

Simulation explicite de l'incertitude dans un modèle d'océan et implications pour l'assimilation de données

Jean-Michel Brankart,
Florent Garnier, Christophe Calone,
Emmanuel Cosme et Pierre Brasseur

Laboratoire de Glaciologie et de Géophysique de l'Environnement

Equipe de Modélisation des Ecoulements Océaniques Multi-échelles



**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),
cité par Richard Hamming (1997)

Plan de l'exposé

1

Introduction

2

Simulation explicite des incertitudes

3

Quelques exemples océanographiques

4

Comparaison aux observations et assimilation de données

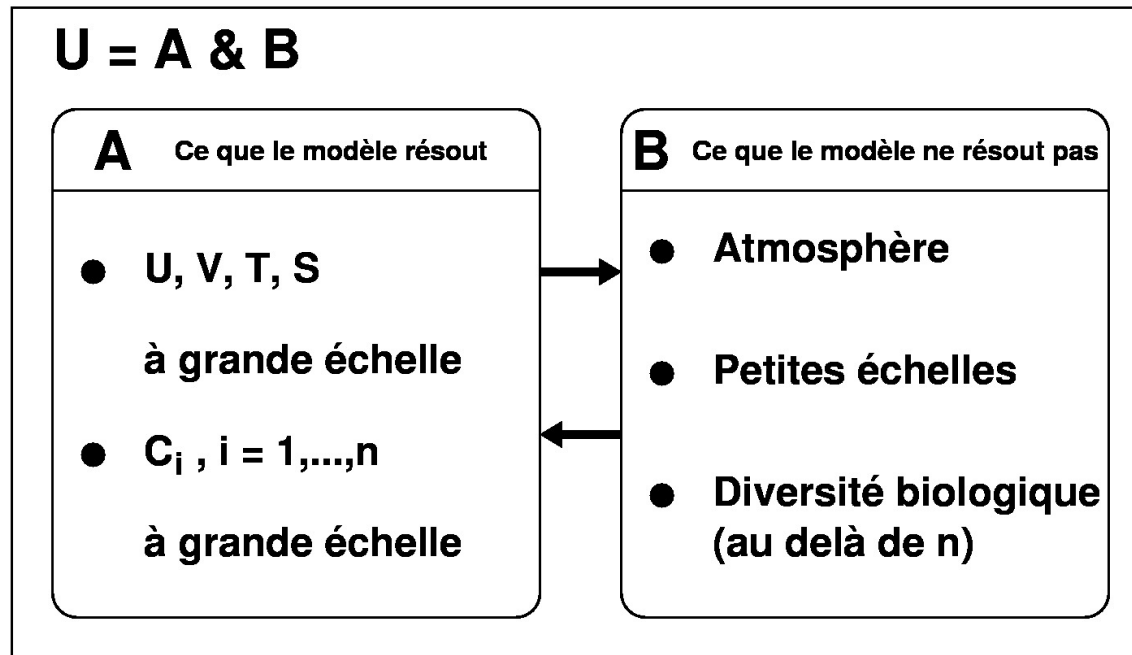
5

Conclusions

1

Introduction

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 (\rightarrow forçage atmosphérique), soit, que l'effet de **B** peut être paramétré (\rightarrow paramétrisation des échelles ou de la diversité biologique non-résolues).
- \rightarrow **B est la principale source d'incertitude dans le modèle.**

Motivations pour une approche probabiliste

L'approche déterministe n'est plus toujours suffisante pour décrire le comportement du système: circulation aux fines échelles, biogéochimie marine

L'approche probabiliste permet de mieux comparer simulations et observations

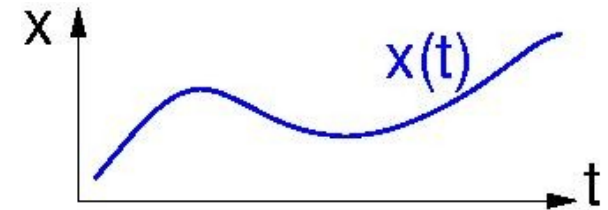
Bien connaître la confiance que l'on peut donner au modèle est nécessaire pour résoudre les problèmes d'assimilation de données

