# A generic approach to explicit simulation of uncertainty in the NEMO ocean model

Jean-Michel Brankart
Florent Garnier, Pierre Brasseur

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

Grenoble, 20-21 April 2017

**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)

# Motivations for a probabilistic approach

**The deterministic approach is not always sufficient to describe the dynamical behaviour of the system**

**Comparison between simulations and observations is easier with the probabilistic approach**

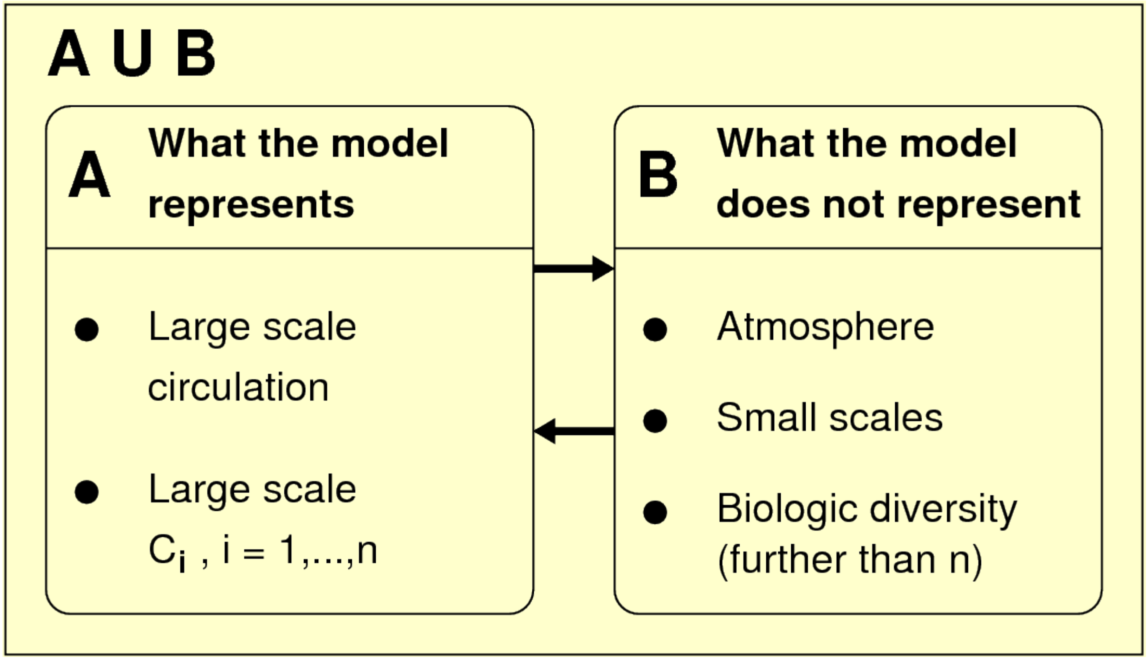
**A good knowledge of model accuracy is necessary to solve data assimilation problems**

# Outline

# 1

# Introduction

# 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 the main source of uncertainty in the model.**

# Uncertainty, as a key component of our systems

**What are the uncertain components of our systems ?**

**How to describe uncertainties ?**

**How does it participate to the solution of inverse problems ?**

# 2

# Explicit simulation of uncertainties

# A first simple implementation based on autoregressive processes (1)

**<u>Method</u>: explicitly simulate uncertainties in the model using *random numbers***

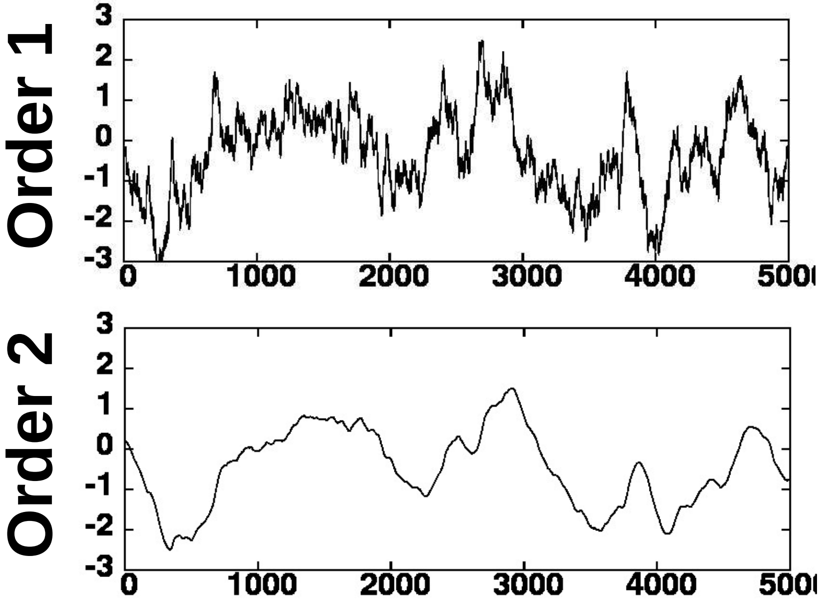
At every model grid point (in 2D or 3D), generate a set of **independent Gaussian autoregressive processes:**

$$\xi(t_k) = a\,\xi(t_{k-1}) + b\,w + c$$

where ***w*** is a Gaussian white noise (→ order 1 process) or an autoregressive process of order n-1 (→ order n process)

Parameters ***a, b, c*** to specify:

mean, standard deviation and correlation timescale



# A first simple implementation based on autoregressive processes (2)

## Introduce a spatial correlation structure

by applying a spatial filter to the map of autoregressive processes:

$\tilde{\xi} = \mathcal{F}[\xi]$ (filtering operator)

$\mathcal{L}[\tilde{\xi}] = \xi$ (elliptic equation)

which can easily be made flow dependent if needed

## Modify the marginal probability distributions

by applying anamorphosis transformation to every individual Gaussian variable:

