Improved statistical method for hydrographic climatic records quality control

Jérôme Gourrion and Tanguy Szekely (R&D Coriolis / CNRS)

Contact: jerome.gourrion@ifremer.fr

A new strategy : why ?

- The input T/S dataset to the CMEMS distribution still contains erroneous data despite the currently implemented automatic detection criteria

  → ***insufficient detection efficiency***

- At the visual inspection step, the operator observes that low occurence oceanic events are often classified as erroneous data

  → ***the amount of false alarms may often reach 70 %***

Local comparison to historical knowledge

To detect outliers, offsets or drifts in T/S data, **a geographically-based criterion** is used, comparing the observation to a reference validity interval derived from available historical information. The NRT and DT CORIOLIS processing chains use validity intervals based on local first and second order moments, i.e. ***mean +/- N*std***, with N constant.

Detection objectives:

① ***maximize the number of "good" detections***

② ***minimize the number of false alarms***

Local comparison to historical knowledge

To detect outliers, offsets or drifts in T/S data, **a geographically-based criterion** is used, comparing the observation to a reference validity interval derived from available historical information. The NRT and DT CORIOLIS processing chains use validity intervals based on local first and second order moments, i.e. ***mean +/- N*std***, with N constant.

Detection objectives:

① ***maximize the number of "good" detections***

② ***minimize the number of false alarms***

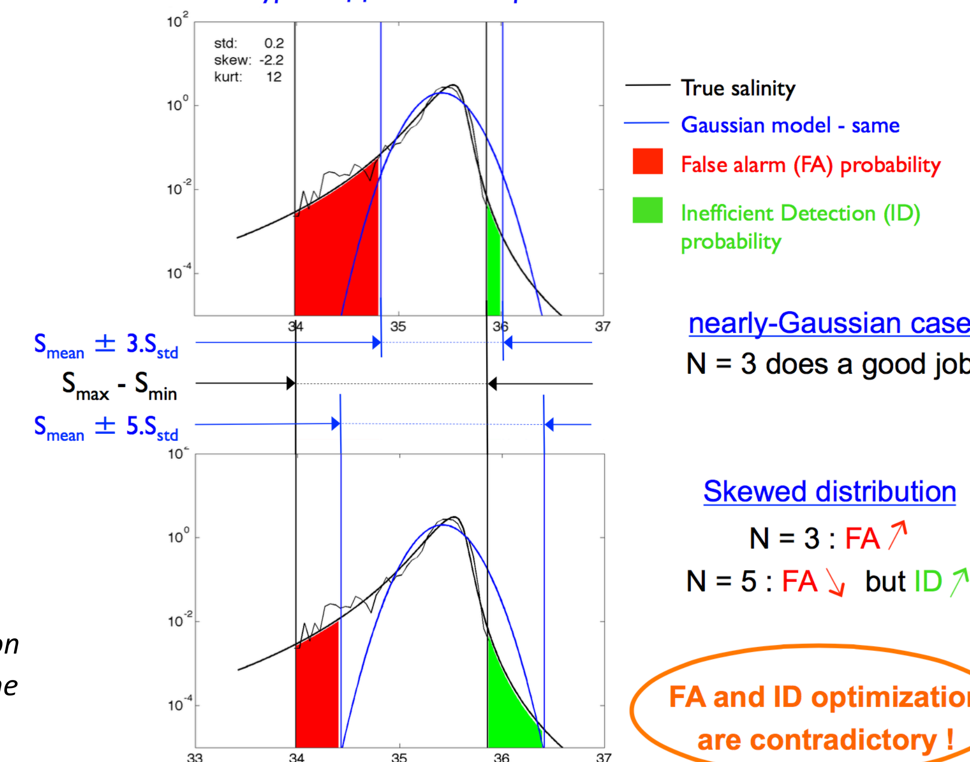
Main assumptions on the data distribution:

1. ***unimodal***
2. ***symmetric*** *i.e. without skewness → over-detection on one side of the distribution and under-detection on the other*
3. ***homogeneously non-Gaussian*** i.e. *constant kurtosis*

→ *non-homogeneous degree of detection efficiency*

Impact of nearly-Gaussian assumption :a skewed salinity distribution

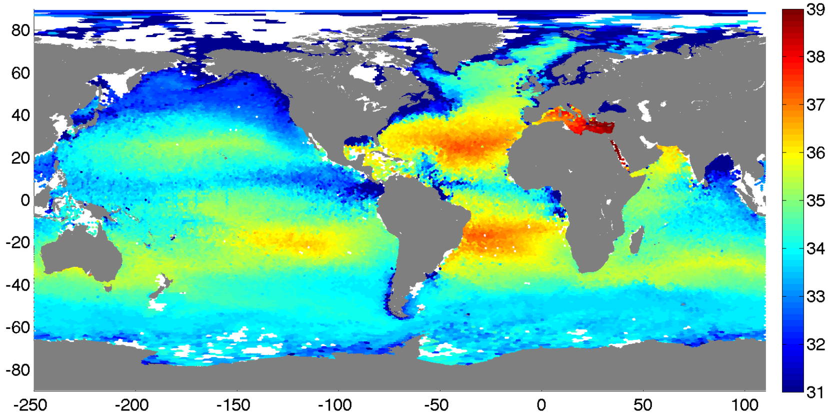
typical of fresh water input or advection area



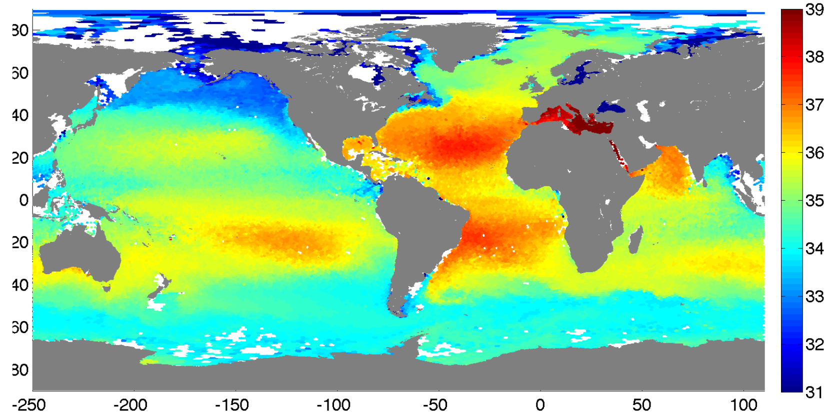
FA and ID optimization are contradictory !