Une approche probabiliste de modélisation océanique

Dynamique océanique stochastique, explicitant les incertitudes

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

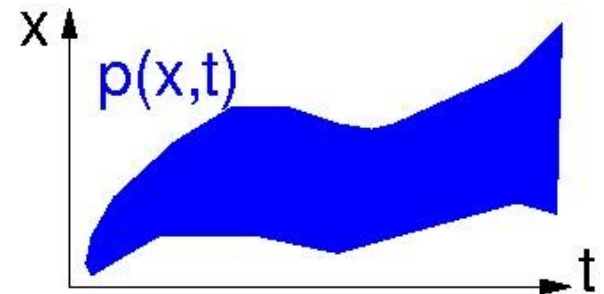
$$\text{où } \mathbf{x} = [x_1, \dots, x_N]$$



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)

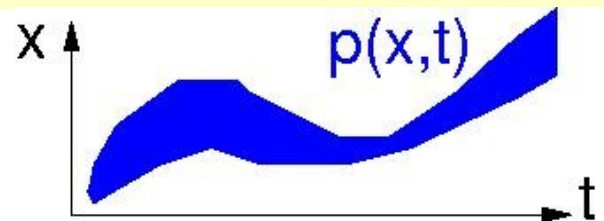
$$\frac{\partial p(\mathbf{x}, t)}{\partial t} = - \sum_{i=1}^N \frac{\partial}{\partial x_i} [\mathcal{M}_i(\mathbf{x}, t)p(\mathbf{x}, t)] + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \frac{\partial^2}{\partial x_i \partial x_j} [D_{ij}(\mathbf{x}, t)p(\mathbf{x}, t)]$$

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



Conditionnement aux observations, pour réduire les incertitudes

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



L'incertitude, en tant qu'élément constitutif de nos systèmes

**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 ?**

2

Simulation explicite des incertitudes

Formulation stochastique du modèle

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



Descrivre 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)

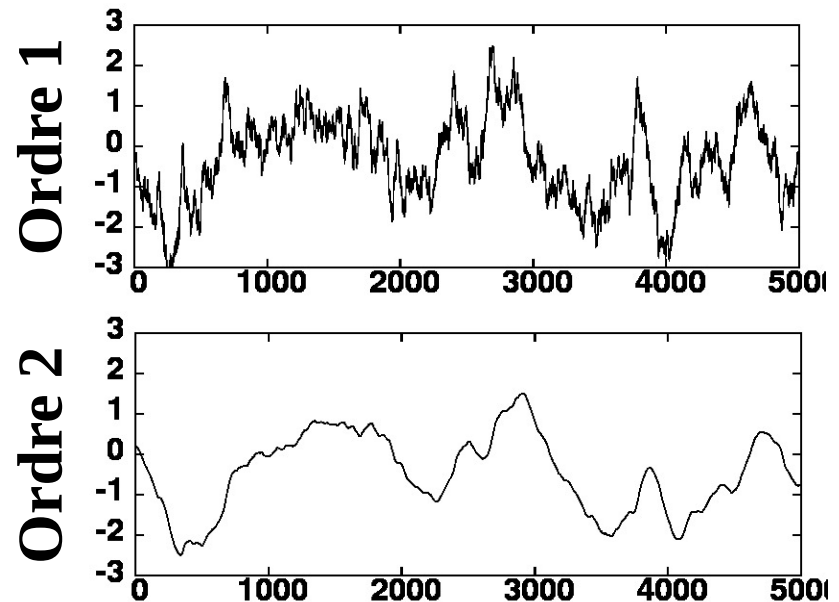
En chaque point de grille du modèle (2D ou 3D), on génère des **processus autorégressifs gaussiens et indépendants**:

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

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

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

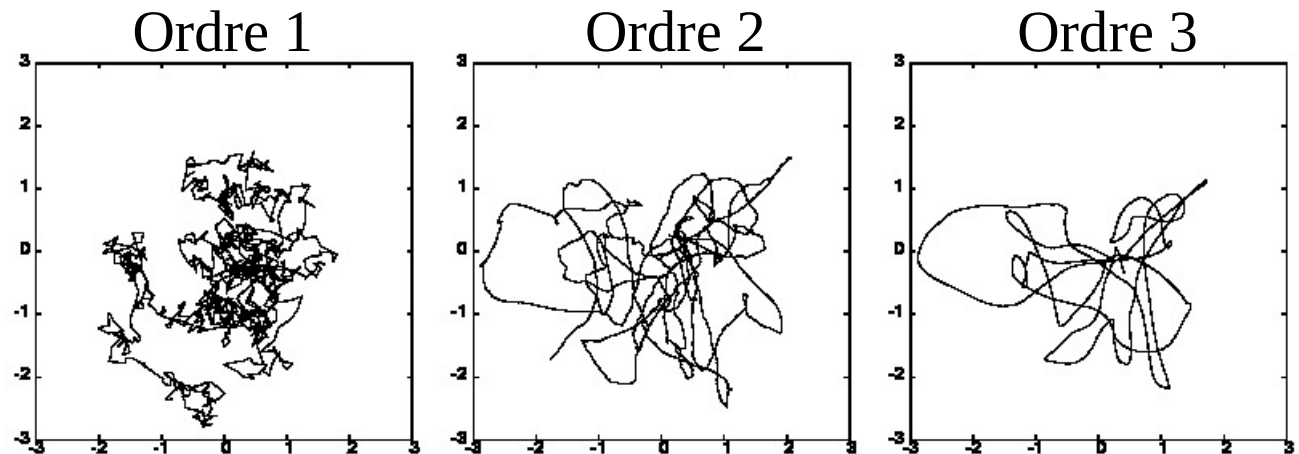
→ Utiliser des cartes de processus autorégressifs comme ξ (de moyenne 1), en spécifiant la structure de corrélation et la distribution marginale de ξ .

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

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

Genérer 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.

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

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.**

3

Quelques exemples océanographiques

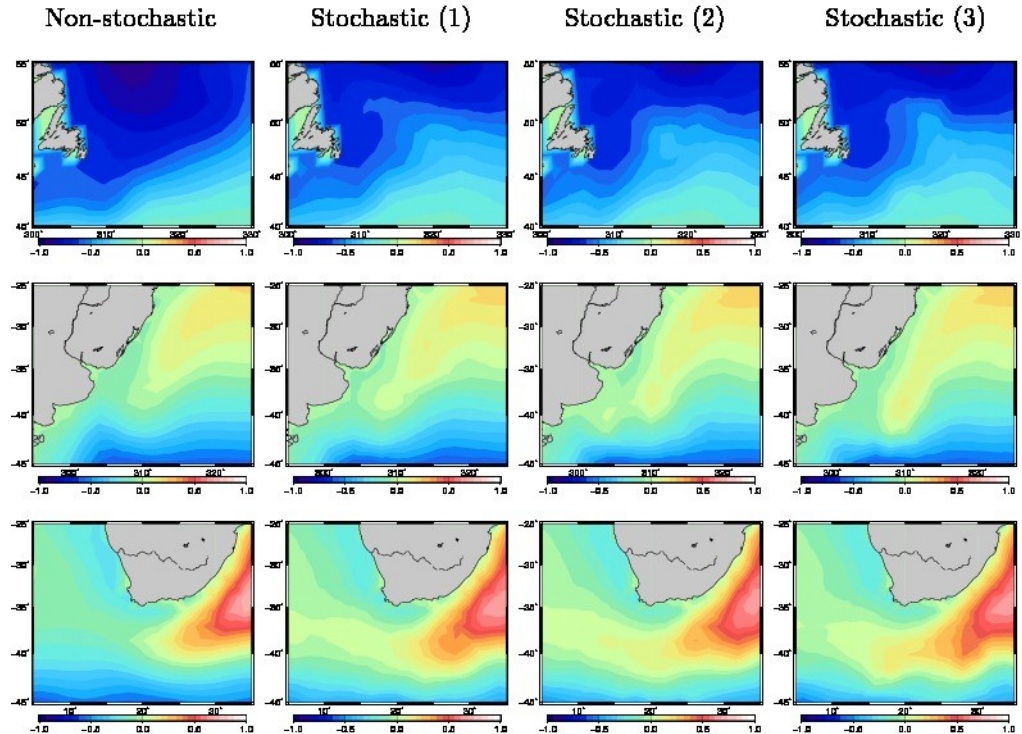
Exemple 1: effet des échelles non-résolues

