Uncertainties in Ocean Forecasting Systems: non-Gaussian and multiscale issues

Jean-Michel Brankart

Institut des Geosciences de l'Environnement
Equipe de Modélisation des Ecoulements Océaniques Multi-échelles

Marine Data Assimilation Workshop
Toulouse, 5-6 September 2018

Since important decisions must rely on simulations, it is essential that its validity be tested, and that its advocates be able to describe the level of authentic representation which they achieved.

Summer Computer Simulation Conference (1975), cited by Richard Hamming (1997)

Outline

1

Introduction

The 3 rules of system engineering

(from Richard Hamming, 1997: The Art of Doing Science and Engineering)

Rule 1: If you optimize the components you will probably ruin the system performance.

Acknowledge that the components of the system are imperfect.
The emerging performance of the system is what matters.
Do not look for the best model, make it suitable to the system.

Rule 2 : Part of systems engineering design is to prepare for changes so that they can be gracefully made and still not degrade the other parts.

Do not think that you will find the final solution to the problem.
Be prepared to involve more complex components:
unexpected model dynamics or observation operators.

Rule 3: The closer you meet the specifications the worse the performance when overloaded

Do not build a system that is specifically and optimally designed for the present model and present observations.
Do not optimize your response to end-users' requests, but prepare to meet their future needs.

Uncertainty, as a key feature of our systems

What are the uncertain components of our systems ?

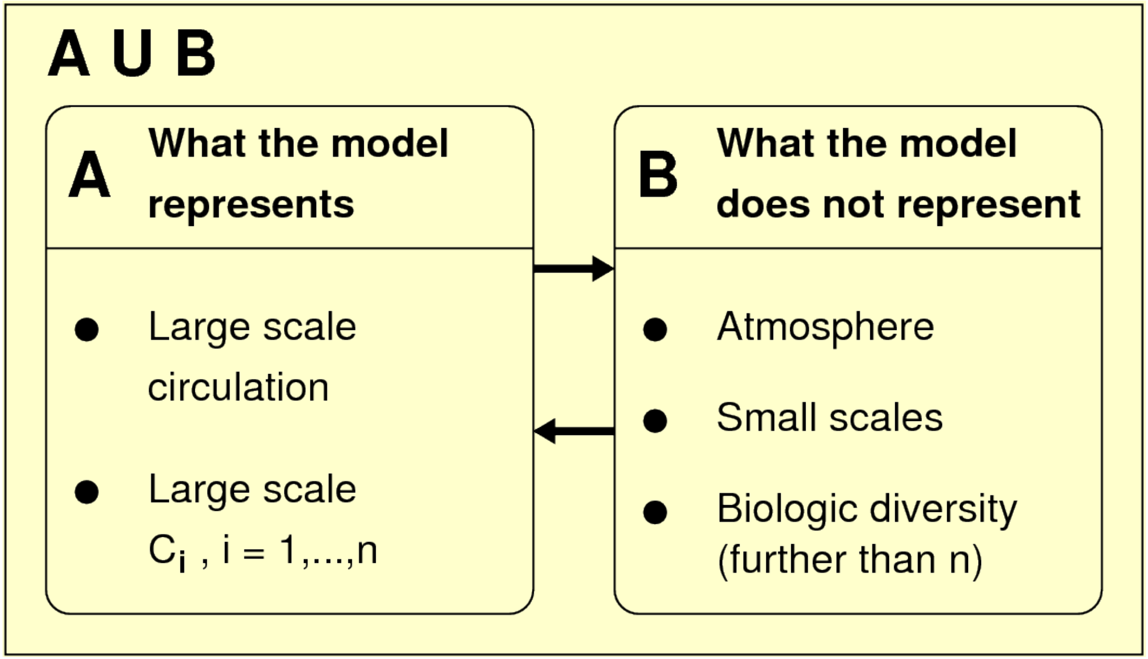
How to describe uncertainties ?

How does it participate to the solution of assimilation problems ?

2

Uncertainties in the system

Sources of uncertainties in ocean models



- Even if the dynamics of **U** can be assumed deterministic,
  the system **A** alone **cannot be assumed deterministic**.

- To obtain a deterministic model for **A**, one must assumed, either
  that **B** is known (→ atmospheric forcing), or
  that the effect of **B** can be parameterized (→ paramétrisation of unresolved scales or unresolved biologic diversity).

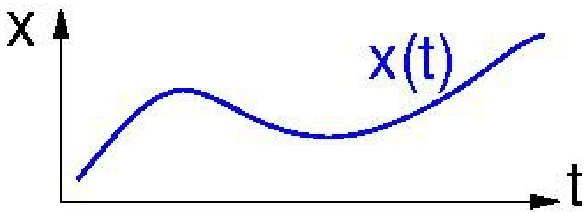
→ **B is always an essential source of uncertainty in the model.**

From deterministic to probabilistic ocean simulations

Stochastic ocean dynamics, with explicit simulation of uncertainties

$$d\mathbf{x} = \mathcal{M}(\mathbf{x},t)dt + \Sigma(\mathbf{x},t)d\mathbf{W}_t$$

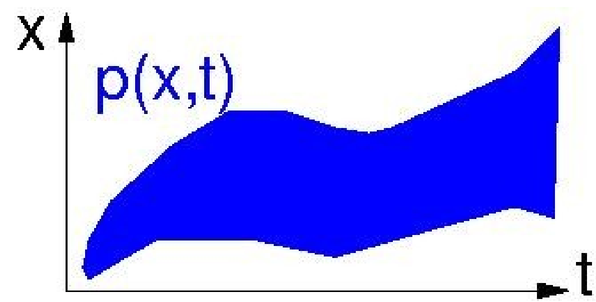
where $\mathbf{x} = [x_1, \ldots, x_N]$



Fokker-Planck equation, for the probability distribution p(**x**,t),

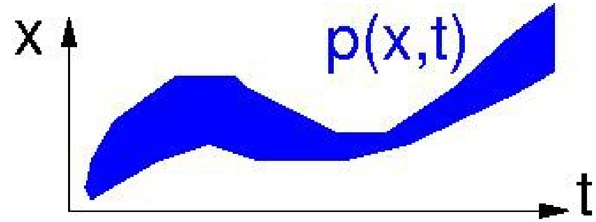
$$\frac{\partial p(\mathbf{x},t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left[\mathcal{M}_i(\mathbf{x},t)p(\mathbf{x},t)\right] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j}\left[D_{ij}(\mathbf{x},t)p(\mathbf{x},t)\right]$$

