QUALITY INFORMATION DOCUMENT

For Global Delayed Mode Insitu Dataset

INSITU_GLO_TS_REP_OBSERVATIONS_013_00 1_b
called CORA

Issue: 1.6

Contributors: T. Szekely

Approval Date by Quality Assurance Review Group: 23 January 2016
## CHANGE RECORD

<table>
<thead>
<tr>
<th>Issue</th>
<th>Date</th>
<th>§</th>
<th>Description of Change</th>
<th>Author</th>
<th>Checked By</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>20/06/2013</td>
<td>all</td>
<td>First version of document</td>
<td>Antoine GROUAZEL</td>
<td>S Pouliquen</td>
</tr>
<tr>
<td>1.1</td>
<td>20/12/2013</td>
<td></td>
<td>Update for V4.0 or CORA-GLOBAL</td>
<td>Antoine GROUAZEL</td>
<td>S Pouliquen</td>
</tr>
<tr>
<td>1.2</td>
<td>14/02/2014</td>
<td></td>
<td>Update after QUARG review</td>
<td>S Pouliquen</td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>01/02/2014</td>
<td></td>
<td>Update for CORA-GLOBAL V4.1</td>
<td>T. Szekely</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>10/02/2015</td>
<td>All</td>
<td>Revision after V5 acceptance</td>
<td>T. Szekely</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>May 1 2015</td>
<td>all</td>
<td>Change format to fit CMEMS graphical rules</td>
<td>T. Szekely</td>
<td>L. Crosnier</td>
</tr>
<tr>
<td>1.6</td>
<td>19/01/2016</td>
<td></td>
<td>Update for CORA-GLOBAL V4.2</td>
<td>T. Szekely</td>
<td></td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

I  Executive summary ........................................................................................................................................ 4
   I.1 Products covered by this document ............................................................................................................ 4
   I.2 Summary of the results .............................................................................................................................. 4
   I.3 Estimated Accuracy Numbers .................................................................................................................. 5

II  Production Subsystem description .................................................................................................................. 6
   II.1 Where do CORA measurements come from? .......................................................................................... 6
   II.2 What kind of data does CORA contains? ................................................................................................. 8
   II.3 What is the difference with MyOcean INS TAC insitu distribution? ....................................................... 13

III  Validation framework .................................................................................................................................. 15
   III.1 Detection and correction of repeated problems ...................................................................................... 16
   III.2 Detection and correction of punctual problems .................................................................................... 17
   III.3 Chronology of the tests .......................................................................................................................... 17
      III.3.1 Before extraction of CORA data in NetCDF format ........................................................................ 17
      III.3.2 After extraction of CORA data in netcdf format ............................................................................ 17
   III.4 Content of the second pass of generic tests .......................................................................................... 19

IV  Validation results .......................................................................................................................................... 23
   IV.1 Duplicates profiles suppression results .................................................................................................. 23
      IV.1.1 Results for CORA-GLOBAL-04.2 .................................................................................................... 23
   IV.2 XBT correction results ........................................................................................................................... 23
      IV.2.1 Summary of the method .................................................................................................................. 24
      IV.2.2 Results for CORA-GLOBAL-4.2 ..................................................................................................... 27
   IV.3 Objective analysis .................................................................................................................................. 28
      IV.3.1 Results for CORA-GLOBAL-4.2 ..................................................................................................... 28

V  Validation synthesis ........................................................................................................................................ 30

VI  Bibliography .................................................................................................................................................. 31
I EXECUTIVE SUMMARY

I.1 Products covered by this document

This document aims to give a detailed picture of the processes and tools used to validate the datasets CORA INSITU_GLO_TS_REP_OBSERVATIONS_013_001_b.

<table>
<thead>
<tr>
<th>Short Description</th>
<th>Product code</th>
<th>Area</th>
<th>Delivery Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global REP</td>
<td>INSITU_GLO_TS_REP_OBSERVATIONS_013_001_b</td>
<td>Global</td>
<td>Yearly</td>
</tr>
</tbody>
</table>

CORA INSITU_GLO_TS_REP_OBSERVATIONS_013_001_b, also called “Global Ocean-Full CORA-Insitu Observations Yearly Delivery in Delayed Mode”, is a global dataset of in situ temperature and salinity measurements. Latest version (called CORA-GLOBAL-04.2) of the product covers the period 1950-2014. The main new features of CORA-GLOBAL-04.2 is the incorporation of profiles and time series measured during 2014 and the addition of an historical dataset collected by the Service hydrographique de la Marine (SHOM). This dataset covers the period 1950-2009 and involves more than 2.4 million profiles distributed around the global ocean. The profiles extracted from the EN.4 dataset have also been updated for the period 1950-1989.

