

Traitement des incertitudes en océanographie

Jean-Michel Brankart
Ingénieur de Recherche CNRS

Laboratoire de Glaciologie et de Géophysique de l'Environnement

Equipe de Modélisation des Ecoulements Océaniques Multi-échelles



Traitement des incertitudes en océanographie

1

Introduction

2

Incertitudes sur les modèles

3

Incertitudes sur les observations

4

Réduction des incertitudes

5

Conclusions

1

Introduction

L'équipe MEOM: une approche intégrée de la recherche en océanographie

Observations

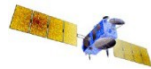


In situ :

profils T/S, courants

Spatiales :

Altmétrie, couleur de l'eau,
SST, SSS, vents....



OSTST / SARAL-Altika

Modèles

Modèles simplifiés,
aux équations primitives



Assimilation

Filtres de Kalman
Méthodes variationnelles

MERCATOR / NEMO-ASSIM

Nos motivations d'océanographe

Décrire

le comportement de l'océan

Comprendre

les mécanismes sous-jacents

Prévoir

l'évolution du système océanique

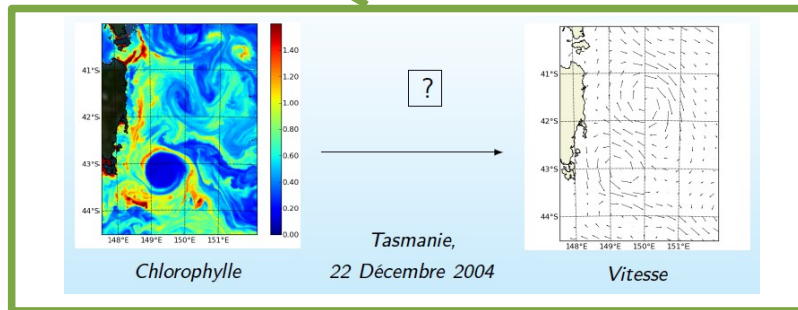
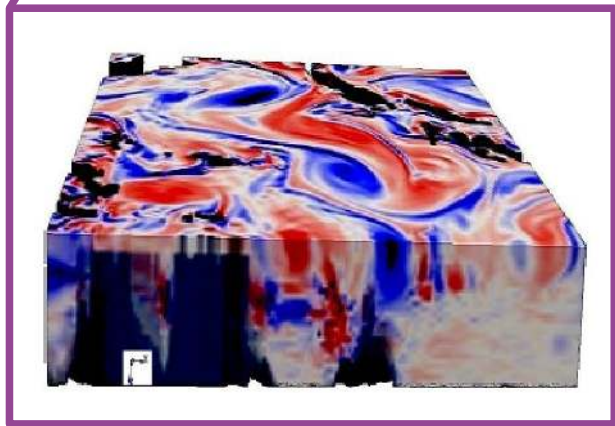
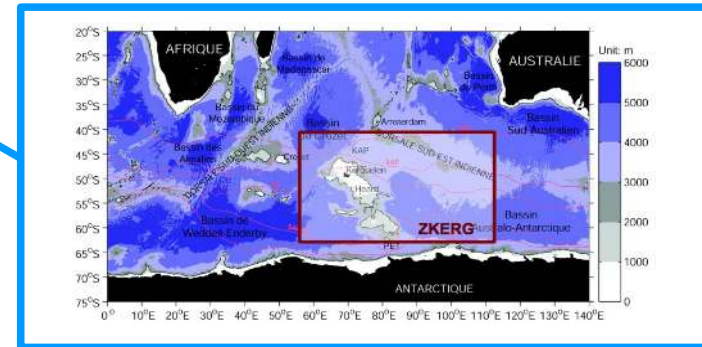
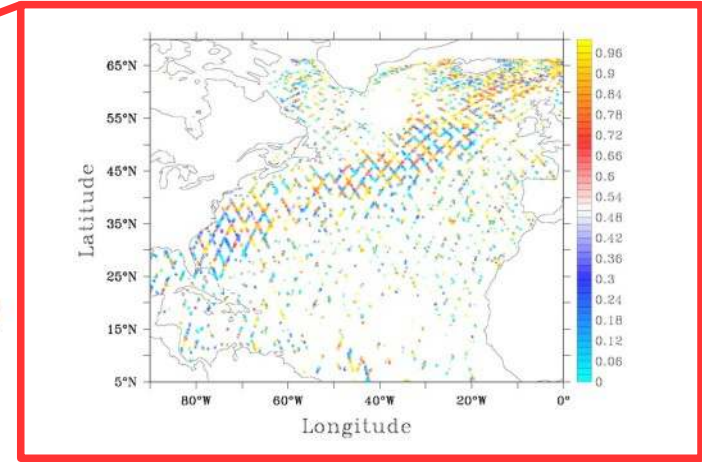
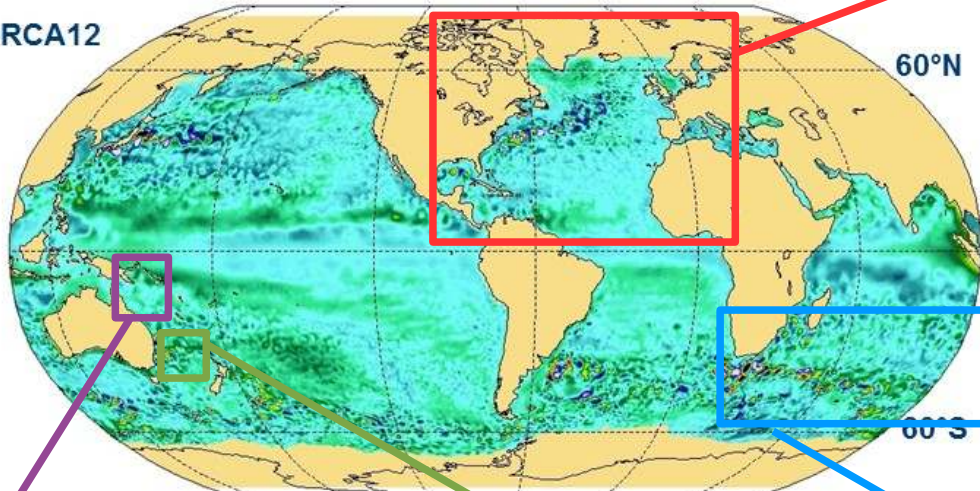
Océanographie numérique



→ systèmes de simulation
de l'océan intégrant
modèles, observations,
assimilation de données

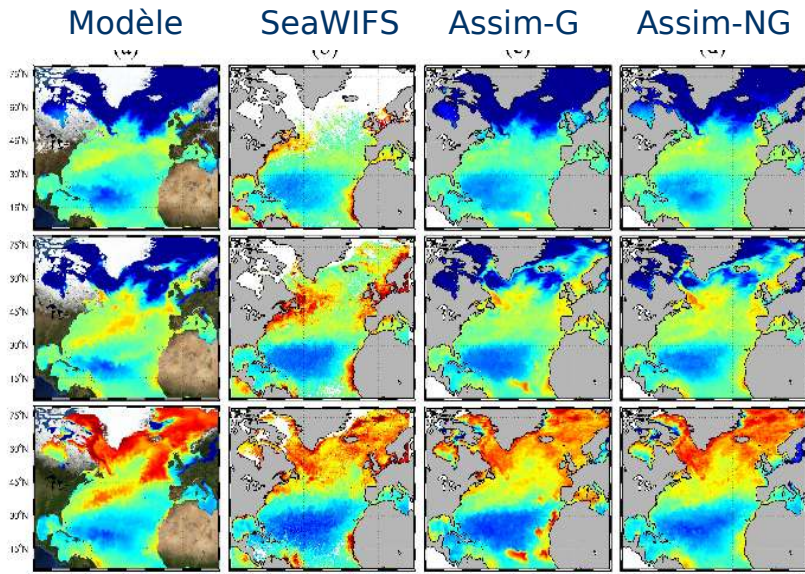
De l'océan global aux laboratoires régionaux

ORCA12

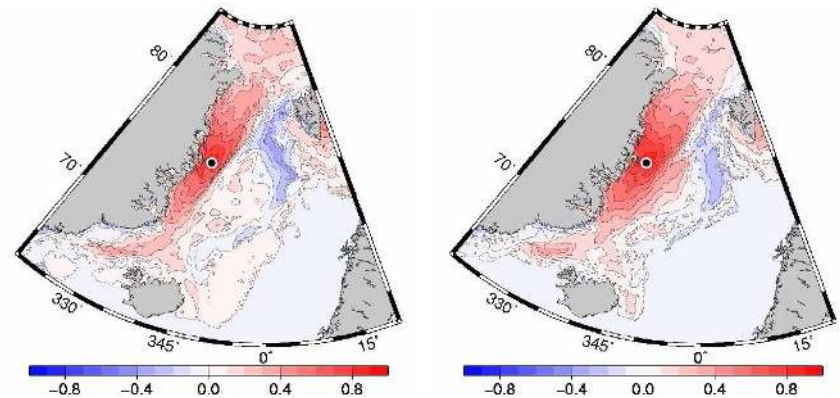


De l'océan bleu à l'océan vert et blanc

Assimilation de données couleur de l'eau SEAWIFS dans un modèle couplé circulation/écosystème



Vers l'assimilation de données dans un modèle de glace de mer



Ingénieur de recherche en calcul scientifique

**Ma place
dans l'équipe
MEOM**

**Mes
responsabilités**

**Développements méthodologiques
et techniques**

pour l'assimilation de données

**Encadrement méthodologique,
technique, scientifique**

étudiants en thèse, postdoctorants

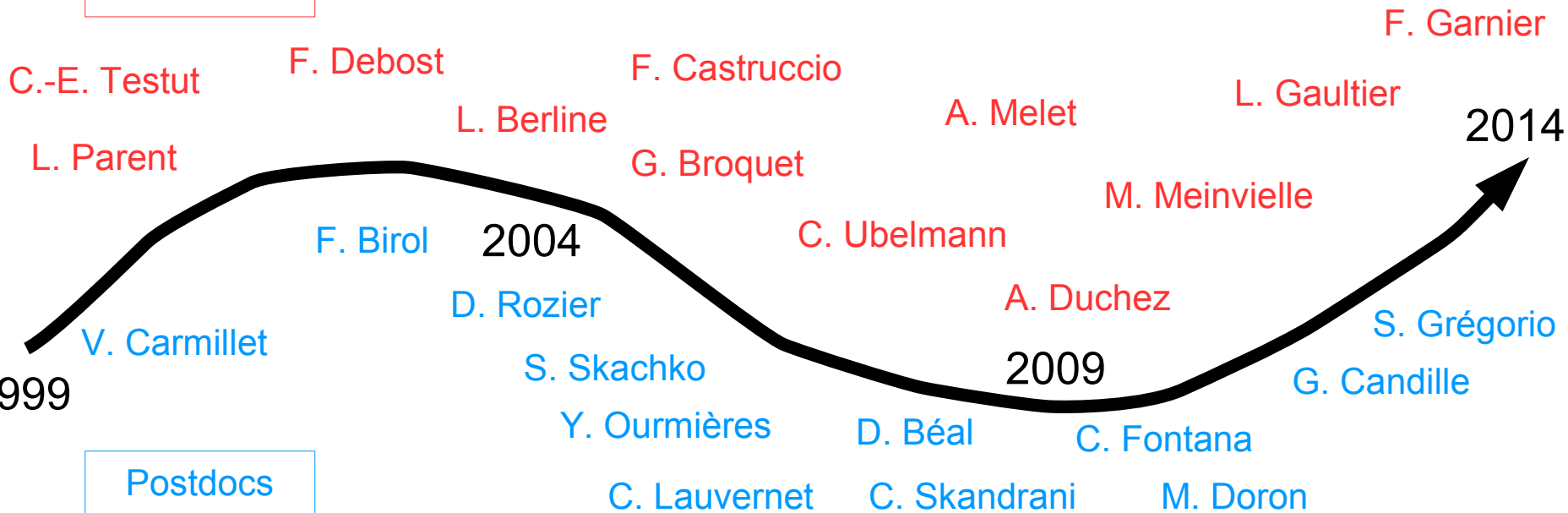
Participation aux projets de recherche

expertise scientifique et technique

Avec qui et pour qui ?