where $\mathbf{D} = \mathbf{\Sigma}\mathbf{\Sigma}^T$



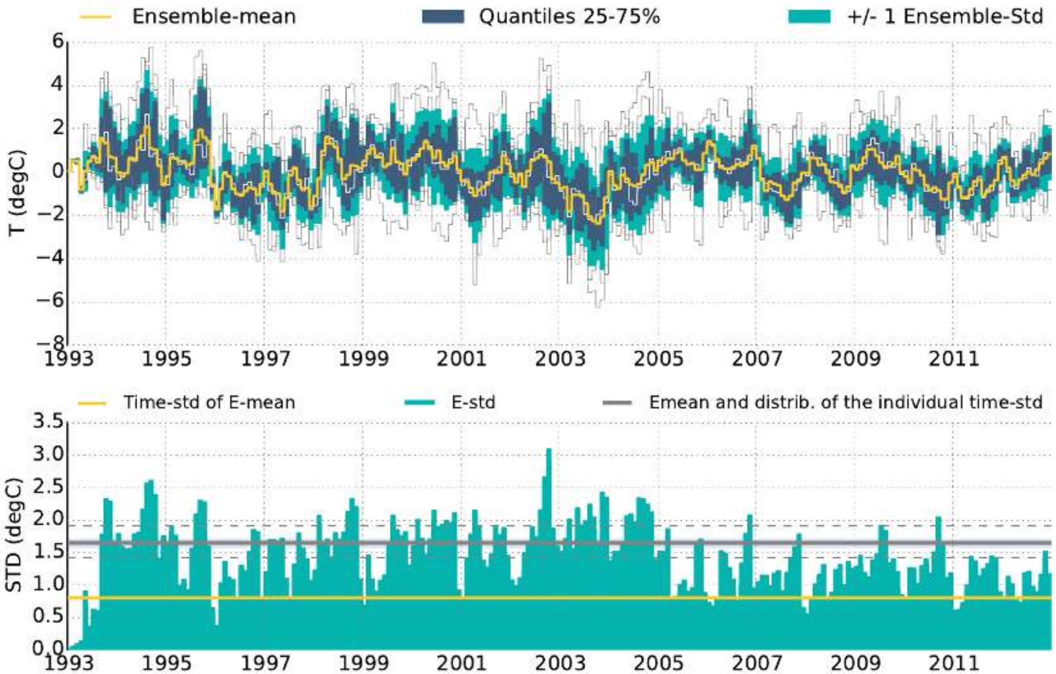
Conditioned to observations, to reduce uncertainties

by an appropriate data assimilation method



Example 1: intrinsic ocean variability (OCCIPUT project)

A 50-year and 50-member ensemble simulation
with a global 1/4° ocean model (NEMO/ORCA025)



Monthly mean temperature at 93 m depth
(in the Gulf Stream region)

The ensemble spread
only results
from small perturbations
in the initial conditions

→ study the chaotic
behaviour of the system

All members see the
same atmosphere

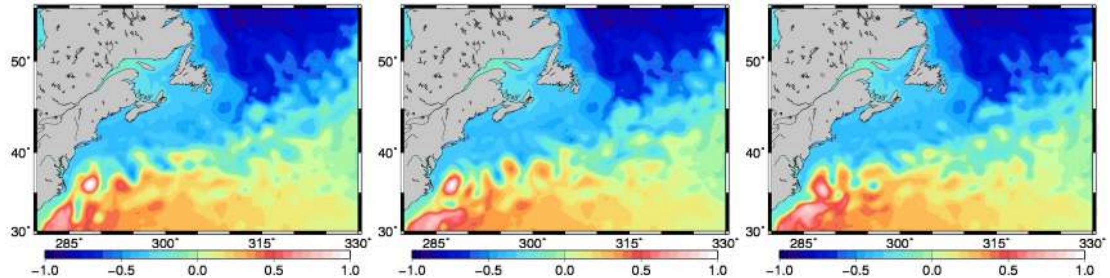
→ identify the variability
that is intrinsic
to the chaotic ocean

No data assimilation is involved in this study.
A probabilistic approach is needed to understand ocean dynamics.

From Bessières et al., 2016. Dévelopment of a probabilistic system with NEMO at eddying resolution, Geoscientific Model Developments.

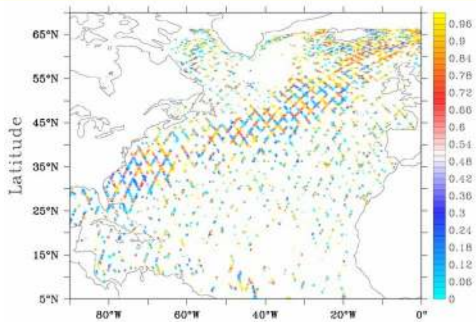
Example 2: assimilation of altimetry (SANGOMA project)

96-member ensemble with NATL025, with explicit simulation of uncertainties in the EOS

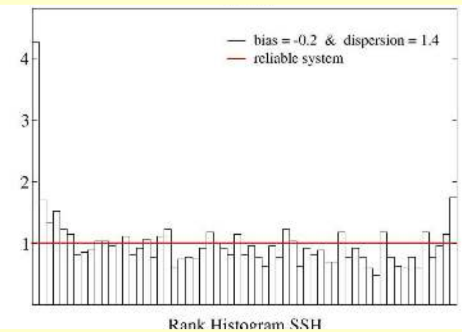


→ **comparison to observation (rank histogram, CRPS score, etc.)**

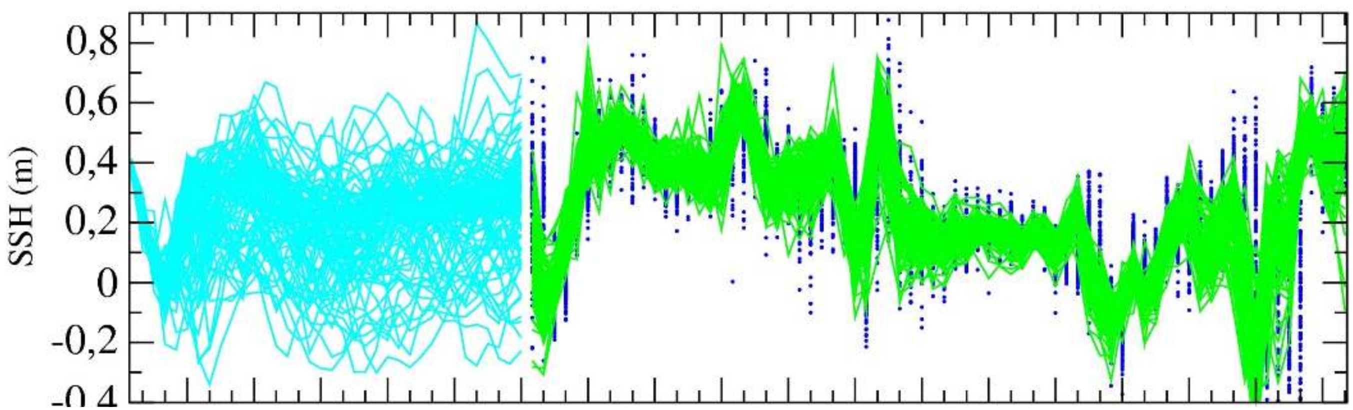
<u>Map of ranks</u> of JASON altimetric data



