Proceedings of the Eigth
EuroGOOS International Conference

3-5 October 2017, Bergen, Norway

OPERATIONAL OCEANOGRAPHY

Serving Sustainable Marine Development

EuroGOOS
European Global Ocean Observing System

www.eurogoos.eu

Published by:

EuroGOOS AISBL

231 Avenue Louise

1050 Brussels

Belgium

www.eurogoos.eu

To be quoted as follows:

Operational Oceanography serving Sustainable Marine Development.

Proceedings of the Eight EuroGOOS International Conference.

3-5 October 2017, Bergen, Norway.

E. Buch, V. Fernández, D. Eparkhina, P. Gorringe and G. Nolan (Eds.)

EuroGOOS. Brussels, Belgium. 2018.

D / 2018 / 14.040 / 1

ISBN 978-2-9601883-3-2. 516 pp.

CORA 5.0: GLOBAL IN SITU TEMPERATURE AND SALINITY DATASET

T. Szekely [1], J. Gourrion [1], S. Pouliquen [2] and G. Reverdin [3]

[1] IUEM, UMS3113, CNRS-UBO-IRD, Plouzané, Brest
[2] IFREMER, Plouzané, France
[3] CNRS-INSU, UMR7159, LOCEAN/IPSL (CNRS-UPMC-IRD-MNHN), Paris, France

Abstract

The ability of the scientific community to monitor and understand the oceanic variability is widely based on the quality and the availability of ocean measurements. The particular feature of the CORA dataset (Coriolis Ocean Dataset for Reanalysis) is to distribute all types of in situ temperature and salinity measurements with a maximal sampling, including high frequency profilers (Argo, CTD, etc...) surface and sub-surface timeseries (Thermosalinographs and surface drifters, etc...). The current version of the CORA dataset (CORA5.0) stands out from the previous version by including millions profiles from the historical period (1950-1990) and the addition of year 2015 profiles from Coriolis. A very careful validation process is performed on the CORA measurements since the probably erroneous profiles are individually checked by an oceanographer which changes the data quality flags if necessary. This work flow reduces the amount of unnecessary flags leading to a better estimation of the ocean variability. The CORA dataset is distributed by the Copernicus Marine and Environment Monitoring Service online catalogue: http://marine.copernicus.eu/services-portfolio/access-to-products/

Keywords: global dataset, in situ, temperature and salinity profiles, CMEMS

1. Introduction

A critical field in the ocean studies is to collect, validate and distribute ocean measurements to provide the best working framework for the research community. The CORA dataset is a global delayed-time validated dataset similar to datasets EN4 (Good *et al.*, 2013, www.metoffice.gov.uk) and Word Ocean Database (WOD, Levitus *et al.*, 2013, https://www.nodc.noaa.gov). Despite very similar data sources, differences in the data organisation leads to divergences in the distributed datasets. In this environment the CORA dataset stands out by distributing a dataset closer to the actual measurements, with a full resolution, a careful validation at all levels and a maximized number of profile distributed.

We will first list the data sources of the CORA measurements in section 2. An overview of the dataset structure and a comparison with EN4 and WOD is given in section 3. A description of the CORA data space and time repartition will be given in section 4. The measurements quality control procedure will be described in section 5.
A discussion on the data validation results is given in section 6.

2. Data providers

The CORA 5.0 dataset is an incremental version of the previous CORA datasets, distributed by COPERNICUS Marine Environment Monitoring Service (CMEMS. Most of the CORA profiles are first collected by the Coriolis data centre.

Coriolis data Centre is a Global Data Assembly Centre (DAC) for the Argo program (Roemmich *et al.*, 2009). It collects Argo profiles from the regional Data Assembly Centres (DACs) and distributes them to the community. Coriolis also collects XBTs, CTDs and XCTDs measurements from French and Europeans research programs as well as from the Global Communication System (GTS), Voluntary Ship System (VOS), subtropical moorings measurements (TAO/TRITON/RAMA/PIRATA programs from the PMEL). Major efforts have also been made to include smaller datasets to the Coriolis dataset such as the ITP and CTD profiles from the ICES program, Sea Mammals measurements and Surface drifters. Delayed time mode measurements have also been integrated from the Word Ocean Database (WOD13) and the French Service Hydrographique de la Marine (SHOM), covering the period 1950 to the present.

