WP4: Benchmarks

Lead: CNRS-LGGE
Contributions: all partners

SANGOMA final meeting – November 5-6, 2015

OUTLINE

Purpose
Why benchmarks in SANGOMA?

Benchmarks
What kind of benchmarks in SANGOMA?

Tasks / deliverables
What has been produced / delivered?

Recommendations
What do we advise to operational centers?

PURPOSE

Why benchmarks in SANGOMA?

Comparison and assessment
of data assimilation methods

Definition of appropriate metrics

Evaluate the importance
of non-linear/non-Gaussian behaviours

Provide recommendations for future developments
in the Copernicus operational systems

BENCHMARKS

What kind of benchmarks in SANGOMA?

A hierarchy of systems of different complexity
from small-scale to realistic large-scale
close to operational configuration

Three benchmarks:

small case: portable Lorenz model

medium case: portable ocean case of NEMO
(double gyre configuration)

large case: realistic configuration of NEMO
(North Atlantic at 1/4° resolution)

TASKS / DELIVERABLES

TASK 4.1: Detailed specification of benchmarks
- **DL4.1:** Benchmark definition (mo 12)
- **DL4.2:** Benchmark implementation (mo 24)

TASK 4.2: Defintion of metrics
- **DL4.3**: Report on metrics (mo 30)

TASK 4.3/4.4: Running small and medium benchmarks
- **DL4.4**: Metrics obtained with these benchmarks (mo 48)

TASK 4.5: Diagnostic of non-Gaussian behaviours in large case benchmark

TASK 4.6: Running large case benchmark
- **DL4.5**: Metrics obtained with large benchmark (mo 48)

TASK 4.1: Detailed specification of benchmarks

This task was fulfilled during the first two years of the project through 2 deliverables:

DL4.1: Benchmark definition (Mo 12)

specification of the model configurations:
- Lorenz, NEMO double gyre, NEMO North Atlantic

specification of the assimilation problem:
- time settings, uncertainties in the system, observations

distribution of the model configurations

DL4.2: Benchmark implementation (Mo 24)

implementation plan for every SANGOMA partner:
- small case: AWI, GHER, CNRS/LGGE
- medium case: AWI, GHER, CNRS/LGGE, TUDelft
- large case: GHER, CNRS/LGGE

defintion of the assimilation scheme used by each partner

Stochastic Assimilation for the Next Generation Ocean Model Applications

Benchmarks

Comparison and assessment of impacts of assimilation methods on systems of different complexity:

- Small case benchmark: Lorenz-40 model
- Medium case benchmark: double-gyre NEMO configuration
- Large case benchmark: North-Atlantic 1/4° NEMO/LOBSTER configuration

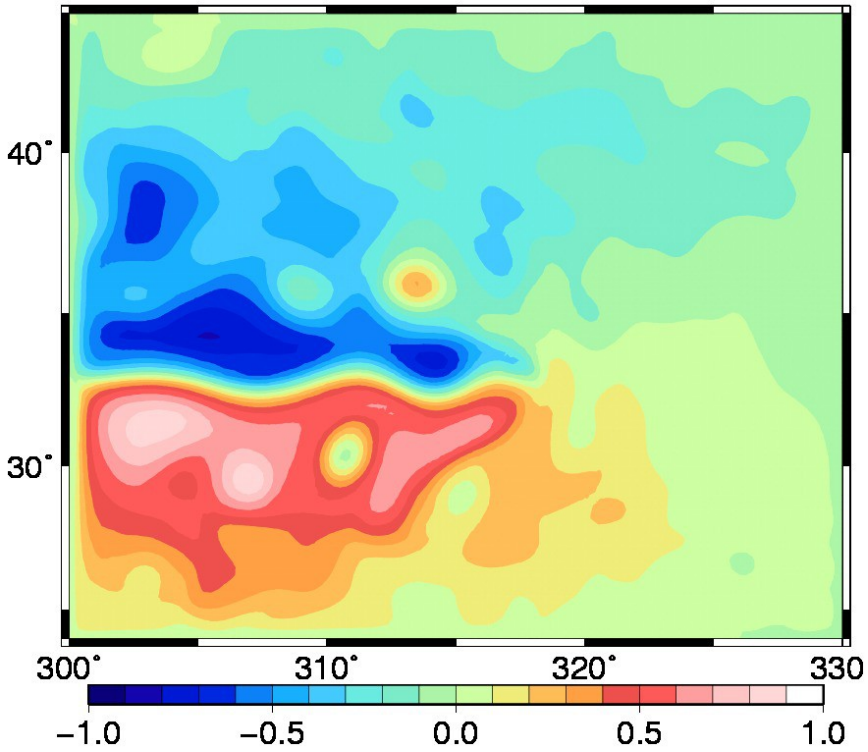
The benchmarks include (i) the detailed specification of the model configurations and assimilation alogrithm, (ii) the definition of a set of metrics to assess the performance of the assimilation systems, and (iii) the eveluation of the results of the experiments:

- Detailed specification of benchmarks
- Definition of metrics
- Evaluation of the results

SANGOMA Stochastic Assimilation for the Next Generation Ocean Model Applications

Medium case benchmark: Double-gyre NEMO configuration

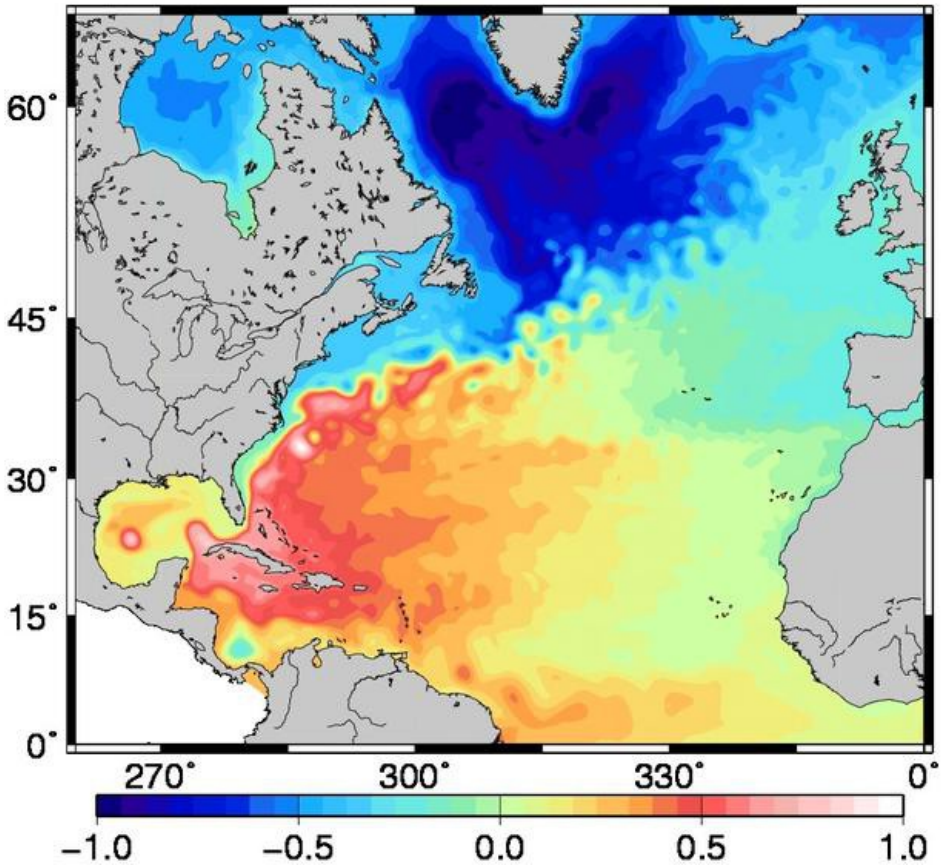
The medium case benchmark is based on an idealized configuration of the NEMO ocean model: a square and 5000-meter deep flat bottom ocean at mid latitudes (the so called square-box or SQB configuration).