<u>Rank histogram</u> to check relability



→ **assimilation of altimetric data (from Candille et al., 2015)**



without assimilation	with assimilation of Jason1 and Envisat

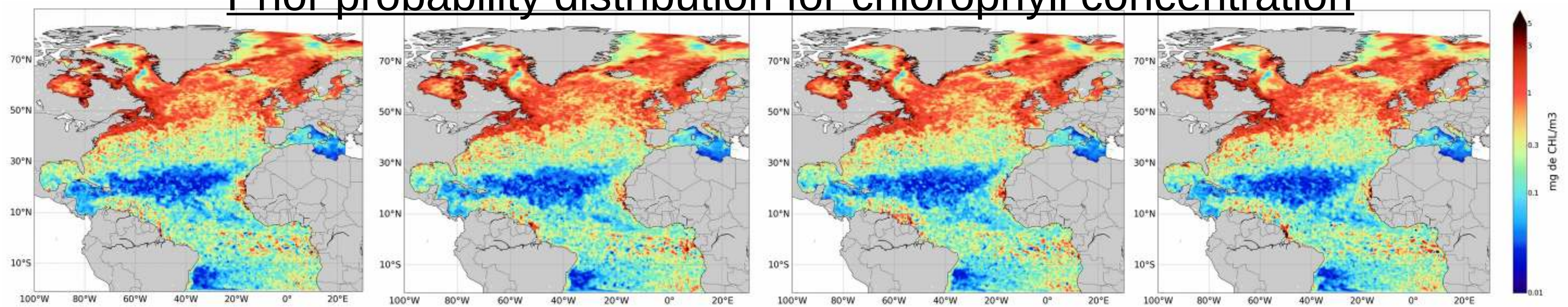
<u>Time evolution of the pdf</u>

from June 2005 to December 2006

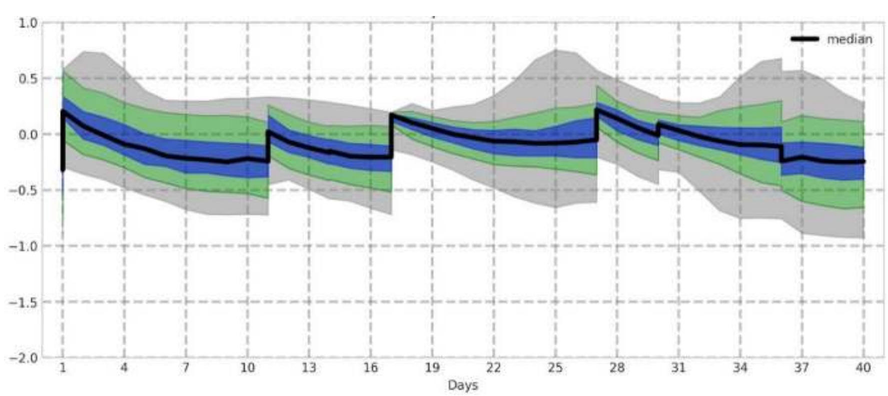
Example 3: assimilation of ocean colour observations (1)

<u>50-member ensemble simulation with NATL025/PISCES</u>
with explicit simulation of uncertainties
(parameters and unresolved scales)

<u>Prior probability distribution for chlorophyll concentration</u>



member 1 | member 2 | member 3 | member 4

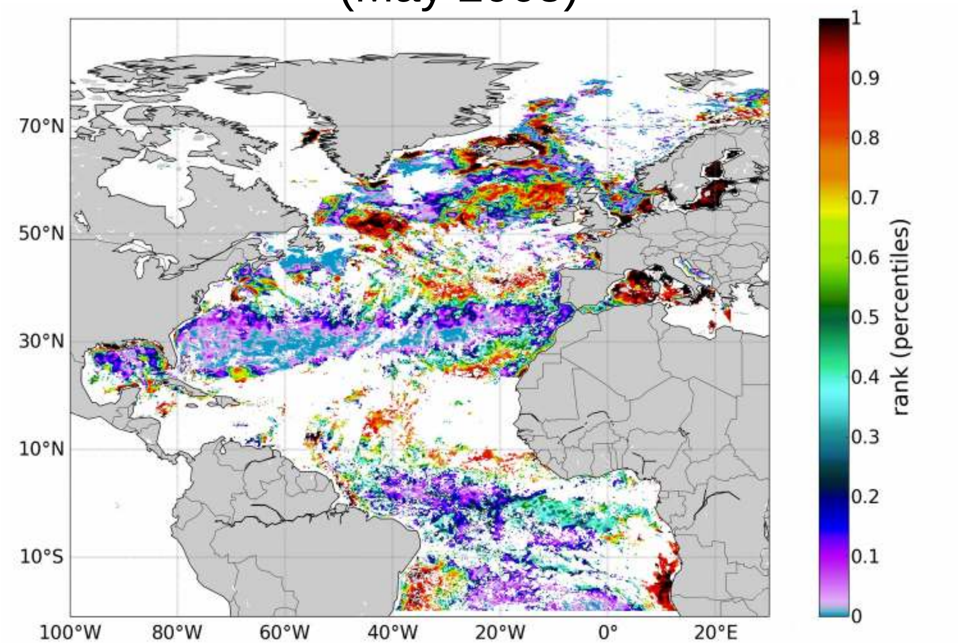


<u>Time evolution of the pdf for phytoplankton</u>

From Garnier et al. 2016 (stochastic parametrization and probabilistic comparison to observations) and Santana Falcon et al. 2018, in prep. (assimilation of ocean colour)