The CORA-GLOBAL-04.2 measurements comes from many different sources collected by Coriolis data centre. It also contains Argo profiler measurements, CTDs, XBTS casted by opportunity vessels or research ships, sea mammal’s data, gliders data, moorings data and drifter data.

CORA INSITU_GLO_TS_REP_OBSERVATIONS_013_001_b is a yearly updated product. It means that each year the dataset is not built from scratch, we update existing data, we add new data and we delete removed data from the GLOBAL INS TAC.

I.2 Summary of the results

The accuracy of the in situ observation depends on the platforms and sensors that have been used to acquire them (see next §I.3). The observations are aggregated by the In Situ Thematic center and include the input from the SeaDataNet II data set. The data set is provided to users together with metadata information on the platforms that were used to perform the observations. The quality of the observation is tested using automatic procedures and comparison to climatology. Quality flags are set to inform the users of about the level of confidence attached of each observation.

The In Situ TAC relies on observing systems maintained by institutes that are not part of the In Situ TAC and MyOcean project is not contributing to the maintenance and setting up of the observing systems it uses.

- The variety of platforms available to monitor the status of the ocean is very diverse within the different regions.

- In some regions the number of available platforms is on a critical low level to provide an adequate representative overall view of the state of the ocean.
• Some Data are obtained by regular vessel cruises or dedicated scientific expeditions. The availability of data from these scientific expeditions is often delayed so they are not available for the RT data stream. Consequently, data are not available for assimilation of the operational models.

• The percentage of data flagged as ‘good data’ is varying from region to region.

In the 90’s low numbers of available observations impacts the provision of a good and representative real time data product is detected within the temporal frame of the project.

I.3 Estimated Accuracy Numbers

Table 1 summarizes the accuracy of the T/S measurements that can be expected depending of the platforms and sensors. The definition of the reference values is obtained from different sources. The platform specific references that differ from the common ones are given for the specific value.

<table>
<thead>
<tr>
<th>Data-type</th>
<th>Temperature $^1$ [°C]</th>
<th>Salinity $^1$ [PSU]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTD</td>
<td>0.005-0.001</td>
<td>0.02-0.003</td>
</tr>
<tr>
<td>XBT</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>XCTD</td>
<td>0.02</td>
<td>0.003</td>
</tr>
<tr>
<td>PFL (profiling floats)</td>
<td>0.01$^2$</td>
<td>0.01$^2$</td>
</tr>
<tr>
<td>Moored buoy data:</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>Drifting buoy data</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Marine mammals</td>
<td>0.005</td>
<td>0.01</td>
</tr>
<tr>
<td>Glider</td>
<td>0.005</td>
<td>0.02</td>
</tr>
<tr>
<td>Underway (Ferrybox, Research vessel TSG)</td>
<td>0.001-0.1$^3$</td>
<td>0.003-0.2$^3$</td>
</tr>
</tbody>
</table>

*Table 1 Accuracy numbers for temperature and salinity observations for the different platforms data is obtained from in the INS TAC $^1$, NOAA, 2009$^2$ ARGO Buoys $^3$, depending on sensor type*
II PRODUCTION SUBSYSTEM DESCRIPTION

II.1 Where do CORA measurements come from?

As Global DU for the GLOBAL INSITU TAC, the Coriolis data centre aggregates all the data circulating within the GLOBAL INSITU structure. The main sources of data for the CORA dataset

GLOBAL INSITU TAC gathers and archives data coming from many different networks of physical oceanic observations. It collects data from real-time flows such as: Argo network (as a Data Assembly Centre), thermostalinograph data from research and opportunity ships (also as global DAC on GOSUD network), GTSPP flow, French research ships, MEDS-Canada-data, voluntary observing and merchant ships, moorings (TAO-TRITON-PIRATA-RAMA networks plus coastal moorings). GLOBAL INSITU TAC also punctually integrates data in delayed mode coming from various sources, such as World Ocean Database (for CTD only) sea mammals, research cruise from the ICES database, data from SeaDataNet project and last but not least the ROOses data.

In the latest version of the CORA product, over 2.4 million measurements from the SHOM have been collected by the GLOBAL INSITU TAC and integrated into CORA.

CORA is an extraction of the Coriolis database at a given date each year. It is a set of NetCDF files in Argo format. Data in CORA are ordered by date and type of data: nomenclature of the files is CO_code_YYYYMMDD_PR_TT.nc where:

- **code** is the name of the analysis performed: DMQCGL01
- **YYYYMMDD** is the date of the data,
- **PR** stands for vertical Profile, **TS** stands for TimeSeries
- **TT** is the type of file (data)

![Figure 1: Yearly number of files associated to each datatypes available in CORA4.2](image)
The yearly number of file per types collected by CORA is given on figure 1. The file type gives an indication on the file profiles instrument types and provider.

File types content:

- **PR_PF files**: data from Argo floats directly received from the Argo DACS (real Time and delayed mode if available). These data have a nominal accuracy of 0.01° and 0.01 PSU and are transmitted with full resolution.