Stochastic Assimilation for the Next Generation Ocean Model Applications

Home
Overview
Partners
Tools
Benchmarks
Documents
Events
Jobs
Links
Intranet

Large case benchmarks: North-Atlantic 1/4° NEMO/LOBSTER configuration



TASK 4.2: Definition of metrics

This task was fulfilled during the first two years of the project through deliverable 4.3 :

List of probabilistic metrics in DL4.3 (Mo 30):

- Rank Histogram
- Reduced Centered Random Variable (RCRV)
- Continuous Ranked Probability Scores (CRPS)
- Brier score & Entropy

Implementation of the metrics in the benchmarks:

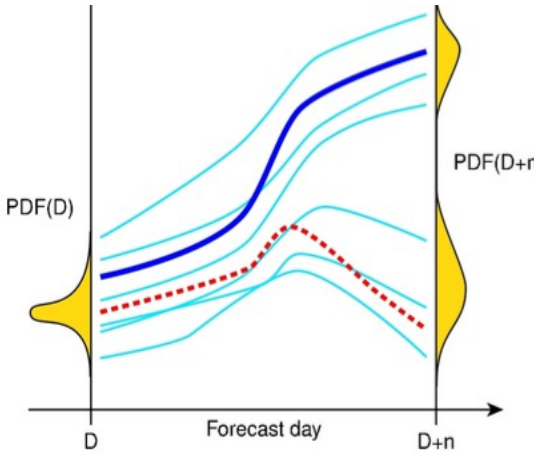
- manual to use/interpret the probabilistic metrics
- distribution of codes implementing the metrics
- included in SANGOMA toolbox (→ WP2)

TASK 4.2: Definition of metrics

Metrics are based on the following ideas:

Metrics consider the **probability distribution**
(as described by the ensemble),
not only the mean and standard deviation
(→ deal with non-Gaussian behaviours)

Probablistic evaluation includes
reliability (consistency with verification data), and
resolution (is the system informative?)



$$\text{CRPS} = E\left[\int_{\mathbb{R}} (F_p(\xi) - F_o(\xi))^2 \, d\xi\right] \quad \text{(global score)}$$

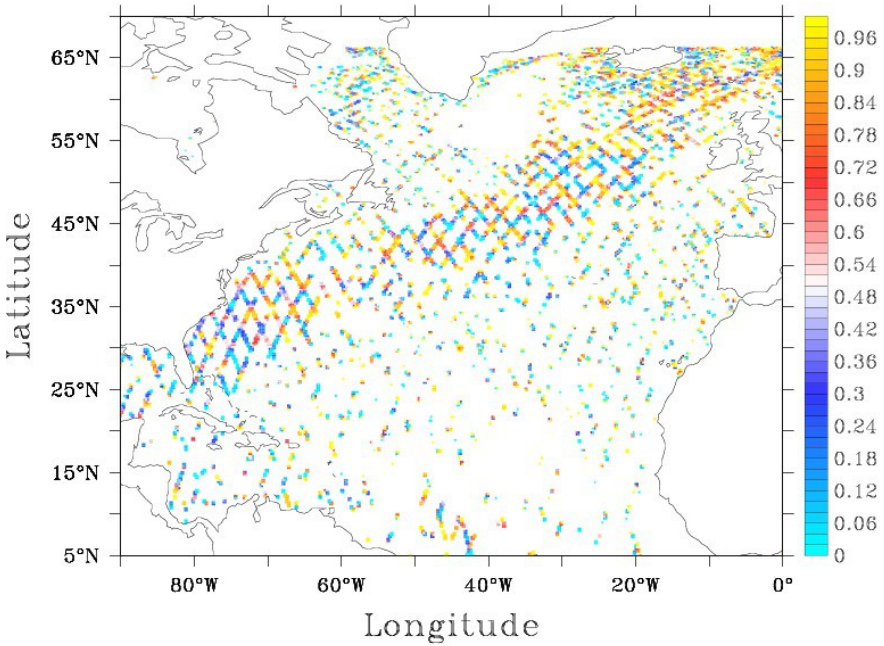
$$\textit{Reli} \quad \text{(reliability)} \quad \equiv y = \frac{o-m}{\sigma} \approx \text{rank histogram}$$

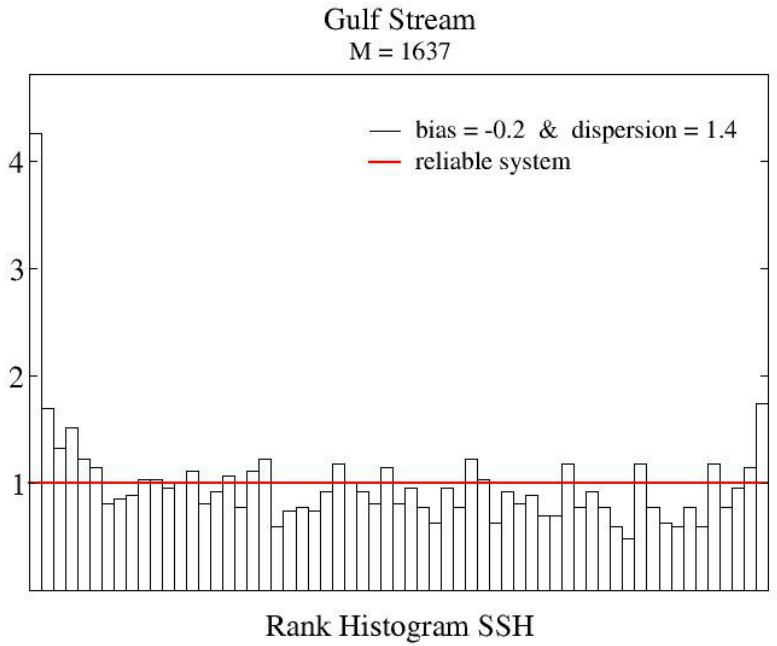
$$\text{bias} : b = E[y]$$

$$\text{dispersion} : d^2 = E[y^2] - b^2$$

$$\textit{CRPS}_{pot} \quad \text{(resolution)}$$

Example: rank histogram, with JASON-1 observations





Rank of JASON-1 altimetric observations in the ensemble simulation

Histogram of ranks in the Gulf Stream region

TASK 4.3/4.4: Running small and medium benchmarks

Many applications of small and medium benchmarks have been performed by all partners, with various existing and new assimilation schemes.