$\tilde{\xi} = \mathcal{T}[\xi]$ (nonlinear function)

for instance to transform the Gaussian variables into lognormal or gamma variables if positive noise is needed

→ **This provides a generic technical way of implementing a wide range of stochastic parameterizations**

# Technological approach: a stochastic module in NEMO

These processes are generated using a **new module in NEMO**,
and **can be used in any component** of the model (Brankart et al., 2015):
circulation model, ecosystem model, sea ice model

**Algorithm 1** sto_par

**for all** (map $i = 1, \ldots, m$ of autoregressive processes) **do**
  Save map from previous time step: $\xi_- \leftarrow \xi_i$
  **if** (process order is equal to 1) **then**
    Draw new map of random numbers $w$ from $\mathcal{N}(0, 1)$: $\xi_i \leftarrow w$
    Apply spatial filtering operator $\mathcal{F}_i$ to $\xi_i$: $\xi_i \leftarrow \mathcal{F}_i[\xi_i]$
    Apply precomputed factor $f_i$ to keep SD equal to 1: $\xi_i \leftarrow f_i \times \xi_i$
  **else**
    Use previous process (one order lower) instead of white noise: $\xi_i \leftarrow \xi_{i-1}$
  **end if**
  Multiply by parameter $b_i$ and add parameter $c_i$: $\xi_i \leftarrow b_i \times \xi_i + c_i$
  Update map of autoregressive processes: $\xi_i \leftarrow a_i \times \xi_- + \xi_i$
**end for**

→ Generic and flexible technological approach
→ Model independent implementation
→ Possible to simulate many kinds of uncertainty

**Algorithm 2** sto_par_init

Initialize number of maps of autoregressive processes to 0: $m \leftarrow 0$
**for all** (stochastic parameterization $k = 1, \ldots, p$) **do**
  Set $m_k$, the number of maps of autoregressive processes required for this parameterization
  Increase $m$ by $m_k$ times the process order $o_k$: $m \leftarrow m + m_k \times o_k$
**end for**
**for all** (map $i = 1, \ldots, m$ of autoregressive processes) **do**
  Set order of autoregressive processes
  Set mean ($\mu_i$), standard deviation ($\sigma_i$) and correlation timescale ($\tau_i$) of autoregressive processes
  Compute parameters $a_i, b_i, c_i$ as a function of $\mu_i, \sigma_i, \tau_i$
  Define filtering operator $\mathcal{F}_i$
  Compute factor $f_i$ as a function of $\mathcal{F}_i$
**end for**
Initialize seeds for random number generator
**for all** (map $i = 1, \ldots, m$ of autoregressive processes) **do**
  Draw new map of random numbers $w$ from $\mathcal{N}(0, 1)$: $\xi_i \leftarrow w$
  Apply spatial filtering operator $\mathcal{F}_i$ to $\xi_i$: $\xi_i \leftarrow \mathcal{F}_i[\xi_i]$
  Apply precomputed factor $f_i$ to keep standard deviation equal to 1: $\xi_i \leftarrow f_i \times \xi_i$
  Initialize autoregressive processes to $\mu + \sigma \times w$: $\xi_i \leftarrow \mu + \sigma \xi_i$
**end for**
**if** (restart file) **then**
  Read maps of autoregressive processes and seeds for the random number generator form restart file (thus overriding the initial seed)
**end if**

# List of uncertainties that have been implemented in NEMO using this generic stochastic module

## In the circulation model

Effect of unresolved scales in the equation of state
Uncertainties in the parameterized tendencies (SPPT scheme)
Uncertainties in the horizontal momentum diffusion
Uncertainties in vertical diffusion (lognormal distribution)
Uncertainties in bulk parameters $C_D$, $C_E$, $C_H$ (gamma distribution)

## In the ecosystem model

Effect of unresolved scales in the SMS terms of the equations
Uncertainties in model parameters (primary production and grazing)

## In the sea-ice model

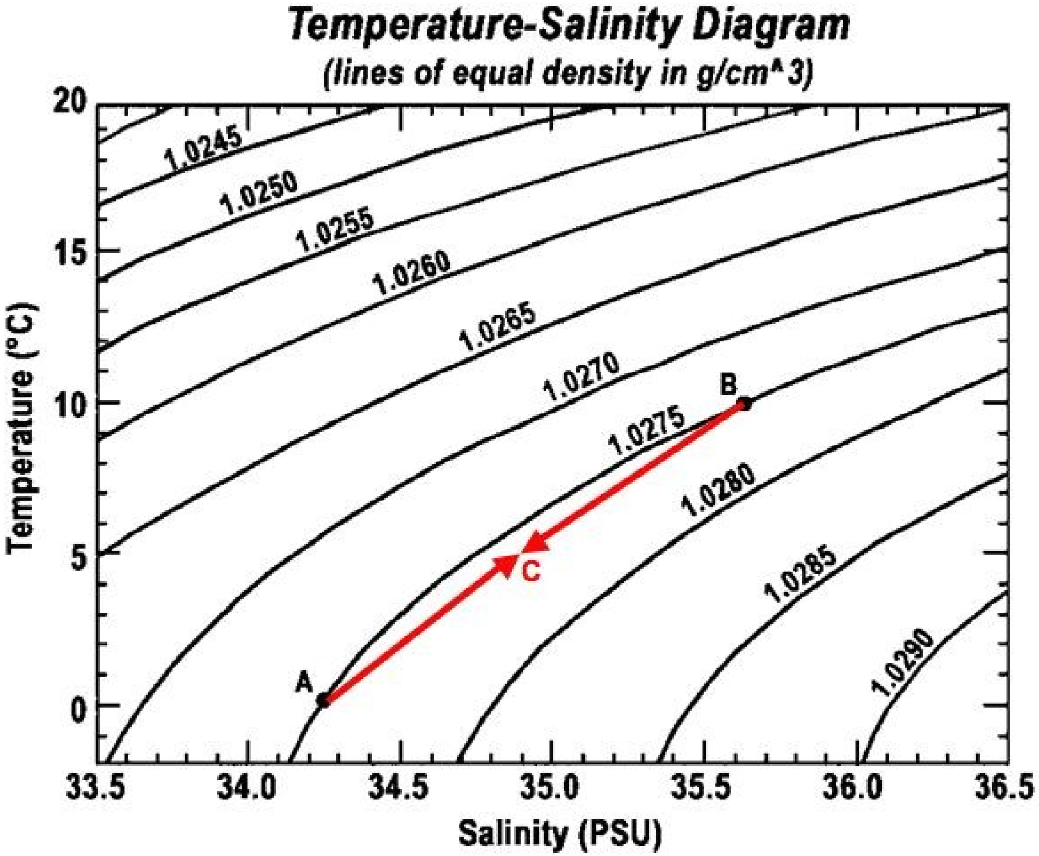
Uncertainties in ice strength (gamma distribution)
Uncertainties in ice/atmosphere drag (gamma distribution)
Uncertainties in ice/ocean drag (gamma distribution)
Uncertainties in ice albedo (beta distribution)