- **PR_XB files**: XBT or XCTD data received from research and opportunity vessels have accuracy within 0.03° to 0.1° for temperature and 0.03 to 0.1 PSU for salinity.

- **PR_CT files**: contains CTD data from research vessels (accuracy on the order of 0.002° for temperature and 0.003 PSU for salinity after calibration) but also data from sea mammals equipped with CTD (accuracy is on the order of 0.01° for temperature and 0.02 PSU for salinity but can be lower depending of the availability of reference data for post-processing, (see Boehme et al, 2009) and received from MNHN and some sea Gliders.

- **PR_OC files**: others CTD and XCTD data coming from the high resolution CTD dataset of the World ocean database 2009 (WOD09).

- **PR_MO files**: Mooring data with an accuracy generally comparable to Argo floats (except for S near surface). Warning this data type is under migration within the Coriolis database. It means that this data type will probably disappear in the next release since moorings data are now classified in TS_MO.

- **PR_TE and BA files**: The two last categories are for all the data transmitted through the GTS (data from Argo floats not yet received at the DACS, mooring, XBT,...). This transmission system imposes limitation on the accuracy: data is truncated two and one places beyond decimal point for TE and BA type respectively.

- **PR_HF files**: High Frequency data coming from coastal moored buoys (max 3 levels of immersion, max depth<200m, with less than 0.01° of variation in position and more than one measure per day. HF files are no more validated since CORA3.4.

- **PR_IC files**: CTD from ICES dataset gathered by Danish CIUEM. Those profiles complete the CTDs coverage on the period 1990-2011.

- **PR_ME files**: CTD from SISMER database, coming from French oceanographic campaigns. First loading in GLOBAL INS TAC in July 2013 (about 60.000 CTDs)

- **PR_SD files**: CTD from SeaDataNet project. More specifically it represents about 10.000 CTD coming from the first exchange between SeaDataNet partners and MyOcean TAC dated back to end of 2012.

---

1 Principal Investigators have to provide a delayed mode validation on their floats within 6months after the date of observation
II.2 What kind of data does CORA contains?

CORA thus contains data from different types of instruments: mainly Argo floats, XBT, CTD, XCTD, and moorings. Refer to the Product User Manual to have details about the content of the CORA dataset.

The figure 2 gives the yearly number of collected profiles associated to each datatype. Most of the data acquired during the historical period (before 1990) measurements are MBT profiles. The XBTs and CTD have been developed during the mid 60s. The XBTs have been widely deployed by military and scientific programs since the early 70s. The CTDs deployment rate however increases slowly since the higher cost
of CTD probes deployment. The number of collected profiles decreases in the early 1990 with the lower contribution of SHOM dataset and EN.4 dataset to CORA-GLOBAL-04.2. The development of worldwide moorings programs (TAO/RAMA/PIRATA) in the mid 1990s is associated to a sharp increase of the moorings profiles number. Finally, the number of profiles increases in the early 2000s thanks to the ARGO program.

The ocean sampling rates given on figure 3 are calculated by counting the monthly sampling rate of 3°N per 5°W gridded ocean basin. It shows that most of the North Atlantic Ocean is well sampled in the early 70s. However the other basins sampling rates are lower. It increases slowly from 1950 to 1990. The ocean sampling rate decreases in the north Atlantic basin and in the north Pacific basin during the 90s. This decrease is however compensated by a better sampling at depth (see fig. 4 and 5). In the early 2000s, the development of the ARGO program provides a quick increase of the ocean sampling rate at a global scale. The ocean sampling scale reaches a top over 40% in the Antarctic basin and over 80% for the others basins.

---

Figure 3: Time variation of ocean basins and global ocean (red) sampling rate.

---

Fig: 3 and 4 gives an overview of the depth repartition of CORA measurements among time. During the 1950-1965 period, most of the observations are MBTs temperature measurements with a few bottle salinity measurements. The increasing number of T4 XBT (250 m depth XBT) after 1965 gives a better ocean coverage at low depth for temperature only. The T7 XBTs (1000m depth XBT) development slowly gives a better sampling at depth. During the same time, the use of CTD sensors provides good temperature and salinity measurements at lower depth. The number of CTD profiles being however lower. The ocean sampling at depth is gradually improved during after the early 1990s with the WOCE program (early 90s), the TAO/RAMA/PIRATA program (late 90s and the ARGO program (mid 2000s).

