HAL
open science

Assimilation variationnelle de données altimétriques dans le modèle océanique NEMO : Exploration de l'effet des non-linéarités dans une configuration simplifiée à haute résolution

Pierre-Antoine Bouttier

▶ **To cite this version:**

Pierre-Antoine Bouttier. Assimilation variationnelle de données altimétriques dans le modèle océanique NEMO : Exploration de l'effet des non-linéarités dans une configuration simplifiée à haute résolution. Optimisation et contrôle [math.OC]. Université de Grenoble, 2014. Français. NNT : . tel-00995305

HAL Id: tel-00995305

https://theses.hal.science/tel-00995305

Submitted on 23 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathématiques Appliquées**

Arrêté ministériel : 7 août 2006

Présentée par

Pierre-Antoine Bouttier

Thèse dirigée par **Éric Blayo**
et codirigée par **Jacques Verron**

préparée au sein **du Laboratoire de Glaciologie et de Géophysique de l'Environnement (LGGE, équipe MEOM, CNRS) et du Laboratoire Jean Kuntzmann (LJK, équipe MOISE, INRIA)**
et de **l'école doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique**

Assimilation variationnelle de données altimétriques dans le modèle océanique NEMO

Exploration de l'effet des non-linéarités dans une configuration simplifiée à haute résolution

Thèse soutenue publiquement le **4 février 2014**,
devant le jury composé de :

Pierre Brasseur
Directeur de Recherche CNRS (LGGE), Président
Fabienne Gaillard
Chercheuse senior IFREMER, Rapporteur
Mark Asch
Professeur Université de Picardie (LAMFA), Rapporteur
Claire Lévy
Ingénieure de Recherche CNRS (LOCEAN), Examinatrice
Arthur Vidard
Chargé de Recherche INRIA (LJK), Examinateur
Anthony Weaver
Chercheur senior CERFACS, Examinateur
Éric Blayo
Professeur UJF (LJK), Directeur de thèse
Jacques Verron
Directeur de Recherche CNRS (LGGE), Co-Directeur de thèse

Remerciements

Sans les personnes qui l'entourent, un travail de thèse n'aurait aucun sens. Le mien n'y fait pas exception.

Tout d'abord, je voudrais profondément remercier mes deux directeurs de thèse, Éric Blayo et Jacques Verron, pour la confiance qu'ils m'ont accordée, leur sympathie, leur écoute, leurs aiguillages judicieux. C'est auprès d'eux que j'ai commencé à apprendre à *faire* de la recherche scientifique, à tous les niveaux. Ils m'ont appris également qu'on ne finit jamais tout à fait d'apprendre.

La recherche scientifique, dont les thèses en sont des éléments, n'existe pas sans une communauté et le jugement par ses pairs. Ainsi, je remercie les membres de mon jury de thèse, que ce soit les rapporteurs, Fabienne Gaillard et Mark Asch, les examinateurs, Claire Lévy, Anthony Weaver et Arthur Vidard, ou le président du jury, Pierre Brasseur. Je les remercie de leur regard critique et bienveillant.

Ce travail n'a jamais été isolé. Il s'est construit entouré de nombreuses personnes, au sein des équipes MOISE du LJK, MEOM du LGGE et NEMO System Team du LOCEAN. Je remercie ainsi, dans un désordre assez confus, Pierre, Julien, Thierry, Jean-Michel, Jean-Marc, Bernard, Manu, Lucie, Gildas, Quam, Natacha, Sammy, Guilhem, Florent, Lucile, Josiane, Anne, Maëlle, Nico, Eugène, Laurent, Florian, Bertrand, Émilie, Vincent, Manelle, Claire, Gurvan, Rachid et tous ceux que j'ai le malheur d'oublier. Une mention toute particulière pour Arthur et Jérémy, pour nos discussions légères ou absurdes, autour de bières bien réelles.

Merci aux amis de toujours Étienne, Fab, Julie, Héloïse, Léa, Dada, Maxime, Mathilde, Guillaume et les autres. Merci à Vivi, Steph' et Simon et leur petite Élise, à Jean-Louis et Catherine. Merci à mon frère, Matthieu. Ce travail et plus généralement ce que je suis dois tellement à mes parents qu'un simple merci en paraît presque inconvenant. Malgré tout, merci. Infiniment.

J'ai une pensée émue pour mes grands-parents, partis pendant l'aventure.

Deux choses magnifiques sont apparues pendant ce travail : Merlin et Roxane. J'hésite à vous remercier pour ce que vous avez apporter à cette thèse, par contre, je vous porte un amour et une tendresse infinis pour ce que vous me donnez à chaque instant.

Enfin, ce travail n'aurait sûrement jamais existé sans toi, Séverine, puisque tu es dans chacun de mes mots.

Résumé

Un enjeu majeur des modèles océaniques est de représenter fidèlement les circulations méso- et subméso-échelles afin de simuler leur importante contribution dans la circulation générale et dans le budget énergétique de l'océan. La poursuite de cet objectif se traduit par une augmentation de la résolution spatiale et temporelle à la fois des modèles et des réseaux d'observation de l'océan. Cependant, à ces petites échelles, la dynamique de l'écoulement revêt un caractère fortement turbulent ou non-linéaire. Dans ce contexte, les méthodes actuelles d'assimilation de données (AD), variationnelles en particulier, sont généralement moins performantes que dans un contexte (quasi-) linéaire.

L'objectif de cette thèse est d'explorer sous divers aspects le comportement des méthodes variationnelles d'AD dans un modèle d'océan non-linéaire. Pour ce faire, nous avons réalisé une série d'expériences dites « jumelles » en assimilant des données altimétriques simulées suivant les caractéristiques des satellites altimétriques Jason-1 et SARAL/AltiKA . À l'aide de ces expériences, nous analysons sous différents angles les problématiques posées par les non-linéarités à l'AD. Enfin, nous ouvrons plusieurs pistes d'amélioration de l'efficacité du système d'AD dans ce contexte.

Ce travail est basé sur le logiciel de modélisation océanique NEMO, incluant la configuration de bassin océanique turbulent idéalisé SEABASS, à différentes résolutions spatiales. Dans la continuité de la plateforme de recherche en AD avec NEMO, NEMO-ASSIM, nous avons utilisé et contribué au développement de cet ensemble d'outil, comprenant, entre autre, opérateur d'observation, modèles linéaire tangent et adjoint de NEMO, permettant de mener à bien notre étude. Le système d'AD variationnelle utilisé est le logiciel NEMOVAR.

Les résultats présentés tentent de lier les échelles caractéristiques des structures d'erreurs d'analyse et l'activité aux petites échelles. Pour ce faire, nous avons utilisé une large gamme de diagnostics, e.g. erreur quadratique moyenne spatiale et temporelle, caractéristiques des fonctions coûts, caractérisation de l'hypothèse linéaire tangente, PSD des champs d'erreurs d'analyse.

Nos expériences montrent que le 4DVAR incrémental contrôle efficacement la trajectoire analysée au 1/4° pour de longues fenêtres d'AD (2 mois). Lorsque la résolution augmente, la convergence de l'algorithme apparaît plus lente voire inexistante sous certaines conditions. Cependant, l'algorithme permet encore de réduire convenablement l'erreur d'analyse. Enfin, l'algorithme 3DFGAT se révèle beaucoup moins performant, quelle que soit la résolution.

De plus, nous montrons également l'importance de l'adéquation entre la circulation simulée et l'échantillonnage altimétrique, en terme d'échelles spatiales représentées, pour obtenir de

meilleures performances. Enfin, nous avons exploré la stratégie de minimisation dite « progressive », permettant d'accélérer la convergence du 4DVAR à haute résolution.

Asbtract

A current stake for numerical ocean models is to adequately represent meso- and small-scale activity, in order to simulate its crucial role in the general ocean circulation and energy budget. It is therefore also a challenge for data assimilation (DA) methods to control these scales. However this small-scale activity is strongly linked to the nonlinear or turbulent character of the flow, whereas DA methods are generally much less efficient in such contexts than in (almost) linear ones. For variational DA methods as incremental 4DVAR, non-linearities imply convergence difficulty, the cost functions to be minimised presenting multiple local minima.

The purpose of this thesis is to address this problem specifically, by exploring the behaviour of variational DA methods in a non-linear ocean model. To achieve this objective, a series of "twin" experiments assimilating simulated altimeter data, following the characteristics of altimetric satellite Jason-1 and SARAL/AltiKA, are analyzed. We also find different ways to improve efficiency of variational algorithms applied to turbulent circulations.

This work is based on oceanic modelisation software called NEMO, including a idealized turbulent oceanic basin configuration, SEABASS, and DA components (e.g. Observation operator, Linear Tangent and Adjoint Models). Thanks to NEMO-ASSIM research platform, we have used and developed this set of tools. The used variational DA system itself is NEMOVAR.

We present results characterizing scales and structures of the analysis error along the assimilation process, as well as tentative links with small scale activity. To study both the algorithm convergence and the analysis and forecast errors in a qualitative and quantitative way, a large spectrum of systematic diagnostics has been employed, e.g. spatial and temporal RMSE, cost function characteristics, projection of error fields on EOFs, validity of the tangent linear hypothesis, PSD of error fields.

In our experiments, it appears that the incremental 4DVAR algorithm proved to be quite robust for long DA windows at eddy-permitting resolution. When the model horizontal resolution increases, the convergence of the minimisation algorithm is poorer but the 4DVAR method still controls efficiently analysis error. It has also been shown that the 4DVAR algorithm is clearly more performant than 3DFGAT for both considered resolutions.

Moreover we investigate some strategies for DA in such nonlinear contexts, with the aim of reducing the analysis error. We performed so-called progressive incremental 4DVAR to improve the algorithm convergence for longer assimilation windows. Finally, we show that the adequation in represented flow scales between the model and the altimetric sampling is crucial to obtain the best error analysis reduction.

Table des matières

Introduction générale **13**

I L'assimilation variationnelle de données en océanographie : les problématiques actuelles et futures **23**

1 Un océan turbulent **25**
1.1 L'importance de la circulation méso-échelle 25
1.1.1 Des fluctuations et des instabilités 25
1.1.2 Vers la subméso-échelle... 28
1.1.3 Un monde de non-linéarités 29
1.2 Observation de la méso- et subméso-échelle 33
1.2.1 Les mesures *in-situ* 33
1.2.2 L'altimétrie 34
1.3 Conclusion 38

2 ... face à l'assimilation variationnelle de données. **41**
2.1 Objectifs et ingrédients de l'assimilation de données 41
2.1.1 Les particules élémentaires 41
2.1.2 Les différentes approches du problème 44
2.2 L'approche variationnelle 48
2.2.1 Un problème d'optimisation 48
2.2.2 Zoologie des formulations variationnelles 49
2.2.3 Caractéristiques des méthodes variationnelles 54
2.3 Les grands défis posés par l'océanographie 54
2.3.1 La problématique de la dimension 54
2.3.2 Réseau d'observation et conditionnement de la fonction coût 55
2.3.3 L'assimilation de données face à la non-linéarité 55
2.4 Conclusion 58

II Implémentation d'un système d'assimilation variationnelle de données pour une configuration idéalisée d'un bassin océanique **59**

3 Description du modèle de circulation océanique NEMO et du système d'assimilation variationnelle de données associé NEMOVAR **61**

3.1 Présentation de NEMO, framework de modélisation océanographique 61
3.1.1 Introduction générale 61
3.1.2 Le modèle de circulation océanique OPA 62
3.2 NEMO-ASSIM, un ensemble d'interfaces pour l'assimilation de données 65
3.3 NEMOTAM : les modèles linéaire tangent et adjoint de NEMO 73
3.3.1 Introduction and history 73
3.3.2 Methodology and choices for NEMOTAM 75
3.3.3 Validation 76
3.4 NEMOVAR : Système d'assimilation variationnelle de données pour NEMO 80
3.4.1 Introduction 80
3.4.2 Modélisation des matrices de covariances d'erreurs 80
3.4.3 Algorithme de minimisation 82

4 La configuration SEABASS de NEMO : un modèle académique pour l'assimilation de données 85
4.1 Présentation de SEABASS 85
4.1.1 Motivations 85
4.1.2 Descriptions physique et numérique de SEABASS 87
4.2 Description de la circulation océanique simulée par SEABASS 90
4.2.1 Comportement général 92
4.2.2 Statistiques de l'écoulement dans SEABASS 94
4.2.3 Indicateurs de non-linéarité 100
4.3 Conception des expériences d'assimilation de données avec SEABASS 102
4.3.1 Ébauche et état vrai 102
4.3.2 Génération des observations 103
4.3.3 Paramétrisation de NEMOVAR 104

III Étude qualitative et quantitative de l'algorithme 4D-VAR incrémental appliqué à la modélisation océanographique 107

5 Applications du 4D-VAR incrémental à SEABASS au 1/4° 109
5.1 Analyse d'une expérience d'assimilation « bien réglée » 109
5.1.1 Évolution des fonctions coûts 109
5.1.2 Etude de l'erreur d'analyse 113
5.1.3 Etude de l'erreur de prévision 120
5.1.4 Comparaison 3D-FGAT/4DVAR 122
5.2 Sensibilité des performances du système d'AD aux différents paramètres de minimisation en contexte non-linéaire 127
5.2.1 Longueur de la fenêtre temporelle d'assimilation 127
5.2.2 Influence du nombre d'itérations de boucle interne 132
5.3 Conclusion 134

6 Influences de l'intensification des non-linéarités sur le 4D-VAR incrémental 135
6.1 Analyses d'expériences pour SEABASS au 1/12° 135
6.1.1 Comparaison SEABASS4 vs SEABASS12 135

6.1.2 Rôle de la longueur de fenêtre d'AD 137
6.1.3 Impact des paramètres de minimisation 140
6.1.4 4DVAR vs. 3DFGAT 141
6.1.5 Synthèse 142
6.2 Améliorations des performances du 4DVAR pour la haute résolution 143
6.2.1 Diminuer la distance entre l'ébauche et l'état vrai 143
6.2.2 Accorder le réseau d'observation au système d'AD 147
6.2.3 Considérations méthodologiques : l'assimilation progressive ou quasi-statique 153
6.3 Conclusion 160

7 Vers la très haute résolution : exploration des difficultés de l'assimilation de données 163
7.1 Considérations générales sur la très haute résolution 163
7.2 Caractéristiques physiques et numériques de SEABASS au 1/100° 165
7.2.1 Description succinte de la configuration 165
7.2.2 Quelques considérations de coût numérique 165
7.2.3 Variabilité et aspects énergétiques de SEABASS100 167
7.3 Extrapolation des résultats obtenus pour la très haute résolution 170
7.3.1 Méthodes variationnelles d'AD et chaoticité 170
7.3.2 Très haute résolution et flux d'information 174
7.4 Synthèse 174

Conclusion 177

Introduction

Une histoire d'interactions L'assimilation de données et l'océanographie physique sont aujourd'hui deux disciplines scientifiques fortes d'un historique respectable, d'une communauté vivante et d'une connaissance solidement construite, dont la richesse de leurs perspectives respectives et conjointes confine au vertige.

Pour bien comprendre les difficultés posées par les développements présents et futurs de l'océanographie physique aux méthodes d'assimilation de données, il faudrait pouvoir embrasser à la fois les fondements et enjeux des deux domaines. Il faut pourtant en présenter un, puis l'autre. Le sujet d'étude étant l'assimilation de données, il paraîtrait naturel de l'aborder dès le début, au risque d'y perdre de vue les enjeux portés par le champ d'application qui nous intéresse. Commençons alors par l'océanographie, ses questionnements et ses défis, en nous autorisant à perdre potentiellement de vue leurs impacts sur l'assimilation de données…

Cette première difficulté, complètement formelle, qui ne devrait regarder que moi, illustre pourtant un aspect problématique, complexe et passionnant tout à la fois, de la réunion de systèmes complexes : les interactions qui en résultent. Le mot « interaction » peut se lire dans de nombreuses situations :

- L'océan présente plusieurs régimes d'écoulements, régis par de nombreux processus. La circulation générale océanique est le résultat des interactions, souvent non-linéaires, entre ces différentes dynamiques.
- D'un point de vue méthodologique, les spécificités des sciences géophysiques induisent de nombreuses contraintes sur l'assimilation de données.
- A contrario, la rétroaction des méthodes d'assimilation de données sur le système géophysique est dépendante desdites méthodes.

Ainsi, les interactions, imbrications, rétroactions de dynamiques d'écoulements et de processus physiques d'une part et de la réunion de l'océanographie physique et de l'assimilation de données d'autre part, sont au cœur de notre travail.

Un océan complexe à étudier En considérant par exemple à la fois la circulation thermohaline s'étendant sur des milliers de kilomètres, à des vitesses de l'ordre de 1 $cm.s^{-1}$ sur plusieurs dizaines d'années, et les courants de marées, induisant un transport des masses d'eau sur quelques kilomètres à des vitesses de 1 $m.s^{-1}$, nous entrevoyons alors que l'océan est le terrain de jeu de dynamiques physiques complexes, s'étendant sur un large spectre d'échelles spatiales et temporelles.

Au cours des trois dernières décennies, un aspect crucial pour la compréhension de la cirulation générale et ses applications est clairement apparu : l'océan est empli de processus dynamiques évoluant à des échelles spatiales de l'ordre de quelques dizaines de kilomètres [1] : la

1. De l'ordre du premier rayon interne de déformation

dynamique méso-échelle. Elle est principalement de nature turbulente. Physiquement, des instabilités de grandes échelles vont provoquer l'apparition de structures de plus petites échelles (tourbillons, fronts, filaments) qui à travers leurs dynamiques transféreront de l'énergie aux plus petites échelles ou la réinjecteront dans les grands courants océaniques.

Autrement dit, comprendre la circulation océanique implique non seulement de comprendre les dynamiques évoluant à différentes échelles spatiales et temporelles : il faut également comprendre leurs interactions.

D'un point de vue mathématique, le fait que la somme des différents mécanismes physiques ne décrit pas la circulation générale et le fait que l'évolution d'un modèle n'est pas forcément proportionnelle à ses forçages (variabilité intrinsèque de l'océan) signifient que ce système n'est pas linéaire. Par conséquent, pour traduire la complexité théorique de la physique de l'océan entrevue ici, nous utiliserons régulièrement le terme de *non-linéarité*. En s'intéressant aux équations continues décrivant l'écoulement océanique, il apparaît que le principal terme non-linéaire correspond au terme d'advection $\vec{u}\nabla\vec{u}$ ($\vec{u}$ étant le champ de vitesse). C'est précisément ce terme qui va introduire les rétroactions entre les différentes dynamiques de l'écoulement.

Si bien d'autres processus, tels que les forçages atmosphériques (vents, radiation solaire) ou les effets de bathymétrie contraignent les écoulements océaniques, la dynamique méso-échelle et son influence dans la circulation générale est majoritairement gouvernée par le terme non-linéaire de l'équation de quantité de mouvement évoqué précédemment. C'est pour cette raison que nous nous concentrerons sur ses effets, à la fois dans l'établissement de nos questions scientifiques ainsi que dans la mise en place de notre cadre expérimental.

Il faut également souligner que ces dernières années, un intérêt fort est porté à la dynamique subméso-échelle (*i.e.* d'échelles spatiales inférieures au rayon de déformation interne), dynamique pouvant avoir un impact certain sur les processus biogéochimiques océaniques, par exemple, correspondant aux mêmes échelles par exemple.

Pour comprendre, expliquer et prévoir la circulation océanique, il faut pouvoir la décrire. Pour ce faire, deux outils, hétérogènes mais complémentaires sont disponibles : les observations et la modélisation.

Un océan complexe à observer et modéliser. Un outil d'expérimentation communément utilisé en océanographie physique prend la forme des équations décrivant l'écoulement océanique : les modèles. Ils sont le réceptacle de la connaissance des lois régissant la caractérisation et la circulation des masses d'eau de l'océan. L'état dynamique de l'océan est représenté par des « variables d'état » : la température de l'eau, sa salinité, sa vitesse et sa hauteur en surface. Ces équations sont déduites des lois de conservation de masse, d'énergie et des lois de la thermodynamique et, à ce jour, elles n'admettent pas de solution analytique. Elles sont alors discrétisées sur une grille numérique afin de pouvoir simuler l'évolution de l'état de l'océan. Nous pouvons noter d'ores et déjà que cette source d'information sur l'océan est entâchée d'erreurs :

- Les équations utilisées en océanographie résultent d'hypothèses simplificatrices dont la validité peut être mise à mal selon les échelles considérées.
- Le problème discrétisé est par nature différent du problème continu, ce qui constitue une deuxième source d'erreur.

Comme nous l'avons expliqué précédemment, pour bien comprendre la circulation océanique dans son ensemble, il est nécessaire de représenter des dynamiques évoluant sur une large gamme d'échelles spatiales et temporelles. Ceci induit que le système océan (et plus largement de nombreux systèmes géophysiques) présente un nombre important de degrés de liberté. Plus particulièrement, en voulant représenter les phénomènes méso- (et subméso-) échelles, ce nombre s'accroît encore. En terme de modélisation, ceci a deux implications fortes :

- *Des difficultés liées à l'augmentation de la résolution*
  Pour comprendre particulièrement ces échelles, il faut les représenter. En d'autres termes, cela revient à augmenter la résolution de la discrétisation. Ceci a un fort impact sur les modèles, que ce soit au niveau des propriétés des schémas numériques employés ou du coût numérique des simulations.
- *La paramétrisation « sous-maille »*
  Au vu des problèmes soulevés ci-dessus par l'augmentation de la résolution, il n'est aujourd'hui pas possible de résoudre numériquement toutes les échelles, particulièrement les plus petites (de la subméso-échelle à l'échelle moléculaire). Seulement, ces échelles jouent un rôle clé dans le circuit énergétique de l'océan dynamique, notamment en tant que principal puit d'énergie via la dissipation moléculaire, par exemple. Comme ces échelles ne sont pas représentables, ces mécanismes sont alors paramétrisés de façon à prendre en compte leurs effets sous-maille (dissipation, par exemple) dans le modèle.

Historiquement, la connaissance de la circulation océanique s'est construite (et se construit encore) en premier lieu à partir d'une autre source d'information : l'observation de l'océan. Aujourd'hui, les instruments de mesure de l'océan peuvent être distingués en deux grandes catégories :

- *Les observations in-situ*
  Ce type d'observations consiste à effectuer des mesures à l'aide d'instruments en contact direct avec l'océan, via des réseaux de mouillages, des campagnes de mesure, des instruments embarqués sur des navires d'opportunité, ou des flotteurs (comme le réseau ARGO, par exemple). Le grand avantage de ce type d'observations est l'accès à la dimension verticale de l'océan, la plupart de ces instruments relevant des mesures sur une partie conséquente de la colonne d'eau. En revanche, le coût des flotteurs, inhérent à tout appareil de mesure *in situ*, fait que ces types de mesures ne sont que très partielles comparativement à la taille de l'océan.
- *Les observations satellitaires*
  Depuis une trentaine d'années, des satellites observent la surface de l'océan. En mesurant principalement la hauteur de l'eau, ces satellites successifs (programmes TOPEX/Poséidon, Envisat, Jason 1 et 2, ...) fournissent un grand nombre de données, denses en espace et en temps. Les satellites encore opérationnels aujourd'hui réalisent des mesures altimétriques le long de traces espacées d'une dizaine à quelques centaines de kilomètres. D'autres programmes satellites permettent d'observer d'autres quantités telles que la température de surface, la salinité ou la couleur de l'eau. Pour répondre aux enjeux de la compréhension de l'océan turbulent, grâce aux développements de nouveaux instruments de mesure, le programme altimétrique SWOT (lancement prévu en 2019) prévoit de fournir des données d'une résolution de quelques dizaines de mètres seulement sur de larges fauchées et non plus seulement le long de la trace satellite.

Il est à noter qu'il existe également des sources d'observations non-reliées directement à l'océan dynamique mais dont l'information peut être utile aux océanographes. Par exemple, les satellites observant la couleur de l'eau, qui nous renseigne sur l'océan biogéochimique (appelé aussi *océan vert*) peut devenir une source d'information pour la circulation dynamique des masses d'eau, par exemple à l'aide de modèles couplés ou de méthodes d'assimilation d'images basées sur la reconnaissance de structures (*cf.* Gaultier et al. (2013), par exemple).

Ainsi, nous venons de voir que l'enjeu propre à l'océanographie d'accéder aux interactions de processus physiques complexes évoluant sur une large gamme d'échelles spatiales et temporelles impacte fortement les deux sources d'informations disponibles sur l'état dynamique de l'océan.

Et l'assimilation variationnelle de données dans tout ça ? La science de l'assimilation de données (*AD* dans la suite) a pour mission de réunir les informations provenant des observations et des modèles afin de construire une vision la plus précise possible de l'océan passé, présent et futur, pour une large gamme d'applications sociétales.

Plus précisément, pour réaliser une prévision par exemple, le principe est le suivant : il « suffit » d'intégrer un modèle à partir d'une condition initiale donnée, la plupart du temps obtenue à partir d'une prévision antérieure. Le problème est que cette condition initiale n'est qu'une estimation imparfaite de l'état vrai. Les méthodes d'AD, dans ce cas précis, auront pour objectif d'estimer le plus précisément possible cet état initial à partir des observations, des équations physiques via le modèle numérique et des informations disponibles sur les erreurs associées à ces deux sources. Les méthodes d'AD peuvent cependant s'étendre à bien d'autres applications que la recherche de la condition initiale : *e.g.* contrôle des conditions limites, des paramètres numériques et physiques du modèles.

Un classement usuel en AD divise les méthodes en deux grandes catégories :

- Les méthodes de filtrage, basées principalement sur le filtre de Kalman étendu à des problèmes non-linéaires. Ces méthodes nécessitent de manipuler des matrices de covariances dont la taille est conditionnée par la dimension du modèle. Ainsi, pour des raisons d'applicabilité, des versions sous-optimales, basées sur des techniques de réduction de rang, ont été et sont encore développées. Ces méthodes sont aujourd'hui performantes, permettant d'obtenir une prévision de bonne qualité mais aussi une estimation de l'erreur qui lui est associée. Elles sont ainsi utilisées dans de nombreux centres opérationnels. Aujourd'hui la tendance est à une génération ensembliste de ces méthodes.
- Les méthodes fondées sur la théorie du contrôle optimal, appelées communément *méthodes variationnelles*. À chaque condition initiale du modèle est associée une trajectoire du modèle à partir de laquelle on calcule l'écart aux observations. L'objectif est de minimiser une fonction *coût* représentant cet écart en fonction de la condition initiale, en introduisant les statistiques d'erreurs (semblable à celle formulées dans les méthodes de filtrage), qui peuvent être vues comme la « confiance » que l'on va accorder aux différentes sources.

Dans ce travail de doctorat, nous nous intéressons aux méthodes variationnelles dont le principal avantage est de construire un système d'optimalité rassemblant toute l'information disponible. Largement répandues dans les centres de météorologie opérationnels, l'utilisation des méthodes variationnelles est réelle en océanographie opérationnelle (*e.g. ECMWF, UK Met-Office*) mais reste plus confidentielle si l'on considère les algorithmes dérivés du 4D-VAR.

Les études passées des méthodes variationnelles d'AD en océanographie sont toutefois nom-

breuses et portent sur plusieurs axes. Tout d'abord, de nombreux travaux œuvrent à l'amélioration méthodologique des algorithmes d'AD variationnelle, à plusieurs niveaux :

- L'amélioration des différents composants des méthodes variationnelles comme la modélisation de la matrice B de covariances d'erreur d'ébauche, pour l'océanographie, avec les travaux de Weaver et Courtier (2001), qui se sont prolongés avec notamment Weaver et al. (2005), Ricci et al. (2005) et Mirouze et Weaver (2010). D'autres stratégies se basent sur la réduction d'ordre du problème, comme par exemple celles présentées dans Robert et al. (2005a) et Cao et al. (2007). Nous pouvons également citer les travaux sur la méthode d'application de l'incrément *Incremental Analysis Update* de Bloom et al. (1996).
- L'amélioration de l'algorithme de minimisation proprement dit, avec le travail sur les classes de préconditionneurs à mémoire limité, présenté dans Tshimanga et al. (2008)
- Le développement de nouvelles stratégies de minimisation, dans le contexte océanique, avec la présentation de l'assimilation progressive (appelée quasi-statique pour Pires et al. (1996)) dans Luong et al. (1998).

Ensuite, de nombreuses études présentent l'application et les performances de systèmes variationnels d'AD appliqués à des modèles d'océan de différentes natures :

- Sur des configurations idéalisées, basées sur les équations quasi-géostrophiques, comme dans Moore (1991), Ishikawa et al. (2001) ou Rémy et al. (2002).
- Plus récemment, plusieurs travaux, basés sur différents modèles océaniques aux équations primitives peuvent être signalés : citons entre autre ceux prenant appui sur ROMS-4DVAR, comme par exemple Powell et al. (2008) ou Moore et al. (2011), ceux basés sur le modèle NEMO, pour une configuration nord-atlantique par Ferron (2011) ou méditerranéenne Dobricic et Pinardi (2008).

À travers les travaux fondateurs en assimilation de données de Le Dimet et Talagrand (1986) pour les méthodes variationnelles ou Pham et al. (1998b) pour les méthodes de filtrage, aux améliorations méthodologiques apportés par Luong et al. (1998) pour les méthodes variationnnelles et Cosme et al. (2010) pour les méthodes séquentielles, ainsi qu'à la forte implication dans différents projets et études d'AD appliqués à l'océanographie physique, ce travail s'inscrit logiquement dans le contexte géographique Grenoblois, articulé autour des équipes MEOM (anciennement appartenant au LEGI [2] et, depuis 2013 au LGGE [3]) et MOISE (LJK [4]). Le projet ANR VODA (2009-2012) s'est inscrit dans la continuité de ces travaux cités plus haut. Il visait à contribuer au développement un système d'AD variationnelle autour du logiciel de modélisation océanique NEMO et a été le cadre de ce travail de doctorat.

Les évolutions effectives et potentielles des observations et des modèles présentées précédemment ont des répercussions directes sur les méthodes d'AD, en particulier variationnelles. Les principaux défis à relever aujourd'hui sont les suivants :

- La dimension (au sens mathématique) des objets manipulés par les algorithmes s'accroît sans cesse.
- Les statistiques d'erreurs sur les modèles et les observations sont mal connues.
- Le domaine de validité théorique de ces méthodes est relativement restreint.

Il faut noter que ces problématiques ne sont pas nouvelles. Cependant, l'évolution présente et future de l'océanographie physique, notamment en ce qui concerne la dimension des modèles,

2. Laboratoire des Écoulements Géophysiques et Industriels
3. Laboratoire de Glaciologie et de Géophysique de l'Environnement
4. Laboratoire Jean Kuntzmann

les rendent encore plus concrètes qu'auparavant.