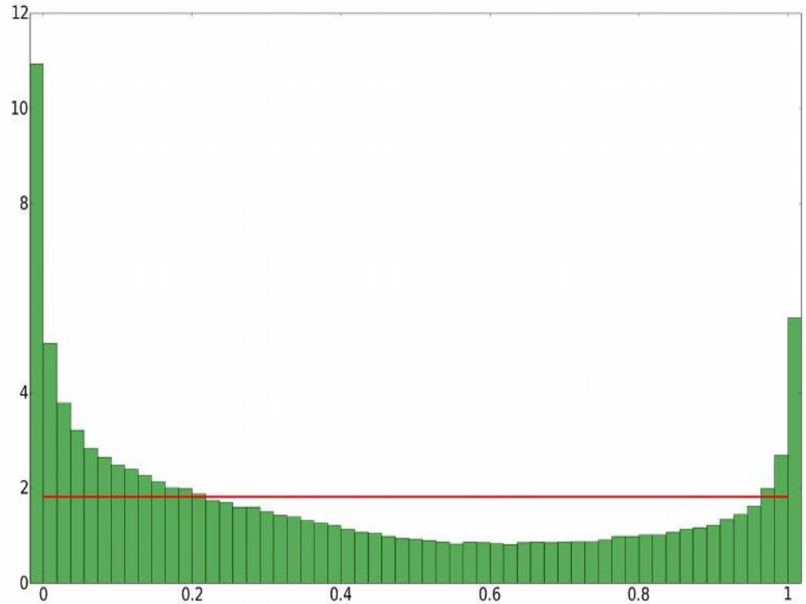
Example 3: Comparison to ocean colour observations (2)

<u>Rank of SeaWifs observations</u>
in the ensemble simulation
(May 2005)



The ensemble spread is already sufficient to include more than 80% of the observations (accounting for a 30% observation error)

<u>Rank histogram for SeaWifs</u>
over the whole domain



The ensemble is not far from being reliable, even if still underdispersive (too many observations in the external ranges of the ensemble)

→ objectively **test the consistency of simulations** as compared to observations
→ prerequisite to ocean colour **data assimilation**

Uncertainties are inherent to ocean models

This is not just a trick to generate more spread and make ensemble data assimilation work

As atmospheric models, ocean models will become stochastic to cope with dynamical uncertainties

Objective comparison to observations (reliability, resolution) and verification of the products

How to make stochastic parameterizations more and more consistent with real uncertainties ?

References: Buizza et al. 1999 (first ECMWF implementation), Berloff 2005, Shutts 2005, Wilks 2005, Palmer al. 2005, 2009, 2014 (atmospheric developments), Juricke et al. 2013 (sea ice), Brankart 2013 (EOS), Brankart et al. 2015 (NEMO implementation)

3

Non-Gaussian issues

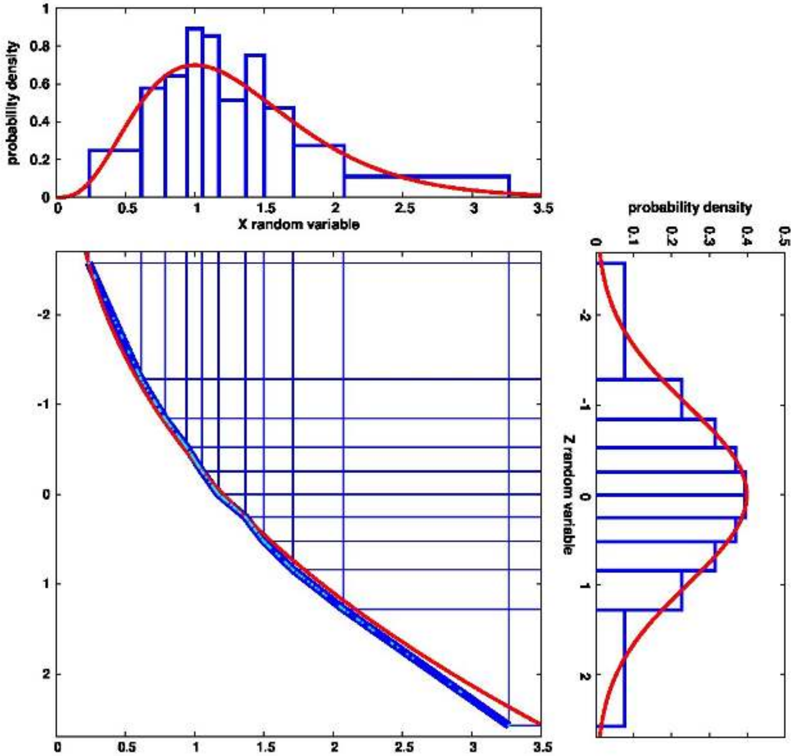
Anamorphosis transformation

A nonlinear change of variables to deal with the non-Gaussianity of the marginal distribution

Transform X into Z:
$Z=G^{-1}[F(X)]$, $X=F^{-1}[G(Z)]$
where
F is the cdf of X
G is the cdf of N(0,1)

