SANGOMA: Stochastic Assimilation for the Next Generation Ocean Model Applications
EU FP7 SPACE-2011-1 project 283580

Deliverable 6.4: Ph.D workshop 1 report

Due date: 31/10/2013

Delivery date: 30/04/2014

Delivery type: Report , Public

J.-M. Beckers A. Barth Y. Yan M. Canter
University of Liège, BELGIUM

P.-J. Van Leeuwen S. Vetra-Carvalho
University of Reading, UK

L. Nerger P. Kirchgessner
Alfred-Wegener-Institut, GERMANY

A. Heemink N. van Velzen M. Verlaan U. Altaf
Delft University of Technology, NETHERLANDS

P. Brasseur J.-M. Brankart G. Candille S. Metref
CNRS-LEGI, FRANCE

P. de Mey
CNRS-LEGOS, FRANCE

L. Bertino F. Counillon
NERSC, NORWAY

Chapter 1

Content

To be completed. In annex presentations and summary.

SANGOMA

Work Package 6
Outreach

on the MoorSPICE cruise
March 2014

Annual SANGOMA Meeting, April 2014

MoorSPICE cruise, 2014

Program : SPICE international program

Location : Salomon sea

Research Vessel : T.G. Thompson

<u>Main objectives</u> :

- Understand air-sea fluxes and oceanic currents in Salomon Sea
- Describe local and global effects of water mass transformations
- Estimate impacts of water redistribution from the subtropics to the equator and the Southern Ocean (e.g. ENSO modulations)

Workshop participants

Participants are specialized in *in situ* observations. This workshop is an introduction to data assimilation and SANGOMA's work.

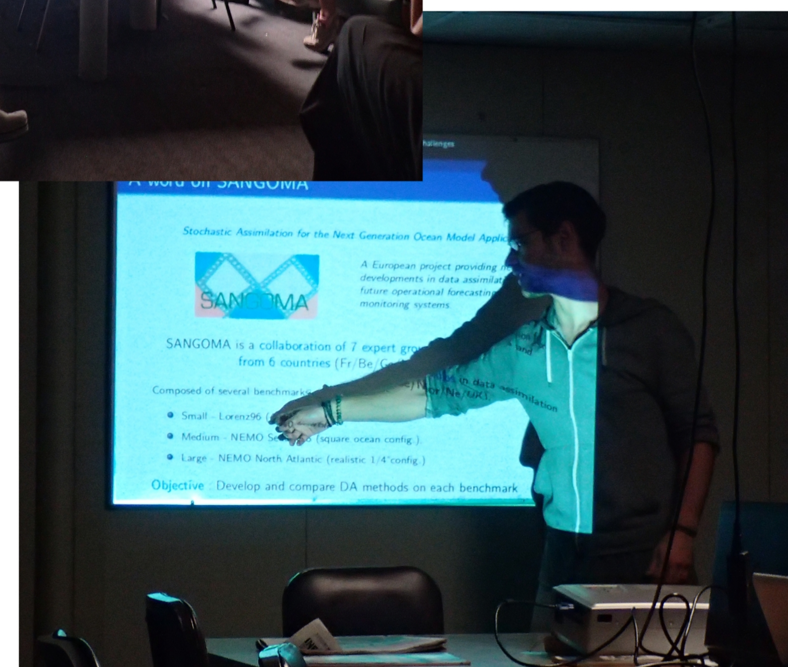
Participants :

- **2 chief scientists** (SCRIPS, San Diego ; IRD, Nouméa)
- **9 students** (University of Fiji ; University of Papua-New Guinea ; IRD, Nouméa ; SCRIPS, San Diego ; LEGOS, Toulouse ; MEOM, Grenoble)
- **5 engineers** (CNRS-INSU, Brest ; SCRIPS, San Diego ; IRD, Nouméa)
- **6 marine technicians** (University of Washington)

SANGOMA Workshop
Between observations and modelization

<u>Workshop part I</u> :

