Update of the reference dataset in the Mediterranean and Black Seas for DMQC activity of Argo data

by

A. Gallo and G. Notarstefano

Approved for release by:

Dr. Cosimo Solidoro
Director, Department of Oceanography

Table of Contents

1. INTRODUCTION

This document describes the activities related to the conductivity-temperature-depth (CTD) reference dataset used for the Delayed-Mode Quality Control (DMQC) of the MedArgo fleet (figure 1) in the Mediterranean and Black Seas.
OGS is responsible, in the framework of the MedArgo Regional Center (ARC), for the DMQC of temperature and salinity data of floats deployed in these two marginal seas. The DMQC analysis is also part of the work planned in tasks 2.4 and 3.2 of the H2020 Euro-Argo RISE project.

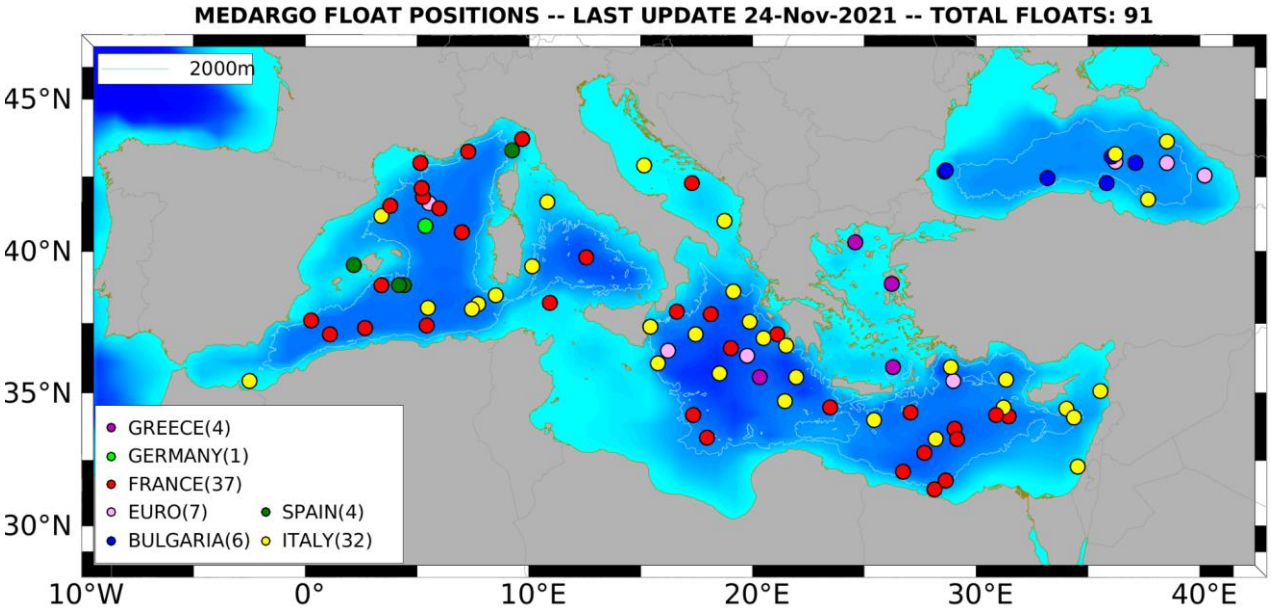

Figure 1. MedArgo float positions as of 24 November 2021 (updated daily).

The reference dataset plays an important role in the DMQC analysis because the accuracy of the float data has been assessed by comparison with high-quality shipboard measurements. OGS, as responsible of the DMQC activities in the Mediterranean and Black Seas, reviews and improves on a regular basis the availability of high-quality ship-based CTD reference data for the quality control of core and deep Argo physical data. It collects CTD data in complement of the official CTD reference dataset, provided by the Coriolis Global Data Assembly Center (GDAC), using mainly two approaches: personal contacts and regional data services. This collection is particular important for the Mediterranean Sea which is characterized by a complex bathymetry and geography and where water masses can go through dramatic changes over the years

(Notarstefano G., 2019). For this reason, it is crucial to have an up-to-date reference dataset with the best co-location (in space and time) between the CTD reference profiles and Argo profiles.
Two different CTD reference datasets are provided and updated on a regular basis. One reference dataset, provided by the Coriolis GDAC, consists of CTD data collected worldwide by several research institutes and the other one by the main European Marine Services (hereby referred as to **CTD reference dataset**). Once a new version of the CTD reference dataset is ready, it is available for downloading at a dedicated FTP address. Once downloaded, data were converted in mat format to be used in OWC delayed-mode procedure (Cabanes et al., 2016). A quality control is applied such as an additional visual check to avoid spike or duplication. Another type of reference dataset is built using the Argo CTD profiles that respect a set of predefined criteria (hereby referred as to **Argo CTD reference dataset**), as stated in the Argo quality control manual for CTD and trajectory data, version 3.5 (Wong et al, 2021).
This report describes the work done to improve the official CTD reference dataset with other new and more updated CTD data in the marginal seas context like the Mediterranean and Black Seas.

2. REFERENCE DATASETS

In this chapter each reference dataset is described. Files are organized following the World Meteorological Organization (WMO) boxes denomination that consists of geocodes that divides the chart of the world into 10° latitude by 10° longitude grid cells (figure 2).