Local comparison to historical knowledge

Minimum



Maximum

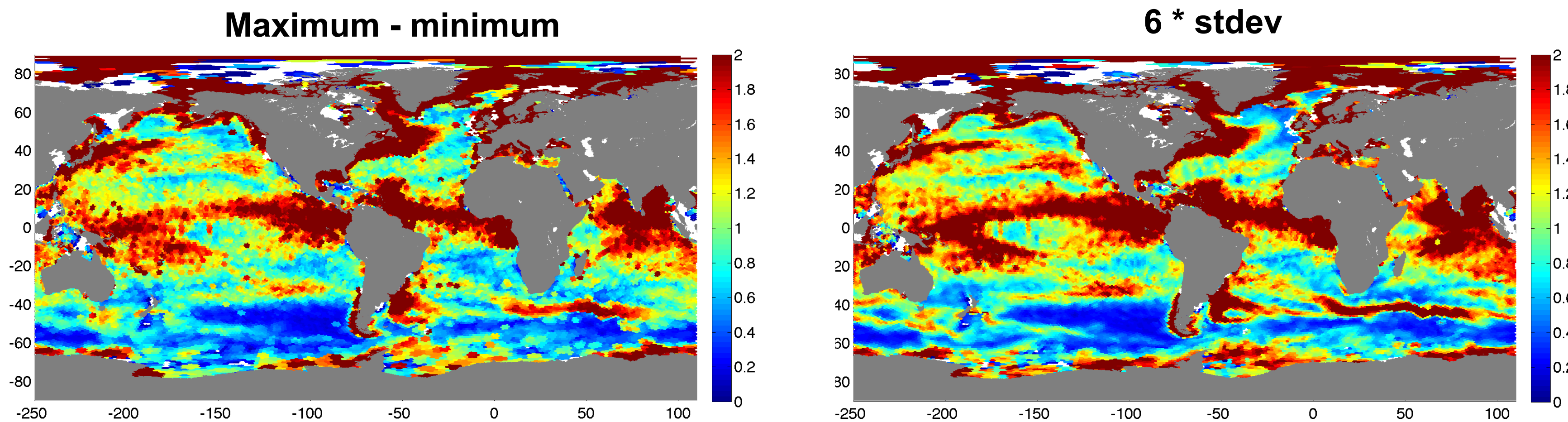


- Min / Max statistics are accumulated at global scale over 20 db thick layers from 0 to 2000 db. Typical horizontal resolution: 110 km (330 km)

- Several datasets have been used:
  - ✓ Profiles from ARGO up to March 2014
  - ✓ CTD from World Ocean Database 2013, ICES and SISMER
  - ✓ CTD mounted on Sea Mammals (MEOP database, F. Roquet)

- All datasets were manually QCed to reach the required quality level for the robust estimation of such statistical parameters

Local comparison to historical knowledge



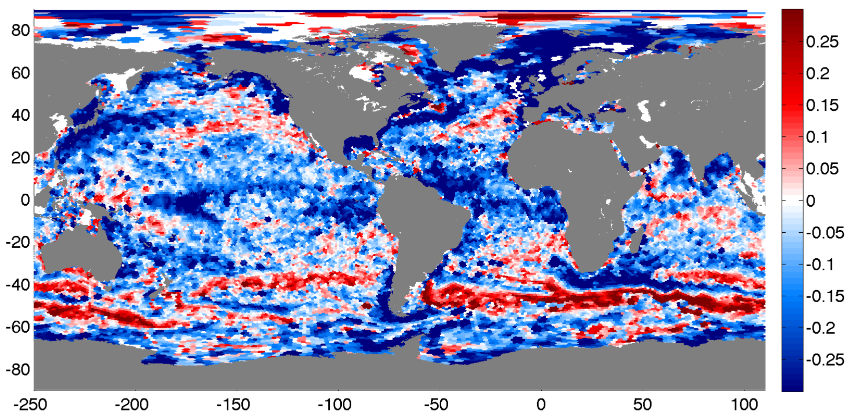
- Min / Max statistics are accumulated at global scale over 20 db thick layers from 0 to 2000 db. Typical horizontal resolution: 110 km (330 km)

- Several datasets have been used:
  - ✓ Profiles from ARGO up to March 2014
  - ✓ CTD from World Ocean Database 2013, ICES and SISMER
  - ✓ CTD mounted on Sea Mammals (MEOP database, F. Roquet)

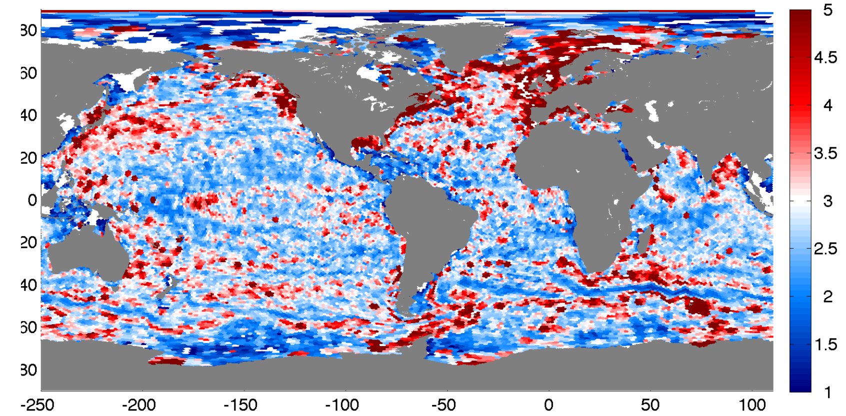
- All datasets were manually QCed to reach the required quality level for the robust estimation of such statistical parameters