Figure 1: Time depth diagram of decimal logarithm of the number of good quality (QC = 0, 1 or 2) temperature measurements per month and per level

Figure 3: Time depth diagram of decimal logarithm of the number of good quality (QC = 0, 1 or 2) salinity measurements per month and per level

Figure 5: Spatial repartition of CORA profiles during June 1990. Top, left: SHOM profiles (purple), CTD measurements (red), XBT measurements (green), TESAC measurements from GTS (blue), OceanSites measurements (black). Top right: ferry boat measurements (red), TSG from research ship (green), drifters buoy timeseries (blue), SVP buoy timeseries (black). Bot. left: mooring measurements (red squares), mooring timeseries (red circles), high frequency coastal moorings (red crosses). Bot. Right: profilers measurements (red), gliders measurements (green), sea mammals measurements (blue).
Figure 6: Spatial repartition of CORA profiles during June 1995. Top, left: SHOM profiles (purple), CTD measurements (red), XBT measurements (green), TESAC measurements from GTS (blue), OceanSites measurements (black). Top right, ferry boat measurements (red), TSG from research ship (green), drifters buoy timeseries (blue). SVP buoy timeseries (black). Bot. left: mooring measurements (red squares), mooring timeseries (red circles), high frequency coastal moorings (red crosses). Bot. Right: profilers measurements (red), gliders measurements (green), sea mammals measurements (blue).

Figure 7: Spatial repartition of CORA profiles during June 2000. Top, left: SHOM profiles (purple), CTD measurements (red), XBT measurements (green), TESAC measurements from GTS (blue), OceanSites measurements (black). Top right, ferry boat measurements (red), TSG from research ship (green), drifters buoy timeseries (blue). SVP buoy timeseries (black). Bot. left: mooring measurements (red squares), mooring timeseries (red circles), high frequency coastal moorings (red crosses). Bot. Right: profilers measurements (red), gliders measurements (green), sea mammals measurements (blue).
Figure 8: Spatial repartition of CORA profiles during June 2005. Top, left: SHOM profiles (purple), CTD measurements (red), XBT measurements (green), TESAC measurements from GTS (blue), OceanSites measurements (black). Top right, ferry boat measurements (red), TSG from research ship (green), drifters buoy timeseries (blue), SVP buoy timeseries (black). Bottom left: mooring measurements (red squares), mooring timeseries (red circles), high frequency coastal moorings (red crosses). Bottom Right: profilers measurements (red), gliders measurements (green), sea mammals measurements (blue).

Figure 9: Spatial repartition of CORA profiles during June 2010. Top, left: SHOM profiles (purple), CTD measurements (red), XBT measurements (green), TESAC measurements from GTS (blue), OceanSites measurements (black). Top right, ferry boat measurements (red), TSG from research ship (green), drifters buoy timeseries (blue), SVP buoy timeseries (black). Bottom left: mooring measurements (red squares), mooring timeseries (red circles), high frequency coastal moorings (red crosses). Bottom Right: profilers measurements (red), gliders measurements (green), sea mammals measurements (blue).
II.3 What is the difference with MyOcean INS TAC insitu distribution?

The product called **INSITU_GLO_NRT_OBSERVATIONS_013_030** comes from the GLOBAL INS MyOcean Distribution Unit (monthly, latest or historical repositories). It is also a NetCDF distribution of the Coriolis database but with two main differences with respect to CORA dataset: the update of MyOcean distribution is continuous in the time and the content of MyOcean/monthly Distribution is an image of GLOBAL INSITU TAC database contrarily to CORA which is a frozen image of this database corresponding to a given date. And the second main difference is that MyOcean distribution is organized by platforms (each file corresponds to a platform). Whereas CORA dataset is designed to be assimilated by models thus it is organized into daily files.
Figure 11: Data source and validation process for CORA GLOBAL
III VALIDATION FRAMEWORK

In this paragraph we will describe the systems that enable the assessment of CORA dataset on observed levels. The procedure evolves each year therefore this document is focused on the current procedure used for the construction of CORA-GLOBAL-04.2.

The production system that handles and manages the validation of CORA dataset involves both the Coriolis Data centre and “Coriolis Research and development team”.

The validation chain of CORA begins within the Coriolis database with real time automatic checks. Those tests control that meta-data are consistent with archiving, then objectives analyses are ran on the data for real-time and near-real-time controls.

After extraction from the Coriolis database (i.e GLOBAL TAC myocean) data are re analysed through a delayed mode assessment process especially dedicated to CORA dataset. It contains a series of 14 generic tests plus extra specific corrections: duplicates suppression and XBT correction. And at the end of the process we perform a comparison of the data with an ultimate objective analysis that guarantees the consistency between each measurements in time and space.

Corrections on measures are avoided as much as possible in Copernicus INSITU TAC, and when performed they are provided aside to the raw measurements to let the users choose the level of processing they want to use. Results of validation are realized in editing the Quality flags associated to the measures (see Table 2).