Sammy Metref



Introduction to data assimilation

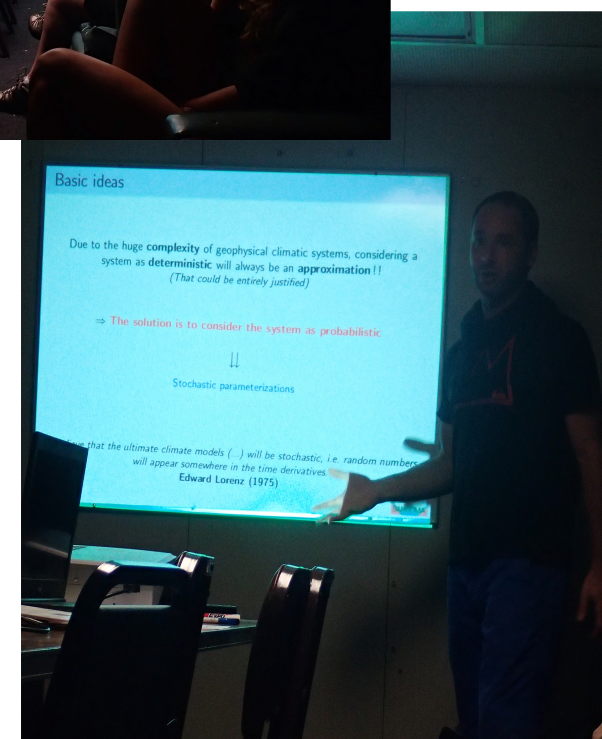
- Introduction and challenges in DA
- SANGOMA presentation
- Training (SANGOMA demo webpage)

SANGOMA Workshop
Between observations and modelization

<u>Workshop part II</u> :

Florent Garnier

Toward data assimilation :
The use of stochastic parametrizations



- Presentation of ensemble simulations
- Basics on stochastic parametrizations
- Results on large benchmark

Workshop SANGOMA - Part I
Between observations and modelization
Introduction to data assimilation

Sammy Metref, Florent Garnier

MEOM Team - LGGE, Grenoble, France
European project : SANGOMA

**MoorSPICE cruise
Noumea, March,
2014**

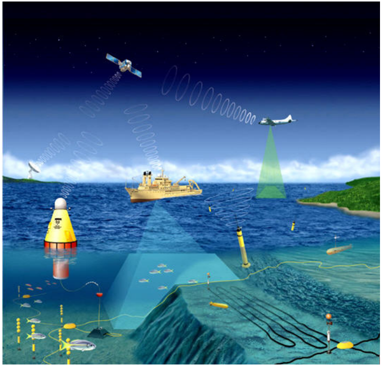
Laboratoire de Glaciologie et Géophysique de l'Environnement

What is data assimilation (DA) ?

Methods estimating a set of unknowns in a system by optimally combining different sources of information :

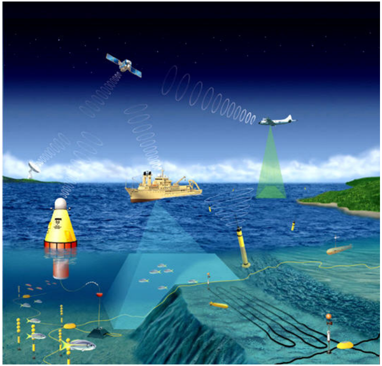
- model equations
- observations, data
- background, prior information
- uncertainties (statistics)

The observations



- Satellites
- Scientific cruises
- Voluntary merchant
- Moored buoys
- Surface drifters
- Argo profiling floats
- Gliders
- Tide gauges
- Sea mammals
- Airplanes

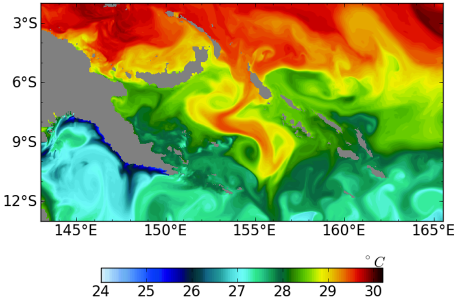
The observations



- Satellites
- Scientific cruises
- Voluntary merchant
- Moored buoys
- Surface drifters
- Argo profiling floats
- Gliders
- Tide gauges
- Sea mammals
- Airplanes

$\Rightarrow$ Observation errors = Measurement errors + Representation errors

The models (SST, 1/36th model [N. Djath])



Characteristics of a model :

- Fluid dynamic
- Thermodynamic
- Geochemistry
- Biology
- Numerical schemes

Error sources :

- Approx. in model equations
- Parameters
- Forcings
- Initial conditions
- Numerical discretization

What is DA for ?