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



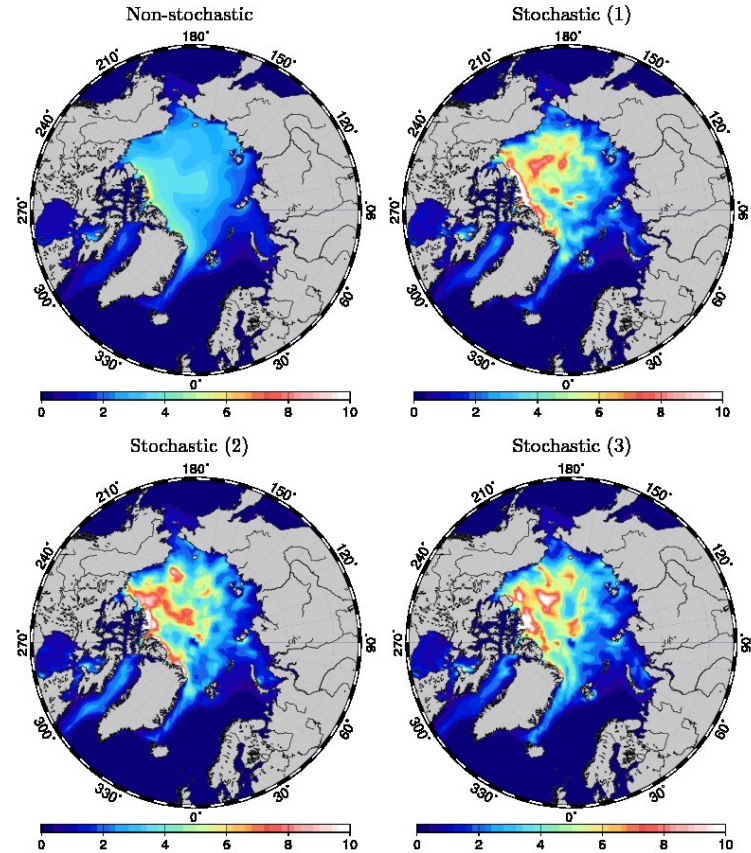
- Effet considérable sur la circulation moyenne (dans ORCA2)
- Stimulation d'une variabilité intrinsèque à grande échelle

Exemple 2: 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^*):