La malédiction de la dimension en assimilation de données Comme nous l'avons évoqué précédemment, les systèmes géophysiques sont par nature de grande dimension et la tendance est d'augmenter la résolution des modèles et des observations, ce qui mène logiquement à un accroissement du volume de données à traiter pour les méthodes d'AD. Ceci pose clairement plusieurs problèmes pour celles-ci :

- Le premier, le plus évident, est l'accroissement du coût numérique de ces méthodes. Aujourd'hui, la dimension d'un état de l'océan simulé oscille entre 10^6 et 10^9, et celle d'un jeu d'observation de l'ordre de 10^6. Pour les algorithmes les plus sophistiqués, le coût approximatif d'une expérience d'assimilation de données est de quelques dizaines de fois celle d'une simulation du modèle seul. Ainsi, il faut multiplier le coût numérique des modèles par ce facteur afin d'estimer le coût d'une méthode d'assimilation de données. Selon le point de vue, pessimiste ou optimiste, sur l'évolution des moyens de calcul dans les années à venir, il est possible d'envisager que les méthodes actuellement utilisées ne soient plus applicables sur des modèles futurs.
- Les espaces de grande dimension limitent également le choix des algorithmes mathématiques applicables.
- Des observations de plus en plus nombreuses, spatialement et temporellement posent paradoxalement des problèmes pour les méthodes variationnelles. Il est connu que la forte densité spatiale et temporelle des observations peuvent dégrader le conditionnement de la fonction coût, la rendant plus ardue à minimiser.

Des incertitudes sur les erreurs Avec l'augmentation de la densité spatiale et temporelle des observations disponibles et de la résolution des modèles, une difficulté historique des sytèmes d'assimilation se renforce : la connaissance et la quantification des erreurs sur les deux sources d'informations propres aux méthodes d'assimilation.

Comme nous l'avons déjà évoqué, la complexité théorique et numérique des modèles augmentant, la dimension des structures d'erreurs s'accroît également. Il est également raisonnable de penser que la prise en compte de nouveaux phénomènes encore mal connus injectera de nouvelles sources d'erreurs difficiles à spécifier dans les méthodes d'AD.

En voulant conjuguer les observations et les modèles, l'assimilation de données se retrouve confrontée à un défi de taille : l'erreur de représentativité. En effet, si l'on pense au programme SWOT, qui fournira des données d'une résolution horizontale de quelques dizaines de mètres, nous observerons dans un avenir proche des processus physiques évoluant dans cette gamme d'échelles spatiales. Cependant, les modèles, pour des raisons partiellement évoquées précédemment, ne pourront pas représenter ces processus.

Ainsi, lors de l'application des méthodes d'assimilation de données, il faut tenir compte des différences de niveaux de représentations existants entre les observations et le modèle.

Les méthodes variationnelles face aux non-linéarités Les processus méso- et submésoéchelles et leurs interactions avec les autres échelles de l'écoulement océanique présentent un fort caractère non-linéaire voire chaotique. La *chaoticité* se définit comme la grande sensibilité

de la trajectoire d'un système numérique donné à ses conditions initiales, au sens large (condition initiale de l'état du modèle, paramétrisation du modèle, sensibilité aux perturbations numériques).

Pour l'AD, qui peut être vue comme un souhait de contraindre cet écoulement selon les différentes informations dont on dispose, notamment en contrôlant la condition initiale, ceci a des impacts méthodologiques forts. En effet, la validité de la plupart des algorithmes d'AD est basée sur des hypothèses de linéarité et de gaussianité des objets physiques à contrôler. Pour l'AD variationnelle en particulier, les non-linéarités dues à la nature turbulente de la circulation océanique méso- (et subméso-) échelle met à mal directement plusieurs hypothèses clés des algorithmes.

La non-linéarité, dans les modèles d'océan, peut se traduire par un temps de décorrélation temporelle court. En d'autres termes, la corrélation entre un état à l'instant t et l'état à l'instant $t+T$ va diminuer avec l'augmentation du degré de non-linéarité du système. Ainsi, par exemple, pour l'algorithme 3D-FGAT, ne prenant pas en compte l'évolution temporelle de la trajectoire simulée en regard des instants d'observations, le contrôle de la condition initiale pour des fenêtres longues d'assimilation ne sera plus pertinent.

Sans rentrer maintenant dans les détails, les algorithmes dérivant du 4DVAR prennent en compte l'évolution temporelle du modèle pour la construction de la fonction coût et du calcul de son gradient. Cependant, pour un modèle et un opérateur d'observations non-linéaires, cette fonction coût n'est plus quadratique : elle n'est plus convexe et peut admettre plusieurs minimums locaux, ces propriétés s'accentuant avec l'augmentation de la non-linéarité du système considéré. Conséquemment, le système de minimisation n'est plus optimal et la convergence de l'algorithme n'est plus assurée.

L'algorithme du 4DVAR incrémental, consistant à approcher la fonction coût non-quadratique par une succession de fonctions coût quadratiques, repose sur l'hypothèse que la propagation d'une perturbation par le modèle direct peut être correctement approximée par sa propagation par le modèle linéarisé, pour une fenêtre temporelle d'assimilation donnée. Plus le modèle est non-linéaire, plus la divergence entre les deux propagations s'accélère. Ceci n'autorise plus les fenêtres longues d'assimilation, pourtant essentielles au contrôle des modes stables et instables se développant dans le modèle, contrôle garantissant une bonne trajectoire analysée.

Les problématiques abordées Les problèmes présentés ici concernent de nombreux domaines scientifiques. Il serait illusoire de s'attaquer de front à toutes ces questions, qu'elles concernent la méthodologie de l'AD, l'océanographie physique, la modélisation numérique, etc.

Nous avons choisi de nous intéresser à l'impact de la turbulence de la circulation méso-échelle sur un système d'AD variationnelle, plus particulièrement d'un point de vue méthodologique. Ainsi, nous avons considéré un bassin océanique académique à diverses résolutions horizontales pour accéder à différentes précisions de la représentation de cette turbulence. L'aspect académique de la configuration nous enlève un certain nombre de contraintes de modélisations, qui auraient pu éventuellement brouiller le signal en sortie du système d'assimilation d'AD pour notre étude, comme par exemple les effets de topographie ou les forçages atmosphériques.

Au cours de ce travail, nous avons cherché, à travers une série d'expériences, à diagnostiquer les réponses propres du système d'AD aux non-linéarités :

- Quelles sont les performances du système d'AD, notamment en terme d'erreurs d'analyse et de prévision, pour un système non-linéaire ? Pour répondre à cette question, nous avons établi quelques indicateurs en tentant de caractériser quantitativement et qualitativement

ces non-linéarités.
- Quelles sont les paramétrisations optimales de ce système face à un niveau de turbulence donné et quelles peuvent être les pistes méthodologiques d'amélioration des algorithmes étudiés ?

Un détour par l'ingénierie Étant ingénieur de formation, j'ai très vite été intéressé par le défi technique présenté par l'AD en océanographie. La grande dimension, des algorithmes complexes, des modèles conséquents, des entrées/sorties volumineuses, le tout agrémenté de calcul parallèle haute performance, c'est une aubaine pour quelqu'un versé dans le calcul numérique au sens pratique du terme.

Au cours de ma thèse, j'ai eu la chance de pouvoir postuler et d'être reçu à un concours d'ingénieur de recherche (IR) CNRS dont la mission est d'assurer entre autre le développement et la maintenance d'outils d'assimilation de données avec le modèle NEMO.

NEMO [5] est un cadre logiciel de modélisation pour l'océanographie et ses applications, permettant de :
- Résoudre numériquement les équations de la dynamique de l'océan (partant d'un système aux équations primitives), de la glace de mer, de la biogéochimie et la physique que leur est associée
- Et ceci dans des environnements donnés (comprenant zones géographiques, topographie, forçages atmosphériques, conditions limites)

L'assimilation de données requiert un certain nombre d'outils pouvant être rattachés au coeur numérique du modèle utilisé : l'opérateur d'observation (OBS dans NEMO), un opérateur d'ajout de perturbation (ASM dans NEMO). Pour certaines méthodes d'assimilation de données variationnelle, les modèles tangent et adjoint sont requis. En tant qu'IR, un aspect de mon rôle est de développer, documenter et maintenir ces derniers.

À travers ma volonté de rester proche de la technique, qui recèle aussi des défis passionants, ce travail comporte une forte dimension d'ingénierie, dont la traduction concrète est le poste que j'ai l'honneur d'occuper actuellement.

Plan du mémoire Dans une première partie, nous exposerons quelques généralités au sujet du présent et du futur proche de l'océanographie physique et de l'assimilation de données. Nous expliquerons dans le premier chapitre plus en détail les mécanismes physiques importants aujourd'hui pour développer et affiner la compréhension des dynamiques de l'océan ainsi que les enjeux qu'ils induisent en terme de modélisation et d'observation. Dans le second chapitre, après une introduction sur l'assimilation de données dans son ensemble et un exposé des fondements et des limites théoriques et numériques des méthodes variationnelles en particulier, nous détaillerons les impacts potentiels des évolutions actuelles et futures de l'océanographie précédemment décrites sur les systèmes d'AD.

Dans notre étude, nous nous focaliserons sur les effets des non-linéarités principalement associés aux dynamiques méso-échelles sur un système d'AD variationnelle. Ainsi, dans une seconde partie, nous présenterons et justifierons le cadre expérimental choisi :
- Le modèle numérique de dynamique océanique NEMO et ses outils d'assimilation de données NEMO-ASSIM (opérateur d'observation, application de l'incrément, modèles linéaire tangent et adjoint)

5. http ://www.nemo-ocean.eu/

- Le système d'AD variationnelle utilisé, NEMOVAR
- Et enfin la configuration académique de bassin océanique idéalisé à haute résolution horizontale SEABASS.

Enfin, dans la troisième partie, à l'aide d'une série d'expériences d'AD réalisée dans ce cadre numérique, en considérant plusieurs résolutions horizontales dégageant chacune différents « degrés » de non-linéarité, nous analyserons physiquement, statistiquement et méthodologiquement l'impact des processus non-linéaires méso-échelles sur les performances des algorithmes d'AD variationnelle.

Dans un dernier temps, nous étudierons et proposerons plusieurs pistes (de minimisation, de traitement des observations) pour tenter de répondre aux problématiques posées par la représentation des processus méso- et subméso-échelles dans les modèles, les observations et donc dans les systèmes d'assimilation de données.

Première partie

L'assimilation variationnelle de données en océanographie : les problématiques actuelles et futures

Chapitre 1

Un océan turbulent

Aujourd'hui, l'amélioration de la compréhension de la circulation océanique méso-échelle (circulation régie par des mécanismes physiques de quelques dizaines de kilomètres) apparaît comme un enjeu crucial de l'océanographie physique. Dans ce chapitre, nous exposerons brièvement les principaux mécanismes régissant cette circulation et leurs impacts sur la dynamique océanique de façon générale, ses applications et surtout sur les sources d'information de l'océan : la modélisation et les observations.

Notre objectif n'est pas de faire une présentation exhaustive de chacun des domaines abordés dans ce chapitre. Voici une liste de références, non-exhaustive, qui peuvent éclairer le lecteur sur chacun de ceux-ci :

- Sur la circulation océanique, le livre de Colling (2004) donne les éléments de bases pour appréhender les grands mécanismes physiques mis en jeu. Pour un niveau plus avancé, et plus généralement sur la mécanique des fluides géophysiques, nous renvoyons à Pedlosky (1987)
- Pour la turbulence, nous renvoyons à Lesieur (1984)
- En ce qui concerne la paramétrisation des modèles de circulation océanique, Chassignet et Verron (1998) nous semble un bon point d'entrée
- Enfin, Chassignet et Verron (2006) traite des différents aspects de la production à l'exploitation en passant par le traitement des observations océanographiques.

Nous nous attarderons sur les enjeux propres à la modélisation au cours des chapitres suivants.

1.1 L'importance de la circulation méso-échelle

1.1.1 Des fluctuations et des instabilités

En tout endroit de l'océan, le courant observé est généralement une combinaison de différentes composantes de l'écoulement : une onde inertielle peut par exemple se superposer à un courant piloté par le vent, dans une région où l'activité de la marée est intense. La rotation de la terre, les effets de topographie, la stratification du fluide, la gravité, les forçages atmosphériques, entre autres, sont à l'origine des nombreux processus présents dans l'océan et qui interagissent entre eux : *e.g.* des ondes se propageant à différentes vitesses de phase et de groupe ou des tourbillons cohérents, d'échelles variables en temps et en espace. La Figure 1.1 illustre la variété des processus océaniques et les échelles qui leurs sont associées.

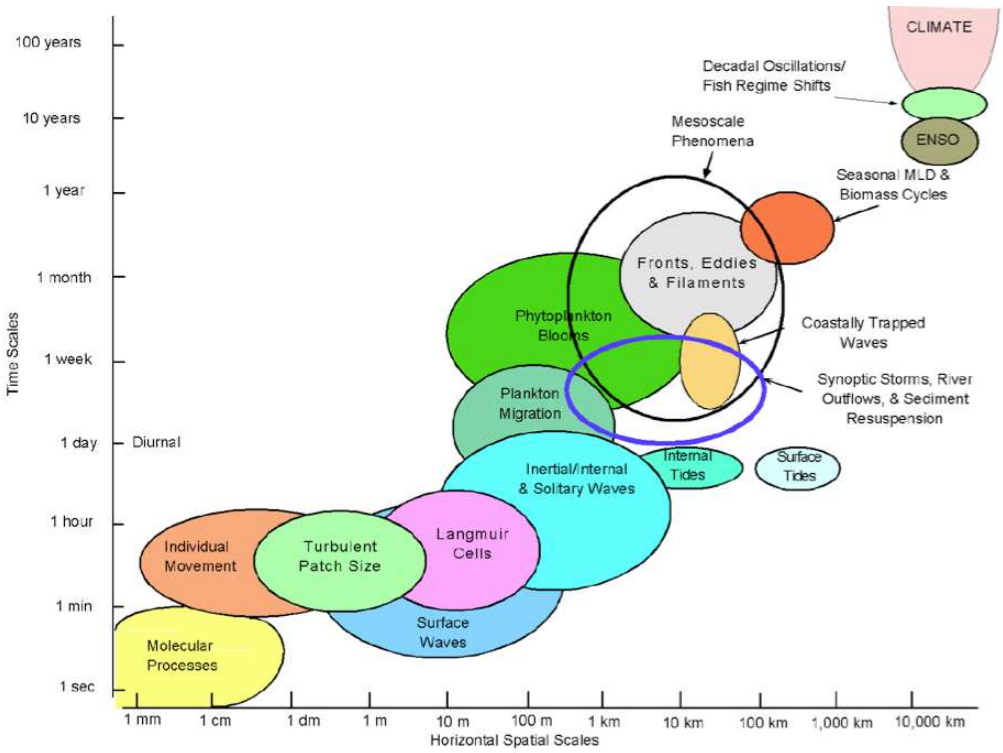

Figure 1.1 – *Illustration de différents processus physiques et biogéochimiques rencontrés dans le milieu océanique, classés selon leurs échelles temporelles et spatiales respectives. D'après Dickey et al. (2006)*

Ces différents mécanismes et surtout leurs interactions font apparaître dans l'écoulement océanique une **variabilité intrinsèque**, là encore à différentes échelles. Même en supposant que les forçages externes de l'océan soient stationnaires, les propriétés et les mouvements des masses d'eau montrent des fluctuations importantes autour de l'état moyen, qui lui aussi peut présenter une certaine variabilité. Au premier rang de ces fluctuations, se trouve la circulation méso-échelle.

Elle se définit comme une classe de phénomènes énergétiques d'échelles spatiales de l'ordre de grandeur du premier rayon interne de Rossby [1] (de quelques dizaines de kilomètres en hauturier à quelques kilomètres en milieu côtier) et d'échelles temporelles comprises entre quelques jours et plusieurs mois. **Son niveau d'énergie cinétique est d'un ordre de grandeur supérieur à celui de la circulation océanique moyenne**. Ses sources sont principalement les instabilités de la circulation générale, et dans une moindre mesure l'interaction des courants avec la bathymétrie et le forçage direct par le vent.

Via les transferts énergétiques induits par la turbulence, ces fluctuations vont fortement impacter la dynamique grande échelle de l'océan et ont donc une importance cruciale pour l'étude

1. Le premier rayon de déformation barocline ou interne est l'échelle spatiale naturelle à laquelle s'ajustent les processus physiques comme par exemple les fronts, les courants de pentes, les tourbillons :

$$\delta_{R'} = \frac{\sqrt{g' h}}{f}$$

où $g' = g\dfrac{\Delta\rho}{\rho}$ est appelée la gravité réduite. **Les échelles spatiales inférieures à ce rayon sont appelées fort logiquement subméso-échelle.**

de la circulation moyenne de l'océan et de son impact climatique.

Instabilités Les sources de variabilité de moyenne échelle sont multiples. Toutes les zones de mouvements intenses dans l'océan sont susceptibles d'être des lieux de formation de structures de moyenne échelle : les zones de forts courants, tel le bord ouest des océans, et plus généralement toutes les zones créatrices de flux de vorticité, liées à la bathymétrie, mais aussi les zones de mélange entre deux masses d'eau aux propriétés différentes.

Dans un écoulement en mouvement non-uniforme, il peut apparaître deux types d'instabilité à grande échelle : une instabilité barocline [2], essentiellement liée au cisaillement vertical de l'écoulement et à la stratification, et une instabilité barotrope [3] liée au cisaillement horizontal des courants sur la verticale. L'instabilité barocline entraîne une libération d'énergie potentielle, tandis que l'instabilité barotrope libère de l'énergie cinétique. Une instabilité sera souvent mixte barotrope-barocline, liant les caractéristiques des deux types. La circulation océanique est instable dans certaines régions, en particulier près des bords ouest et dans les zones de recirculation. Les instabilités créent des tourbillons de méso-échelle qui influencent en retour la circulation.

De nombreuses autres types de déstabilisations existent. C'est par exemple le cas de l'interaction d'ondes de natures différentes, ou encore du contact entre deux masses d'eaux aux propriétés thermohalines [4] différentes. Dans ce dernier cas, par exemple, l'instabilité va se traduire par une déformation de l'interface entre les deux masses d'eau. Selon l'intensité et la longueur d'onde préférentielle du phénomène, cette déformation peut aller de la simple filamentation avec intrusion d'une masse d'eau dans l'autre jusqu'au détachement de structures cohérentes.

Une dimension importante de l'activité méso-échelle est l'existence de structures cohérentes associées à des instabilités de courants de plus grande échelle. Ces structures cohérentes sont capables de se propager sur des distances de plusieurs milliers de kilomètres parfois, et ainsi propager des eaux issues d'une zone particulière de l'océan à de très grandes distances de leur lieu de formation. On entrevoit ici l'importance de ces phénomènes puisqu'ils participent activement au transport des propriétés dans l'océan, y compris d'un point de vue climatique.

Considérations énergétiques Pour comprendre un peu mieux l'importance de l'activité méso-échelle dans la circulation océanique, il faut faire un arrêt sur ses aspects énergétiques.

La théorie de la quasi-géostrophie, proposée par Charney (1971), Gill et al. (1974), Salmon (1978), modifiée par Held et Larichev (1996) et Smith et Vallis (2002), est une description simplifiée de l'écoulement océanique pour des échelles temporelles supérieures à la période de Coriolis $\frac{2\pi}{f}$ (*i.e.* pour un nombre de Rossby $\frac{U}{fL}$ petit devant 1). Son propos est de décrire l'évolution de l'écoulement sous ces hypothèses comme une succession d'équilibres géostrophiques [5]. Un de ses grands succès a été de mettre en évidence les mécanismes de l'instabilité barocline. Cette

2. Un fluide est dit barocline lorsque les lignes d'égale pression (isobare) croisent celles d'égale densité (isopycne) dans celui-ci.

3. Un fluide est dit barotrope quand les lignes d'égale pression sont parallèles à celles d'égale densité (isopycne)

4. La circulation thermohaline est la circulation à grande échelle dans l'océan mondial liée aux gradients de température et de salinité des masses d'eau.

5. L'équilibre géostrophique est le régime permanent d'écoulement océanique dans lequel les forces de pression horizontales sont en équilibre avec la force de Coriolis.

théorie permet également de mettre en évidence les grands mécanismes théoriques de transfert énergétique dans l'océan :

- L'instabilité barocline injecte de l'énergie potentielle dans les échelles supérieures au premier rayon de déformation interne
- Cette énergie potentielle est transférée vers les petites échelles par l'action du mode barotrope
- À l'échelle du premier rayon interne de déformation, il y a transfert d'énergie potentielle en énergie cinétique, de type barotrope.
- En dessous du rayon de déformation, il y a transfert d'une partie de l'énergie cinétique vers les plus petites échelles jusqu'à la dissipation moléculaire : la **cascade énergétique directe**.
- Au dessus du rayon de déformation, une partie de l'énergie cinétique barotrope est transférée vers les grandes échelles : la **cascade énergétique inverse**.
- À grande échelle, l'énergie cinétique est dissipée par la friction de fond.

Ainsi, la nature turbulente de l'océan à méso-échelle est ici un intermédiaire essentiel dans ces différents mécanismes de transfert et de conversion énergétiques.

La quasi-géostrophie est une description relativement simplifiée de la dynamique océanique qui reste qualitativement valide dans l'océan intérieur. Mais elle peut être en défaut, notamment sur la dynamique de surface méso- et subméso-échelles comme les travaux Le Traon et al. (2008) et Klein et al. (2008) l'indiquent. À la vue de son rôle clé en tant que prescripteur important des dynamiques biogéochimiques, l'approfondissement de la connaissance de la surface de l'océan passe par une plus fine compréhension des processus d'échelles inférieures au premier rayon interne de déformation, communément appelées la subméso-échelle.

1.1.2 Vers la subméso-échelle...

Les fronts et les filaments subméso-échelles jouent un rôle crucial dans l'océan. En premier lieu, ils constituent des barrières dynamiques qui peuvent isoler partiellement des structures mésoéchelles et plus généralement « contrôler » leur évolution et décomposition (*e.g.* Capet et al. (2008)). Ensuite, ils ont un rôle essentiel dans l'échange vertical des traceurs [6] à la surface de l'océan. Ils sont à l'origine d'une pompe verticale qui impacte profondément l'activité biogéochimique dans de nombreuses régions de l'océan global.

La Figure 1.2 illustre bien l'émergence de structures subméso-échelle (fronts, tourbillons, filaments) aux interfaces des structures méso-échelles.

6. Les traceurs sont des quantités physiques, biologiques ou chimiques (*e.g.* salinité, température) caractérisant les masses d'eaux de l'océan

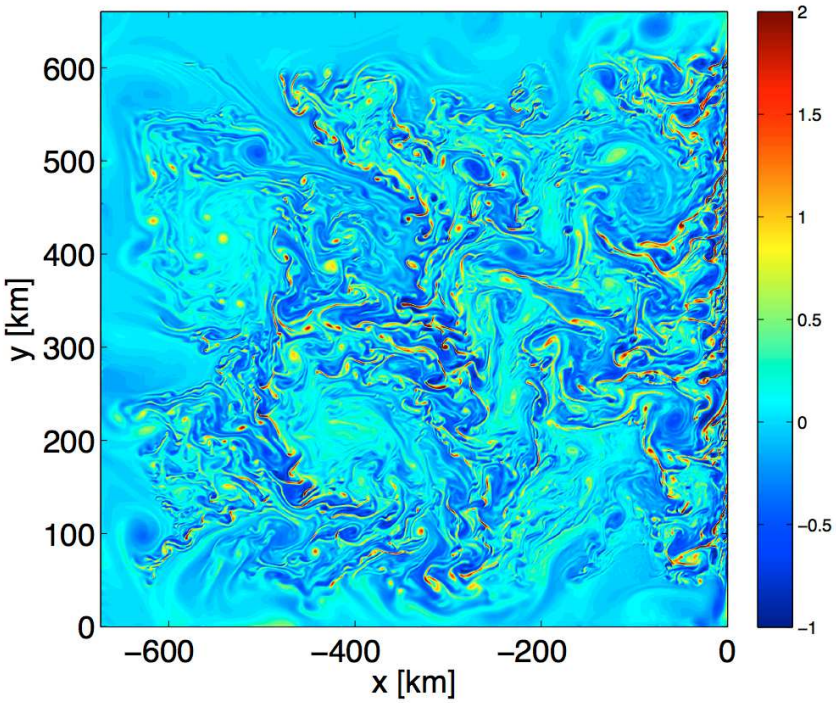

Figure 1.2 – *Vorticité verticale simulée normalisée par f dans le courant californien. On peut identifier clairement les structures subméso-échelles se développant aux interfaces des structures méso-échelles. D'après Capet et al. (2008)*

1.1.3 Un monde de non-linéarités

Non-linéarité et turbulence

- En mathématiques, la **non-linéarité** peut se définir comme le caractère d'un système dont le comportement n'est pas linéaire, c'est-à-dire ne satisfaisant pas le principe de superposition, ou, dit différemment, dont la sortie n'est pas directement proportionnelle à l'entrée. Autrement dit, est non-linéaire quelque chose qui n'est pas linéaire (!).
- En mécanique des fluides, la **turbulence** est un mode d'écoulement dans lequel se superpose au mouvement moyen un mouvement d'agitation aléatoire. Dans un écoulement turbulent, la vitesse à un instant donné peut varier, de façon importante et imprévisible, pour des points rapprochés.

En d'autres termes, l'écoulement turbulent ne peut pas se caractériser comme la superposition de processus physiques d'échelles spatiales et temporelles différentes : leurs interactions participent de façon conséquente à la nature de l'écoulement et à la construction de sa moyenne.

Quantifier les non-linéarités Au vu de ces définitions, il paraît clair que la turbulence est reliée à la notion de non-linéarité. La circulation océanique méso-échelle, comportant une part turbulente importante, est donc le terrain de jeux de non-linéarités. S'il est difficile de caractériser précisément la notion de non-linéarité [7], en revanche, **quelques indicateurs**, chiffrés ou non, permettent d'en mesurer et d'en décrire les effets ainsi que leur intensité.

7. Ceci reviendrait à proposer une quantification tous les objets physiques et mathématiques qui ne sont pas linéaires. Vaste programme.

Mesures de la variabilité Une première façon de mesurer les effets de la non-linéarité et par là même de la quantifier est de mesurer la variabilité qu'elle engendre dans le système observé.

Dans le cas de la circulation méso-échelle, certains diagnostics physiques traduisent cette variabilité, comme l'énergie cinétique turbulente. Cette dernière réprésente l'énergie cinétique par unité de masse des fluctuations de vitesse dans un écoulement turbulent. Elle s'exprime en $m^2.s^{-2}$. Elle se calcule selon la formule :

$$EKE(x,y) = \frac{1}{2T}\int_{t=0}^{T}(u(x,y,t) - \bar{u}(x,y))^2 dt \tag{1.1}$$

où $\bar{u}(x,y) = \dfrac{1}{T}\displaystyle\int_{t=0}^{T} u(x,y,t)dt$.

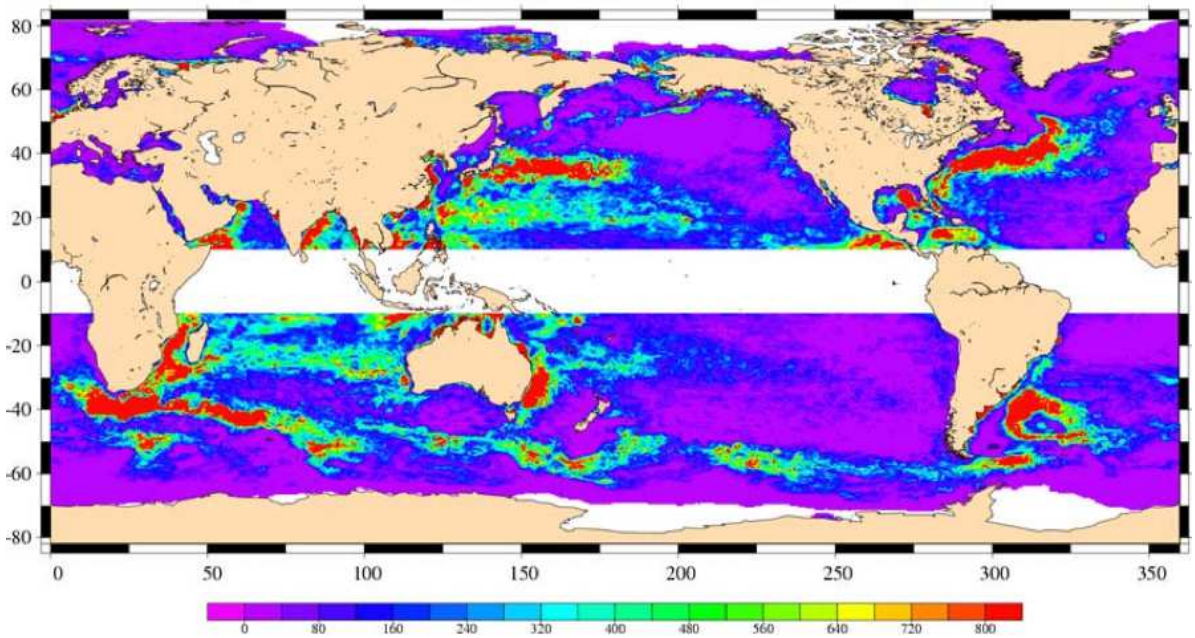

Figure 1.3 – *EKE calculée à partir de produits altimétriques issus des missions T/P, Envisat, Jason-1 et GFO. D'après Pascual et al. (2006)*

Pour un écoulement variable, il peut être intéressant de décomposer les principales structures spatiales de la variabilité naturelle de l'écoulement. Ceci peut être réalisé à l'aide d'une décomposition en composantes principales (appelées *EOFs* dans la suite [8]).

Les EOFs sont utilisées dans de nombreuses études en océanographie (voir par exemple Korres et al. (2000) ou Alvera-Azcárate et al. (2005)). En assimilation de données, elles ont été également utilisées, notamment pour l'algorithme SEEK (voir Pham et al. (1998a)) ou dans l'étude d'un algorithme de type 4DVAR de rang réduit (Robert et al. (2005b)).

Le calcul des EOFs consiste à diagonaliser une matrice de covariance construite à partir d'états instantanés du modèle. Les vecteurs propres traduisent les modes de variabilité du modèle. À titre d'illustration, considérons ici un calcul d'EOFs monovariées, ne prenant en compte

8. Selon le domaine scientifique, une telle décomposition admet de nombreux noms : Principal Orthogonal Decomposition (POD), Principal Component Analysis (PCA), Empirical Orthogonal Function (EOF). En océanographie, cette dernière dénomination est le plus souvent employée, c'est pourquoi nous la retenons ici.

qu'une variable d'état du modèle. Soient $\mathbf{h}(t_i, x, y),\ i \in [1, n]$, une série temporelle d'instantanés de la hauteur d'eau simulée. Si, pour chaque instant t_i, nous mettons le champ de hauteur d'eau sous forme de vecteur, duquel est retirée la moyenne temporelle, nous pouvons former une matrice $F \in \mathcal{M}_{n,m}$ où m est le nombre de degrés de liberté spatiaux pour cette variable d'état. Les vecteurs propres de la matrice de covariance $C = F.F^T$ sont ainsi les EOFs du modèle, sur la période considérée. Les valeurs propres λ_i nous renseignent sur la variabilité du modèle expliquée par les structures présentées dans chaque EOF_i. Ainsi, le pourcentage de variabilité du modèle expliquée par le vecteur propre i est donné par :

$$v_i = \frac{\lambda_i}{\sum_{j=0}^{n} \lambda_j}$$

Bien souvent, les EOFs seront classées par ordre décroissant de variabilité expliquée. Cependant, des classifications plus sophistiquées sont parfois mises en place (Preisendorfer, 1988).

Décorrélation temporelle Une autre caractéristique des écoulements non-linéaires est la décorrélation temporelle de l'état de l'écoulement. Autrement dit, si l'on prend un état $\mathbf{x}(t)$ caractérisant la circulation à un instant t, il existe un temps T_c tel que la corrélation statistique entre les états $\mathbf{x}(t)$ et $\mathbf{x}(t + T_c)$ n'est plus significative.