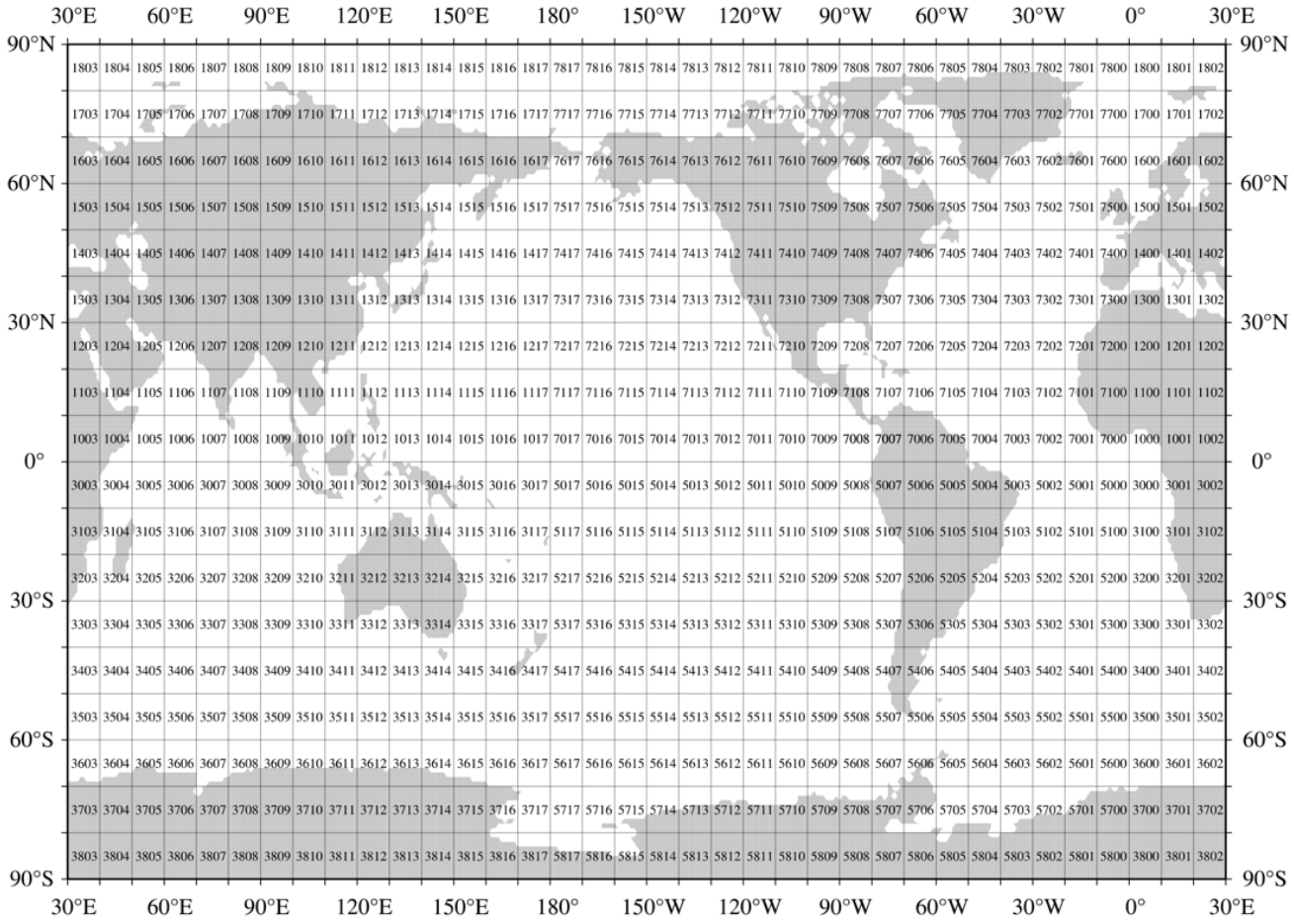

Figure 2. World Meteorological Organization (WMO) 10-degree squares.

The 10-degree squares that cover the Mediterranean and Black Seas area are the following:

1300, 1301, 1302, 1303, 1400, 1401, 1402, 1403, 1404, 7300, 7400

2.1. CTD reference dataset

The CTD reference dataset is provided by the Coriolis in-situ Service for operational oceanography and it is usually updated once or twice a year. At the time of this report, the latest version available is "CTD_for_DMQC_2021V01". Hence, the above files are downloaded by the Coriolis FTP address, uncompressed and saved in MatLab (.mat) format. The spatial and temporal distribution of these CTD profiles is shown in figure 3 and figure 4, respectively. The data are distributed from 1972 to 2011. As in Notarstefano (2019) the temporal distribution exhibits a consistent increase until the year 1995 and then a sharp decrease is evident. Some areas are highly under sampled, like the Southern Ionian and Algerian, and the Thyrrhenian Sea.

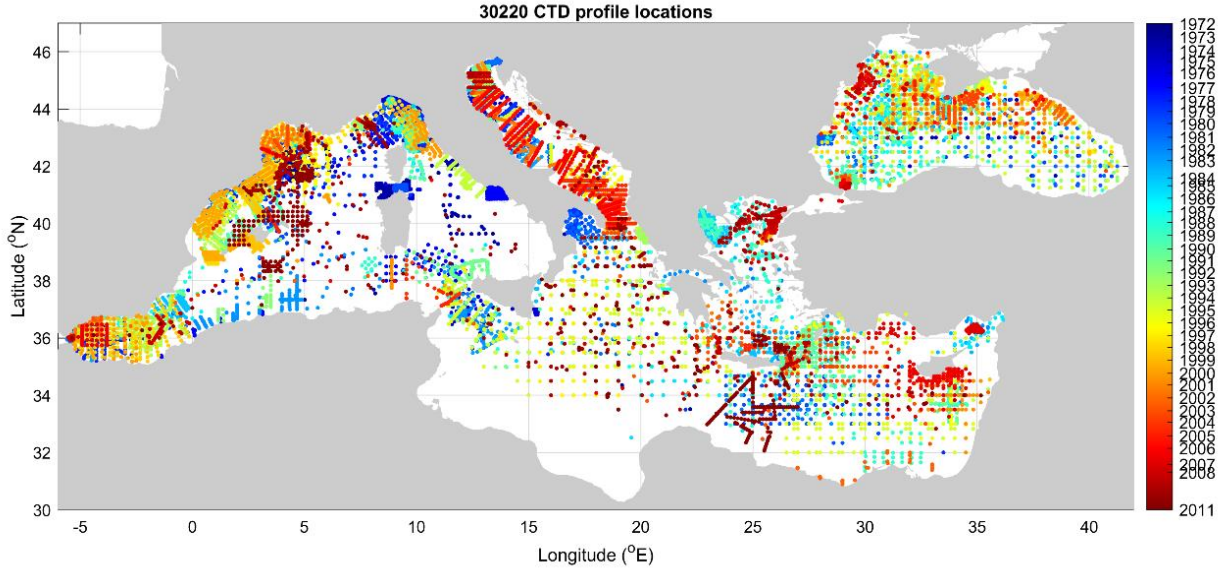

Figure 3. Spatial distribution, color-coded for time, of the CTD profiles available in the CTD reference dataset version "CTD_for_DMQC_2021V01".