Historically : meteorology. Later, oceanography.

Today, many other fields :

- glaciology,
- seismology,
- nuclear fusion,
- medicine,
- agronomy,
- etc

What are the goals of DA ?

Historically : initial state estimation, for weather forecasting.

Today, many other applications :

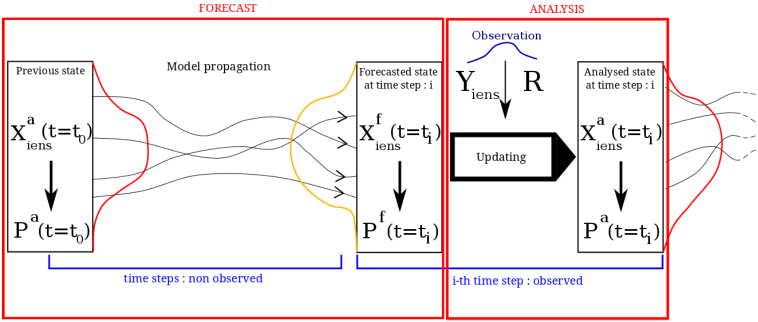
- initial conditions for predictions,
- calibration and validation,
- observing system design, monitoring and assessment,
- reanalysis,
- better understanding (model errors, data errors, physical process interactions, parameters, etc),
- etc

Two main approaches : Stochastic or Variational

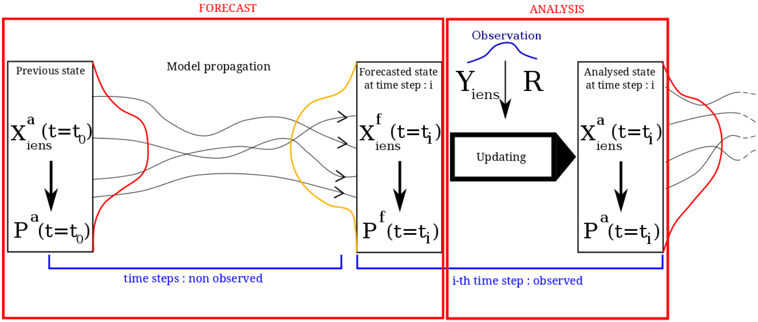
- Stochastic approach (e.g. EnKF) :
  $\rightarrow$ Optimal stochastic estimation theory
  $\rightarrow$ Sequential correction of the model state (at each observed time)
  $\rightarrow$ Minimizes the uncertainty of the estimated solution (in the least-squares sens)
  $\rightarrow$ Knowledge of the observation and background errors is needed

- Variational approach (e.g. 4D-var) :
  $\rightarrow$ Optimal control theory.
  $\rightarrow$ Minimization of a cost function J
  $\rightarrow$ J measures the quadratic distance between a set of ata and a trajectory of the model.
  $\rightarrow$ J can be controled by initial conditions, boundary conditions, model parameters ...

Stochastic scheme



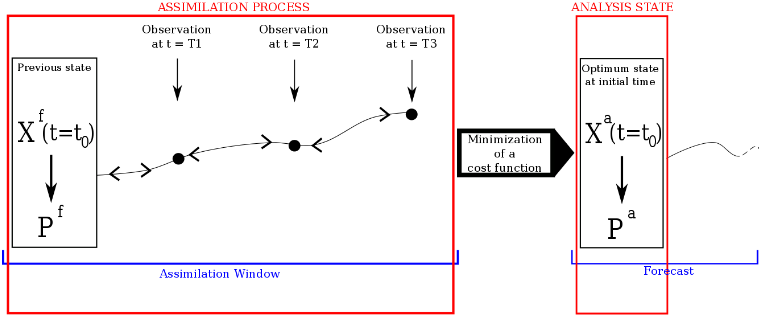
Stochastic scheme



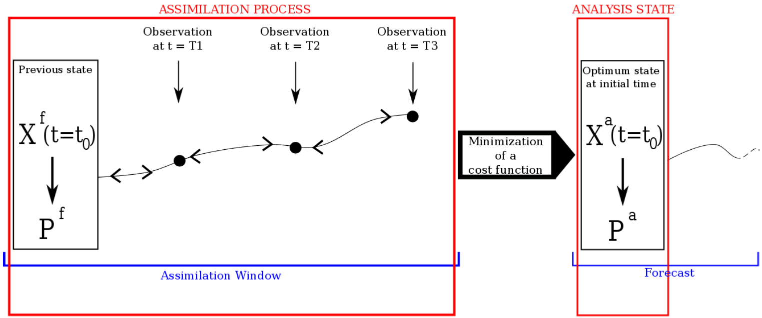
Drawbacks of classical sequential stochastic DA :