Many ways of estimating the transformation function

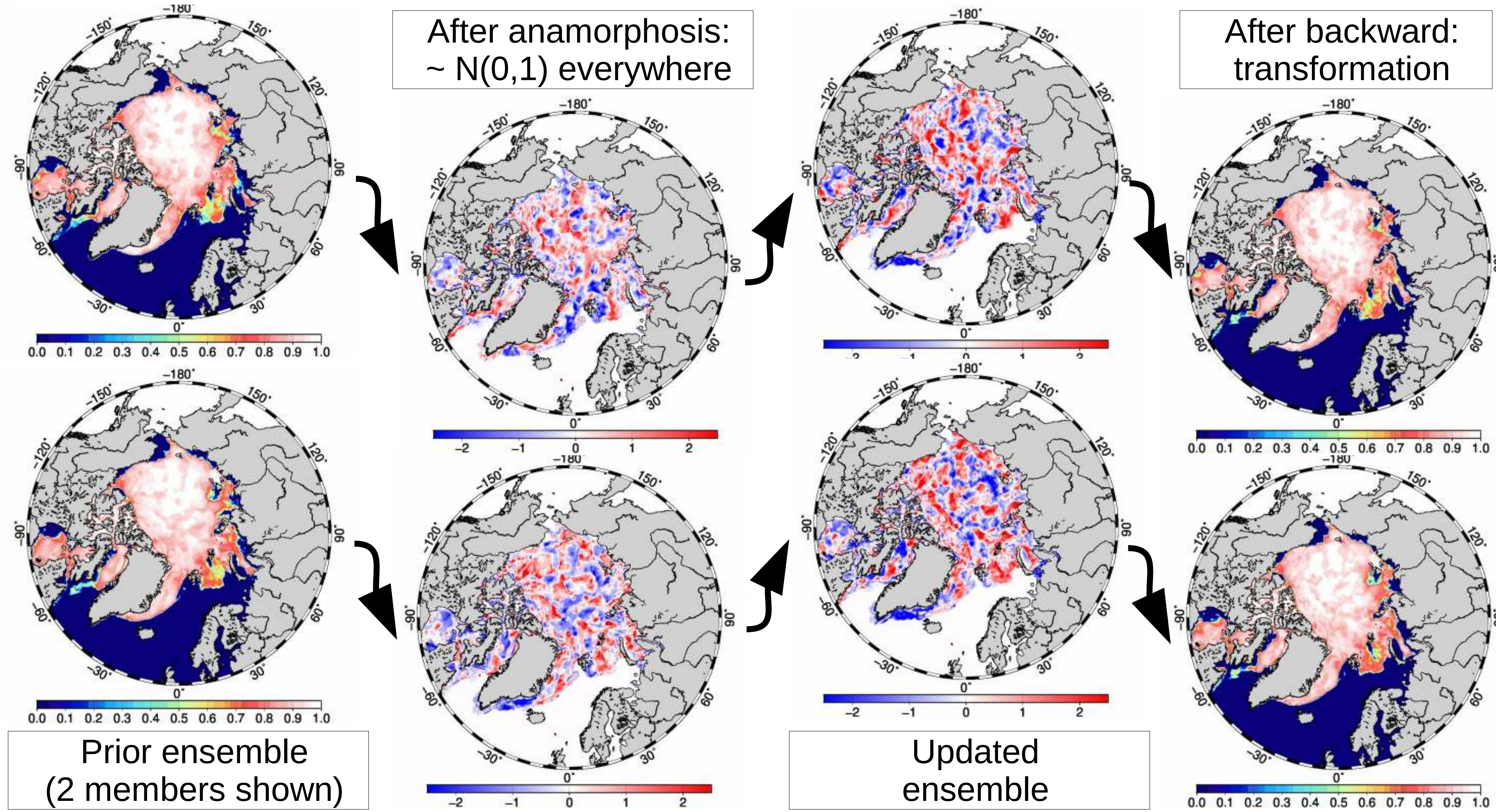
<u>Simple piecewise linear algorithm:</u>
compute ensemble quantiles,
linearly interpolate bewteen
the corresponding quantiles of N(0,1)



<u>References</u>: Wackernagel 2003 (multivariate geostatistics),
Bertino et al. 2003, Simon et al. 2009 (first ocean application of the method),
Béal et al. 2010, Brankart et al. 2012 (simplified piecewise linear algorithm)

Example 1: in the white ocean

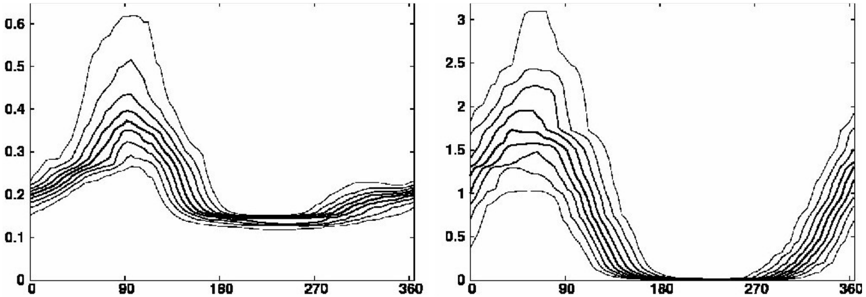
Application to sea ice concentration in the CREG4 model configuration



The prior ensemble becomes marginally Gaussian (with mean=0 and std=1).
Updated variables are kept inside their bounds (here between 0 and 1).

Example 2: in the green ocean

Application to NATL025/PISCES biogeochemical fields



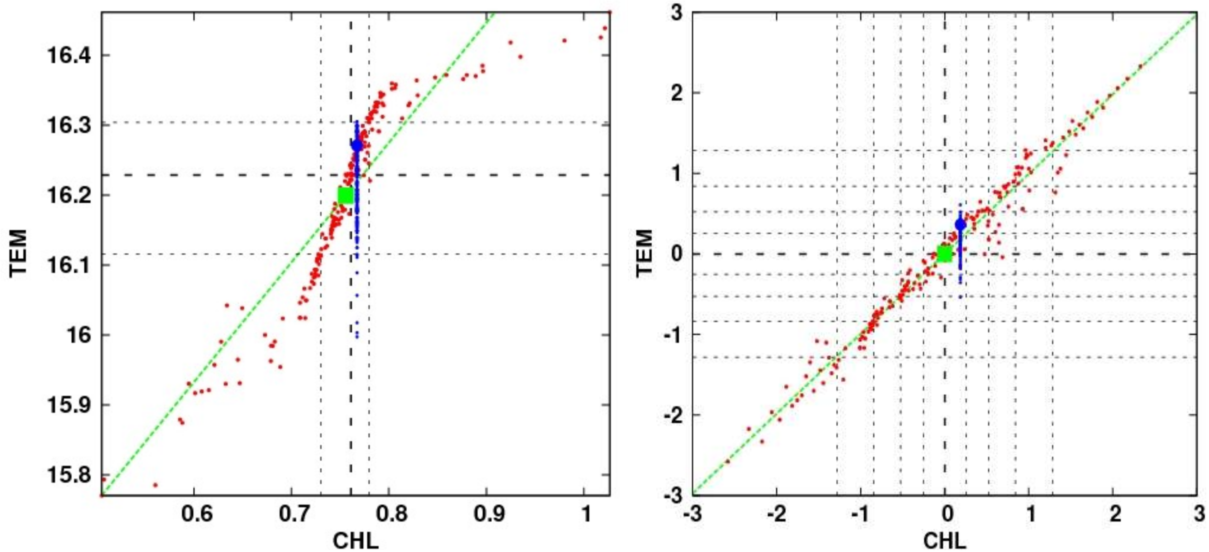
Ensemble deciles for **phytoplankton** and **nitrate** over one year

Marginal distributions can be strongly non-Gaussian

There can be large probability densities close to the bound

Transformation of joint distribution for chlorophyle and temperature

Updated ensemble (in blue) assuming a perfect observation of chlorophyle



Anamorphosis improves the linear dependency between the 2 variables.
The spread of the updated ensemble is thus smaller.
It is like using a nonparametric measure of correlation.

