SOURCE software's reprocessing and merging of different sea temperature and salinity time series data collections from SeaDataCloud and CMEMS

Paolo Oliveri [1] **Simona Simoncelli** [1] **Pierluigi di Pietro** [1] **Gelsomina Mattia** [1]

paolo.oliveri@ingv.it

Goal

Harmonize, integrate and reprocess observational fixed platforms data coming from different european infrastructures.

Benefits

1. Increase the number of platforms;
2. increase the data coverage;
3. increase the available fields.

Applications

- Better evaluate Ocean Circulation Model data (Cal/Val) (analysis, reanalysis, etc.);
- show, download and analyze good observational data;
- continuous monitoring of coastal environment;
- develop services of integrated coastal monitoring systems;
- set up early-warning systems for coastal environmental protection and preservation.

Problems

1. Observational data being fragmented between different repositories and infrastructures!
2. data collation and reprocessing requires experience and specific skills!

SeaDataCloud ODV Database

- Global Ocean dataset released in binary, user friendly Ocean Data View format;
- database management plan originally based for profiles, but adapted for time series;
- analyzed database: Mediterranean Sea and part of the Atlantic Ocean from the Canary Islands to the Gulf of Biscay;
- Only time series data from fixed platforms analyzed.

Inner database fragmentation

Most of the time series in the database were split into a large amount of subsets, sometimes **one platform for each time step**!

SeaDataCloud pre processing procedure

- Find and aggregate all broken time series using likeness in ID parameter strings (LOCAL_CDI_ID, EDMO_code, etc.);
- organize metadata;
- time units correction;
- filter by area of interest or instrument type;
- produce information on the original Q/C scheme by SeaDataNet infrastructure;
- produce log files with all the problems encountered (missing time, depth, data, wrong Q/C variables, etc.).

