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\(^1\)

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

Benefits
- Increase the number of platforms;
- Increase the data coverage;
- 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
- Observational data being fragmented between different repositories and infrastructures!
- Data collation and reprocessing requires experience and specific skills!

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 log files with all the problems encountered (missing time, depth, data, wrong Q/C variables, etc.).

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 and merging
- 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 log files with all the problems encountered (missing time, depth, data, wrong Q/C variables, etc.).

Merging conditions
- 500 m of horizontal proximity;
- Likeness in CMEMS platform_code and SDN LOCAL_CDI_ID;
- Likeness in platform name (if available);
- Likeness 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 \( z_1, \ldots, z_n, z'_1, \ldots, z'_m \), such that:

\[
\begin{align*}
  z & = \begin{cases} 
  z' & \text{if } z' \in \{z_1, \ldots, z_n\}, \\
  z & \text{if } z \in \{z'_1, \ldots, z'_m\}, \\
  \text{otherwise},
  \end{cases}
\end{align*}
\]

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

SOURCE Q/C procedures
- Global range check;
- Spike test;
- Stuck value test;
- Statistical iterative test based on KDE (Kernel Density Estimation).

Test case subject
- Mediterranean sea fixed platforms near real time data from CMEMS in situ TAC ftp://art.cmems-du.eu/;
- SeaDataCloud time Series Data Collection (unpublished version TODD)
- Temperature and salinity fields;
- All available data range.


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