Explicit simulation of uncertainties in ocean models, and application in SANGOMA benchmarks

Jean-Michel Brankart

Laboratoire de Glaciologie et de Géophysique de l'Environnement
Equipe de Modélisation des Ecoulements Océaniques Multi-échelles

SANGOMA final meeting – November 5-6, 2015

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.

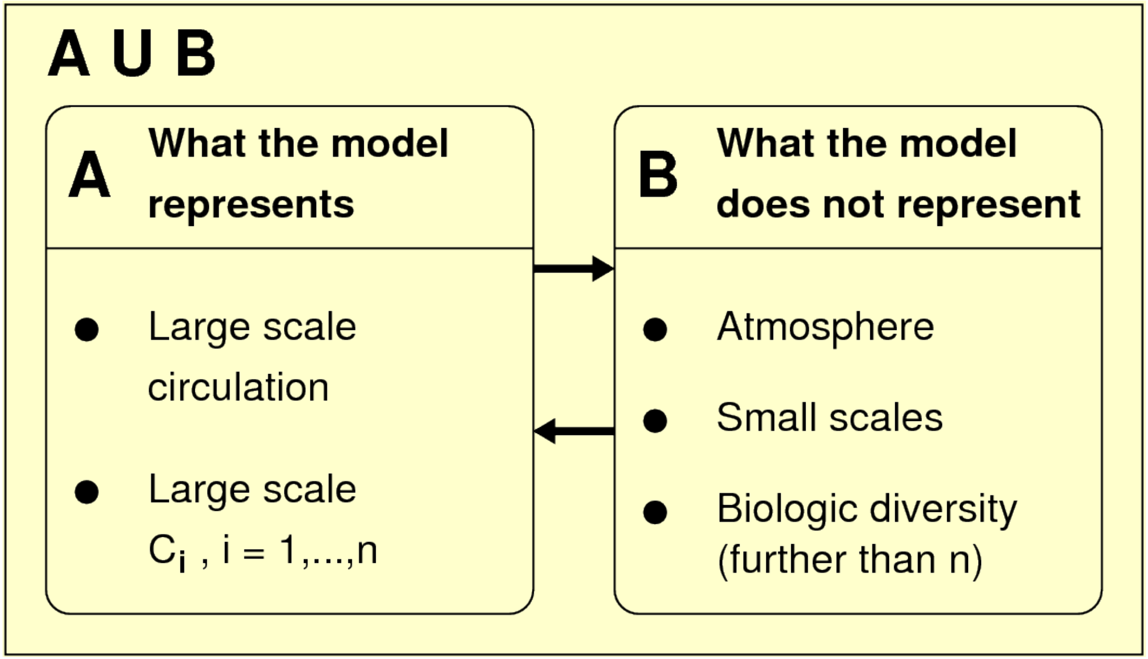
Summer Computer Simulation Conference (1975),
cited by Richard Hamming (1997)

Outline

1

Introduction

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.

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 probabilstic approach

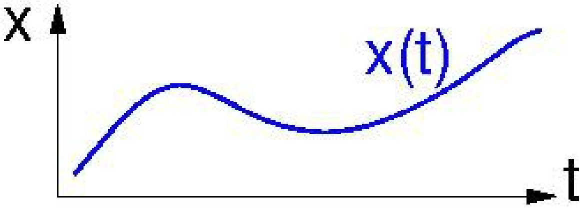
A good knowledge of model accuracy is necessary to solve data assimilation problems

Probabiliste approach to ocean modeling

Stochastic ocean dynamics, explicitly simulating uncertainties

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

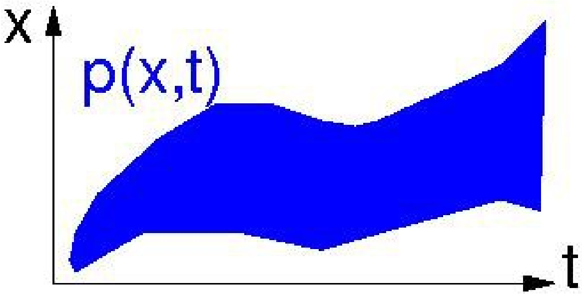
où $\mathbf{x} = [x_1, \ldots, x_N]$



Fokker-Planck equation, for the probability distribution p(**x**,t),
following ideas at the origin of the Ensemble Kalman filter (Evensen, 1994)

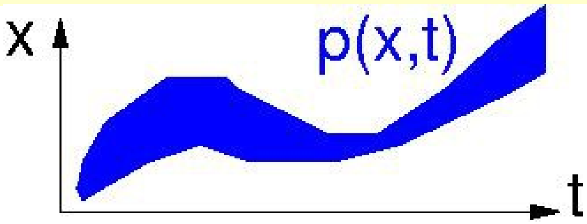
$$\frac{\partial p(\mathbf{x},t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left[\mathcal{M}_i(\mathbf{x},t)p(\mathbf{x},t)\right] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial x_i \partial x_j}\left[D_{ij}(\mathbf{x},t)p(\mathbf{x},t)\right]$$

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



Conditioning to observations, to reduce uncertainties

using an appropriate data
assimilation method



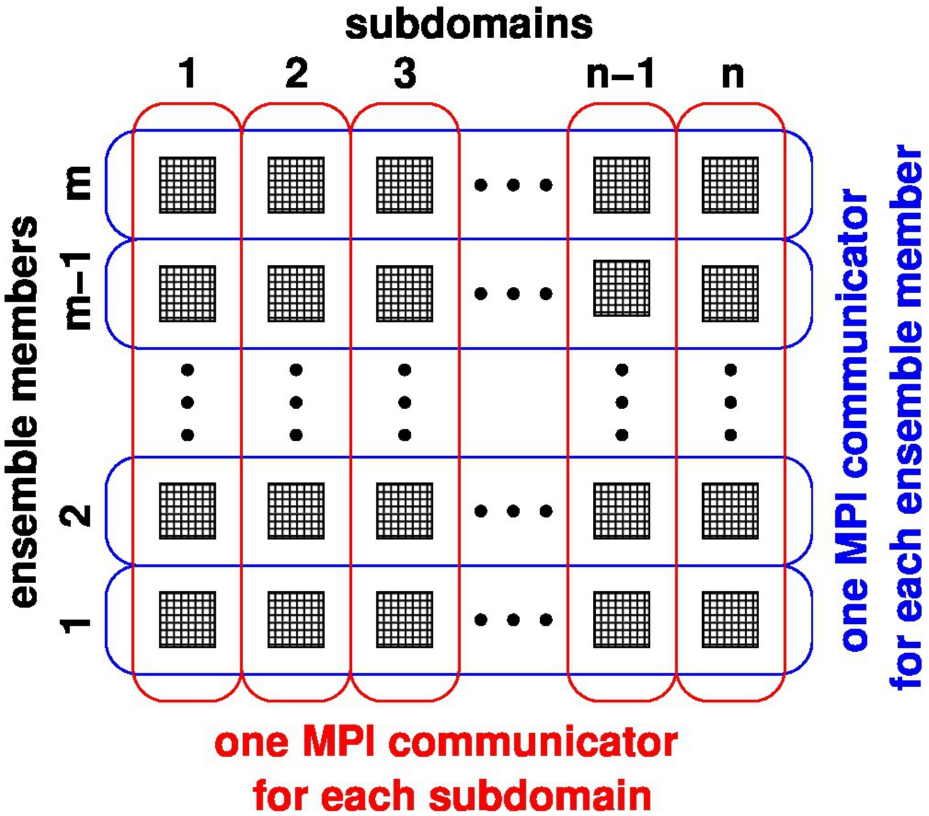
Technological approach: ensemble simulations

