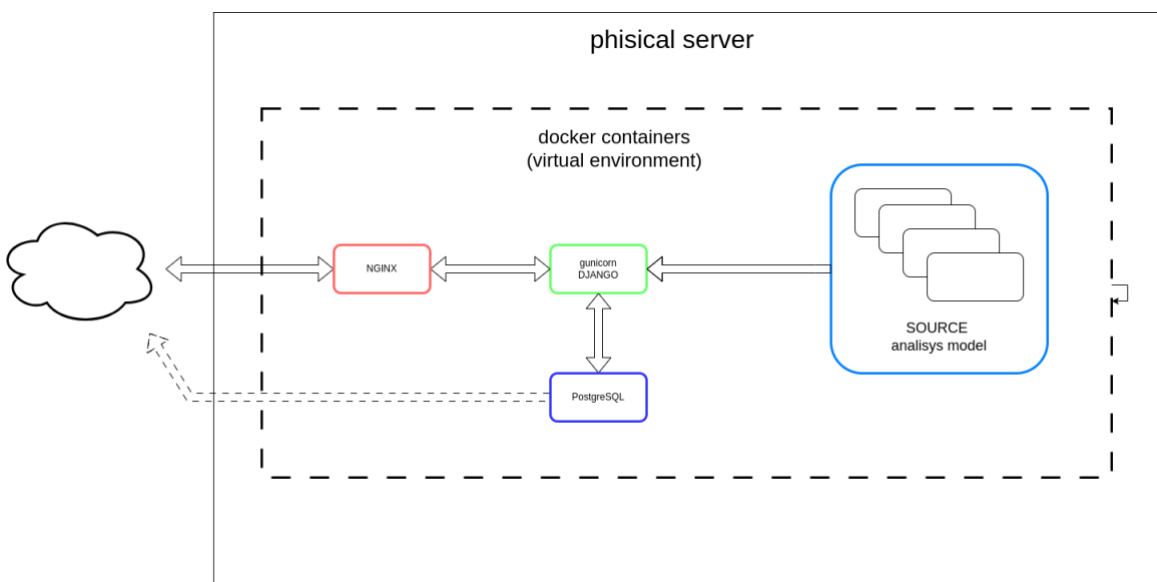


Figure 11 - Schematic of the INGV server hosting the prototype web application.

The Docker platform (<https://www.docker.com/>) has been installed on the physical server, enabling the creation, deployment, and execution of applications within containers, i.e. isolated and portable environments that encapsulate all the necessary components for running an application (code, libraries,

dependencies, and configurations). Containerization allows one or more applications to run in isolated containers, preventing conflicts between different runtime environments.

Docker shares only the host operating system's kernel, compared to other virtualization technologies, making containers more lightweight and faster to start. This approach ensures consistent application behavior regardless of the underlying host system (Windows, Linux, or macOS).

Specifically, a microservices architecture was adopted, separating the NGINX front-end from the back-end, where Gunicorn and Django are combined into a single microservice, as is the PostgreSQL database. The SOURCE analytics itself has been also sub-divided into multiple microservices corresponding to its different processing phases.

The preliminary web page is available at <https://itineris.bo.ingv.it/>.

6. SUMMARY AND FUTURE DEVELOPMENTS

A preliminary version of the prototype web application has been implemented by INGV-BO in a dedicated server. This web application is based on SOURCE analytics tool, first released and described by Oliveri et al. (2022) to compare INGV ocean prediction models or model reanalysis with independent moored observations from the Copernicus Marine Service (CMS) Near Real Time product.

INGV-BO PU committed to adapt the SOURCE tool to process ITINERIS in situ data with open access models from CMS. In situ data from EMSO SS fixed observatories have been used as proof of concept to build the automatic workflow to access, to process in situ and model data and to store the output in a PostgreSQL database for its easy visualization in a webGIS application.

There are many tasks that need fine tuning:

1. to further verify that the secondary QC procedure is fully effective on deep observations, presenting very different variability from coastal observations considered in the initial SOURCE implementation;
2. to improve the <https://itineris.bo.ingv.it/> web page design through user's engagement (e.g. EMSO facility managers and modelers) in order to better fit their needs;
3. to integrate the prototype within the IT-IOOS Marine Data Portal.

The prototype integration process in IT-IOOS started with the discussion of a common strategy with the WP5 coordinator and the MDP developers. The first step to accomplish this is to access the data through the unique MDP entry point, once ready, which will enhance data and service interoperability. Next, we will explore the possibility of replicating and exposing the PostgreSQL service in the Marine Data Store to integrate the SOURCE output visualization into the IT-IOOS portal.

Additional functions could be added to the SOURCE analytics tool to include ITINERIS ocean models among the CMS products and to extend it to other RIs in situ data time series, i.e. Jerico.

Bibliography

Oliveri P, Simoncelli S, Di Pietro P, Fratianni C, Mattia G, Delrosso D and Guarnieri A (2022) SOURCE: Sea Observations Utility for Reprocessing, Calibration and Evaluation. Front. Mar. Sci. 8:750387. doi: 10.3389/fmars.2021.750387

Paolo Oliveri, & Simona Simoncelli. (2022). SOURCE v1.4.0 (1.4.0). Zenodo. <https://doi.org/10.5281/zenodo.6319836>