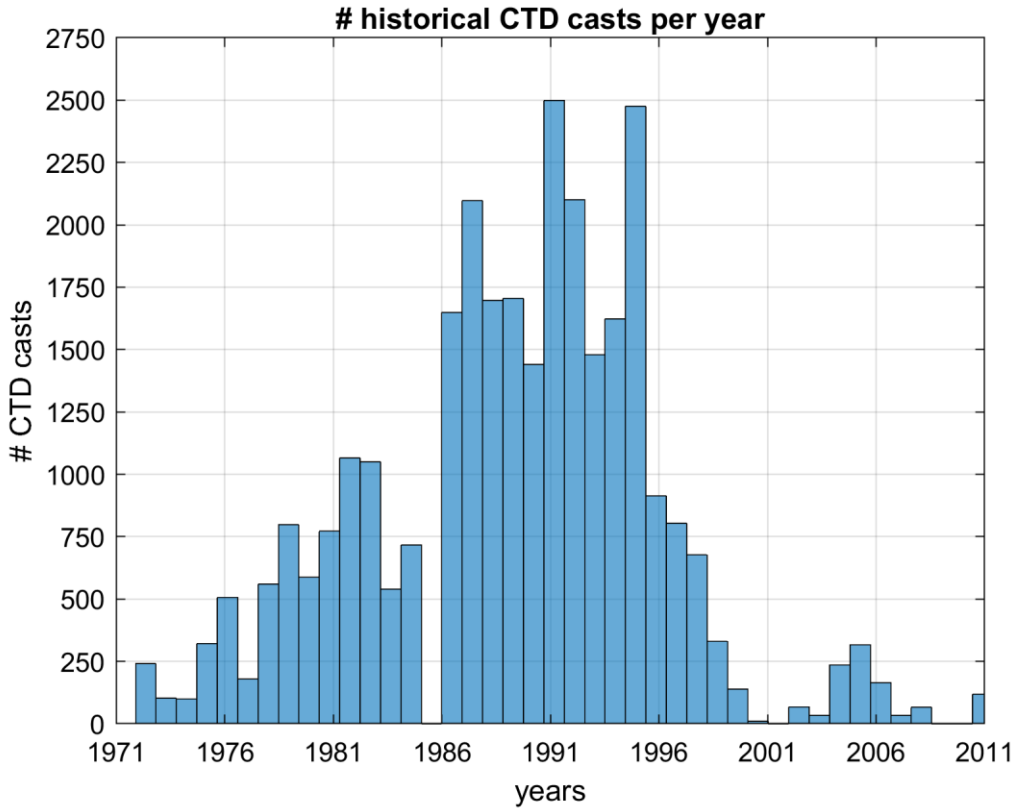

Figure 4. Temporal distribution of the CTD profiles available in the CTD reference dataset version "CTD_for_DMQC_2021V01".

2.2. Argo CTD reference dataset

The Argo CTD reference dataset is also provided by the Coriolis in-situ service for operational oceanography and files are organized in the same way as for the CTD reference dataset. The latest version available at Coriolis is "ARGO_for-DMQC_2020V03".

This dataset has been greatly updated with respect to the previous version. Figure 5 shows the spatial distribution in the several sub-basins of the Mediterranean Sea and also in the Black Sea. Shallow areas, where Argo floats are not often deployed, don't have an accurate coverage (Sicily Channel, Aegean and Middle-North Adriatic). The temporal distribution (figure 6) shows that the Argo CTD profiles are quite well homogeneously distributed with an increment between 2012 and 2014, especially in 2013.

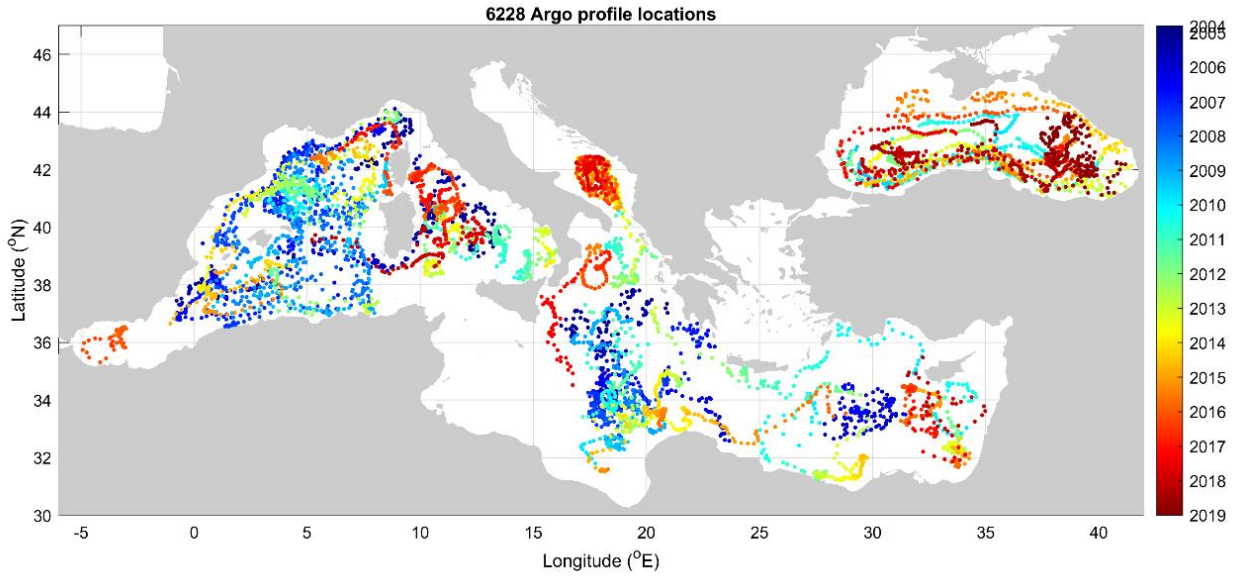

Figure 5. Spatial distribution, color-coded for time, of the Argo CTD profiles available in the Argo CTD reference dataset version "ARGO_for_DMQC_2020V03".