- Observation and background errors are assumed Gaussian
- Model and observation operator are assumed linear

Variational scheme



Variational scheme



Drawbacks of classical variational DA :

- Observation and background errors are assumed Gaussian
- Model needs to be linearized during minimization

SANGOMA : Data Assimilation Demo

A word on twin experiments : A methodological tool

- A simulation of reference = Truth
  $\rightarrow$ Diagnose the exact performance of the assimilation
- Creating observations from the truth
  $\rightarrow$ Controle the spat/temp repartition + observation error
- Creating background from the truth
  $\rightarrow$ Controle the background error

DA examples

DA new concerns

Data assimilation techniques are now used for a range of geophysical state estimation problems (e.g. land surface, ocean, atmospheric constituents)

and are applied to the atmosphere on scales from global to convective.

Coupling between the state components (e.g. ocean and atmosphere) has also become an important area of research.

Foreword and Symposium Summary
6th WMO data assimilation symposium

DA new concerns

Data assimilation techniques are now used for a range of geophysical state estimation problems (e.g. land surface, ocean, atmospheric constituents)
$\Rightarrow$ Nonlinear and non-Gaussian problems
and are applied to the atmosphere on scales from global to convective.
$\Rightarrow$ Parameter estimation, Multiscale DA, Adaptative grids ...
Coupling between the state components (e.g. ocean and atmosphere) has also become an important area of research.
$\Rightarrow$ Coupling DA

Foreword and Symposium Summary
6th WMO data assimilation symposium

SANGOMA Project

Stochastic Assimilation for the Next Generation Ocean Model Applications

A European project providing new developments in data assimilation for future operational forecasting and monitoring systems.

SANGOMA is a collaboration of 7 expert groups in data assimilation from 6 countries (Fr/Be/Ge/Nor/Ne/UK).

Composed of several benchmarks :

- Small - Lorenz96 (40 variables),
- Medium - NEMO SeaBASS (square ocean config.),
- Large - NEMO North Atlantic (realistic 1/4°config.)

Objective : Develop and compare DA methods on each benchmark

What about me ? - My interests

Data assimilation in a non-Gaussian context :
Methodology and applications to marine bio-geochemistery.

Objectives :

- Define and caracterize non-linearity and non-gaussianity
- Diagnose non-linearity and non-gaussianity in a system
- Methodologically compare DA methods
- Define/Caracterize/Diagnose/Compare on an experimental marine bio-geochemistery 1D model

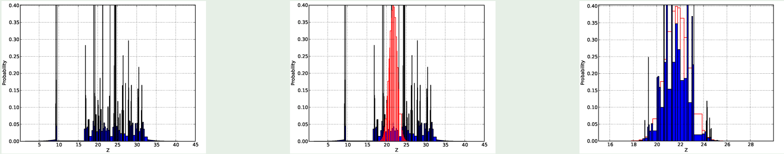
What about me ? - Methodological investigation

S. Metref, E. Cosme, C. Snyder, P.Brasseur

Creation of a new method with no Gaussian hypothesis !

The Rank Histogram Filter (Anderson, 2010) :

Bayes' theorem : $p(z|z^o) \propto p(z)p(z^o|z)$



$$p(z) \quad \times \quad p(z^o|z) \quad \propto \quad p(z|z^o)$$

Idea ⇒ Develop this concept to multidimensional problems : MRHF

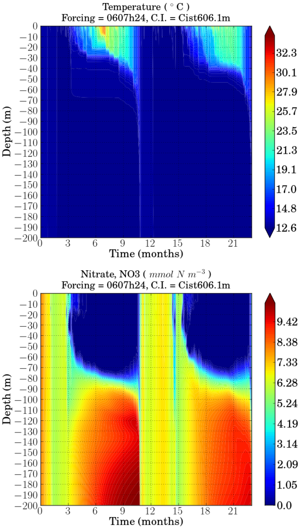
[Metref et al., in revision]

What about me ? - Biogeochemical adventures

S. Metref, P. Brasseur, E. Cosme, J.-M. Brankart, S. Grégorio

MODECOGel : 1D vertical marine biogeochemistery model (5 dyn. + 12 bio.)