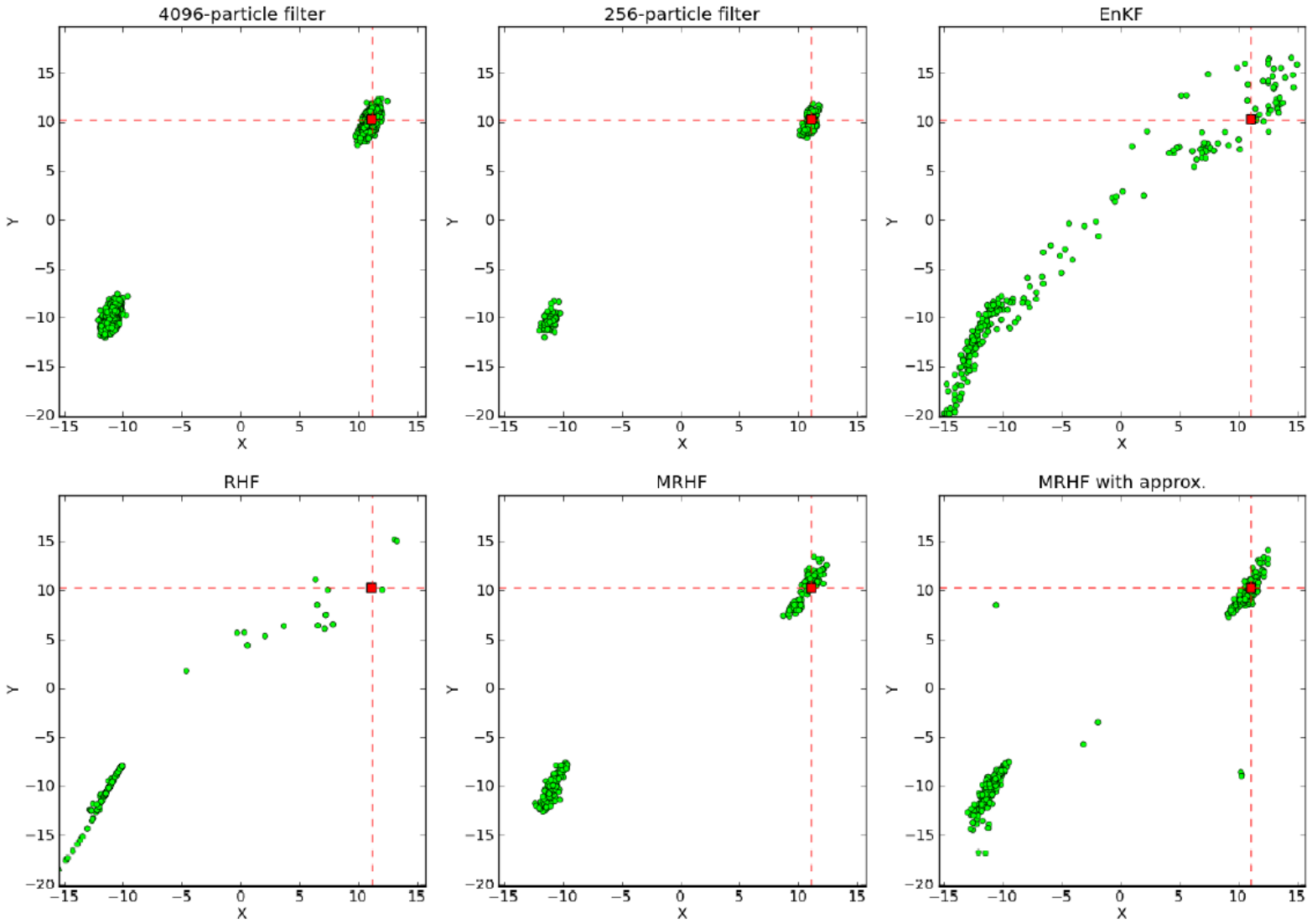
DL4.4 summarizing these results is being prepared.

Here are a few examples of what has been done:

- -Example 1: MRHF with small case benchmark
- -Example 2: 4DVAR with medium case benchmark