Local comparison to historical knowledge

Normalized interval center shift



Interval width ratio

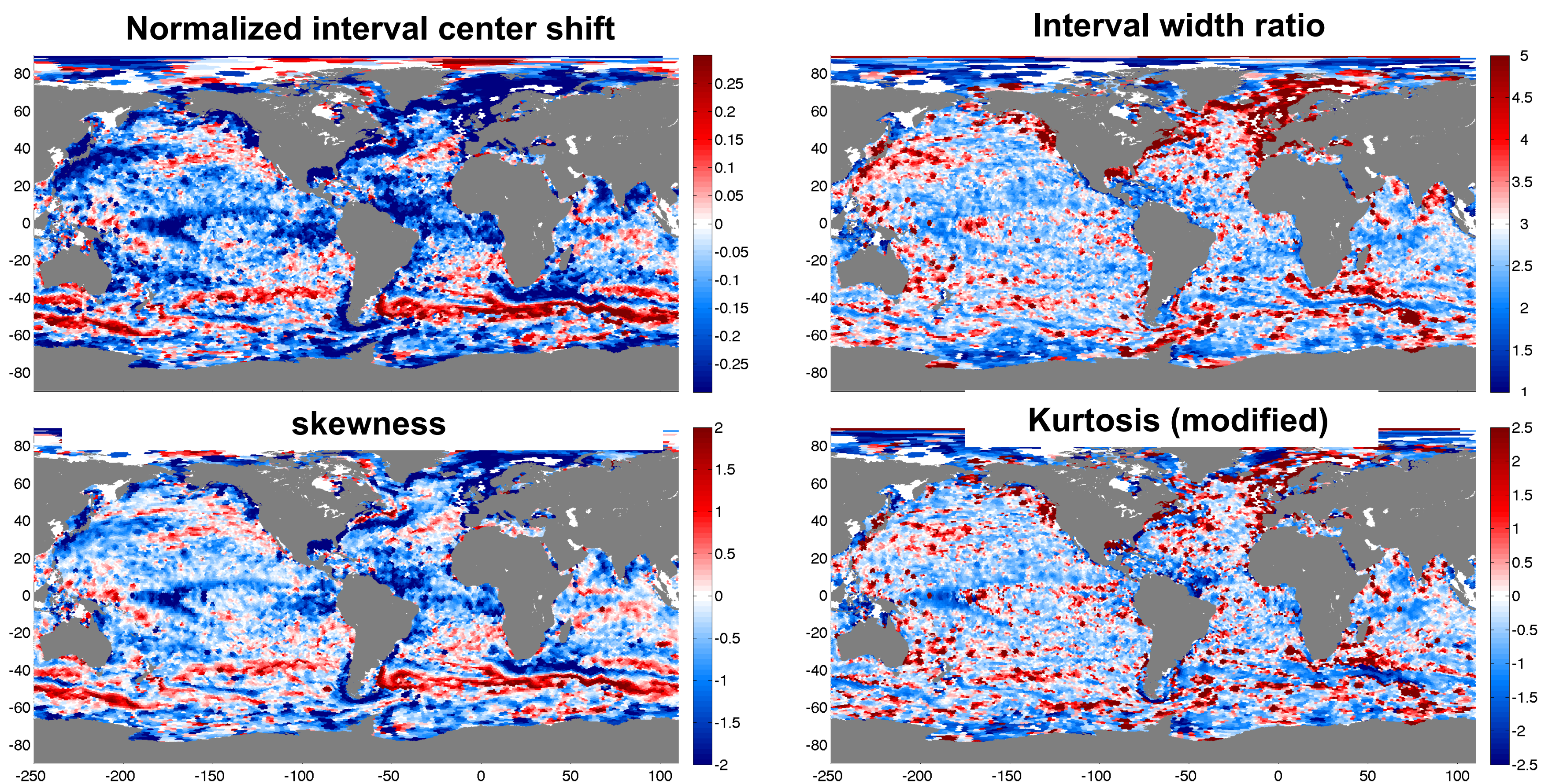


- Min / Max statistics are accumulated at global scale over 20 db thick layers from 0 to 2000 db. Typical horizontal resolution: 110 km (330 km)

- Several datasets have been used:
  - ✓ Profiles from ARGO up to March 2014
  - ✓ CTD from World Ocean Database 2013, ICES and SISMER
  - ✓ CTD mounted on Sea Mammals (MEOP database, F. Roquet)

- All datasets were manually QCed to reach the required quality level for the robust estimation of such statistical parameters

Local comparison to historical knowledge

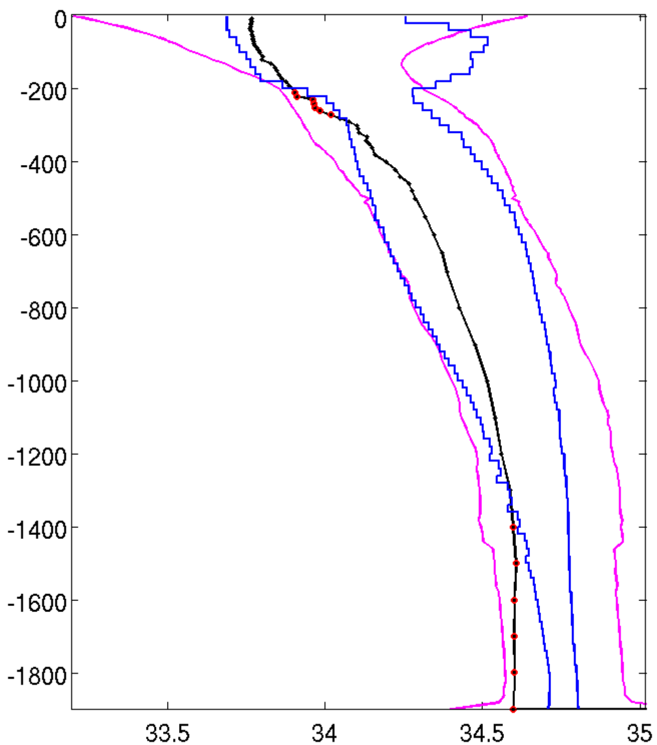


Skewness may shift the validity interval by as much as 0.5 psu !