Solution of the Fokker-Planck equation using **a Monte Carlo method**
→ **ensemble NEMO simulation** → échantillon de p(**x**,t)

1 ensemble simulation
=
1 single call to NEMO

***m* times more processors**
to run
***m* members together in parallel**

To each processor is given
one of the *n* subdomains
from one of the *m* members



1 MPI communicator for each member
→ each member lives its own life, as in standard NEMO
1 MPI communicator for each subdomain
→ “online” computations of any feature of p(**x**,t)

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

2.1 Stochastic formulation of NEMO

<u>Objective</u>: transform a *deterministic* model into a *probabilistic* model

- **Describe the non-deterministic nature of the system**
- **Allow objective comparison to observations**
- **Introduce a weak model constraint in data assimilation**

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

- **Propose a generic and flexible technical approach**
- **Develop a first simple implementation**
- **External forcing
Unresolved scales
Unresolved diversity**

Autoregressive processes (1)

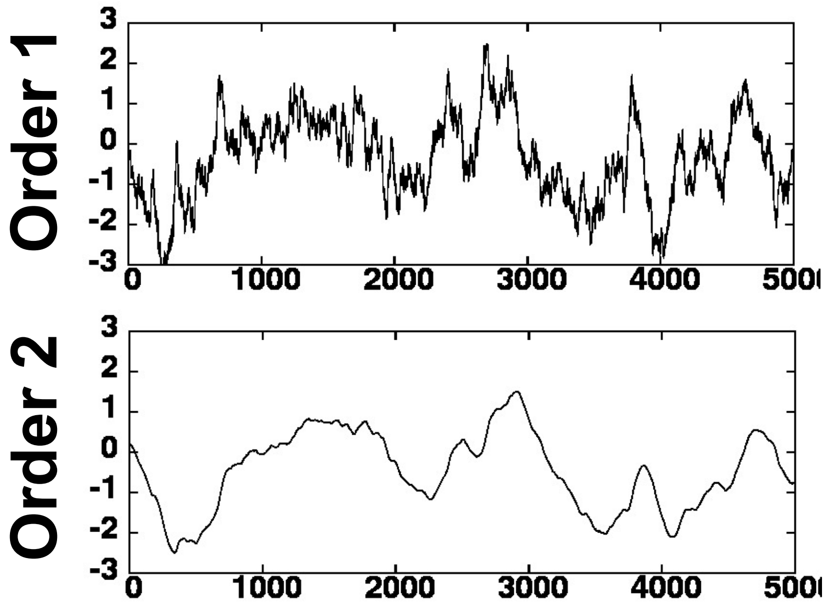
At every model grid point (in 2D or 3D), generate a set of **independent Gaussian autoregressive processes:**

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

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



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**

Technolgical approach: a new 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 ξ_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 (μ_i), standard deviation (σ_i) and correlation timescale (τ_i) of autoregressive processes
  Compute parameters a_i, b_i, c_i as a function of μ_i, σ_i, τ_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 ξ_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

Example 1: Stochastic perturbation of parameterized tendencies

With this generic implementation, we can reproduce the SPPT scheme proposed by Buizza et al. (1999)

Separate model operator
in NP (non-parameterized) et P (parameterized)

Assume that P is uncertain and
simulate uncertainty by a multiplicative noise ξ

$$\frac{d\mathbf{x}}{dt} = \mathcal{NP}\left[\mathbf{x}, \mathbf{u}(t), \mathbf{p}\right] + \mathcal{P}\left[\mathbf{x}, \mathbf{u}(t), \mathbf{p}\right] \boldsymbol{\xi}(t)$$

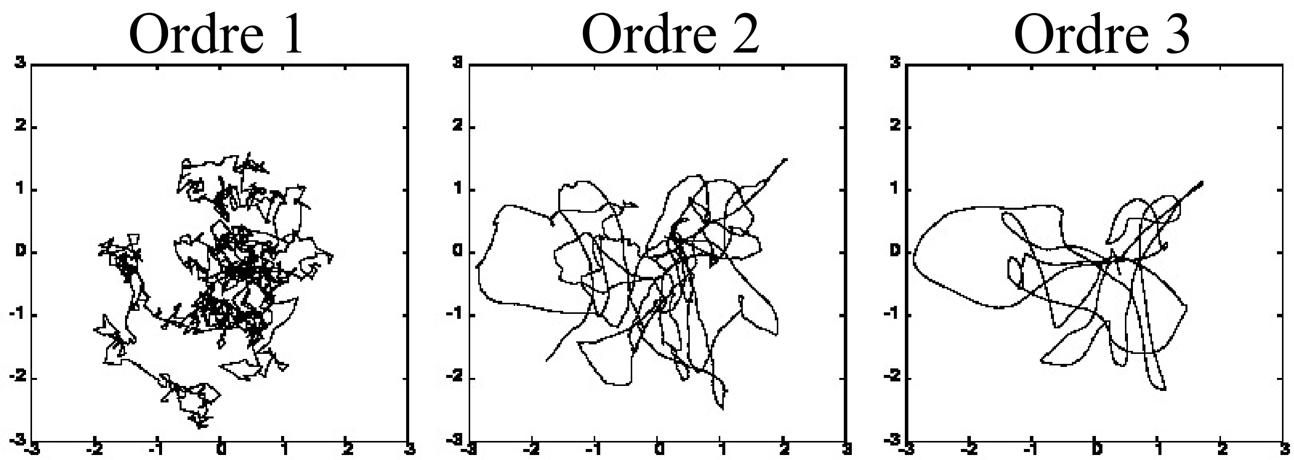
→ Use maps of autoregressive processes as ξ (with mean 1), and specify the correlation structure and marginal distribution of ξ.