**Au sein de l'équipe MEOM (au LEGI puis au LGGE),
sous la direction de Jacques Verron et Pierre Brasseur**

Thèses



Applications du filtre SEEK

Carmillet et al., 2001
Testut et al., 2003
Brankart et al., 2003
Birol et al., 2004, 2005

Extension ensembliste,
estimation du forçage

Broquet et al., 2008
Skachko et al., 2009
Skandrani et al., 2009

Extensions
non-gaussiennes

Lauvernet et al., 2009
Béal et al., 2010
Brankart et al., 2012
Gaultier et al., 2013

Incertitudes
sur le modèle

Doron et al.,
2011, 2013
Brankart, 2013

Dans quel cadre ?

1998

Filtre SEEK

projets DIADEM/TOPAZ
collaboration MERCATOR
développement de SESAM

2004

Estimation de forçages et de paramètres

projet MERSEA

Simulation de systèmes d'observation

missions altimétriques

2008

Extensions non-gaussiennes

projets MyOcean/Sangoma
gaussiennes tronquées,
anamorphose,
chaînes de Markov

Incertitudes sur les modèles

projets MyOcean2/Sangoma
modélisation probabiliste,
paramétrisations
stochastiques

2015

2013

Les 3 règles de l'ingénieur-système

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

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.

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

(d'après Richard Hamming, 1997: The Art of Doing Science and Engineering)



Imaginer des méthodes flexibles et assez génériques pour traiter les problèmes sur le long terme.

Traiter nos systèmes comme des systèmes complexes faits de composants incertains et imparfaits.

L'incertitude, en tant qu'élément constitutif
des systèmes de simulation de l'océan

**Quels sont les composants incertains
dont nos systèmes sont faits ?**

**Comment décrire
ces incertitudes ?**

**Comment cela participe-t-il
à la résolution de problèmes inverses ?**

Comment je vais discuter de ces questions
au cours de cet exposé

→ **Essayer de donner un point de vue synthétique,**
sous un angle surtout méthodologique, non-chronologique

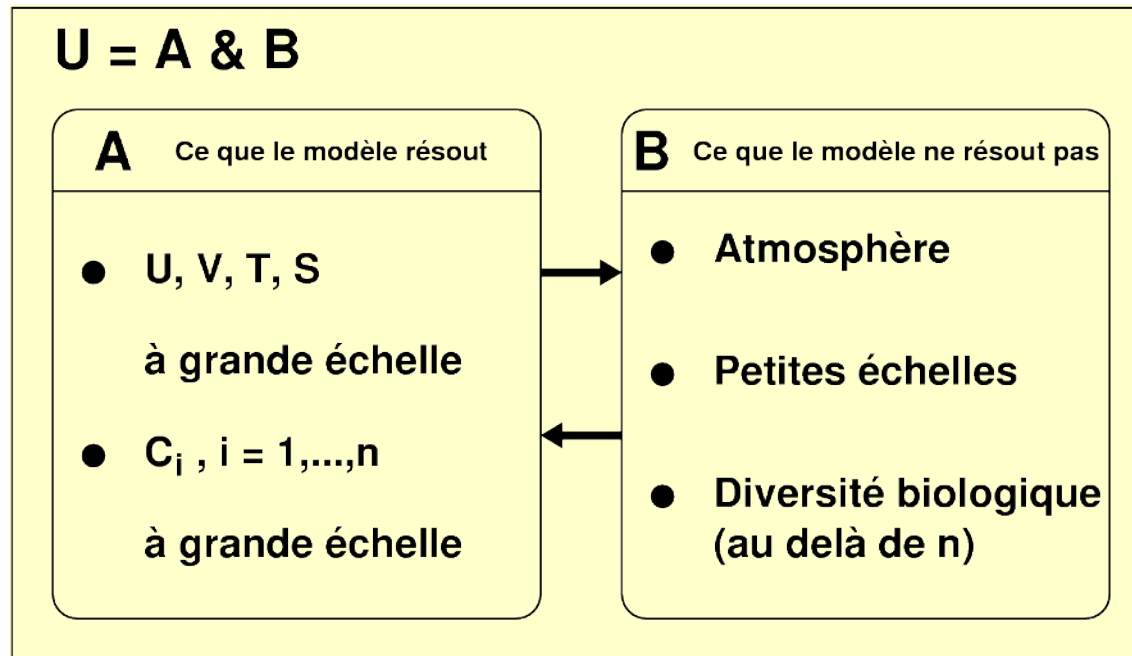
→ **Montrer l'aspect transverse de ces questions,**
entre modélisation et assimilation de données,
pour les océans bleu, vert, et blanc

→ **A travers ce qu'on a fait dans l'équipe,**
sans pourtant entrer dans le détail
des différentes applications

2

Incertitudes sur les modèles

Origines des incertitudes sur le modèle



- Même si la dynamique de **U** peut-être supposée déterministe, le système **A** seul **ne peut être supposé déterministe**.
- Pour obtenir un modèle déterministe pour **A**, il faut supposer soit, que **B** est connu (→ forçage atmosphérique), soit, que l'effet de **B** peut être paramétré (→ paramétrisation des échelles ou de la diversité biologique non-résolues).

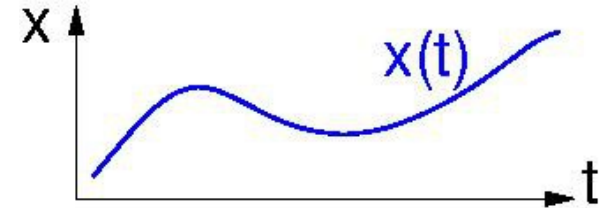
→ **B est la principale source d'incertitude dans le modèle.**

Une approche probabiliste de modélisation océanique

Dynamique océanique stochastique, explicitant les incertitudes

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

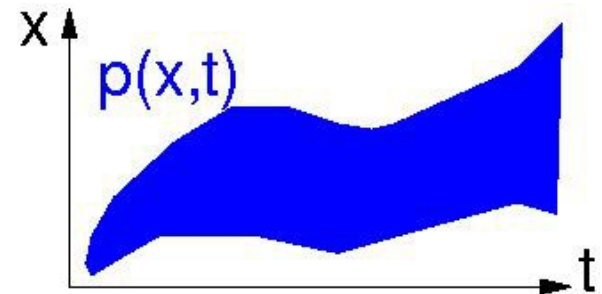
$$\text{où } \mathbf{x} = [x_1, \dots, x_N]$$



Equation de Fokker-Planck, pour la distribution de probabilité $p(\mathbf{x}, t)$,
en suivant les idées à l'origine du filtre de Kalman d'ensemble (Evensen, 1994)

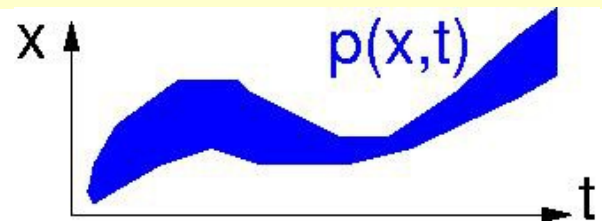
$$\frac{\partial p(\mathbf{x}, t)}{\partial t} = - \sum_{i=1}^N \frac{\partial}{\partial x_i} [\mathcal{M}_i(\mathbf{x}, t)p(\mathbf{x}, t)] \\ + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \frac{\partial^2}{\partial x_i \partial x_j} [D_{ij}(\mathbf{x}, t)p(\mathbf{x}, t)]$$

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



Conditionnement aux observations, pour réduire les incertitudes

Par une méthode d'assimilation
de données appropriée
→ voir la troisième partie



2.1 Formulation stochastique de NEMO

**Objectif: transformer le modèle *déterministe*
en un modèle *probabiliste***



**Décrire la nature
non-déterministe
du système**



**Permettre une
comparaison
objective
aux observations**



**Introduire une
contrainte faible
pour l'assimilation
de données**

**Méthode: simuler explicitement l'incertitude sur le modèle
en utilisant des *nombres aléatoires***



**Proposer une
approche technique
générique et flexible**



**Développer une
première
implémentation
simple**



**Forçage extérieur
Echelles non-résolues
Diversité non-résolue**

Processus autorégressifs (1)

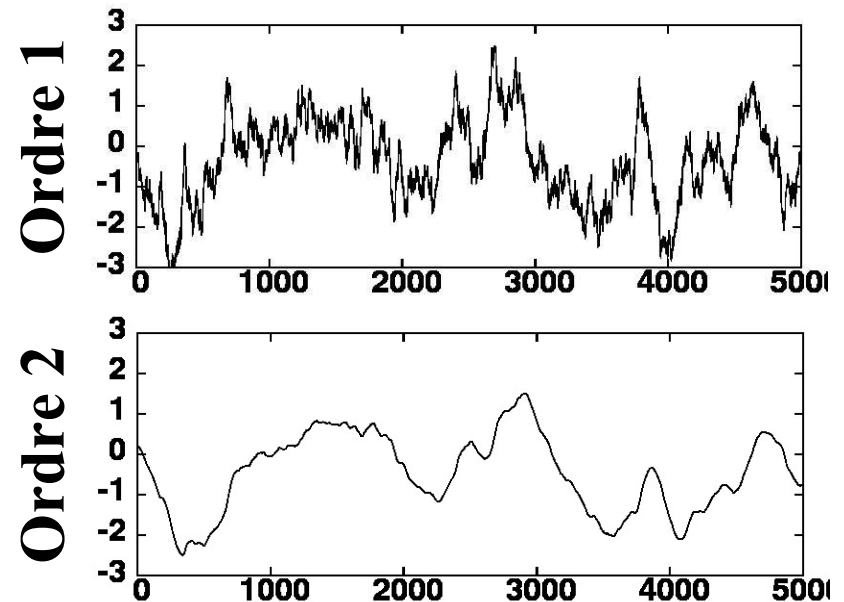
En chaque point de grille du modèle (2D ou 3D), on génère des **processus autorégressifs gaussiens et indépendants**:

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

où w est un bruit blanc gaussien (\rightarrow processus d'ordre 1)
ou un processus d'ordre $n-1$ (\rightarrow processus d'ordre n)