Un diagnostic de cette décorrélation temporelle peut être calculé comme suit. Soit $\mathbf{x}_m(t_i) = \frac{1}{|\Omega|} \int_\Omega \mathbf{x}(t_i, x, y, z) d\Omega$ la moyenne spatiale d'une variable d'état de notre modèle. Le facteur d'autocorrélation temporelle $\rho(T)$ de $\bar{\mathbf{x}}$ est défini par :

$$\rho(T) = \frac{\sum_{i=1}^{N} (\mathbf{x}_m(t_i) - \bar{\mathbf{x}}_m)(\mathbf{x}_m(t_i + T) - \bar{\mathbf{x}}_m)}{\sqrt{\sum_{i=1}^{N} (\mathbf{x}_m(t_i + T) - \bar{\mathbf{x}}_m)^2} \sqrt{\sum_{i=1}^{N} (\mathbf{x}_m(t_i) - \bar{\mathbf{x}}_m)^2}} \tag{1.2}$$

où $\bar{\mathbf{x}}_m$ est la moyenne temporelle de $\mathbf{x}_m$. Un exemple en est donné sur la Figure 1.4.

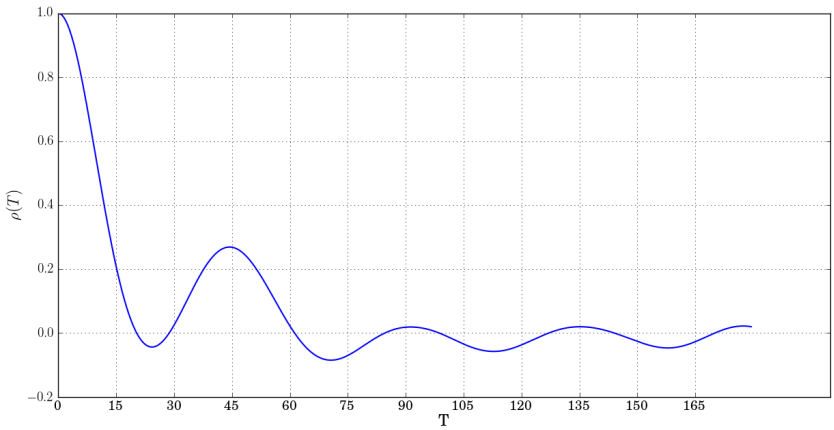

Figure 1.4 – *Exemple d'autocorrélation temporelle, calculée à partir de la SSH, pour un modèle donné au 1/4° (abscisse en jours)*

Évolution de perturbation Enfin, lorsque nous disposons d'un outil de modélisation, l'influence des termes non-linéaires sur l'état de l'écoulement peut être directement étudié en comparant la propagation d'une perturbation de l'état initial par le modèle non-linéaire, et d'autre part par ce même modèle linéarisé autour de l'état courant.

Soient $h(t)$ et $\tilde{h}(t)$ les trajectoires d'une même perturbation initiale respectivement propagée par le modèle non-linéaire et par le modèle linéarisé. On peut définir un *indice de similarité* comme suit :

$$S(t) = \frac{(h(t), \tilde{h}(t))}{\|h(t)\| \|\tilde{h}(t)\|} \tag{1.3}$$

On utilise ici le produit scalaire euclidien. Ce diagnostic peut être vu comme une mesure de « l'angle » à l'instant t entre les deux propagations. Un exemple en est donné sur la Figure 1.5.

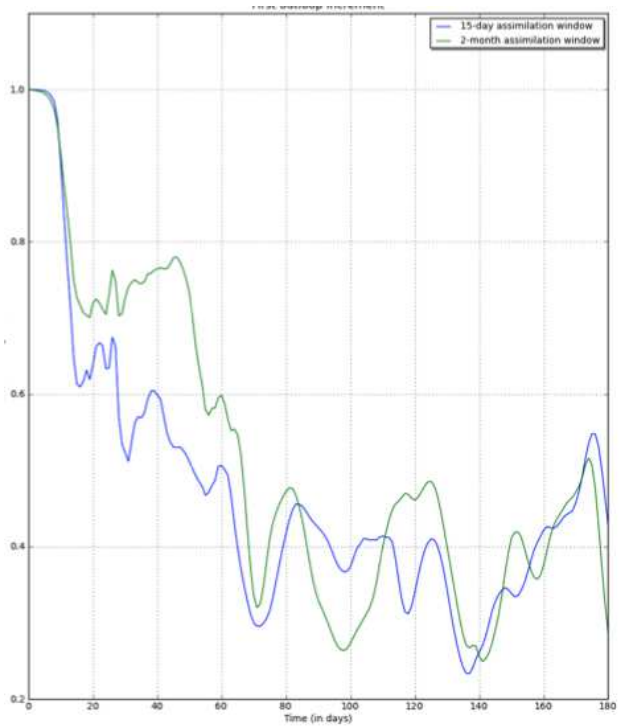

Figure 1.5 – *Indice de similarité en fonction du temps (en jours) pour deux perturbations d'amplitudes différentes (la plus élevée en bleu).*

L'évolution de l'océanographie physique mène donc à s'intéresser de plus en plus aux phénomènes méso- et subméso-échelles. Les deux moyens d'accès à la dynamique de l'océan que sont les observations et la modélisation en sont impactés directement. Nous allons maintenant donner un aperçu des enjeux qu'elles doivent affronter pour explorer les plus petites échelles dynamiques de l'océan.

1.2 Observation de la méso- et subméso-échelle

Si l'observation des océans est très ancienne, l'existence de réseaux synoptiques [9] permettant d'observer la variabilité de la circulation à l'échelle planétaire et permettant le développement de l'océanographie opérationnelle est beaucoup plus récente. Systématisées depuis une soixantaine d'années, les mesures *in-situ* ont été densifiées avec le déploiement du réseau ARGO, au début des années 2000. Depuis une trentaine d'années les observations satellitaires apportent une vision quasiment synoptique de la surface océanique. Ces dernières ont été cruciales pour accéder à la compréhension de la circulation méso-échelle et le restent encore. Après une courte présentation de réseaux d'observations *in situ*, nous nous attarderons sur les principes de l'altimétrie satellitaire et sur les implications pour les observations de l'intérêt croissant pour les petites échelles de la circulation océanique.

1.2.1 Les mesures *in-situ*

Les mesures *in-situ*, bien que peu nombreuses en regard du nombre de degrés de liberté du système océanique, sont un atout précieux pour les océanographes, car elle donne une information sur la dimension verticale de l'océan et, pour certains réseaux, de façon synoptique.

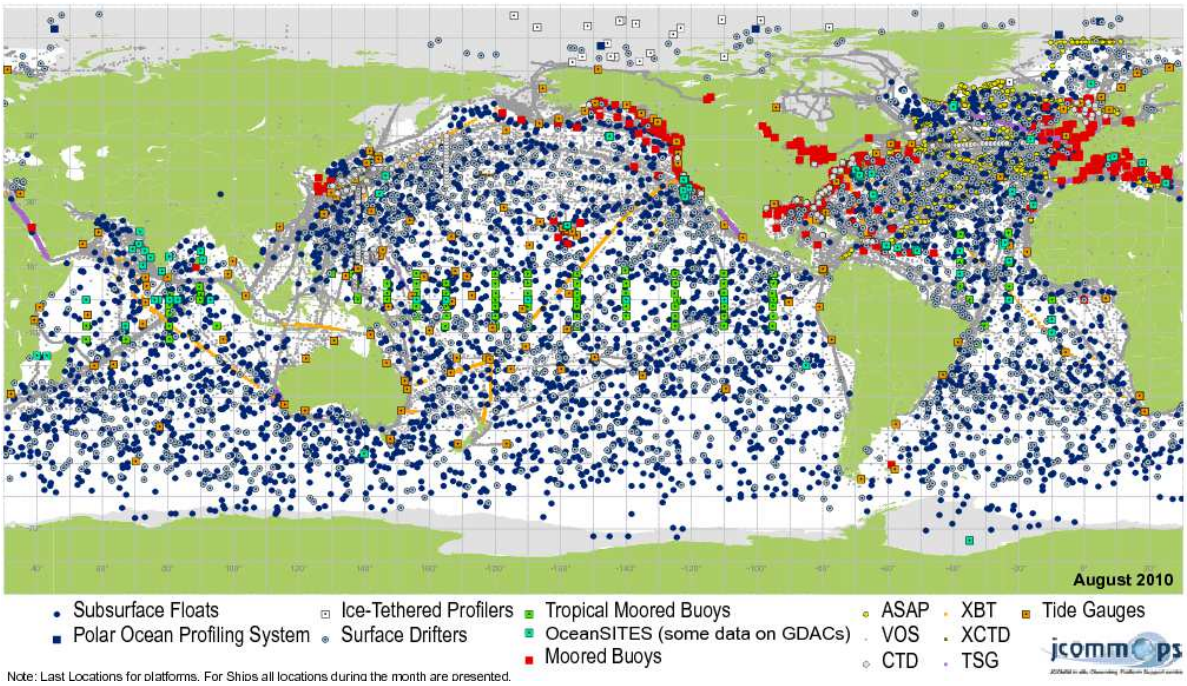

Figure 1.6 – *Carte des positions des différents appareils de mesure in-situ en activité en août 2010. Crédits : Jcommops*

Les mesures *in-situ* se sont véritablement systématisées à partir des années 1960 avec les mouillages, c'est à dire la mise en place de stations dans l'océan, relevant des mesures temporellement quasi-continues de salinité et de température. Dans les régions tropicales, le réseau de mouillages TAO/TRITON (Tropical Atmosphere Ocean) du programme international

9. Qui permet d'embrasser, de saisir d'un même coup d'œil les diverses parties d'un ensemble, qui en offre une vue générale. *Définition Wiktionnaire.*

TOGA (Tropical Ocean Global Atmosphere) est opérationnel depuis 1992 dans le Pacifique. Il a été étendu avec le réseau PIRATA dans l'Atlantique tropical, composé de 20 bouées, et plus récemment dans l'océan indien avec le réseau RAMA.

Les mesures XBT [10] sont issues d'instruments de mesure de la température tirés par des navires scientifiques ou d'opportunité (*e.g.* cargo). Bien qu'elles présentent d'importants biais thermiques (voir par exemple Gouretski et Reseghetti (2010)), elles sont aujourd'hui utiles pour reconstruire les tendances océaniques à moyen terme, étant une source majeure de données entre 1960 jusqu'à aujourd'hui.

Depuis l'an 2000, la création du réseau ARGO (profileurs dérivants) par la Commission Océanographique intergouvernementale de l'Unesco et de l'Organisation Météorologique Mondiale assure un réseau dense d'observations in situ de température et salinité jusqu'à 2000m de profondeur pour certains profileurs (PROVOR). Lorsque qu'ils sont à une profondeur appelée « profondeur de parking », les profileurs dérivent durant une période préprogrammée. Ils remontent ensuite vers la surface en effectuant des mesures de température et de salinité en fonction de la pression. A la surface les mesures ainsi que la position est transmise par le système ARGOS. Les trajectoires des appareils sont difficilement prévisibles, à cause de la nature turbulente de l'écoulement dans lequel ils sont plongés. Les mesures peuvent être plus fréquentes dans les régions de convergence.

1.2.2 L'altimétrie

Principes généraux Un satellite altimétrique mesure la distance qui le sépare de la surface océanique. Avec la connaissance de la position du satellite par rapport à une surface de référence (appelée ellipsoïde de référence, voir Figure 1.7) on déduit la hauteur de mer (SSH [11]) par rapport à cette ellipsoïde.

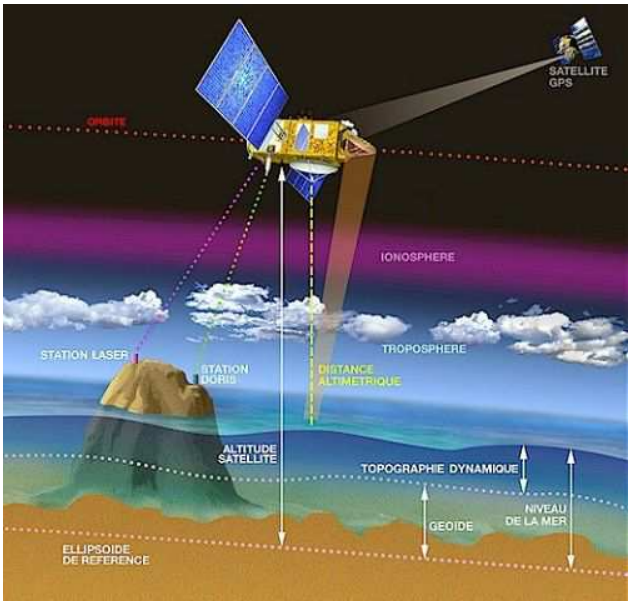

Figure 1.7 – *Décomposition de la SSH. Crédits : site web AVISO*

10. Expendable BathyThermograph
11. Sea Surface Height

D'un point de vue purement physique, la SSH, parfois notée η peut se décomposer en deux termes :

- La hauteur du geoïde par rapport à l'ellipsoïde, qui se définit sur toute la surface terrestre à une constante près, indépendamment du temps. Elle correspond à la topographie de surface qu'aurait l'océan si la quantité d'eau et la température de cette eau étaient constantes, et si l'océan n'était affecté par aucune perturbation (vent, marées, courants...). Cette topographie suit une surface d'isopotentiel qui n'a pas exactement la forme d'une ellipsoïde parce que la répartition des masses n'est pas homogène à l'intérieur de la terre. Elle varie aussi très légèrement dans le temps à cause de redistributions globales de masses, mais cette variation est négligeable à court terme.
- La hauteur appelée topographie dynamique, variable dans l'espace et dans le temps, qui est directement liée aux caractères de la circulation et aux variations des circulations océaniques.

Mais la connaissance du géoïde est plus imprécise, surtout à petite échelle, que la mesure de la distance à l'ellipsoïde de référence (de l'ordre de 2 cm). Il en résulte évidement une imprécision sur η. Les faibles variations temporelles du géoïde devant celles de la circulation océanique conduisent souvent à décomposer la SSH comme suit :

- La SSH moyenne sur une période de temps la plus longue possible, appelée MSSH
- L'anomalie par rapport à la MSSH, appelée SLA [12].

L'avantage de cette décomposition est que la SLA est mesurée avec la précision des satellites altimétriques. Mais les caractères permanents de la circulation, ou d'évolution temporelle plus longue que l'intervalle de temps sur lequel est calculée la MSSH, n'ont pas de signature sur cette variable. Autrement dit, la prise en compte de la SLA lors des processus d'assimilation de données est dépendante du calcul de la MSSH.

Missions passées, présentes et futures A partir de 1978, la brève (3 mois) mission satellite SEASAT marque le début de l'altimétrie satellitaire en démontrant la faisabilité des mesures d'une précision relative de 5 cm. Toutefois, l'erreur sur l'évolution de l'orbite étant importante, les mesures n'étaient utilisables que pour des régions de forte activité tourbillonnaire comme les courants de bord ouest (Fu, 1983). À partir de 1985, le satellite GEOSAT a fourni la première base de données altimétriques conséquente sur une durée de trois ans. Deux catégories de satellites sont ensuite apparues : les satellites de type ERS ou ENVISAT, dédiés à l'observation conjointe de la terre et des océans, et les satellites de type TOPEX/POSEIDON, GFO et JASON, plus spécifiquement conçus pour l'observation des océans. Les observations issues de ces difFérents programmes ont été largement utilisées dans la communauté océanographie, à la fois pour la recherche et les applications opérationnelles.

Cette année, le satellite SARAL a été lancé, sur la même orbite qu'ENVISAT pour assurer une certaine continuité des observations. La nouveauté de ce satellite est d'embarquer un altimètre en bande Ka large bande (35.75 GHz, 500MHz), premier altimètre océanographique utilisant une fréquence et une résolution *along-track* aussi élevée, caractéristiques intéressantes pour représenter la méso-échelle dans les observations. Cette haute fréquence permet d'accroître la précision des mesures dans les régions côtières. La Figure 1.8 illustre les mesures altimétriques de SARAL comparativement au satellite Jason-2. Nous pouvons voir que ce dernier à une fréquence spatiale d'échantillonnage plus faible que celle de SARAL qui, en contrepartie, a une période de

12. Sea Level Anomaly

répétitivité plus longue.

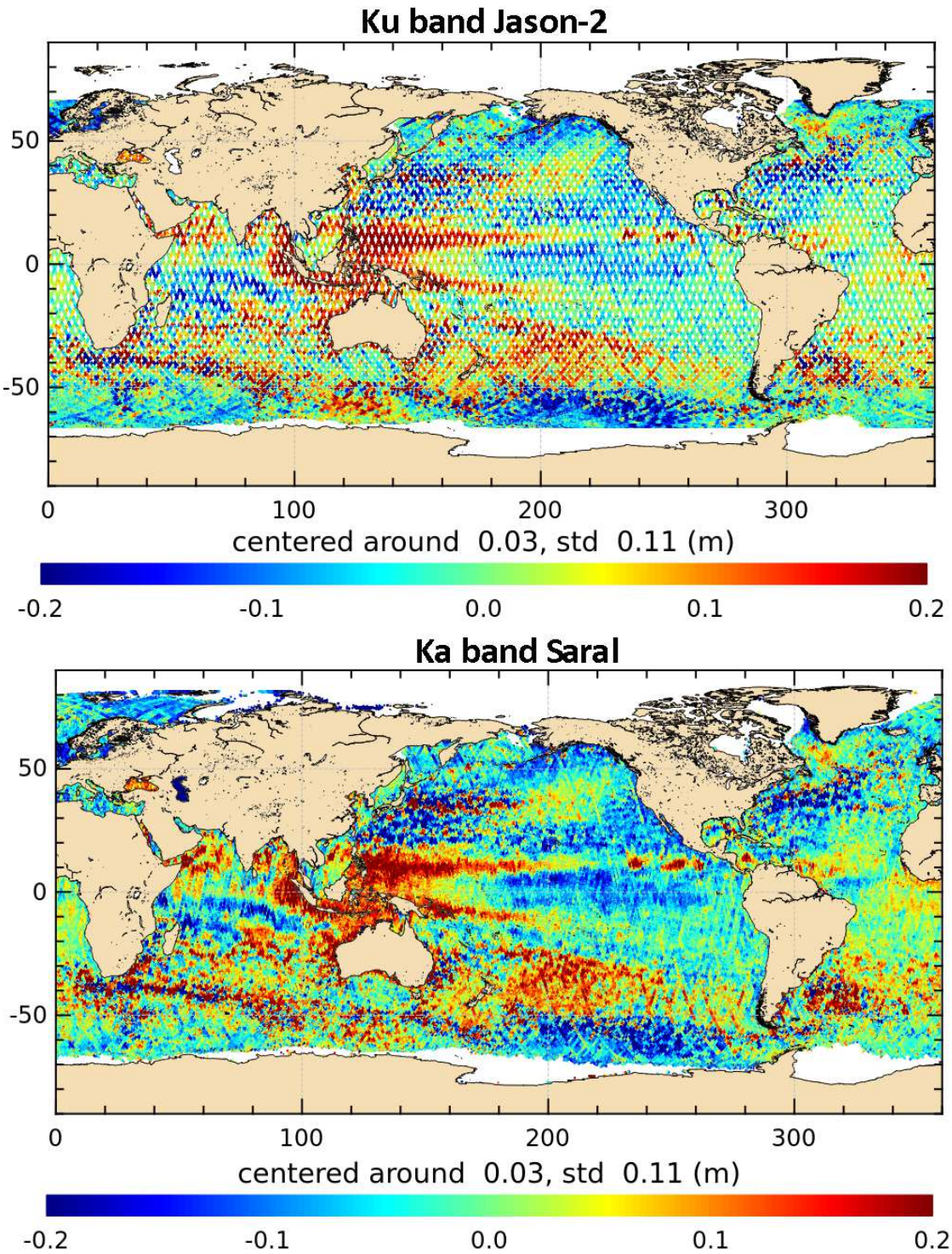

Figure 1.8 – *Produits altimétriques along-track issus de Jason-1 (en haut) et de SARAL (en bas), sur 35 jours. Crédits : AVISO/CNES.*

À l'horizon 2020, le programme satellitaire NASA/CNES SWOT projette le lancement d'un

altimètre dont les objectifs seraient les suivants [13] :

- Mesurer la SSH et les hauteurs d'eaux continentales sur une fauchée de 120 km de large avec une lacune de données au nadir [14] sur une tranchée de +/-10 km
- Sur l'océan, mesurer la SSH tous les 1 km le long de chaque fauchée, avec une précision de 2,5 cm.
- Sur les continents, obtenir une résolution spatiale de 100 m sur les rivières et de 100 km sur les lacs, les zones humides et les réservoirs. La précision à atteindre serait de 10 cm pour la hauteur d'eau de 1 cm/ 1 km pour la pente
- Couvrir au moins 90% des surfaces du globe

Cette résolution n'a jamais été atteinte et constitue une véritable révolution dans l'altimétrie satellitaire.

En aval des instruments de mesures, les données sont bien évidemment traitées, à plusieurs niveaux, pour transformer les observations brutes en produits exploitables par la communauté scientifique. Ces différents traitements que nous ne détaillerons pas ici requièrent un haut niveau d'ingénierie statistique, informatique et de traitement du signal.

La variété de la constellation de satellites altimétriques n'a cessé de croître au cours des deux dernières décennies, à la fois quantitativement (par le nombre de satellites) et qualitativement (orbites différentes et donc échantillonnages différents, précision accrue). Le Tableau 1.1 montre les caractéristiques des quelques acteurs majeurs de cette constellation.

Satellite	Période de production	Résolution spatiale	Résolution temporelle
T/P, Jason-1/2	1992-	2.8°	9.91j
GFO	1998-2008	1.5°	17.05j
ERS 1/2, Envisat, SARAL	1991-	0.72°	35j
SWOT	2019-	1.22°	20.87j

Table 1.1 – *Liste non-exhaustive de satellites altimétriques passés, présents et futurs. La résolution spatiale est l'intertrace en longitude des traces montantes. La résolution temporelle représente le temps d'un cycle.*

Les autres observations satellitaires Aujourd'hui, d'autres données issues de réseaux d'observations satellitaires sont disponibles :

- La température de surface (SST) depuis plusieurs décennies.
- La salinité de surface (SSS), via les programmes SMOS lancé en 2009 ou AQUARIUS lancé en 2011.
- La couleur de l'eau (SeaWIFS), mesurée à l'aide d'instruments embarqués sur ENVISAT, entre autres, qui est un indicateur de l'activité biogéochimique via la production de chlorophylle, elle-même révélatrice des conditions de surface de l'océan et d'éventuelles remontées d'eaux profondes.

Synthèse Comme nous venons brièvement de le voir, depuis l'avènement de l'altimétrie, les sources d'observations se sont multipliées, le système ARGO favorisant le développement des réseaux *in-situ* tandis que la constellation de satellites altimétriques n'a cessé de se diversifier et

13. Source : http ://www.aviso.oceanobs.com/fr/missions/missions-futures/swot.html
14. *i.e.* à la verticale du satellite

de s'améliorer, que ce soit en terme d'échantillonnage, de densité d'observations, de précision de mesure ou de résolution. Ceci a plusieurs conséquences directes pour l'assimilation de données :

- La densité des observations induit une représentation de dynamiques qui ne sont pas forcément encore modélisées, notamment aux petites échelles. Il faut pouvoir alors les filtrer sans dénaturer le signal afin de rendre compatibles les observations et les modèles. La densité d'observation a d'autres impacts plus spécifiques aux méthodes variationnelles d'AD qui seront explicitées ultérieurement.
- L'influence de l'échantillonnage des différents réseaux d'observations sur les méhodes d'AD est difficilement prévisible a priori et peut être spécifique à chaque méthode.
- Le volume de données croissant impacte les structures d'erreurs associées, quantitativement (la taille des matrices de covariances d'erreur d'observation croît également) et qualitativement (sur des données altimétriques à très haute résolution, l'hypothèse de décorrélation spatiale des erreurs d'observations n'est plus forcément raisonnable). Il faut également mettre en place des méthodes permettant d'exploiter au mieux toute l'information disponible.
- Enfin, l'altimétrie ne s'intéresse qu'à la surface de l'océan. Comment les méthodes d'AD peuvent-elles alors contrôler efficacement l'océan vertical ?

L'altimétrie, comme toutes les observations satellites, est donc cruciale pour l'océanographie physique et ses applications opérationnelles et posent de nombreuses questions pour l'assimilation de données. Dans ce travail, nous avons donc choisi d'aborder les problématiques spécifiques à cette source d'observation.

1.3 Conclusion

Ainsi, nous venons de voir que la compréhension de la circulation méso-échelle est un enjeu majeur pour l'océanographie physique et ses applications. Ce défi a des impacts importants sur les deux sources d'information sur l'océan que l'assimilation de données a pour devoir de conjuguer : les observations et les modèles

Les principaux impacts sur les premières sont les suivants :

- *L'augmentation du nombre de mesures*
  La tendance depuis l'avènement de l'altimétrie est donc d'augmenter la densité spatiale et temporelle des observations. Avec le projet SWOT et son altimètre à large fauchée, celle-ci se confirme clairement dans un futur proche. Cependant, ce phénomène ne concerne que la surface de l'océan. Sa verticale ne sera pas sensiblement plus observée, laissant la majeure partie de l'océan inconnue.
- *Le traitement des données*
  Le corollaire du point précédent est l'augmentation du volume de données exploitables. Au delà du stockage et de la transmission de ces données, il faut réfléchir aux méthodes optimales d'exploitation de ces informations, notamment en assimilation de données
- *Les corrélations d'erreurs*
  Les méthodes d'assimilation de données font régulièrement l'hypothèse que les erreurs d'observations ne sont pas corrélées dans l'espace, pour des raisons de simplifications théoriques et numériques des différents algorithmes utilisés. Cependant, il est connu que cette hypothèse est parfois par trop abusive. Avec l'augmentation de la densité spatiale

des observations, cette hypothèse peut poser problème et n'est pas applicable au futur de l'altimétrie avec le projet SWOT, mesurant la hauteur de l'eau sur de larges fauchées. Le problème peut également être soulevé pour la corrélation temporelle des erreurs d'observations, notamment altimétriques.

En ce qui concerne les modèles, la tendance est donc d'augmenter constamment leur résolution spatiale. Aux échelles visées (de l'ordre du premier rayon interne de déformation et en dessous), la nature de l'écoulement océanique est de plus en plus turbulente. Ceci se traduit directement par l'augmentation de la prépondérance des effets non-linéaires sur la circulation simulée, à travers, par exemple, une sensibilité croissante à la condition initiale et une décorrélation temporelle plus rapide de l'écoulement océanique. L'autre effet secondaire de l'augmentation de la résolution porte sur le coût numérique des simulations.

Chapitre 2

...face à l'assimilation variationnelle de données.

Dans ce chapitre, nous présenterons d'abord succinctement les principes généraux et théoriques de l'assimilation de données, d'un point de vue formel, en décrivant les différentes approches existantes et leurs problématiques communes. Dans un deuxième temps, nous nous centrerons sur les méthodes variationnelles, en détaillant leurs fondements théoriques ainsi que les différentes formulations existantes. Les problématiques spécifiques soulevées par l'application de ces méthodes à l'océanographie physique (et plus largement aux sciences géophysiques) seront ensuite exposées.

2.1 Objectifs et ingrédients de l'assimilation de données

La science de l'assimilation de données (AD) consiste à élaborer des méthodes permettant de **combiner** les différentes **informations** disponibles, qualitativement et quantitativement différentes, sur un système donné, afin **d'estimer** le plus précisément possible son **état**, passé, présent et/ou futur.

Dans cette section, nous introduirons les différents ingrédients nécessaires à l'AD et les notations associées.

2.1.1 Les particules élémentaires

L'assimilation de données conduit à manipuler de nombreux objets mathématiques. Comme il n'est pas évident de s'y retrouver au premier coup d'œil ou à l'aide de définitions ponctuelles, il devient essentiel de s'armer de notations consistantes et récurrentes, que nous présentons ici. La plupart de ces notations (indices, exposants, typographie) respectent les conventions détaillées dans Ide et al. (1997).

D'un point de vue formel, les variantes typographiques présentées dans le tableau 2.1 sont associées aux différents objets mathématiques utilisés dans la suite.

Typographie	Signification
x	Valeur scalaire
X	Opérateur non-linéaire
$\mathbf{x}$	Vecteur
$\mathbf{X}$	Opérateur linéaire ou matrice

Table 2.1 – *Significations formelles des typographies utilisées*

Les matériaux de base de l'assimilation de données sont donc les sources d'informations disponibles sur un système donné. Dans le cadre des sciences géophysiques en général et de l'océanographie en particulier, ces informations proviennent des modèles et des observations, tous deux entâchés d'erreurs.

Les modèles Les modèles de dynamique océanique permettent de simuler l'évolution de l'état de l'océan au cours du temps. Cet état est représenté par la distribution spatiale de la température, de la salinité, de la hauteur d'eau et des vitesse des courants océaniques. Classiquement en AD, ces variables sont regroupées dans un vecteur $\mathbf{x} \in R^n$, appelé **vecteur d'état** de dimension n. n est dépendant de la discrétisation du système d'équations du modèle. Dans la suite, $\mathbf{x}_i$ décrit l'état de l'océan à l'instant t_i.

Le modèle est un ensemble d'équations mathématiques décrivant les lois physiques régissant la dynamique océanique. Une fois discrétisées, ces équations permettent donc de représenter l'évolution temporelle du vecteur d'état. Ainsi, l'état prédit par le modèle, à l'instant t_{i+1}, sera :

$$M_{i,i+1}(\mathbf{x}_i) \tag{2.1}$$

où $M_{i,i+1}$ désigne l'opérateur d'évolution temporelle du système entre les instants t_i et t_{i+1}.

Que ce soit théorique ou numérique, un modèle comportera des erreurs issues des approximations réalisées ou des phénomènes non pris en compte dans le modèle. Cette erreur du modèle est symbolisée par une perturbation statistique η_i. Ainsi, l'état vrai à l'instant t_{i+1} sera :

$$\mathbf{x}_{i+1} = M_{i,i+1}(\mathbf{x_i}) + \boldsymbol{\eta}_i \tag{2.2}$$

$\bar{\boldsymbol{\eta}}_i = E[\boldsymbol{\eta}_i]$ est l'espérance de ce processus stochastique et sa matrice de covariances associée est notée $\mathbf{Q}_i$ telle que :

$$\mathbf{Q}_i = E[(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}_i)(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}}_i)^T] \tag{2.3}$$

Les erreurs $\boldsymbol{\eta}_i$ sont décorrélées dans le temps : les matrices $\mathbf{Q}_i$ sont indépendantes.

Dans notre travail, nous nous placerons dans un cadre idéalisé où le modèle est supposé parfait. Nous supposerons donc cette erreur nulle.

Les observations Comme nous l'avons évoqué au chapitre précédent, nous possédons un certain nombre d'observations de l'océan, de natures différentes. En assimilation de données, ces observations, pour une période temporelle donnée, sont regroupées dans le **vecteur d'observation** $\mathbf{y} \in R^p$, de dimension p. En océanographie, p est couramment inférieur d'au moins un ordre de grandeur à la dimension du vecteur d'état n.