- Experiment period : 2006/2007 (spinup+DA)
- Ensemble creation : Perturbation on wind intensity forcing
- Observations : T/S/Phyto/Ni profilers, ocean color, SST/SSS
- Control variables : T and S
- DA methods : EnKF, ETKF, RHF, MRHF, PF ...



Thank you !

Sammy Metref

LGGE/CNRS/SANGOMA

SANGOMA

Laboratoire de Glaciologie et Géophysique de l'Environnement

Workshop SANGOMA-Part II
Between observations and modelization

Towards data assimilation in a realistic configuration: the use of stochastic parametrizations

Florent Garnier, Sammy Metref, P.Brasseur, J-M.Brankart, J.Verron

MEOM Team-LGGE, GRENOBLE

Moor SPICE campaign
March 2014

Laboratoire de Glaciologie et Géophysique de l'Environnement

PLAN

1. Introduction : Quick overview of data assimilation in oceanography
   - Why ?
   - Where ?
   - Main objectives of the presentation
2. Toward a probabilistic system : presentation of stochastic parametrization concepts
   - Basic ideas
   - One exemple : Impacts of unresolved scales uncertainties
   - Definition of the system
3. My work : Stochastic parametrizations of biogeochemical uncertainties
   - The coupled Physical-biogeochemical model
   - Stochastic parametrizations of unresolved processes
     - Formulations
     - Some results
4. Conclusions

Why ?

As already introduced in the previous presentation data assimilation in Oceanography is mainly used for :

- The reconstruction of the past dynamics (Reanalysis)
  - **Better Understand the Physical/Biogeochemical processes**
  - Improving climatological studies
- The forecast :
  - Identification of initial conditions

2 Approaches :

- **Stochastic approach**
- Variational approach

Where ?

Scientific research

- Assimilation methods development for a 3D turbulent “green and blue“ ocean
  - Strongly non linear and non gaussian
  - With various scales phenomena

Operationnal developpment

- Export new oceanographic tools and data
- Match operationnal R&D needs

Long term

- Contribute to the comprehension of the earth system component evolutions in the climate perspective $\leftrightarrow$ observations
  - Satellite data (altimetry : Jason,SARAL/altika, ocean colour : SeaWIFS, Meris)
  - In situ data

Main objectives of the presentation

The more probable is the message, the less information it gives.
Cliches, for example, are less illuminating than great poems...
Norbert Wiener (1894-1964)

$$\Downarrow$$

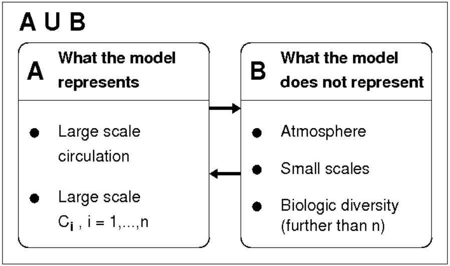
The use of **stochastic parametrizations** (i.e the inclusion of randomness in a model) bring some **statistical informations** on the distribution of the model **uncertainties** able to :

- Consider a system as **probabilistic** instead of **deterministic**
  $\Rightarrow$ Work with ensemble simulations
- Improve the **efficiency of data assimilation**

SANGOMA

Basic ideas

Stochastic parametrization must simulate processes not resolved →**We must identify the main sources of uncertainties**



- Even if the dynamic of **U** can be assumed deterministic, the system **A** alone cannot be assumed **deterministic**
- To consider **A** as deterministic we assume that :
  - **B** is know (e.g atmospheric forcing)
  - the effects of **B** can be parameterized (e.g effects of unresolved scales)

⇒ **B is the main source of uncertainty in the model**

Basic ideas

Due to the huge **complexity** of geophysical climatic systems, considering a system as **deterministic** will always be an **approximation** ! !
(That could be entirely justified)