Paramètres a , b , c
pour spécifier:

moyenne, écart-type
et échelle de corrélation



Processus autorégressifs (2)

Introduire une structure de corrélation spatiale

en appliquant un filtre spatial aux cartes de processus autorégressifs:

$$\tilde{\xi} = \mathcal{F}[\xi] \text{ (opérateur de filtrage)}$$

$$\mathcal{L}[\tilde{\xi}] = \xi \text{ (équation elliptique)}$$

qui peuvent être rendus dépendants de l'écoulement si nécessaire

Modifier les distributions de probabilité marginale

en appliquant une transformation anamorphique à chaque variable gaussienne:

$$\tilde{\xi} = \mathcal{T}[\xi] \text{ (fonction non-linéaire)}$$

par exemple pour transformer les variables gaussiennes en variables lognormales ou gamma si elles doivent être positives

→ Ceci procure un moyen générique d'implémenter un large spectre de paramétrisations stochastiques

Exemple 1: Perturbation stochastique des tendances paramétrées

Avec cette implémentation générique, nous pouvons reproduire le schéma SPPT proposé par Buizza et al. (1999)

Séparer l'opérateur modèle
en NP (non-paramétré) et P (paramétré)

Supposer que P est incertain et
modéliser l'incertitude par un bruit multiplicatif ξ

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

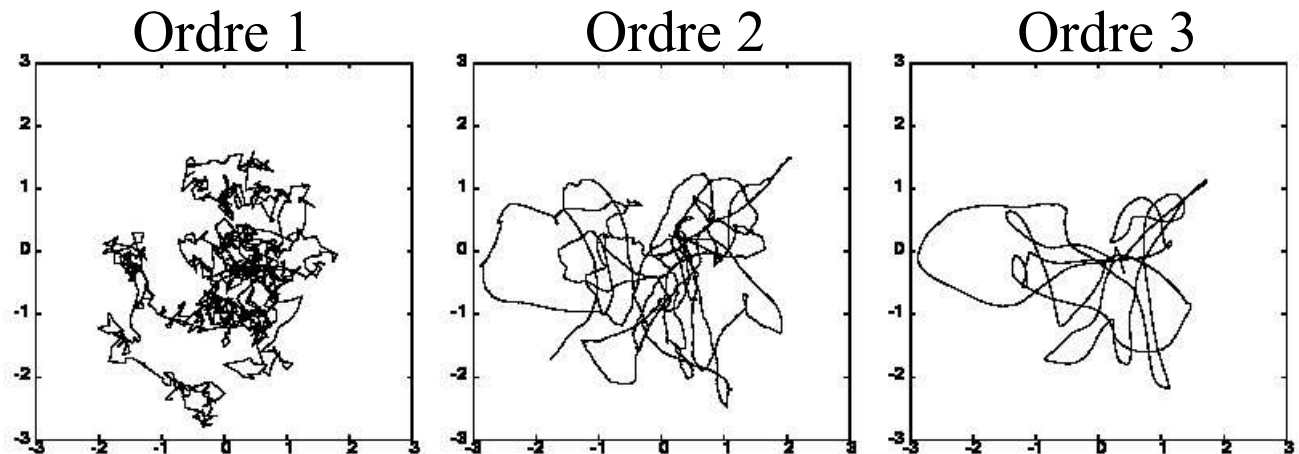
→ Utiliser des cartes de processus autorégressifs comme ξ
(de moyenne 1), en spécifiant la structure de corrélation
et la distribution marginale de ξ .

Exemple 2: Paramétrisation stochastique des fluctuations non-résolues

Pour simuler l'effet des fluctuations non-résolues dans les termes non-linéaires des équations du modèle

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

Générer des fluctuations à partir de **marches aléatoires** autour de chaque point de grille



→ Utiliser des cartes de processus autorégressifs pour les composantes x, y, z des marches aléatoires. Spécifier la structure de corrélation spatio-temporelle.

Exemple 3:. Paramétrisation stochastique de la diversité non-résolue

Pour simuler l'effet de la diversité non-résolue des comportements du système (par ex. la diversité biologique,...)

Cela revient à supposer que le système inclut simultanément une variété de comportements possibles, qui ne peuvent être décrits par une seule valeur de chaque paramètre.

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

Par exemple, l'écosystème contient en général de nombreuses espèces différentes de phytoplancton ou de zooplancton, chacun avec son propre comportement, tandis que le modèle ne peut résoudre que quelques classes d'espèces.

→ Utiliser des cartes de processus autorégressifs ξ comme bruit multiplicatif pour les paramètres, en spécifiant leur structure de corrélation et leur distribution marginale.

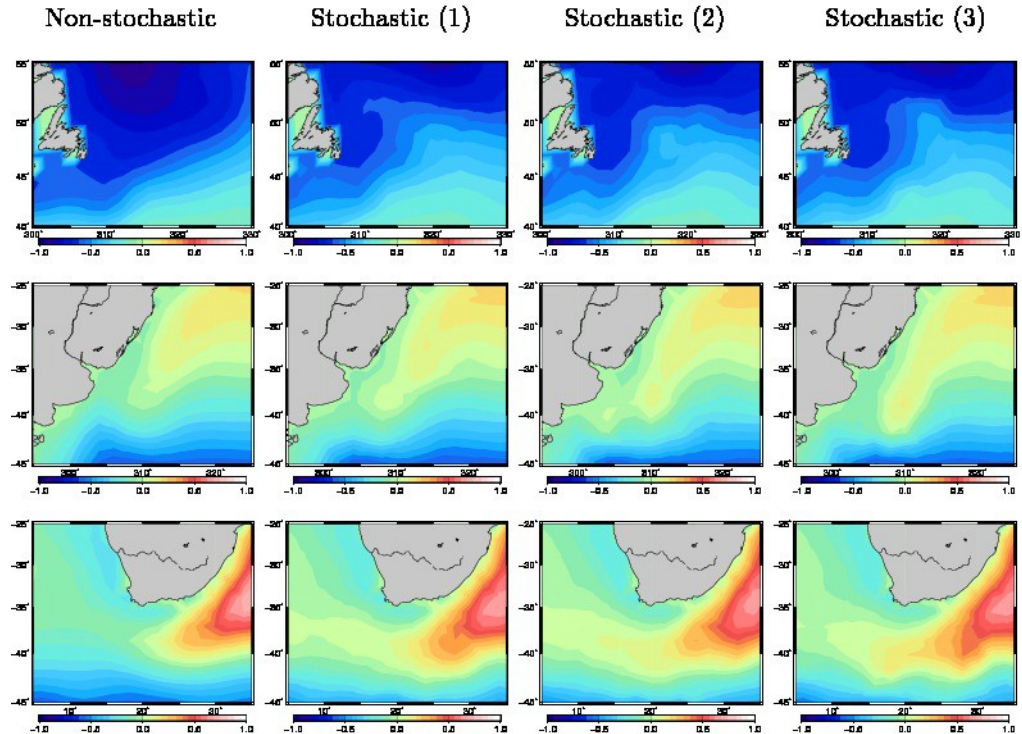
2.2 Modèle stochastique de circulation

En raison de la non-linéarité de l'équation d'état, la densité à grande échelle dépend aussi des échelles non-résolues

Equation d'état stochastique

$$\rho = \frac{1}{p} \sum_{i=1}^p \rho [T + \Delta T_i, S + \Delta S_i, p_0(z)]$$

processus stochastiques
simulant explicitement
l'effet des échelles
non-résolues de T et S
(Brankart, 2013)

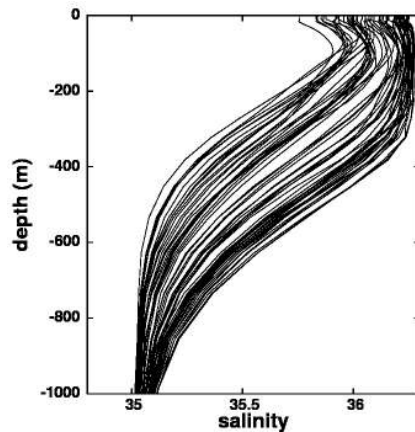
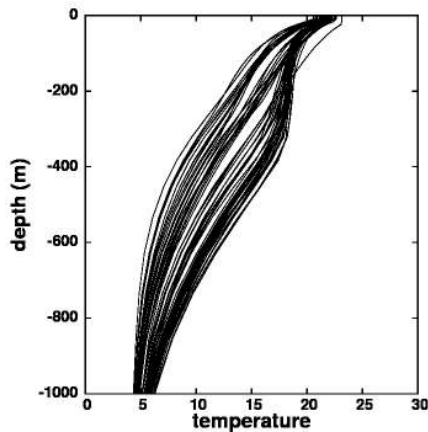
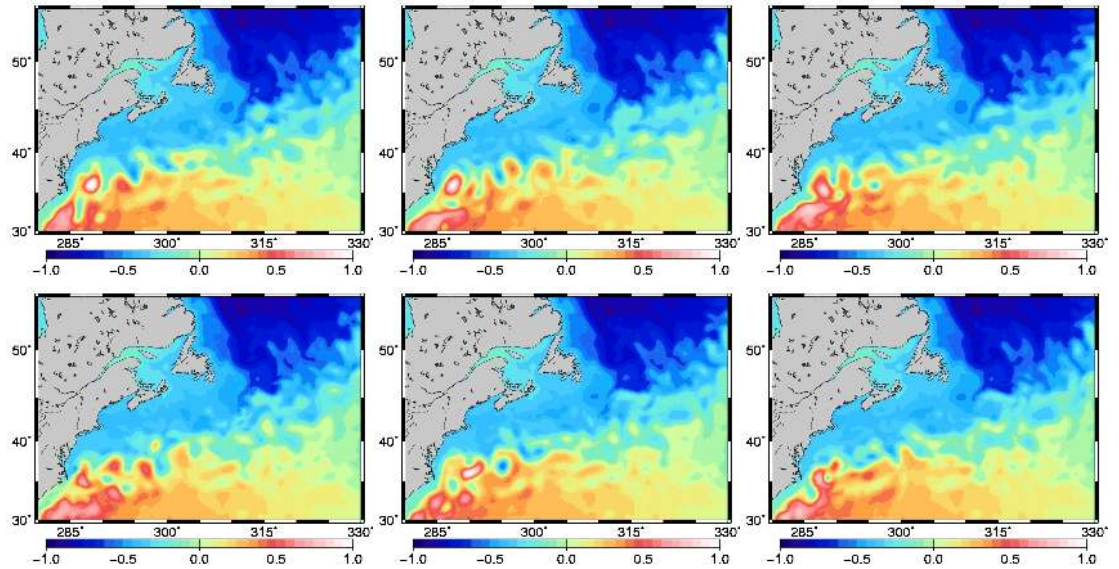


- Effet majeur sur la circulation moyenne (dans ORCA2)
- Stimulation d'une variabilité intrinsèque à grande échelle

Application à la configuration NATL025

Dispersion de
l'écoulement à
mésoséchelle,

illustré ici par la SSH
(6 membres parmi 96)



Distribution de probabilité
correspondante
pour les profils de
température et salinité