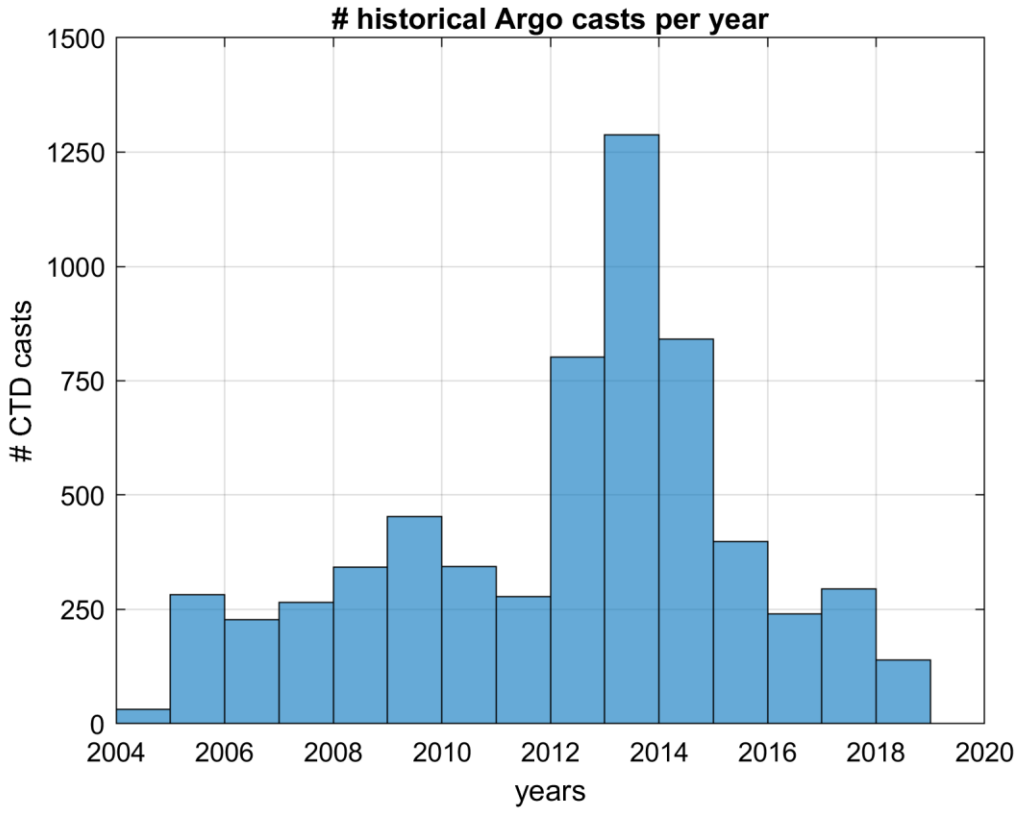

Figure 6. Temporal distribution of the Argo CTD profiles available in the Argo CTD reference dataset version "ARGO_for_DMQC_2020V03".

3. IMPROVING OF THE CTD REFERENCE DATASET

New CTD data have been collected to improve the official CTD reference dataset and try to fill some gaps in temporal and spatial coverage in some areas.

3.1. CTD profiles obtained through dedicated services

The Copernicus Marine Environment Monitoring System (CMEMS) provides a great quantity of CTD data that follows a multiple level quality control procedure. These data are not integrated in the CTD reference dataset by the Coriolis in-situ Service. Black Sea and Mediterranean Sea data are taken from the INSITU_GLO_TS_REP_OBSERVATIONS_013_001_b product that is the delayed mode product designed for reanalysis purposes that integrates the best available version of in situ data for temperature and salinity measurements. It is updated on a yearly basis and the last review has been done in 2019, 30th of June.
The CMEMS files are in NetCDF format and are available through a dedicated FTP server. The profiles are grouped per platform and data type. The CTD files were extracted in two different data repository due to the large amount of data. For the same reason, the procedure to convert the files from NetCDF to MatLab format has been done separately to collect the CTD data in the Mediterranean Sea (figures 7 and 8) and in the Black Sea (figures 9 and 10), respectively.

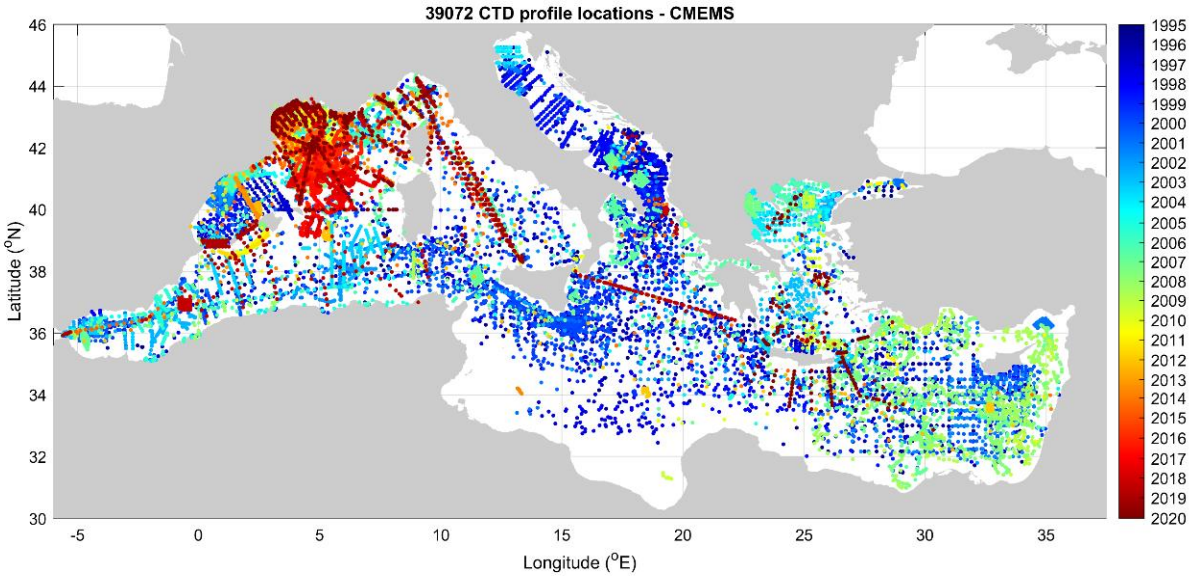

Figure 7. Spatial distribution, color-coded for time, of the CTD profiles collected through the CMEMS portal and used as a complement of the CTD reference dataset of the Mediterranean Sea.

