

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

Paolo Oliveri, Istituto Nazionale di Geofisica e Vulcanologia (Italy), paolo.oliveri@ingv.it
Simona Simoncelli, Istituto Nazionale di Geofisica e Vulcanologia (Italy), simona.simoncelli@ingv.it
Pierluigi di Pietro, Istituto Nazionale di Geofisica e Vulcanologia (Italy),
pierluigi.dipietro@ingv.it
Gelsomina Mattia, Istituto Nazionale di Geofisica e Vulcanologia (Italy), gelsomina.mattia@ingv.it

Several European, regional and global marine data, and data product initiatives provide access to marine data (Martín Míguez et al., 2019). CMEMS and SeaDataNet initiatives aggregate, archive, quality assess, update and distribute the data with a different scope (<https://www.seadatanet.org/About-us>). The CMEMS in situ TAC gathers both Near Real Time (NRT) and Delayed Mode (DM) observations to serve the EU Monitoring and Forecasting Centers, while the SeaDataNet diffuse network of National Oceanographic Data Centers mainly aggregate DM marine data coming from observing systems and research programs. This implies that the integration of the their DM data collections is of paramount importance to assure the usage of the largest data sets for scientific applications.

The data originating from a specific data provider may be available only in one of the two infrastructures or their availability could be fragmented in terms of temporal and spatial data coverage and resolution or in terms of the distributed parameters. This may result in partial data availability, with great impact for the researchers who wants to access as much data as possible.

The objective of this work is to analyze the mooring time series data from CMEMS in situ TAC and SeaDataNet, compare and integrate them when necessary in order to obtain the most complete data set for temperature and salinity over the Mediterranean Sea.

Due to the different data management procedures, it was necessary to first convert all the data in a unique common data format (netCDF). Then it has been applied an accurate quality control analysis, a duplicates detection and process.

This data collation and reprocessing requires specific skills on how to manipulate the data, that final users might not have or it would be very costly for them.

INGV team developed a software called **SOURCE** (Sea Observations Utility for Reprocessing, Calibration and Evaluation) written in Python. SOURCE handles the pre-processing of the data coming from the different infrastructures, their merging in one final database and the post-processing by filtering out the duplicates. The final data quality assurance and consistency process consists of several tools that follow the ocean best practices guidelines.

The software was already ready to handle the CMEMS in situ TAC Mediterranean NRT data set with the main target to quickly acquire and publish data into an operational chain already in place. Then a first data merging test has been performed between SeaDataCloud V2 and CMEMS temperature and salinity data collections.

The SeaDataCloud time series data collection is a Global Ocean soon-to-be-published dataset that could represent a reference for the users. It has been released in binary, user friendly Ocean Data View format, originally implemented for vertical profiles, but successively adapted for time series. The pre-processing of those data was very difficult in order to concatenate and aggregate the different time series belonging to the same platform through likeness of ID parameter strings, organizing metadata,

harmonizing time units, filtering the data by area of interest or instrument type, producing information on the original SeaDataNet QC scheme and producing log files to track the problems encountered during the processing steps (i.e. missing time, depth, data, wrong Q/C variables).

The reconstructed SeaDataCloud time series, divided by parameter and organized in a fit for purpose dataset, gave the possibility to start the integration process with the already pre-processed CMEMS time series in order to obtain the final dataset and maximize the data coverage. Statistics have been computed per each platform in order to further QC the temperature and salinity time series and discard the data anomalies.

The data has been processed and published through a new INGV data access and visualization service that allows to visual inspect the content of the obtained aggregated database and its statistics. Moreover the INGV web service allows to visualize the synthetic time series extracted from INGV model analyses and reanalyses at each mooring location. This functionality could also be expanded to the available CMEMS prediction model products.

This work allows a continuous monitoring of the coastal environment through the all available existing moorings using both observational and model data displaying high quality data but it represent a continuous calibration/validation tool to assess the skill of the existing models This is the base from which to start to develop new services and applications to serve a variety of different users that perform operations at sea, integrated coastal monitoring systems, early-warning system for coastal environmental protection and preservation.