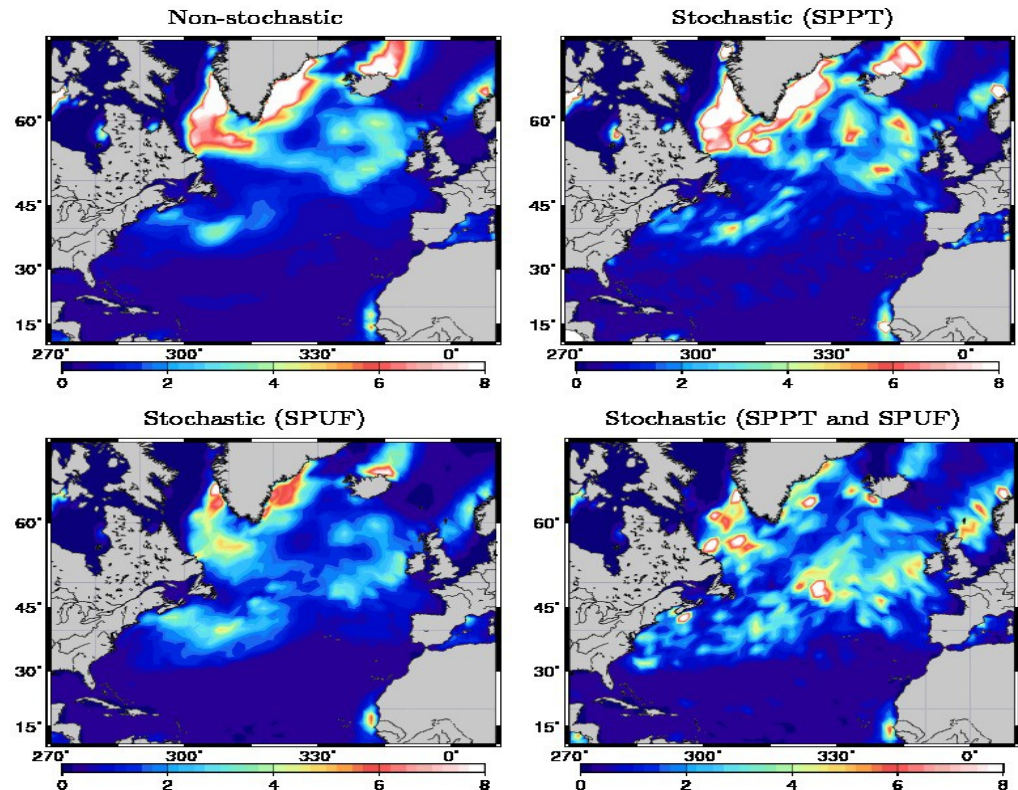
→ Prêt à être confronté aux observations, et pour assimiler des données altimétriques (Candille et al., 2014)

2.3 Modèle stochastique d'écosystème

Incertitudes multiples dans les modèles d'écosystème:
diversité biologique non-résolue, échelles non-résolues, etc.
Simulées par exemple dans Doron et al. (2011, 2013),
par des perturbations des paramètres du modèle.

Diversité non-résolue
bruit multiplicatif
dans les termes SMS
du modèle

Echelles non-résolues
processus stochastiques
simulant explicitement
les fluctuations
non-résolues des C_i



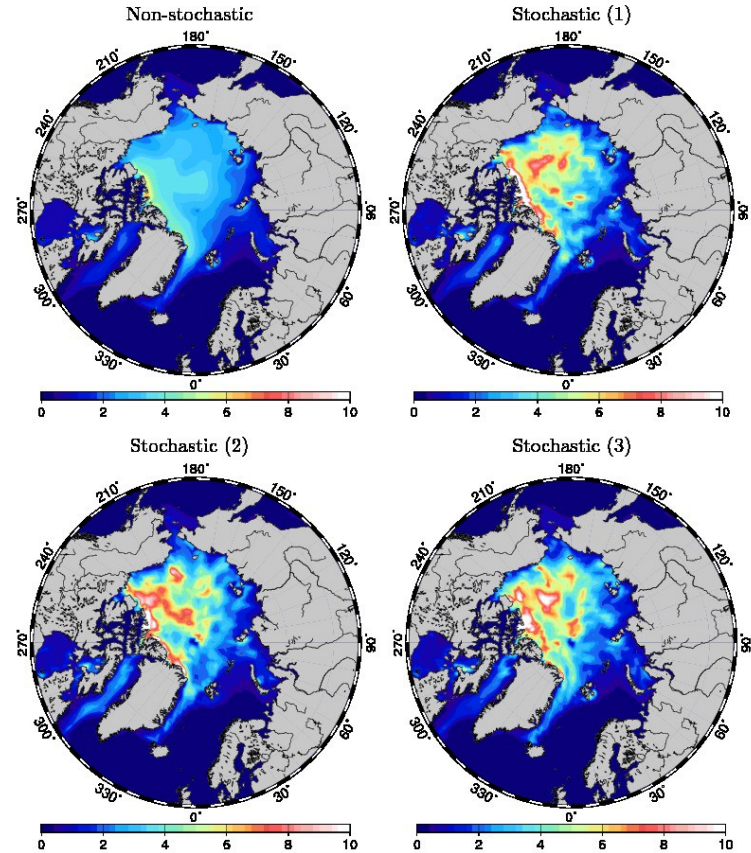
- Effet majeur sur le comportement moyen
- Augmentation de la microstructuration (↔ couleur de l'eau)

2.4 Modèle stochastique de glace de mer

Une difficulté essentielle des modèles de glace de mer est la diversité non-résolue des comportements dynamiques de la glace

Un des paramètres incertains les plus sensible est la résistance de la glace (P^*):

bruit multiplicatif
appliqué à P^*
(paramétrisation de
Juricke et al., 2013,
recodée dans NEMO)



- Effet majeur sur l'épaisseur moyenne de glace (ORCA2)
- Stimulation d'une variabilité interannuelle intrinsèque

2.5 Modèle probabiliste et simulations d'ensemble

**Le modèle NEMO devient probabiliste;
il est traité comme un système complexe,
fait de composants explicitement incertains**

→ L'objectif du modélisateur est alors de produire un modèle aussi informatif que possible et au moindre coût.

**Cette description probabiliste appelle la
réalisation de simulations d'ensemble**

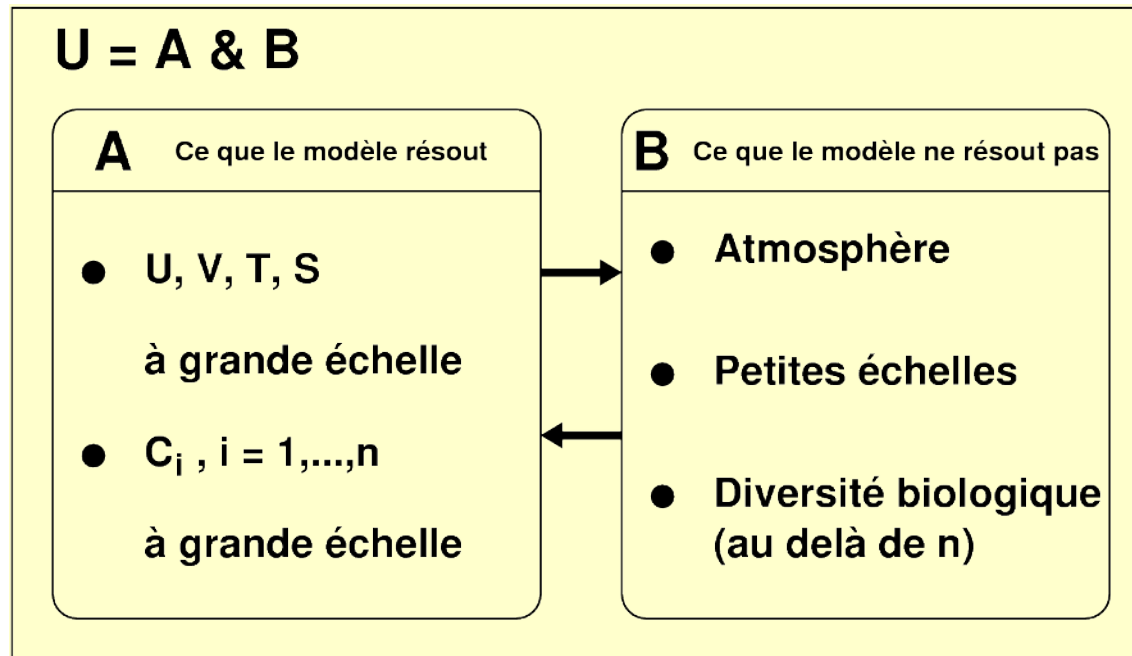
→ Comparaison objective du modèle aux observations

→ Prise en compte automatique des incertitudes sur le modèle dans les systèmes d'assimilation de données océaniques

3

Incertitudes sur les observations

3.1 Origines des incertitudes sur les observations



- Opérateur d'observation $\mathbf{y} = H(\mathbf{x}) \rightarrow$ équivalent observé pour le système **A**
 - L'incertitude sur les observations $p(\mathbf{y}^o | \mathbf{y})$ inclut
 - les erreurs de mesure (liées à l'instrument ou à des effets parasites),
 - l'incertitude de représentativité (cad tout ce qui se passe dans **B**)
- **L'approche probabiliste est ici toute naturelle**
- **Une approche stochastique serait aussi possible pour H, par ex:**
- pour simuler les échelles non-résolues \rightarrow échelles observées
 - pour simuler la diversité non-résolue \rightarrow obs. de couleur de l'eau

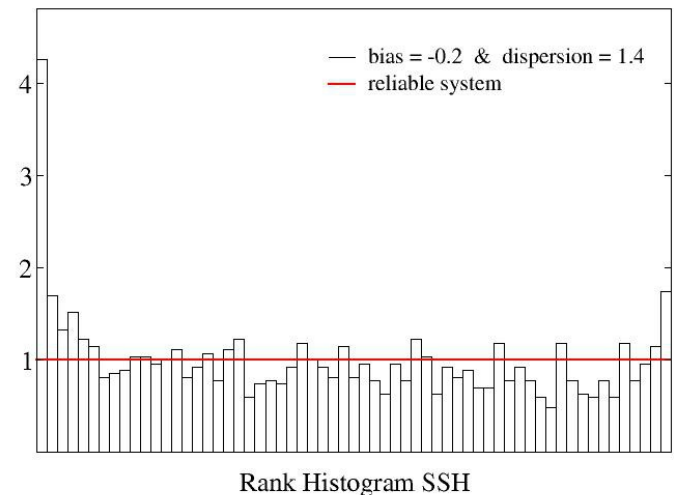
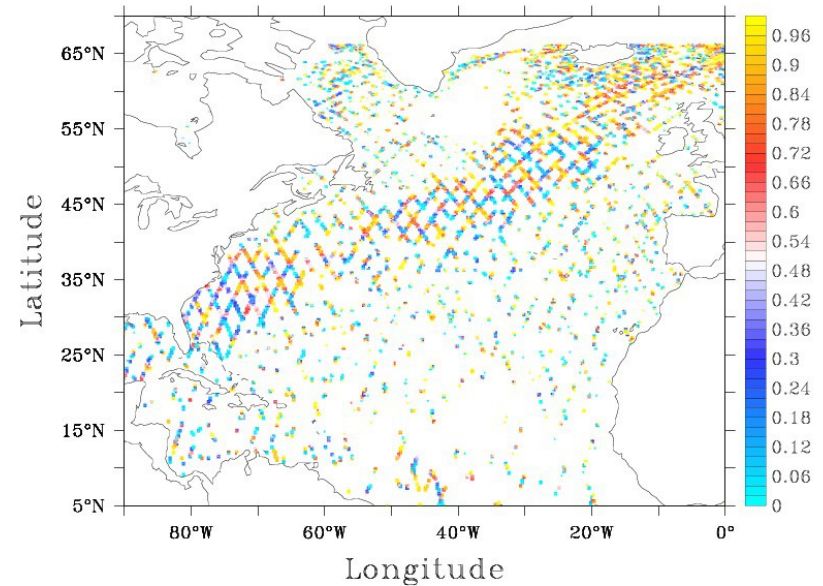
3.2 Test de cohérence entre modèle et observations