$\Rightarrow$ **The solution is to consider the system as probabilistic**

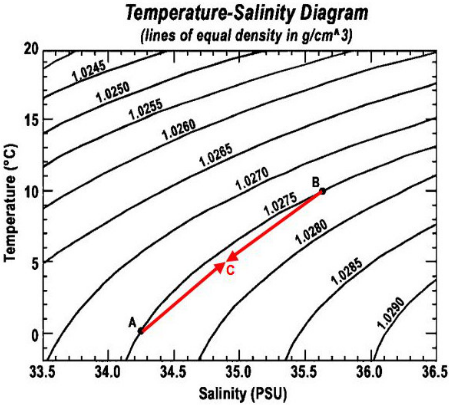
$\Downarrow$

Stochastic parameterizations

I believe that the ultimate climate models (...) will be stochastic, i.e. random numbers will appear somewhere in the time derivatives.
Edward Lorenz (1975)

Exemple of unresolved scales uncertainty : the computation of the density

In the model, the large scale density is computed from large scale temperature and salinity, using the sea-water equation of state $\rho(T, S, p)$



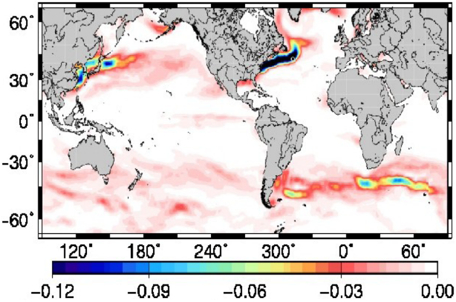
Averaging T and S on the equation systematically overestimates the density. Know in reality as the cabella effect

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

SANGOMA

Exemple of unresolved scales uncertainty : the computation of the density

From a T&S reananlysis data at $1/4^\circ$ we use an averaging operator to downscale data to 2° resolution



The figure (Brankart et al, 2013) present the density misfit ($\delta\rho$) between applying the averaging operator **before and after** the equation of state.
$\delta\rho = \overline{\rho(T,S)} - \rho(\overline{T},\overline{S})$

This unresolved scale uncertainty is only **mathematical** issued from the non linearity of the equation.

→ The **stochastic solution** is to consider $\rho(\overline{T} + \delta\overline{T}, \overline{S} + \delta\overline{S})$

Definition of the system

Objective : Transform a deterministic model into a probabilistic model

- Describe the non-deterministic nature of the system
- Allow objective comparison with observation
- Introduce a weak model constraint in data assimilation systems

Method : explicitely simulate model uncertainties using random numbers

- Propose a generic and flexible technical approach
- Develop a first simple implementation
- • atmospheric forcing
  • unresolved scales
  • unresolved diversity

SANGOMA

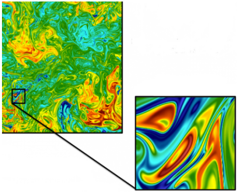
Definition of the system

- With this stochastic methods, the prior error covariance matrix ($P^a(t = t_0)$ in the previous presentation) is directly related to **the model formulations** instead of the only sensivity of initial conditions

- We expect a more consistent evaluation of model uncertainties
  $\Rightarrow$ **improve data assimilation**

SANGOMA

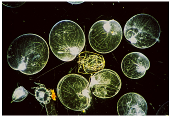
General idea about some of my work



The Idea is to implement **stochastic** parametrization in the **NEMO/PISCES** configuration to simulate some of the main biological unresolved processes.
The results of this simulations would allow to perform assimilation of ocean colour data (SeaWifs)

I focused on 2 main points :

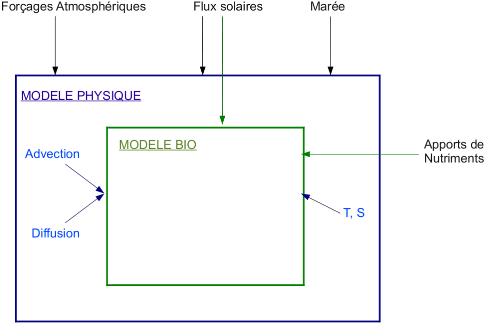
- The non linearity of biological equations (small scale effects)
- The uncertainty of biological formulations (diversity of species, biological adaptation)



As one simulation including random numbers can only represent one possible state of the system we perform ensemble simulations.