# 3

# Stochastic circulation model

# Uncertainties in the computation of density

**In the model, the large-scale density is computed form large-scale temperature and salinity, using the sea-water equation of state.**




**(a)**
**Mixing waters of equal density but different T&S systematically increases density (cabbeling)**

**(b)**
**Averaging T&S equations systematically overestimates density (in a fluctuating, non-deterministic way)**

**Because of the nonlinearity of the equation of state, unresolved scales produce an average effect on density.**

# Stochastic equation of state for the large scales

**Stochastic parameterization (Brankart, 2015)**

using a set of random T&S fluctuations
$\Delta T_i$ et $\Delta S_i$ , i=1,...,p
to simulate unresolved T&S fluctuations

$$\rho = \frac{1}{p} \sum_{i=1}^{p} \rho \left[ T + \Delta T_i, S + \Delta S_i, p_0(z) \right] \quad \text{with} \quad \sum_{i=1}^{p} \delta T^{(i)} = 0 \, , \, \sum_{i=1}^{p} \delta S^{(i)} = 0$$

No effect if the equation of state is linear.
Proportional to the square of unresolved fluctuations.

**Correction $\Delta\rho$ applied in the thermal wind equation, as in the semi-prognostic method of Greatbatch et al. (2004)**

No direct modification of T&S; no enhanced diapycnal mixing.
T&S only modified indirectly through a modification of the main currents

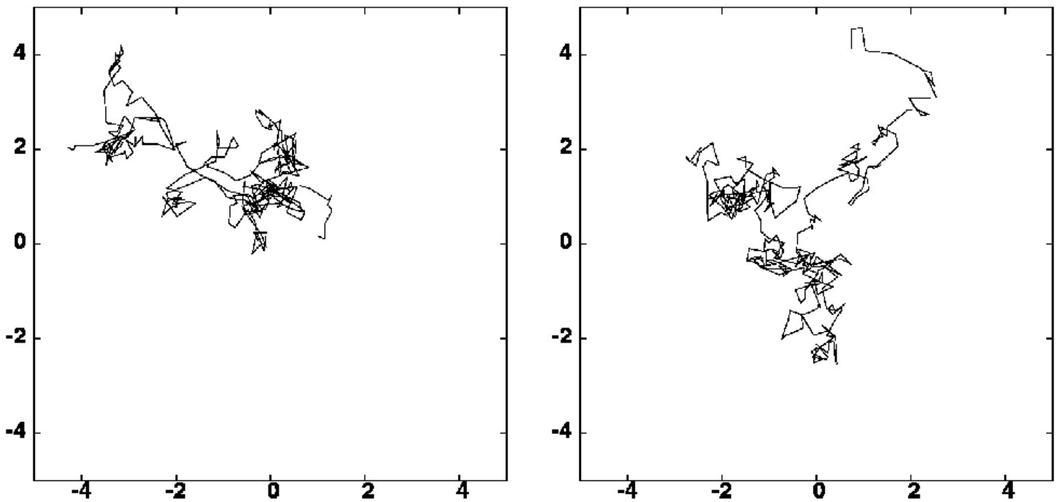
# Random walks to simulate unresolved temperature and salinity fluctuations

**Computation of the random fluctuations $\Delta T_i$ et $\Delta S_i$**

as a scalar product of the local gradient with random walks $\xi_i$

$$\Delta T_i = \boldsymbol{\xi}_i \cdot \nabla T \quad \text{and} \quad \Delta S_i = \boldsymbol{\xi}_i \cdot \nabla S$$