L'approche probabiliste permet un test de cohérence objectif entre modèle et observations

Tenant compte des incertitudes sur le modèle, $p(y)$ caractérise l'information que le modèle contient à propos de y (sa résolution)

Tenant compte en plus des incertitudes sur les observations, $p(y^o)$ caractérise la distribution de probabilité pour le résultat des mesures (\rightarrow test de fiabilité)

\rightarrow Condition suffisante d'invalidité du modèle (par ex. l'histogramme de rang)



Rang des observations JASON dans une simulation d'ensemble NATL025

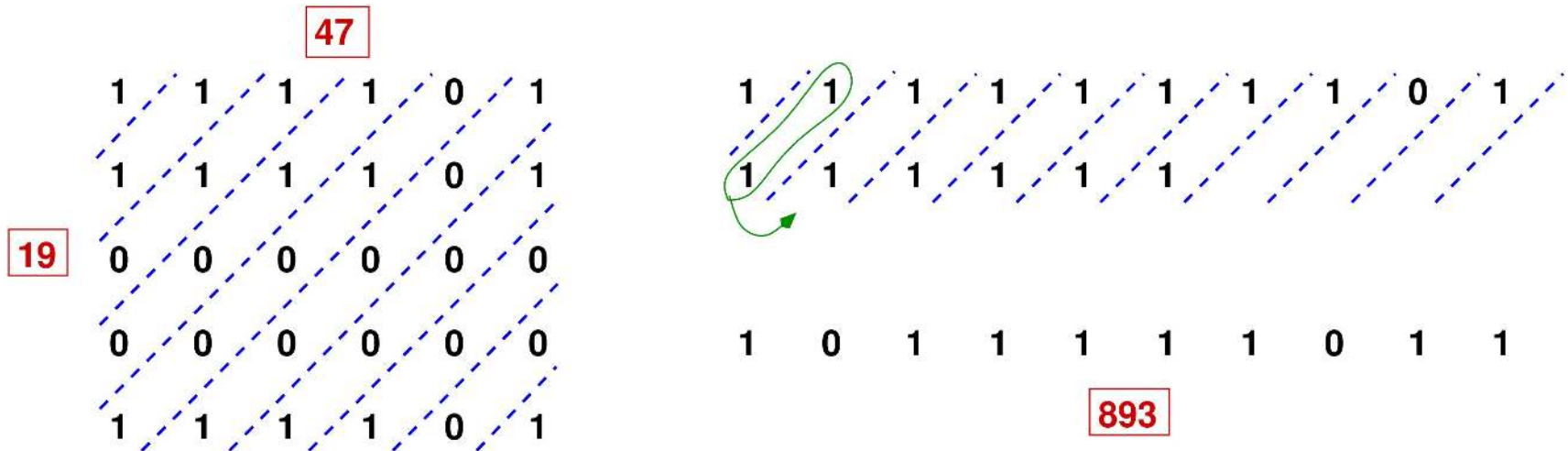
4

Réduction des incertitudes

Problème inverse et malédiction des dimensions

Problème inverse: remonter aux causes à partir des conséquences

Exemple: multiplication et factorisation



Multiplication: n^2 opérations binaires

Factorisation: pour l'instant insoluble en temps polynomial

→ Aucun nombre de plus de 768 bits n'a jamais été factorisé

Cette « malédiction des dimensions » est la difficulté essentielle:

→ pas de solution générale qui rende le problème traitable

→ besoin d'approximations spécifiques adaptées à nos problèmes

4.1 Hypothèses pour rendre le problème traitable

L'approche probabiliste conduit aux 2 hypothèses les plus couramment utilisées en assimilation de données

Théorème de Bayes: $p^a(\mathbf{x}) \sim p^b(\mathbf{x}) p(y^o|\mathbf{x})$

Hypothèse gaussienne
pour $p^b(\mathbf{x})$ et $p(y^o|\mathbf{x})$



calcul explicite de la moyenne
et de la covariance de $p^a(\mathbf{x})$



filtres de Kalman

Hypothèse d'unimodalité
pour $p^a(\mathbf{x})$



recherche itérative du mode
par méthode variationnelle



3DVAR, 4DVAR

Ces 2 hypothèses rendent le problème traitable, mais sont parfois insuffisantes en océanographie → besoin d'aller au-delà

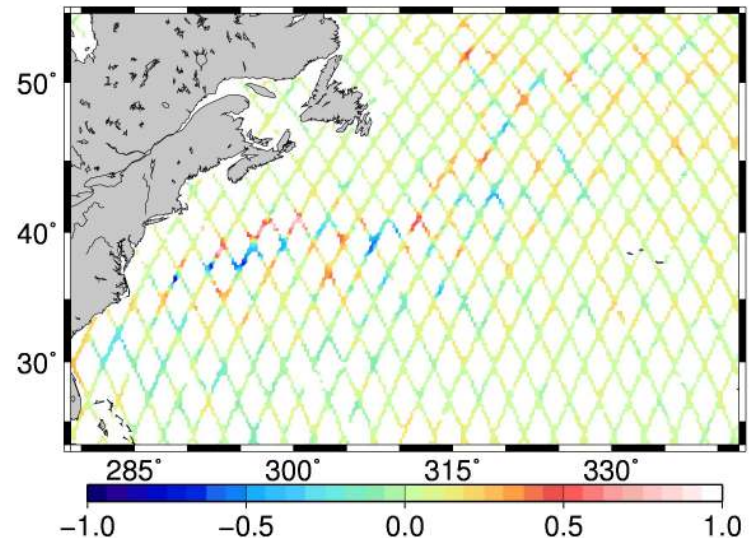
Exemple d'application de l'hypothèse gaussienne, pour l'assimilation des observations altimétriques de JASON1 et ENVISAT dans NATL025

Approche probabiliste :
ensemble de 96 membres

Mise à jour de l'ensemble par
l'algorithme SEEK, avec localisation

Expérience réalisée dans le cadre
du projet européen SANGOMA
(Candille et al., 2014)

Observations JASON le 17/9/2005

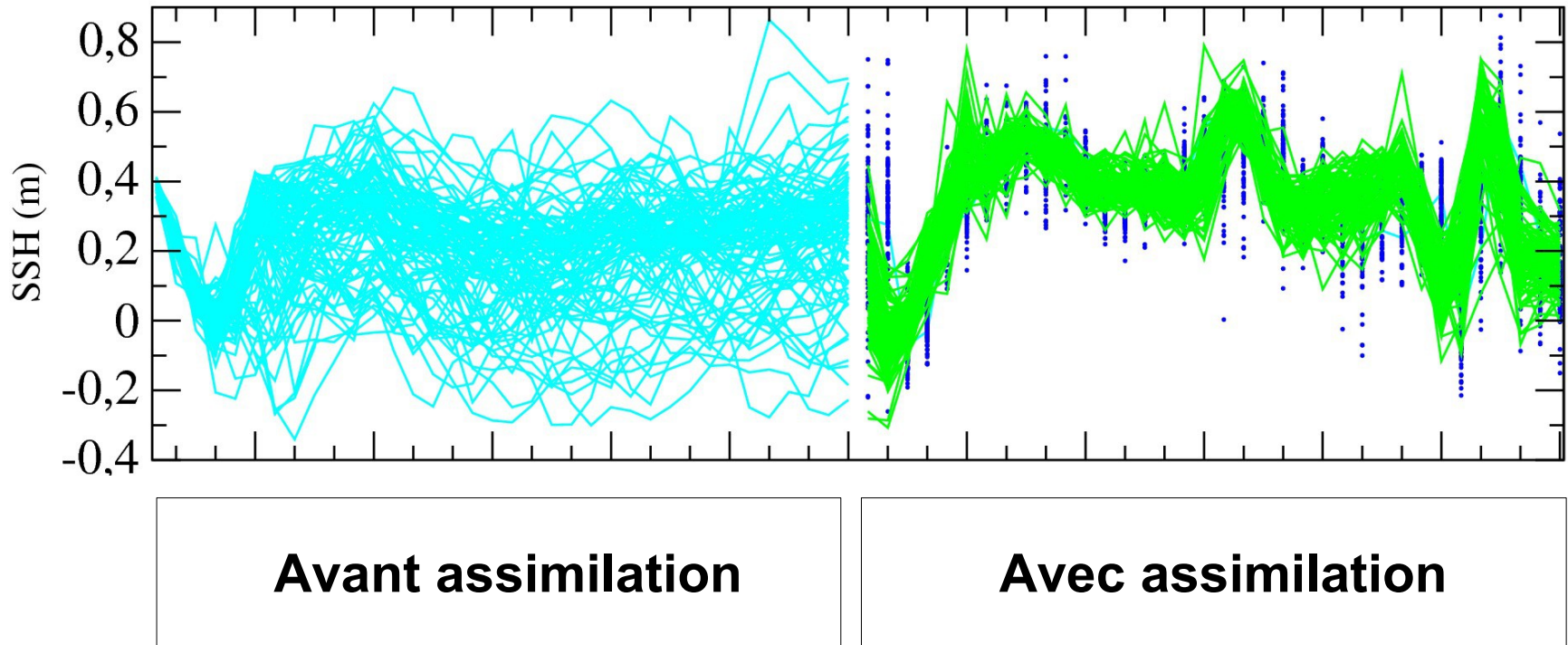


Caractéristiques importantes de la méthode:

- 1) Paramétrisation stochastique des incertitudes sur le modèle**
(→ pas d'inflation des prévisions d'ensemble)
- 2) Calcul de l'équivalent observé de chacun des membres de l'ensemble au bon temps** (→ on utilise la covariance 4D)
- 3) Schéma IAU (incremental analysis update) pour chaque membre**
(→ pas de discontinuité temporelle dans l'ensemble a posteriori)