Finally a comparison of the CORA profile positions with the EN4 dataset (www.metoffice.gov.uk) has shown that some of the profiles distributed in EN4 were not in CORA previous versions. Within CMEMS, a partnership with the EN4 teams was instrumental to identify and to import most of these profiles. Over 5 millions profiles have been imported this way, covering the period 1950-2015. However, for the measurements being directly imported from the EN4 database, the seawater pressure measurement is not provided, a water depth being distributed instead and the profiles maximum vertical levels is set to 400 instead of the full resolution while integrated

from WOD or National Oceanographic Data Centers (NODCs). Those issues will be addressed in the future versions of CORA by recovering the datadirectly from the data provider.

3. Dataset structure

Table I. Description of the datasets validation and distribution charts.

	DISTRIBUTION	COVERAGE	VALIDATION	DATA TYPES	DISTRIBUTED DATA
CORA 5.0	marine.copernicus.eu	1950-2015	- All levels - Semi automatic error detections - Visual control	- Profiles - Timeseries	- All profiles and timeseries - Associated flags
EN4	metoffice.gov.uk	1900-2015	- 400 levels max - Automatic error detections	- Profiles only	- Best profiles - Meta profiles - Data flags
EN4	nodc.noaa.gov	1772-2015	- Standard levels only - Automatic error detections	- Profiles - Timeseries	- All profiles - Standard levels flags

Table 1 gives an overview of WOD13, EN4 and CORA 5.0 dataset structure and distribution. It shows the different ways chosen to validate and distribute the measurements. Despite a data distribution beginning only in 1950, the CORA dataset stands out by providing measurements with a full resolution and validated at all levels. In addition to that, the CORA validation framework is based on the detection of probably erroneous measurements by automatic tests and the visual control of the detected profiles by an oceanographer. Last, it provides not only profiles at the measurement levels but also timeseries of surface and sub-surface measurements such as surface drifters temperature and thermosalinographs providing worldwide high-frequency temperature and salinity measurements.

4. Dataset description

The different data types in CORA vary widely in time (Fig. 1). Most of the profiles measured before 1965 are mechanical bathythermographs (MBT) measurements or bottle-sampled measurements. Between 1970 and 1990, the most common instrument is the expendable bathythermographs (XBT), developed during the 60s and widely used by the community. Most of the XBT profilers deployed during this period are T4 type sensor, measuring temperature above 460 meter depth. The development of the Sippican T-7 instrument with a maximum depth of 1000m sowly increases the measurement number below the sub-surface during the 1980s (see Fig. 2 for the dataset measurements distribution with depth). The conductivity temperature

depth sensor (CTD), a popular instrument capable of measuring both temperature and conductivity leading to an accurate estimation of sea salinity and sea water density was developed in the early 1960. The amount of CTD profiles in the CORA dataset slightly increases then and reaches a plateau of about 20000 yearly profiles in the early 1990s.

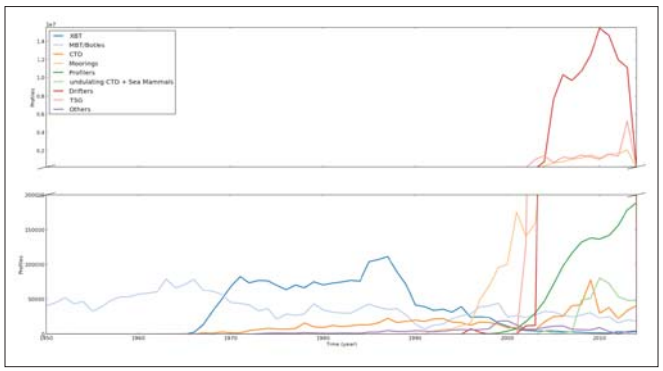

Fig.1. Yearly number of distributed profiles, sorted by instruments types.

The measurements provided are however deeper than the previous decade, leading to a better coverage below 500m depth (Fig. 2).

The number of profiles then exponentially increases since the development of the TAO/RAMA/PIRATA equatorial mooring program in the mid-late 1990s. The ocean sampling rate exploded in the early 2000 thanks to the development of autonomous profilers and the worldwide Argo program.

See the simultaneous increase in the number of measurements below 1000m depth.