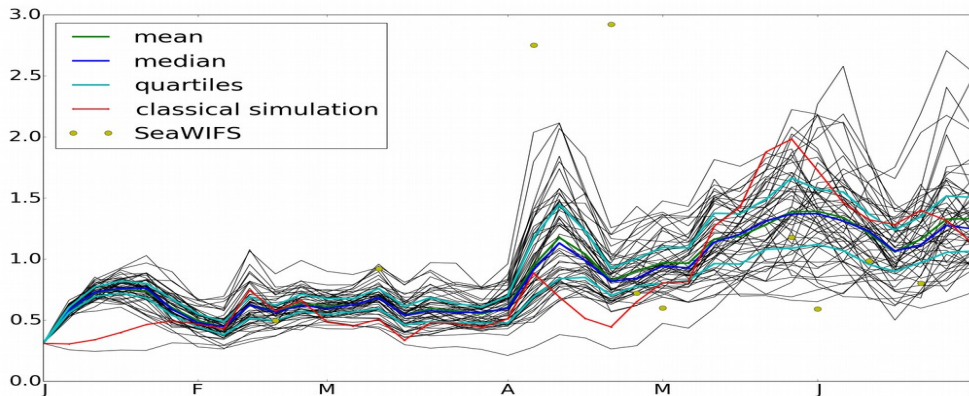
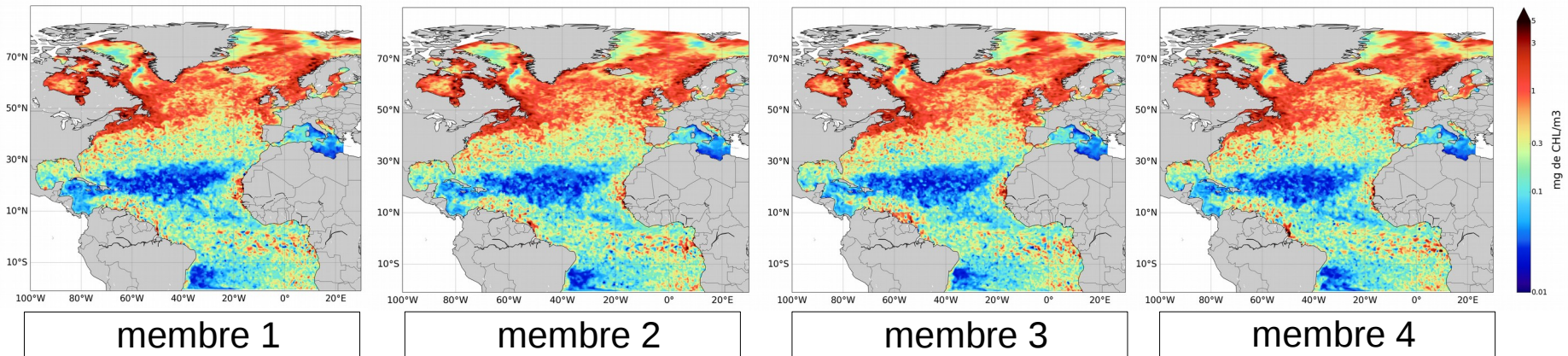
bruit multiplicatif appliqué à P^*
(paramétrisation de Juricke et al., 2013, recodée dans NEMO)



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

Exemple 3: ensemble de simulations de l'écosystème

Distribution de probabilité pour la concentration en chlorophylle
simulée ici par un ensemble NATL025/PISCES (4 membres parmi 50)
avec paramétrisation stochastique des incertitudes



Evolution temporelle
de la pdf
du phytoplancton
de janvier à juin 2005

→ assimilation de données de couleur de l'océan
(projets FP7-MyOcean2/SANGOMA, Garnier et al., 2016)