The Physical-Biogeochemical coupling

- The configuration used for the study is the 1/4° NEMO/PISCES coupled model



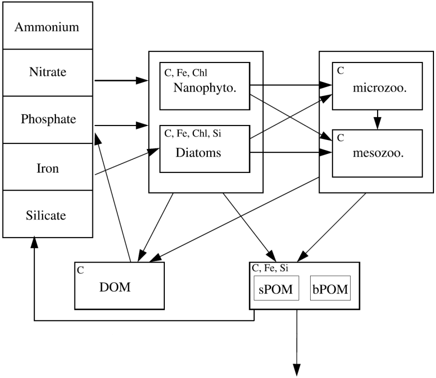
- The biogeochemical variables are tracers forced on-line by the physical model
- There is no biological feedback on the physics

$$\frac{\partial C_i}{\partial t} = \overbrace{-\nabla.(u.C_i)}^{Advection} - \underbrace{A_\ell.\nabla^2 C_i}_{\text{Diffusion Horizontale}} + \overbrace{\frac{\partial}{\partial z}(K_\ell.\frac{\partial C_i}{\partial z})}^{\text{Vertical Diffusion}} + \underbrace{SMS(C_i)}_{Sources-Puits}$$

SANGOMA

The PISCES model

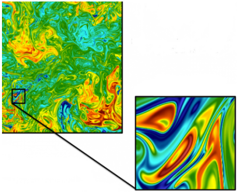
The PISCES model (Pelagic Interaction Scheme for Carbon and Ecosystem Studies)



The PISCES model [*Aumont, 2006*]

- 24 state variables
  - 4 living species
    - 2 Phytoplanktons
    - 2 Zooplanktons
  - 5 limiting nutrients
    - NH4, NO3, Fe, SI, PO4
  - 3 non living compartments
    - semilabile DOM
    - 2 organic particles

Stochastic parametrizations of unresolved processes : formulation



Fluxes perturbation :

$$\frac{\partial C}{\partial t} \rightarrow SMS(C, u, p, t) = \sum_{i=1}^{\alpha} F_i(C, u, p, t)$$

$$\equiv \sum_{i=1}^{\alpha - n} F_i(C, u, p, t) + \sum_{i=\alpha-n}^{\alpha} F_i(C, u, p, t).\xi_i(t)$$

Unresolved scales perturbations :

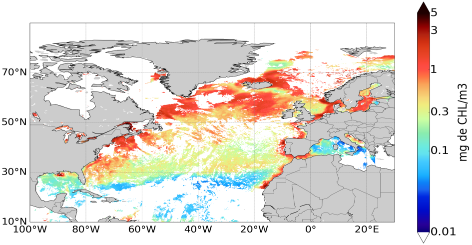
$$SMS(\overline{C}) \neq \overline{SMS(C)} \Rightarrow \frac{\partial C}{\partial t} = \frac{1}{m}.\sum_{i=1}^{m} SMS(\overline{C} + \delta \overline{C}_i, u, p, t) \quad \text{with}$$

$$\delta C \rightarrow \xi(t).C \quad \textit{and} \quad \sum_{i=1}^{m} \delta C_i = 0$$

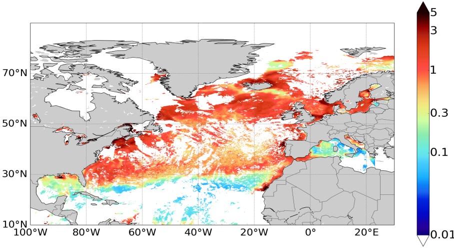
SANGOMA

Stochastic parametrizations of unresolved processes : Some results

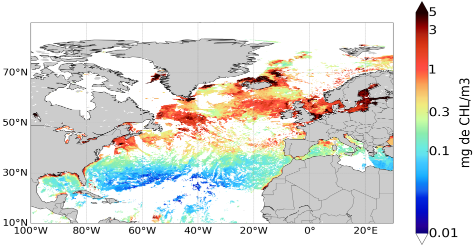
We perform a **14** member ensemble simulation with the coupled uncertainties



Minimum surface chlorophyll of the ensemble, 15 May 2005



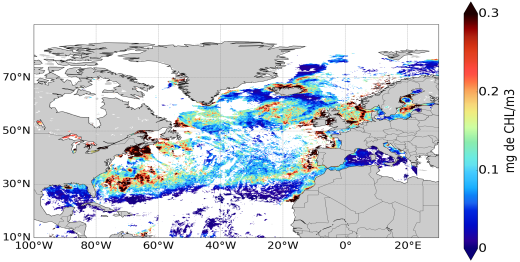
SeaWiFS surface chlorophyll, 15 May 2005