Example 2: Stochastic parameterization of unresolved fluctuations

To simulate the effect of unresolved fluctuations in the nonlinear terms of the model equations

$$\frac{d\mathbf{x}}{dt} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{M}\left[\mathbf{x} + \delta\mathbf{x}_i(t), \mathbf{u}(t), \mathbf{p}\right] \quad \text{with} \quad \sum_{i=1}^{n} \delta\mathbf{x}_i(t) = 0$$

Generate fluctuations using **random walks** around every grid point



→ Use maps of autoregressive processes as components x, y, z of the random walks. Specify space and time correlation structure.

Example 3: Stochastic parameterization of unresolved diversity

To simulate the effect of the unresolved diversity of system behaviours (e.g. biological diversity,...)

This assumes that the system simultaneoulsy includes a variety of possible behaviours, which cannot be described by one single value of each parameter.

$$\frac{d\mathbf{x}}{dt} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{M}\left[\mathbf{x}, \mathbf{u}(t), \mathbf{p} + \delta\mathbf{p}_i(t)\right]$$

For instance, the ecosystem usually includes many different species of phytoplankton and zooplankton, each with its own behaviour, while the model can only resolve a few classes of species.

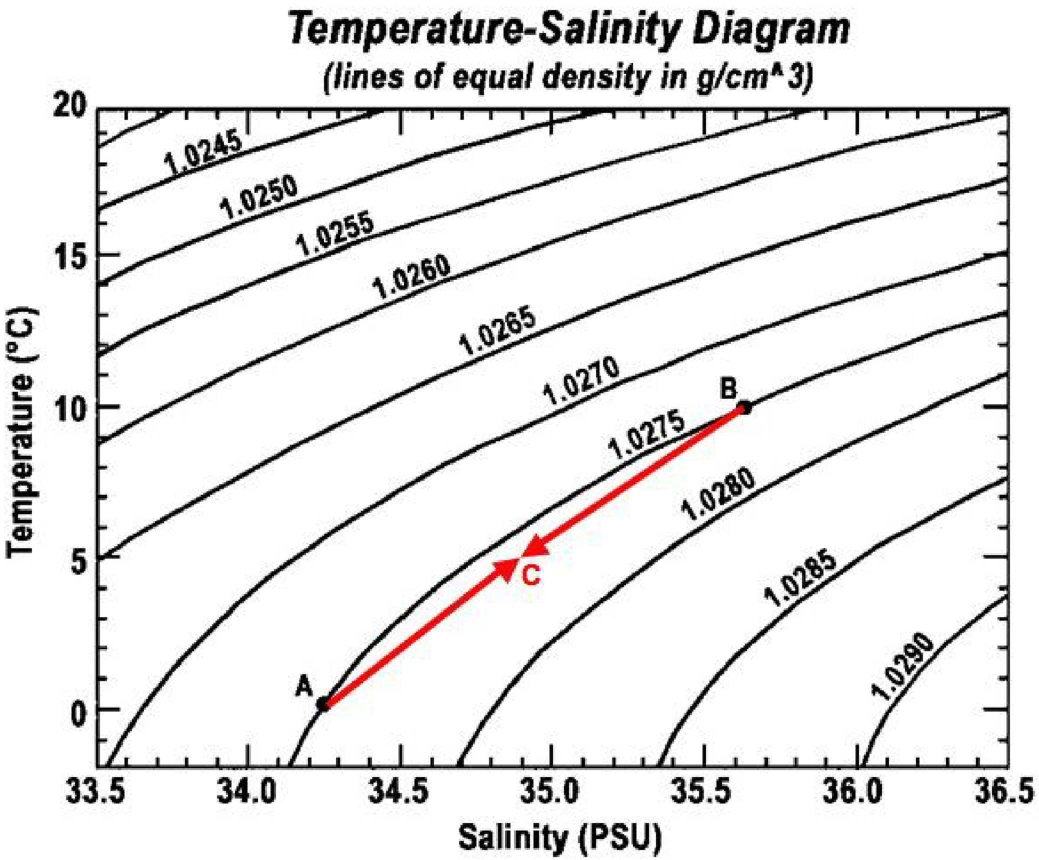
→ Use maps of autoregressive processes ξ as multiplicative noise for the parameters, and specify their correlation structure and their marginal distribution.

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.

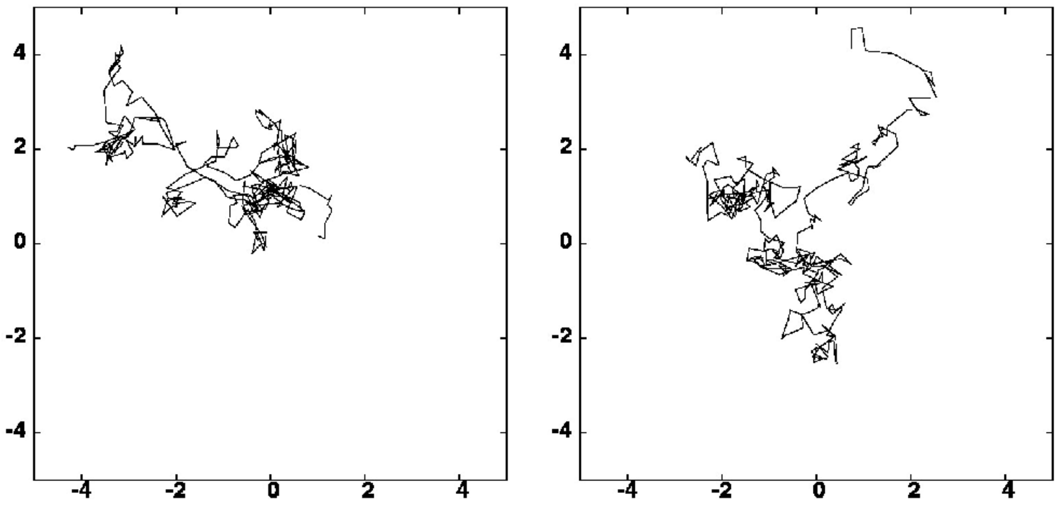
Random walks to simulate unresolved temperature and salinity fluctuations

Computation of the random fluctuations ΔT_i et ΔS_i

as a scalar product of the local gradient with random walks ξ_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

Stochastic equation of state for the large scales

Stochastic parameterization (Brankart, 2015)

using a set of random T&S fluctuations
ΔT_i et ΔS_i , i=1,...,p
to simulate unresolved T&S fluctuations