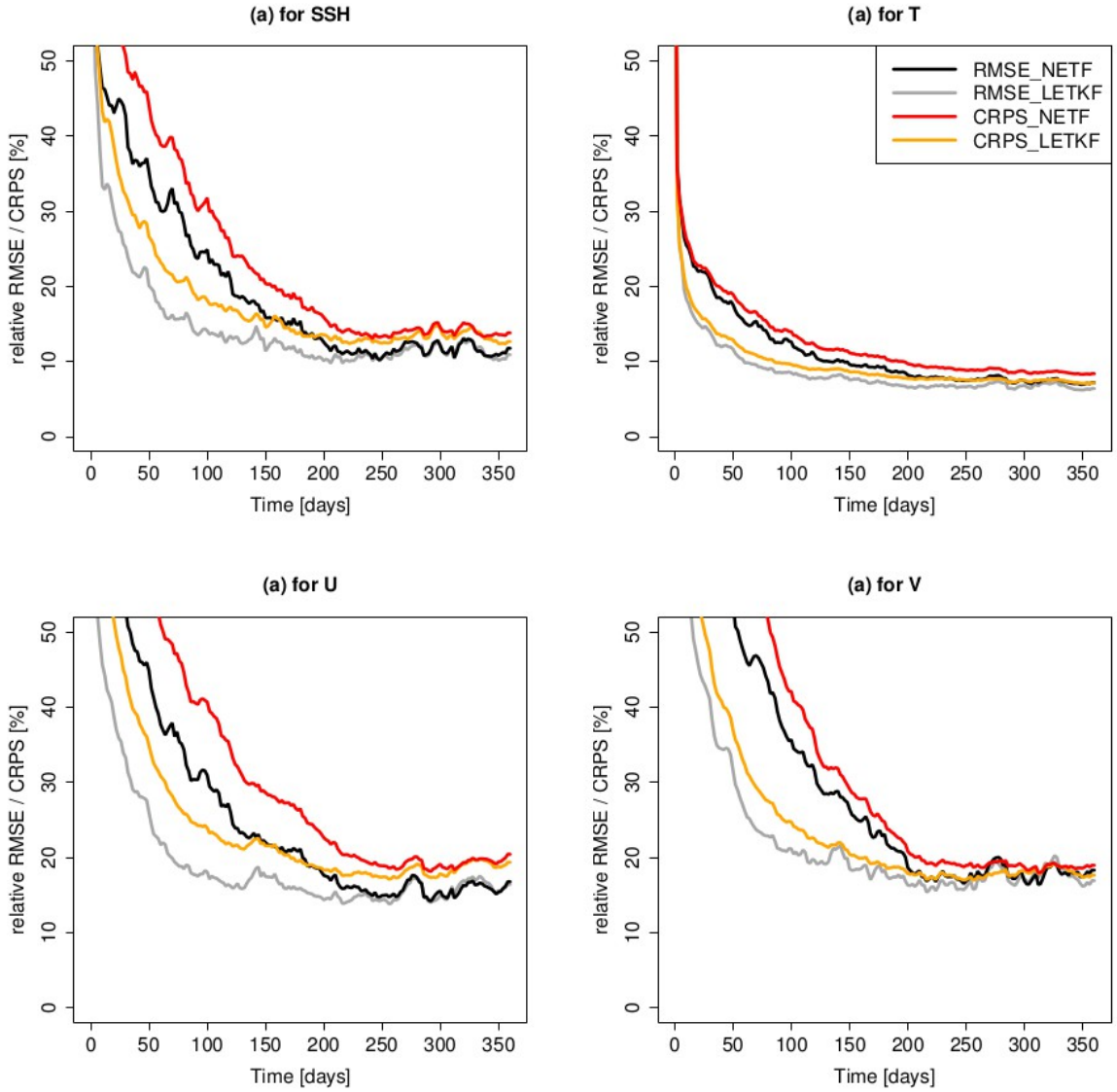
TASK 4.3/4.4: Running small and medium benchmarks

Example 1: Application of MRHF to **small benchmark**



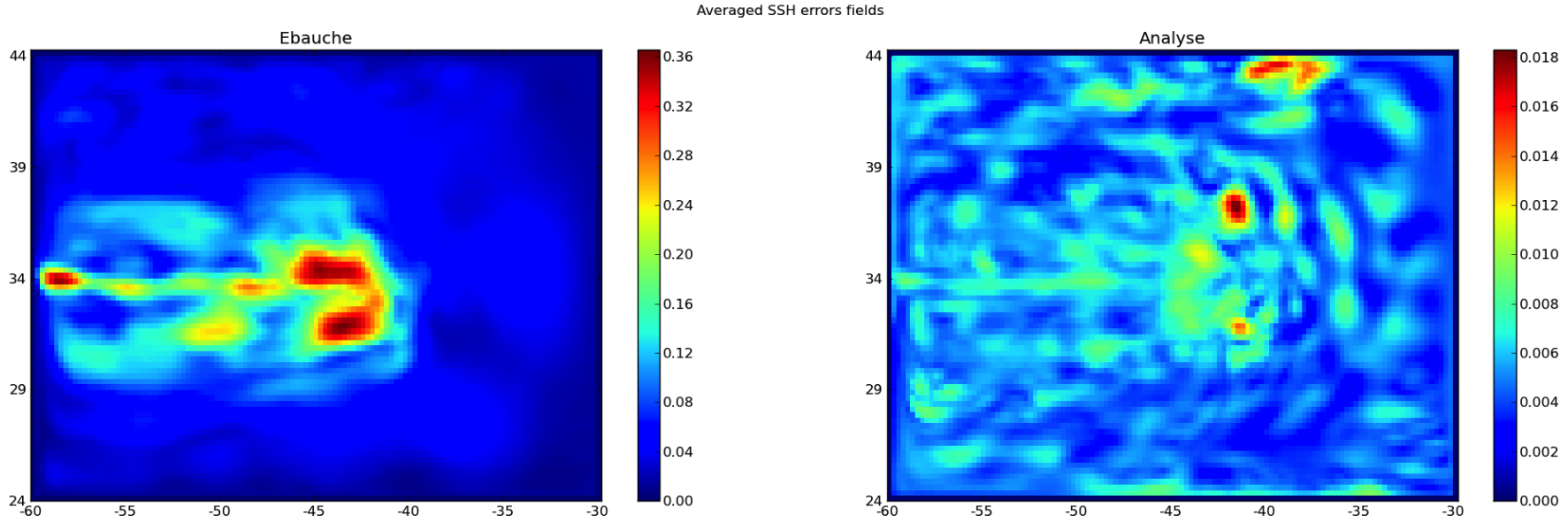
TASK 4.3/4.4: Running small and medium benchmarks

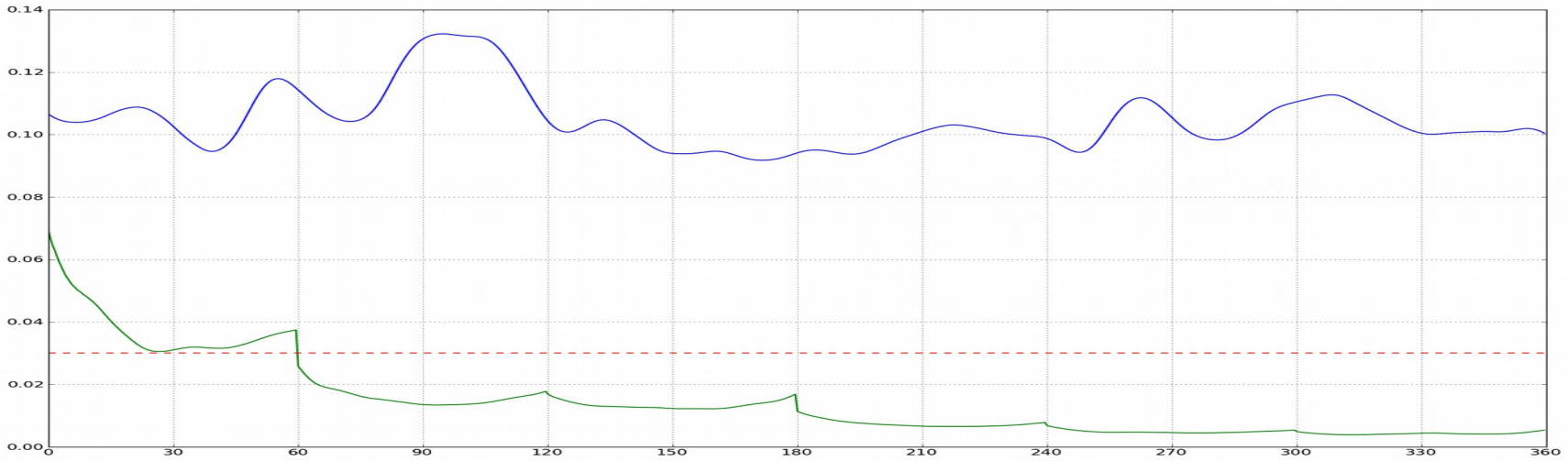
Example 2: Comparison LETKF/NETF in **medium benchmark**