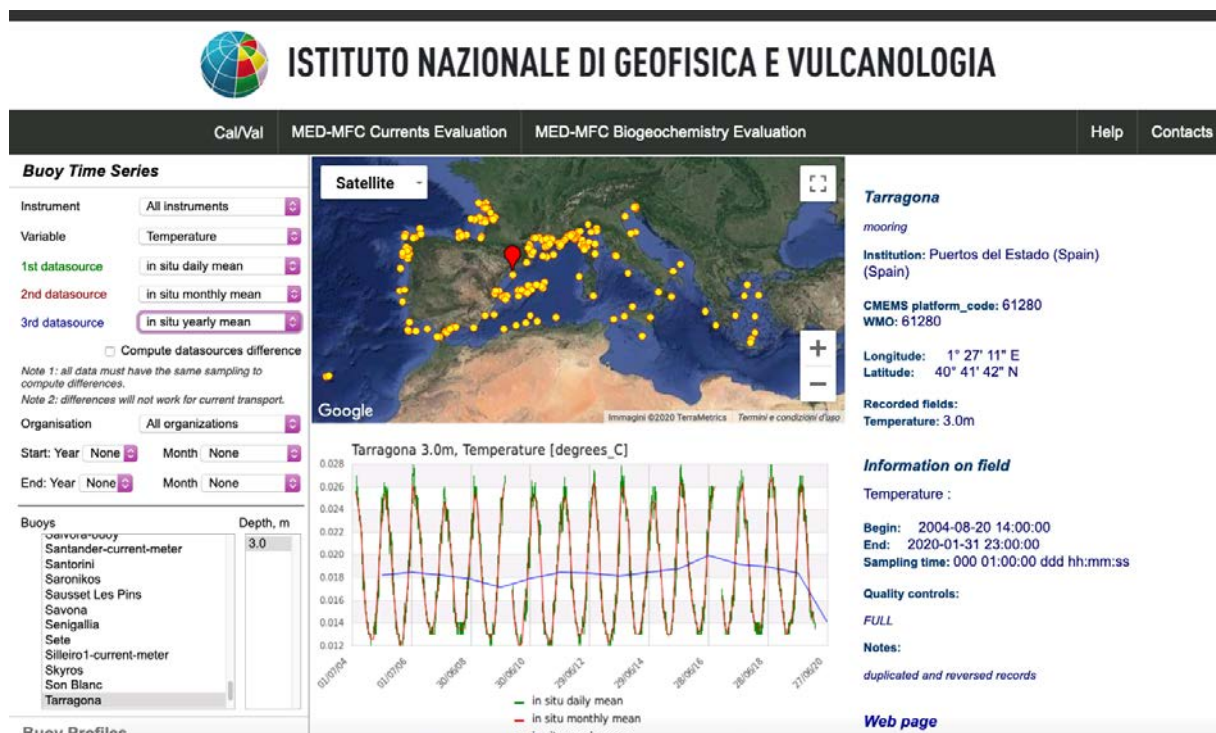


Figure 1. Snapshot of the new INGV service portal for visualization and evaluation of processed in situ observations from fixed platforms.

Martín Míguez B, Novellino A, Vinci M, Claus S, Calewaert J-B, Vallius H, Schmitt T, Pititto A, Giorgetti A, Askew N, Iona S, Schaap D, Pinardi N, Harpham Q, Kater BJ, Populus J, She J, Palazov AV, McMeel O, Oset P, Lear D, Manzella GMR, Gorringer P, Simoncelli S, Larkin K, Holdsworth N, Arvanitidis CD, Molina Jack ME, Chaves Montero MM, Herman PMJ and Hernandez F (2019) The European Marine Observation and Data Network (EMODnet): Visions and Roles of the Gateway to Marine Data in Europe. *Front. Mar. Sci.* 6:313. doi: 10.3389/fmars.2019.00313