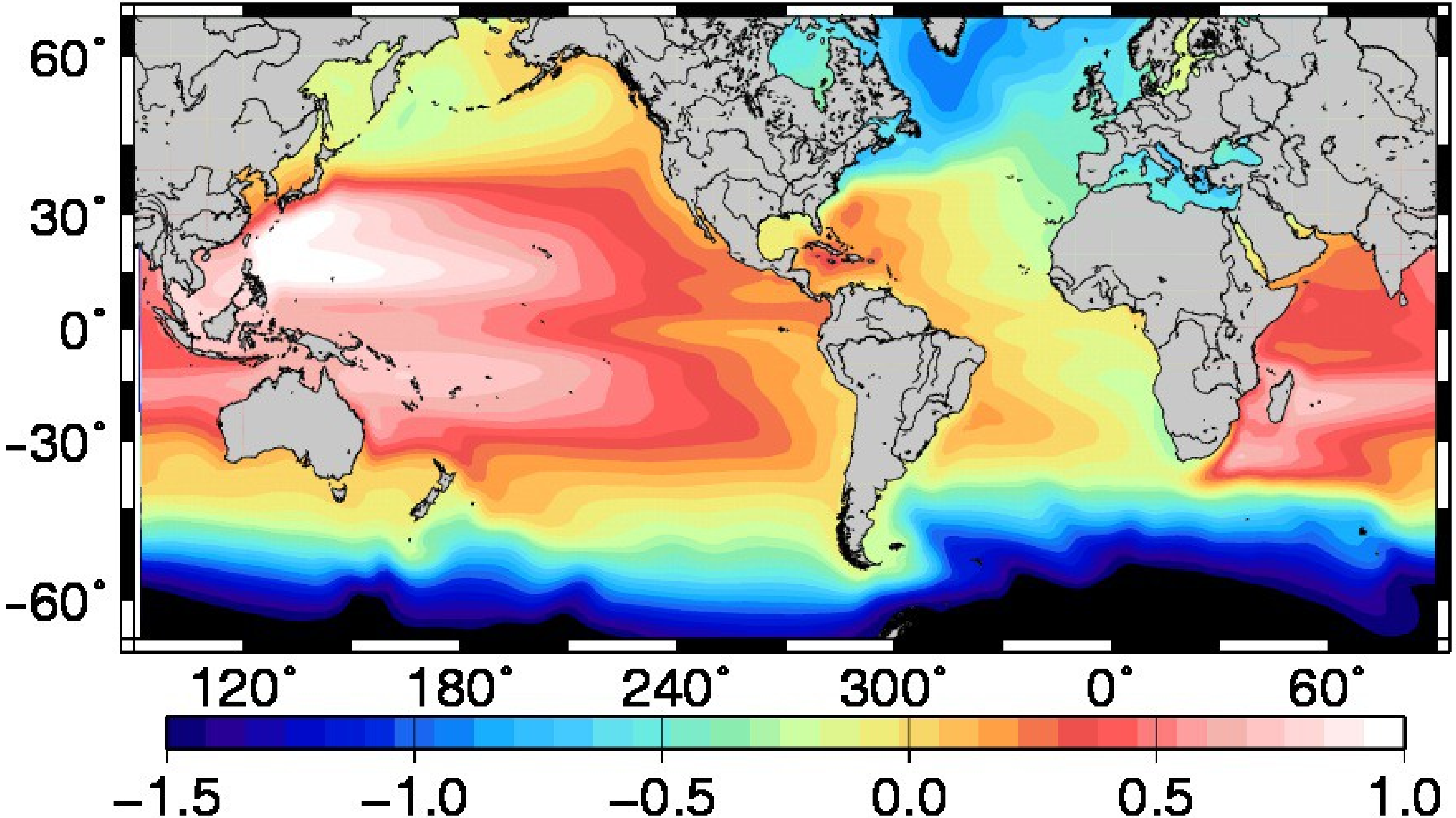
$$\rho = \frac{1}{p} \sum_{i=1}^{p} \rho \left[T + \Delta T_i, S + \Delta S_i, p_0(z) \right] \quad \text{with} \quad \sum_{i=1}^{p} \delta T^{(i)} = 0 \, , \, \sum_{i=1}^{p} \delta S^{(i)} = 0$$

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

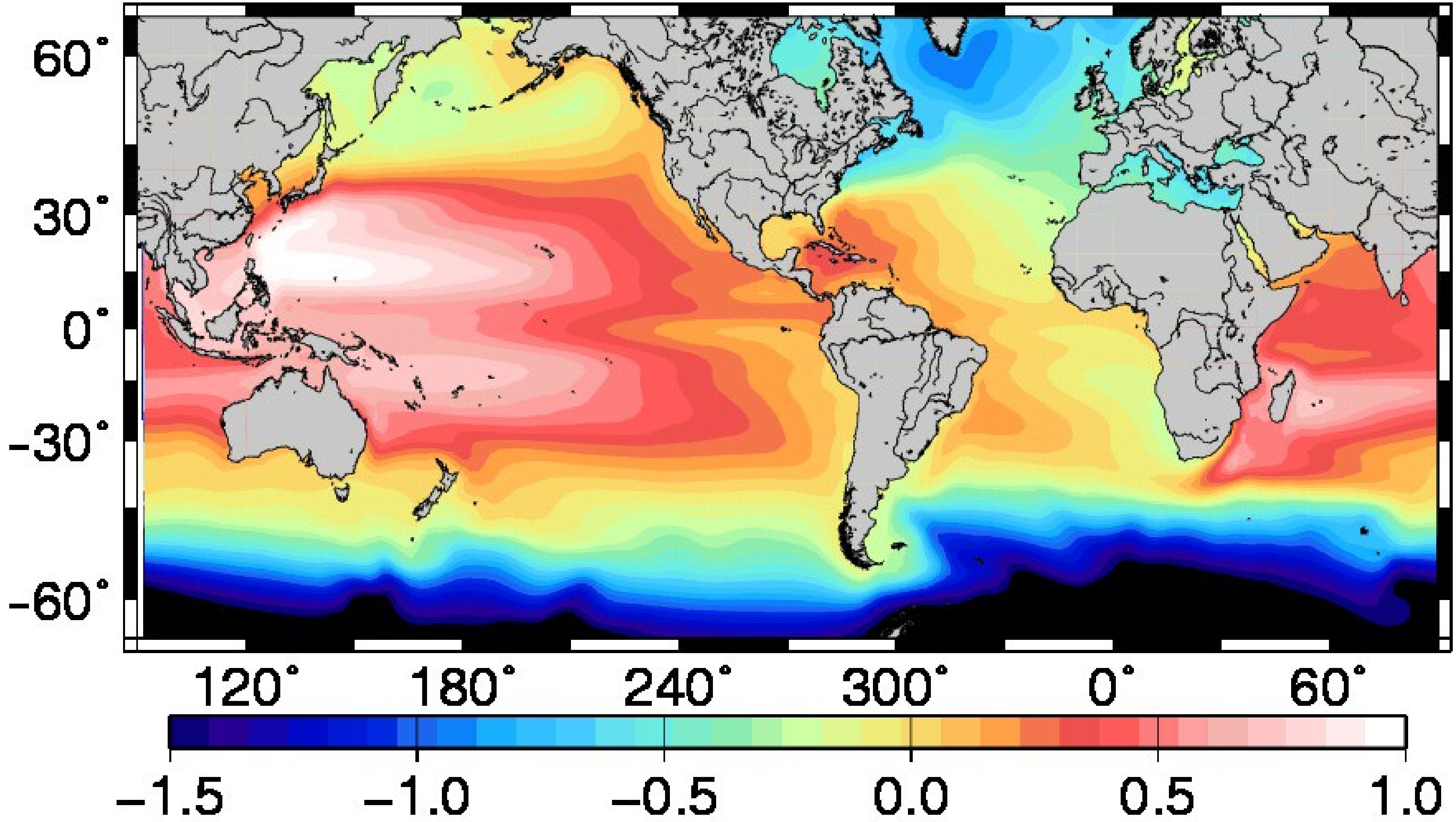
Correction $\Delta\rho$ applied in the thermal wind equation, as in the semi-prognostic method of Greatbatch et al. (2004)