TASK 4.3/4.4: Running small and medium benchmarks

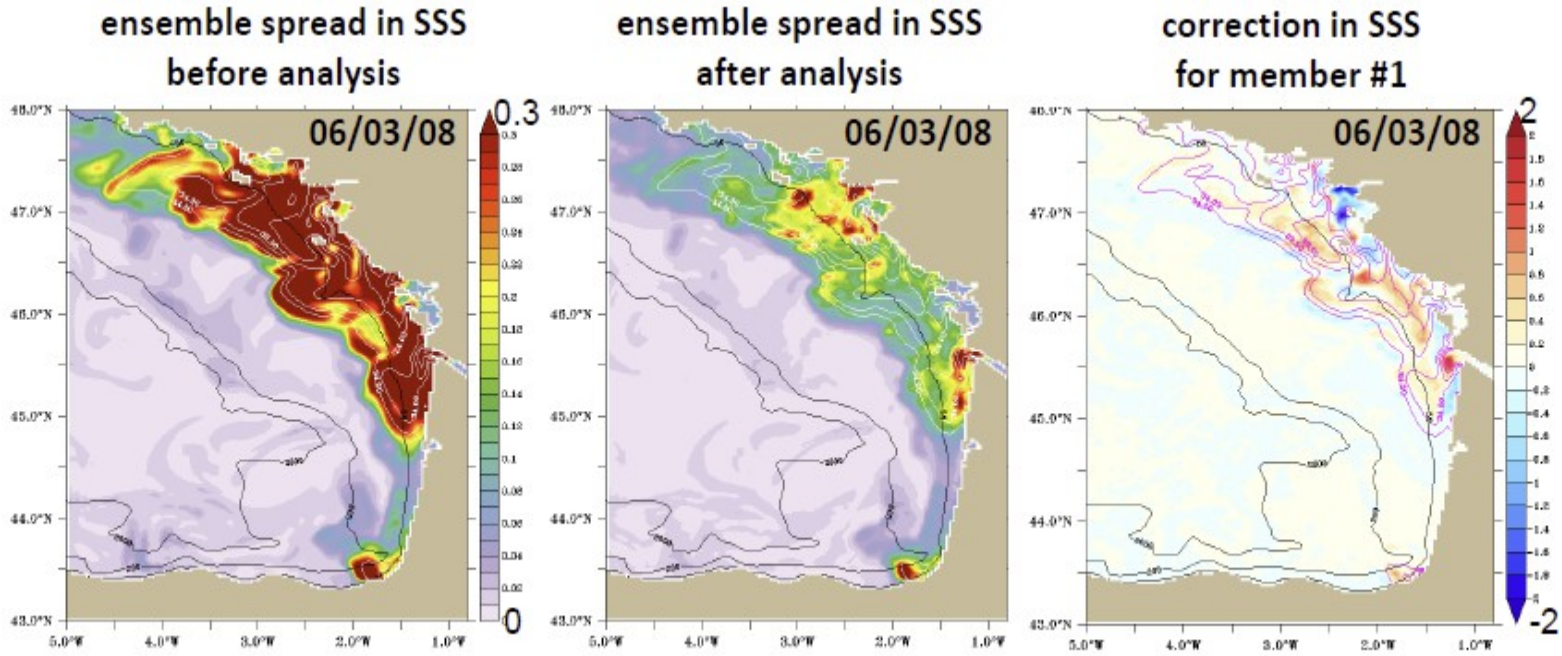
Example 3: Application of 4DVAR to **medium benchmark**





TASK 4.3/4.4: Running small and medium benchmarks

New coastal SANGOMA benchmark: Bay of Biscay shelf
Application of EnKF to coastal **benchmark** (CNRS/LEGOS)



(Ayoub et al., 2015)

- Twin experiments with EnKF assimilating SST
- Assimilation code: SDAP (one of SANGOMA DA codes)
  https://sourceforge.net/projects/sequoia-dap/
- SDAP is available for transferring to SANGOMA or external partners.

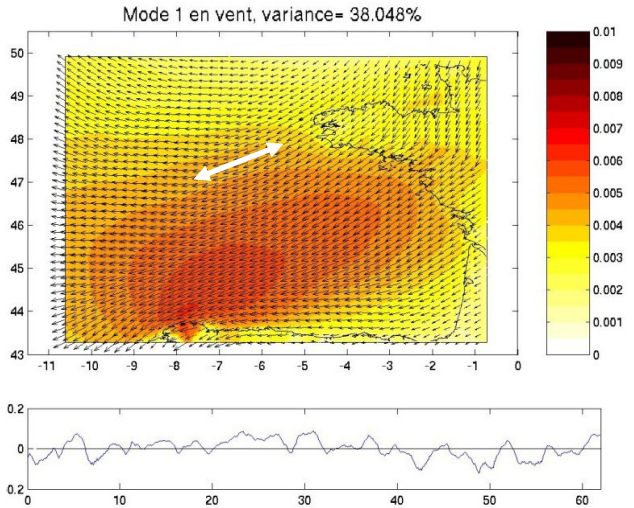
TASK 4.3/4.4: Running small and medium benchmarks

Coastal benchmark: Wind stress perturbations

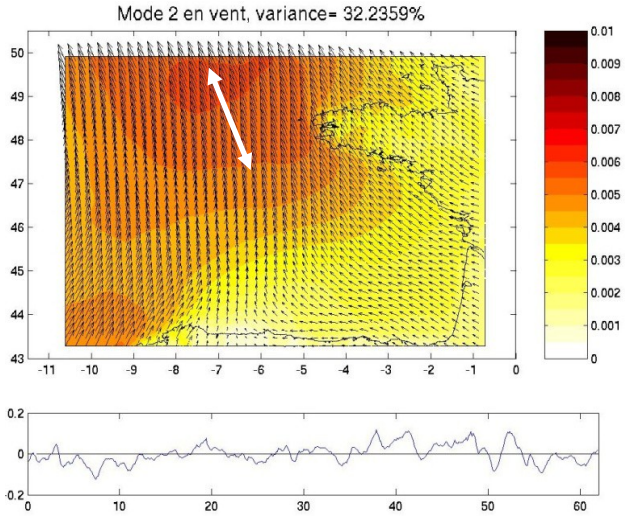
Ocean ensemble generation:

- Generate samples of surface atmospheric variables by randomly combining 10 bivariate ($\mathbf{U}_w$) variability EOFs (Auclair *et al.*, 2003)
- One set of Gaussian random coefficients every 5 days
- Integrate O(10)-O(100) ocean members, depending on case, providing samples of oceanic and atmospheric surface variables

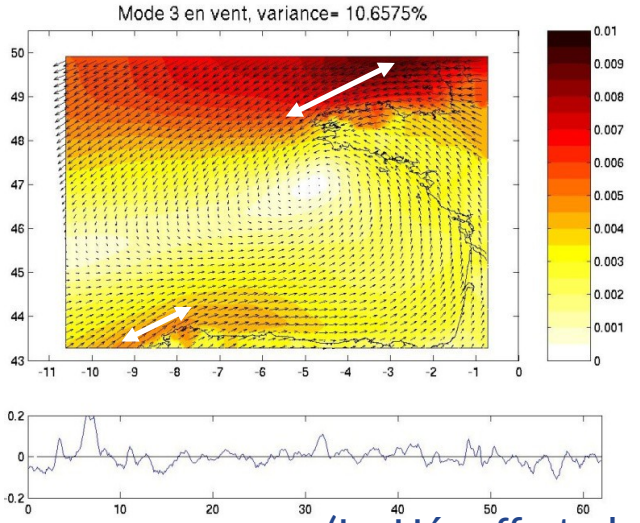
Mode 1: 38% of variance



Mode 2: 32% of variance



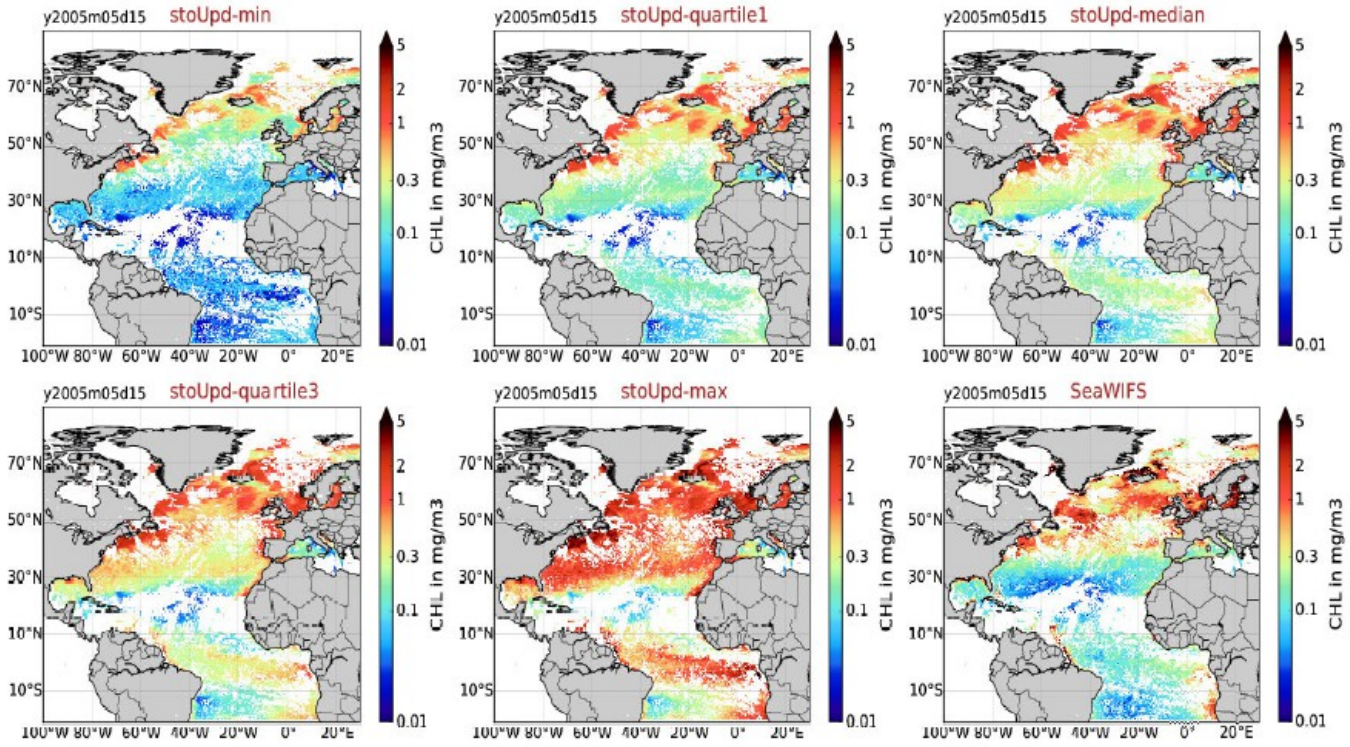
Mode 3: 11% of variance



(Le Hénaff et al., 2007)

TASK 4.5: Diagnostic of non-Gaussian behaviours In large case benchmark

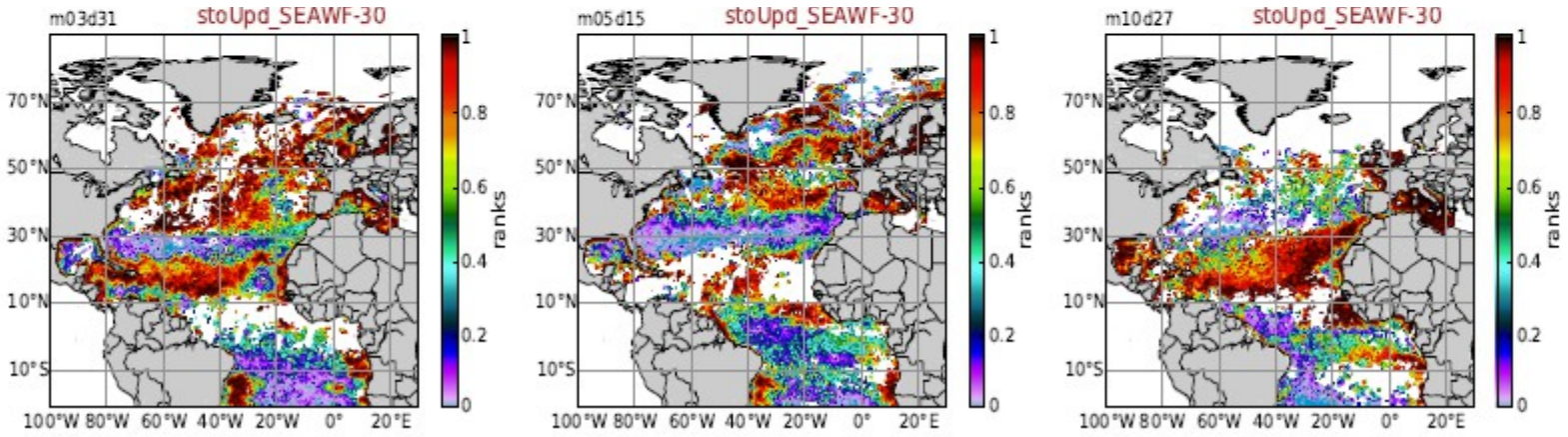
This task has been fulfilled with the large case benchmark, in coupled mode, with the **PISCES ecosystem model:**