<table>
<thead>
<tr>
<th>Quality Code</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No QC was performed</td>
</tr>
<tr>
<td>1</td>
<td>Good data</td>
</tr>
<tr>
<td>2</td>
<td>Probably good data</td>
</tr>
<tr>
<td>3</td>
<td>Bad data that are potentially correctable</td>
</tr>
<tr>
<td>4</td>
<td>Bad data</td>
</tr>
<tr>
<td>5</td>
<td>Value changed</td>
</tr>
<tr>
<td>6</td>
<td>Not used</td>
</tr>
<tr>
<td>7</td>
<td>Not used</td>
</tr>
<tr>
<td>8</td>
<td>Interpolated value</td>
</tr>
<tr>
<td>9</td>
<td>Missing value</td>
</tr>
</tbody>
</table>
III.1 Detection and correction of repeated problems

What we call “repeated problems” are issues that occur very frequently on the data and where there is no doubt on the truthfulness of the alert raised. For example, it is frequent to have some levels that appear twice in a profile of salinity. When a problem is detected on a measure or on meta data in CORA a log file is produced. This file is sent to the GLOBAL INSITU TAC in order to validate the alert and correct the data at the source.

Those “repeated problems” are numerous and the alert itself does not need any confirmation thus the correction applied is automatic. In CORA dataset, detection and correction are done in delayed mode directly into CORA NetCDF files and then those files are resubmitted to the GLOBAL TAC in order to replace former measures.

Table 2 : Quality code meaning

<table>
<thead>
<tr>
<th>Quality Code</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No QC was performed</td>
</tr>
<tr>
<td>1</td>
<td>Good data</td>
</tr>
<tr>
<td>2</td>
<td>Probably good data</td>
</tr>
<tr>
<td>3</td>
<td>Bad data that are potentially correctable</td>
</tr>
<tr>
<td>4</td>
<td>Bad data</td>
</tr>
<tr>
<td>5</td>
<td>Value changed</td>
</tr>
<tr>
<td>6</td>
<td>Not used</td>
</tr>
<tr>
<td>7</td>
<td>Not used</td>
</tr>
<tr>
<td>8</td>
<td>Interpolated value</td>
</tr>
<tr>
<td>9</td>
<td>Missing value</td>
</tr>
</tbody>
</table>

© EU Copernicus Marine Service – Public
III.2 Detection and correction of punctual problems

On the contrary some problems are rare and needs that an operator confirms the truth of the alert and change carefully the values of quality flags. For example some spikes in the data are rather hard to highlight. In such a case the detection ran on a measure or on metadata in CORA permits to create a log file containing of the suspicious measures. Fortunately the number of “punctual problems” is much inferior to “repeated problems” and operators from the TAC can visualize each of them to confirm/dis-confirm the alert and then correct or let the quality flags as it.

III.3 Chronology of the tests

III.3.1 Before extraction of CORA data in NetCDF format

The measurements extracted from the Coriolis database are subject to a first set of quality checks.

1. Automatic checks at data loading in GLOBAL INSITU TAC database
2. Comparison with Objective analysis on 21 days frame in real time
3. Comparison with Near Real Time objective analysis once a month on the last 30 days
4. First validation run on MyOcean GLOBAL distribution Unit (monthly repository) with feedback to the GLOBAL INSITU TAC

The tests listed above are exhaustively described in Coatanoan and Petit de la Villeon, 2005 and discussed by Cabanes, 2013.

III.3.2 After extraction of CORA data in netcdf format

A second set of tests is set up after the extraction of the measurements from Coriolis.

1. Second validation run on CORA files with feedbacks (as log files or NetCDF resubmission) to the GLOBAL INSITU TAC and correction of “repeated problems” in files or re extraction of corrected data from the GLOBAL INSITU TAC database
2. Integration of other feedbacks from different partners: (Mercator-Ocean, CLS, Altran, Meteo-France,...)
3. Suppression of duplicated measurements
4. Correction of temperature and depth on XBT measurements
5. Last objective analysis ran on the whole dataset with feedbacks to the GLOBAL INSITU TAC and correction of “repeated problems” in files
6. CF compliance checks on CORA RAW files and GRIDDED files.
Figure 12: bar chart of validated profiles (red), profiles visualised by an oceanographer (green) and QC flagged profiles (blue)

Figure 13: Validations performed on CORA dataset
### III.4 Content of the second pass of generic tests