## Random walks




## Assumptions

AR1 random processes

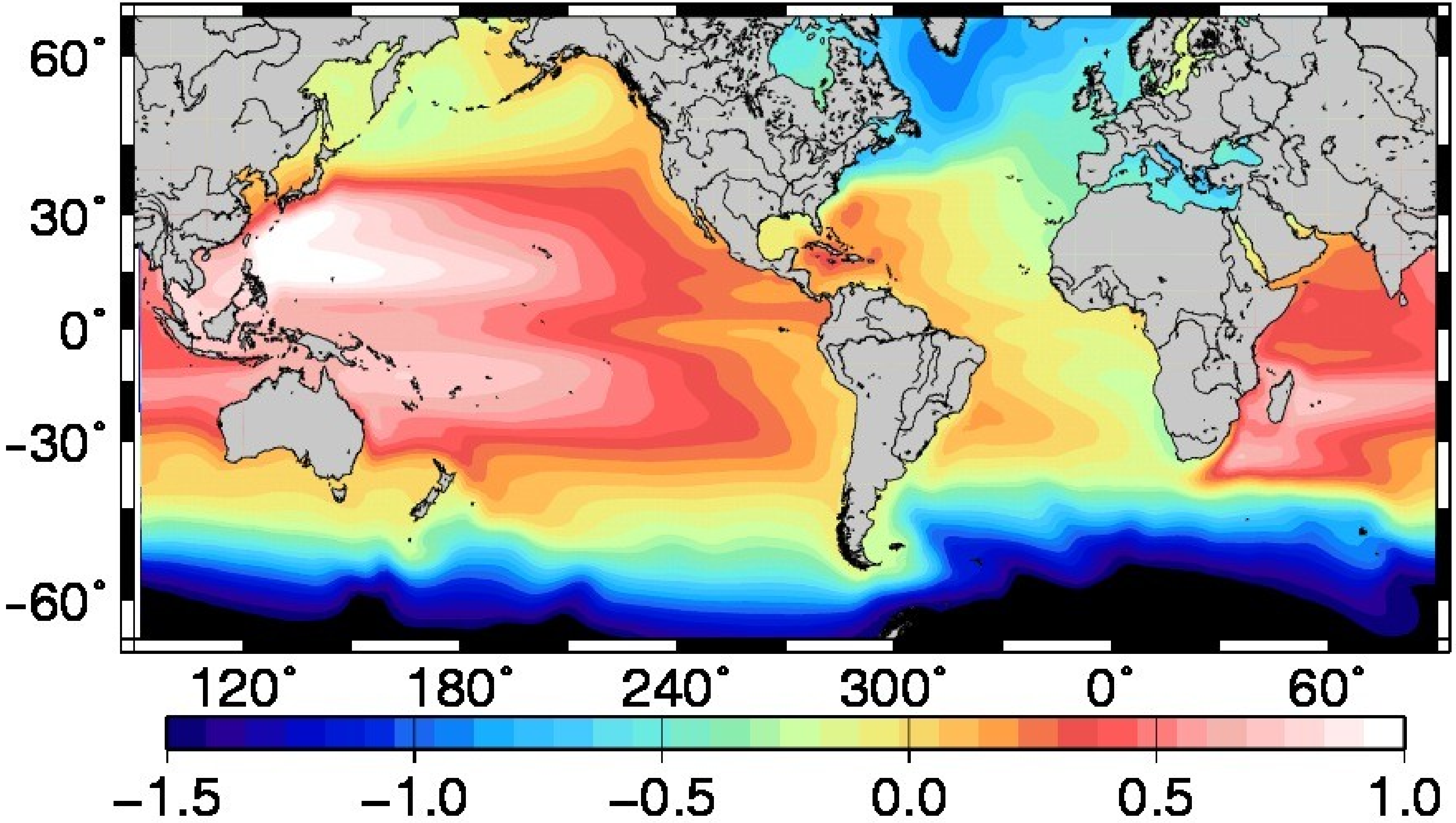
uncorrelated on the horizontal

fully correlated
along the vertical

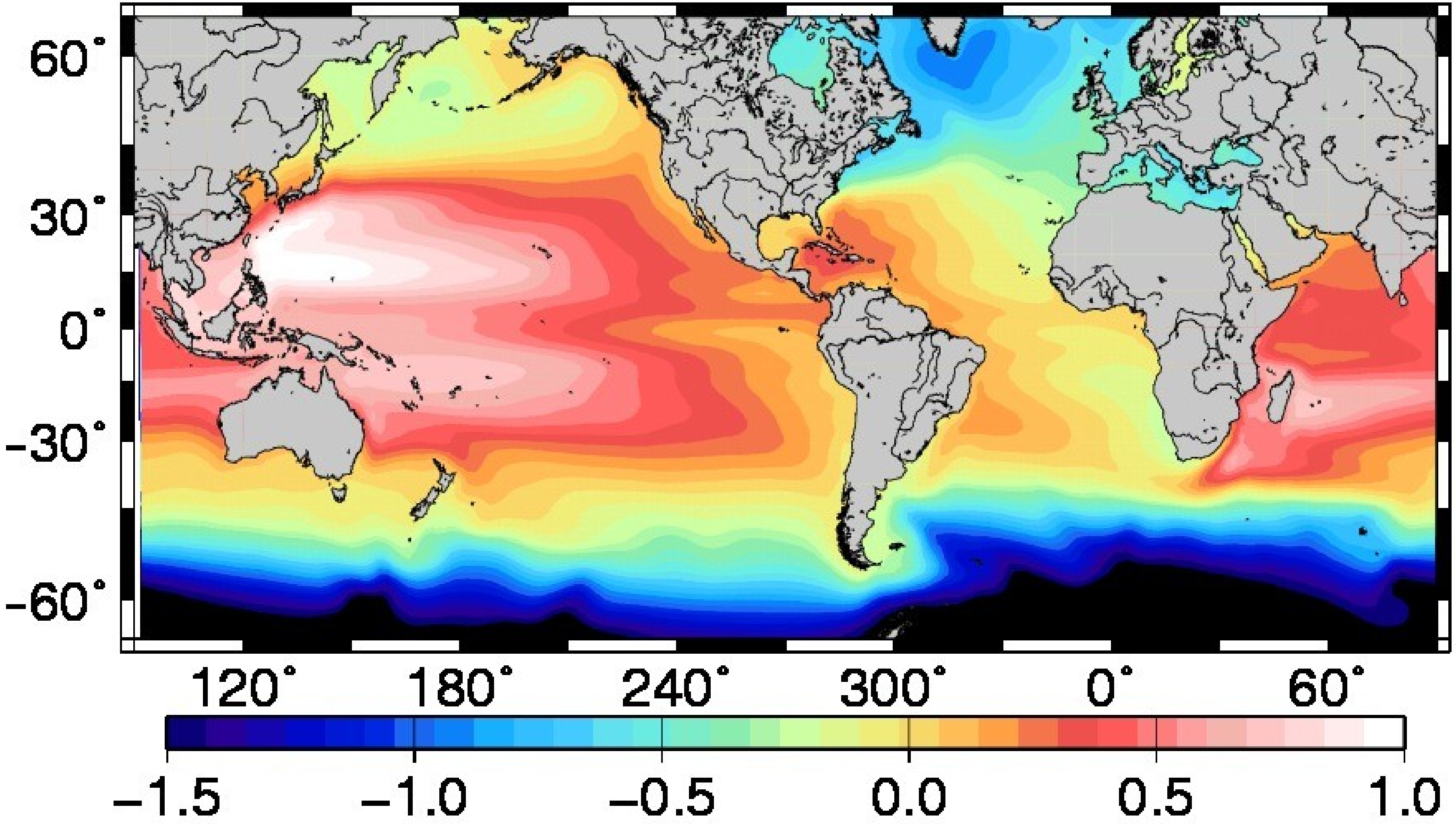