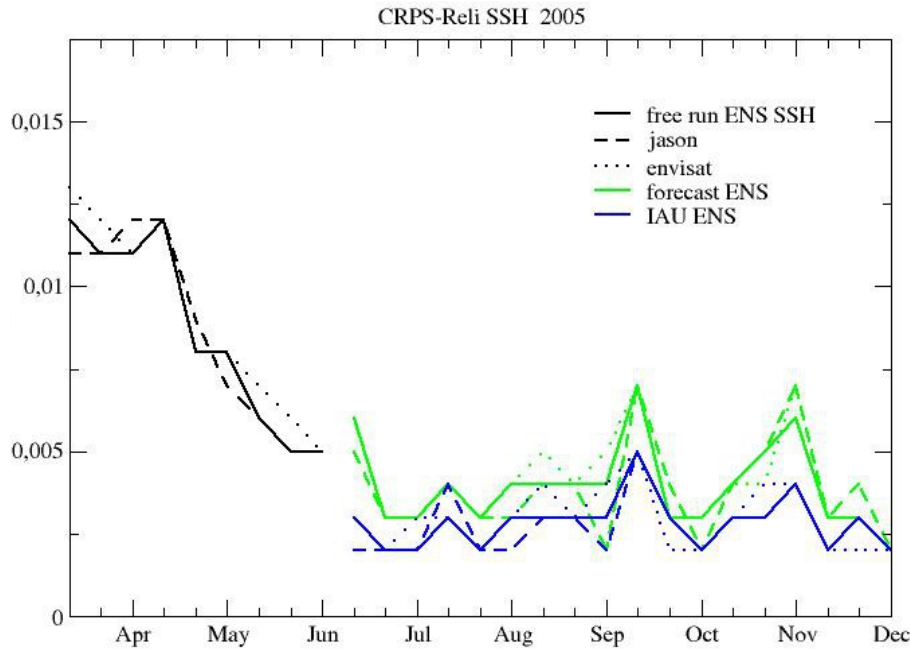
Evolution de la dispersion de l'ensemble (pour la SSH)

(A): 68.5E 35.5N

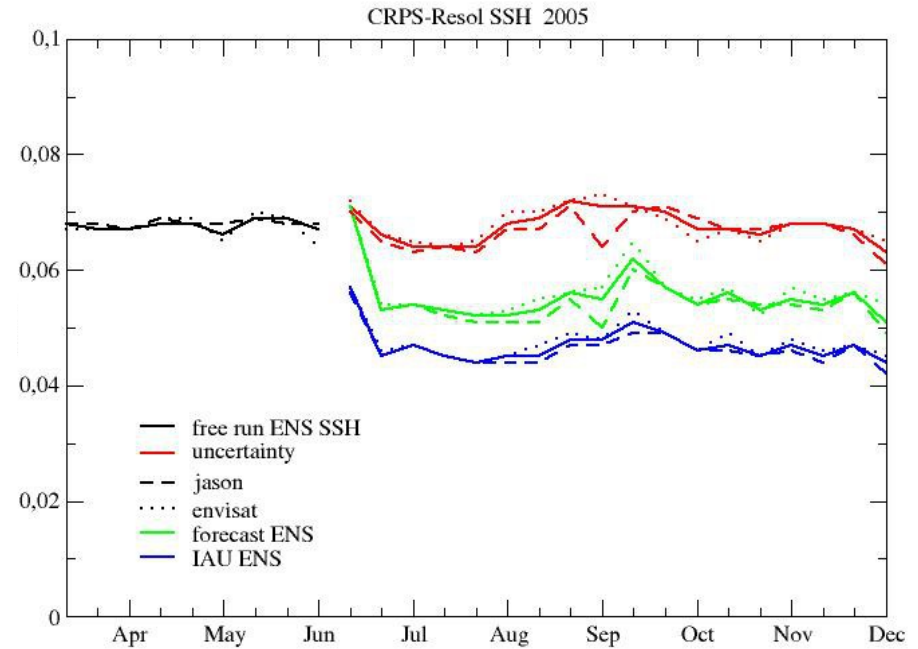


- La dispersion de l'ensemble est suffisante pour expliquer les observations altimétriques, du moins dans le Gulf Stream (histogramme de rang, vu précédemment).
- On passe d'un ensemble climatologique (sans info sur la mésoéchelle) à un ensemble qui informe sur l'écoulement à mésoéchelle réel (tel que vu par les observations).

Evaluation probabiliste de l'ensemble (métrique CRPS)



FIABILITE

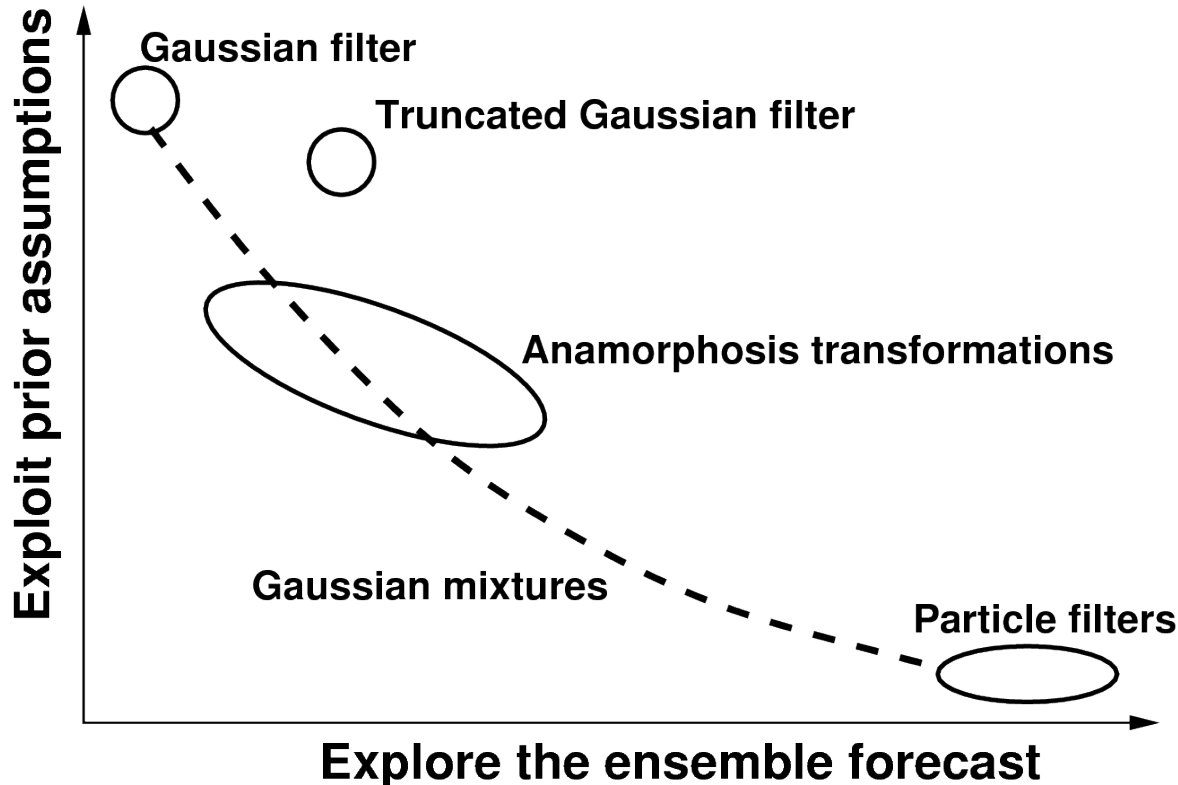


RESOLUTION

→ **Par rapport à l'ensemble libre, nous améliorons la résolution sans perdre de fiabilité**

4.2 Au delà de l'hypothèse gaussienne

Modélisation des incertitudes: exploitation d'hypothèses a priori vs exploration d'une prévision d'ensemble



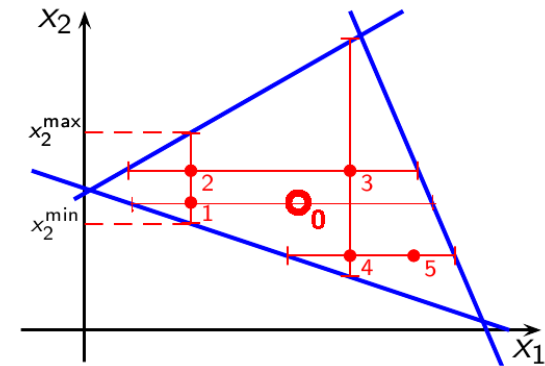
L'hypothèse doit être suffisamment générale pour être réaliste, et suffisamment simple pour rendre le problème traitable.

Au delà de l'hypothèse gaussienne: gaussiennes tronquées

Contraintes d'inégalité → hypothèse gaussienne tronquée (GT)

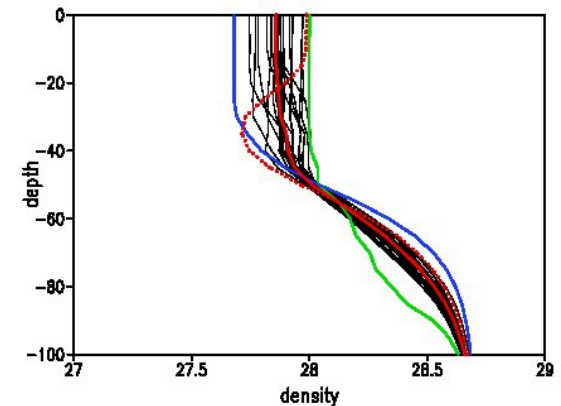
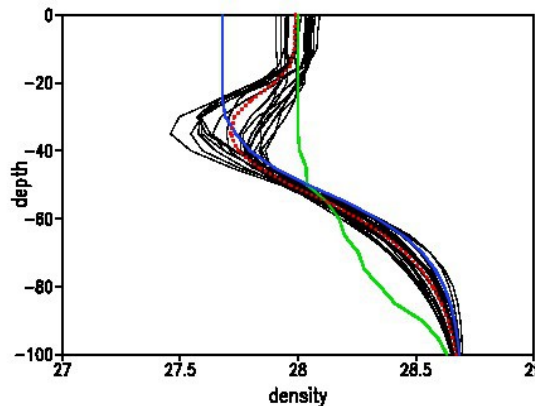
L'application de la formule de Bayes se fait avec le même algorithme et au même coût que pour une distribution gaussienne

La GT s'échantillonne facilement par l'échantillonneur de Gibbs (si pas trop de dimensions)



Echantillonnage de la GT par échantillonneur de Gibbs

Application à un modèle 1D de couche de mélange océanique (Lauvernet et al., 2009)