<table>
<thead>
<tr>
<th>Name of the validation</th>
<th>description</th>
<th>Alert Validation</th>
<th>Correction applied in CORA</th>
<th>Method of Correction applied in GLOBAL TAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure on earth</td>
<td>Compare position to bathymetry, to reject on land positions. Position more than 5km distant position from nearest coastline with elevation above 50m</td>
<td>Operator visualisation</td>
<td>Position QC edited in file</td>
<td>Visualizing and manual QC position edition</td>
</tr>
<tr>
<td>Date check</td>
<td>Checks that the date correspond to the name of the file.</td>
<td>automatic</td>
<td>Quality flag edited in file</td>
<td>Visualizing and manual QC position edition</td>
</tr>
<tr>
<td>Parameter Range check</td>
<td>Check that TEMP PSAL PRES and DEPH have acceptable values -2.5&lt;TEMP&lt;45 0&lt;PSAL&lt;50 -2.5&lt;PRES&lt;20000 -2.5&lt;DEPH&lt;20000</td>
<td>automatic</td>
<td>Quality flag edited in file</td>
<td>Visualizing and manual QC position edition</td>
</tr>
<tr>
<td>Constant check</td>
<td>Check that TEMP PSAL PRES and DEPH have different values along the vertical. All the values must be identical.</td>
<td>automatic</td>
<td>Quality flag edited in file</td>
<td>Visualizing and manual QC position edition</td>
</tr>
<tr>
<td>Ascending immersion check</td>
<td>Check that PRES and/or DEPH are increasing.</td>
<td>automatic</td>
<td>Quality flag edited in file</td>
<td>CORA corrected files loaded into GLOBAL INSITU TAC database to replace former one</td>
</tr>
<tr>
<td>Duplicate levels check</td>
<td>Check that immersion levels are not duplicated.</td>
<td>automatic</td>
<td>Quality flag edited in file</td>
<td>CORA corrected files loaded into GLOBAL INSITU TAC database to replace former one</td>
</tr>
<tr>
<td>Minmax test</td>
<td>Comparison of the profiles to a minimum and a maximum reference field.</td>
<td>Operator visualisation for doubtful profiles and automatic for large errors</td>
<td>Extraction + concatenation of corrected data from GLOBAL INSITU TAC</td>
<td>Visualizing and manual QC position edition</td>
</tr>
<tr>
<td>Name of the validation</td>
<td>description</td>
<td>Alert Validation</td>
<td>Correction applied in CORA</td>
<td>Method of Correction applied in GLOBAL TAC</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>---------------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>Spike check</td>
<td>spot spikes on profiles when TEMP ΔP/ΔZ&gt;1.5 PSAL ΔP/ΔZ&gt;0.5</td>
<td>Operator visualisation</td>
<td>Extraction + concatenation of corrected data from GLOBAL INSITU TAC</td>
<td>Visualizing and manual QC position edition</td>
</tr>
<tr>
<td>Quality Flag relevance</td>
<td>Control that a given QC is relevant with the associated measure. For Example if a measure is at fill value then corresponding QC must be have the quality flag fill value</td>
<td>automatic</td>
<td>Quality flag edited in file</td>
<td>CORA corrected files loaded into GLOBAL INSITU TAC database to replace former one</td>
</tr>
<tr>
<td>Depth wrote in pressure field</td>
<td>Check that depth measurements are not written in PRES field (especially for XBT and X-CTD)</td>
<td>automatic</td>
<td>Quality flag edited in file</td>
<td>Independent correction in database (coming soon)</td>
</tr>
<tr>
<td>Duplicate profile</td>
<td>Detect duplicate profiles (see table 3 and 4 for the detection criterion and decision chart).</td>
<td>Operator visualisation</td>
<td>Suppression of worst duplicated profile</td>
<td>CORA corrected files loaded into GLOBAL INSITU TAC database to replace former one</td>
</tr>
<tr>
<td>XBT correction</td>
<td>Empirical correction of depth bias and temperature offset on XBTs. Method of M. Hamon et al based on co-localisation with CTD profiles. Correction of temperature offset Correction of depth bias</td>
<td>Operator visualisation</td>
<td>Measure and quality flag edited in file</td>
<td>No feedback to GLOBAL INSITU TAC for this correction</td>
</tr>
<tr>
<td>Assimilation feedback</td>
<td>Alerts on profiles raised by a too strong innovation value when assimilated in a model.</td>
<td>Alerts from Mercator-Ocean validated by operator (fake alert declared by operators are then analysed back at Mercator-Ocean)</td>
<td>Quality flag edited in file</td>
<td>Visualizing and manual QC position edition</td>
</tr>
</tbody>
</table>
### QUality Information Document

**INS_GLO_TS_REP_OBSERVATIONS_013_001_b**

- **Ref:** CMEMS-INS-QUID-013-001_b
- **Date:** 23/01/2015
- **Issue:** 1.6

### Table: Validation Description

<table>
<thead>
<tr>
<th>Name of the validation</th>
<th>description</th>
<th>Alert Validation</th>
<th>Correction applied in CORA</th>
<th>Method of Correction applied in GLOBAL TAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison to ENSEMBLE dataset</td>
<td>A comparison to En3v2 (ECMWF insitu T&amp;S dataset) is performed to check the temporal and spatial coverage of data.</td>
<td>N/A</td>
<td>N/A</td>
<td>Missing datasets in CORA are identified and we try to integrate them in the next release. See 15</td>
</tr>
<tr>
<td>Ultimate objective analysis</td>
<td>Last step of the process that guarantee a global (spatial and temporal) consistency of the dataset by computing a residual value between each measure and the “background” given by the climatology and the other measurements. 2 kind of alerts can be raised by the analyse objective: Standardisation alert (spikes, climatology differences) Residual alert (differences between measure and field computed)</td>
<td>automatic</td>
<td>Quality flag edited in file</td>
<td>Visualizing alert logs and manual QC position edition</td>
</tr>
</tbody>
</table>