5-day time correlation

horizontal std: 2-3 grid points
vertical std: <1 grid point

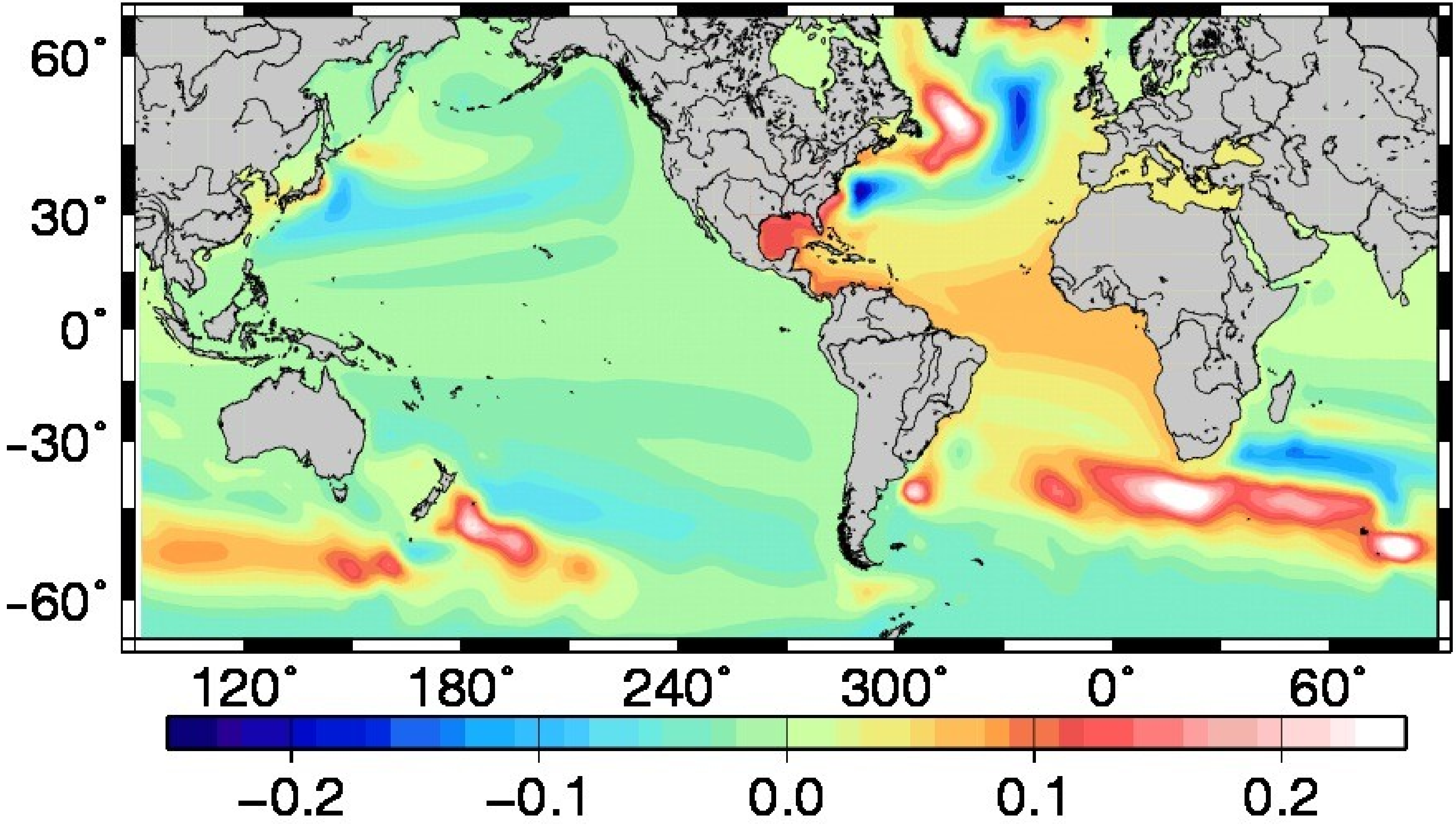
# Mean sea surface elevation (standard)