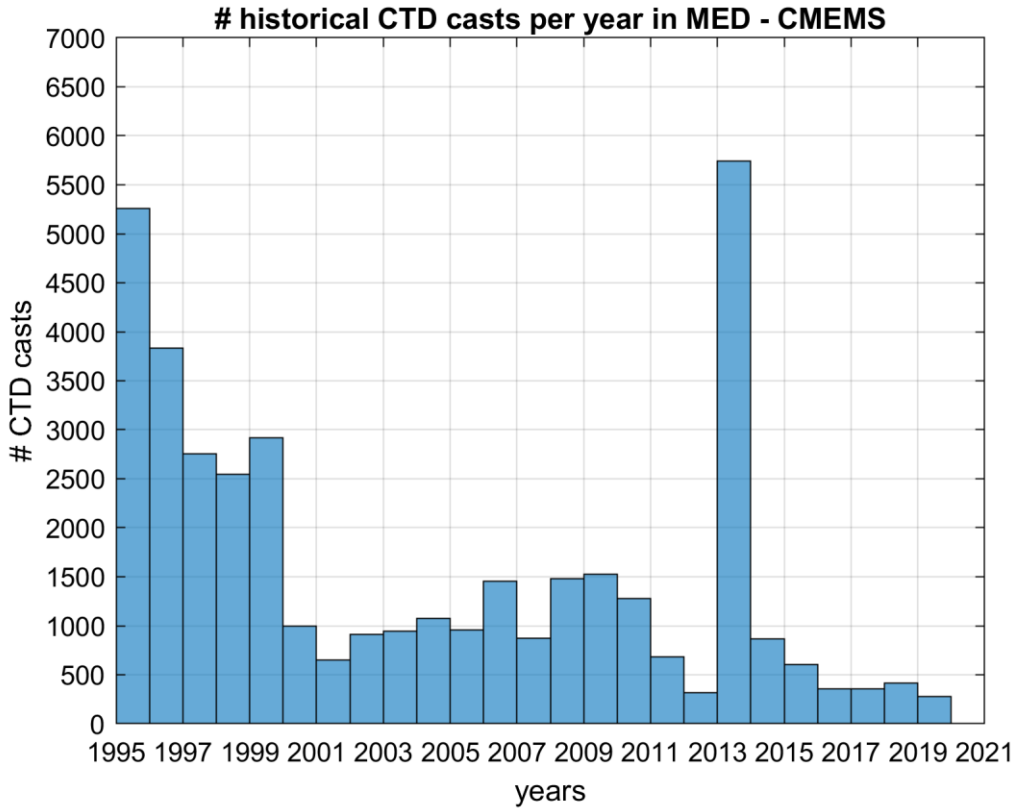

Figure 8. Temporal distribution of the CTD profiles collected through the CMEMS portal and used as a complement of the CTD reference dataset of the Mediterranean Sea.

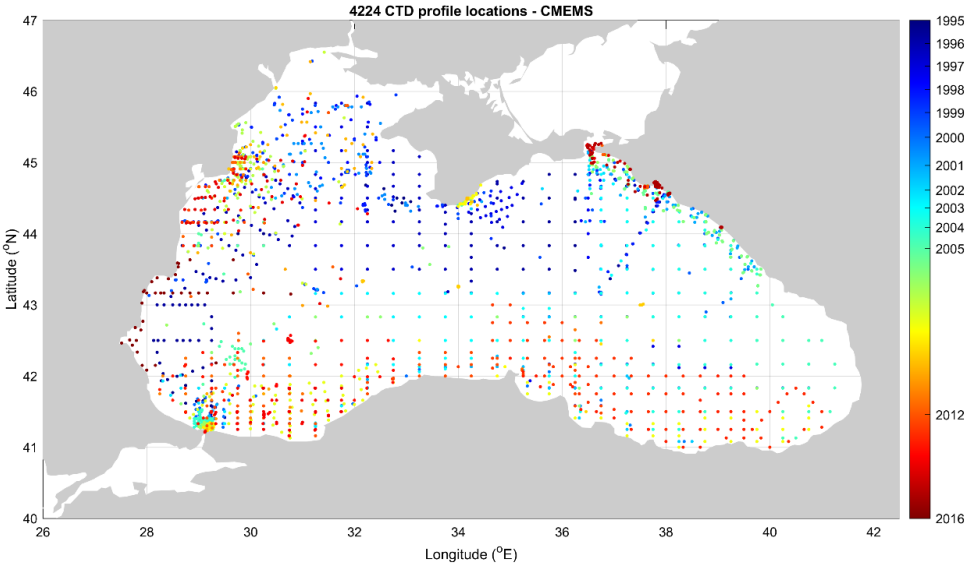

Figure 9. Spatial distribution, color-coded for time, of the CTD profiles collected through the CMEMS portal and used as a complement of the CTD reference dataset of the Black Sea.

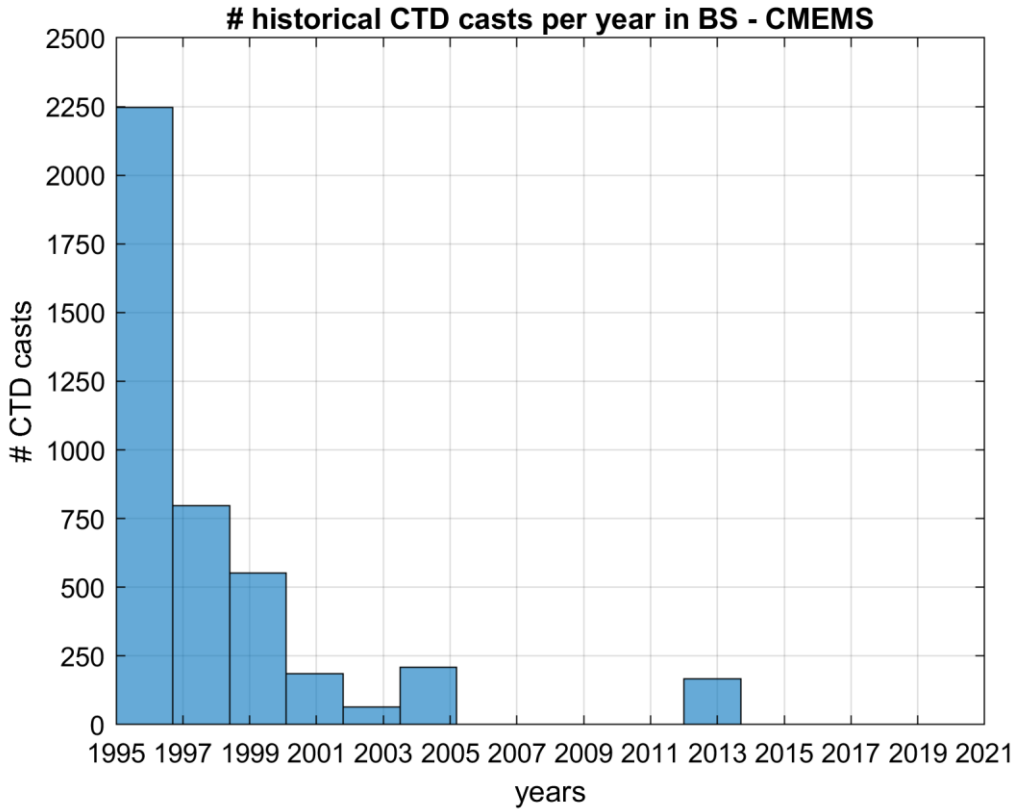

Figure 10. Temporal distribution of the CTD profiles collected through the CMEMS portal and used as a complement of the CTD reference dataset of the Black Sea.