Presence of kurtosis explains most of the local N parameter variations

Illustration through examples

Example of an ARGO profile with offset



ARGO profile with detected anomalies

Mean +/- 5 Std – based validity interval

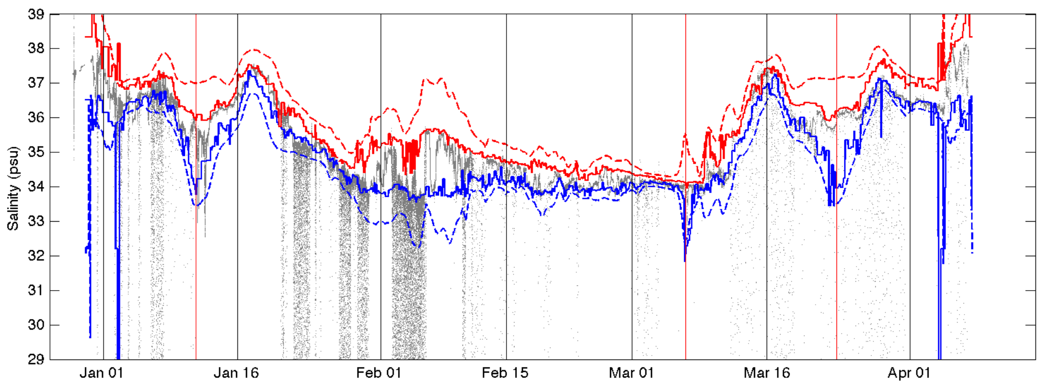
Min / Max – based validity interval

→ **Detection is more efficient, especially at depth**

Illustration through examples

Example of a CTD onboard a round-the-world sailing ship

Courtesy of J.Salat (ICM/CSIC,Spain)



Max from min/max

Min from min/max

Max from mean/std

Min from mean/std

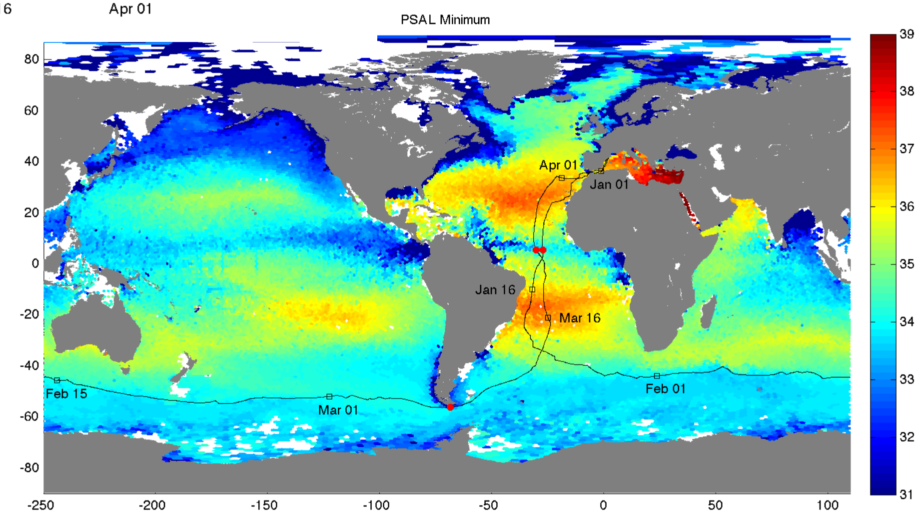
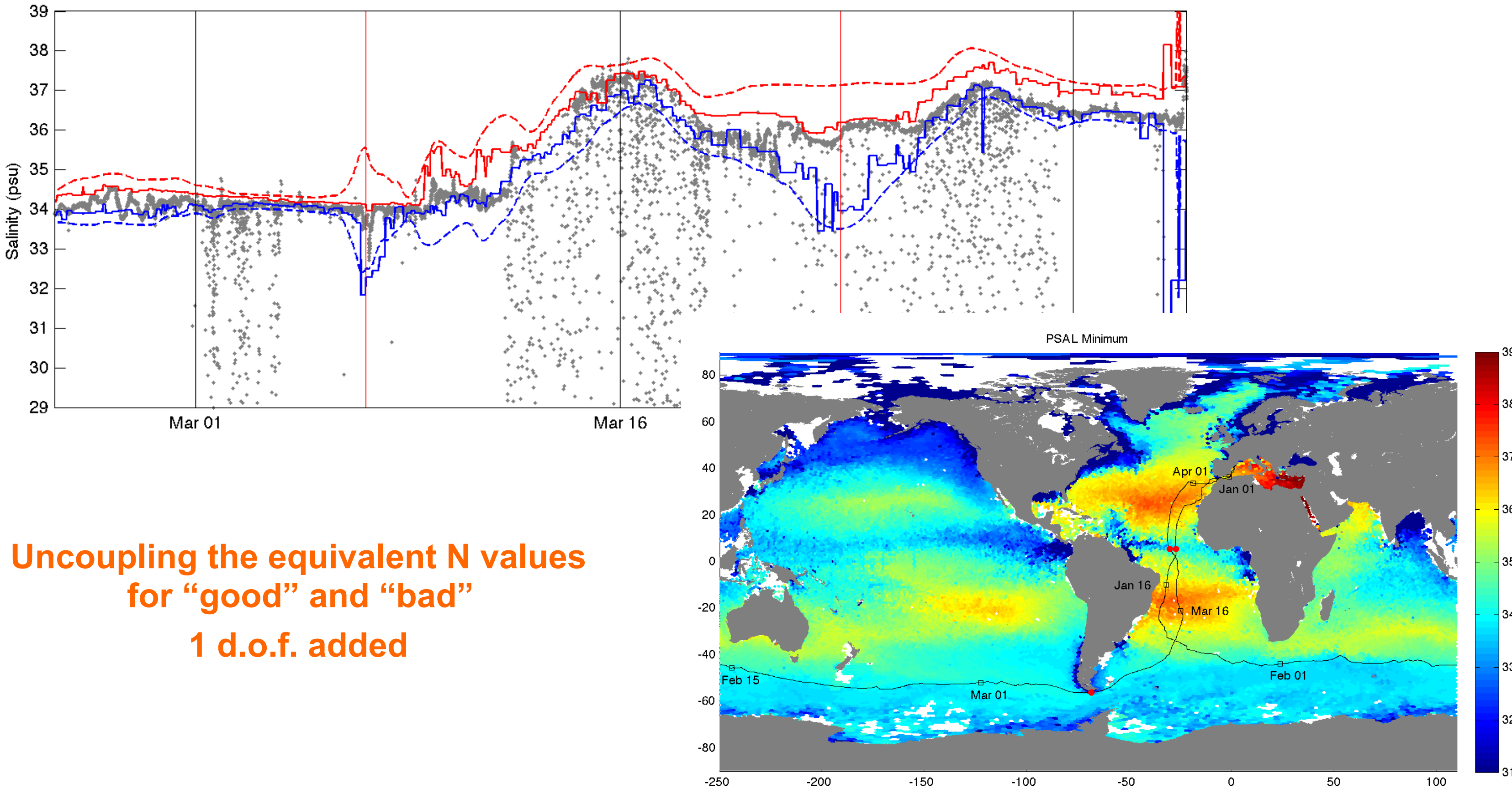