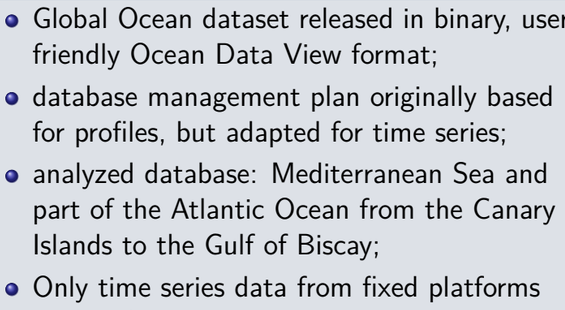


Figure 1: Merged platforms per data originator

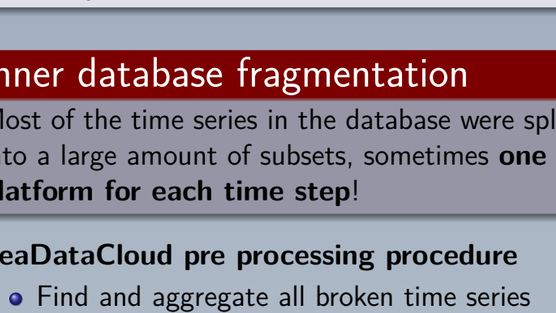
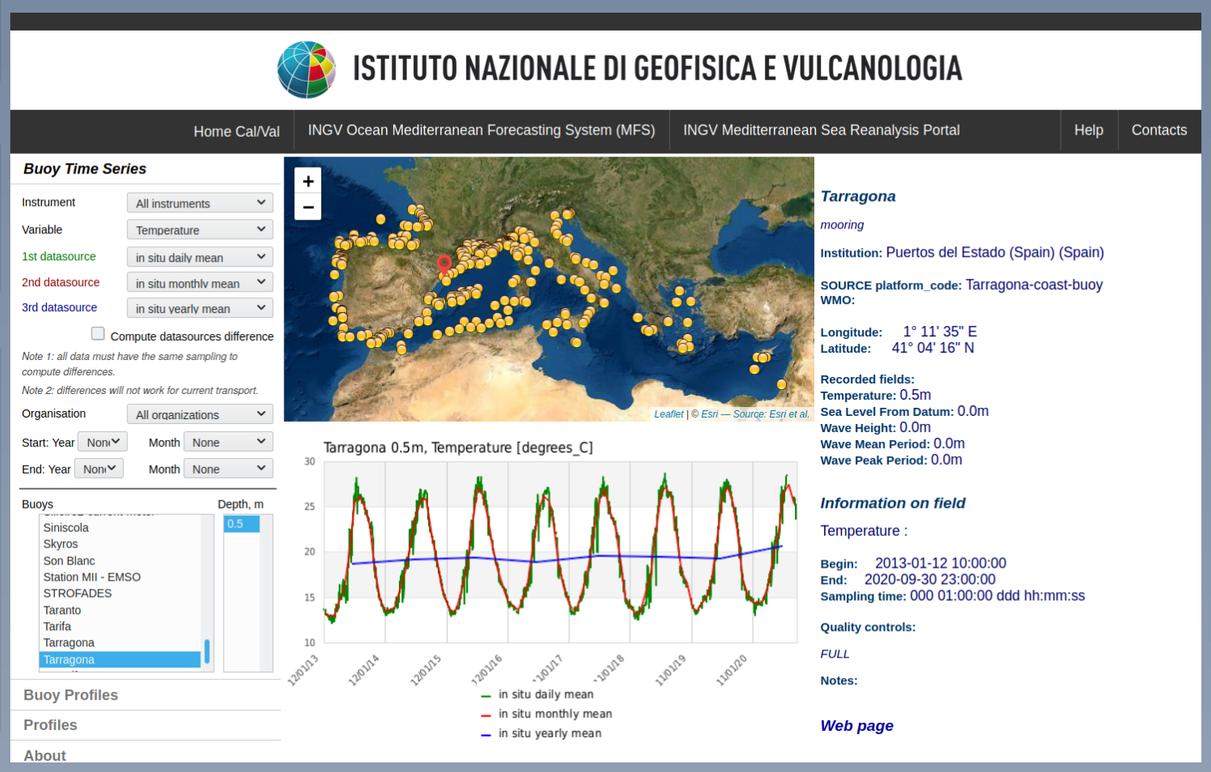
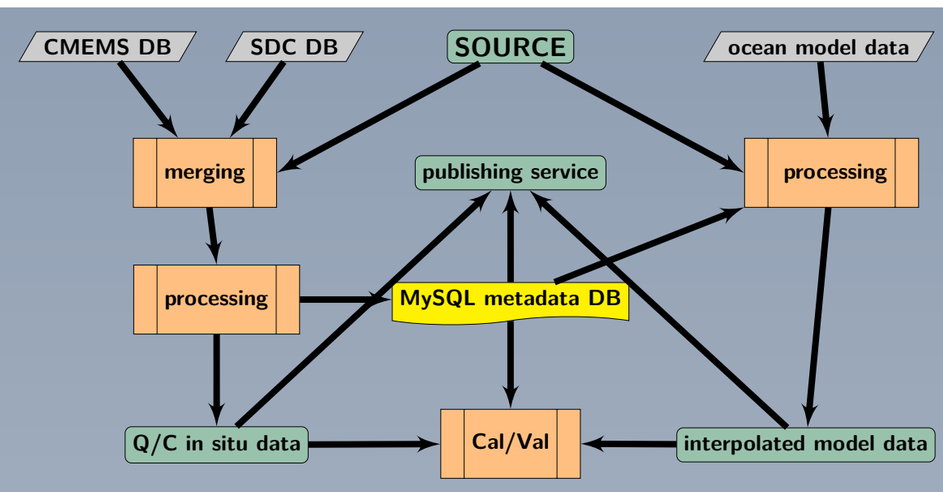