État vrai, ébauche, analyse et prédiction Le but de l'AD est de reconstruire précisément l'état réel du système étudié, à l'aide de ces deux sources d'information, sur une période temporelle et sur un domaine géographique donnés. La méthodologie courante de l'AD dans les sciences géophysiques suit le schéma suivant :

- Le modèle est initialisé et paramétré *a priori*, *i.e.* sans tenir compte directement des observations, donnant une première approximation $\mathbf{x}^b$, couramment qualifiée **d'ébauche**, de l'état **vrai** recherché, $\mathbf{x}^t$.
- À l'aide des observations et des méthodes d'assimilation de données, l'initialisation (ou la paramétrisation) du modèle est alors affinée, de façon à ce que l'évolution de l'état nouvellement produit, noté $\mathbf{x}^a$ et couramment appelé état **analysé** soit le plus proche possible des observations.
- Une fois trouvée la condition initiale remplissant le critère ci-dessus, le modèle peut alors **prédire** l'évolution de l'état du système, c'est à dire simuler son évolution sur une période où aucune observation n'est disponible.

Ces différentes *catégories* d'état du système (ébauche, analyse ou prédiction) seront indiquées par des exposants. Le Tableau 2.2 résume les différents exposants ainsi que leurs significations.

Exposant	Signification
t	État vrai (*true*)
b	État du modèle avant assimilation ou ébauche (*background*)
a	État après AD pendant la période d'assimilation, ou état *analysé*
f	État après AD après la période d'assimilation, ou état prédit (*forecast*)

Table 2.2 – *Significations des différents exposants utilisés*

Vecteur de contrôle et ébauche Le vecteur de contrôle est l'ensemble des paramètres du modèle que l'on cherche à estimer par les méthodes d'AD. Il peut par exemple regrouper des termes de paramétrisation du modèle (conditions limites, viscosité, schémas numériques) et/ou la condition initiale $\mathbf{x}_0$. C'est cette dernière qui est classiquement considérée dans les applications opérationnelles de l'AD, les systèmes de prévision géophysiques y étant très sensibles.

L'ébauche $\mathbf{x}^b$, évoquée précédemment, est une estimation *a priori* du vecteur de contrôle. Ainsi, ici, il s'agira d'une estimation de l'état initial du modèle. Cette estimation introduit une erreur par rapport à l'état initial vrai, l'erreur d'ébauche $\boldsymbol{\epsilon}^b = \mathbf{x}^b - \mathbf{x}_0^t$. Elle est en général supposée de moyenne nulle et caractérisée par sa matrice de covariances $\mathbf{B}$:

$$\mathbf{B} = E[\boldsymbol{\epsilon}^b.\boldsymbol{\epsilon}^{b^T}] \tag{2.4}$$

L'opérateur d'observation Dans le schéma général d'AD présenté auparavant, il vient qu'il faut pouvoir comparer l'état issu du modèle et les observations. Ceci pose plusieurs problèmes :

- Comme nous l'avons signalé, $\mathbf{x}$ et $\mathbf{y}$ n'ont pas les mêmes dimensions n et p. Les observations ne sont pas nécéssairement prises aux pas de temps et sur la grille du modèle
- Les observations et les variables simulées sont parfois reliées de façon indirecte

Ainsi, il est nécessaire de définir un opérateur H qui permettra de passer de l'espace du modèle à l'espace des observations. Cet opérateur peut être non-linéaire et inclure une dépendance

temporelle. On le définit comme suit :

$$\mathbf{y}_i = H_i(\mathbf{x}_i^t) + \boldsymbol{\epsilon}_i \tag{2.5}$$

où $\mathbf{y}_i \in R^{p_i}$ sont les observations à l'instant t_i. De façon analogue à l'erreur modèle, ϵ_i représente l'erreur d'observation, caractérisée par sa moyenne $\bar{\epsilon}_{t_i} = E[\epsilon_{t_i}]$ qui, à l'instar de l'erreur d'ébauche, est généralement supposée nulle, et sa matrice de covariances $\mathbf{R}_i$:

$$\mathbf{R}_i = E[(\epsilon_i - \bar{\epsilon}_i)(\epsilon_i - \bar{\epsilon}_i)^T] \tag{2.6}$$

Ces erreurs sont dues d'une part à la précision des instruments de mesure et d'autre part à l'erreur de représentativité, c'est à dire la représentation des mesures dans l'espace du modèle (typiquement des mesures peuvent rendre compte de phénomènes non-modélisés).

Les erreurs d'analyse et de prévision Les états analysés $\mathbf{x}^a$ et prédits $\mathbf{x}^f$ présentent des erreurs par rapport à l'état vrai $\mathbf{x}^t$. Ces erreurs, notées respectivement ϵ^a et ϵ^f, comme celles évoquées jusqu'à maintenant peuvent être vues comme des processus stochastiques, caractérisés par leurs moyennes $\bar{\epsilon}^a$ et $\bar{\epsilon}^f$ et leurs matrices de covariances $\mathbf{P}^a$ et $\mathbf{P}^f$.

Hypothèses communes sur les erreurs

Après ce petit tour d'horizon des différents ingrédients de l'AD, il apparaît que le modèle et les observations ne sont pas les seuls protagonistes : les erreurs qui leur sont associées doivent être inévitablement prises en compte. Cependant, il est courant en assimilation de données de faire les deux hypothèses suivantes, partiellement évoquées auparavant :

- L'erreur modèle est considérée comme négligeable, *i.e.* le modèle est une approximation suffisamment précise de la réalité.
- Les erreurs d'observation sont supposées décorrélées en espace et non biaisées. Ceci se traduit par le fait que $\mathbf{R}_i$ est diagonale et $\bar{\epsilon}^o = 0$.

Ces deux hypothèses peuvent être bien entendues discutées. La validité de la première dépend bien évidemment de la précision du modèle, de façon absolue, mais aussi et surtout comparativement aux autres sources d'erreurs. Dans notre étude, basée sur des expériences académiques, où l'état vrai est connu puisque simulé et où nous réalisons des expériences jumelles, nous avons choisi de ne pas prescrire d'erreur modèle.

L'hypothèse sur les erreurs d'observation est elle aussi discutable : dans le cas de l'altimétrie, par exemple, l'instrument de mesure et la chaîne de traitement de l'information sont toujours les mêmes. Ainsi, rien ne garantit que les erreurs ne soient pas corrélées spatialement **et** temporellement. Ce problème peut potentiellement s'accentuer avec l'avènement des satellites à large fauchée, de type SWOT, fournissant des observations très denses en espace. Cependant, dans notre travail, là encore, les observations sont simulées et nous avons également choisi de ne pas prescrire une corrélation spatiale des erreurs d'observations.

2.1.2 Les différentes approches du problème

Comme nous l'avons évoqué, l'AD de données est donc un ensemble de méthodes qui conjuguent les informations données par le modèle et les observations afin d'estimer de façon optimale l'état vrai (passé, présent et/ou à venir) du système considéré. Trois grandes classes de méthodes coexistent aujourd'hui :

- Les *méthodes variationnelles*, basée sur la théorie du contrôle optimal, dont le principe général consiste à chercher une estimation du vecteur de contrôle qui minimisera une **fonction coût**, représentant l'écart entre l'état analysé et l'ébauche d'une part et celui entre l'état analysé et les observations d'autre part.
- Les *méthodes séquentielles stochastiques* basées sur la théorie de l'estimation statistique optimale. Le principe de ces méthodes est d'effectuer une correction de la trajectoire du modèle à chaque nouvelle donnée disponible en deux étapes, une dite de correction suivie d'une étape de prévision.
- Les méthodes de rappel aux observations, dont le principe est d'introduire dans les équations un terme supplémentaire tirant la solution vers les données.

Nous allons maintenant rentrer un peu plus dans le détail des trois approches et identifier leurs avantages et inconvénients respectifs. Il est à noter que dans le cas linéaire (*i.e.* où l'opérateur d'observation et le modèle sont des opérateurs linéaires) et gaussien, les deux premières approches sont optimales et fournissant un même état analysé à la fin de la période d'assimilation. Les méthodes variationnelles, qui sont au coeur de notre étude, seront détaillées plus avant dans la section suivante.

Les méthodes de rappel aux observations

L'algorithme du nudging (appelé aussi relaxation newtonienne) consiste à introduire dans le système d'équations dynamiques du modèle un terme de relaxation vers les observations, pondéré par une matrice de gain $\mathbf{K}$. Cette méthode, premièrement utilisée en météorologie, a été introduite avec succès en océanographie, notamment par Verron et Holland (1989), Verron (1992), et Blayo et al. (1994). Un inconvénient de cette méthode simple est qu'elle ne tient pas compte des erreurs respectives du modèle et des observations. Plus récemment, l'algorithme BFN (Back and Forth Nudging), introduit par Auroux et Blum (2005), permet d'identifier un état initial à partir d'observations réparties dans le temps. L'algorithme résout d'abord les équations avec le terme de nudging pour arriver à un état final. Le modèle est ensuite utilisé de façon rétrograde en temps, avec un terme de rappel opposé à celui du nudging direct. On obtient ainsi une réestimation de l'état initial. Après répétition de ces deux étapes, l'état initial est censé converger vers l'état recherché.

L'avantage principal de cette méthode est la simplicité de sa mise en oeuvre et son faible coût numérique, surtout en comparaison des autres méthodes d'AD. En revanche, dans le cadre d'un modèle océanique aux équations primitives, aucun résultat théorique n'assure la qualité de la trajectoire analysée ou prédite. Et certains problèmes numériques peuvent parfois faire obstacle à sa mise en oeuvre [1].

Les méthodes statistiques séquentielles

Le principe des méthodes séquentielles d'AD est de corriger l'état simulé par le modèle à chaque instant où une observation est disponible.

1. Le modèle direct présente souvent des termes de diffusion. Le modèle rétrograde présente alors des termes d'antidiffusion, termes qui peuvent le rendre instable numériquement.

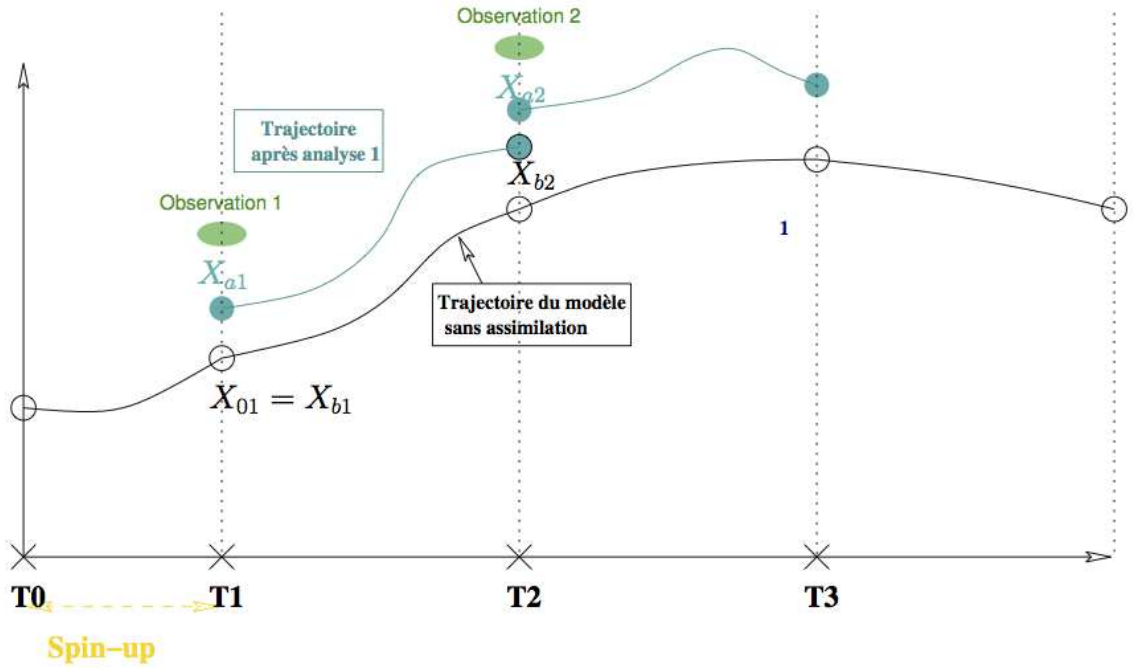

Figure 2.1 – *Illustration du principe des méthodes séquentielles.*

À partir d'un ensemble d'observations prises à un instant donné et des statistiques d'erreurs de prévision et d'observation, les méthodes stochastiques calculent un nouvel état considéré comme étant le plus probable. Dans le cadre de l'océanographie, il apparaît que l'erreur de prévision est corrélée spatialement, ce qui peut être intuitivement compris en regardant la variété du spectre des échelles spatiales et temporelles mises en jeu dans la circulation océanique.

Le filtre de Kalman Les méthodes statistiques séquentielles sont basées sur la théorie de l'estimation statistique optimale, particulièrement de l'algorithme du filtre de Kalman. Cet algorithme se décompose en deux étapes, la première, *l'étape d'analyse*, consistant à produire un état analysé tenant compte de la trajectoire simulée passée et des observations. Vient ensuite *l'étape de prévision.*

L'étape d'analyse se déroule comme suit. Connaissant les matrices des covariances d'erreur de prévision $\mathbf{P}^f$ et d'observation $\mathbf{R}$, sous l'hypothèse que les erreurs suivent une loi gaussienne, la meilleure estimation de l'état analysé se calcule suivant la formule suivante :

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \tag{2.7}$$

avec :

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} \tag{2.8}$$

La matrice $\mathbf{K}$ est appelée *le gain de Kalman.* La matrice des covariances d'erreur d'analyse se calcule comme suit :

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f \tag{2.9}$$

L'étape de prévision consiste alors à propager l'état analysé précédemment trouvé, pris à

l'instant k, par le modèle pour disposer d'un nouvel état simulé à l'instant $k+1$.

$$x_{k+1}^f = \mathbf{M}_{k,k+1} x_k^a \tag{2.10}$$

La matrice de covariances d'erreur de prévision se calcule alors :

$$\mathbf{P}_{k+1}^f = \mathbf{M}_{k,k+1} P_k^a \mathbf{M}_{k,k+1}^T + \mathbf{Q}_{k+1} \tag{2.11}$$

où $\mathbf{Q}_{k+1}$ est la matrice de covariances d'erreur modèle entre les instants k et $k+1$.

Dans **le cas linéaire** (H et M sont des matrices), comme l'algorithme est décrit ici, le filtre de Kalman est optimal et fournit la même solution à l'instant final que les méthodes variationnelles. En l'état, ce filtre peut être étendu aux cas non-linéaires (Jazwinski (1970)), en utilisant des approximations par des développements de Taylor autour de l'état courant. L'étape 2.7 devient :

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - H(\mathbf{x}^f) \tag{2.12}$$

et l'étape 2.10 devient :

$$x_{k+1}^f = M_{k,k+1}(x_k^a) \tag{2.13}$$

Cependant, plus l'ordre de l'approximation pris en compte est élevé, plus le coût d'application du filtre de Kalman « étendu » est élevé. Il faut noter que ces algorithmes supposent de disposer de bonnes estimations de $\mathbf{P}_o^f$, $\mathbf{Q}$ et $\mathbf{R}$.

Avantages et contraintes Les méthodes statistiques séquentielles présentent plusieurs **avantages** :

- **La simplicité d'implémentation** : comparativement aux méthodes variationnelles, comme nous le verrons par la suite, les méthodes basées sur le filtre de Kalman sont relativement simples à implémenter.
- Ces algorithmes permettent d'accéder à une estimation des **erreurs a posteriori**.

Le principal obstacle à l'application de ces méthodes est la dimension de l'espace d'état. Pour appliquer le filtre de Kalman, il faut en effet pouvoir calculer, stocker et manipuler les matrices de covariances d'erreur d'analyse et de prévision. Or, en océanographie, de telles matrices peuvent contenir rapidement 10^{12} éléments. Pour pallier cette contrainte, des variantes du filtre de Kalman ont été développées :

- Le filtre de Kalman d'Ensemble (Evensen (1994), Evensen (2003)). Cet algorithme basé sur une approche de type Monte-Carlo calcule les covariances d'erreur de prévision en intégrant en temps un ensemble de trajectoires dont les états initiaux et les forçages ont été perturbés aléatoirement. L'analyse des états perturbés est effectuée dans les filtres les plus utilisés, en utilisant des observations également perturbées.
- Les filtres racine carrée de rang réduit (Pham et al. (1998b) ; Brasseur (2006) ; Brasseur et Verron (2006) ; Brankart et al. (2010)) sont des algorithmes basés sur une décomposition en valeurs propres des matrices de covariances d'erreur. Les vecteurs propres associés aux plus grandes valeurs propres de la matrice de covariances d'erreur d'analyse sont utilisés pour former un ensemble de vecteurs d'état perturbés qui sont ensuite intégrés en temps par le modèle. Cet ensemble a la propriété que, dans le cas d'un modèle linéaire, les covariances dans le sous-espace engendré par les prévisions de l'ensemble sont exactes. Dans le cas d'un modèle non linéaire, la méthode est approchée et comparable au filtre de Kalman étendu.

Il faut remarquer que la non-linéarité et la grande dimension de l'espace d'état sont également des contraintes pour les méthodes variationnelles comme nous le verrons par la suite.

Un autre aspect intrinsèque de ces méthodes, qui peut être considéré comme un désavantage, est de proposer des trajectoires analysées discontinues en temps. Cependant, une stratégie telle que *l'Incremental Analysis Update* (IAU), décrite plus loin, peut pallier partiellement ce problème.

2.2 L'approche variationnelle

2.2.1 Un problème d'optimisation

Formulation du problème d'optimisation

Les méthodes variationnelles reposent sur la minimisation d'une fonction dite **coût**, notée J qu'il faut donc définir. Généralement, cette fonction coût se décompose en deux parties :

- La première traduit l'écart de l'état courant à l'ébauche, pondéré par la matrice de covariances d'erreur d'ébauche. Ce terme est noté J_b.
- La deuxième traduit l'écart de la trajectoire courante aux observations, pondéré par la matrice de covariances d'erreur d'observation. Ce terme est noté J_o.

On a donc $J = J_b + J_o$. **Une fois la fonction coût définie, le problème variationnel est complètement défini et sa solution existe**. Il faut « simplement » mettre en oeuvre des techniques pour la trouver.

Les matrices de covariances d'erreur peuvent se voir comme des indices de confiance associés à chaque source d'information. Nous évoquerons par la suite les problématiques qui s'y rattachent. Cette minimisation est basée sur des méthodes de descente, requérant des calculs de gradient (ou de variations) de la fonction coût J pour être mises en place, d'où le nom de *méthodes variationnelles*.

Techniques d'optimisation par méthode de descente

Principe général Comme nous venons de l'évoquer, la minimisation de la fonction coût J est réalisée à l'aide de méthodes de descente, algorithmes itératifs estimant l'état optimal $\mathbf{x}_0$ utilisant l'information portée par le gradient de J :

$$\mathbf{x}_{0,k+1} = \mathbf{x}_{0,k} + \alpha_k G_k(\nabla J(\mathbf{x}_{0,k})) \tag{2.14}$$

où $\mathbf{x}_{0,k}$ est le k-ième itéré, $G_k(\nabla J(\mathbf{x}_{0,k}))$ la direction de descente et α_k le pas de descente. Les algorithmes dérivant de cette classe de méthodes différent par le calcul de G_k et α_k. Nous pouvons citer notamment les méthodes de Quasi-Newton qui considèrent G_k comme une approximation de l'inverse de la Hessienne et calculent α_k le long de la direction de descente par une recherche linéaire.

Comme nous le détaillons par la suite, nous utilisons dans ce travail l'algorithme de 4DVAR incrémental. Dans ce contexte, nous minimisons une suite de fonctions coûts quadratiques. Pour cette classe de problèmes, les méthodes dites de *gradient conjugué* sont plus indiquées.

La méthode du gradient conjugué Considérons une fonction coût quadratique de la forme suivante :

$$J = \frac{1}{2} < \mathbf{x}_0, \mathbf{A}\mathbf{x}_0 > + < \mathbf{b}, \mathbf{x}_0 > \tag{2.15}$$

où $\mathbf{A}$ est une matrice définie positive et symétrique par rapport au produit scalaire $<,>$. La méthode du gradient conjugué consiste à rechercher les directions de descente $G_k(\nabla J(\mathbf{x}_{0,k}))$, notées $\mathbf{d}_k$ par la suite, menant à la solution exacte du problème d'optimisation. Chaque direction de descente doit être $\nabla^2 J$-orthogonale aux directions précédentes. Le pas de descente α_k est construit de telle façon que l'erreur réalisée à l'itéré k $(\hat{\mathbf{x}}_0 - \mathbf{x}_{0,k-1})$ soit orthogonale à la direction de descente $\mathbf{d}_k$. Pour les fonctions quadratiques définies positives, l'erreur initiale est une combinaison linéaire des directions de descente $\mathbf{d}_k$. L'algorithme élimine alors une composante de cette erreur initiale par itération, et si $\mathbf{x}_0 \in \mathbb{R}^n$, l'algorithme converge alors en n itérations.

Initialisation : $k = 0$, $\mathbf{d}_0 = \mathbf{r}_0 = G_k(\nabla J(\mathbf{x}_{0,k}))$;
while $\|r_k\| > \epsilon$ *or* $k < n$ **do**
  $\mathbf{q}_k = \mathbf{A}\mathbf{d}_k$;
  $\alpha_k = \dfrac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{d}_k^T \mathbf{d}_k}$;
  $\mathbf{x}_{0,k+1} = \mathbf{x}_{0,k} + \alpha_k \mathbf{d}_k$;
  $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{q}_k$;
  $\beta_{k+1} = \dfrac{\mathbf{r}_{k+1}^T \mathbf{r}_{k+1}}{\mathbf{r}_k^T \mathbf{r}_k}$;
  $\mathbf{d}_{k+1} = \mathbf{r}_k + \beta_{k+1} \mathbf{d}_k$;
  $k+ = 1$;
end

Algorithm 1: Algorithme du gradient conjugué

Il est à noter que pour les problèmes de grandes tailles, pour des raisons de coût numérique, seules quelques itérations sont possibles.

2.2.2 Zoologie des formulations variationelles

Les méthodes variationnelles regroupent de nombreuses variantes. Cependant, chacune de ces variantes est caractérisée par la définition bien précise d'une fonction coût (ou d'un ensemble de fonctions coûts). Nous introduisons ici les différentes variantes des méthodes variationnelles et leurs caractéristiques théoriques, numériques et pratiques.

Le 4DVAR

L'algorithme du 4DVAR consiste à minimiser la fonction coût suivante :

$$J(\mathbf{x}_0) = \underbrace{\frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b)}_{J_b} + \underbrace{\frac{1}{2} \sum_{i=0}^{N} (H_i(\mathbf{x}_i) - \mathbf{y}_i)^T \mathbf{R}_i^{-1} (H_i(\mathbf{x}_i) - \mathbf{y}_i)}_{J_o} \tag{2.16}$$

où $\mathbf{x}_0$ est la condition initiale que l'on cherche à contrôler, $\mathbf{x}_0^b$ l'ébauche, $\mathbf{y}_i$ le vecteur d'observations à l'instant t_i, $\mathbf{B}$ la matrice des covariances d'erreurs d'ébauche, $\mathbf{R}_i$ la matrice des covariances d'erreurs d'observation et H_i l'opérateur d'observation non-linéaire à l'instant t_i.

Le terme J_b traduit l'écart du vecteur de contrôle courant à l'ébauche choisie. Le terme J_o traduit l'écart de la trajectoire courante du modèle aux observations, dans l'espace d'observation et aux instants d'observations. Si l'on remplace le terme $\mathbf{x}_i$ par $M_{0 \to t_i}(\mathbf{x}_0)$, nous voyons que

cette fonction coût dépend de la condition initiale, notre vecteur de contrôle ici, transportée par le modèle.

Si H et M sont linéaires, J est quadratique et admet alors un seul minimum. En revanche, J n'est plus quadratique si M ou H sont non-linéaires. Il vient alors que J peut avoir plusieurs minima locaux et ne plus être convexe. Les algorithmes de minimisation par méthode de descente ne garantissent plus alors de trouver le minimum global, solution de notre problème.

Le système d'optimalité Minimiser J revient à résoudre $\nabla J_{\mathbf{x}_0}(\mathbf{x}_0^a) = 0$. Nous devons donc calculer le gradient de J pour le vecteur de contrôle courant. Pour les dimensions du problème en océanographie, ce calcul par accroissement fini est prohibitif. La méthode de l'adjoint proposée par Lions (1968) et appliquée pour la première fois en météorologie par Le Dimet et Talagrand (1986) permet le calcul de ∇J de façon beaucoup plus accessible.

Appliquons une perturbation $\delta\mathbf{x}_0$. On introduit la dérivée de Gâteaux $\hat{J}(\mathbf{x}_0, \delta\mathbf{x}_0)$ ou dérivée directionnelle, définie par :

$$\hat{J} = \lim_{\alpha \to 0} \frac{J(\mathbf{x}_0 + \alpha\mathbf{x}_0) - J(\mathbf{x}_0)}{\alpha} \tag{2.17}$$

Cette dérivée directionnelle est liée au gradient de J par la relation :

$$\hat{J}(\mathbf{x}_0, \delta\mathbf{x}_0) =< \nabla_{\mathbf{x}_0} J, \delta\mathbf{x}_0 > \tag{2.18}$$

Ainsi, nous voyons qu'obtenir le gradient de J peut revenir à calculer sa dérivée directionnelle pour une perturbation donnée, et à utiliser l'équation 2.18 pour identifier le gradient.

En posant :

$$\begin{cases} \hat{\mathbf{x}}_{i+1} = \dfrac{\partial M_{i,i+1}}{\partial \mathbf{x}} \hat{\mathbf{x}}_i = \mathbf{M}_{i,i+1} \hat{\mathbf{x}}_i \\ \hat{\mathbf{x}}_0 = \delta\mathbf{x}_0 \end{cases} \tag{2.19}$$

$\hat{J}$ s'écrit :

$$\hat{J}(\mathbf{x}_0, \delta\mathbf{x}_0) = \sum_{i=0}^{N-1} [\hat{\mathbf{x}}_i]^T \mathbf{H}_i^T \mathbf{R}_i^{-1} [H_i(x_i) - \mathbf{y}_i] + [\delta\mathbf{x}_0]^T \mathbf{B}^{-1} [\mathbf{x_0} - \mathbf{x}^b] \tag{2.20}$$

Soit une variable $\mathbf{p}$, vérifiant :

$$\sum_{i=0}^{N-1} [\hat{\mathbf{x}}_i]^T \mathbf{p}_i = \sum_{i=0}^{N-1} [\hat{\mathbf{x}}_i]^T \mathbf{p}_{i+1} + \delta\mathbf{x}_0^T \mathbf{p}_0 \tag{2.21}$$

Or

$$\sum_{i=0}^{N-1} [\hat{\mathbf{x}}_{i+1}]^T \mathbf{p}_{i+1} = \sum_{i=0}^{N-1} [\mathbf{M}_{i,i+1} \hat{\mathbf{x}}_i]^T \mathbf{p}_{i+1} = \sum_{i=0}^{N-1} [\hat{\mathbf{x}}_i]^T [\mathbf{M}_{i,i+1}]^T \mathbf{p}_{i+1} \tag{2.22}$$

$\mathbf{p}$ vérifie alors :

$$\sum_{i=0}^{N-1} [\hat{\mathbf{x}}_i]^T \mathbf{p}_i = \sum_{i=0}^{N-1} [\hat{\mathbf{x}}_i]^T [\mathbf{M}_{i,i+1}]^T \mathbf{p}_{i+1} \tag{2.23}$$

On définit p de façon à remplacer le premier terme de l'égalité 2.20 en fonction de $\delta\mathbf{x}_0$:

$$\begin{cases} \mathbf{p}_i = \mathbf{M}_{i,i+1}^T \mathbf{p}_{i+1} + \mathbf{H}_i^T \mathbf{R}_i^{-1} [H_i(\mathbf{x}_i) - \mathbf{y}_i] \\ \mathbf{p}_N = 0 \end{cases} \tag{2.24}$$

$\mathbf{M}^T$ et $\mathbf{H}^T$ sont le modèle et l'opérateur d'observation adjoints. $\mathbf{p}$ est la variable adjointe à $\mathbf{x}$. L'équation 2.23 devient :

$$\sum_{i=0}^{N-1} [\hat{\mathbf{x}}_i]^T \mathbf{H}_i^T \mathbf{R}_i^{-1} [H_{t_i}(\mathbf{x}_i) - \mathbf{y}_i] = \delta \mathbf{x}_0^T \mathbf{p}_0 \tag{2.25}$$

En réécrivant les équations 2.18 et 2.20, on obtient le gradient de J en $\mathbf{x}_0$:

$$\nabla_{x_0} J(\mathbf{x}_0) = \mathbf{p}_0 + \mathbf{B}^{-1}[\mathbf{x}_0 - \mathbf{x}^b] \tag{2.26}$$

Ainsi, accéder au gradient de J revient à résoudre le système d'Euler-Lagrange suivant :

$$\begin{cases} \mathbf{x}_{i+1} = M_{i,i+1}(\mathbf{x}_i) \\ \mathbf{x}_0 = \mathbf{x}_0^a \\ \mathbf{p}_i = \mathbf{M}_{i,i+1}^T \mathbf{p}_{i+1} + \mathbf{H}_i^T \mathbf{R}_i^{-1}[H_i(\mathbf{x_i}) - \mathbf{y}_i] \\ \mathbf{p}_N = 0 \\ \mathbf{x}_0^a = \mathbf{x}^b + \mathbf{B}\mathbf{p}_0 \end{cases} \tag{2.27}$$

Les deux premières égalités représentent le modèle d'origine, appelé parfois *modèle direct*. Vient ensuite le *modèle adjoint*. Comme les indices de la variable $\mathbf{p}$ le montrent, le modèle adjoint est intégré de façon rétrograde en temps.

Ainsi, le calcul du gradient de la fonction coût se fait via l'intégration rétrograde du modèle adjoint. Nous verrons par la suite comment celui-ci est obtenu en pratique.

En résumé La méthode du 4DVAR consiste à estimer itérativement l'état $\mathbf{x}_0^a$ minimisant J, définie par l'égalité 2.16, à l'aide de la valeur de son gradient ∇J, via des méthodes de descente :

$$\mathbf{x}_{0,k+1} = \mathbf{x}_{0,k} - \alpha_k G_k(\nabla J(\mathbf{x_{0,k}})) \tag{2.28}$$

où $\mathbf{x}_{0,k}$ est le k-ième itéré, $G_k(\nabla J(\mathbf{x_{0,k}}))$ la direction de descente et α_k le pas de descente. Idéalement, le critère d'arrêt est $\nabla J(\mathbf{x_{0,k}}) = 0$.