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

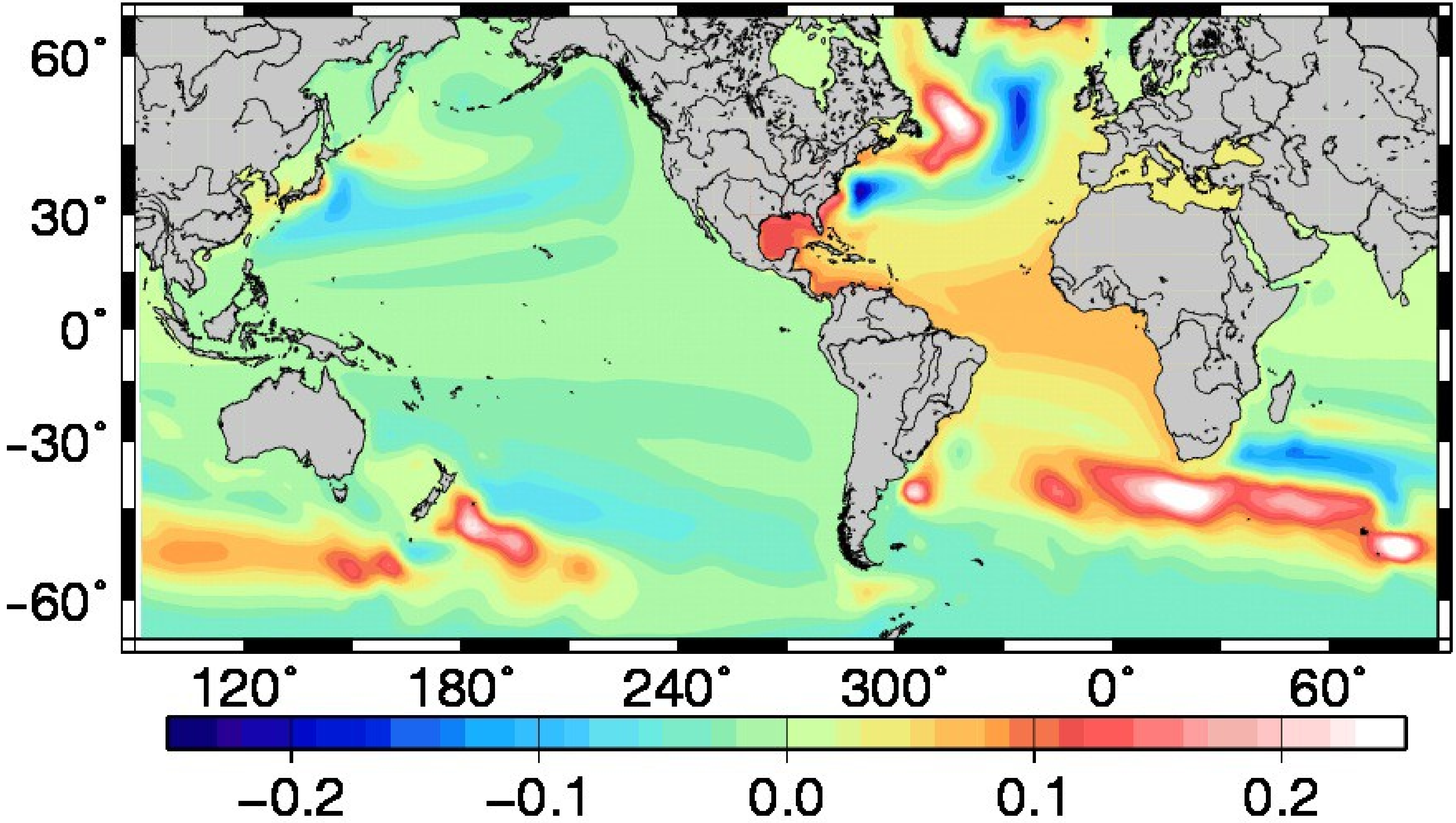
Mean sea surface elevation (standard)



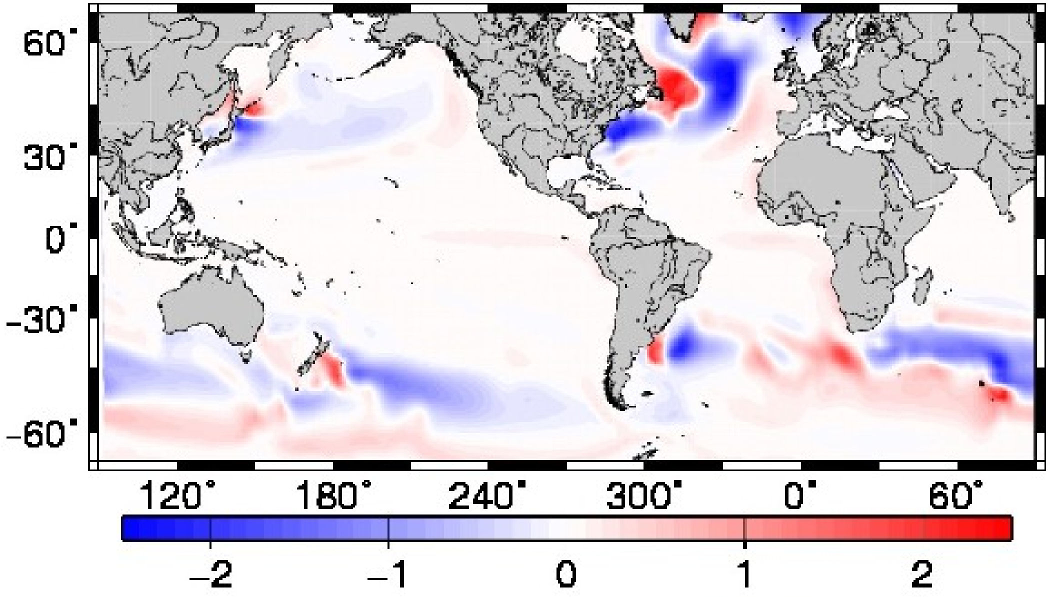
Mean sea surface elevation (stochastic)