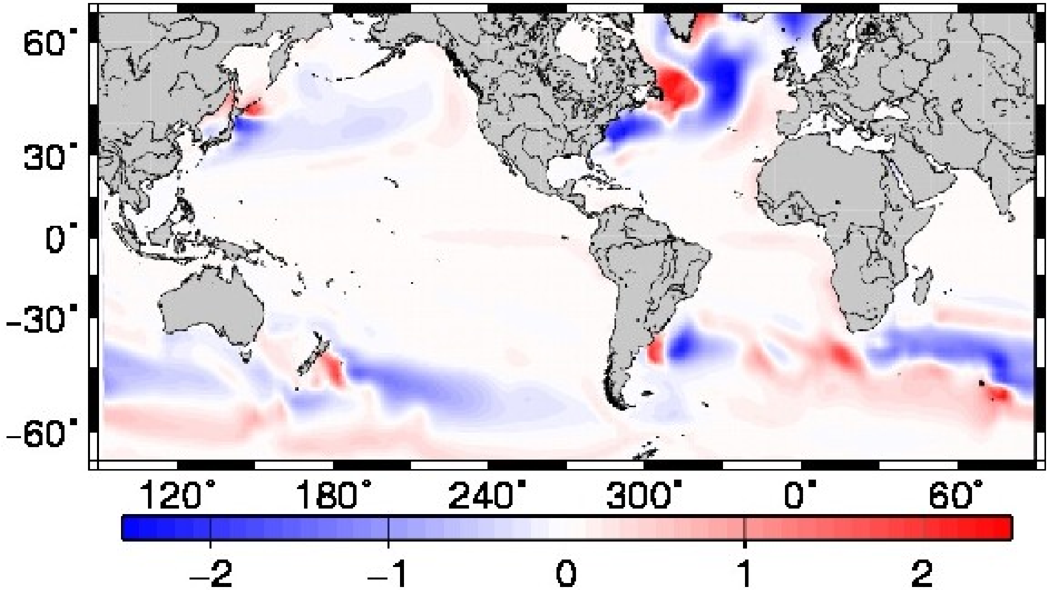
# Mean sea surface elevation (stochastic)

