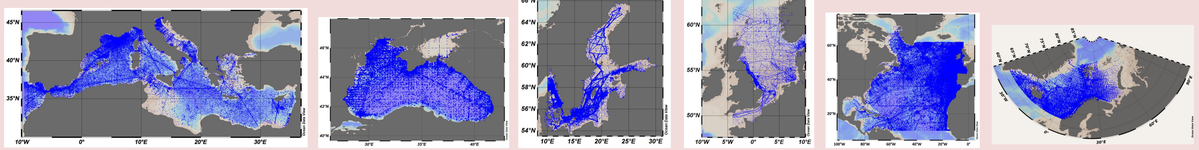
The SeaDataNet data products
regional temperature and salinity historical data collections

S. Simoncelli[1], C. Coatanoan[2], O. Bäck[3], H. Sagen[4], S. Scory[5], V. Myroshnychenko[6], D. Schaap[7], R. Schlitzer [8], S. Iona[9,10], M. Fichaut[2]

*(1) **simona.simoncelli@ingv.it**, INGV, (2) Ifremer, (3) SMHI, (4) IMR, (5) RBINS, (6) METU, (7) MARIS, (8) AWI, (9) HCMR, (10) Univ. of Thessaly*

INTRODUCTION

Qualified historical data collections of all unrestricted temperature and salinity measurements contained within SeaDataNet database were produced for the European sea basins (Mediterranean Sea, Black Sea, Baltic Sea, North Sea, North Atlantic Ocean, Arctic Sea) covering the time period 1900-2014.

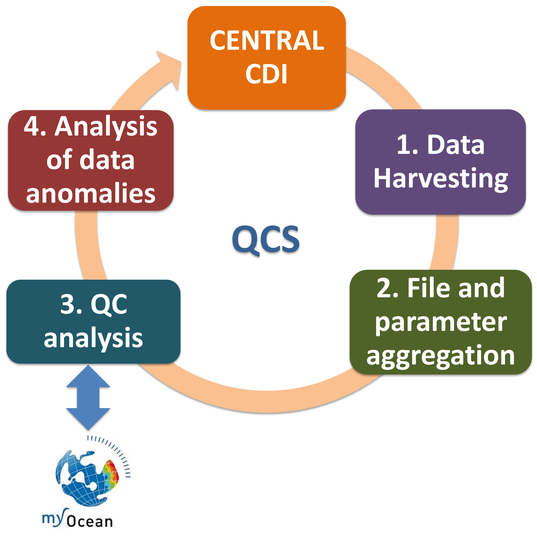


Two versions of the data collections have been published within the framework of SeaDataNet II project and they represent a snapshot of the database content at two different times: **V1.1** (January 2014) and **V2** (March 2015). They are available through the SeaDataNet web catalog at http://sextant.ifremer.fr/en/web/seadatanet/ and are all minted by permanent doi. V1.1 regional data sets have been used to compute regional monthly climatologies of T and S fields also available on the web catalogue (http://www.seadatanet.org/Products/Aggregated-datasets).

SeaDataNet Quality Control Strategy (QCS)

SeaDataNet II project implemented and continuously refined a Quality Control Strategy aiming at improving the quality of the database content and creating the best products deriving from it. The QCS was originally implemented in collaboration with MyOcean2 and MyOcean Follow On projects in order to develop a true synergy at regional level to serve operational oceanography and climate change communities. The QCS consists of four main phases:

The **data anomalies** were organized in files by EDMO code and sent to the data providers asking for inspection and correction. Guidelines were provided to explain the information in the anomalies files, how to make corrections on the data and how to send back the resulting information in detailed reports. The confirmed anomalies were corrected within the CDI improving the quality of the database.



Quality Control analysis of T&S collections has been conducted at regional level by Regional Coordinators (RC) and an harmonized procedure was designed in collaboration with MyOcean In-situ TAC (Thematic Assemble Centre) to facilitate data and information flow. RC and In-situ TAC both identified data anomalies through their QC procedures and defined a common plan to improve the quality of SDN database content and consequently of the next data collection release.

Data Harvesting of T and S files was performed through a CDI Robot, a robot user that uses the CDI (Common Data Index) Data Discovery and Access Service to query, shop and retrieve data sets from the SDN distributed data centres in automatic way. Data query consisted in searching for all data sets with T&S, whose access was unrestricted or under SDN License. The Robot was triggered to start harvesting the related ODV files from the distributed data centres through the general CDI shopping mechanism RSM-DM (Request Status Manager and Download Manager).

Data (SDN/ODV format) and csv files with CDI metadata were processed using ODV software in 2 steps. **File aggregation**: all data files were analysed determining the type of data (profile, trajectory or time-series) and the set of parameters. Data were added to the metadata enriched ODV collection importing the full set of CDI metadata, instrument information and references. **Parameter aggregation**: ODV software aggregated parameters using up-to-date versions of the P35 vocabulary and the AWI unit conversion database.