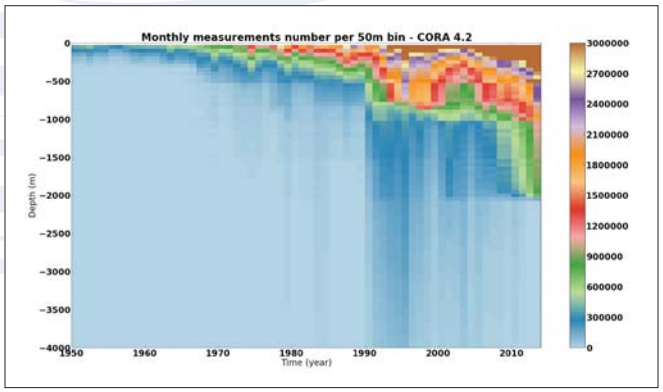

Fig. 2. Yearly number of distributed measurements per 20 meter depth bin.

In the Antarctic Ocean, the sampling rate increases in 2005-2006 as a consequence of Argo deployment south of 55°S and the deployment of sea mammals mounted CTD in the vicinity of the Kerguelen islands (Roquet *et al.*, 2011).

It must also be emphasised that a fraction of the number of profiles sharp increase of the early 2000s is a consequence of the development of high frequency measurements devices such as the ocean drifters, the thermo salinographs (TSGs) or towed undulating CTD (gliders, scanfish, seasoar...). Each towed undulating CTD profile and each independent TSG or drifter measurement is treated as an independent profile in the CORA dataset. This lead to a more homogeneous dataset file structure but it can be misleading when estimating the number of profiles. This is also true for the mooring measurements since some of them are high frequency devices. A consequence is also the increase of measurements originating from mooring data at 250m depth and 500m depth and from TSG and drifting buoys at the surface.

5. Data quality control

The Coriolis data centre checks the data quality and consistency, in order to provide to the scientific community a consistent global dataset of validated measurements.

A description of the near-real time dataset validation in the Coriolis data centre is given by Cabanes *et al.*, 2013.

First, a set of near-real time validation tests is performed within a few days after measurements reception. An additional control of the Argo profiles is performed thanks to an objective analysis on a daily basis (Gaillard *et al.*, 2009). The detected profiles are then visually checked by a PI to flag the erroneous measurements. A quality control based on altimetry comparisons is also performed on a quarterly basis on Argo data to improve the real time validated dataset (Guinehut *et al.*, 2009). A PI investigation is also performed on the altimetry detected profiles.

The corresponding dataset is distributed by the CMEMS catalogue and regularly updated.

The CORA dataset validation is a delayed time validation performed each year with sharper tests than the real time mode validation. A PI visual investigation is performed at each step to reduce the number of erroneous flags.

The first quality check considers the detection the well-known instrumental errors: spikes, constant value, absurd value, etc...The following step of the CORA data validation is designed to detect the profiles diverging from the known ocean variability. Each temperature and salinity profile is compared to a minimum and maximum reference field. The field is a gridded mesh of 1 degree resolution horizontal hexagonal cells of 20m depth. The higher and the lower temperature (resp. Salinity) ever measured by

Argo floats on the period 2002-2013 in a given cell is compared to the profile sampled in this cell. The profilse containing measurements exceeding the reference values are checked by an oceanographer.

Fig. 3 shows the position of the profiles flagged during this process.

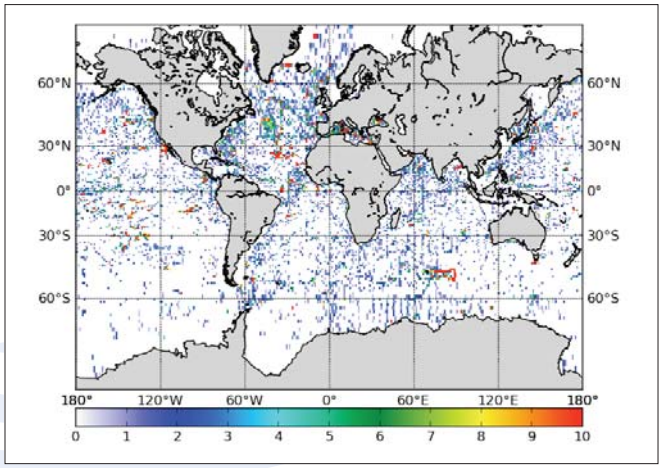

Fig. 3. Number of profiles flagged during the delayed time mode quality control process gridded in a 1° per 10 grid cells.