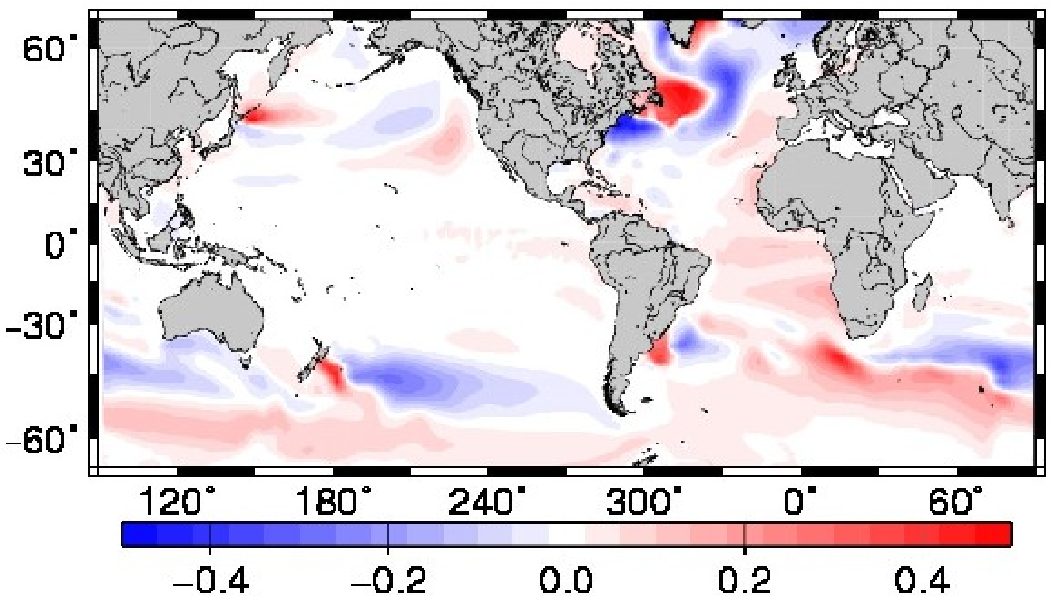


# Mean sea surface elevation difference



# Averaged SST & SSS difference




**Modification of the mean flow**

↓

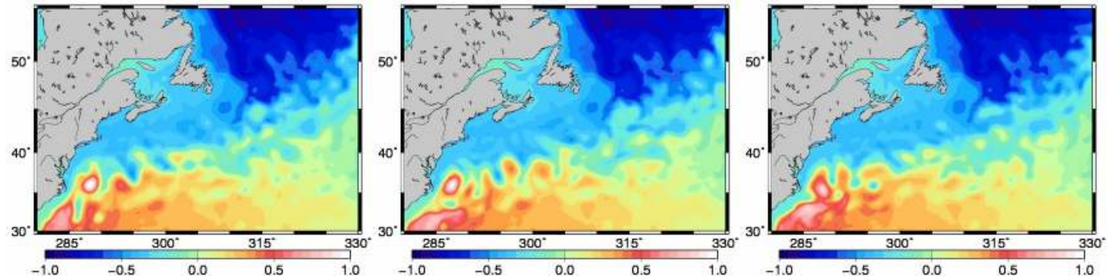
**Modification of the mean SST & SSS**

↓

**Modification of air/sea interactions**

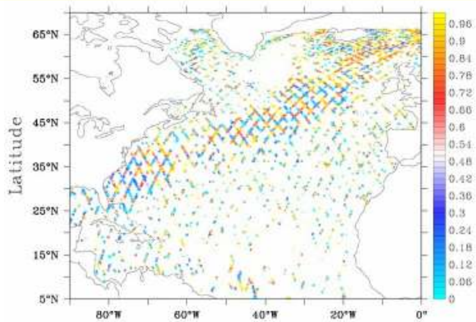
# Ensemble of mesoscale flows

<u>SSH pdf</u>
ensemble NATL025:
3 members among 96

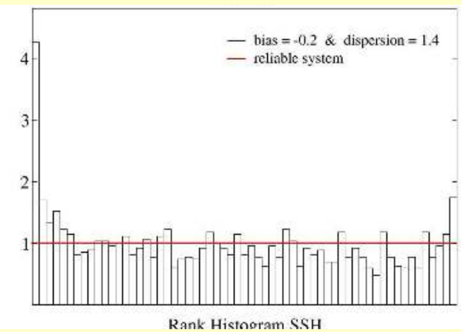



**→ comparison to observation (rank histogram, CRPS score, etc.)**

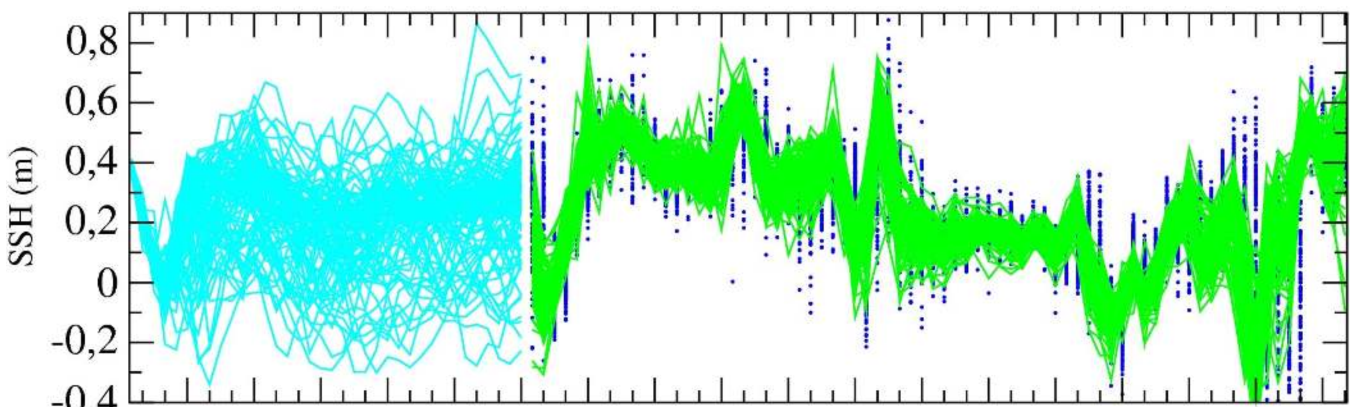
<u>Map of ranks</u>
of JASON
altimetric data




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




**→ assimilation of altimetric data (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

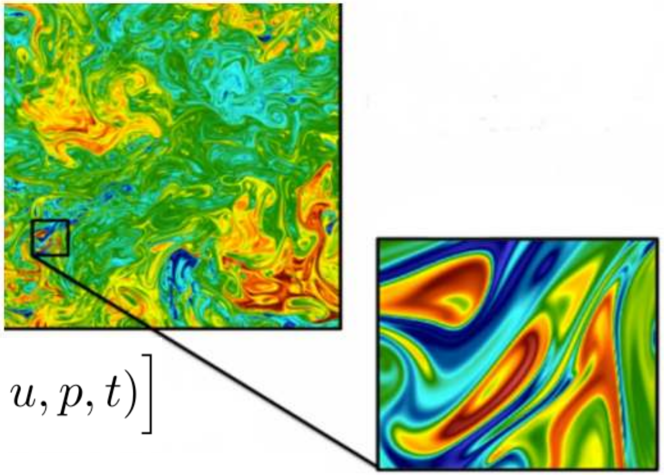
# 4

# Stochastic biogeochemical model

# Stochastic biogeochemical model

**Multiple sources of uncertainty in ecosystem model:**
**<u>unresolved biological diversity</u>, <u>unresolved scales</u>, etc.**

**<u>Unresolved scales</u>**
**stochastic processes**
**explicitly simulating**
**unresolved fluctuations of $C_i$**



$$\frac{\partial C}{\partial t}|_{bio} = \frac{1}{2}\left[SMS(\overline{C}+\delta\overline{C},u,p,t) + SMS(\overline{C}-\delta\overline{C},u,p,t)\right]$$

$$\delta\overline{C} = \xi(t)\overline{C}$$

**<u>Unresolved diversity</u>**
**multiplicative noise**
**in model parameters**

$$\frac{\partial C}{\partial t}|_{bio} = SMS(C,u,p',t) \quad avec \quad p' = \exp[\xi(t)]p$$