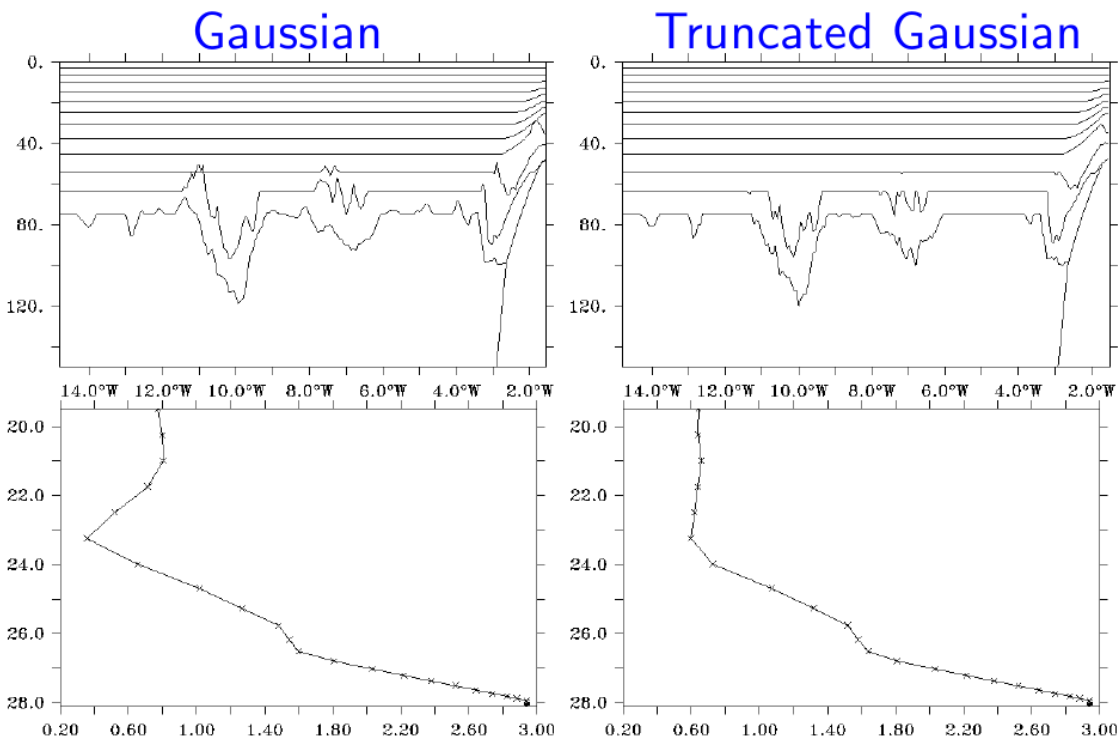
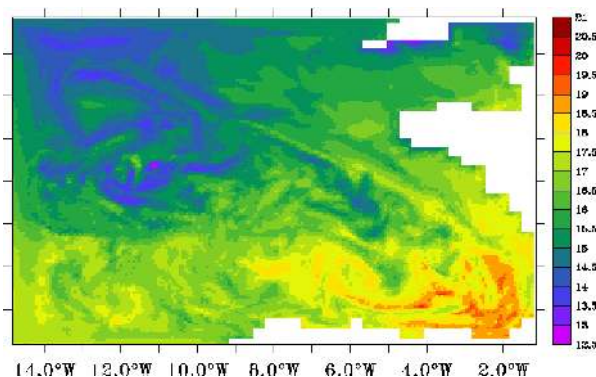


Gaussiennes tronquées: application à une configuration HYCOM du golfe de Gascogne

Modèle: configuration au $1/15^\circ$ du Golfe de Gascogne (Broquet et al., 2008)

Contraintes d'inégalité: $\rho_k - \rho_{k+1} < \varepsilon$, $\rho_k - \rho_{k+1} < 0$, $\rho_k > \rho_k^{\min}(H)$

Assimilation de cartes de SST du Golfe de Gascogne

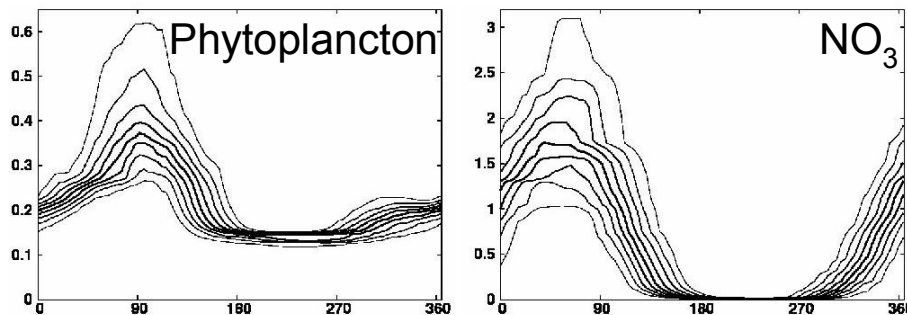


→ La méthode est assez efficace pour être appliquée à un système de taille réaliste

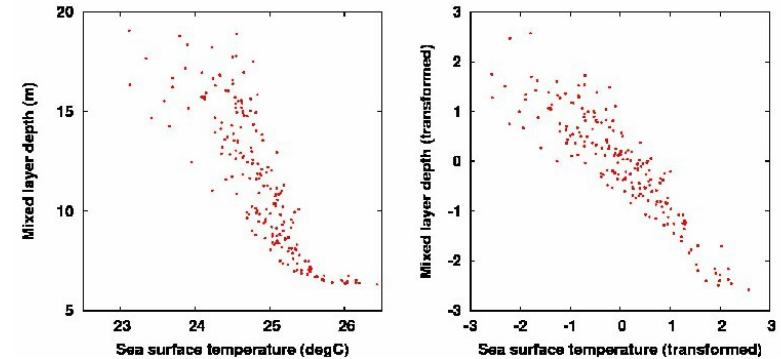
Au delà de l'hypothèse gaussienne: l'anamorphose

Transformation non-linéaire → distributions marginales gaussiennes

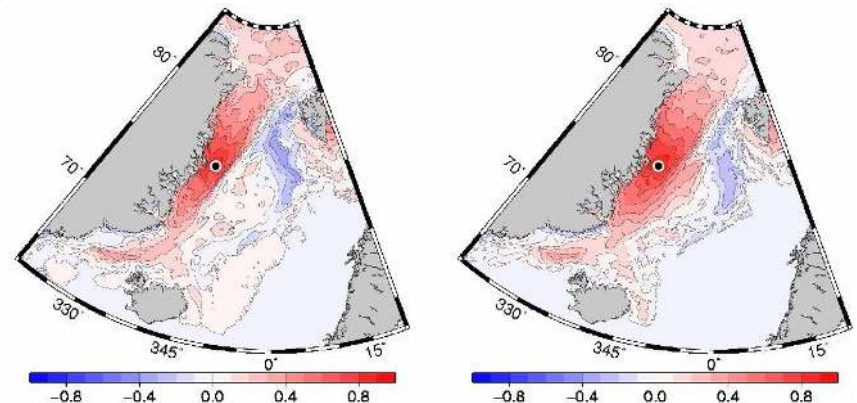
Modèle de circulation
ensemble simulant l'effet d'un forçage incertain dans ORCA2 (Meinvielle et al., 2013)



Modèle de glace de mer
effet de l'anamorphose sur la structure de corrélation (Brankart et al., 2012, à partir de données de Mercator-Océan)



Modèle d'écoystème
cycle annuel des déciles, définissant l'anamorphose (Fontana et al., 2013)

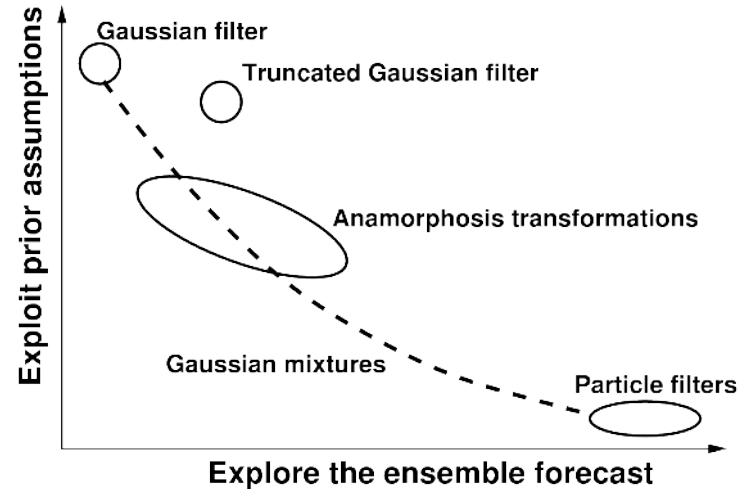


Au delà de l'hypothèse gaussienne: mélange de gaussiennes

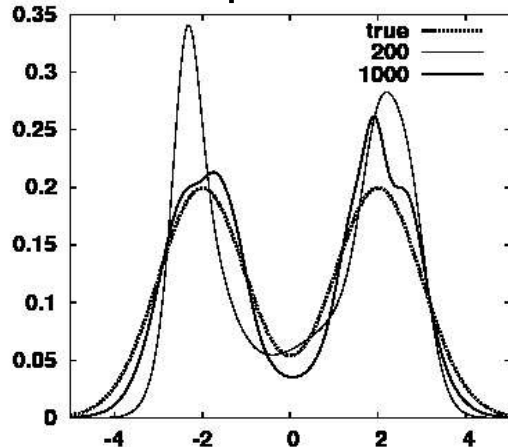
Intérêts de cette méthode

Efficace si appliquée dans un espace de dimension réduite (par ex. avec l'algorithme du filtre SEEK).

Permet de régler le degré d'exploration de l'ensemble, entre le modèle gaussien et le filtre particulaire

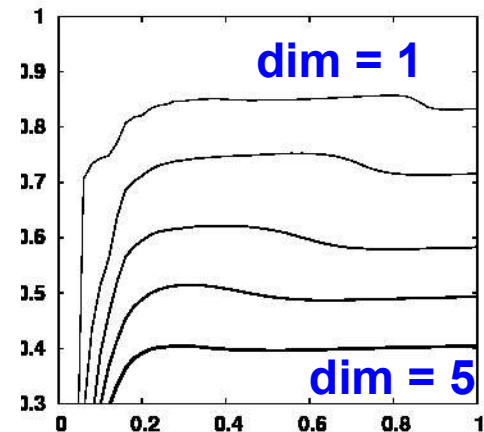


Identification à partir d'un échantillon



Apparition de détails non-significatifs si on explore trop finement l'ensemble

Optimisation du degré d'exploration



En maximisant la vraisemblance croisée des particules

4.3 Au delà de l'hypothèse d'unimodalité

En explorant directement la distribution a posteriori $p^a(\mathbf{x})$, sans décrire explicitement la distribution a priori

- Construire un chaîne de simulations qui ne sont plus indépendantes les unes des autres, mais conditionnées séquentiellement les unes aux autres
 - Moins sensible à la malédiction des dimensions

Les moments clés du processus sont les passages de col, qui permettent de poursuivre l'exploration vers de nouvelles vallées