4

**Comparaison aux observations
et assimilation de données**

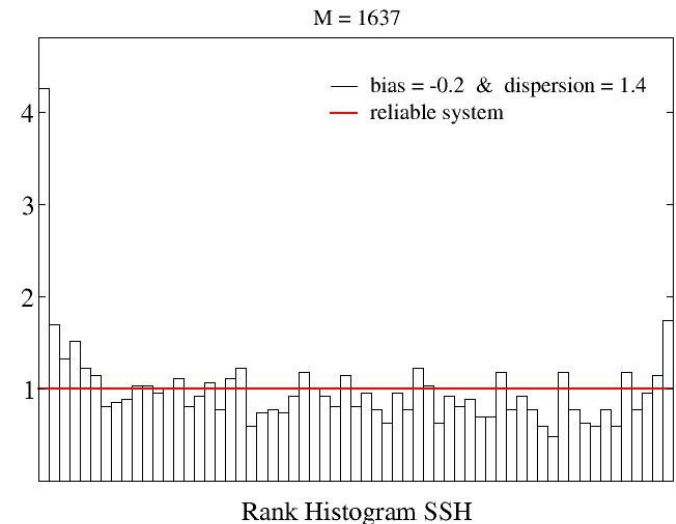
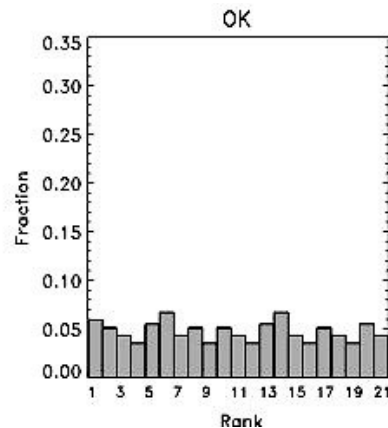
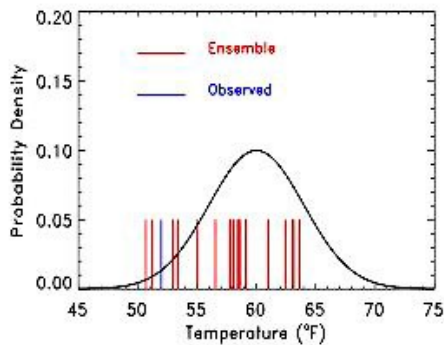
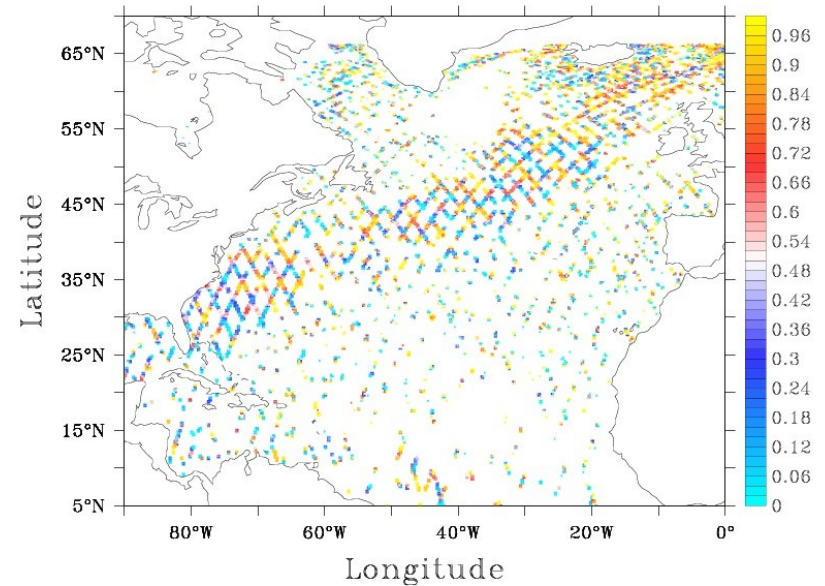
Une condition suffisante d'invalidité du modèle

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

Elle permet de caractériser:

l'information que le modèle contient à propos d'un observable (sa résolution)

la distribution de probabilité pour le résultat des mesures (→ test de fiabilité)