**Errore ! Source du renvoi introuvable.: duplicated profiles detection criterion**

**Errore ! Source du renvoi introuvable.: duplicated profiles decision chart**
IV VALIDATION RESULTS

IV.1 Duplicates profiles suppression results

IV.1.1 Results for CORA-GLOBAL-04.2

Figure 3: Position of potential duplicate observations in CORA-GLOBAL-04.2

IV.2 XBT correction results

A particular attention is paid to apply a correction on the expendables bathythermograph (XBT) measurements. The aim of this work is to avoid the documented influence of XBT measurement bias on the ocean global heat content estimation (Levitus et al, 2009).

Before 1990, a large part of the present dataset is composed of measurements from expendable bathythermographs (XBT). The XBT is a rocket-shaped instrument dropped into the water from a marching ship without slowing down. The temperature sensor is linked to the ship by a capillary wire and transfers the measured temperature while falling. It is then necessary to assess the falling rate of the XBT in order to assess the depth of the measurements. There is an uncertainty of the correction method applied to the XBT profiles since the metadata of XBT profiles are often incomplete in the early period of XBT exploitation. As a consequence, Hamon et al, 2012 have developed a correction method based on the statistical fit of XBT measurements with accurate reference profiles.
IV.2.1 Summary of the method

Different issues with the data of expendable BathyThermograph (XBTs) exist and, if not corrected, they are known to contribute to anomalous global heat content variability (e.g. [12]). The XBT system measures the time elapsed since the probe entered the water and thus inaccuracies in the fall rate equation result in depth errors. There are also issues of temperature offset but usually with little dependence on depth. The correction applied on CORA-GLOBAL-04.2 dataset is an application of the method described in Hamon et al, 2012.

This correction is divided in two parts: first the computation of the thermal offset then the correction of depth. To evaluate the temperature offset and the error in depth the reference used are all the colocalised profiles (e.g. in a 3km ray, +/-15 days temporal frame, a maximum average temperature difference of 1°C and a bathymetric difference inferior to 1000m) that are not XBT and with quality flags different from 3 and 4 (suspicious and bad quality). Those references thus gather CTD, Argos profilers and mooring buoys give the number of XBT with colocalised profiles that can be found for each year.

What is an XBT profile in CORA4.2? It is a profile either in XB, BA or TE files with a WMO_INST_TYPE that refers to an XBT probe (see ). Profiles in XB files with a WMO_INST_TYPE unknown (999) and no salinity data (to avoid XCTD) are also considered as XBT. Profiles with a WMO_INST_TYPE unknown in BA or TE files cannot be qualified as XBT since many different instruments types are gathered in those files. As information on the XBT type is missing for a large part of XBT profiles in the XB files, we decided not to apply the Hanawa fall-rate for XBT depth computed with the old fall rate equation. This differs from Hamon et al, where the Hanawa correction was first applied when possible. Thus, the only correction we made for the XBT in the CORA3 database is statistical.

The temperature offset correction: this correction aims to give a value of correction for each profile as a function of the year and the category of XBT: shallow XBT or deep XBT (e.g. maximal depth >=500m). The values are computed by the difference of each XBT profile with its reference profile in the layer 30-50m (below the mixed layer and where depth errors are not important enough to explain the observed bias). Solely low temperature gradient points (e.g. <0.0025 °C/m) are used to compute those corrections. XBT and reference profiles are interpolated on standard levels from 0 to 1000m and a resolution of 10m before the calculation. The final offset is the median of all those differences.

The depth error correction: Results of this depth correction are second order polynomial coefficients depending of the year, the depth and the category of XBT profile. In this second step there are not two but four different categories: deep and hot, deep and cold, shallow and hot, shallow and cold (e.g. maximal depth 500m and mean temperature 11°C). To evaluate the error in depth we use the following formula:

\[ dZ = \frac{T - T_{ref}}{\partial_z T_{ref}} \]

\[ dZ = \frac{T - T_{ref}}{\partial_z T_{ref}} \]

Some cursors such as a minimum number of collocations per level and a maximum value of depth error allow improving the quality of the median profile gotten from the raw depth errors in each of the four categories. Then we fit a second order polynomial on the median depth errors and we get three coefficients:
\[ Z_{\text{true}} - Z = aZ^2 + bZ + c \]

The \( c \) coefficient is replaced by the mean of the depth error in the layer 30-200m to ignore the noise due to the mixed layer. The values of the coefficient are given in . The difference between the XBT profile and the reference profile before and after applying the correction is plotted in Fig: 17. It shows that the corrected XBT measurements are close from the reference profiles. The comparison works better for data older than 1990 thanks to a higher number of match points as inputs to the statistical correction method. After 1990 the comparison is still better than the original measurements, but the signal to noise ratio is lower.

