Comparison of CORA and EN4 in-situ datasets validation methods, toward a better quality merged dataset.

T. Szekely, R. Killick, A. Piron, J. Gourrion, G. Reverdin
## Global in-situ datasets

### In-situ TEMP & PSAL datasets comparison

<table>
<thead>
<tr>
<th>Provider</th>
<th>Coverage</th>
<th>Validation</th>
<th>Data types</th>
<th>Distributed data</th>
</tr>
</thead>
<tbody>
<tr>
<td>CORA 5.0</td>
<td>Copernicus CMEMS</td>
<td>1950-2015</td>
<td>Semi – Automatic, all levels, Visual control of suspicious data</td>
<td>Profiles + timeseries</td>
</tr>
<tr>
<td>EN.4.1.1</td>
<td>Mettoffice.org</td>
<td>1900-Present days</td>
<td>Automatic, 400 vertical levels max</td>
<td>Profiles only, max 400 levels</td>
</tr>
<tr>
<td>WOD13</td>
<td>Nodc.noaa.gov</td>
<td>1772-2012</td>
<td>Automatic – standard levels validation only</td>
<td>Profiles + timeseries</td>
</tr>
</tbody>
</table>

**Application**: Reanalysis, Climate studies at bassin/global scale, scientific studies (local to global scale), Satellite SST calibration, etc....

Data validation methods and choices in the dataset construction may differ.
Small differences in the dataset format and data providers leads to major differences in data coverage!
Small differences in the dataset format and data providers leads to major differences in data coverage!
CORA/EN4 comparison

CORA profiles 01 - 2015

EN4 profiles 01 - 2015
How to compare CORA/EN4 data Validation methods?

- **Extraction** of 1 year of TEMP & PSAL in-situ measurements

- **Validation** of the whole dataset by **CORIOLIS** (CORA) and **Metoffice** (EN.4)

- **Comparison** of the profile flags

- **Focus** on the tests applicable in both datasets.
Dataset extraction

1 year of global TEMP and PSAL measurements extracted from INSITU_GLO_NRTOBSERVATIONS_013_030

- Focus on profilers, CTD, Sea Mammals and XBT validation
- Various instruments and data sources
Flag status comparison

64% of the EN.4 bad flags are good in CORA
60% of the CORA bad flags are good in EN.4
→ Less than 1% of the dataset
Global comparison

What is the bad flag repartition in CORA and EN.4?

There is a divergence in the flag attribution.
How can CORA learn from the EN.4 methods to improve the CMEMS dataset validation?
Validation comparison

→ 5 tests flag most of the undetected bad measurements in CORA
(See Ingleby & Huddleston, 2005 for the test definition)

Number of Bad flags applied by the top 5 tests by EN.4 and associated CORA flags if any.

How many of the undetected bad measurements are actually wrong measurements?
Validation comparison

→ 5 tests flag most of the undetected bad measurements in CORA

Number of Bad flags applied by the top 5 tests by EN.4 and associated CORA flags if any.

Methodology:

- Profiles with undetected bad measurements have to be visualised by a PI
  - If at least one of the profile measurement is flagged by the PI, the profile is a good detection
    - If not, the profile is a bad detection
  - The test efficiency is ratio
Validation comparison

Tests efficiency rate from the sampled detected profile list, after a visual check, for XBT, CTD, profilers and sea mammals.

- The tests efficiency varies with the instrument types
- The flags are often not set at the detected levels.
- A bad profile may be detected by numerous test in EN.4 but only one in CORA.
- The tests may be implemented in a certain order to maximize the validation efficiency
- Taking the EN4 tests in automatic mode is a bad solution → **To much wrong detections**

- The **best order** is Deph check → Vertical stability → Vertical check → Bayesian check → Bayesian and Buddy

- The visual check efficiency rate is good if we keep the Deph check, Vertical stability and Vertical check, with a PI visual control of the detection.
Toward CORA 5.1

CORA 5.1 release: April 2018

→ Updated QC methods
→ Updated dataset
→ Profiles and time series
Copernicus catalogue

http://marine.copernicus.eu/