The approach is iterative to facilitate the upgrade of the database and it allows versioning of data products through the release of new data collections at the end of each QCS loop and the generation of the derived climatological products after a certain time lag. During SDN II project the QCS loop was completed three times. A first trial loop produced V1 regional data collections that were not released. A second loop permitted to refine each technical phase and improve the quality of both the central CDI content and the data collections that were released as version V1.1. A third QCS loop gave origin to the V2 regional aggregated datasets.

Quality Control Analysis

Temperature and Salinity data collections, one per each European marginal sea (Arctic Sea, Baltic Sea, North Sea, North Atlantic Ocean, Mediterranean Sea, Black Sea) were analysed at regional level to assess and report on their quality. A common and basic **QC analysis** was performed using ODV software (http://www.seadatanet.org/Standards-Software/Software/ODV):

- Analysis of data distribution and data density to identify possible spatial gaps and missing data sets (FIG.1);
- Analysis of temporal (annual/seasonal) data distribution (FIG.1);
- TS scatter plots of the entire dataset and range check;
- Visual control of scatter plot of observations considering various SDN Quality Flags (QF) to identify wrong profiles (outliers, spikes)(FIG.2);
- Analysis Quality Flag statistics;
- Identification of stations falling on land;
- Identification of wrong or missing data.

The scatter plots of the regional data sets highlighted the necessity of applying specific sub-regional checks (FIG.3), per areas and per depth, and stability check on density to point out observations with depth, T and S out of reasonable values. Some good data (QF=1,2) presented values out of ranges and their QFs were modified accordingly. Many data resulted not QCed (QF=0) but looked reasonable and were further analyzed.

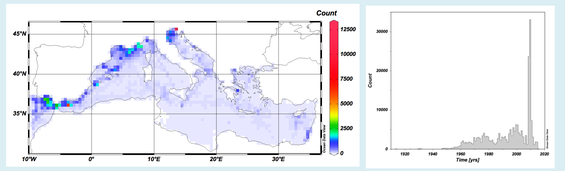


FIG.1 Mediterranean V2 Data Collection: Data density map and annual time distribution

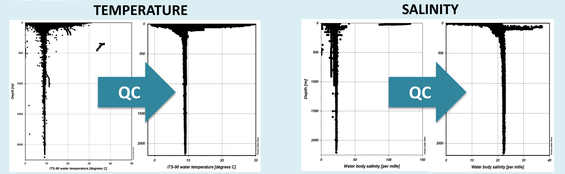


FIG.2 Black Sea V2 Data Collection: (left) Temperature and Salinity (right) scatter plots of measurements with QF=1 before and after additional QC analysis.

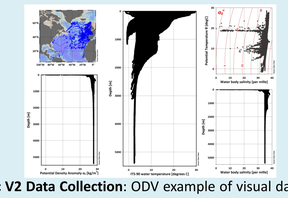


FIG.3 North Atlantic V2 Data Collection: ODV example of visual data inspection (map, σ_0 versus depth, temperature versus depth, potential temperature versus salinity and salinity versus depth).

Conclusions

The production of SDN historical data collections was an extensive and constructive exercise to manage more than 1 Million data and involving many people and institutions, 62 data centres and more than 300 data originators. The implemented **Quality Control Strategy** permitted to identify and correct lots of data and it highly improved the quality of SDN database. The deriving **data collections** present an increasing quality thanks to the additional QC analysis performed at regional level and can be used for further applications. V1.1 collections have been used to compute regional Temperature and Salinity climatologies (V1.1) already available through the SDN web catalogue at http://sextant.ifremer.fr/en/web/seadatanet/, V2 climatologies are in progress.

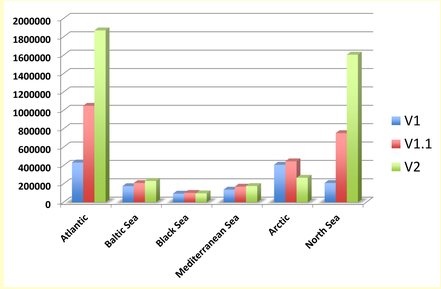


FIG.4 Histogram representing SeaDataNet population increase per each sea basin computed analysing the number of data for V1, V1.1 and V2 regional data collections.

SDN data population increased due to 3 QCS iterations that produced V1, V1.1 and V2 data collections (FIG.4). V2 regional collections show a general increment, especially in the Atlantic and North Sea regions due to the insertion of new data. The Black Sea presents a reduction of number of stations from V1.1 and V2 due to the removal of data from the central CDI after the Ukrainian crisis and duplicates elimination. The data decrease in the Arctic Ocean (-35%) is due to an error in the ODV format of the Norwegian time series detected through a more strict control by ODV software. Norwegian files have been corrected to be completely compliant with format specifications.

http://www.seadatanet.org/