Illustration through examples

Example of a CTD onboard a round-the-world sailing ship



Uncoupling the equivalent N values for "good" and "bad"

1 d.o.f. added

Statistical validation

“Good” and “bad” detection statistics: impact of the reference dataset

- **Mean +/- 5 * std**
- **ARGO data from 2012 through 2015**
- **Truth given by CORA 4.2 flags**

WOA-01

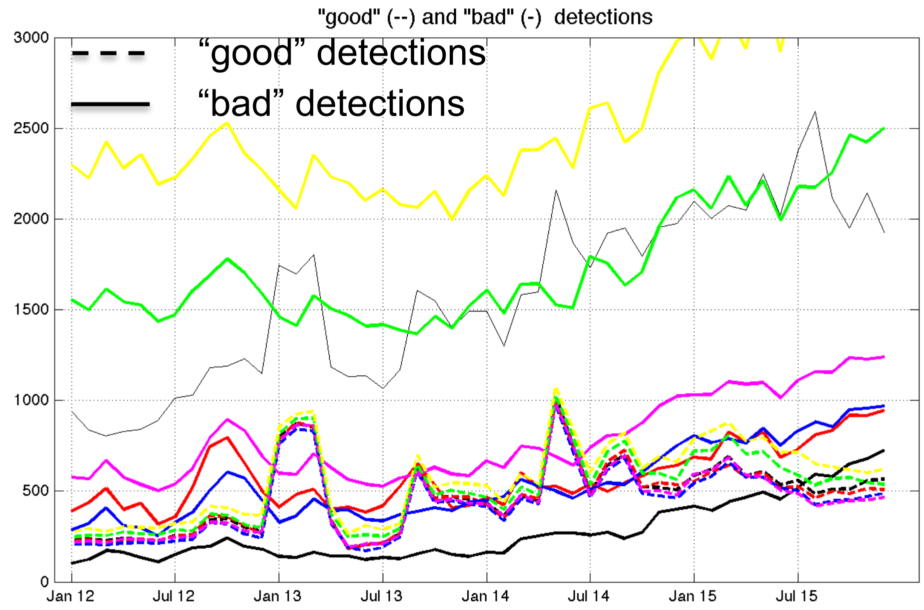
WOA-05

WOA-09

ARIVO

WOA-13

Present dataset



- ***“good” detections : little dependent on the reference dataset***
- ***“bad” detections : highly dependent on the reference distribution robustness***
- ***the present dataset seems to better describe the variability at global scale***
- ***since early 2014, statistics degrade for all datasets (atypical variability)***

Statistical validation

“Good” and “bad” detection statistics: impact of the method

- **Present dataset**
- **ARGO data from 2012 through 2015**
- **Truth given by CORA 4.2 flags**