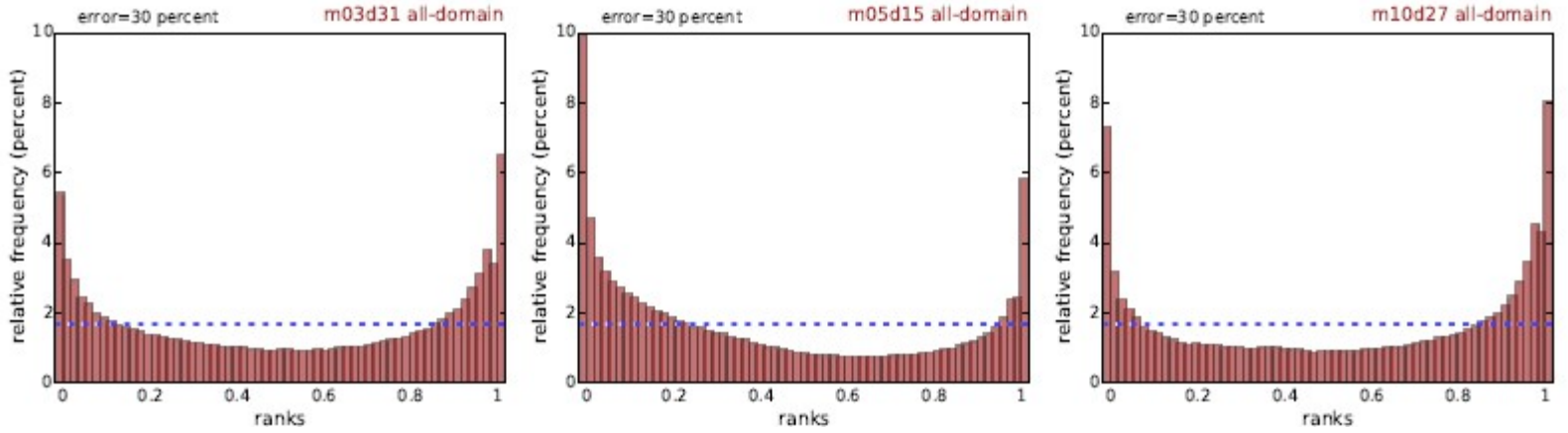
Ensemble simulation (60 members) with explicit simulation of ecosystem uncertainties (Garnier et al., J. Mar. Syst., 2015)

TASK 4.5: Diagnostic of non-Gaussian behaviours In large case benchmark

Rank of SeaWifs ocean observations in the ensemble:

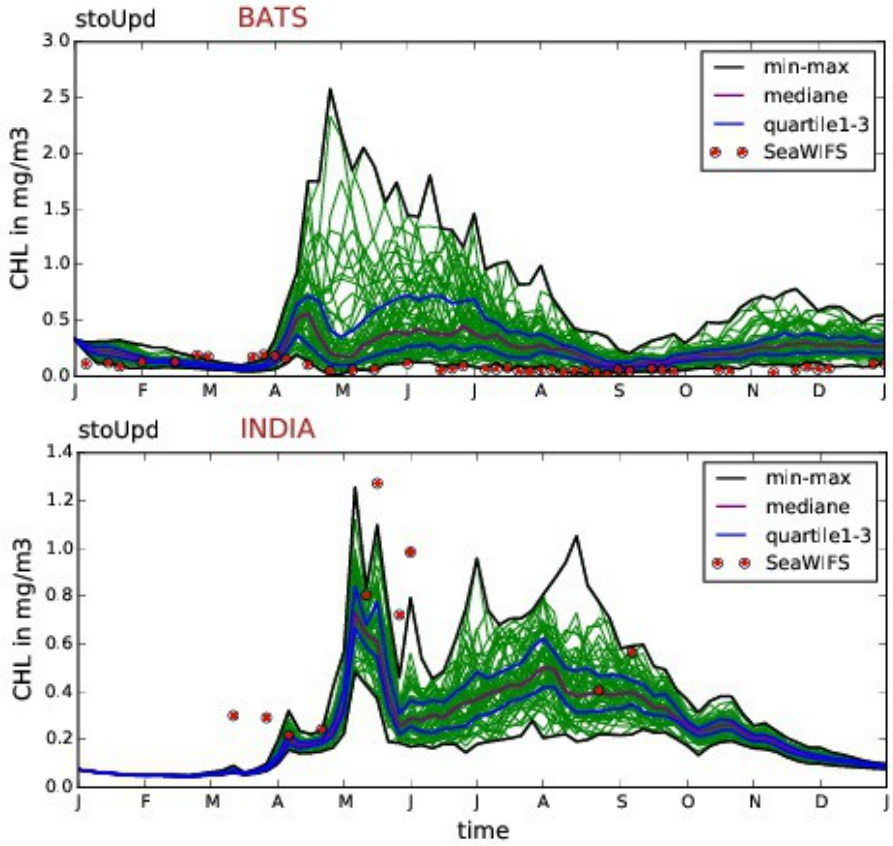


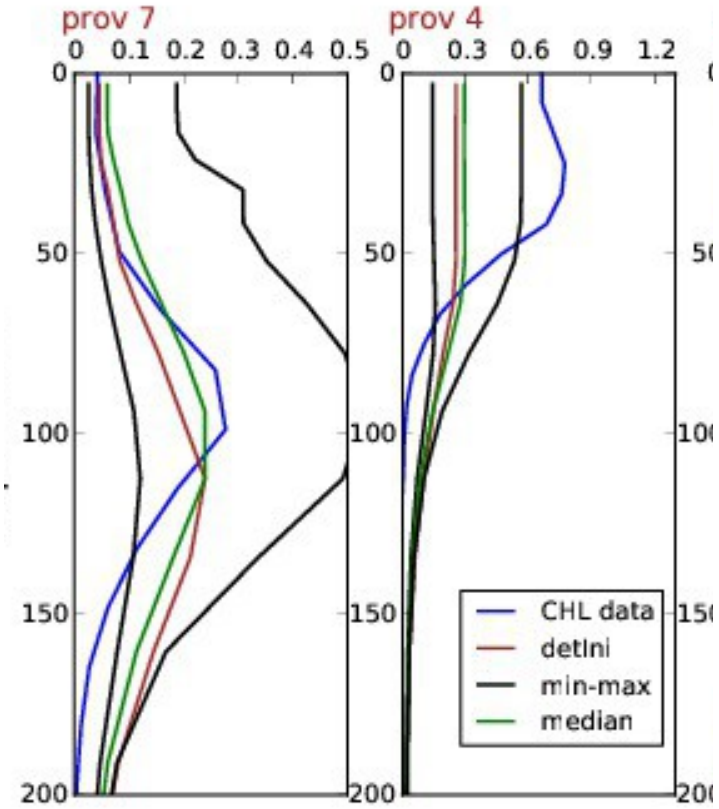
Rank histogram to check ensemble reliability:



TASK 4.5: Diagnostic of non-Gaussian behaviours In large case benchmark

The ensemble displays important non-Gaussian behaviours:





Local anamorphosis transformations have been applied to perform ensemble observational update using ocean colour observations (Garnier, PhD thesis, 2015)

TASK 4.6: Running large case benchmark

The large case benchmark has been run by 2 partners, with 2 different model perturbation strategies:

Partner GHER (Y. Yan):

- add realistic noise in the atmospheric forcing
(wind, air temperature, long and short wave radiation flux)
- growing perturbation during 6 months (1/1 → 29/6/2005)

→ *Yan et al., J. of Geophys. Res. 120, 5134-5157, 2015.*

Partner CNRS/LGGE (G. Candille):

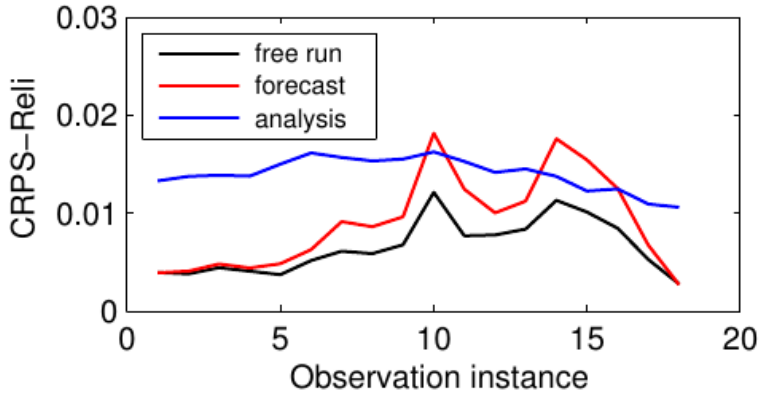
- simulate the effect of unresolved scales in the seawater equation of state
- growing perturbation during 6 months (1/1 → 29/6/2005)

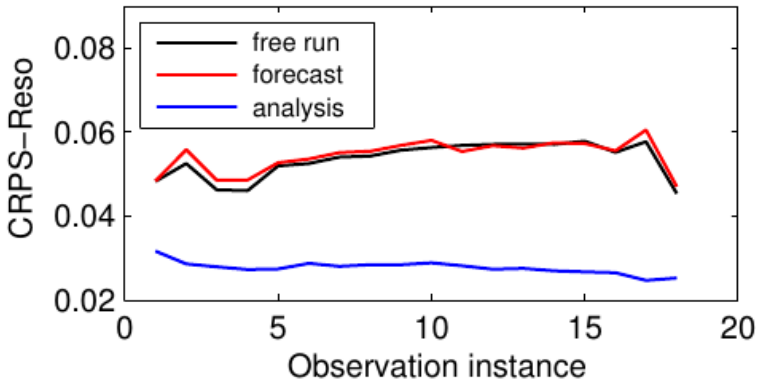
→ *Candille et al., Ocean Science, 11, 425-438, 2015.*

TASK 4.6: Running large case benchmark

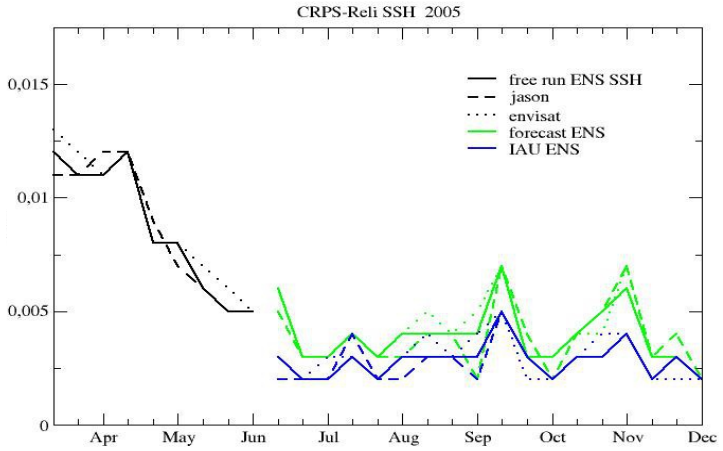
Probablistic metrics discussed in **DL4.5** (Mo 48):

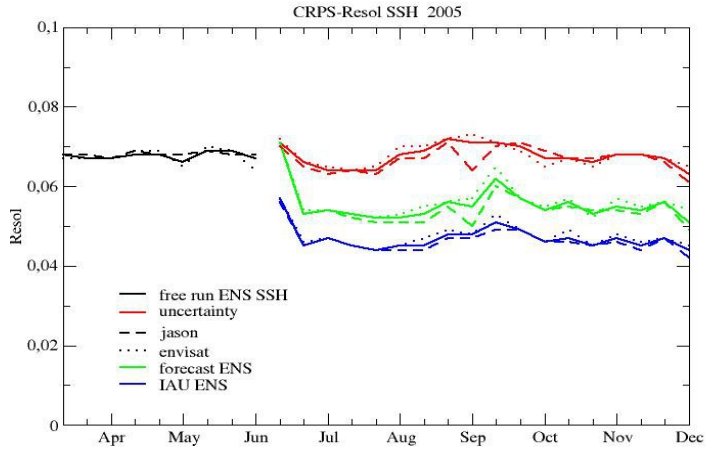
- Partner GHER (Y. Yan), CRPS score for SSH:





- Partner CNRS/LGGE (G. Candille), CRPS score for SSH:





RECOMMENDATIONS

What do we advise Copernicus operational centers?

Make progress in the explicit simulation of model uncertainties using a stochastic approach

Progressively move to a probabilistic description of the operational products (using ensembles)

Generalize the use of probabilistic metrics to evaluate the quality of the products and their impacts on end-user applications