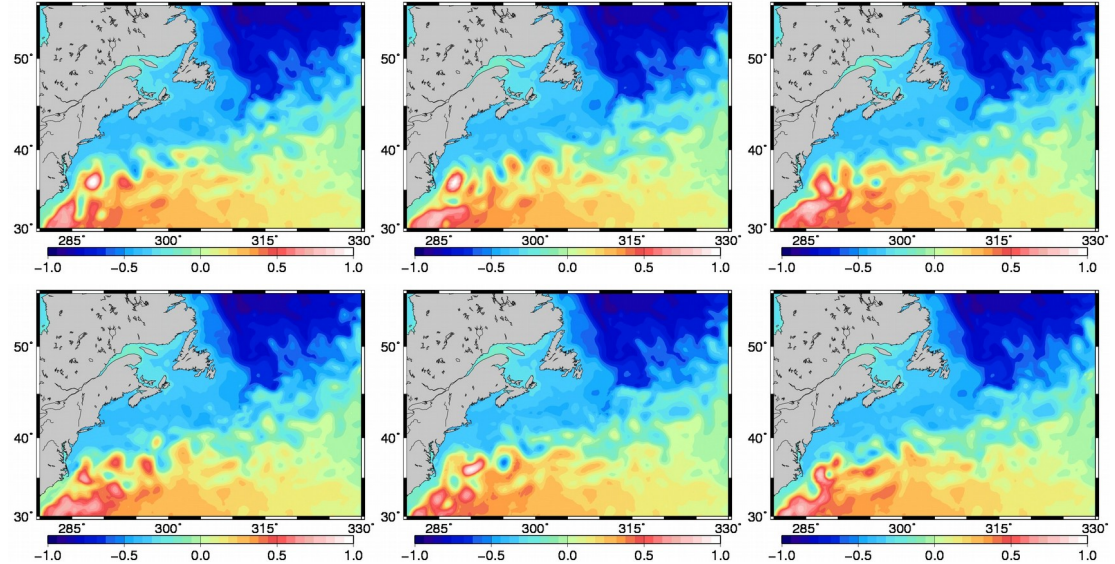


(Candille et al., 2015)

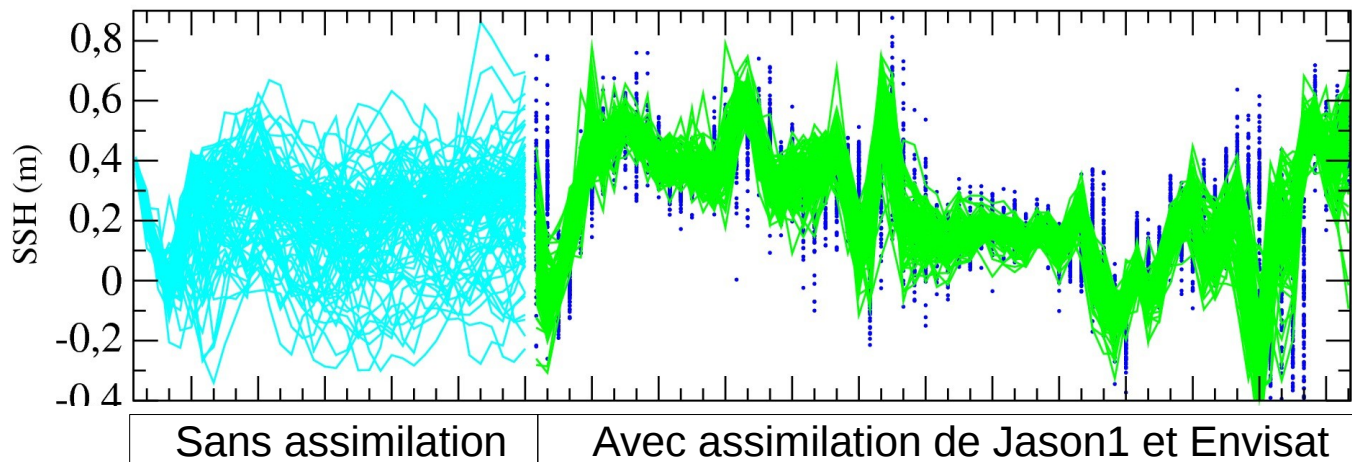
Gagner de l'information par assimilation de données

Distribution de probabilité pour la SSH

simulée ici par NATL025 d'ensemble (6 membres parmi 96)