Non-Gaussian behaviours are ubiquitous in the ocean

The anamorphosis transformation is cheap (computation of quantiles of the ensemble)

A good way to deal with bounded variables (→ biogeochemical and sea ice variables)

Often improves linear correlations

Non-Gaussianity of joint distributions is still much more expensive to account for (particle filters, MCMC algorithms,...)

4

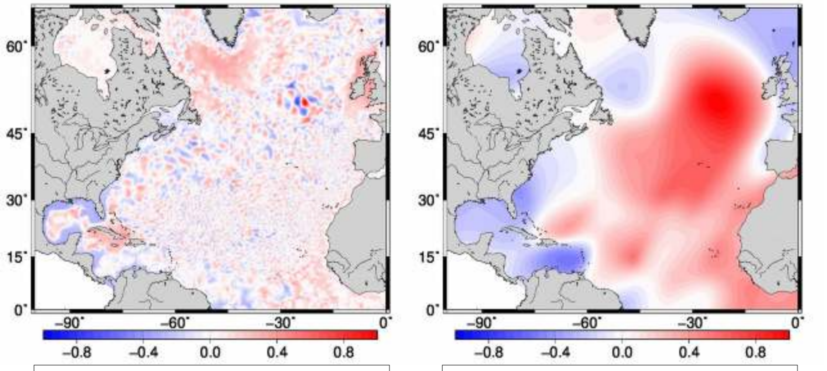
Multiscale issues

More and more degrees of freedom in the system

Covariance localization is needed
to make the probem tractable

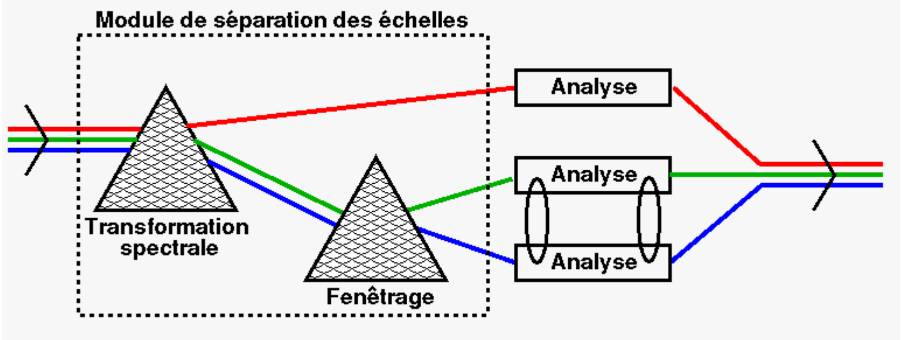
The long-range correlation structure
for the large-scale signal
is not used by the system

<u>Ensemble correlation structure</u>



All scales | Large scales

→ **How can we perform a seamless observational update from the largest scales to the finest scales ?**



Apply transformation operators
to separate scales

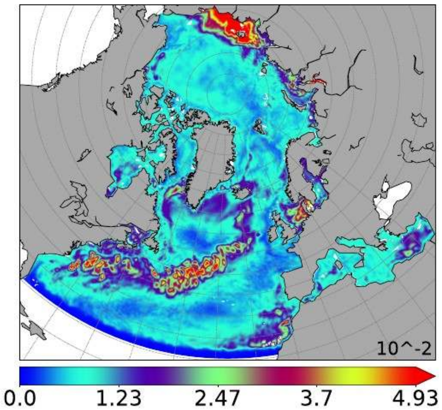
Localization is applied
to trandormed variables

→ **Direct control of the large scales (usually well observed)**

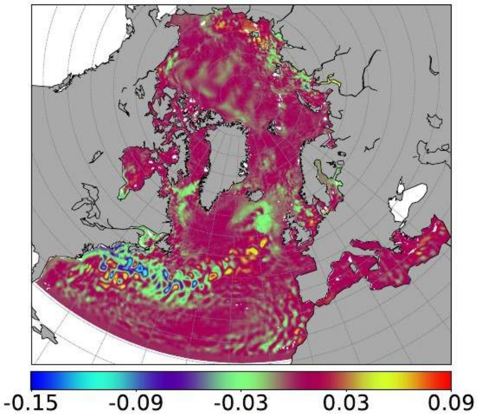
<u>References</u>: Zhou et al. 2008, Zhang et al. 2009 (spatial localization, using several scales),
Buehner 2012 (spatial/spectral localization, using wavelet transform),
Tissier et al. 2018, in prep. (spectral localization, using spherical harmonics).

Example 1: in the blue ocean (1)