All the available CTD data were taken in account, for example: CTD data from research vessels (accuracy on the order of 0.002° for temperature and 0.003 PSU for salinity after calibration), CTD and XCTD data coming from the high resolution CTD dataset of the World ocean database 2009 (WOD09), CTD from ICES dataset gathered by Danish CIUEM, CTD from SISMER database, coming from French oceanographic campaigns, etc (Skekely, 2020).

3.2. CTD profiles obtained through personal contact

Another CTD data source is CTD profiles acquired by European colleagues from different research institutes during regular cruises or in the framework of projects. A lot of work has been done in finding the right contacts, email exchanging and gathering the data because some local research institutes don't share they data because they are not part of dedicated infrastructures or international projects.

The files were usually received in different formats and hence file-reading MatLab scripts were adapted accordingly. All profiles that sampled deeper than 60 dbar are selected to take into

account profiles in shallow areas. A quality control has been applied to remove any residual outlier and spikes. Many CTD data collected were policy free and hence they will be shared with the Coriolis in-situ Service and integrated in the "official" CTD reference dataset. The data policy is discussed with the owner of the data and as soon as they are declared "not restricted", they become part of the "official" dataset. Figures 11 and 12 shows the spatial and temporal distribution of this CTD dataset.

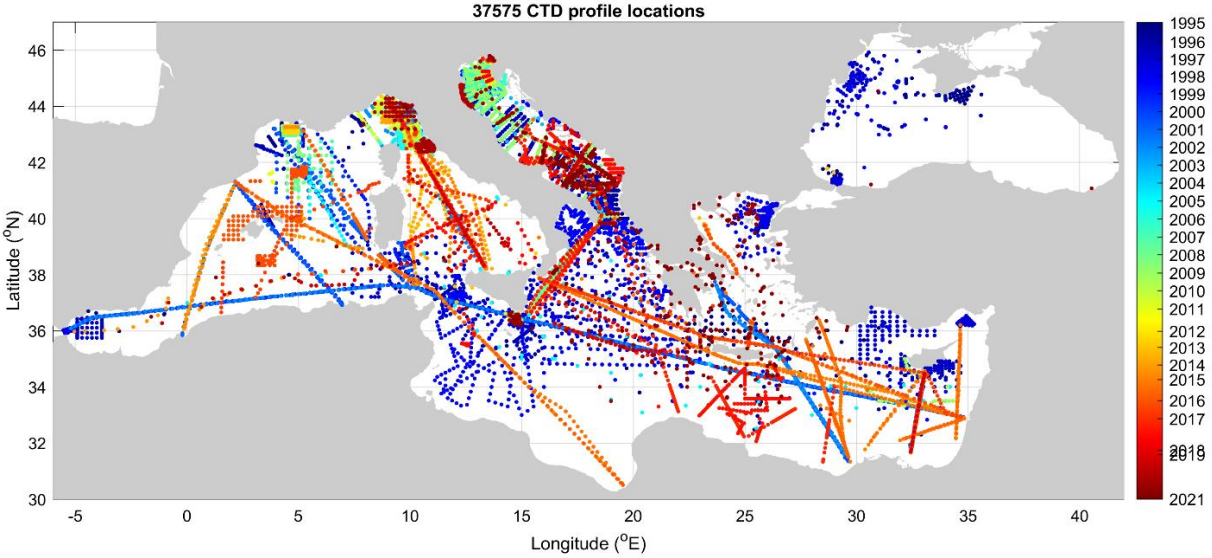

Figure 11. Spatial distribution, color-coded for time, of the CTD profiles collected and used as a complement of the CTD reference dataset.

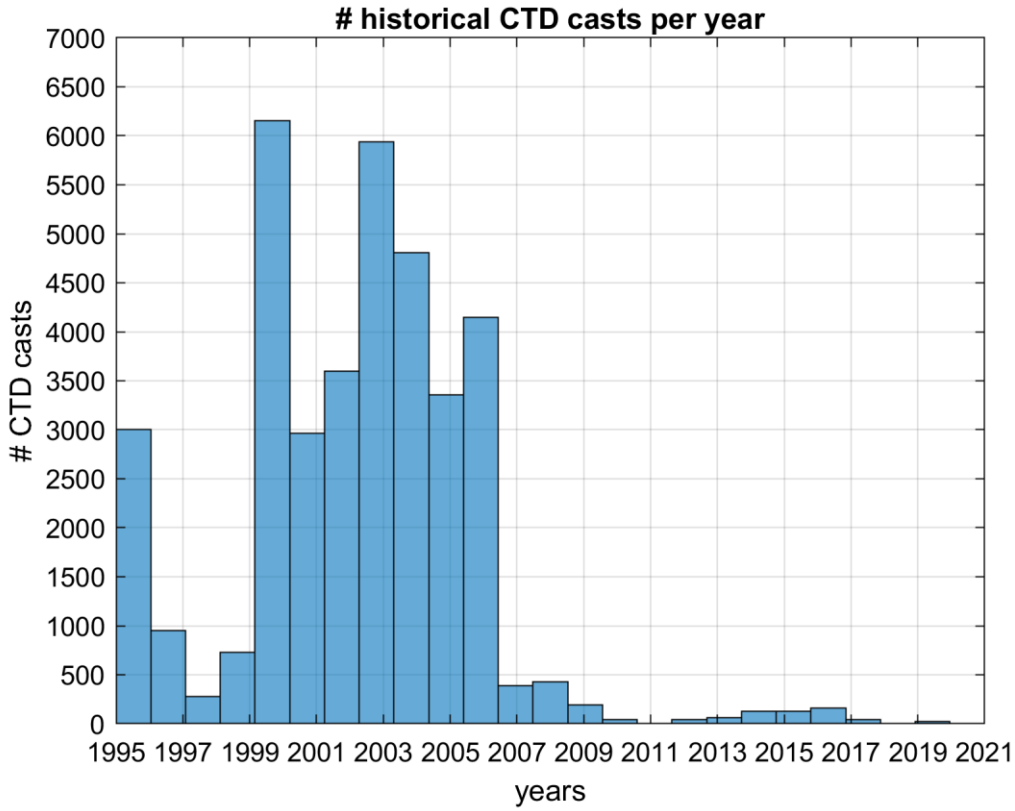

Figure 12. Temporal distribution of the CTD profiles collected and used as a complement of the CTD reference dataset.