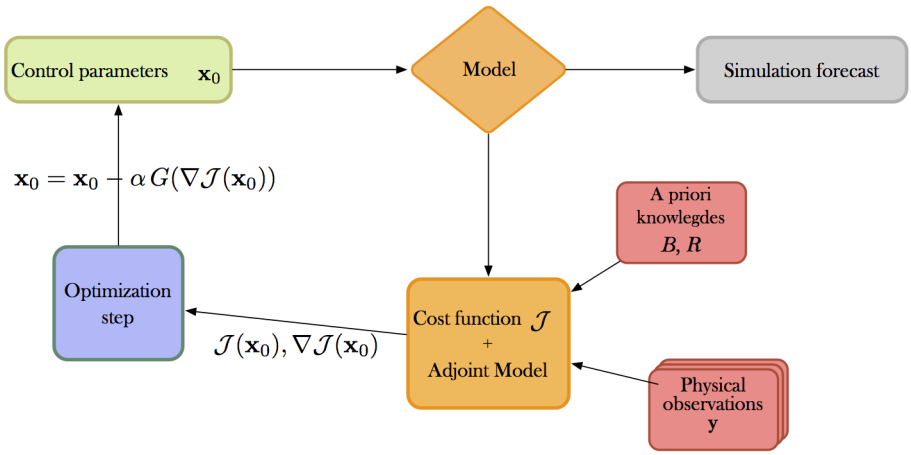

Figure 2.2 – *Schéma illustratif du procédé itératif de la méthode 4DVAR*

La Figure 2.2 illustre l'ordonnancement de la méthode 4DVAR.

Cependant, comme nous l'illustrerons par la suite, si H et/ou M sont non-linéaires, J n'est pas quadratique et peut présenter de nombreux minima locaux. La convergence de l'algorithme de minimisation par méthode de descente n'est alors plus garanti. C'est pour pallier ce problème que l'algorithme du 4DVAR incrémental a été développé.

Le 4DVAR incrémental

Le principe général de l'algorithme du 4DVAR incrémental consiste à approcher la fonction coût du 4DVAR par une série de fonctions coûts quadratiques définies comme suit :

$$J^{(k+1)}(\delta\mathbf{x}_0) = \underbrace{\frac{1}{2}\delta\mathbf{x}_0^T\mathbf{B}^{-1}\delta\mathbf{x}_0}_{J_b^{(k+1)}} + \underbrace{\frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i^{(k)}\delta\mathbf{x}_i - \mathbf{d}_i^{(k)})^T\mathbf{R}_i^{-1}(\mathbf{H}_i^{(k)}\delta\mathbf{x}_i - \mathbf{d}_i^{(k)})}_{J_o^{(k+1)}} \tag{2.29}$$

avec $\delta\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}_0^b$, $\mathbf{d}_i^{(k)} = \mathbf{y}_i - H_i(\mathbf{x}_i^{(k)})$. On a également $\delta\mathbf{x}_{i+1} = \mathbf{M}_{i,i+1}^{(k)}\delta\mathbf{x}_i$. $\mathbf{M}^{(k)}$ est le modèle linéaire tangent, linéarisé autour de la trajectoire issue de l'application de l'incrément. Cet incrément a été obtenu à la k-ième boucle externe et est alors ajouté à l'état initial d'ébauche.

Sur la figure 2.3, la courbe rouge illustre un exemple de fonction coût non-linéaire présentant plusieurs minima locaux. Le 4DVAR incrémental consiste à approcher la fonction coût non-linéaire par une fonction coût quadratique autour du vecteur de contrôle courant, puis de minimiser celle-ci. Ce processus est appelé une **boucle externe**. Chaque itération de minimisation au sein d'une boucle externe est appelé **une itération interne**.

Au terme d'une boucle externe, un nouveau vecteur de contrôle est ainsi obtenu, solution du problème de minimisation quadratique précédent. À partir de celui-ci, un nouvel écart aux observations est calculé à l'aide du modèle non-linéaire et de l'opérateur d'observation non-linéaire. Une nouvelle fonction coût quadratique est alors construite, donnant lieu à une nouvelle boucle externe.

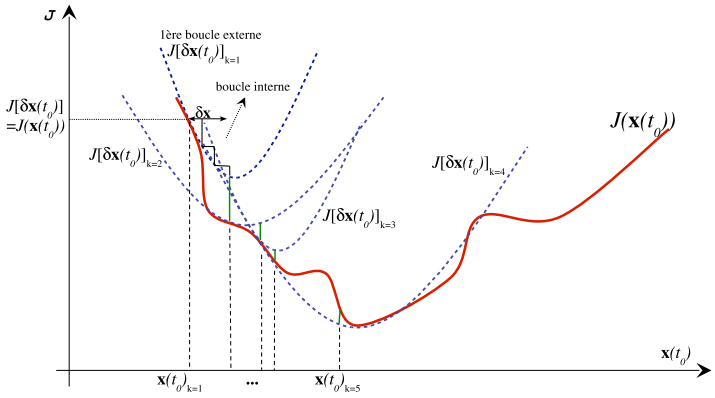

Figure 2.3 – *Illustration du déroulement de l'algorithme du 4DVAR incrémental par approximation par une suite de fonctions coût quadratiques (en pointillés bleus) de la fonction coût non-linéaire (en rouge)*

Pour résumer, de l'équation (2.29), nous pouvons noter plusieurs points a priori, concernant la transition d'une boucle externe k à la suivante $k+1$:

- Au début de chaque cycle, on a $\mathbf{x}_0 = \mathbf{x}_0^b$ et donc $\delta\mathbf{x}_0 = 0$. Le terme J_b est donc nul.
- Le terme d'innovation $\mathbf{d}^{(k+1)}$ ainsi que le modèle linéaire tangent sont mis à jour en fonction de l'incrément à l'étape k trouvé précédemment.

Théoriquement, la convergence de l'algorithme vers l'optimum global du problème non-linéaire n'est pas garanti.

Le 3DFGAT

L'algorithme du 3D-FGAT (*First Guess at Appropriate Time*) est une approximation de l'algorithme du 4DVAR incrémental : l'évolution de l'incrément durant la fenêtre d'assimilation est supposée stationnaire. Autrement dit, les modèles linéaire tangent et adjoint sont remplacés par l'identité :

$$J^{(k+1)}(\delta\mathbf{x}_0) = \underbrace{\frac{1}{2}\delta\mathbf{x_0}^T\mathbf{B}^{-1}\delta\mathbf{x}_0}_{J_b^{(k+1)}} + \underbrace{\frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i^{(k)}\delta\mathbf{x}_0 - \mathbf{d}_i^{(k)})^T\mathbf{R}_i^{-1}(\mathbf{H}_i^{(k)}\delta\mathbf{x}_0 + \mathbf{d}_i^{(k)})}_{J_o^{(k+1)}} \tag{2.30}$$

Ainsi, le terme d'innovation est bien calculé au temps d'observation. Cet algorithme est ainsi très proche du 4DVAR incrémental mais ne nécessite pas d'intégration des modèles linéaire tangent et adjoint. Ceci implique deux avantages techniques :

- Si un système d'AD basé sur le 4DVAR incrémental existe, l'implémentation de l'algorithme 3D-FGAT est triviale
- Le coût de calcul du gradient de la fonction coût est drastiquement diminué

Le gain de temps de calcul explique le choix de cet algorithme dans les systèmes d'AD basés sur les méthodes variationnelles aujourd'hui utilisées dans certains centres opérationnels en océanographie (UK-MetOffice, ECMWF).

Cependant, l'hypothèse de stationnarité de l'évolution de l'incrément est forte, et ne permet pas d'appliquer cette méthode sur de longues fenêtres d'assimilation (typiquement de l'ordre de quelques jours, en océanographie), période qui se réduit à mesure que la résolution spatiale, et donc la chaoticité du modèle, augmente.

Incremental Analysis Update Pour un modèle non-linéaire, l'incrément obtenu à l'issue d'une minimisation d'une fonction coût de type 3D-FGAT provoque des instabilités numériques et physiques dans la trajectoire analysée. En effet, celui-ci peut présenter une trop forte amplitude ou des structures spatiales respectant peu la physique du modèle. Pour pallier ce problème, la méthode IAU (*Incremental Analysis Update*) est couramment appliquée. Le principe est d'ajouter progressivement l'incrément $\delta\mathbf{x}_0^a$ au cours de l'étape d'analyse :

$$\mathbf{x}_{k+1} = M_{k,k+1}(\mathbf{x}_k + f(k)\delta\mathbf{x}_0^a) \tag{2.31}$$

où $k \in \{1..N\}$ et $f(k)$ une fonction de pondération de l'application de l'incrément, qui se définit par exemple comme suit :

- $f(k) = \delta_{N/2}$, où δ représente ici la fonction de Dirac. L'incrément est appliqué au milieu de la fenêtre temporelle et non au début.
- $f(k) = \dfrac{1}{N}$. L'incrément est ajouté petit à petit au cours de la fenêtre temporelle. Nous appellerons ceci l'IAU uniforme.

Il est à noter que la correction appliquée ne correspond effectivement pas à la correction calculée. En effet, rien ne garantit que :

$$\mathbf{x}_{k+1}^f - \mathbf{x}_k^a = \delta \mathbf{x}_0^a \tag{2.32}$$

2.2.3 Caractéristiques des méthodes variationnelles

Les méthodes variationnelles dérivant du 4DVAR prennent ainsi en compte l'ensemble des informations disponibles sur le système considéré pour produire une estimation, optimale au sens statistique du terme dans le cas linéaire, du vecteur de contrôle recherché :

- Les écarts entre le modèle et les observations aux temps d'observations
- L'écart à l'ébauche
- Les statistiques d'erreur associées

Le grand avantage est donc de produire cette estimation pour une fenêtre temporelle donnée, et de produire ainsi une trajectoire d'analyse continue à l'intérieur de cette fenêtre. Cependant, contrairement aux méthodes stochastiques séquentielles, les méthodes variationnelles ne fournissent pas, en première approximation, des diagnostics statistiques *a posteriori*.

2.3 Les grands défis posés par l'océanographie

Nous allons maintenant voir les problématiques de l'AD spécifiques à leur application à l'océanographie physique et les défis qui restent à relever. Même si l'utilisation de méthodes d'AD de type 4DVAR reste encore minoritaire dans les centres d'océanographie opérationnelle, leur étude en océanographie n'est pas récente ni confidentielle.

2.3.1 La problématique de la dimension

Comme nous l'avons expliqué au chapitre précédent, les systèmes géophysiques s'étendent sur une large gamme d'échelles spatiales et temporelles. Ceci implique directement des modèles numériques de grande taille ou dimension. Ceci a des répercussions directes sur les méthodes d'AD.

La représentation des matrices de covariances d'erreur

Certains objets mathématiques ne sont pas représentables et/ou manipulables explicitement pour de telles dimensions, comme les matrices de covariances d'erreurs, par exemple. La dimension de l'espace d'état n, dans les modèles de circulation océanique, se définit comme le nombre de points de grille de discrétisation sur lesquels sont calculés chaque variable pronostique du modèle (température, salinité, vitesse, hauteur de la surface de l'eau). Il est courant que n dépasse alors 10^6. Comme nous l'avons vu plus haut, les différentes méthodes d'AD nécessite de prendre en compte les erreurs (d'ébauche, d'analyse, de prévision et d'observations) notamment à travers les matrices de covariances associées. La taille de ces matrices est alors supérieure à $10^6 \times 10^6$. Ces matrices doivent être stockées, manipulées et inversées. Pour des dimensions de cette ordre, ces actions ne peuvent techniquement être réalisées de façon explicite.

Plusieurs méthodes existent pour représenter autrement les matrices de covariances d'erreur. Une première stratégie est de réduire la taille de ces matrices en dégradant le moins possible l'in-

formation qu'elles contiennent. Les matrices de covariances d'erreur sont symétriques définies positives par construction. Obtenir une telle représentation est possible par une décomposition en composantes principales des champs du modèle, comme cela a été réalisé par Robert et al. (2005a).

Une autre stratégie consiste à considérer que l'algorithme d'AD n'a pas besoin de stocker explicitement la matrice $\mathbf{B}$ de rang plein, mais de pouvoir réaliser une série de produits matrice/vecteur. Cette méthode, utilisée dans le cadre de ce travail, sera détaillée plus loin.

Une revue plus complète des autres méthodes d'estimation des statistiques de la matrice de covariances d'erreur d'ébauche est faite dans Mirouze (2010).

2.3.2 Réseau d'observation et conditionnement de la fonction coût

Un autre phénomène propre aux méthodes variationnelles est la sensibilité du conditionnement de la fonction coût J à la configuration du réseau d'observation. Évoqué dans Courtier et al. (1994), ce phénomène peut se décrire en s'intéressant au préconditionnement idéal de J. Il peut être assuré par sa hessienne :

$$J'' = J''_b + J''_o = \mathbf{I} + \mathbf{L}^T \left(\sum_{i=0}^{N} \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i \right) \mathbf{L} \tag{2.33}$$

où $\mathbf{L}$ est un éventuel opérateur de changement variable, $\chi = \mathbf{L}(\mathbf{x} - \mathbf{x}_b)$, χ étant la variable de contrôle de la fonction coût. Ici, $\nabla^2 J$ est déjà préconditionné par $\mathbf{B} = \mathbf{L}\mathbf{L}^T$. Par construction, J''_o est semi-définie positive, d'où $\lambda(J''_o) \geqslant 0$, λ représentant le spectre. On sait également que $\lambda(J'') = 1 + \lambda(J''_o)$. Soit n la dimension de l'espace d'état et p la dimension de l'espace d'observations. Si $p < n$, $\lambda_{min}(J''_o) = 0 \;\Rightarrow\; \lambda_{min}(J'') = 1$, d'où $\kappa(J'') = \lambda_{max}(J'')$, κ représentant le conditionnement.

À titre d'illustration, considérons deux points de grilles, sur lesquels nous disposons de p observations décorrélées sur chaque, avec une erreur d'observation associée σ_o ($\mathbf{R}^{-1} = \mathbf{I}/\sigma_o^2$). Dans ce cas, $\mathbf{H} \in M_{2p\times 2}$, avec p lignes $(0,1)$ puis p lignes $(1,0)$ et $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} = n\mathbf{I}/\sigma_o^2$.

L'erreur d'ébauche est σ_b en chaque point, de corrélation α entre les deux points. Après calcul de $\mathbf{L}$, comme étant la racine carrée symétrique de $\mathbf{B}$, on a :

$$\kappa(J'') = \frac{\sigma_b^2(1+\alpha) + \sigma_o^2/p}{\sigma_b^2(1-\alpha) + \sigma_o^2/p} \tag{2.34}$$

Si ces deux points de grille sont proches, on a $\alpha \approx 1$ et $\kappa(J'') = 2p\dfrac{\sigma_b^2}{\sigma_o^2} + 1$. Ainsi, le conditionnement peut augmenter suivant :

- L'augmentation du nombre p d'observations en chaque point de grille
- L'augmentation du ratio des variances d'erreurs $\dfrac{\sigma_b}{\sigma_o}$

2.3.3 L'assimilation de données face à la non-linéarité

Pour les méthodes variationnelles, si le modèle M et l'opérateur d'observation H sont des opérateurs linéaires, le caractère quadratique de la fonction coût est assuré. Ceci implique sa convexité et l'unicité d'un minimum global. Dans le cas non-linéaire, l'unicité d'un minimum n'est plus garantie, comme nous allons l'illustrer ci-après.

Modèle de Lorenz et minima locaux

Décrit dans Miller et al. (1994), dans le cadre du contrôle optimal d'un modèle dynamique non-linéaire, l'augmentation du temps d'intégration du modèle à contrôler augmente *de facto* le nombre de minima locaux de la fonction coût associée. Nous reprenons ici l'exemple du modèle de Lorenz, décrit par le système d'équation 2.35.

$$\begin{cases} \dfrac{dX}{dt} = -\sigma(X - Y) \\ \dfrac{dY}{dt} = rX - Y - XZ \\ \dfrac{dZ}{dt} = XY - bX \end{cases} \tag{2.35}$$

Discrétisé, le système 2.35 devient :

$$\begin{bmatrix} X_{t+1} \\ Y_{t+1} \\ Z_{t+1} \end{bmatrix} = \begin{bmatrix} 1 - \sigma\tau & \sigma\tau & 0 \\ r\tau & 1 - \tau & -X_t\tau \\ Y_t\tau & 0 & 1 - b\tau \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \\ Z_t \end{bmatrix} \tag{2.36}$$

où τ est le pas de temps d'intégration du modèle, ici, $\tau = 0.01$. Nous prenons ici les paramètres classiques du modèle de Lorenz pour obtenir un système chaotique ($\sigma = 10$, $r = 28$, $b = 8/3$), et les conditions initiales $X_0 = Y_0 = Z_0 = 1$ qui caractérisent l'état vrai. L'opérateur d'observation tient uniquement compte de la composante X du modèle 2.35 de la forme $\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$. Nous définissons la fonction coût des écarts quadratiques moyens suivante :

$$J(\mathbf{x}_0) = \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i)^T(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i) \tag{2.37}$$

où $\mathbf{y}_i$ sont les observations (*i.e.* la composante X) tirées de l'état vrai, $\mathbf{x}_i = [X_i\ Y_i\ Z_i]^T$ et N la période d'intégration du modèle (qui par analogie, serait la période d'assimilation du modèle). En s'intéressant à l'évolution de cette fonction coût en fonction de l'erreur sur la condition initiale Y_0 de la composante Y, nous obtenons la figure 2.4

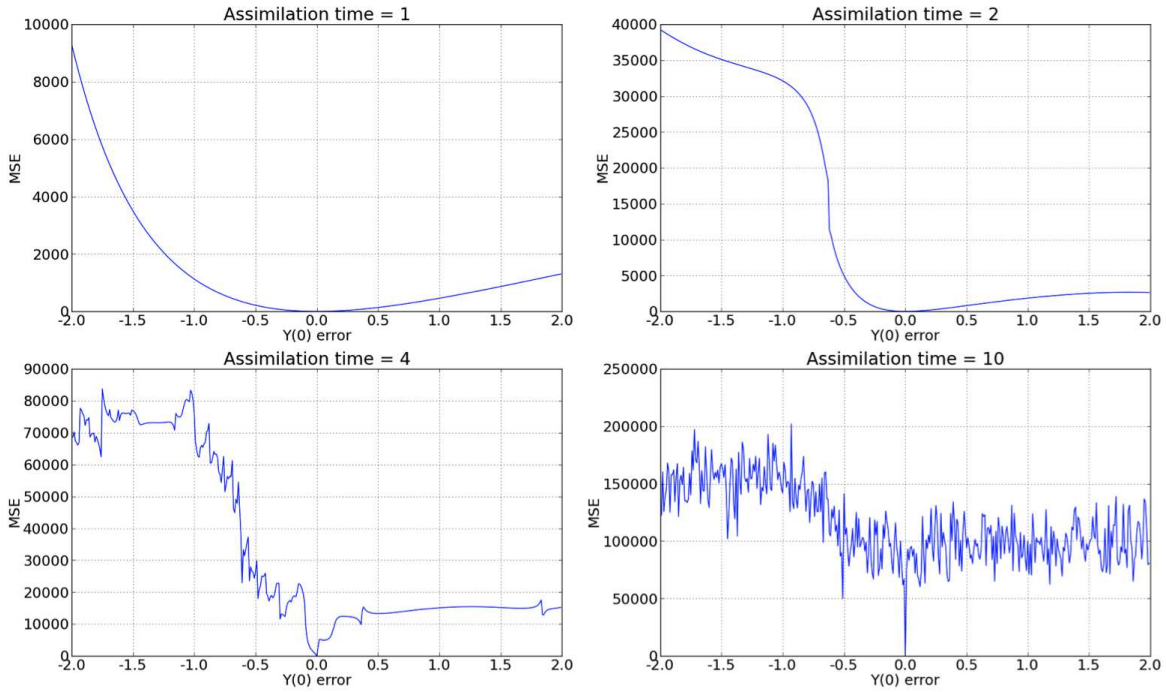

Figure 2.4 – *Exemples de fonctions coût, basées sur le modèle de Lorenz à 3 variables, pour différentes longueurs de fenêtres temporelles. Plus la fenêtre d'agrandit, plus le nombre de minima locaux croît, rendant la minimisation de plus en plus ardue.*

Il apparaît qu'en augmentant la période d'intégration du modèle, cette dernière devient de plus en plus irrégulière (l'unicité du minimum étant déjà perdue pour une fenêtre d'assimilation de longueur 4), jusqu'à présenter de nombreux minima locaux. Les algorithmes de descente deviennent alors lents dans de tels cas irréguliers, voire inefficaces dans les cas extrêmes. Cet exemple académique illustre l'apparition de nombreux minima locaux avec l'augmentation de la période d'assimilation, pour un modèle fortement non-linéaire.

Impacts potentiels des non-linéarités sur le système d'AD variationnel

En océanographie, H et M sont le plus fréquemment non-linéaires. Du point de vue du modèle continu, le principal terme non-linéaire est le terme d'advection, de la forme $U.\nabla U$, U étant la vitesse du courant. Une fois le modèle discrétisé, d'autres processus paramétrisés peuvent être non-linéaires, voire discontinus, comme par exemple la paramétrisation du mélange vertical ou la diffusion le long des isopycnes.

L'algorithme du 4DVAR, dans le cas linéaire, donne à convergence une estimation statistiquement optimale de la solution du problème d'optimisation posé (de la même manière que le filtre de Kalman). Dans le cas non-linéaire, plusieurs problèmes se posent :

- Comme nous nous l'avons vu ci-dessus, rien ne garantit l'unicité d'un extremum global de la fonction coût. De la même manière, rien ne garantit la convergence vers cet extremum s'il existe.
- Dans le cadre du 4DVAR incrémental, le modèle linéaire tangent peut comporter des approximations, le modèle direct n'étant pas forcément complètement différentiable. Cette approximation peut dans certains cas se révéler grossière.

Ces problèmes, tous dûs aux termes non-linéaires du modèle numérique et de l'opérateur d'observation, peuvent amener :

- **La non-convergence de l'algorithme**, pour une période temporelle donnée, à coût de calcul fini. En effet, pour des fonctions coût telles que celle présentée en bas à droite de la Figure 2.4, pour un coût de calcul limité, rien ne garantit la convergence d'un algorithme de descente, même vers un minimum local.
- **La convergence de l'algorithme vers un état sous-optimal.** Soit l'algorithme converge vers un minimum local, soit la limitation du coût de calcul le stoppe avant d'atteindre la convergence numérique. Dans les deux cas, l'état analysé sera sous-optimal.

Ces conséquences seront d'autant plus prégnantes que l'importance des phénomènes non-linéaires du modèle augmentera, notamment avec l'accroissement de la résolution.

2.4 Conclusion

Nous avons présenté ici succinctement les méthodes d'AD utilisées en océanographie en nous centrant sur les méthodes variationnelles. Plusieurs problèmes engendrés par l'augmentation de la résolution des modèles de circulation océanique apparaissent :

- **L'augmentation de la dimension de l'espace modèle**, rendant techniquement impossible la manipulation explicite des objets mathématiques nécessaires à l'AD (*e.g.* matrices de covariances d'erreur), problèmatique commune à toutes les méthodes d'AD. De plus, les algorithmes de minimisation pour des problèmes de grandes tailles sont restreints et présentent un important coût numérique.
- **L'augmentation de la densité spatiale des réseaux d'observation**, qui a deux principaux effets :
  - Dégradation potentielle du conditionnement de la fonction coût
  - Mise à mal de l'approximation de décorrélation spatiale des erreurs d'observation
- **La prépondérance des non-linéarités du modèle direct**, qui, pour les méthodes variationnelles, se traduit essentiellement par l'apparition de difficultés potentielles de convergence numérique vers une solution optimale.

Ainsi, à très haute résolution, l'application du 4DVAR incrémental peut potentiellement rencontrer des difficultés. Notre objectif est ici de les identifier, comprendre leurs origines et ouvrir des pistes de travail pour les dépasser.

Deuxième partie

Implémentation d'un système d'assimilation variationnelle de données pour une configuration idéalisée d'un bassin océanique

Chapitre 3

Description du modèle de circulation océanique NEMO et du système d'assimilation variationnelle de données associé NEMOVAR

Dans ce chapitre, nous présentons le modèle numérique de circulation océanique NEMO dans un premier temps, et le système d'AD variationnelle NEMOVAR ensuite. Ces outils sont ceux utilisés dans le cadre de cette étude. Ils sont aussi au cœur d'une partie de mes missions en tant qu'ingénieur de recherche NEMO-ASSIM, consistant à développer ou coordonner les développements, maintenir et documenter les composantes utiles à l'AD au sein de NEMO.

3.1 Présentation de NEMO, framework de modélisation océanographique

3.1.1 Introduction générale

NEMO (*Nucleus for European Modelling of the Ocean*, Madec (2008)) est un système logiciel de modélisation regroupant différentes composantes, pouvant être couplées :

- **La composante de circulation océanique**, basée sur le modèle OPA (Ocean PArallélisé, Madec et al. (1998)). Elle regroupe les différents schémas numériques simulant l'évolution de la dynamique océanique. Elle inclut également des modules nécessaires à l'intégration d'un système d'AD avec NEMO, tels que l'opérateur d'observation (**OBS**) et l'application d'un incrément à l'état initial d'une simulation (**ASM**). Ces modules font partie de l'environnement **NEMO-ASSIM**, dont les objectifs seront détaillés dans la section suivante.
- **Les modèles tangent et adjoint de OPA**, regroupés sous la composante nommée NEMOTAM (*NEMO Tangent and Adjoint Models*), nécessaires à l'application des méthodes variationnelles d'AD avec NEMO. Ces modèles seront détaillés plus précisément dans la suite.
- **La composante de glace de mer**, basée sur le modèle LIM (*Louvain-la-neuve sea Ice Model*, Fichefet et Maqueda (1997))

- **La composante de traceurs passifs biogéochimiques** basée notamment sur le modèle de biogéochimie marine PISCES (*Pelagic Interaction Scheme for Carbon and Ecosystem Studies*, Aumont et Bopp (2006))

Le développement du système NEMO est organisé par un consortium européen entre le CNRS, Mercator Océan, UKMO [1], le NERC [2], l'INGV [3] et le CMCC [4]. NEMO est aujourd'hui largement utilisé par la communauté oceanographique au niveau européen et international, dans les principaux cadres suivants :

- La recherche en océanographie
- L'océanographie opérationnelle (Mercator Océan, UKMO)
- La prévision saisonnière (ECMWF [5], UKMO et Météo-France) et la projection climatique dans le cadre des expériences de l'IPCC (IPSL [6] et Météo France)

Au cours de ce travail, seuls OPA et NEMOTAM seront utilisés. La composante OPA permet de simuler l'évolution dynamique des masses d'eau. En terme d'architecture logicielle, OPA peut être décomposé en trois classes de modules, largement entremêlées, mais décrivant la philosophie de cette composante :

- **Les outils de modélisation**, *i.e.* les schémas numériques et paramétrisations représentant les processus physiques à l'œuvre dans l'océan (*e.g.* advection, diffusion horizontale ou verticale)
- **L'environnement**, décrivant le domaine spatial et temporel d'intérêt. En océanographie, cette description peut être complexe : bathymétrie, géométrie côtière ou frontières ouvertes (qui impliquent alors des forçages océaniques aux frontières), forçages atmosphériques variables en temps et en espace (vent, radiation solaire, pluviométrie), apports des réseaux hydrographiques, de la dynamique de la glace de mer, *etc.*
- **La gestion informatique** regroupe quant à elle la gestion techniques des entrées/sorties, l'optimisation du code logiciel, la modularité du système et la gestion de la parallélisation des calculs. Scientifiquement moins essentielle, elle n'en reste pas moins indispensable à la bonne marche de NEMO.

En résumé, NEMO, et plus particulièrement OPA, est un système logiciel permettant de mettre en oeuvre **informatiquement** des **outils numérique de modélisation océanique** dans un **environnement donné**.

3.1.2 Le modèle de circulation océanique OPA

La composante OPA (Madec et al., 1998) est basée sur les *équations primitives* de l'océan qui fournissent une description temporelle tridimensionnelle de la circulation océanique. Les variables prognostiques [7] sont la température potentielle T, la salinité S, le champ de vitesse horizontale $\vec{U} = (u,\ v)$ et la hauteur de la surface libre η (appelée aussi *SSH* [8])

1. *United Kingdom Meteorological Office*, service national britannique de météorologie
2. *National Environment Research Council*, agence nationale britannique de recherche scientifique
3. *Istituto Nazionale di Geofisica e Vulcanologia*, agence nationale de recherche italienne
4. *Centro Euro-Mediterraneo sui Cambiamenti Climatici*, centre d'études climatiques italien
5. European Centre for Medium-Range Weather Forecasts, centre européen de prévision météorologique
6. Institut Pierre-Simon Laplace
7. prognostiques est l'adjectif qualifiant les variables directement calculées par le modèle, à l'opposé des variables diagnostiques qui sont calculées à partir des variables prognostiques
8. Sea Surface Height

Équations primitives de l'océan

La circulation océanique est décrite par les équations de la dynamique des fluides géophysiques, à savoir :

- *Les équations de Navier-Stokes*
  Elles traduisant la conservation de la quantité de mouvement :

$$\frac{D\vec{U}}{Dt} = -\frac{\vec{\nabla}P}{\rho} + \vec{g} - \vec{\Omega} \times \vec{U} + \vec{D}^U \tag{3.1}$$

  où l'opérateur $\frac{D}{Dt}$ désigne la dérivée particulaire, $\vec{\nabla}$ l'opérateur généralisé de dérivation vectorielle, t le temps, U le vecteur vitesse, P la pression, ρ la densité, $\vec{g}$ l'accélération de la pesanteur et $\vec{\Omega}$ le vecteur vitesse de rotation de la terre. $\vec{D}^U$ est le terme de paramétrisation sous-maille associé aux processus non-résolus.
- *L'équation de continuité*
  Elle traduit la conservation de la masse :

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}.(\rho\vec{U}) = 0 \tag{3.2}$$

- *Les équations d'évolution de la température potentielle*[9] *et de la salinité*
  Elles traduisent respectivement la conservation de la chaleur et du sel dans l'océan :

$$\begin{cases} \frac{\partial \theta}{\partial t} = & -\vec{\nabla}.(\theta\vec{U}) + D^\theta \\ \frac{\partial S}{\partial t} = & -\vec{\nabla}.(S\vec{U}) + D^S \end{cases} \tag{3.3}$$

  où D^θ et D^S sont les termes de paramétrisations sous-maille pour la température potentielle et la salinité, exprimés comme une diffusion.
- *L'équation d'état de l'eau de mer* Elle relie la densité, la température potentielle, la salinité et la pression :

$$\rho = \rho(\theta, S, P) \tag{3.4}$$

  La densité de l'océan est ici une fonction empirique et non-linéaire de la température potentielle, de la salinité et de la pression *in situ* de l'océan. L'équation d'état majoritairement utilisée dans NEMO est celle de Jackett et McDougall (1995).

Une série d'hypothèses simplificatrices, liées aux échelles caractéristiques de l'océan, sont effectuées :

- *L'approximation de Boussinesq et incompressibilité de l'eau de mer*
  Les variations de densité de l'eau de mer par rapport à la densité moyenne étant faibles, on néglige ces variations, excepté dans les termes de flottabilité (*i.e.* ρg que l'on peut retrouver dans l'équation 3.1). L'application de cette approximation à l'équation de continuité 3.2 transforme cette équation en $\vec{\nabla}.\vec{U} = 0$.
- *L'approximation sphérique*
  Les surfaces géopotentielles sont supposées être des sphères. Ceci implique que $\vec{g}$ est parallèle au rayon terrestre et correspond à la verticale locale.