**<u>Assumptions</u>**

perturb 8 parameters governing primary production and grazing

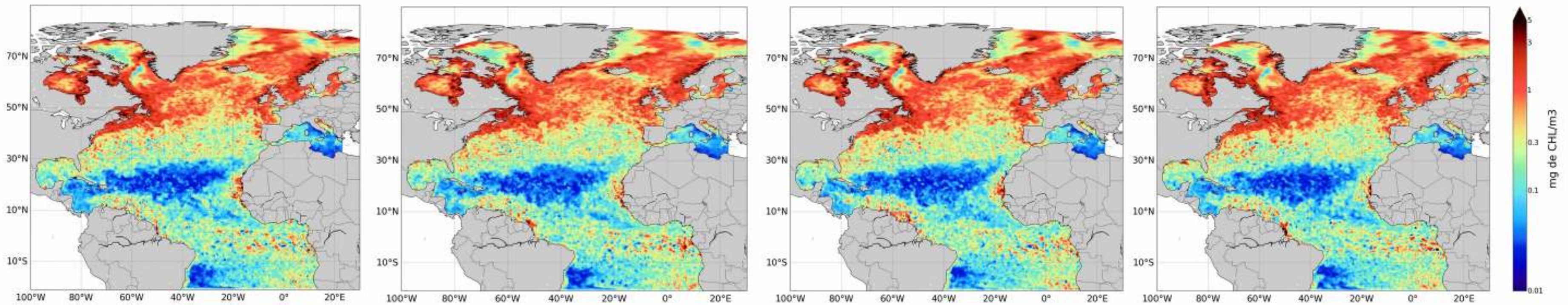
fully correlated along the vertical

30-day time correlation

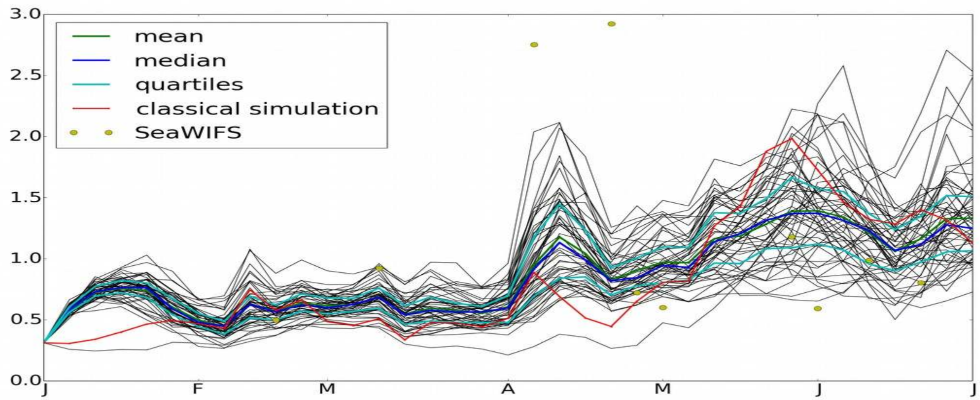
→ **Considerable effect on the mean behaviour of the system**
→ **Increase of patchiness (↔ ocean colour data)**

# Ensemble simulation of the ecosystem

**<u>Probability distribution of chlorophyll concentration</u>**
as simulated here by ensemble NATL025/PISCES
with stochastic parameterization of uncertainty: 4 members among 50




member 1 | member 2 | member 3 | member 4



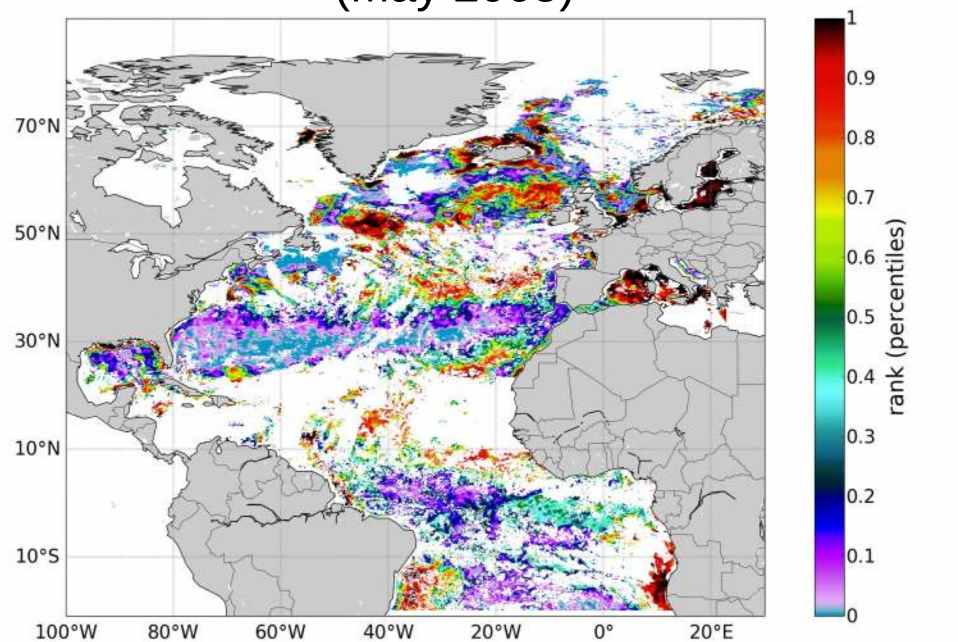

<u>Time evolution</u>
<u>of the pdf</u>
<u>for phtyoplankton</u>

from January to June 2005

→ **description of biogeochemical uncertainties (Garnier et al., 2016)**

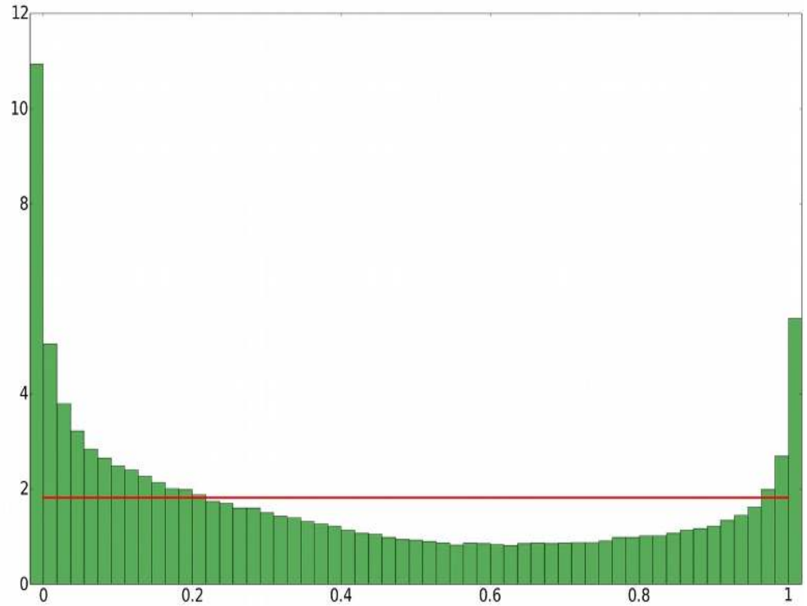
# Comparison to ocean colour observations

<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**

# 6

# Conclusions

# Probabilistic model and ensemble simulations

**The NEMO model becomes <u>probabilistic</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.**

**This probabilistic description requires <u>ensemble simulations</u>**

**→ Objective comparison between simulations and observations**

**→ Deal with model uncertainty in ocean data assimilation systems**

An appropriate simulation of uncertainty is necessary
to make the link between model, observations,
and data assimilation systems

**Uncertainty is bound to become**
**a <u>key constituent of the systems</u>**
**that we are using in oceanography,**
**not something that can be thought**
**separately from the results**

**Properly dealing with uncertainty will require**
**an <u>integrated engineering approach</u>**
**at the interface between oceanography**
**and applied mathematics**

**Error is viewed therefore
not as an extraneous and
misdirected of 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 ».