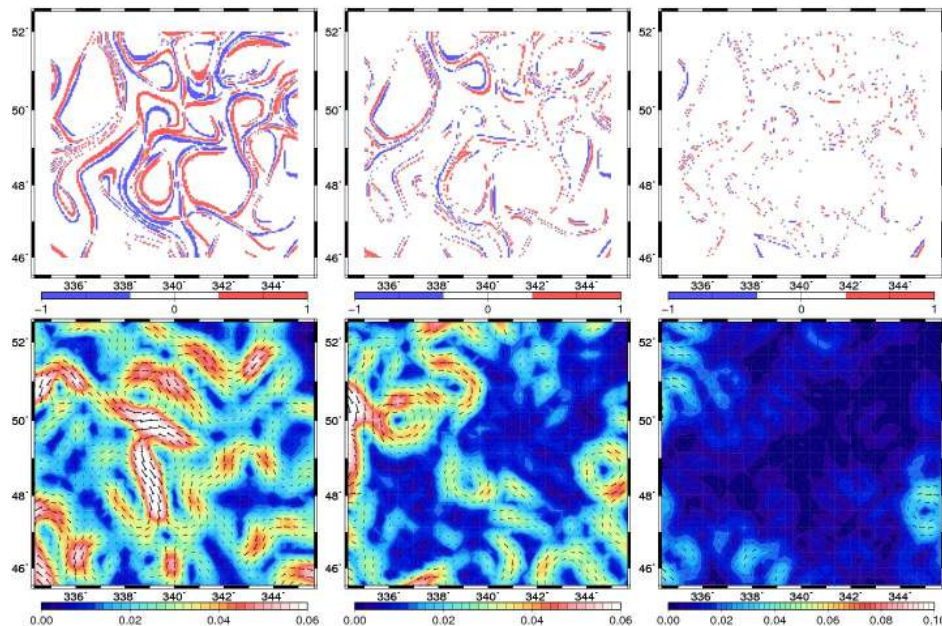
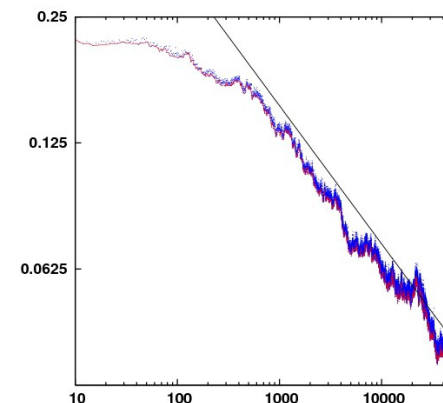
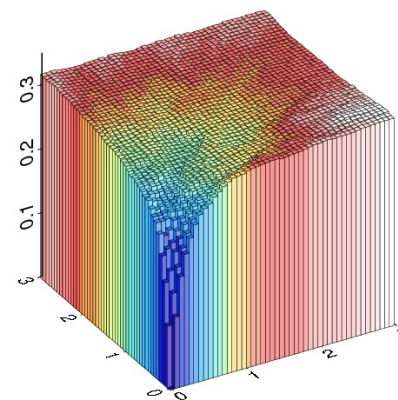


D'un point de vue technique, cela ne pourrait-il pas se construire comme un prolongement des outils variationnels existants, en y introduisant progressivement un peu de stochasticité ?

Au delà de l'hypothèse d'unimodalité: chaînes de Markov

Algorithmes MCMC: refroidi simulé, échantillonneur de Gibbs, algorithme de Métropolis/Hastings,...

Rechercher le champ de vitesse que les structures frontales observées (images de SST ou de couleur de l'eau) rendent le plus vraisemblable (Gaultier et al., 2013)

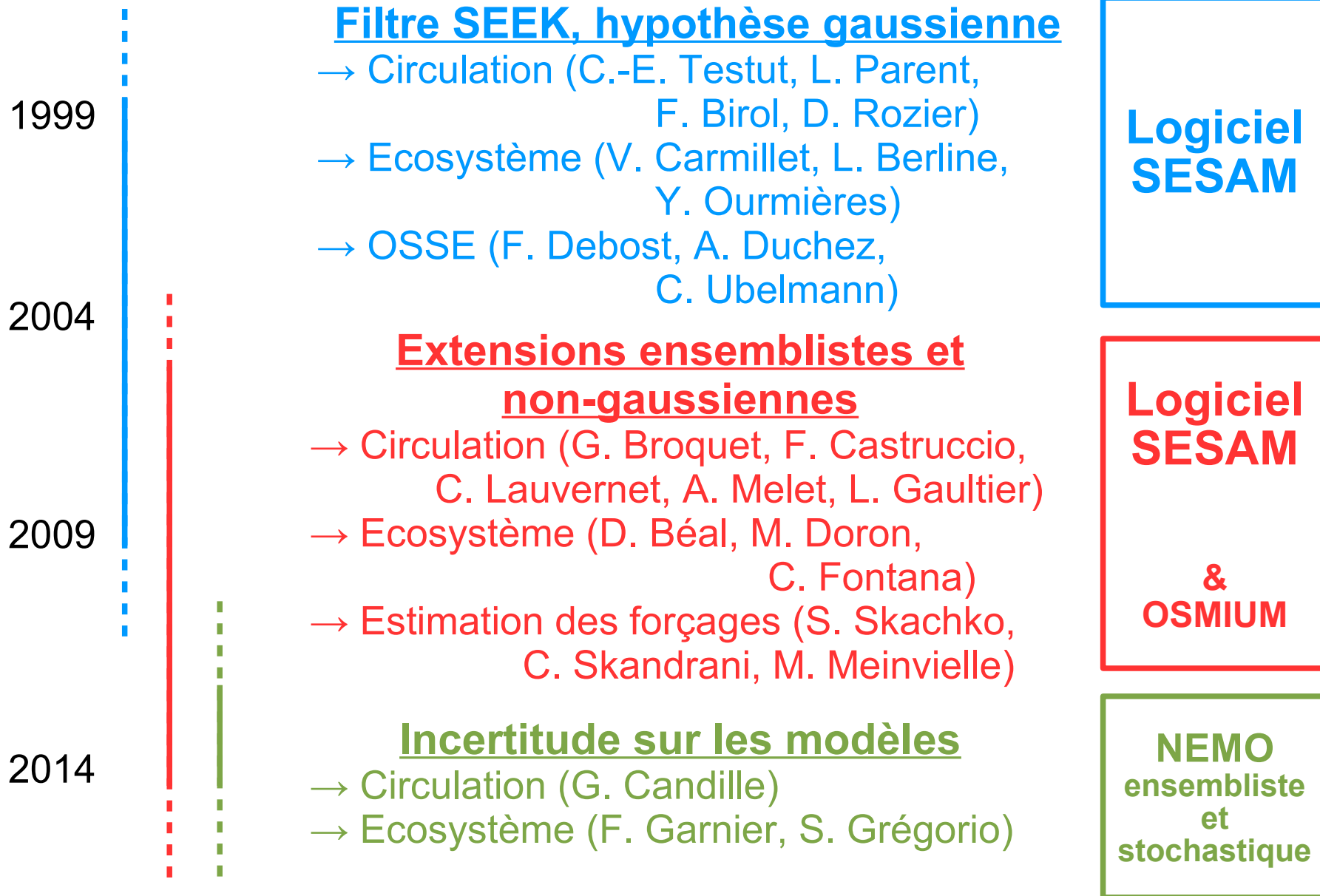


De col en col, au fur et à mesure de la réduction de l'écart aux structures observées, l'erreur sur la vitesse se réduit.

5

Conclusions

Synthèse des réalisations



Les outils que nous avons développés

Le logiciel SESAM

boîte à outils de composants
pour le filtre SEEK
et ses extensions :

analyse SEEK
en base propre

calcul d'EOFs

localisation

schéma adaptatif

échantillonnage de
gaussiennes tronquées

anamorphose

Le logiciel OSMIUM

reconstitution d'un champ
de vitesse à partir
d'une image de traceur

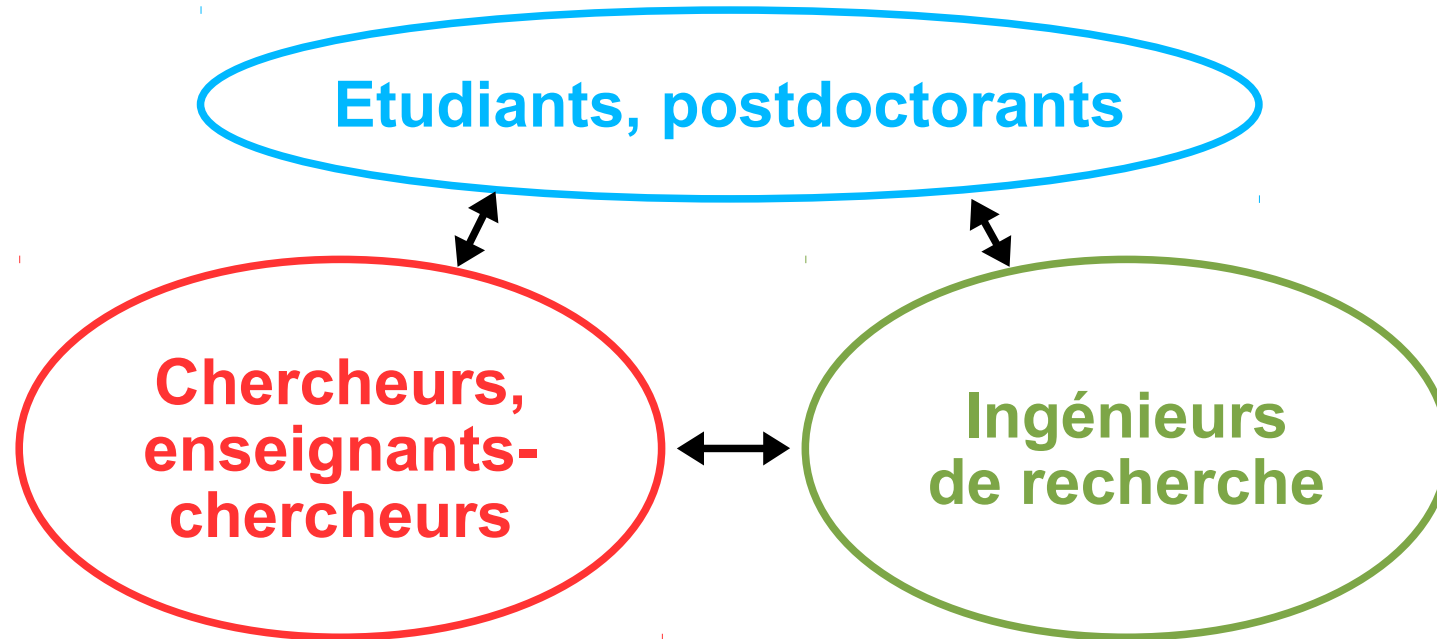
→ refroidi simulé

NEMO ensembliste et stochastique

module de génération
de champs de processus
autorégressifs

simulation en parallèle
de plusieurs membres
d'un ensemble

Mon travail d'ingénieur de recherche



- **Encadrement méthodologique et technique, mené selon une vision et une direction autonome**
- **Recherche méthodologique**
 - **algorithmes et logiciels génériques, flexibles, pouvant faire l'objet de publications**
 - **orientés vers le transfert aux applications opérationnelles (MERCATOR)**

L'incertitude: un élément clé de la prospective d'équipe 2016-2020

Axe scientifique de la prospective MEOM :

fonctionnement de l'océan des petites aux grandes échelles,
à travers les nouvelles données altimétriques
du satellite SWOT

Axe de développement méthodologique :

développement d'une nouvelle génération de modèles
et d'interfaces modèles-observations-assimilation
où les incertitudes seront explicitement intégrées
(océanographie probabiliste)

Ce mouvement vers l'océanographie probabiliste est déjà en cours dans l'équipe

En assimilation de données :

assimilation ensembliste (circulation, écosystème)
méthodes de comparaison objective aux observations

En modélisation :

simulation ensembliste de la variabilité océanique
(projet OCCIPUT, T. Penduff)

Thèse débutant en 2014

assimilation de données dans un modèle de glace de mer

Un traitement approprié des incertitudes est nécessaire pour faire le lien entre modèle, observations, et systèmes d'assimilation de données

L'incertitude est appelée à devenir un élément constitutif des systèmes que nous utilisons en océanographie, et non plus une superstructure distincte des résultats proprement dits.

Son traitement nécessitera de plus en plus une approche intégrée d'ingénierie à l'interface entre océanographie et mathématiques appliquées.