9. Température d'un élément du fluide ramené adiabatiquement à la surface océanique

- *L'approximation couche mince*
  La profondeur de l'océan est faible devant le rayon terrestre, elle est donc négligée dans la coordonnée radiale du système de coordonnées sphériques. Les effets de courbure ne sont alors pas pris en compte.
- *L'approximation hydrostatique*
  Les échelles caractéristiques de l'océan conduisent à négliger les termes d'advection, de Coriolis et d'accélération dans la composante verticale de l'équation de conservation de la quantité de mouvement. Il en résulte un équilibre entre gradient vertical de pression et terme de flottabilité :
  $$\frac{\partial P}{\partial z} = -\rho g \tag{3.5}$$
  La vitesse verticale w n'est donc pas une variable prognostique du modèle mais est calculée à partir de l'expression de la non-divergence du champ de vitesse. Par ailleurs, dans le cadre de cette approximation, les écoulements gravitaires ainsi que les processus convectifs ne sont plus représentés et doivent donc être paramétrés.
- *L'hypothèse de fermeture turbulente*
  Les processus de taille caractéristique égale ou inférieure à la grandeur de la maille utilisée dans le modèle ne peuvent être résolus dans le modèle. Ces processus sont donc paramétrés, et on choisit de le faire en fonction des gradients grande échelle des variables du modèle selon une forme analogue à la diffusion moléculaire.

L'ensemble de ces hypothèses permet d'obtenir le système d'équations dites *primitives* utilisé par bon nombre de modèles globaux de circulation océanique, qui le résolvent numériquement. Pour plus de détails sur les équations de la dynamique océanique, voir par exemple Pedlosky (1987) ou McWilliams (1996).

Implémentation numérique

les schéma implémentés dans NEMO sont des schémas aux différences finies, sur des grilles de type *Arakawa C*, où les variables d'état sont positionnées et indicées selon le tableau 3.1 et la Figure 3.1. Sur la verticale, nous pouvons voir que les traceurs et les vitesses horizontales sont aux mêmes niveaux (points T). Seules les vitesses verticales sont positionnées sur des points de grilles intermédiaires (points W). Les schémas en temps sont de type « saute-mouton » (*leap-frog*) assorti d'un filtre d'Asselin. De nombreux schémas d'advection et de diffusion sont disponibles. Pour plus de détail, voir Madec (2008).

θ	i	j	k
u	$i+1/2$	j	k
v	i	$j+1/2$	k
w	i	i	$k+1/2$

Table 3.1 – *Emplacement des variables selon les indices entiers et demi-entiers sur la grille de discrétisation*

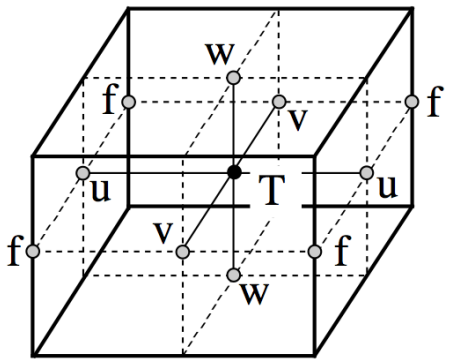

Figure 3.1 – *Placement des variables sur une maille de la grille discrétisée. f indique les points où les vorticités relatives et planétaires sont calculées.*

3.2 NEMO-ASSIM, un ensemble d'interfaces pour l'assimilation de données

Dans cette section, nous reprenons l'article décrivant le contexte, les objectifs et les activités passées, présentes et futures de NEMO-ASSIM.

TOWARD A DATA ASSIMILATION SYSTEM FOR NEMO

By P.-A. Bouttier[(1)], E. Blayo[(2)], J. M. Brankart[(3)], P. Brasseur[(3)], E. Cosme[(3)], J. Verron[(3)] and A. Vidard[(2)]

[1] *LEGI/OSUG, CNRS, Grenoble*
[2] *LJK, INRIA and University of Grenoble*
[3] *LEGI, CNRS and University of Grenoble*

Abstract

In this note, we discuss the project that has been conceived and the first achievement steps that have been carried out to set up a data assimilation system associated to NEMO. Of specific interest here are applications to operational oceanography. This data assimilation system is schematically made of three subcomponents: Interface Components, Built-in Components and External Components. Several elements of this NEMO data assimilation system have already been developed by various groups in France and in Europe and several of them could be introduced in the system (the linear Tangent and Adjoint Model, TAM, is one of the most important of them as far as variational assimilation is concerned), some others will require specific developments. Finally, we introduce the SEABASS reference configuration that is proposed to be the NEMO data assimilation demonstrator and the experimentation and training platform for data assimilation activities with NEMO. These various thoughts take advantage of the advances and discussions that have been carried out by the NEMOASSIM working group.

Introduction

Other contributions of this issue discuss how the NEMO modeling platform can be used to simulate the time evolution of the ocean circulation including its variability from global to regional scales. Due to the non-linearity of the equations governing the ocean dynamics, a wide range of such temporal and spatial scales interact together in such a way that the ocean evolution is partly chaotic and beyond some limit, unpredictable. Therefore, the routine monitoring and forecasting of oceanic variables, which is the essential goal of Operational Oceanography, must be treated as a series of inverse problems that require observed data at regular intervals to re-initialize the model state “close” to the observed ocean using all available data combined with the latest model predictions. In this respect, operational ocean monitoring is similar to numerical weather forecasting.

The terms “Data Assimilation” (DA) designate the range of objective methods enabling optimal combination of observations, model simulation and error statistics, in order to reduce as much as possible the uncertainty of ocean state estimations involved in short-term predictions or more or less long-term reanalyses. Very significant progress has been accomplished in ocean data assimilation during the past 20 years in the framework of a variety of pre-operational projects, such as the French SIMAN/QADRAN program in the nineties (*e.g.* Blayo et al., 1994), the DIADEM, TOPAZ (Brusdal et al., 2003), ENACT (Davey et al., 2006) and MERSEA European projects (Brasseur et al., 2005), and more recently the GMES MyOCEAN and on-going MyOCEAN2 and SANGOMA projects. The choice made in Europe to routinely monitor the ocean down to the meso-scales has strongly guided the first stages of the assimilation strategy in place today in most operational centres. At international level, the effort of the nations involved in the development of DA for Operational Oceanography were coordinated in the framework of GODAE (Cummings et al., 2009), demonstrating the relevance of the concept which is further developed in the framework of the on-going GODAE Oceanview.

Briefly speaking, two different categories of algorithms can be distinguished to solve the DA inverse problems: the optimal control approach, most often based on the variational adjoint method (Le Dimet and Talagrand, 1986) and the stochastic methods mostly derived from the Kalman filter concept. In its 4D-Var formulation, the variational method requires the adjoint of the linear tangent model to compute the gradient of the cost function to be minimized, and in that case can therefore be designated as a « model-dependent » DA method. By contrast, stochastic methods such as the Ensemble Kalman filter introduced by Evensen (1994) or the SEEK filter introduced by Pham et al. (1998) can be considered as « model-independent » DA methods which intensively use the direct model code to propagate ensemble statistics, while the update of these ensemble statistics requires additional « model-independent » algebraic operations. The EU SANGOMA project has been set up during the period 2012-2015 to advance stochastic assimilation methods, focussing on non-linear and non-Gaussian assimilation schemes to be used in the next operational systems of the GMES Marine Core Services.

Due to the fast evolution of ocean models during the past 20 years, thanks mainly to computer power increases, the flexibility of « model-independent » DA methods has been an asset to follow the successive updates of ocean model versions without much recoding. Today, the convergence of some of the oceanographic community in Europe toward the NEMO modeling platform provides the opportunity to revisit the overall assimilation strategy, since a more stable and smoother evolution of the model platform can be expected in the future. This is in essence the primary motivation of the project described in the present paper.

Despite some earlier attempts, no assimilation system had ever been formally included in the NEMO system so far. However, a number of DA frameworks already use NEMO as model component: e. g. SESAM (Brankart et al., 2001), SAM (Drillet et al., 2008), OPAVAR/NEMOVAR (Weaver et al., 2003, Mogensen et al., 2009), OceanVar (Dobricic and Pinardi, 2008) and many papers have been published discussing data assimilation results within OPA/NEMO. Since some common components are required by every system, there was therefore some duplication of

efforts. In order to provide a more structured offer to a large number of users, CNRS proposed in 2009 to set up the so-called NEMOASSIM working group. This group gathers experts involved in the development of such systems (members of CERFACS, CNRS, ECMWF, INGV/CMCC, INRIA, MERCATOR and MetOffice), and members of the NEMO system team. This leads to sketch a strategy for implementation of an assimilation component associated to the NEMO code system and thereby make assimilation tools for NEMO more readily available to the user community.

In this note, we present the current status and perspectives of what could be the different components of a NEMO data assimilation system, which are categorized into Interface Components such as OBS, ASM components which make the link between observations and model-structured information, "Built-In" Components, such as TAM that is the linear tangent and adjoint models of NEMO, which are intrinsically linked to the direct NEMO code, and External Components to be included into NEMO DA system (4DVAR minimizer, square-root or ensemble analysis kernels, singular value decomposition), which are technically independent from NEMO, but nonetheless essential to build a stand-alone assimilation system. All these components (Interface Components, Built-in Components and Externals Components) are thus needed together to implement the NEMO DA system. Then, SEABASS, an academic NEMO configuration will be proposed as NEMO benchmark basis and demonstrator for DA. A number of perspectives will be discussed regarding the future of the NEMO DA system as a conclusion to this note.

NEMO Data Assimilation: Interface Components

The interface components are the modules that are NEMO-dependent, and which connect NEMO to external information independent from the model output.

The first two components to be included in the standard release of NEMO (from version 3.3 onward) were the *observation operator* (OBS) and the *application of the analysis increment* (ASM) since they represent the common interfaces needed to most assimilation schemes relevant to systems like NEMO. Indeed, the fundamental input to any assimilation kernel is the vector of differences between the observations and the reference state. This vector is usually known as the *innovation vector*, and is a direct product of OBS. On the other hand, the fundamental output of any assimilation kernel is a vector of corrections to the background state on the model grid (known as the *analysis increment*). Typically, it consists of corrections to the model current state but may also include corrections to other fields such as the surface forcing fields, model tendencies, or system parameters. The analysis increment is used to initialize or correct the model trajectory, and this is done through the ASM component.

The main purpose of observation operators is to transform model variables on grid points into quantities which can be directly compared with observations. For example, for in situ observations this can simply be interpolation from the model grid to observation points, but more complicated operators involving non-linear transformations of model variables to produce the observable quantity are also possible. Observation operators are mostly independent of the assimilation kernel. Moreover they are also valuable as a diagnostic tool for evaluation of model performance, since they provide the possibility of comparing model variables with observations. The choice proposed by the NEMOASSIM working group was to include, as a first version, the OBS module coming from NEMOVAR. It came in the NEMO 3.3 reference version as a contribution from the MetOffice (which is also a member of the NEMOVAR consortium) along with full documentation (Lea et al., 2012a).

There are several issues to deal with when developing such operator, namely what input file format should be available, how the output will be handled over any assimilation or diagnostic tool and how and where the observation operators are called. For the latter, two choices are available: either it is done off-line, once the model has finished its time integration, but that would potentially require a tremendous amount of extra I/O, or it can also be done on-line, for an extra computing time during the model integration, and that is the adopted solution in NEMO. In the current version, OBS is able to read data such as profiles of temperature and salinity from CORIOLIS or ENACT/ENSEMBLE database, GRHSTT or Reynolds sea surface temperature, AVISO sea level anomaly, sea ice concentration, and TAO/PIRATA/RAMA velocity, both in their original format and in the NEMOVAR feedback format. In the future it is likely that only the latter (with some improvement if needed) will be available in order to avoid maintaining many possible formats. A set of converting tools to produce feedback formatted observation files is already available in the NEMO reference version. The feedback format is also used as the output of the OBS module where the model equivalent to observation at the same time and place is also present, along with all associated information (position, time, QC flags, observation type and instrument, etc). This format represents the interface on the model side, it is then up to the user to build the interface between feedback files and the desired assimilation scheme or diagnostic package.

In most of the assimilation methods relevant to NEMO, the trajectory is controlled by introducing a correction to the model state. This correction, or *increment* is produced by the *analysis step* on variables linked to a control vector. The increment updates the model trajectory either directly, or in a gradual manner over a time period around the analysis date. This latter approach, usually referred to as the *Incremental Analysis Updates* (IAU) methodology (Bloom et al., 1996), tends to reduce shock in the model restart stage and to minimize spurious adjustment processes. From a technical point of view, this latter concept of a correction introduced progressively during the model run is not closely linked to the assimilation methods and could also be seen or interpreted as a use of forcing terms. In practice, these forcing terms correspond to 2D and/or 3D fields (*e.g.* temperature, salinity, or currents) and are applied directly as additional terms in the equations of the NEMO code (Ourmières et al., 2006).

As for the OBS module, the first version of ASM was imported from NEMOVAR as a Met Office contribution (Lea et al., 2012b). Both these choices were driven by the fact that the NEMOVAR code is sufficiently compatible with the NEMO code (same coding convention and practices and limited modification required to the standard reference).

Several possible further developments and improvements to the interfaces (OBS and ASM) are considered today. For instance, for the ASM module, new specificities will be introduced such as new fields (*e.g.* ice concentration), new IAU weight functions and memory usage optimisation. It

would also be appreciated that ASM could manage several increments at the same time (as it is done in SAM). The supervision of their development is under the responsibility of the MetOffice.

Regarding the OBS module, its evolution goes through adding new types of observations. Special care should be devoted to ensure that OBS remains flexible enough to easily include both simple or sophisticated new operators.

NEMO Data Assimilation: Built-In Component

The other current key element of the NEMO DA system is NEMOTAM (stands for NEMO Tangent and Adjoint Model) or simply TAM. The development of tangent and adjoint models is an important step in addressing variational DA problems. In such methods, one minimizes a cost function that is a measure of the model-data misfit, and the adjoint variables are used to build the gradient for descent algorithms. Similarly the tangent model is used in the context of the incremental algorithms to linearize the cost function around a background control. During the ANR–funded project VODA, specific effort was dedicated to the development of the TAM for the ocean engine component of NEMO.

The only needed interface between the direct model and the TAM is the handling of the non-linear trajectory (required for differentiating around). So, once available, the inclusion of the TAM within the NEMO framework is straightforward. A version of TAM was released along with NEMO 3.2.2 and made available to the users on the NEMO website, although it was not fully part of the standard release due to lack of resources. With the support of CNRS, recent efforts have been dedicated to this point that will allow TAM to be soon part of the standard release as of NEMO 3.4 (Bouttier et al., 2012).

Assimilation-wise Tangent-linear and Adjoint Models are mainly used for variational assimilation applications. However they are also powerful tools for the analysis of physical processes. They can indeed be used for sensitivity analysis, parameter identification and for the computation of characteristic vectors (singular vectors, Lyapunov vectors, etc), which in their turn can be used for defining reduced order assimilation schemes.

Sensitivity analysis for instance is the study of how model outputs vary with changes in model inputs. The sensitivity information is given by the adjoint model, which provides the gradient of the outputs w.r.t. the inputs. One can find an example of application of such methods in the study conducted by Vidard et al. (2010) on the GLORYS ¼° global ocean reanalysis. The initial objective was to estimate the influences of geographical areas to reduce the forecast error using an adjoint method to compute sensitivities. They conducted a preliminary study by considering the misfit to observations as a proxy of the forecast error and sought to determine the sensitivity of this misfit to changes in the initial condition and/or in the forcing fields. Without going into the details of this study, one can see an example of sensitivity to initial temperature (surface and 100m) as shown in the two bottom panels of Figure 1. In this example it is clear that the SST misfit (top left) is highly sensitive to changes in surface temperature (bottom left) where the initial mixed layer depth (top right) is low. The opposite conclusion can be drawn from the sensitivity to the initial temperature at 100m (bottom right). This is obviously not a surprise, and is more useful here for the assessment of the model (more precisely of the tuning of vertical mixing) rather than to the original goal of assimilation system improvement.

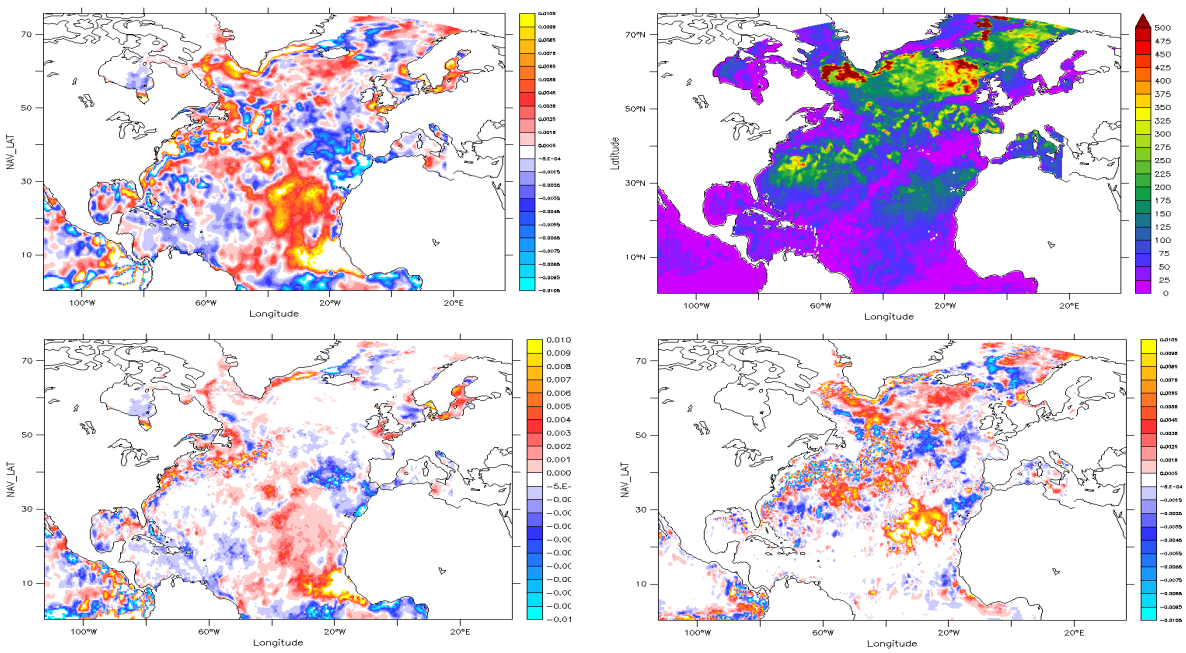

Figure 1: Top: misfit between forecast and observed SST (left) and mixed layer depth (right). Bottom: sensitivity to one-week lead-time SST error with respect to variations in initial surface (left) and 100m (right) temperature.

TAM and associated minimization tools have originally been developed for DA purposes. However, as mentioned above, they could be used for improving the understanding on the model behaviour as well. For example, they can be used to diagnose the differences between model simulations and ocean measurements and to provide valuable information on the robustness of numerical schemes. They can also be used to improve

the balance in initial conditions, particularly for shelf-seas models or to assess the structures of the fastest growing instabilities in any flow.

For the future, it is expected that a number of further developments should be added. Several NEMO functionalities (for example AGRIF or Open Boundary Conditions), are not available yet in the current development of TAM. They will be under consideration for next versions of TAM. Estimation of error statistics is a key product of data assimilation and is not readily (and simply) available in variational schemes. There are some potentialities from the hybridization of variational and stochastic methods to bring some solutions. This should be studied further in the future. In addition, one could think that diagnostic tools should be made available in a coordinated way to the users of the NEMO data assimilation system. Note that many of such tools have already been developed in various groups. Built-in components would include physical diagnostic tools whereas specific methodological diagnostics should be in the following external components.

NEMO Data Assimilation: External Components

The NEMO data assimilation interfaces described before has been developed seeking generality and hopefully should be usable easily by any modular-enough assimilation engine. These interfaces have already been used by at least two major complex DA systems: NEMOVAR and SESAM (precursor of Mercator-SAM2). Although they are not part of the NEMO system per se, they are distributed under the same Cecill licence and can therefore be used in conjunction with NEMO.

These two External Components are quite different in their concept; therefore it is a good illustration of the generic aspect of the interfaces. On the one hand NEMOVAR is based on optimal control techniques where the model trajectory is kept close to the observation through the minimization of an objective function, using a descent algorithm. It uses the so-called incremental formulation where the minimization of the full problem is approximated by a succession of simpler problems, assuming that the evolution of the increment follows the tangent linear model. The algorithm is composed of sequences of two-stage processes: an integration of the non-linear direct model that computes the misfit to observations and possibly outputs the non-linear trajectory (outer iteration) and a minimization of the linearized problem (inner loop). The non-linear trajectory is then updated applying the computed increment during a new outer iteration possibly followed by another inner loop, and so on and so forth. In NEMOVAR, outer iterations and inner loops are separate executables. The outer loop is a standard NEMO model integration making use of OBS (resp. ASM) output (resp. input) files to communicate with the inner loop, while the inner loop is a specific program including the multiplication by the error covariance matrices, the computation of the cost function and its gradient thanks to a tangent-linear and adjoint models integrations and several minimization drivers. Note that NEMOVAR can also perform a 3D-FGAT-type assimilation where the evolution of the increment is assumed stationary in the inner loop, avoiding the need for the tangent and adjoint models integrations.

On the other hand, SESAM is a toolbox of assimilation modules originally developed to implement the SEEK filter algorithm (a reduced rank Kalman filter developed by Pham et al., 1998). In sequential methods, the assimilation algorithm consists of two successive operations: a forecast step to propagate in time the initial probability distribution for the state of the system and an analysis step to update this distribution using available observations. The forecast step is performed using the NEMO model, either by propagating the initial covariance using a linearized model operator (as in the original SEEK algorithm) or by producing an ensemble of model forecasts starting from a sample of the initial probability distribution. The analysis step is implemented in SESAM using the SEEK observational update algorithm, which is especially efficient with large observation vectors and low rank covariance matrices. Moreover, the original algorithm is now extended to work efficiently with localization of the forecast error covariance matrix, non-diagonal observation error covariance matrix and adaptive forecast and observation error statistics. It is still a linear algorithm, but extensions exist in SESAM (providing that an ensemble forecast is available) to account for non-Gaussian distributions using anamorphosis or a truncated Gaussian assumption. From a practical point of view, there is a clear separation between the forecast step, only involving NEMO model integrations, and the analysis step, implemented in the SESAM software. The overall system can thus be written as a master program (a shell script for instance) cycling forecast and analysis steps with quite simple interfaces between them.

For some scientific applications, especially for academic process configurations, it is sufficient to use simple error covariance matrix definitions. The so called B matrix required by 4D-VAR DA can thus be defined more simply than going through NEMOVAR: a simple tool should be developed for this. In the same way, a simplified version of the error observation definition routine (the so-called R) is under consideration.

CNRS manpower could be devoted together with other partners, to work toward the integration of these external components within the overall NEMO DA system.

SEABASS: a NEMO reference configuration for data assimilation

In the same way than the NEMO code has some reference demonstrating configuration (such as ORCA2, GYRE, etc …), it is proposed here to set up a reference assimilating configuration that can be used as a first demonstrator of several assimilation methodologies applied to NEMO. In this first stage, such a configuration has to be flexible, easy to manipulate and to maintain. Physically, it has to represent some key ocean dynamical processes. Moreover, its numerical cost must be reasonable.

An academic ocean basin double-gyre configuration is proposed. The rectangular domain of this configuration, now called Sea Box for Assimilation (SEABASS) extends from 24°N to 44°N and over 30° in longitude. Any horizontal resolution can be simply specified. For a 1/4° horizontal resolution, the grid contains only 121 points in longitude and 81 points in latitude. The ocean is sliced into 11 verticals levels, from surface to 4000

meters, described with a z-coordinate. The domain is closed and has a flat bottom. Lateral boundaries conditions are frictionless and bottom boundary condition exerts a linear friction. The circulation is only forced by a zonal wind. Lateral dissipation is performed on dynamics and tracers with a biharmonic diffusion operator. The salinity is constant over the whole domain and the initial stratification is produced using an analytical temperature profile. Details can be found in Cosme et al. (2010), for example.

Actually, this type of configuration has a long story that comes from ancient QG models implemented earlier to mimic the Gulf Stream system (e. g. Holland & Lin, 1975) and serve as a common basis through years to exemplify dynamical processes typical of the mid latitude like the Gulf Stream system with an unstable central jet, presenting meanders and eddies, as illustrated on Figure 2. Even very simplified, it is statistically meaningful regarding the eddy activity and the non-linearity amplitude of the actual Gulf Stream system. Clearly, the horizontal resolution of SEABASS configuration can be simply modified. As seen on Figure 2, which shows instantaneous relative vorticity at surface for three different horizontal resolutions (1/4°, 1/12° and 1/24°), increases in resolution allow stronger evidence of mesoscales and smaller scales activities.

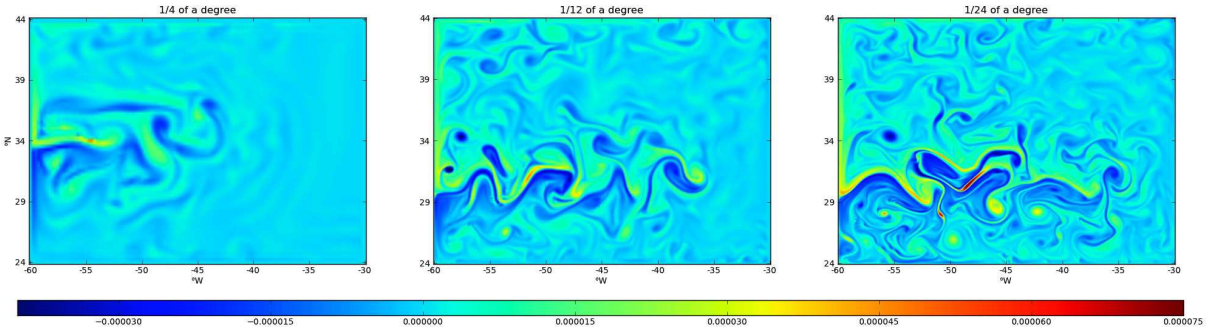

Figure 2: Instantaneous relative vorticity fields from SEABASS configuration, taken at different horizontal resolutions (from left to the right 1/4°, 1/12°, 1/24°)

SEABASS is a compromise between simplicity and a good representation of non-linear dynamical processes, increasing with its horizontal resolution. From a practical point of view, this configuration is easy to maintain across NEMO evolutions. This simplicity ensures that results obtained with different DA systems will probably be robust to most possible numerical evolutions of the NEMO code. In addition for variational DA, this configuration is fully differentiable, which ensures that Tangent and Adjoint Models do not contain approximations of the direct model. Several DA methods have also already been studied and tested on this configuration: SEEK filter and smoother, variational DA methods, nudging, Back-and-Forth nudging, ...

The ambition for SEABASS within NEMO is to have this simple model configuration together with the above DA methodologies and the associated tools and interfaces freely available to users.

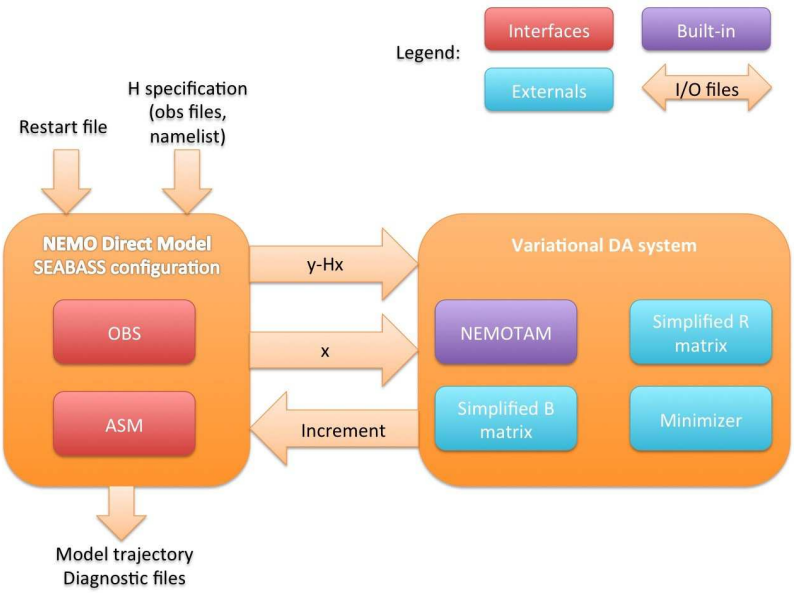

Figure 3a

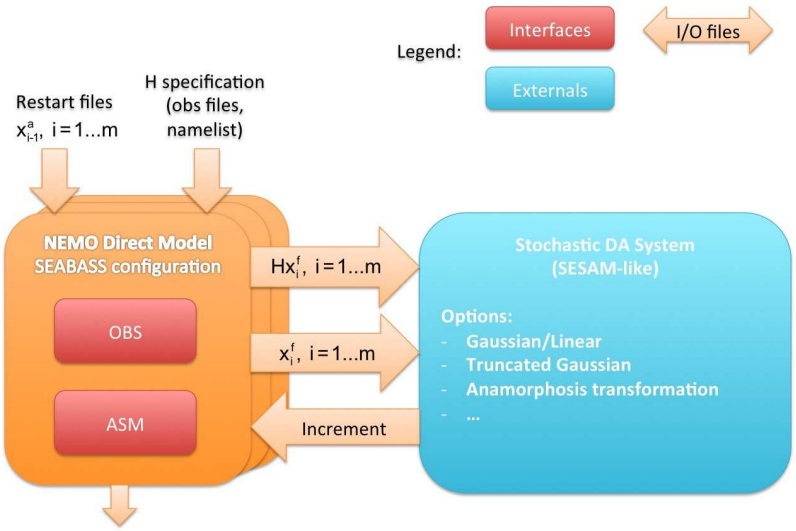

Figure 3b

Figure 3: Sketch of a data assimilation system for the SEABASS demonstrator: a) for a variational DA system, b) for a stochastic DA system

Figure 3 intends to show in a schematic representation how the NEMO data assimilation system presented here could be set up with the present SEABASS demonstrator in the case of a variational DA system (Figure 3a) and of a stochastic DA system(Figure 3b). Note that in this simplified SEABASS test case, some external components have been added: a minimizer, simplified error covariance matrices (B and R) and a stochastic analysis tool.

Conclusion

The goal of this paper was to set up some proposals to rationalize the developments related to data assimilation in NEMO and with the aim of building progressively a flexible data assimilation component designed for the NEMO system. In addition and most importantly, there is the willingness to make all of these freely available to all interested users.

There is a real diversity of data assimilation methods, addressing different scientific and operational challenges. In the future, the accumulated experience in inverse methods and the growing amount of computing resources will even accentuate the use of these different data assimilation methods for multiple applications: operational forecasting, parameter estimation, sensitivity analysis, design of new parameterizations… New observations from space or in-situ will also be a strong incentive for these developments. This will result in a number of new challenges, among which some can probably be anticipated.

From a methodological point of view, it appears clearly that Ensemble approaches encounter increasing interest, either for operational oceanography, for research studies, or for climate prediction. These Ensemble approaches can be either purely stochastic, as EnKF and its derivatives, or hybrid methods, as En-4DVAR under present development in several meteorological operational centres (Buehner et al., 2010).

Another important aspect of future data assimilation systems will be their capacity to take into account totally new types of observations and/or a dramatically higher flux of information. This will be the case with the advent of high resolution altimetry satellites such as with the NASA-CNES SWOT satellite mission project that should be launched around 2019. Relatively new sensors such as gliders also offer prospects that have not been carefully explored. Also, one can question the optimality of the use of existing sensors in assimilating mode, e. g. ocean colour satellite data. The increasing flux of information of SWOT type for example may raise new methodological issues for data assimilation, e.g. in the direction of image data assimilation or of multi-incremental approaches.