Twin experiments with the CREG4 model configuration



Ensemble std

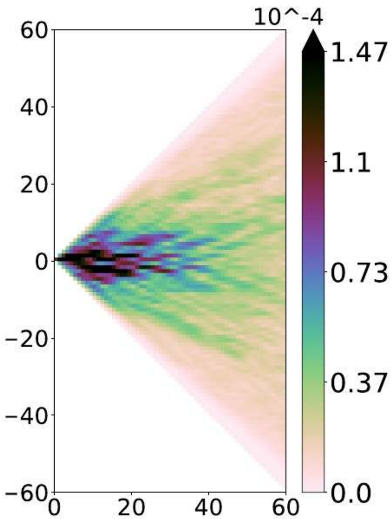


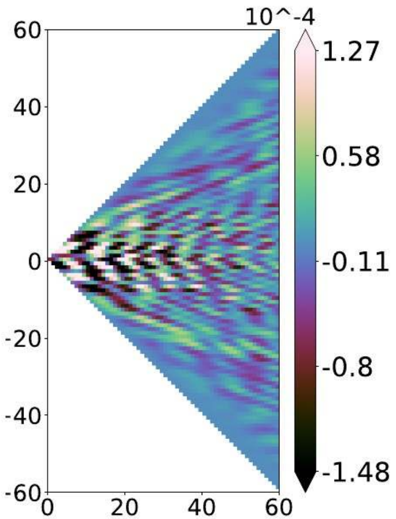
True anomaly

Transformation

Transformation

by projection of anomalies
on the spherical harmonics

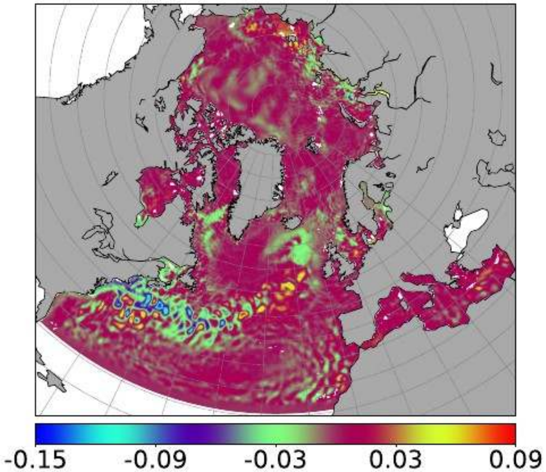




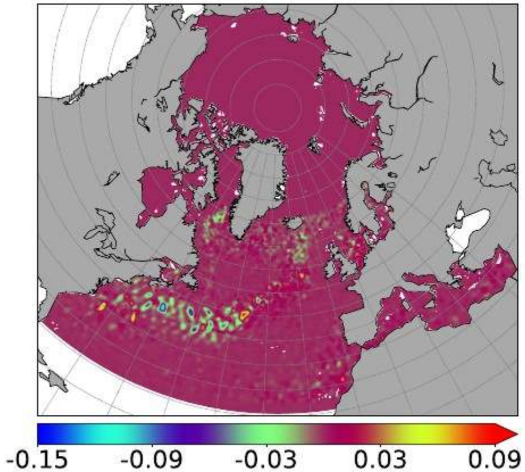
From Tissier et al., 2016. Mutliscale ensemble observational update combining spatial and spectral localization, Ocean Science, in preparation.

Example 1: in the blue ocean (2)

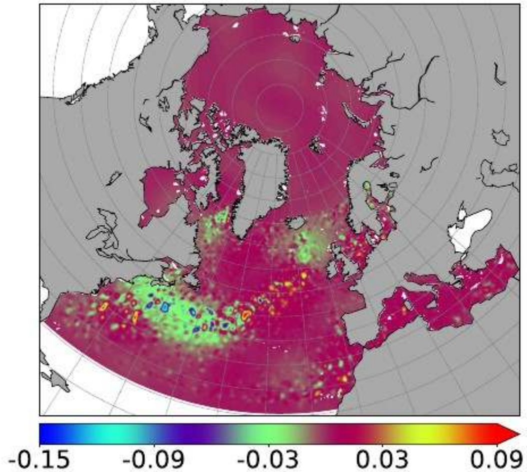
Result of the observational update



True anomaly

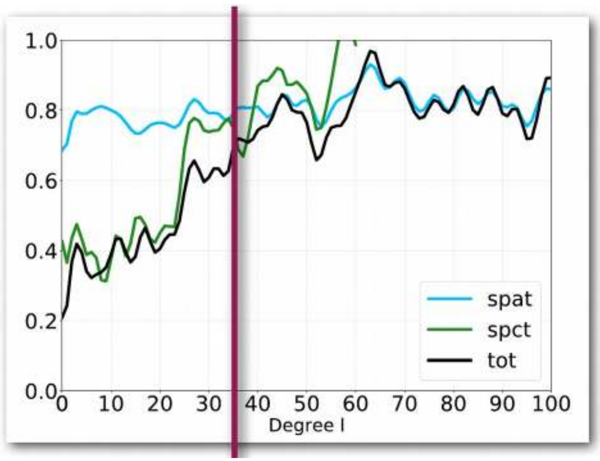


Spatial localization



Spatial+spectral localization

Error reduction according to scale



Spectral localization for the large scales

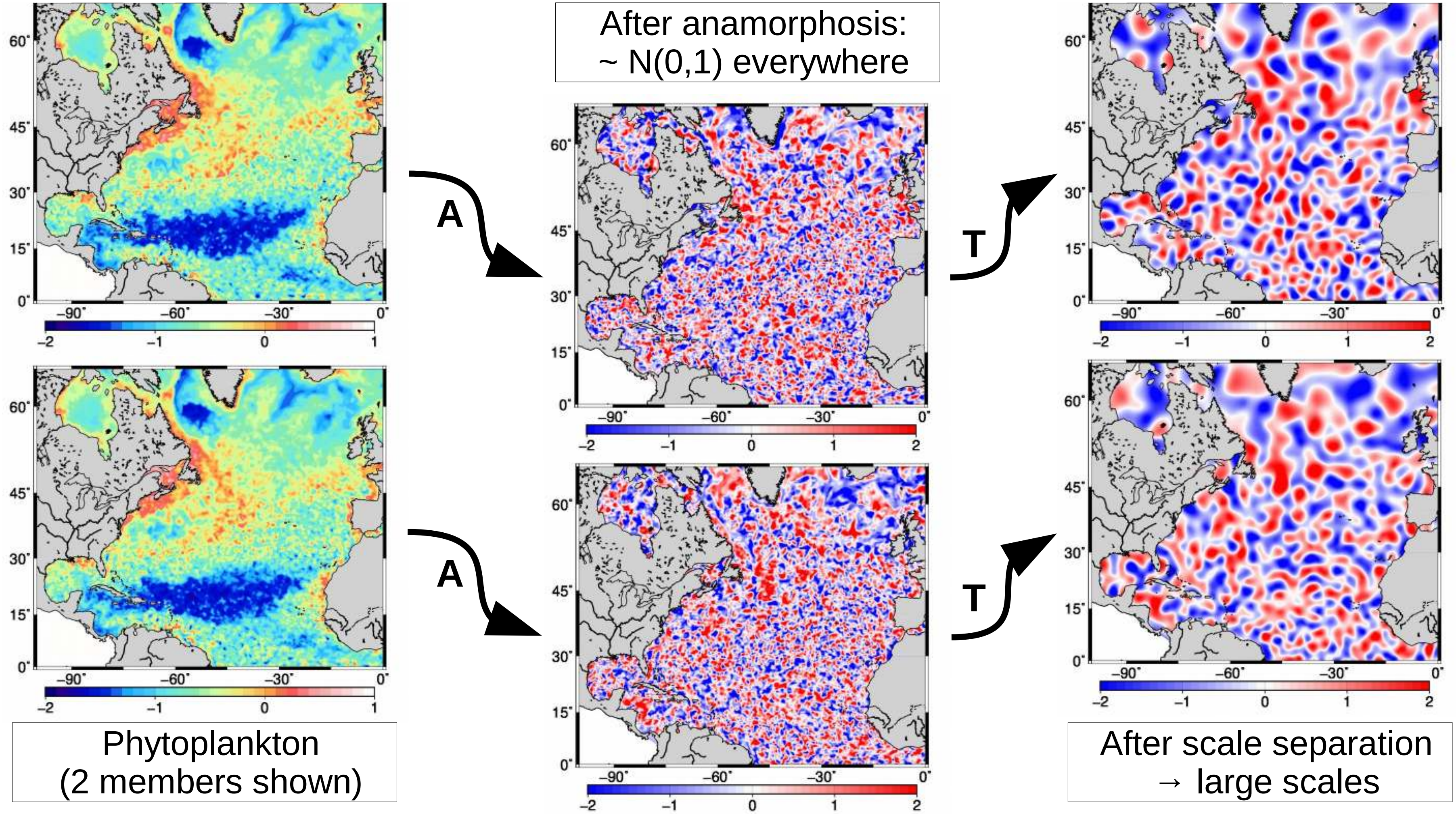
Spatial localization for the small scales

→ Multiscale algorithm

From Tissier et al., 2016. Mutliscale ensemble observational update combining spatial and spectral localization, Ocean Science, in preparation.