6. Discussion on the data validation

A robust rather simple way to confirm the global ocean data flag is to check the efficiency of the dataset to estimate the ocean variability.

Two ocean average 0-10m depth temperature estimations are calculated: a raw estimation which takes into account every measurement without considering the ocean flags and a flagged estimation which only consider the good and probably good quality control flags.

The global ocean is divided in 1° per 1° grid cells with 10m depth from surface to 700m depth. For each cell, the mean temperature measured by each profile located in the cell is taken into account. An interpolated field is calculated following the method presented by Forget and Wunch 2007.

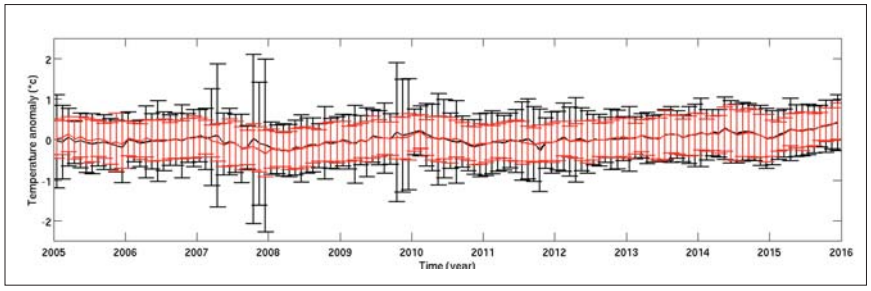

Fig. 4. Mean temperature anomaly between 60°N and 60°S estimated with the CORA dataset.

Fig. 4 gives the mean temperature anomaly in the top 10 meters of the ocean. It shows that the CORA delayed time mode quality control flags allow to reduce significantly the ocean temperature anomaly error bar to an almost constant level with a low impact on the anomaly signal. As a consequence, one can consider that the CORA validation process is consistent within time while limiting the over-flagging to a low level.

References

Cabanes,C., grouazel, A., Von Schuckmann, K., Hamon, M., Turpin, V., Coatanoan, C., Paris, F., Guinehut, S., Boon, C., Ferry, N., De Boyer Montegut, C., Carval, T., Reverdin, G., Pouliquen, S., le Traon, P.-Y. (2013). The CORA dataset: validation and diagnostics of in situ ocean temperature and salinity measurements. *Ocan Science*, 9,1-18.

Forget, G., Wunch, C. (2006). Estimated global hydrographyc variability. *Journal of plysical oceanography*, 37,1997-2008.

Gaillard, F., Autret, E., Thierry, V. Galaup, P., Coatanoan, C. Loubrieu, T. (2009). Quality control of large Argo datasets, *Journal of Atmospheric and oceanic Technology*, 26, 337-351.

Guinehut, S., Coatanoan, C., Dhomps A.-L., Le Traon, P.-Y., Larnicol, G. (2009). On the use of satellite altimeter data in Argo quality control. *Journal of Atmospheric and oceanic Technology*, 26, 395-402.

Good, S. A., M. J. Martin and N. A. Rayner, (2013). EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, *Journal of Geophysical Research: Oceans*, 118, 6704-6716

Levitus S., Antonov J., Baranova, O., Boyer T., Coleman C., Garcia H., Grodsky A. , Johnson D., Locarnini R., Mishonov A., Reagan J., Sazama C., Seidov D., Smolyar I., Yarosh E., Zweng M. (2013) The World Ocean Database, *Data Science Journal*, 12, WDS229-WDS234.

Roemmich, D., Johnson, G., Riser, S., Davis, R. Gilson, J., Owens, W., Garzoli, S., Schmid, C., Ignaszewski, M. (2009). The Argo program: Observing the global ocean with profiling floats. *Oceanography*, 22, 34-43.

Roquet, F., Charrassin, J.-B., Marchand, S., Boehme, L., Fedak, M., Reverdin, G., Guinet, C. (2011). Delayed-mode calibration of hydrographic data obtained from animal-borne satellite relay data loggers. *Journal of Atmospheric and oceanic Technology*, 28, 787-801.