Figure 2: total amount of fragmentation per data originator





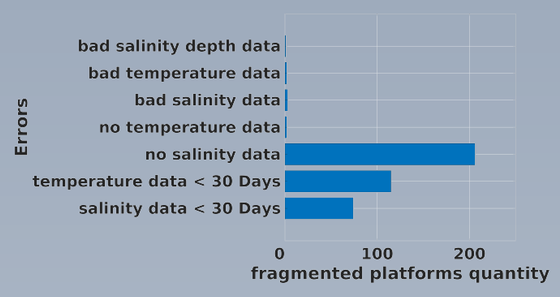
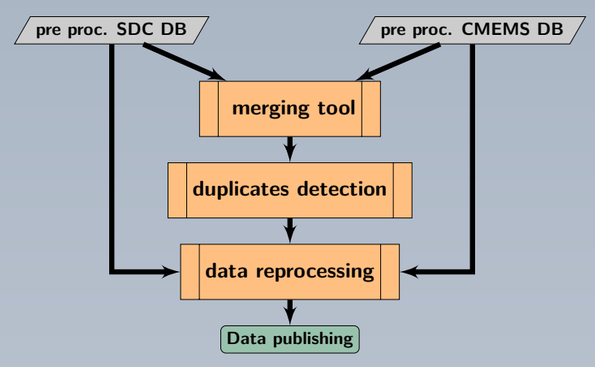


Figure 3: Detected errors in platform data storage

SDC and CMEMS integration



Merging conditions

1. 500 m of horizontal proximity;
2. likeliness in CMEMS platform_code and SDN LOCAL_CDI_ID;
3. likeliness in platform name (if available);
4. likeliness in WMO (if available).

If two platforms are taken for merging, every time series at the same depth level for both platform will be concatenated.
Given $z_1, \ldots, z_n$ and $z'_1, \ldots, z'_m$ the recorded depths for the two platforms, the merged platform will have the depths $z"_1, \ldots, z"_p$, such that:

$$z" = \begin{cases} z' & \text{if } z' \notin \{z_1, \ldots, z_n\}; \\ z & \text{if } z \notin \{z'_1, \ldots, z'_m\}; \\ z(=z') & \text{otherwise,} \end{cases}$$

For each depth, the available data for only one database will be copied.

Merged data before the publishing phase **must** be checked for duplicates and reprocessed by passing several tests:

SOURCE Q/C procedures

- Global range check;
- spike test;
- stuck value test;
- statistic iterative test based on **KDE** (Kernel Density Estimation).

Aggregated database

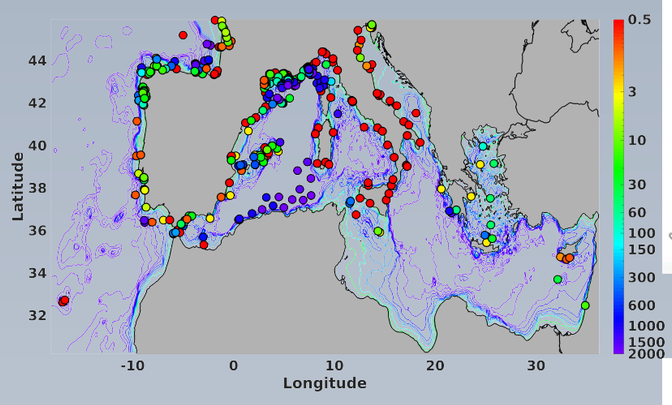


Figure 4: sea temperature platform colored with their deepest depth sensor

Data publishing service

New web service developed, published and maintained at INGV:

- Easy search for data using an open-source geo referenced and intuitive interface;
- Visualize, compare and analyze all available products;
- More functionalities are planned.

The web interface will be redesigned to meet Web 4.0 responsiveness standards and allow the complete data access and download.
In conclusion user will be able to near real time monitor the coastal environment, assess the performances of ocean circulation models and access all available services and applications based on this service.

