

# Improved statistical method for hydrographic climatic records quality control

Jérôme Gourrion, CORIOLIS R&D, CNRS (France), [jerome.gourrion@ifremer.fr](mailto:jerome.gourrion@ifremer.fr)

Tanguy Szekely, CORIOLIS R&D, CNRS (France), [tanguy.szekely@ifremer.fr](mailto:tanguy.szekely@ifremer.fr)

Climate research benefits from the continuous development of global in-situ hydrographic networks in the last decades. Apart from the increasing volume of observations available on a large range of temporal and spatial scales, a critical aspect concerns the ability to constantly improve the quality of the datasets.

In the context of the Coriolis Dataset for ReAnalysis (CORA) version 4.2, a new quality control method based on a local comparison to historical extreme values ever observed is developed, implemented and validated.

In this talk, we present how temperature, salinity and potential density validity intervals are directly estimated from minimum and maximum values from an historical reference dataset, rather than from traditional mean and standard deviation estimates. The concept is presented through examples and empirical validation results.

Such an approach avoids strong statistical assumptions on the data distributions such as unimodality, absence of skewness and spatially homogeneous kurtosis. As a new feature, it also allows addressing simultaneously the two main objectives of a quality control strategy, i.e. increase the number of good detections while reducing the number of false alarms.

The reference dataset is presently built from the fusion of 1) all ARGO profiles up to early 2014, 2) 3 historical CTD datasets and 3) the Sea Mammals CTD profiles from the MEOP database. All datasets are extensively and manually quality controlled. In this communication, the latest method validation results are also presented.

The method has been implemented in the latest version of the CORA dataset and will benefit to the next version of the Copernicus CMEMS dataset.

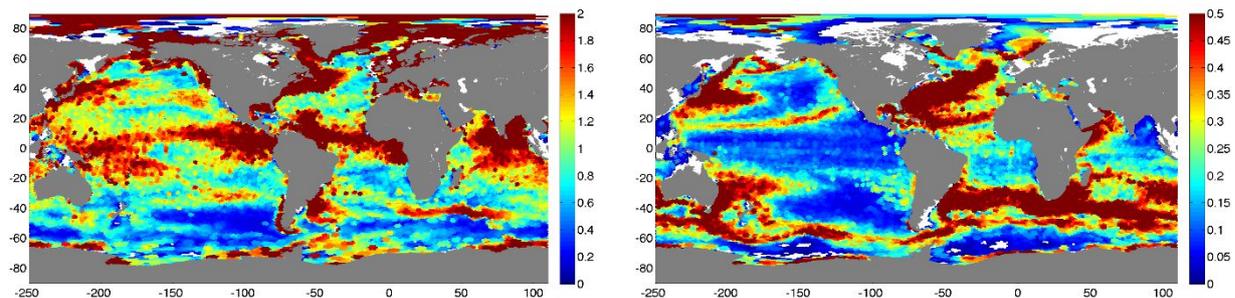


Figure 1: Examples of surface (left) and 500 dbar (right) minimum / maximum salinity interval width used as reference fields for outlier detection.