Example 2: in the green ocean

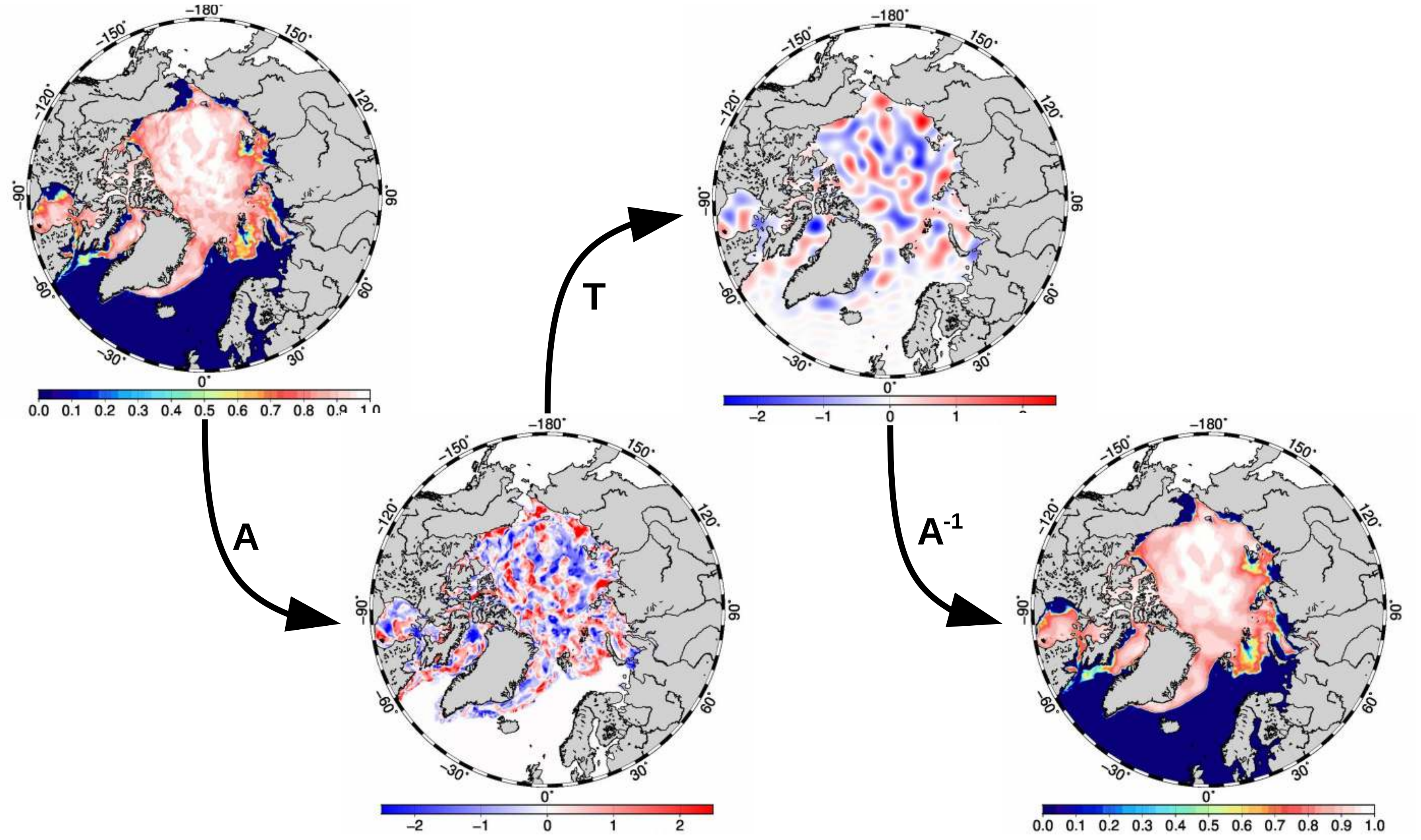
Application to NATL025/PISCES biogeochemical fields



Phytoplankton: heterogeneous statistics, non-Gaussian, limited by bounds
→ scale separation is possible if anamorphosis is applied

Example 3: in the white ocean

Application to sea ice concentration



The large structures (not exactly the large scales) are extracted without going out of bounds and without problems with the ice edge

Scale separation to better use large-scale observations

The spectral transformation is cheap
(if limited to large scales)

Explicit control of the large scales
with spectral localization
Before going to the smaller scales
with spatial localization

Applicable to non-Gaussian variables,
even limited by bounds,
providing anamorphosis is applied

5

Conclusions

Uncertain model and observations → uncertain products

The model becomes <u>non-deterministic</u>;
il is seen as a complex system,
built up from uncertain components

→ **The goal of ocean modelers is then to build a model as informative as possible at the lesser cost.**

<u>Uncertainty</u> must be represented
in ocean data assimilation systems

→ **Adjust the system to the features of the probability distributions (non-Gaussian behaviour, multiscale correlation structure)**

→ **Objective comparison between simulations and observations**

→ **Compute nonlinear diagnostic from the products**

What desirable features for the assimilation systems ?

Uncertainty is bound to become a key feature of the systems that we are using in oceanography, not something that can be thought separately from the results

→ Be prepared for stochastic model dynamics, stochastic observation operator, stochastic forcing, stochastic coupling,...

→ Be prepared to increasing dynamical complexity: more components (optical module, elasto-fragile rheology,...), more scales, more non-Gaussian behaviours,... Be prepared to smoothly upgrade the system accordingly.

→ The real world is uncertain; end-users live in the real world; be prepared for their need of reliable information about uncertainty in the products.

Error is viewed therefore
not as an extraneous and
misdirected or misdirecting accident,
but as an essential part
of the process under consideration.

John von Neuman (1956),
in « Probabilistic logics and the synthesis of
reliable organisms from unreliable components ».