Possible solution

Use of SOURCE's merging tool to process the data and have the most possible data available!

SOURCE

SOURCE (Sea Observations Utility for Reprocessing, Calibration and Evaluation) is written in Python.
SOURCE handles:

- Conversion of all data coming from the different infrastructures in an unique common data format (netCDF);
- merging in one final database using accurate collation procedures and duplicates detection;
- processing and quality control analysis using several tools that follow the ocean best practices;
- model data management for Cal/Val.

Test case subject

- Mediterranean sea fixed platforms **near real time** data from **CMEMS** in situ **TAC** ftp://nrt.cmems-du.eu/;
- **SeaDataCloud** time Series Data Collection (unpublished version TODO)
- **temperature** and **salinity** fields;
- all available date range.

CMEMS DB reprocessing and Cal/Val procedures details here: IMDIS2018_Proceedings.pdf, pp. 339.

Merging example: Lesvos platform

Here we present an example of data merging for the HCMR Lesvos platform.

- Field: sea temperature;
- Depth: 3 m;
- Time range: 2005-01-01 –> 2007-12-31;
- Data type: in situ daily mean.

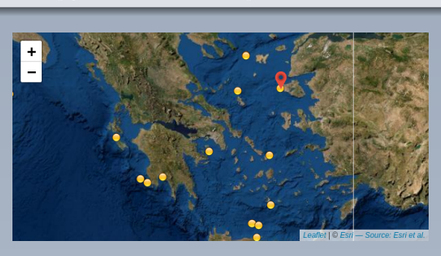


Figure 5: Lesvos platform location.

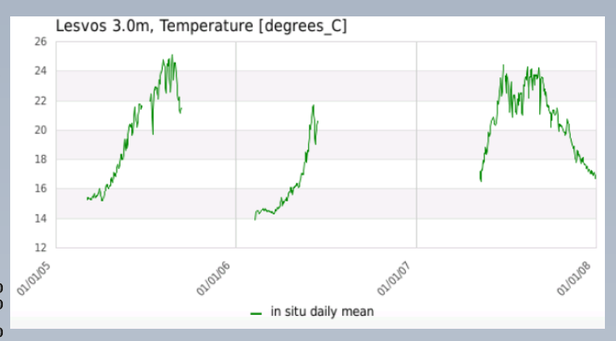


Figure 6: CMEMS available data at Lesvos platform

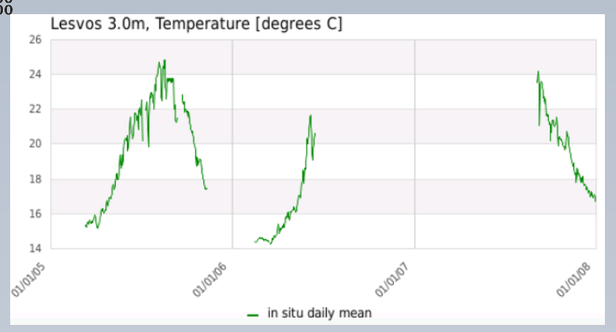


Figure 7: SeaDataCloud available data at Lesvos platform

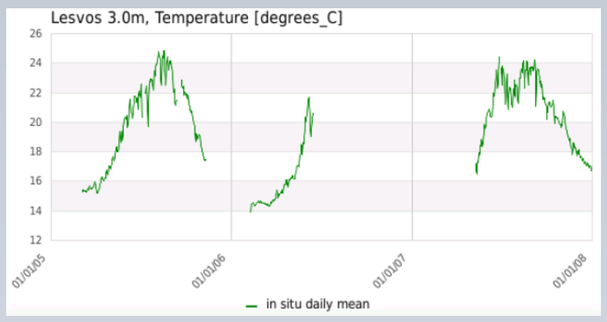


Figure 8: Integrated available data at Lesvos platform

[1] **Istituto Nazionale di Geofisica e Vulcanologia**