In the future the scope of ocean applications making use of data assimilation will broaden. For example, biogeochemical models bring their own specificities to data assimilation. They are generally highly non-linear and statistically not well conditioned (e. g. non-Gaussianity) and include a number of poorly known parameters. In this context, data assimilation tools may not only be used to perform state estimation but also to calibrate those parameters and validate model outputs. The expertise gained on these topics with NEMO will be a valuable asset when going toward more integrated Earth systems, coupling the ocean to components (atmosphere, ice, hydrosphere, …). In this prospect, the NEMO data assimilation environment should be adaptive enough to help the users to properly cope with these present and future scientific and operational issues.

Acknowledgement

Many points presented in the present paper were discussed within the NEMOASSIM working group (M. Bell, S. Drobicic, D. Lea, C. Lévy, M. Martin, K. Mogensen, C. E. Testut and A. Weaver).

References

Blayo E., Verron J. and Molines J.-M., 1994: Assimilation of Topex/Poseidon altimeter data into a circulation model of the North Atlantic. J. Geophys. Res., 99 (C12), 24691-24705.

Bloom S. C., Takacs L. L., DaSilva A. M. and Levina D., 1996: Data assimilation using incremental analysis updates. Mon. Wea. Rev., 124, 1256-1271.

Bouttier P.-A., Vidard A., Vigilant F., 2012: NEMO Tangent & Adjoint Model (NEMOTAM), TAM reference manual & user guide. Technical report, MEOM-LEGI and INRIA report, Grenoble. 44 pp.

Brankart, J.-M., Testut, C.-E., and Parent, L., 2001: An integrated system of sequential assimilation modules. SESAM3.2 reference manual. Technical report, MEOM-LEGI, Grenoble. 85 pp.

Brasseur P., Bahurel P., Bertino L., Birol F., Brankart J.-M., Ferry N., Losa S., Rémy E., Schroter J., Skachko S., Testut C.-E., Tranchant B., Van Leeuwen P.J. and Verron J., 2005: Data Assimilation for marine monitoring and prediction: the MERCATOR operational assimilation systems and the MERSEA developments, Q. J. R. Meteorol. Soc., 131, pp 3561-3582.

Brusdal, K., Brankart, J.-M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P.-J., Dombrowsky, E., and Verron, J. (2003). A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. Journal of Marine Systems, 40-41, 253-289.

Buehner, M.; Houtekamer, P. L.; Charette, C.; Mitchell, H. L. & He, B. Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments. Monthly Weather Review, 2010, 138, 1550-1566

Cummings J., Bertino L., Brasseur P., Fukumori I., Kamachi M., Mogensen K., Oke P., Testut C.-E., Verron J. and Weaver A., 2009: Ocean Data Assimilation Systems for GODAE. Oceanography. vol. 22, n°3, 103-115.

Cosme E., Brankart J.-M., Verron J., Brasseur P. and Krysta M., 2010: Implementation of a reduced square-root smoother for high resolution ocean data assimilation. Ocean Modeling, 33(1-2), 87-100.

Davey M., Huddleston M., Ingleby B., Haines K., Le Traon P.-Y., Weaver A., Vialard J., Anderson D., Troccoli A., Vidard A., Burgers Gerrit J. H., Leeuwenburgh O., Bellucci A., Masina S., Bertino L., and Korn P, 2006: Multi-model multi-method multi-decadal ocean analyses from the ENACT project. CLIVAR Exchanges, 11(3) :22–25, 07 2006.
Dobricic, S. and Pinardi N., 2008: An oceanographic three-dimensional variational data assimilation scheme. Ocean Modeling, 22, 89-105.

Drillet, Y., Bricaud C., Bourdallé-Badie R., Derval C., Le Galloudec O., Garric G., Testut C.-E. and Tranchant B., 2008: The Mercator Ocean global 1/12° operational system: Demonstration phase in the MERSEA context, The Mercator Ocean Newsletter #29, April 2008.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistic. J. Geophys. Res. 99, 10143–10162.

Holland, W. R., and L. B. Lin, 1975: On the origin of mesoscale eddies and their contribution to the general circulation of the ocean. I. A preliminary numerical experiment. Journal of Physical Oceanography, 5, 642-657

Lea D., M. Martin, K. Mogensen, A. Vidard, A. Weaver, 2012a: Observation and model comparison in NEMO ocean engine v3.4 Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace No 27, Chapter 12.

Lea D., M. Martin, K. Mogensen, A. Weaver, 2012b: Apply assimilation increments in NEMO ocean engine v3.4 Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace No 27, Chapter 13.

Le Dimet, F. X. and Talagrand, O., 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus, 38A:97-110.

Mogensen, K., Balmaseda M. A., Weaver A. T., Martin M. and Vidard A., 2009: NEMOVAR: A variational data assimilation system for the NEMO model. ECMWF Newsletter, Summer Edition.

Ourmières Y., Brankart J.-M., Berline L., Brasseur P. and Verron J., 2006 : Incremental Analysis Update implementation into a sequential ocean data assimilation system, J. Atmos. Ocean. Technol, 23(12), 1729-1744.

Pham D. T., Verron J. and Roubaud M. C., 1998: Singular evolutive extended Kalman filter with EOF initialization for data assimilation in oceanography. J. of Marine Systems, 16 (3-4), 323-340.

Vidard, A., Vigilant, F., Rémy, E., Greiner, E., Deltel, C., Benshila, R. 2010: Sensitivity analysis through adjoint method: application to the GLORYS reanalysis. Mercator report 08/D43

Weaver, A.T., J. Vialard, and D.L.T. Anderson, 2003: Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean, Part 1 : formulation, internal diagnostics and consistency checks. Monthly Weather Review, 131, 1360-1378.

3.3 NEMOTAM : les modèles linéaire tangent et adjoint de NEMO

NEMOTAM recouvre les modèles linéaire tangent et adjoint de NEMO ainsi que leurs différentes routines de validation. Comme nous l'avons vu au chapitre précédent, ils permettent d'obtenir le gradient de la fonction coût du 4DVAR incrémental, nécessaire lors de la minimisation de cette dernière.

Dans cette section, nous reprenons la partie descriptive de NEMOTAM, les choix numériques qui ont été faits pour son développement et les différentes méthodes de validation décrits dans un article en préparation, co-écrit par Arthur Vidard [10], Franck Vigilant [11] et moi-même.

3.3.1 Introduction and history

The package made of the tangent linear and adjoint models (TAM in the following) is a powerful modeling tool for forecasting or analysis systems. From the discrete dynamical systems described by numerical models, the tangent linear model provides the Jacobian of the dynamical operators that is tangent linear to a solution trajectory of the nonlinear system. The adjoint of this Jacobian operator provides information about the sensitivity of the system to variations of its input. TAM is therefore widely used for variational assimilation applications, but also for the analysis of physical processes, since it can be used for sensitivity analysis, parameter identification and for the computation of characteristic vectors (singular vectors, Liapunov vectors, etc. see Moore et al. (2004) for an extended review).

This is particularly true for geophysical applications such as meteorology and oceanography where many data assimilation systems rely on the availability of such models. However, only few ocean general circulation models are routinely provided with their TAM. Among them, one can cite MITGCM and ROMS. They are quite symbolic of the two possible route for deriving tangent and adjoint model, either using automatic differentiation tool (MITGCM uses TAF) or hand-coded (ROMS' TAM was developed by a pool of researchers Moore et al. (2004)).

Automatic tools are now mature enough, and they show some significant advantages over the hand-coded route. Differentiation of a numerical code is a tedious and error-prone task, the use of an automatic tool could alleviate this difficulty. Moreover when an updated version of the non linear model is provided the corresponding TAM can in general be obtained effortlessly. Additionally automatic tools offer an important flexibility when one wants to change the variable to differentiate around. Indeed, typically, for data assimilation purpose, the TAM is differentiated around the initial condition, while for parameter estimation it is differentiated around the sought parameter state. Most of the time the obtained TAM are very similar, but in some cases (e.g. for grid related parameter) going from one to another can lead to a significant amount of code changes.

On the other hand, automatic differentiation suffer from some limitations compared to hand-coding. First some newer (or archaic) language features may not be supported at first by the automatic tools, second the numerical performance of automatic derived TAM is still relatively poor compared to that of hand-coded one, in particular, the handling of the parallelization is still an open issue. Third non differentiable part of the original code still require to be delt with

10. Équipe MOISE, INRIA/LJK
11. Ingénieur, Bull

specialist in the field. All these issues can however be overcome by mixing automatic and manual coding.

The NEMO ocean engine (Madec, 2008) was previously known as the OPA model (Madec et al., 1998). It used to have a TAM (called OPATAM), fully hand-coded. OPATAM was initially developed for a Pacific ocean configuration, and targeted at variational data assimilation applications in the framework of OPAVAR (Weaver et al., 2003, 2005). OPATAM/OPAVAR were extended to other regional basins (Mediterranean sea (Rémy, 1999), North Atlantic 1/3° (Forget et al., 2008), South Atlantic 1°), to the global ocean (ORCA 2° (Daget et al., 2009a)), and were used for methodological studies such as control of the 3D model error (Vidard, 2001), control of the surface forcing and open boundary conditions (Deltel, 2002; Vossepoel et al., 2003). OPATAM was also used for sensitivity studies (Sévellec et al., 2008), singular vectors (Moore et al., 2003; Sévellec et al., 2009), etc.

For several reasons, mainly because of lack of workforce, OPATAM, OPAVAR and related developments were not included in the standard release of OPA. As a consequence, synchronisation of OPATAM with OPA's releases could not be achieved on a regular basis, and all developments were on individual branches, without feedback to the OPATAM/OPAVAR system. The pool of potential users was reduced consequently. The complete rewriting of the model during the transition to NEMO rendered OPATAM obsolete. A first prototype of NEMOTAM was obtained by (Tber et al., 2007) using the TAPENADE automatic differentiation tool (Hascoët et Pascual, 2004) for a fixed and somewhat simplified configuration (ORCA 2° with all non-differentiable options switched off (Tber et al., 2007)). Even for this simplified configuration, however, substantial human intervention and additional work was required to obtain a useable product from the raw generated code. Three main drawbacks were identified for this application. First, the memory management and CPU performance of the raw code were rather poor. Second, the current version of TAPENADE generates single-processor code only and cannot handle directives from the C-PreProcessor (CPP keys), which are widespread in NEMO. Third, the technique of binomial checkpointing that is used to handle nonlinearities (see (Tber et al., 2007)) is not compatible, at least in its present implementation, with the incremental algorithm of NEMOVAR, which employs separate executables and (possibly) different resolutions for the outer and inner loops. Improved memory management and extensions to support MPP and CPP keys are planned in future versions of TAPENADE so the first two deficiencies are not fundamental. The third deficiency, however, is more problematic and the trajectory management for nonlinearities in NEMOTAM is done differently from TAPENADE, along the lines of the simpler strategy implemented in OPAVAR. The modifications required to make any AD tool, compatible with the multi-incremental approach are really substantial and cannot be done in a short or medium term. Moreover the numerical performances of the generated TAM do not allow yet their use for "big" configurations and for operational applications. The writing of an adjoint code is not a simple technical task, this explains why Autodiff tools can seldom be used as a black-box. Numerous potential problems can arise and in general they should be dealt with on a case by case basis by someone who masters both the model and differentiation techniques.

From that experience it has been decided to go toward the hand-coding approach, the use of AD tool being left aside for the time being. We may reconsider this in a medium or long term though. An optimal mix of both approaches is likely to be the preferred choice in the short term future and lead to a semi-automatic way of generating NEMOTAM.

3.3.2 Methodology and choices for NEMOTAM

The approach used in developing NEMOTAM is based on the well known differentiation rules as described in Giering et Kaminski (1998). It relies in particular on the definition of active and passive variables. The former depends on the control variables (variable one differentiates around) variations, while the latter are independent of the control (typically model or grid parameters). Active variable have tangent and adjoint counterparts while passive variables do not.

In the current version (3.4.1), only the general circulation component of NEMO is supported by NEMOTAM and a few key components are still missing, namely the variable volume (VVL), open boundary conditions (OBC) and the nesting capabilities (AGRIF). There is no fundamental reason for non supporting these options (support for OBC is actually on its way), but it was not flagged as priority. In particular VVL would require a tremendous amount of coding due to the fact that it makes the size of the surface grid cells depending on the flow (they become then active variables). This is a typical case where a semi-automatic route would be beneficial.

When coding tangent and adjoint models, either automatically or manually, one has to face several difficulties. The main ones are listed in the remaining of this section and discussed in the particular case of NEMOTAM.

Non-differentiabilities : Problem and solutions

In realistic models like those included in NEMO, non-differentiable physics are quite frequent. They can be non-differentiable by essence (e.g., step processes), due to the way they have been programmed or the chosen numerical methods (e.g. non-oscillary schemes). Coding-wise they can be represented by an IF statement with a condition on an active variable (or a max, abs, etc. statement but that is equivalent). In order to manage these non-differentiable parts several options are available :

- Regularisation : either mathematical regularisation (thanks to the introduction of a differentiable connection), or a physical regularisation whenever it is possible (by rewriting the processes in a differentiable manner).
- Approximation : The direct model physics is approximated in TAM in order to transform the non-differentiable part. One should obviously be careful that the approximation is not too strong.
- Simplification : the non-differentiable part is deleted from TAM. This can be done, either by using the same branching (same side of the IF) in the TAM than in the direct model, or by simply removing the concerned part. Again, this should be done very carefully.

The choice of treatment heavily relies on the type of the non-differentiability and it is more a matter of an educated choice. Indeed it requires an important knowledge of both the direct model and differentiation techniques. A typical example of this in NEMO is the vertical mixing schemes included in OPA (TKE in particular, but it is true for the other options as well). The current version is strongly non-differentiable. In NEMOTAM this has been sorted out by simplification of the physics, i.e. some active variables are computed by the direct model and treated as passive variable in the TAM. Adopting this strategy has the advantage to keep the full physics in the non linear model, but is at the expense of an approximation in the TAM. However, most of the physical process it models being quite regular, it may be possible to rewrite the direct version to make it differentiable [12].

12. This was the strategy adopted in ROMS for the KPP vertical mixing scheme

Checkpointing

One of the issue one has to tackle when dealing with TAM is the storage and/or re-computation of the non-linear trajectory that is required for the differentiation of the non-linear terms. This is particularly important for the adjoint model which needs the trajectory in the reverse order as it is produced by the non linear model. A common practice is the so-called checkpointing where well-chosen snapshots of the trajectory (the checkpoints) are stored and between the checkpoints the intermediate variables are re-computed. This is the strategy used in NEMOTAM : a subset of the direct variables are stored every time step and only the required intermediate variables are recomputed. The number of non-linearities in NEMO being relatively small, the need for storage and re-computation is limited. However, for long period of time the amount of required storage may become too important ; in that case it is possible to subsample the output of the trajectory (say one per day) and NEMOTAM will interpolate linearly between checkpoints. Additionnally it can be stored in single precision. The validity of these approximations is discussed in section 3.3.3.

Numerical issues

The numerical characteristics of the direct model are not always conserved in the adjoint model. In particular the two following difficulties are quite common :

- The adjoint of a numerically stable scheme can be unstable.
- The convergence speed of an iterative algorithm may be different for the direct and its tangent/adjoint.

In both cases, a specific solution should be provided. It can go as far as replacing the problematic scheme by a more TAM-friendly one. Once again, in OPATAM this kind of problem arose, and we had to replace the conjugate gradient solver used for the computation of the surface pressure gradient (while self-adjoint, in theory) by a red-black SOR solver. However, this scheme being on the verge of depreciation in NEMO, NEMOTAM will soon adopt the time-splitting-based direct solver.

Stability is not the only numerical issue : the TAM are, by essence, more expensive numerically than the direct model because of the additional computations they require. Moreover, and it is especially true for the adjoint model, the direct code optimisations may not be so optimal for the TAM. Therefore it is important to specifically optimise the TAM code, and this can only be done through a careful performance analysis. As of today, such an optimisation may not be achieved automatically. The tendency toward application at high and very high resolution make the computing cost aspect a crucial issue. In NEMOTAM the same domain decomposition strategy as for NEMO has been adopted and is therefore fully parallel, and some specific targeted optimisation have been performed. Additionally some irrelevant (as far as TAM are concerned) on-line diagnostics have been removed. Thanks to all these effort the additional computing burden has been contained to around twice the non-linear integration (more or less, depending on the application).

3.3.3 Validation

Due to the numerous caveats mentioned in section 3.3.2 validation of TAM is a crucial aspect. Fortunately there are several numerical tests that can be performed to ensure this validity, moreover these tests can be applied at module level.

Numerical validation

In NEMOTAM, modules include both the tangent and adjoint parts, as well as the *validation interface*, for numerical verification of the TAM source codes. The *adjoint test* would always be present, while the *tangent test* could be optional and reserved to specific and problematic routines.

In the following $\mathcal{M}(\mathbf{x})$ will be the full non-linear NEMO model with initial state vector $\mathbf{x}$, $\mathbf{L}(\mathbf{x}) \equiv (\partial\mathcal{M}/\partial\mathbf{x})(\mathbf{x})$ its tangent linear model (possibly simplified), and $\mathbf{L}^T(\mathbf{x})$ the adjoint code of $\mathbf{L}$.

Adjoint validation

The adjoint part is actually relatively easy to check. Indeed, by definition of the adjoint one gets :

$$(\mathbf{L}\delta\mathbf{x}, \delta\mathbf{y}) = \langle \delta\mathbf{x}, \mathbf{L}^T\delta\mathbf{y}\rangle \tag{3.6}$$

where $\langle ., . \rangle$ and $(., .)$ denote the appropriate dot products. Equation 3.6 being exact, the relative error between the two computed scalar products must be close to zero barring rounding errors. In NEMOTAM, the actual test performed is

$$(\mathbf{L}\delta\mathbf{x})^T\mathbf{W}\delta\mathbf{y} = \delta\mathbf{x}^T\mathbf{L}^T\mathbf{W}\delta\mathbf{y} \tag{3.7}$$

where $\delta\mathbf{x}$ is a random vector, $\mathbf{W}$ is a diagonal matrix of scale factors and $\delta\mathbf{y} = \mathbf{L}\delta\mathbf{x}$. This test only ensures that the adjoint is indeed the adjoint of the tangent linear, which in turn has to be validated.

Tangent validation

Validation of tangent modules are more tricky since there is generally no affordable exact tests available. A classical method of testing a numerical tangent linear model $\mathbf{L}$ is to compare the evolution of a perturbation by $\mathbf{L}$ with the difference of difference of two evolutions, with and without the perturbation, by the full nonlinear model $\mathcal{M}$.
Considering a perturbation vector $\gamma\delta x$, where γ is a scale parameter, we have for a fixed perturbation :

$$\mathcal{M}(\mathbf{x} + \gamma\,\delta\mathbf{x}) = \mathcal{M}(\mathbf{x}) + \gamma\,\mathbf{L}(\mathbf{x})\,\delta\mathbf{x} + O(\gamma^2) \tag{3.8}$$

$\mathcal{N}(\mathbf{x}_0, \gamma\delta\mathbf{x}, t_0, t)$ denotes the non linear evolution of a perturbation and we use the simplified notation $\mathcal{N}(\gamma\delta\mathbf{x}, t)$,

$$\mathcal{N}(\gamma\delta\mathbf{x}, t) = \mathcal{M}(\mathbf{x} + \gamma\delta\mathbf{x}, t) - \mathcal{M}(\mathbf{x}, t) \tag{3.9}$$

The linearization error $\mathcal{E}(\gamma\delta x, t)$ is defined by :

$$\mathcal{E}(\gamma\delta\mathbf{x}, t) \quad = \quad \mathcal{N}(\gamma\delta\mathbf{x}, t) - \gamma\mathbf{L}(t)\delta\mathbf{x} \tag{3.10}$$

$\mathcal{E}(\gamma\delta\mathbf{x}, t)$ behaves in $O(\gamma^2)$. The first order accuracy index ϵ_γ may be defined as :

$$\epsilon_\gamma = \frac{\|\mathcal{E}(\gamma\delta\mathbf{x}, t)\|}{\|\mathbf{L}(t)\gamma\delta\mathbf{x}\|} \tag{3.11}$$

ϵ_γ tends to zero when γ tends to zero which validates $\mathbf{L}(t)$. Moreover when γ is small enough, $\mathcal{N}$ enters a linear regime and ϵ_γ converges with a rate γ.

One can also define the second order accuracy index :

$$\varepsilon_\gamma(t) = \frac{\|\mathcal{E}(\gamma\delta\mathbf{x}, t)\|}{\gamma^2} \tag{3.12}$$

which tends to a constant when γ tends to 0.

Table 3.2 shows an example of such tests on a single routine. It illustrates nicely the expected behaviour of ϵ_γ which gains one digit in precision when γ is divided by ten.

Routine (**L**)	γ	ϵ_γ	ε_γ
bn2_tan	1E+00	.999961862090	7.29241788399E-07
bn2_tan	1E-01	.999995878199	7.29211958733E-07
bn2_tan	1E-02	.999999584740	7.29209298692E-07
bn2_tan	1E-03	.999999958442	7.29209023679E-07
bn2_tan	1E-04	.999999995846	7.29208626213E-07

Table 3.2 – *Tangent validity tests for the bn2 routine (computation of the Brünt-Väisälä frequency)*

Estimation of the approximations error

As mentioned above, when differentiating realistic models, approximations have to be done. To estimate the effect of these approximations on the numerical tangent linear model $\mathbf{L}$, we must first estimate the truncated part, $O((\gamma\delta\mathbf{x})^2)$, of the Taylor expansion (see equation 3.8). In order to do this, following Lawless et al. (2003), one can write the Taylor expansion of

$$\mathcal{E}(\gamma\delta\mathbf{x}, t) = \frac{\gamma^2}{2}\partial^2\mathcal{M}\delta\mathbf{x}^2 + \frac{\gamma^3}{6}\partial^3\mathcal{M}\delta\mathbf{x}^3 + O(\gamma^4) \tag{3.13}$$

On another hand, from two nonlinear perturbations :

$$\begin{aligned} \mathcal{N}(\delta\mathbf{x}, t) &= \mathcal{M}(t_0, t)\delta\mathbf{x} + O((\delta\mathbf{x})^2, t) \\ \mathcal{N}(\gamma\delta\mathbf{x}, t) &= \mathcal{M}(t_0, t)\gamma\delta\mathbf{x} + O((\gamma\delta\mathbf{x})^2, t) \end{aligned} \tag{3.14}$$

one can compute

$$\mathrm{E}(\gamma\delta\mathbf{x}, t) = \frac{\mathcal{N}(\gamma\delta\mathbf{x}, t) - \gamma\mathcal{N}(\delta\mathbf{x}, t)}{\gamma^2 - \gamma} \tag{3.15}$$

whose Taylor expansion reads

$$\mathrm{E}(\gamma\delta\mathbf{x}, t) = \frac{\gamma^2}{2}\partial^2\mathcal{M}\delta\mathbf{x}^2 + \frac{\gamma^3 + \gamma^4}{6}\partial^3\mathcal{M}\delta\mathbf{x}^3 + O(\gamma^4) \tag{3.16}$$

For small values of γ and $\delta\mathbf{x}$, we can compare E and $\mathcal{E}$. That way we built-up the estimator of the numerical tangent linear model error as follows :

$$\hat{\mathcal{E}} = 100\left(1 - \frac{\|\mathrm{E}\|}{\|\mathcal{E}\|}\right) \tag{3.17}$$

Moreover, in NEMO, the vast majority of the nonlinearities are quadratic ones, meaning the third order and above derivatives vanish from the Taylor expansion and we get $\mathrm{E} = \mathcal{E}$.

This diagnostic is very valuable when comparing different simplifications made to the tangent linear model. Table 3.3 shows $\hat{\mathcal{E}}$ for two configurations of NEMO : an academic test case that is fully differentiable (SEABASS) and 2°-resolution global realistic configuration (ORCA2). This allows to measure the effect of the two approximations mentioned in section 3.3.2 and 3.3.2

Configurations	$\hat{\mathcal{E}}$
SEABASS no simplification	negligible
SEABASS interpolated checkpoint (1 per day)	5%
ORCA2 simplified vertical mixing	2%
ORCA2 idem + interpolated checkpoint (1 per day)	4-5%

Table 3.3 – *Approximation error in the tangent linear model for different configurations*

Validity of the tangent linear model

Once the numerical code has been validated using the consistency checks described in the previous section, another question remains : is the TAM an appropriate tool for the particular problem I am interested in ? Obviously the answer to this question depends on the type of application TAM is intended for. Most of the time however it relies on the ability of the tangent linear model to behave similarly to the non linear model. This in turn depends on the size of the perturbation $\delta\mathbf{x}$, on the length of the considered time-window, and obviously on the nonlinearity of the full nonlinear model $\mathcal{M}$.

In NEMOTAM two indicators can be used. Both are based on the comparison of the behaviour of two opposite perturbations. Indeed, looking on the Taylor expansion of the sum of the non linear evolutions of these perturbations, the linear part vanishes and only remains the non-linear part

$$\mathcal{N}(\gamma\delta\mathbf{x}, t) + \mathcal{N}(-\gamma\delta\mathbf{x}, t) = \frac{2\gamma^2}{2}\delta\mathbf{x}^T\mathcal{H}\delta\mathbf{x} + O(\gamma^4) \tag{3.18}$$

Neglecting the higher order terms, equation 3.18 gives a good estimate of (twice) the second order term of the full evolution of a perturbation. This leads to define the relative non-linear ratio

$$\Theta(\gamma\delta\mathbf{x}, t) = \frac{\|\mathcal{N}(\gamma\delta\mathbf{x}, t) + \mathcal{N}(-\gamma\delta\mathbf{x}, t)\|}{2\|\mathbf{L}(t)\gamma\delta\mathbf{x}\|} \tag{3.19}$$

When Θ is small, this means that the linear regime is dominant. As t becomes larger, the non linearities will develop, Θ will increase showing that the tangent linear is not behaving anymore as a good approximation of the full non linear model.

The use of the tangent linear is time-limited by the diffusion rate of nonlinearity. This spread can be monitored with the similarity indice $l(\gamma\delta\mathbf{x}, t)$ defined as :

$$l(\gamma\delta\mathbf{x}, t) = -\frac{< \mathcal{N}(\gamma\delta\mathbf{x}, t), \mathcal{N}(-\gamma\delta\mathbf{x}, t) >}{\|\mathcal{N}(\gamma\delta\mathbf{x}, t)\|\|\mathcal{N}(-\gamma\delta\mathbf{x}, t)\|} \tag{3.20}$$

The residual of the Taylor expansion at first order cannot be neglected when $l(\gamma\delta\mathbf{x}, t)$ is below 0.7. (see (Gilmour, 2001) and (Buizza, 1995)). Thus, the tangent linear assumption is not valid.

3.4 NEMOVAR : Système d'assimilation variationnelle de données pour NEMO

3.4.1 Introduction

NEMOVAR est un framework logiciel d'assimilation variationnelle de données construit autour du framework de modélisation océanique NEMO, et plus particulièrement sur sa composante de modélisation de la dynamique océanique OPA. Son développement est basé sur le système OPAVAR, décrit par Weaver et al. (2005). Originellement, OPAVAR permettait de réaliser des expériences d'AD, basées sur les algorithmes du 3D-FGAT ou du 4D VAR incrémental, puisqu'il incluait les modèles linéaire tangent et adjoint de NEMO. OPAVAR comportait également la modélisation des matrices de covariances d'erreur.

Cependant, OPAVAR présentait plusieurs inconvénients :

- Il était basé sur une version obsolète de NEMO
- Son application à des configurations pour différentes résolutions horizontales n'était pas triviale
- La parallélisation à mémoire distribuée n'était pas applicable

Pour ses raisons, il a été décidé de développer NEMOVAR. Plusieurs contributeurs sont aujourd'hui réunis dans le consortium NEMOVAR : CERFACS, ECMWF, Met Office, et l'INRIA.

Dans NEMOVAR, les principaux développements réalisés ont été les suivants :

- Un opérateur d'observation OBS, gérant le contrôle qualité des observations en temps réel. Cette composante est aujourd'hui intégrée officiellement dans NEMO
- L'actualisation des modèles linéaire tangent et adjoint (TAM) parallélisés pour des versions récentes de NEMO, aujourd'hui intégrés officiellement comme composante de NEMO,
- Évolution de la modélisation des matrices de covariances d'erreur et amélioration des algorithmes de minimisation proprement dit.

Nous allons présenter succinctement ici les principes généraux de modélisation des matrices de covariances d'erreur ainsi que les algorithmes de minimisation présents dans NEMOVAR.

3.4.2 Modélisation des matrices de covariances d'erreurs

Matrices de covariances d'erreurs d'ébauche

Comme nous l'avons évoqué au §2.3.1, pour des systèmes de grandes tailles tels que ceux rencontrés en océanographie, la matrice de covariance d'erreur d'ébauche ne peut être explicitement stockée et manipulée, d'autant plus que les algorithmes d'AD variationnels nécessitent son inversion.

Il est cependant à noter que, dans le déroulement numérique des méthodes variationnelles, ce n'est pas la matrice en elle-même qui nous intéresse, mais son produit avec un vecteur $\mathbf{x}$ défini. Si l'on formule $\mathbf{B}$ comme un opérateur de covariance à partir d'une suite de m opérateurs linéaires (notés $\mathbf{A}$) appliqués au vecteur $\mathbf{x}$, on peut alors calculer le produit $\mathbf{Bx}$ suivant :

$$\mathbf{Bx} = \mathbf{A}_m \mathbf{A}_{m-1} ... \mathbf{A}_1 \mathbf{x} \tag{3.21}$$

Cependant, dans ce cas, rien ne garantit que $\mathbf{B}$ soit symétrique définie positive. On peut alors décomposer $\mathbf{B}$ ainsi :

$$\mathbf{B} = \mathbf{B}_{1/2} \mathbf{B}_{1/2}^T \tag{3.22}$$

où $\mathbf{B}_{1/2}$ peut être assimilé à un opérateur « racine carrée » de $\mathbf{B}$. $\mathbf{B}_{1/2}^T$ peut être considéré comme la forme matricielle de l'opérateur adjoint de $\mathbf{B}_{1/2}$.

$\mathbf{B}$ doit représenter les covariances spatiales de chaque variable mais également les covariances entre les variables. En se rappelant de la définition de B comme une suite d'opérateurs linéaires, il peut être intéressant de décomposer ce problème multivarié en une succession de problèmes univariés indépendants afin de réduire la dimension desdits opérateurs linéaires.

Afin de décomposer le problème multivarié en plusieurs problèmes monovariés, Derber et Bouttier (1999) proposent alors de séparer chaque variable en une partie dite *équilibrée* et une partie dite *non-équilibrée*, à l'exception d'une qui servira de point de départ à la définition des parties équilibrées des autres variables. Chaque partie équilibrée des autres variables peut alors être exprimée en fonction de la variable totale grâce à des équations d'équilibre dynamique et de conservation de masse d'eau (voir Ricci et al. (2005) et Weaver et al. (2005)).