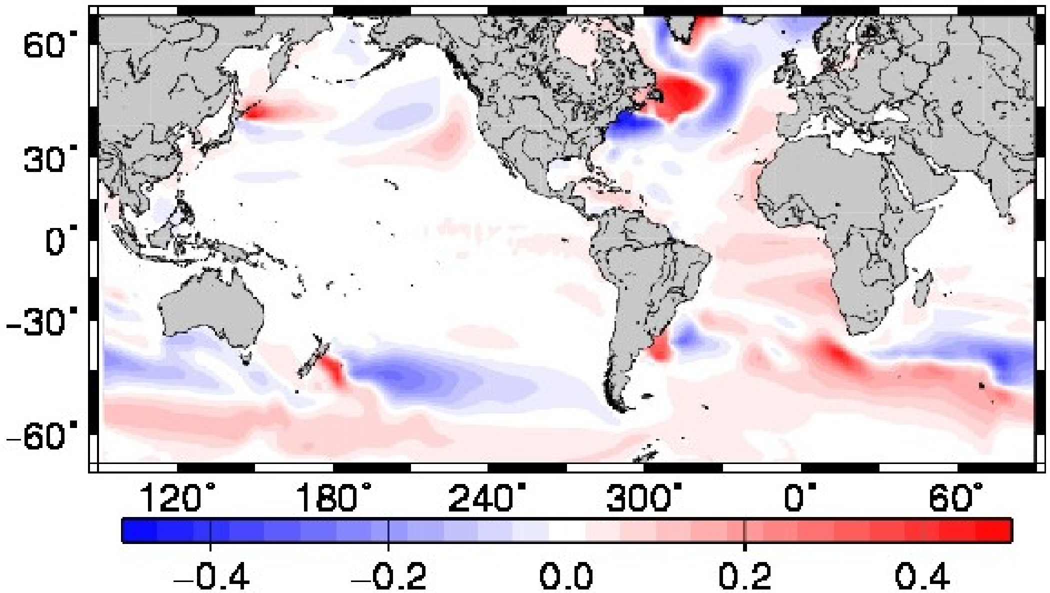


Mean sea surface elevation difference



Averaged SST & SSS difference





Modification of the mean flow

↓

Modification of the mean SST & SSS

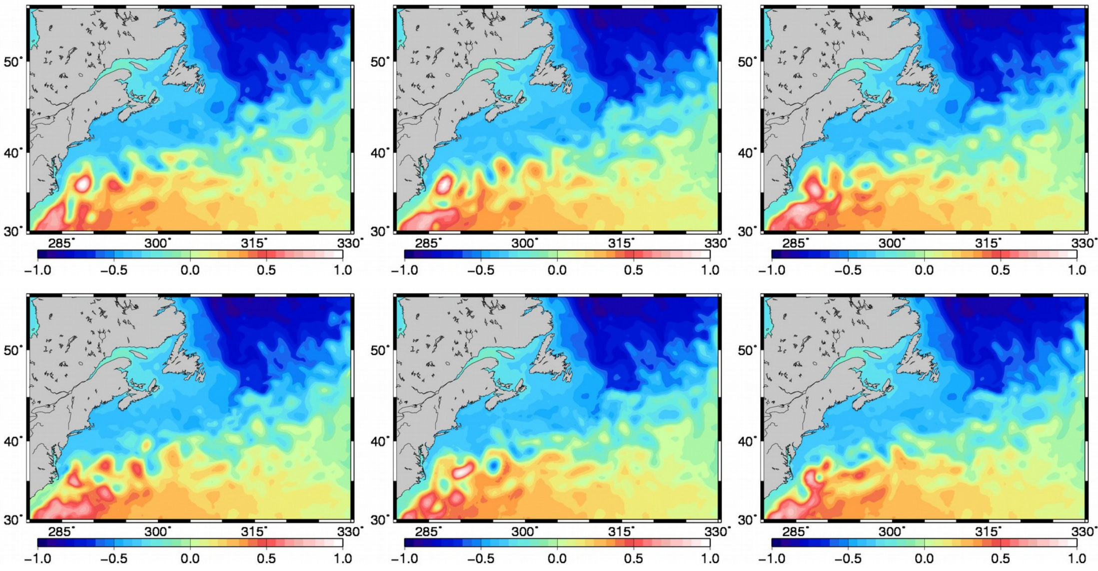
↓

Modification of air/sea interactions

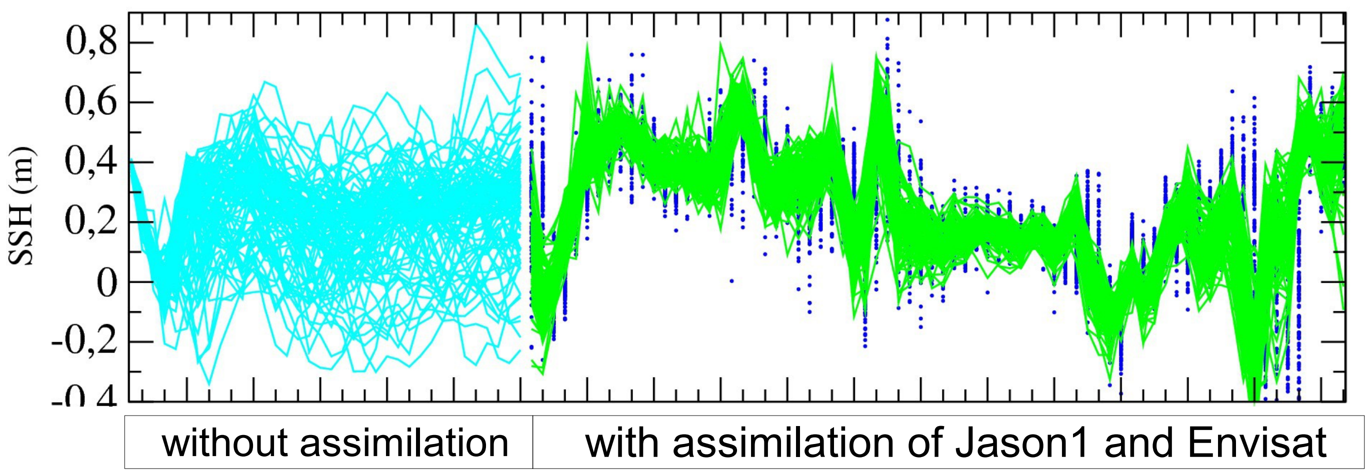
Ensemble of mesoscale flows

<u>Probability distribution for SSH</u>

as simulate here by ensemble NATL025 (large case SANGOMA benchmark): 6 members among 96