4. MERGER OF THE DATASETS

4.1. Checking for duplicates and CTD data merger

All CTD profiles collected are compared to the CTD reference dataset and checked to remove duplicates taking into account thresholds of 10 minutes and 100 meters for time and location respectively. CTD profiles whose difference in time and space is less than the above predefined thresholds are considered duplicates and hence removed from the dataset. Due to the different nature of the existing water masses and to the geography of the Mediterranean, data are then divided in the various sub-basins separately according to the climatological areas of the Mediterranean Sea (figure 13). This final step is done to obtain a more reliable results by OWC delayed-mode procedure (Cabanes et al., 2016) and to avoid selecting historical data for calibration coming from completely different oceanographic regions.

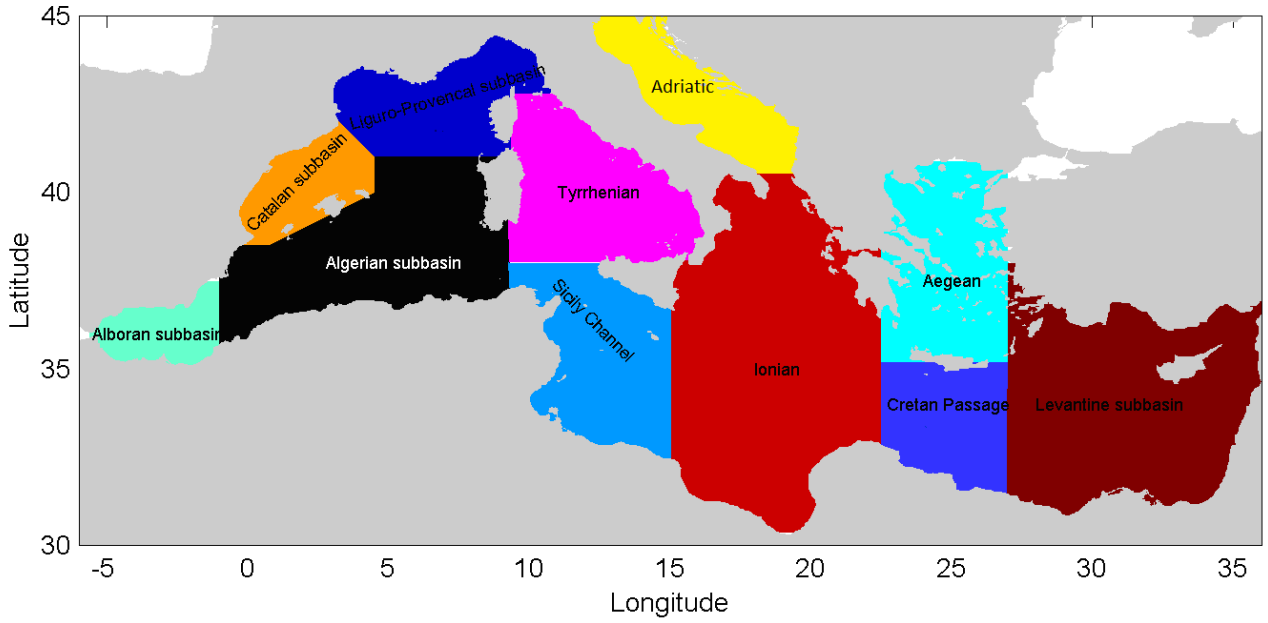

Figure 13. Sub-basins of the Mediterranean Sea, based on the climatological areas defined by the EU/MEDAR-MEDATLAS II project.

4.2. Final CTD reference dataset

The updated reference dataset consists of 67483 CTD profiles, more than about 10000 CTD compared to the previous one. The spatial distribution (figure 14) is greatly improved, and each sub-basin is well sampled, at least spatially. There has been a substantial improvement especially in the Aegean and Levant sub-basins. Data before year 1995 were discarded (figure 15) because considered too old for quality control purposes. In 2020 CTD profiles are scarce due to the pandemic.

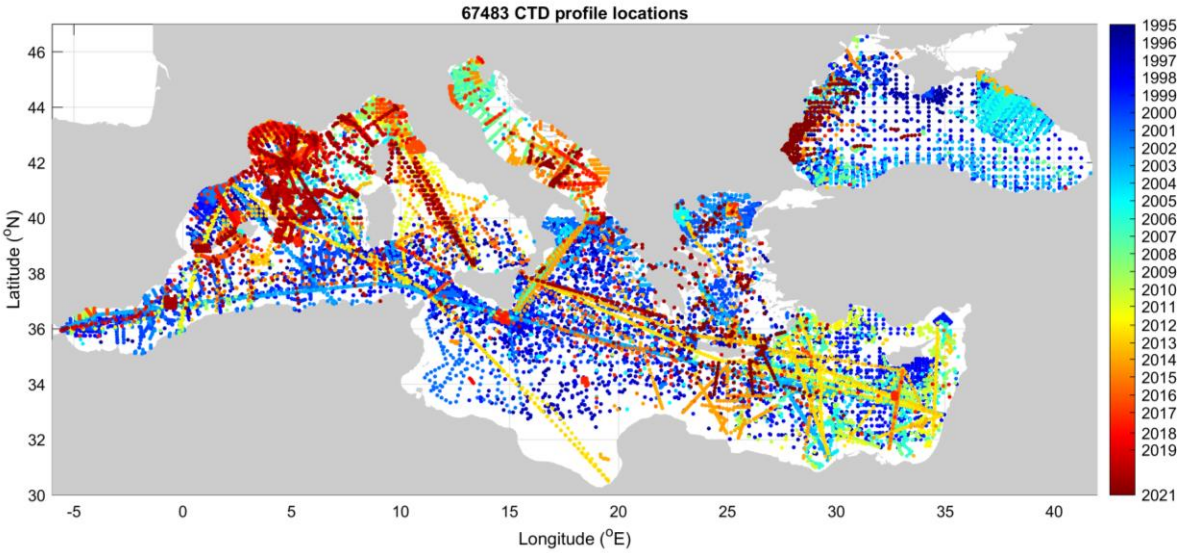

Figure 14. Spatial distribution, color-coded for time, of the CTD profiles in the final version of the CTD reference dataset of the Mediterranean and Black Seas.