→ **assimilation de l'altimétrie (projet FP7-SANGOMA, Candille et al., 2014)**

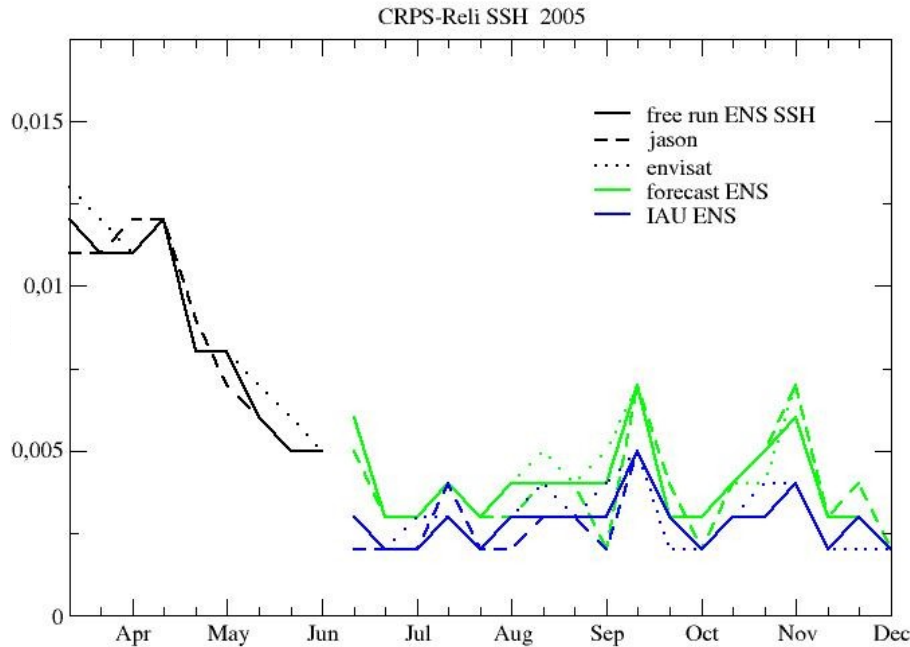


Evolution temporelle de la pdf

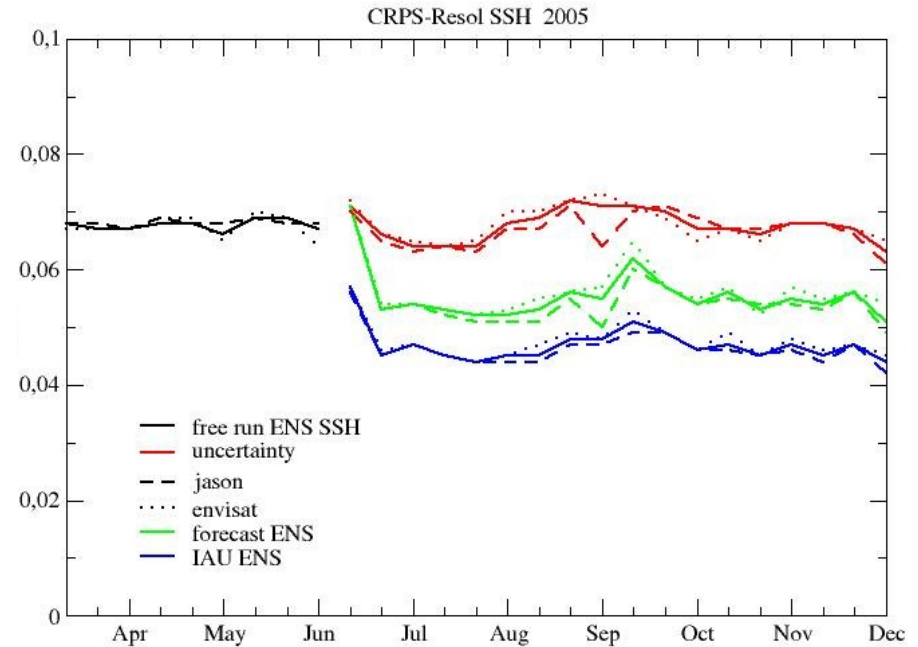
de juin 2005 à décembre 2006

Quels scores pour mesurer ce gain ?

Mesurer l'écart entre distributions observée et simulée



CRPS: fiabilité



CRPS: résolution

→ **Par rapport à l'ensemble libre (en noir), on améliore la résolution, sans perdre en fiabilité.**

5

Conclusions

**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

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.

**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 ».