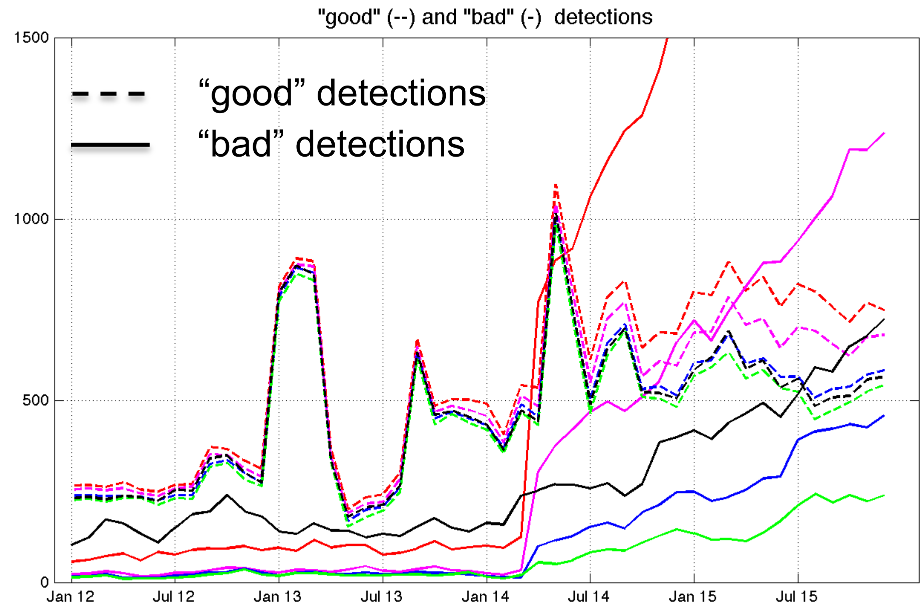
Mean/Std, 330 km

Min/Max, 330 km

Min/Max, 330 km, +/- 0.01 psu

Min/Max, 550 km, +/- 0.01 psu

Min/Max, 770 km, +/- 0.01 psu



“bad” detections :

- ***discontinuity after early 2014***
- ***clear improvement when accounting for instrument noise***
- ***further improvement assuming ergodicity***

Summary

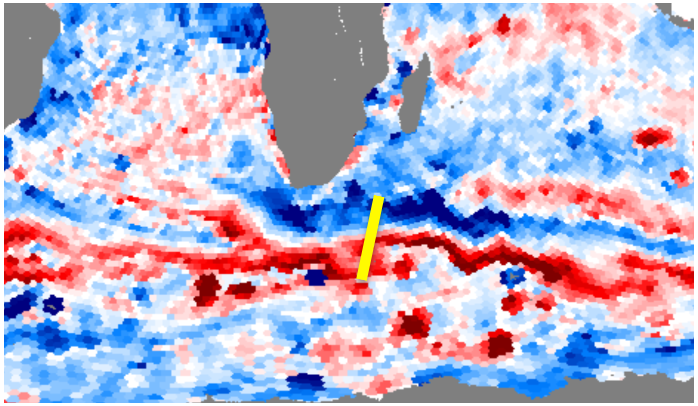
- A new statistical approach for automatic QC based on local comparison to historical observations
- One additional degree of freedom in regard of the “classical” mean/std approach
- Allows to optimize simultaneously “good” and “bad” detections
- Severe efficiency increase in terms of human power during delayed-time QC
- Reference fields need to be updated from new variability observations

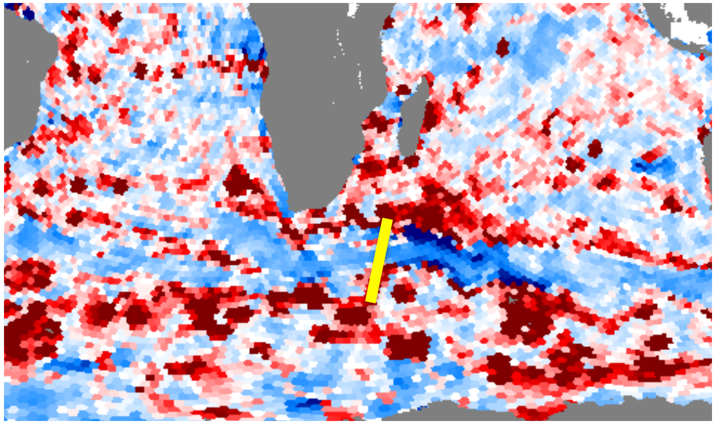
On going work and perspectives

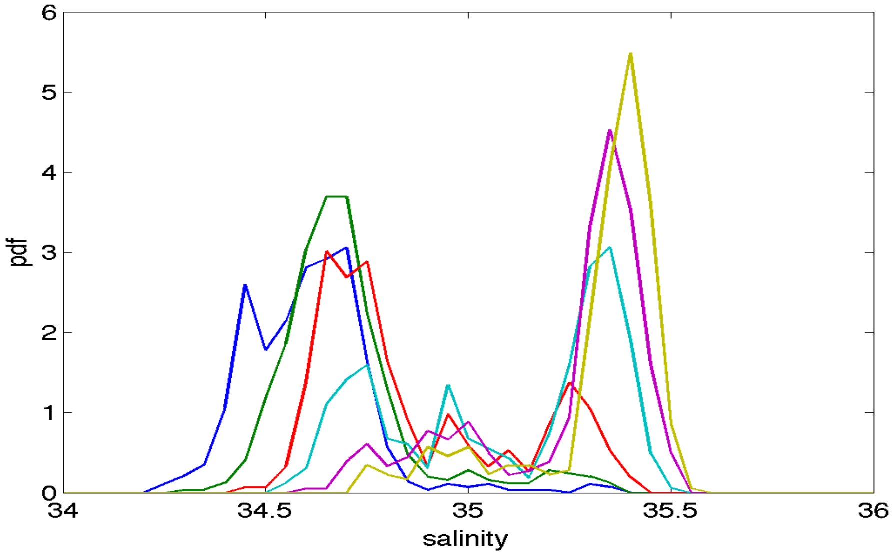
- Extend reference dataset up to late 2015
- Min/Max Implementation at regional scale (Arctic Ocean in CMEMS-INSTAC)
- Min/Max implementation at MERCATOR QC level
- Min/Max implementation in the real-time QC at CORIOLIS (CMEMS-INSTAC-GLOBAL)
- Focus on the multimodal cases: refined characterization of an historical dataset

Thank you for your attention !