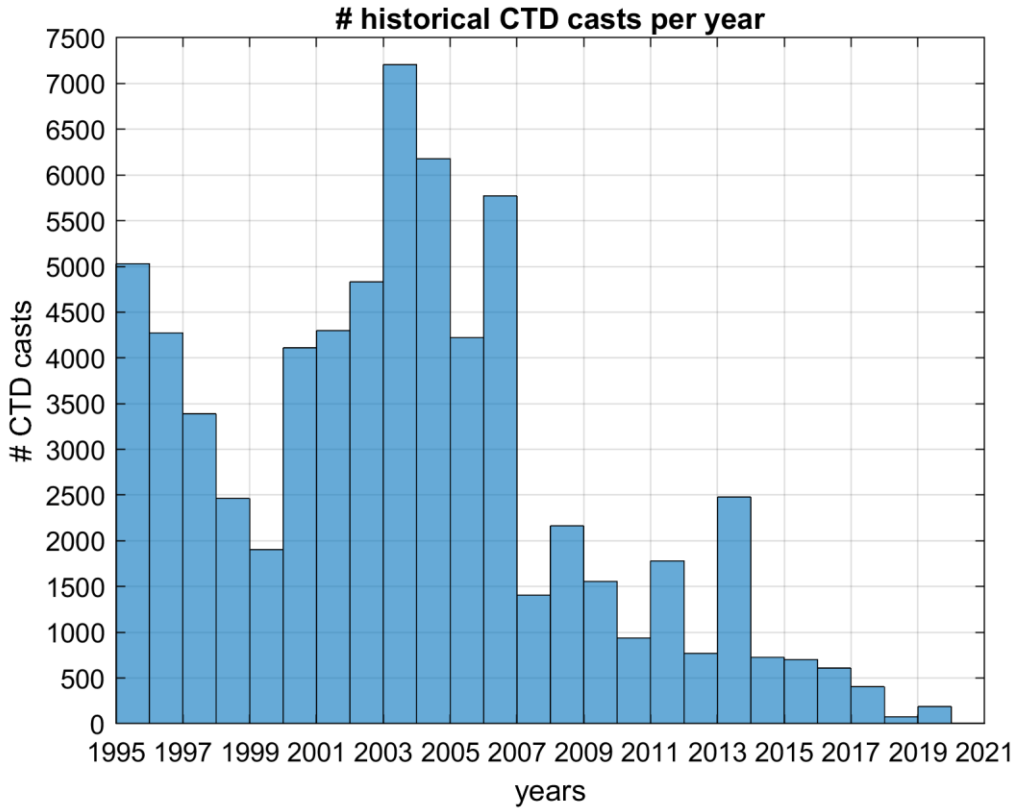

Figure 15. Temporal distribution of the CTD profiles in the final version of the CTD reference dataset of the Mediterranean and Black Seas.

5. CONCLUSIONS

CTD data are used for an advanced quality control of physical data collected from floats deployed mainly by Mediterranean and Black Sea riparian countries. Within the MedArgo program and as part of the H2020 Euro-Argo RISE project, OGS, as responsible of QC activities, has reviewed and improved the availability of high-quality ship-based CTD reference data for QC of Argo float data in the Mediterranean and Black seas. In the marginal seas context, it is important to improve the official CTD reference dataset (provided by the Coriolis GDAC) with other new and updated CTD data to have the best as possible distribution in time and space. A demanding work has been done to collect data from different sources and make them compatible for use in DMQC analysis. Data were collected from several research institutes at regional level and the main European Marine Services. Data were converted in mat format to be used in OWC procedure. A quality control was applied such as an additional visual check to avoid spike or duplication. Data was merged and divided in subsets of WMO boxes according to the climatological areas of the Mediterranean Sea. The reference dataset in the Mediterranean and Black Sea for the DMQC activity has therefore been updated. A substantial improvement has been achieved and gaps in temporal and spatial coverage have been filled in some areas such as the Levantine and the Aegean.
The aggregated data policy will be further discussed in order to make the data available through the Coriolis in situ Service for DMQC activities only.

6. TABLE OF FIGURES

7. REFERENCES

CABANES, C., THIERRY, V., & LAGADEC, C. (2016). IMPROVEMENT OF BIAS DETECTION IN ARGO FLOAT CONDUCTIVITY SENSORS AND ITS APPLICATION IN THE NORTH ATLANTIC. DEEP-SEA RESEARCH PART I: OCEANOGRAPHIC RESEARCH PAPERS, 114, 128–136. *HTTPS://DOI.ORG/10.1016/J.DSR.2016.05.007*

NOTARSTEFANO G. (2019) THE REFERENCE DATASET IN THE MEDITERRANEAN AND BLACK SEAS FOR DMQC ACTIVITY. 2019/15 OCE 3 MAOS 17 PP. *HTTP://MAOS.INOGS.IT/PUB/THE%20REFERENCE%20DATASET%20IN%20MED%20AND%20BS%20FOR%20DMQC%20ACTIVITY.PDF*

SZEKELY T. (2020) FOR GLOBAL DELAYED MODE INSITU DATASET INSITU_GLO_TS_REP_OBSERVATIONS_013_001_B CALLED CORA. 44 PP. *HTTPS://CATALOGUE.MARINE.COPERNICUS.EU/DOCUMENTS/PUM/CMEMS-INS-PUM-013-001-B.PDF*

WONG A., KEELEY R., CARVAL T. AND THE ARGO DATA MANAGEMENT TEAM (2021) ARGO QUALITY CONTROL MANUAL FOR CTD AND TRAJECTORY DATA. *HTTP://DX.DOI.ORG/10.13155/33951*