→ **assimilation of altimetric data (in SANGOMA, Candille et al., 2014)**



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

from June 2005 to December 2006

4

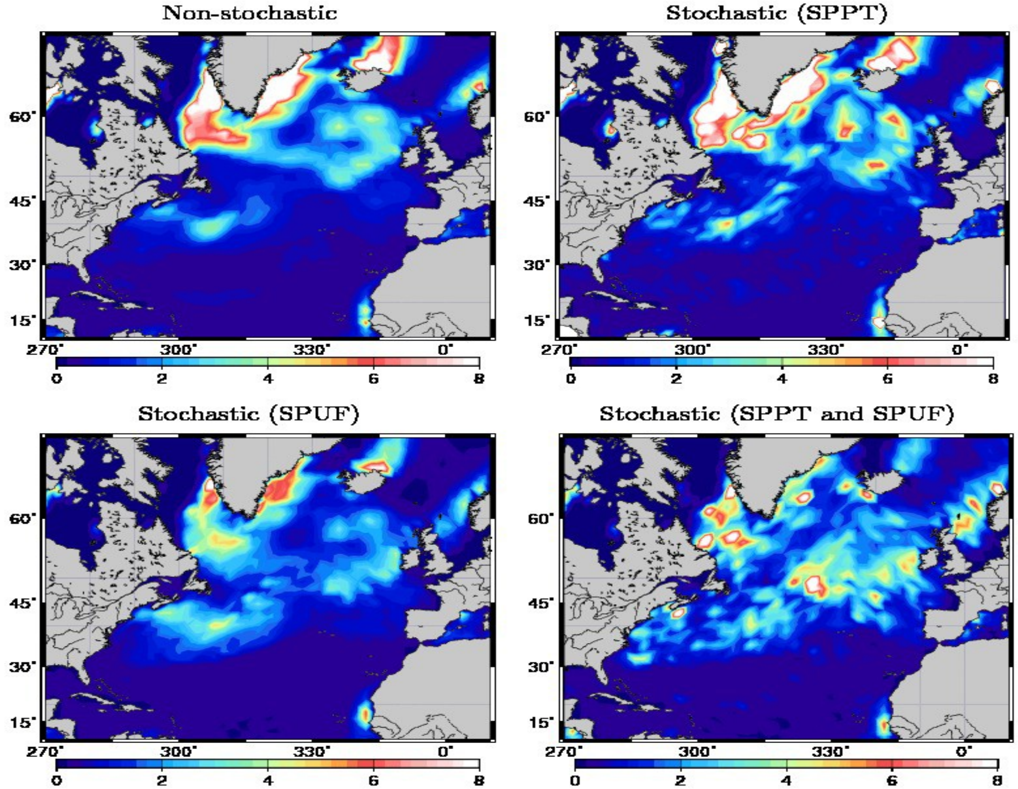
Stochastic ecosystem model

Stochastic ecosystem model

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

<u>Unresolved diversity</u>
multiplicative noise in the SMS terms of the model

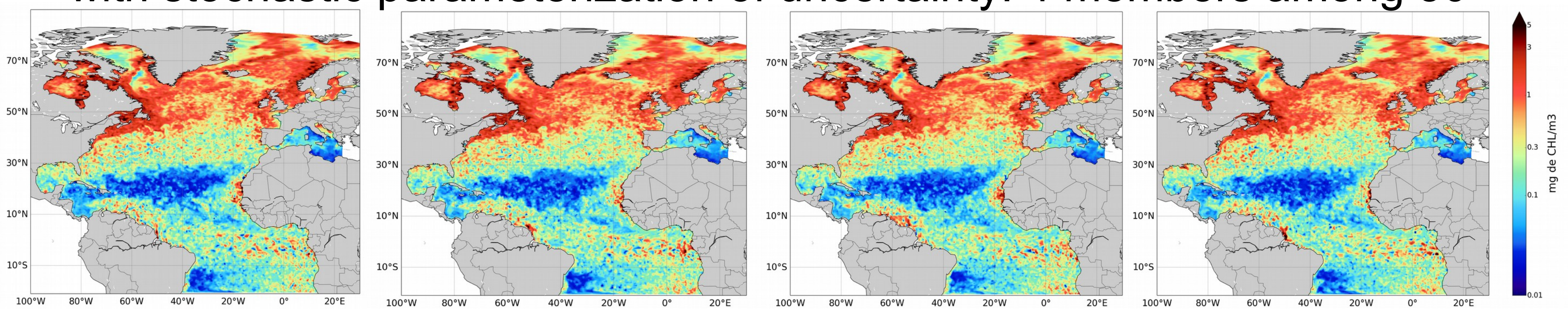
<u>Unresolved scales</u>
stochastic processes explicitly simulating unresolved fluctuations of C_i



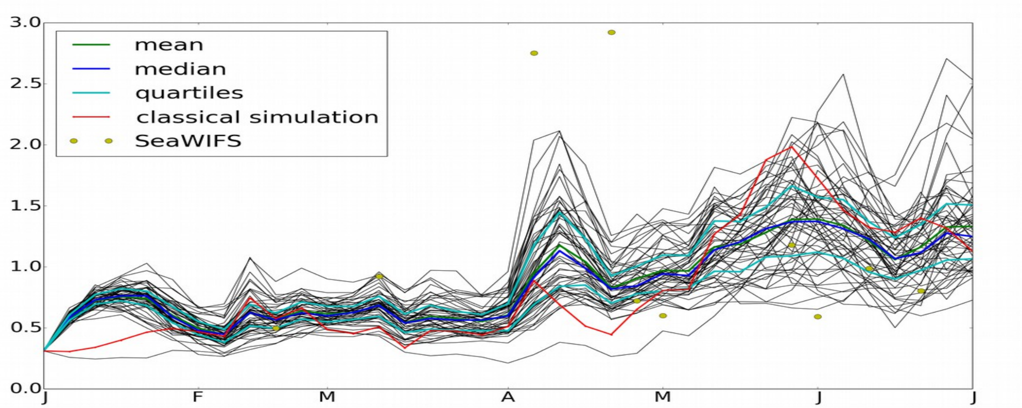
→ Considerable effect on the mean behaviour of the system
→ Increase of patchiness (↔ ocean colour data)