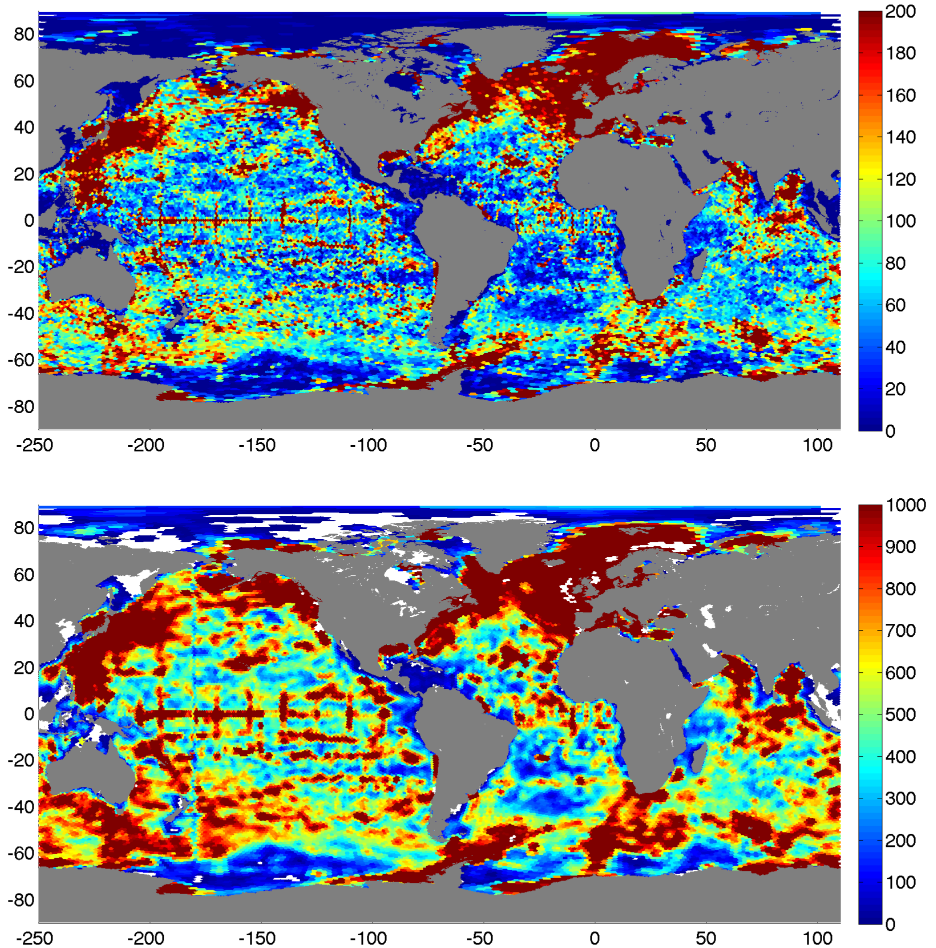
Contact: jerome.gourrion@ifremer.fr






6
5
4
3
2
1
0
pdf
34
34.5
35
35.5
36
salinity

Number of samples



Global scale strategy : present status

Copernicus IN SITU TAC - CMEMS

Coriolis

N. Ferry's detection method *(from Cabanes et al. 2013)*

Based on this 17-yr-long reanalysis, "innovation" (i.e. observation minus model background) statistics for in-situ temperature and salinity profiles were collected and used to detect suspicious profiles and provide a black list of observations present in CORA3 dataset. This observation screening is known as background quality control. The probability density functions (PDFs) of the innovations are calculated as a function of spatial location (x, y, z) in the global ocean. We find that in most places, innovation PDFs are very close to a normal distribution. Therefore, we assume that innovations have a Gaussian distribution and that the tails of the probability density function contain suspicious observations. First, the collected innovations are binned on a $5° \times 5°$ grid on the horizontal, the model vertical grid, and the season. In each cell of this 4-dimensional grid, we estimate two parameters, which are the mean M and standard deviation STD. These parameters are used to define the following space- and season-dependent threshold value:

$$T = |M| + N \times \mathrm{STD}, \qquad (2)$$

with N being an empirical parameter.

In a second stage, we perform the observation screening for each profile. At a given depth, an observation is considered suspicious if the following two criteria are satisfied:

1. $|\mathrm{innovation}| > T$
2. $|\mathrm{obs\text{-}clim}| > 0.5\ |\mathrm{innovation}|$

The first criterion diagnoses whether the innovation is abnormally large, which would most likely be due to an erroneous observation. Condition 2 avoids rejecting "good" observations (i.e. those that are close to the climatology) in the case of a biased model background. In the case of a good observation and a biased model background, LHS of criterion 2 is small and RHS is large, implying that the condition is not satisfied. This criterion significantly reduces the number of good observations that may be rejected (false alarms). The threshold value (0.5) in test 2 is empirical and has been tuned in order to minimize the false alarms. A small threshold value implies fewer false alarms, but also fewer detections of bad profiles.

Ocean Sci., 9, 1–18, 2013

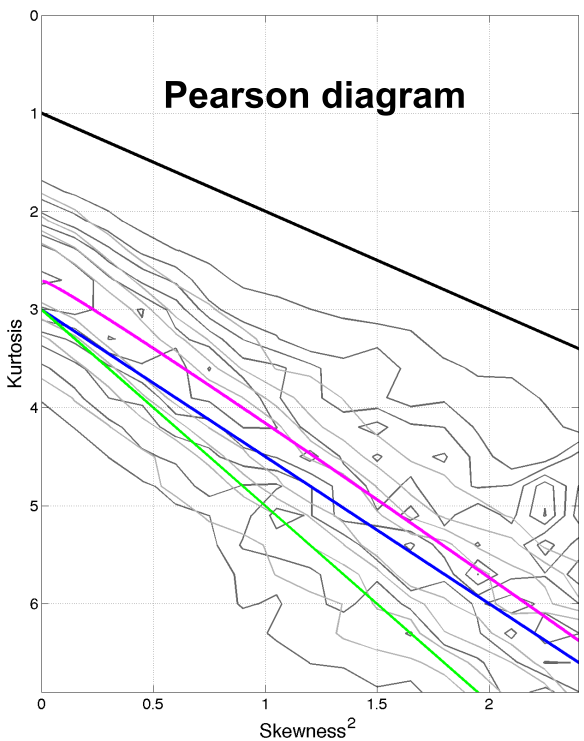
All measurements are tested.