![Temperature anomaly between XBT and reference profiles](image)

**Figure 14**: Yearly temperature anomaly between XBT and reference profiles before (A) and after (B) temperature offset and depth corrections in CORA4.2.

### IV.2.2 Results for CORA-GLOBAL-4.2

Over 1 million XBT profiles have been corrected thanks to the Hamon et al 2012 correction method on the period 1970-2014. The earliest XBT profiles (1965-1969) being too sparse to perform reliable statistics, they have not been corrected. Figure 15 gives the yearly mean difference between the measured temperature field and the colocalised XBT profiles before and after the XBT correction. The temperature anomaly between XBTs and associated reference profiles decreases after the XBT correction. The anomaly overreaches 0.5 °C in the early period and in the vicinity of 1000m depth in the 2000s. The early period overstated anomaly is a consequence of the low number of deep XBT profiles (>500m) between 1975 and 1980. The late period anomaly is also caused by a lack of deep XBT profiles (T-5a dn fast deep) in the 2000s.
IV.3 Objective analysis

The software ISAS\(^2\) used to perform the last Objective Analysis is divided in 2 steps:

1) A first standardization that raises alerts on climatological test and spike test.

2) The objective analysis itself that computes a residual value between the measure and the field build with all the data available in the time-space area for each measure. We raise an alert on the following criteria:

\[ R > 4 \]

\[ R = \text{RESIDUAL}/\sqrt{\text{ERR_UR}^2 + \text{ERR_ME}^2} \]

RESIDUAL is the difference between a given observation and the analysis interpolated value.

ERR_UR is the error linked to unresolved scales. It is an a priori parameter of the analysis.

ERR_ME is the error of the measure and it is also given as a parameter for each type of measurement.

You can find more information on how ISAS (see 31) is used on CORA dataset in the user manual of the CORA product.

Note that the present version of ISAS cannot handle time series data containing only one measure vertically, so drifters, ferry-box, moorings are not analysed in this objective analysis nor displayed in gridded fields.

IV.3.1 Results for CORA-GLOBAL-4.2

The objective analysis raises alerts on profiles that have a strong innovation. These profiles are individually controlled by a scientist. Figure 16 gives the space repartition of warning profiles raised by ISAS. Many alerts are associated to WOCE profiles in the 1990s since the first guess of the objective analysis often fail to present a basin scale variability in the 1990s.

Figure 5: Position of CORA profiles alerts raised by the objective analysis on temperature from CORA4.2 dataset
V VALIDATION SYNTHESIS

Although we cannot provide an exhaustive comparison of the in situ measurements with an adequate reference since in situ measurements do not have any references but themselves) we can provide a global idea of the quality expected in CORA dataset through the residual value (difference between a measure and the corresponding field computed by the objective analysis).

Generic tests on meta-data, values ranges, and quality flag consistency are very powerful tools and the repetition of those checks, made by different programs, is useful since special case sometime falls through the cracks. We particularly recommend to use carefully coastal observations, typically observations located closer than 50km off coasts.

About the integration of time-series, most of the validation relies on providers (drifters, ferrybox, TSG, moorings) but a reinforced validation on those special temporal features will be developed for the next release of CORA.

The numerous objective analysis performed on measures guarantee that delayed-mode data are coherent together.

The duplicate profiles suppression has given good results with independent validation of the method. However, it is important for the users to know that this detection of duplicates is deliberately non exhaustive in order to guarantee that no relevant/unique data would be removed from the dataset.

XBT correction follows Hamon et al, 2011 empirical method. This method was first applied in CORA-GLOBAL-03.2. It has not been applied in CORA-GLOBAL-04.0 since the number of XBT profiles was not sufficient to apply the statistical method. This problem has been solved in CORA-GLOBAL-04.2 and corrected XBT profiles are now available. The correction of XBT measurements extracted from EN4 (before 1989) is consistent with the correction applied after 1990.

Even if we also have some specific diagnostics on specific instrument/platform (for instance Argo profilers), we cannot have the same expertise than the PIs. Thus as a final gatherer of data, GLOBAL INS TAC and R&D Coriolis team mostly trust on the data providers to guarantee a level of quality. Moreover our additional checks, performed on CORA dataset as a reanalysis, guarantee the global consistency of the dataset but it cannot prevent from any error in data nor meta data, that is why GLOBAL INS TAC encourages its users to send feedback on the datasets to improve their quality.
VI BIBLIOGRAPHY