Ensemble simulation of the ecosystem

<u>Probability distribution of chlorophyll concentration</u>
as simulated here by ensemble NATL025/PISCES
(large case Sangoma becnhmark)
with stochastic parameterization of uncertainty: 4 members among 50



member 1 | member 2 | member 3 | member 4



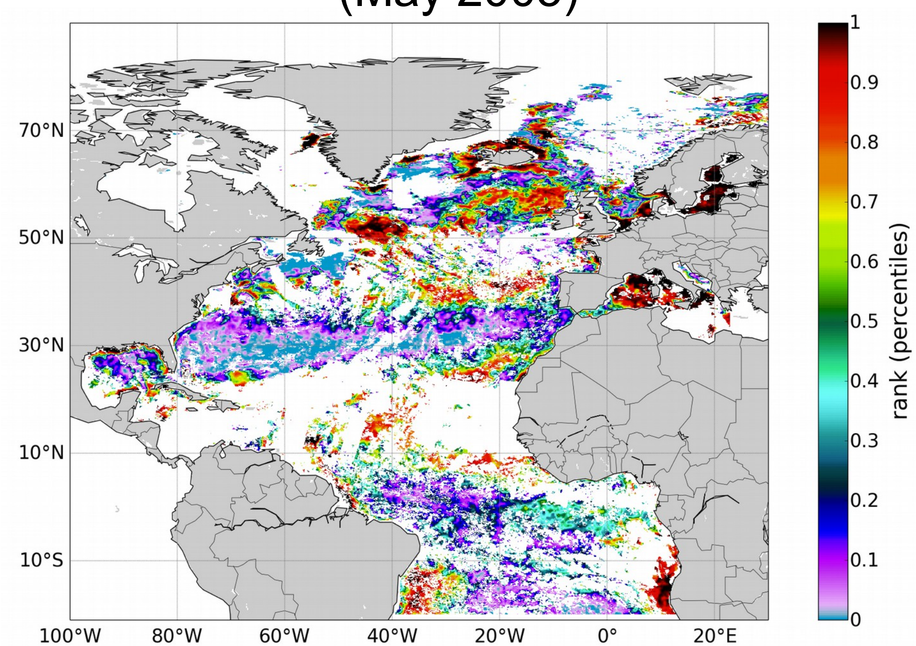
<u>Time evolution of the pdf for phtyoplankton</u>

from January to June 2005

→ assimilation of ocean colour observations (projects FP7-MyOcean2 and SANGOMA, Garnier et al., 2015)

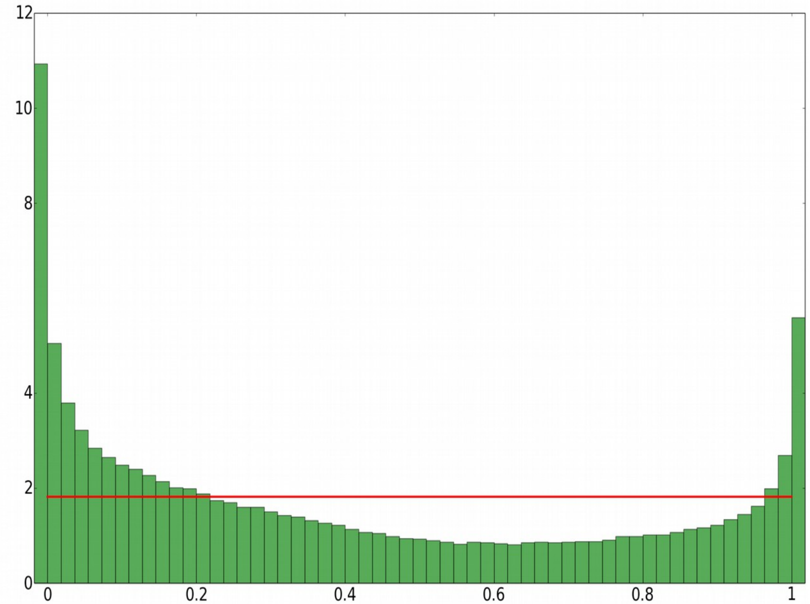
Comparison to ocean colour observations

Rank of SeaWifs observations
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)

Rank histogram for SeaWifs
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 consistency bewteen simulations and observations
→ prerequisite to ocean colour **data assimilation de données**

5

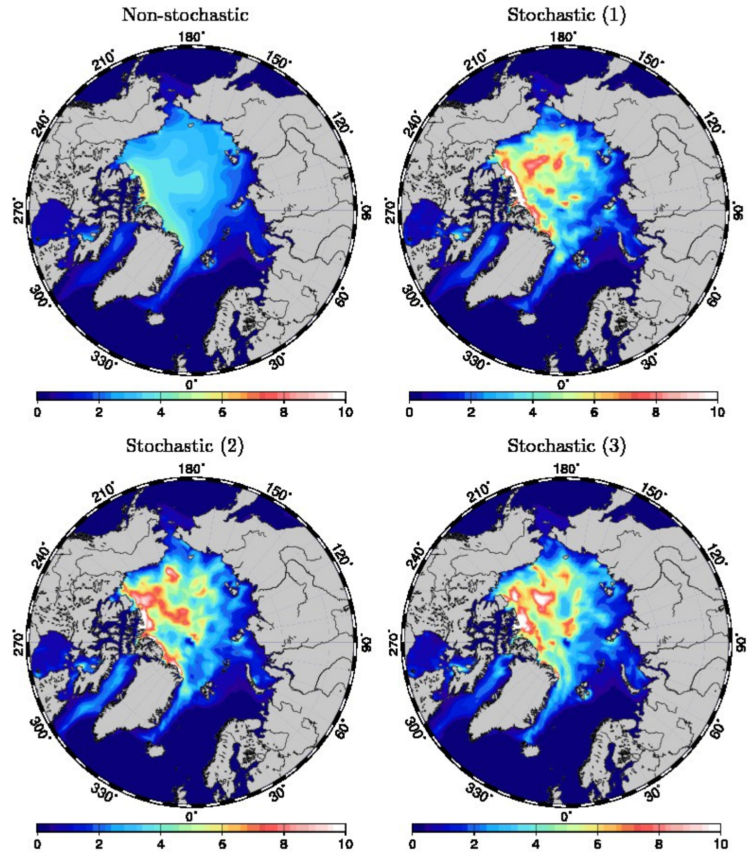
Stochastic sea ice model

Stochastic sea ice model

An important difficulty of sea ice model
is the unresolved diversity of dynamical sea ice behaviours

One of the most sensitive sea ice parameters is <u>ice strength</u> (P*):

multiplicative noise applied to P* (parametrization of Juricke et al., 2013, implemented in NEMO)



→ Considérable effect on the mean ice thickness (in ORCA2)
→ Intrinsic interannual variability is stimulated

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 key constituent of the systems
that we are using in oceanography,
not something that can be thought
separately from the results

Properly dealing with uncertainty will require
an integrated engineering approach
at the interface between oceanography
and applied mathematics