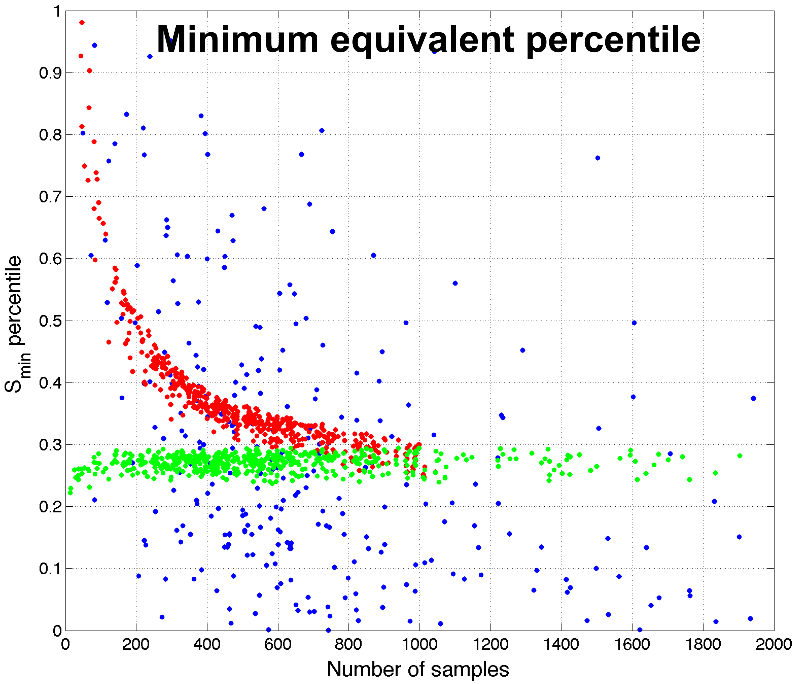
- Innovation is computed as the difference between data and model background
- An innovation climatology (5°x5°) is built
- A climatological threshold **T** is built
- Test 1 (T1) and Test 2 (T2) are combined and applied

Min / Max : equivalent percentile

Requires choosing an analytical description of the local distributions …

Using empirical S/K

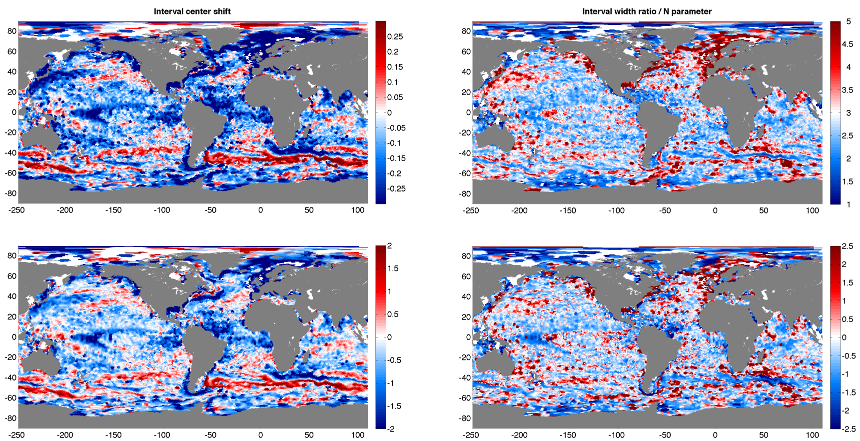


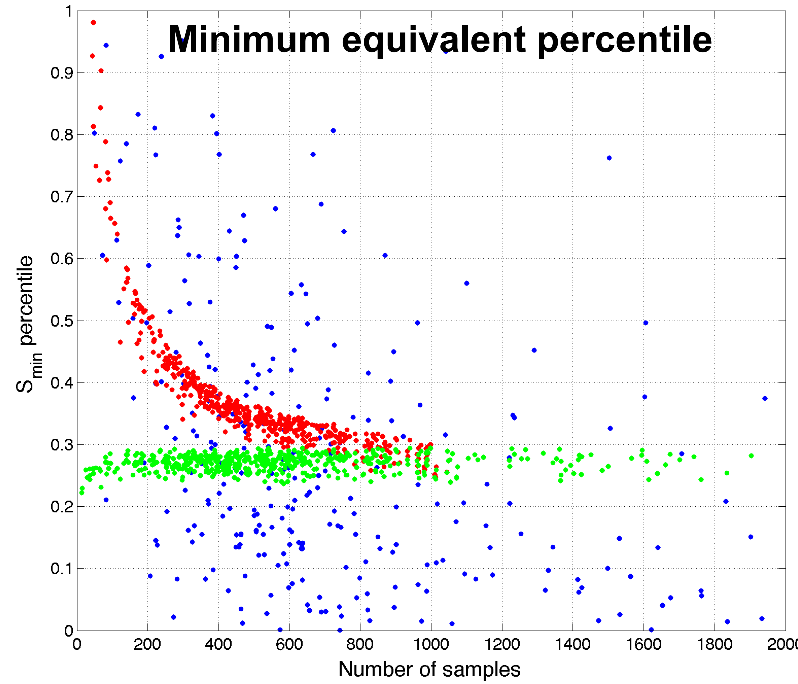


Min / Max : equivalent percentile

Requires choosing an analytical description of the local distributions …

Estimating S/K from Min/Max





Min / Max : equivalent percentile

Requires choosing an analytical description of the local distributions …

Correcting Min/Max from number of samples