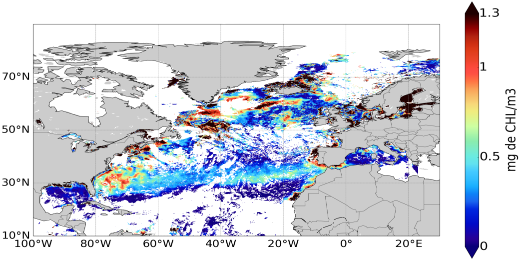
Maximum surface chlorophyll of the ensemble, 15 May 2005

- An important, **well spread dispersion** is generated
- Over most of the domain, the ensemble spread **includes the observations**

Stochastic parametrizations of unresolved processes : Some results



Surface chlorophyll standard deviation of the ensemble, 15 May 2005



Surface anomaly between the SeaWiFS data and the ensemble mean, 15 May 2005

Strong spatial **coherence** between the patterns of higher anomalies and the maximum of standard deviations

$\Rightarrow$ **Dispersion is higher where anomalies are stronger**

SANGOMA

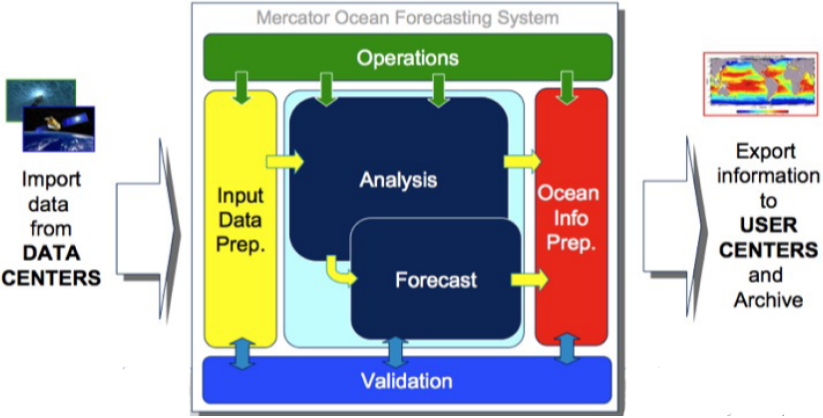
Conclusions

- I hope i convinced you that people in modelling don't think their model are the perfect truth and that they might also be interested by the observations (at least for data assimilation).....
- Taking into account the uncertainties (even with random numbers) allows to bring informations and then to improve model representations.
- There is a strong link between the knowledge of model uncertainties and the efficiency of data assimilation
  $\rightarrow$Stochastic parametrizations are a relevant tool for this
- The future of oceanography (for modellers) is probably with ensemble simulations and high resolution simulations ! !

SANGOMA

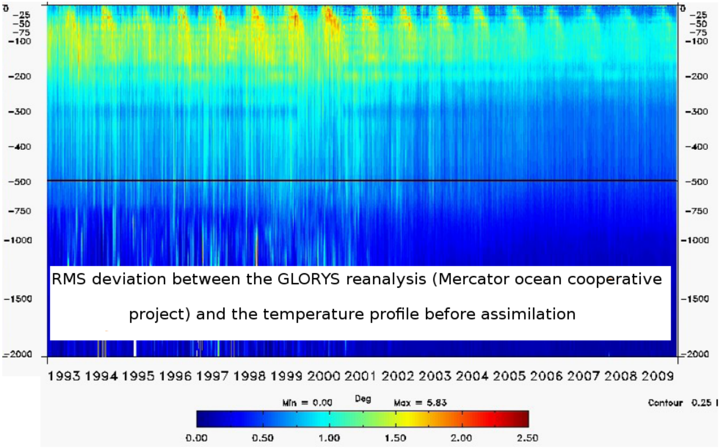
Thanks for your attention

Any questions ?


Import data from DATA CENTERS
Mercator Ocean Forecasting System
Operations
Input Data Prep.
Analysis
Forecast
Ocean Info Prep.
Validation
Export information to USER CENTERS and Archive

Assimilation T/P, ERS, Envisat, Jason,GFO (SSALTO/DUACS)

+ ARGO



RMS deviation between the GLORYS reanalysis (Mercator ocean cooperative project) and the temperature profile before assimilation