En partant de la température T, chacune des autres variables peut être décrite comme la somme de sa composante non-équilibrée (indicée U, pour *unbalanced*) et de sa composante équilibrée (indicée B, pour *balanced*), correspondant à une transformation K_{xy} d'une variable y en une variable x. On peut alors écrire les relations suivantes :

$$\begin{cases} T = T \\ S = S_B + S_U = K_{ST}[T] + S_U \\ \eta = \eta_B + \eta_U = K_{\eta\rho}[\rho] + \eta_U \\ u = u_B + u_U = K_{u\rho}[p] + u_U \\ v = v_B + v_U = K_{v\rho}[p] + v_U \end{cases} \tag{3.23}$$

avec

$$\begin{cases} \rho = K_{\rho T}[T] + K_{\rho S}[S] \\ p = K_{p\rho}[\rho] + K_{p\eta}[\eta] \end{cases} \tag{3.24}$$

En linéarisant ces opérateurs de transformation, on définit l'opérateur d'équilibre $\mathbf{K}$ tel que :

$$\mathbf{B} = \mathbf{K}\mathbf{B}_U\mathbf{K}^T \tag{3.25}$$

où $\mathbf{B}_U$ est la matrice diagonale par bloc des covariances spatiales univariées.

Maintenant que nous avons réduit notre problème multivarié en plusieurs problèmes monovariés, la taille de notre problème est effectivement réduite. Cependant, reste à estimer les covariances spatiales monovariées, incluses dans la matrice $\mathbf{B}_U$.

Par définition, les covariances spatiales sont le produit des variances et des corrélations. Ainsi, on peut réécrire $\mathbf{B}$:

$$\mathbf{K}\mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2}\mathbf{K}^T \tag{3.26}$$

où $\mathbf{D}$ est la matrice diagonale des variances et $\mathbf{C}$ la matrice diagonale par bloc des corrélations spatiales.

La spécification des variances et des longueurs de corrélation est alors laissée au choix de l'utilisateur. Les corrélations elle-mêmes sont modélisées par un opérateur de diffusion. Cette méthode est décrite exhaustivement dans Weaver et Courtier (2001), Weaver et al. (2005) et Mirouze et Weaver (2010).

Matrice de covariances d'erreurs d'observations

Dans la version de NEMOVAR utilisée dans ce travail, la matrice de covariances d'erreur d'observation est diagonale. Les corrélations spatiales des erreurs d'observation sont négligées, ainsi que les covariances d'erreurs entre les différents types possibles d'observations. Cette hypothèse a longtemps été faite dans les systèmes d'AD et le reste aujourd'hui pour beaucoup d'applications opérationnelles. Seules les erreurs (*i.e.* écarts-types de l'erreur de mesure) pour chaque source d'observation sont à spécifier par l'utilisateur.

3.4.3 Algorithme de minimisation

Nous avons exposé au chapitre précédent la forme de la fonction coût effectivement minimisée dans NEMOVAR, décrite dans l'équation 2.29. Nous la réécrivons ici, simplifiée :

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x_0}^T\mathbf{B}^{-1}\delta\mathbf{x}_0 + \frac{1}{2}(\mathbf{G}\delta\mathbf{x} - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{G}\delta\mathbf{x} - \mathbf{d}) \tag{3.27}$$

avec :

$$\mathbf{G} = \begin{bmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_i\mathbf{M}_i...\mathbf{M}_1 \\ \vdots \\ \mathbf{H}_N\mathbf{M}_N...\mathbf{M}_1 \end{bmatrix} \tag{3.28}$$

Nous avons vu plus haut que $\mathbf{B}$ pouvait s'écrire comme le produit de deux matrices « racine carrée » $\mathbf{B} = \mathbf{B}_{1/2}^T\mathbf{B}_{1/2}$. Derber et Bouttier (1999) proposent alors le changement de variable suivant :

$$\mathbf{v} = \mathbf{B}_{1/2}^{-1}\delta\mathbf{x} \tag{3.29}$$

L'équation 3.27 peut s'écrire alors :

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}(\mathbf{G}\mathbf{B}_{1/2}\mathbf{v} - d)^T\mathbf{R}^{-1}(\mathbf{G}\mathbf{B}_{1/2}\mathbf{v} - d) \tag{3.30}$$

Le gradient de cette fonction coût s'écrit alors :

$$\nabla_v J = \mathbf{v} + \mathbf{B}_{1/2}^T\mathbf{G}^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{G}_{1/2}\mathbf{v} - d) \tag{3.31}$$

Notre problème consiste à trouver $\mathbf{v}_a$ vérifiant $\mathbf{A}\mathbf{v}_a = \mathbf{b}$ avec :

$$\begin{cases} \mathbf{A} = \mathbf{I}_n + \mathbf{B}_{1/2}^T\mathbf{G}^T\mathbf{R}^{-1}\mathbf{G}\mathbf{B}_{1/2} \\ \mathbf{b} = \mathbf{B}_{1/2}^T\mathbf{G}^T\mathbf{R}^{-1}\mathbf{d} \end{cases} \tag{3.32}$$

Ce changement de variable peut être considéré comme un préconditionnement du problème initial.

L'algorithme du gradient conjugué, présenté au chapitre précédent, peut alors être appliqué à ce problème, avec un coût numérique dépendant de la dimension de $\mathbf{A}$. En pratique, deux variantes de l'algorithme du gradient conjugué sont utilisées :

- Le premier utilise un préconditionneur à mémoire limitée. Pour plus de détails, voir Tshimanga et al. (2008). Dans ce cas, un autre préconditionnement est utilisé.
- Le deuxième se base sur une ré-orthogonalisation des directions de descente. Pour plus de détails, voir Fisher (1998).

Le principal objectif de ces deux méthodes est d'accélérer la convergence de l'algorithme pour diminuer le nombre d'itérations nécessaire à la convergence du gradient conjugué.

Chapitre 4

La configuration SEABASS de NEMO : un modèle académique pour l'assimilation de données

4.1 Présentation de SEABASS

4.1.1 Motivations

Afin d'analyser l'impact de la résolution horizontale et donc de la non-linéarité sur le comportement des méthodes d'assimilation variationnelle de données, il est important de choisir une configuration de bassin océanique appropriée à cette étude, basée sur NEMO. Les premières questions qui se posent quant à ce choix portent sur différents points :

- Modèle réaliste ou idéalisé
- Modèle d'océan global ou régional
- Coût numérique de l'intégration du modèle
- Facilité de prise en main et de maintenance
- Disponibilité de la même configuration à différentes résolutions horizontales

Nous présentons ici les arguments qui ont conduit au choix de la configuration SEABASS [1], utilisée dans Cosme et al. (2010) pour mener notre étude à bien.

Tout d'abord, **d'un point de vue technique**, un système d'AD couplé à un framework de modélisation océanique est techniquement et scientifiquement complexe. Considérer une configuration réaliste rajoute encore une complexité non-négligeable à ce système, sur plusieurs plans :

- Comme décrit dans la Section 3.1, une configuration globale ou régionale réaliste suppose de prendre en compte des aspects variés et non-triviaux : bathymétrie, forçages atmosphériques, périodicité des conditions aux limites ou frontières ouvertes, par exemple. Ceci se traduit par un nombre non-négligeable d'entrées/sorties, ainsi que par une complexité accrue du coeur numérique mis en oeuvre.
- Ces éléments de complexité supplémentaires impactent directement le système d'AD. Premièrement, ils se reportent directement sur les modèles linéaire tangent et adjoint, comme

1. Acronyme de *SEA Box for ASSimilation*

évoqué dans la Section 3.3. A ceci se rajoutent les algorithmes non-différentiables qui induisent des approximations numériques dans NEMOTAM, issus du traitement des forçages atmosphériques et des différents schémas numériques requis dans une configuration réaliste (*e.g.* schémas de fermeture turbulente). En ce qui concerne NEMOVAR, la spécification des matrices de covariances d'erreurs doit aussi prendre en compte les frontières de notre domaine.
- Une telle configuration, à résolution numérique *eddy-permitting* ou *eddy-resolving*, peut potentiellement présenter un coût numérique élevé, coût démultiplié par le système d'AD [2].
- Changer la résolution horizontale d'une configuration réaliste n'est pas trivial. En effet, les différentes paramétrisations du modèle doivent être finement choisis pour garantir une bonne restitution de l'écoulement observé. À ceci se rajoute le traitement spécifique à chaque résolution des fichiers d'entrées du modèle.

Ensuite, **d'un point de vue scientifique**, notre étude vise à mettre en évidence l'influence des phénomènes méso-échelles sur les performances des méthodes variationnelles d'AD, dérivant du 4DVAR. Sur une configuration réaliste, d'autres aspects du modèle tels que la bathymétrie, en contrôlant fortement l'écoulement, peuvent potentiellement atténuer l'effet des non-linéarités sur les sorties du système d'assimilation. Ainsi, choisir une configuration idéalisée plutôt qu'une configuration réaliste permet de s'affranchir de ce risque, quitte à « sur-évaluer » l'impact des non-linéarités sur le système d'AD.

Enfin, l'assimilation variationnelle de données a été largement explorée en météorologie, tant du point de vue méthodologique (citons ici les études théoriques de Le Dimet et Talagrand (1986), Derber (1987) et des articles de revues) que du point de vue opérationnel (voir notamment les descriptions et études des systèmes opérationnels d'AD variationnels par Gauthier et al. (1999), Lorenc et al. (2000) ou l'étude complète de l'algorithme du 4DVAR incrémental implémenté à l'ECMWF faite dans Trémolet (2007a)). Il n'en est pas de même en océanographie. Les recherches concernant cette famille d'algorithmes d'AD pour ce domaine d'application sont plus récentes et moins nombreuses :
- Quelques études méthodologiques menées sur des modèles QG sont présentées par Blum et al. (1998), Luong et al. (1998) ou Rémy et al. (2002).
- Dans le cadre de NEMOVAR, les avancées sur la modélisation de la matrice de covariance d'ébauche sont exposées dans Weaver et al. (2003) et Weaver et al. (2005) et ses applications (voir notamment Ricci et al. (2005), Daget et al. (2009b) ou Mirouze et Weaver (2010)).
- D'autres études réalistes ont été réalisées sur le modèle ROMS, par Di Lorenzo et al. (2007), Powell et al. (2008) ou encore Moore et al. (2011). D'autres systèmes d'AD sont basées sur le modèle ECCO comme Stammer et al. (2002). Pour NEMO, nous pouvons signaler les études sur l'atlantique nord de Ferron (2011) et la méditerranée par Dobricic et Pinardi (2008).
- En océanographie opérationnelle, l'algorithme 3DFGAT est maintenant utilisé à l'ECMWF (voir Balmaseda et al. (2013)) ou au Met-Office UK.

Devant le peu de recul relatif sur l'utilisation des algorithmes dérivant du 4DVAR, qui plus est à haute résolution avec NEMO (l'étude de Ferron (2011) étant réalisée sur une configuration de l'atlantique nord au 1/3°), nous avons choisi de nous restreindre à une application académique,

2. Une configuration *régionale* aurait pu répondre à cette contrainte. Cependant, un tel choix porte les autres contraintes d'une configuration réaliste

mettant en évidence uniquement les aspects que nous voulons étudier.

Ainsi, pour toutes ces raisons, notre choix s'est porté sur une configuration de bassin océanique idéalisé, SEABASS, présentant les caractéristiques générales suivantes :
- Prépondérances des phénomènes non-linéaires associés à la turbulence
- Coût numérique modéré
- Changement de résolution horizontale aisé
- Géométrie du domaine la plus simple possible, pour limiter au maximum les traitements potentiellement complexes des conditions au limites, que soit au niveau algorithmique ou des entrées/sorties
- Configuration déjà éprouvée, simple à prendre en main et à maintenir
- Schémas numériques différentiables pour éviter les approximations dans NEMOTAM

4.1.2 Descriptions physique et numérique de SEABASS

SEABASS utilise les schémas numériques implémentés dans NEMO et décrits dans la Section 3.3. Nous présentons ici les caractéristiques propres de cette configuration, à savoir la description du domaine spatial, des conditions aux limites et des paramétrisations dépendantes de la résolution.

Le domaine horizontal Il est rectangulaire sur une sphère, s'étendant de 24°N à 44°N en latitude (soit sur environ 2226 km) et de 60°O à 30°O en longitude (soit une longueur comprise entre 3000km et 3500km, selon la latitude considérée). Nous avons étudié cette configuration pour plusieurs résolutions horizontales : le quart, le douzième et le vingt-quatrième de degré. Le tableau 4.1 indique le nombre de points de grille sur chaque axe horizontal et sur le domaine tout entier selon la résolution considérée ($N_{level} = 11$, voir ci-dessous).

Résolution	N_{lon}	N_{lat}	Total ($N_{lat}N_{lon}N_{level}$)
$1/4°$	121	81	107 811
$1/12°$	361	241	957 011
$1/24°$	721	481	3 814 811

Table 4.1 – *Nombre de points de grille en longitude, latitude et au total en fonction de la résolution horizontale*

La configuration au 1/4° sera parfois appelée *SEABASS4*, au 1/12°, *SEABASS12*, etc.

Le domaine vertical Il s'étend sur environ 4500 m. Il est scindé en onze couches dont les profondeurs (selon points T et W) sont notées dans le tableau 4.2 et ce, quelle que soit la résolution. Ce choix a été fait pour plusieurs raisons :
- Techniquement, cela limite l'accroissement du coût de calcul lorsque l'on augmente la résolution horizontale
- L'augmentation du nombre de niveaux verticaux aurait introduit des modes baroclines supplémentaires, dont les échelles horizontales caractéristiques, de l'ordre du kilomètre voire moins, n'auraient pas été résolues.

– Ce nombre de niveaux suffit pour étudier ce qui nous intéresse ici, à savoir la propagation verticale de l'information, aspect très important dans notre contexte d'assimilation de données altimétriques. Cependant, pour étendre les conclusions des travaux présentés ici à des résolutions verticales plus réalistes, il faudra plus tard traiter des discrétisations comportant plusieurs dizaines de couches.

Niveau	Grille T	Grille W
0	152.0232	0
1	459.9266	305.244
2	775.0345	616.392
3	1100.861	936.329
4	1442.611	1269.334
5	1807.974	1621.735
6	2208.243	2002.858
7	2659.891	2426.359
8	3186.681	2912.054
9	3822.335	3488.320
10	4613.589	4195.000

Table 4.2 – *Profondeurs (en m) des différents niveaux verticaux selon les deux grilles associées aux différentes variables*

Les conditions aux limites

Dans cette section, nous décrirons les conditions aux limites de la configuration SEABASS. Nous noterons Γ_b la frontière correspondant au fond du bassin, Γ_s la surface et Γ_l la réunion des frontières nord, est, sud et ouest du bassin. ∂_n désignera la dérivée normale au bord considéré. $\vec{n}$ désigne la normale sortante à Γ et $\vec{\tau}$ son vecteur tangent.

Sur Γ_b Au fond, nous supposons nulle la vitesse verticale (condition d'imperméabilité sur la frontière, $\vec{U}.\vec{n} = 0$), et la vitesse horizontale U_h est soumise à une condition de friction linéaire :

$$\partial_n U_h = r U_h \tag{4.1}$$

où r est le coefficient de friction. Dans notre cas, $r = 3.955.10^{-4}\ m^{-1}$.

Sur Γ_l Aux frontières latérales du domaine, la vitesse horizontale vérifie une condition de glissement libre (*free slip*), qui revient à annuler la dérivée normale de la vitesse tangentielle au bord $\partial_n(U_h.\vec{\tau}) = 0$. Elle vérifie également une condition de flux nul (condition d'imperméabilité), $\vec{U_h}.\vec{n} = 0$.

Sur Γ_s La vitesse verticale est nulle en surface et nous « forçons » la surface avec un vent stationnaire, constant en longitude, sinusoïdal en latitude. Si l'on considère que le domaine s'étend sur $(0, l)$ en latitude, la tension de vent est donnée en fonction de la latitude y par la formule

suivante (illustrée dans le Figure 4.1) :

$$\begin{aligned} \tau_x(x,y) &= -0.1\cos\left(\frac{2\pi y}{l}\right) \\ \tau_y(x,y) &= 0 \end{aligned} \tag{4.2}$$

Il est à noter que ce forçage, le seul dans notre configuration, est symétrique par rapport à la latitude médiane de notre domaine horizontal.

Sur $\Gamma_l \cup \Gamma_b$ Aux interfaces océan-continent, les flux de sel et de chaleur sont supposés nuls.

$$\partial_n \theta = \partial_n S = 0 \tag{4.3}$$

Nous ajouterons ici qu'initialement, nous spécifions un profil vertical de température, montré dans la Figure 4.2. La salinité est initialement constante sur tout le domaine et le reste au cours d'une simulation, aucun forçage de salinité n'étant appliqué.

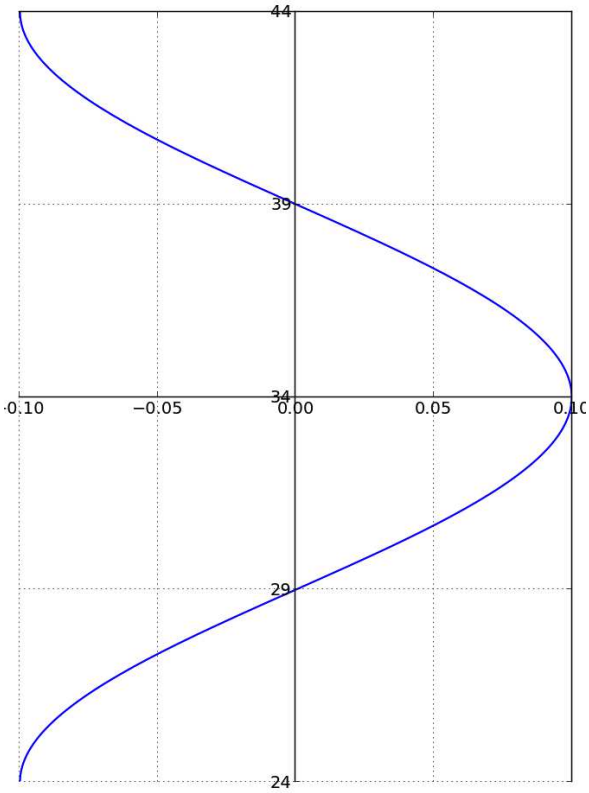

Figure 4.1 – *Tension zonale de vent (en abscisse, en $N.m^{-2}$) en fonction de la latitude (en ordonnée)*

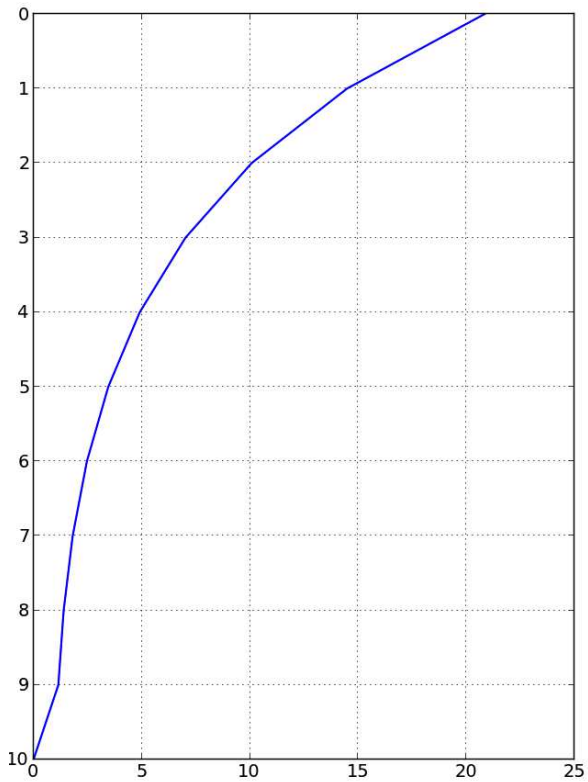

Figure 4.2 – *Profil initial de température (en abscisse, en °C) en fonction des numéros de niveaux (en ordonnée)*

Paramétrisations Pour répondre aux contraintes de stabilité (CFL) des schémas numériques (si Δx diminue, Δt doit diminuer également), le pas de temps d'intégration est dépendant de la résolution. Afin de mieux représenter les phénomènes non-linéaires méso-échelles, la diffusion des traceurs et la viscosité pour les équations dynamiques sont modélisées par un opérateur biharmonique (choix expliqué dans la Section 3.1) dont le coefficient (identique pour les deux)

est dépendant de la résolution. Le tableau 4.3 récapitule les valeurs de ces paramètres en fonction de la résolution.

Résolution	Pas de temps (en s)	Viscosité et coefficient de diffusion (en $m^4.s^{-1}$)
$1/4°$	900	-8.10^{10}
$1/12°$	300	$-0.89.10^{10}$

Table 4.3 – *Valeurs des paramètres dépendant de la résolution.*

Coût numérique Dans le Tableau 4.4 sont présentés quelques indicateurs du coût numérique d'une simulation de SEABASS4 et de SEABASS12. Les paramètres limitants de cette configuration sont en premier lieu le temps CPU et le coût de stockage des sorties du modèle. La mémoire vive utilisée reste raisonnable. Comme nous le verrons par la suite, ces coûts sont démultipliés par le système d'AD.

Résolution	Processeurs	Mémoire vive (en Mb)	Temps CPU (en min)	Stockage (en Gb)
$1/4°$	16	218	21	7Go
$1/12°$	64	637	30	10Go

Table 4.4 – *Coût numérique de SEABASS, pour une simulation annuelle, avec des sorties instantanées journalières, pour SEABASS4 et SEABASS12, sur un supercalculateur composé de 16 noeuds de calcul bi-socket comprenant chacun 2 processeurs Nehalem (Xeon X5550) et 24 gigas de RAM, connectés par un réseau de type Infiniband*

Résolution	D_h^U/D_h^T	τ	Δx	Δt	r_l	r_b	D_v^T	D_v^U
1/4°	-8.10^{10}	0.1	27	900	4.10^{-4}	$3.955.10^{-4}$	$1.2.10^{-5}$	$1.2.10^{-4}$
1/12°	$-0.89.10^{10}$	0.1	8	300	4.10^{-4}	$3.955.10^{-4}$	$1.2.10^{-5}$	$1.2.10^{-4}$

Table 4.5 – *Résumé des paramètres numériques pour SEABASS à différentes résolutions. τ est la tension de vent, r_l et r_b respectivement les frictions latérales et de fond (en m^{-1}), D_h^U et D_h^T respectivement la viscosité et le coefficient de diffusion sur les traceurs (en $m^4.s^{-1}$), en xy, et D_v^U et D_v^T (en $m^2.s^{-1}$) en z*

4.2 Description de la circulation océanique simulée par SEABASS

Afin de bien paramétriser l'algorithme d'assimilation et d'en comprendre son comportement vis-à-vis du modèle étudié, il est essentiel d'analyser les caractéristiques de la circulation océanique présentée par SEABASS.

Le modèle étant initialisé avec un champ de vitesse nul, il est nécessaire de le laisser évoluer pendant une durée suffisamment longue pour qu'un régime permanent s'établisse, après une phase transitoire (appelée phase de *spinup*).

Sur la Figure 4.3, l'énergie cinétique turbulente augmente fortement au tout début de l'intégration libre puis décroît progressivement, jusqu'à se stabiliser (au plus tard pour la couche de surface) aux alentours de l'année 80 de simulation autour d'un équilibre « statistiquement

stationnaire ». La latitude de décollement du jet zonal se stabilise aux alentours de l'année 30 (*cf.* Figure 4.4).

C'est pourquoi tous les résultats présentés ici sont considérés après 80 ans de simulation pour les deux résolutions étudiées (1/4° et 1/12°).

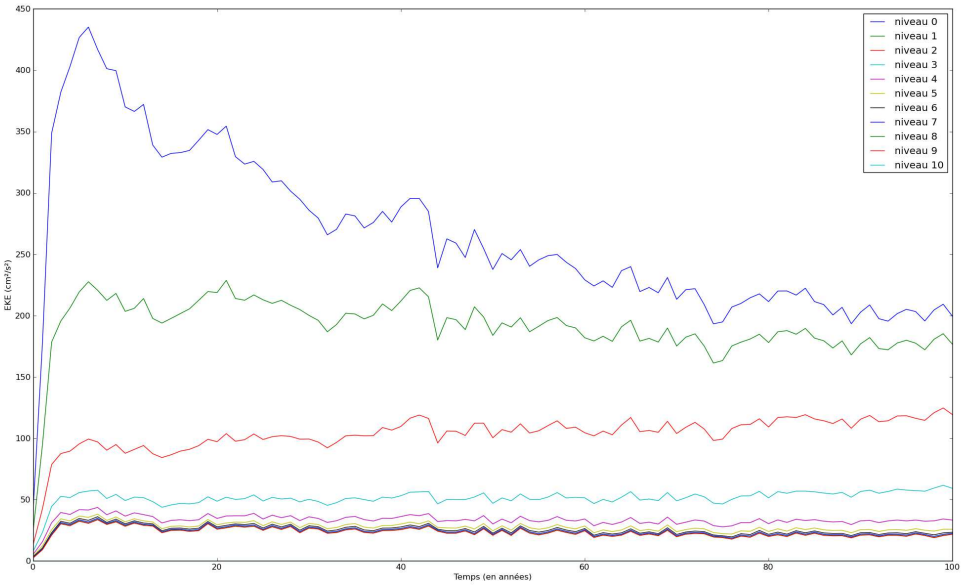

Figure 4.3 – *Évolution temporelle (en année) de la moyenne spatiale de l'énergie cinétique turbulente (en $cm^2.s^{-2}$), sur chaque niveau horizontal, pour 100 ans de simulation partant d'un état initial au repos, dans SEABASS4*

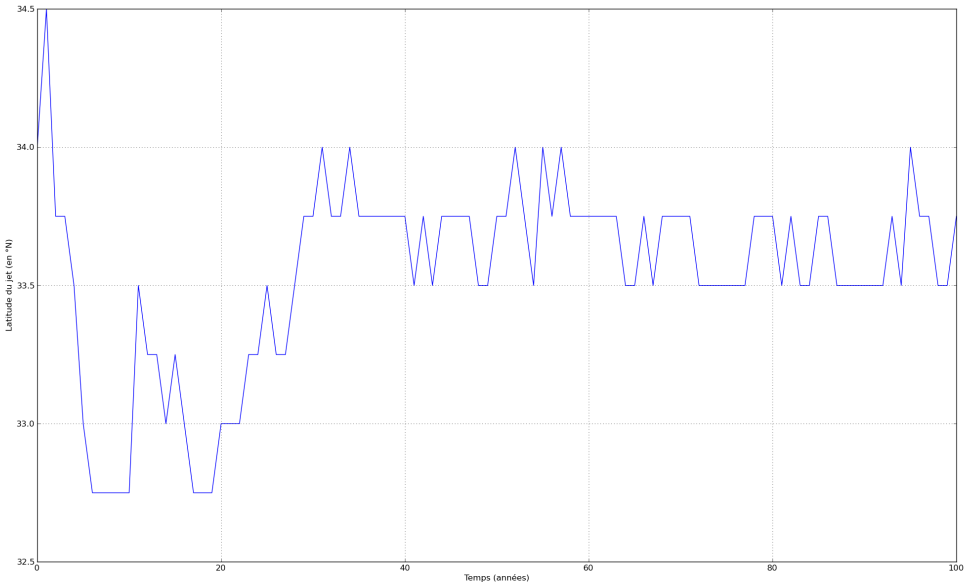

Figure 4.4 – *Évolution temporelle de la latitude de décollement du jet zonal pour 100 ans de simulation partant d'un état initial au repos, dans SEABASS4*

4.2.1 Comportement général

La circulation dans le bassin est de type *double-gyre*. Elle se caractérise par deux cellules de recirculation, une au nord et une au sud, dans le sens antihoraire pour la première, dans le sens horaire pour la seconde. À la confluence de ces deux cellules, un jet se forme, où la vitesse zonale est maximale. Comme nous le verrons par la suite, la variabilité naturelle de SEABASS est dépendante de la résolution horizontale considérée.

Sur les Figures 4.5 et 4.6, nous pouvons voir 9 champs de SSH, à intervalles de 5 jours, pour SEABASS4 et SEABASS12, respectivement. Tout d'abord, nous obervons que le jet décolle bien quasiment au centre du domaine, à 34°N, au 1/4°, légèrement décalé au sud de deux points de grille en moyenne. En revanche, au 1/12°, le décollement se produit plus au sud du domaine (30°N). Un résultat analogue pour une configuration similaire, de type double-gyre sur une grille tournée, est remarqué dans Lévy et al. (2010). Il faut noter que les équations ne sont pas symétriques : nous ne faisons pas d'approximation *beta-plan*, notre domaine étant placé sur une sphère aux latitudes moyennes. Nous discuterons ce point plus avant dans le chapitre 7.

Ensuite, quelle que soit la résolution, des structures tourbillonnaires s'échappent régulièrement de l'extrémité est du jet et recirculent vers l'ouest du domaine. Cependant, pour SEABASS12, de plus fines structures évoluent quant à elles tout autour du jet, contribuant de manière significative à la différence de forme et d'extension à l'est du jet entre le 1/4° et le 1/12°.

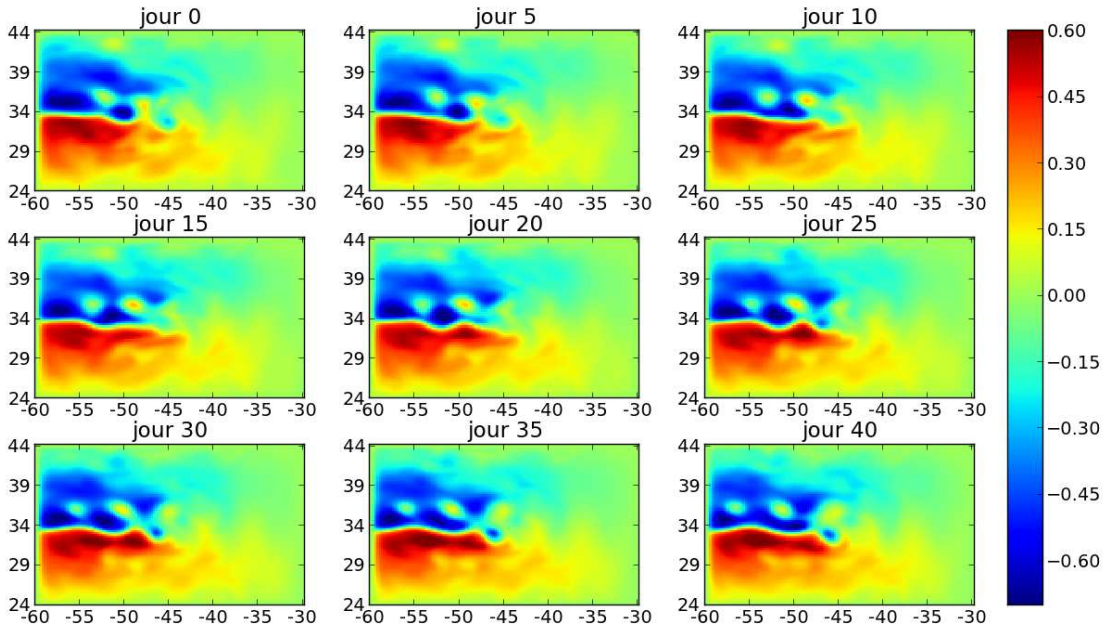

Figure 4.5 – *Champs instantanées de SSH (en m), tous les 5 jours pour SEABASS4. On voit clairement des tourbillons se détachant du jet central et recirculant vers l'ouest*

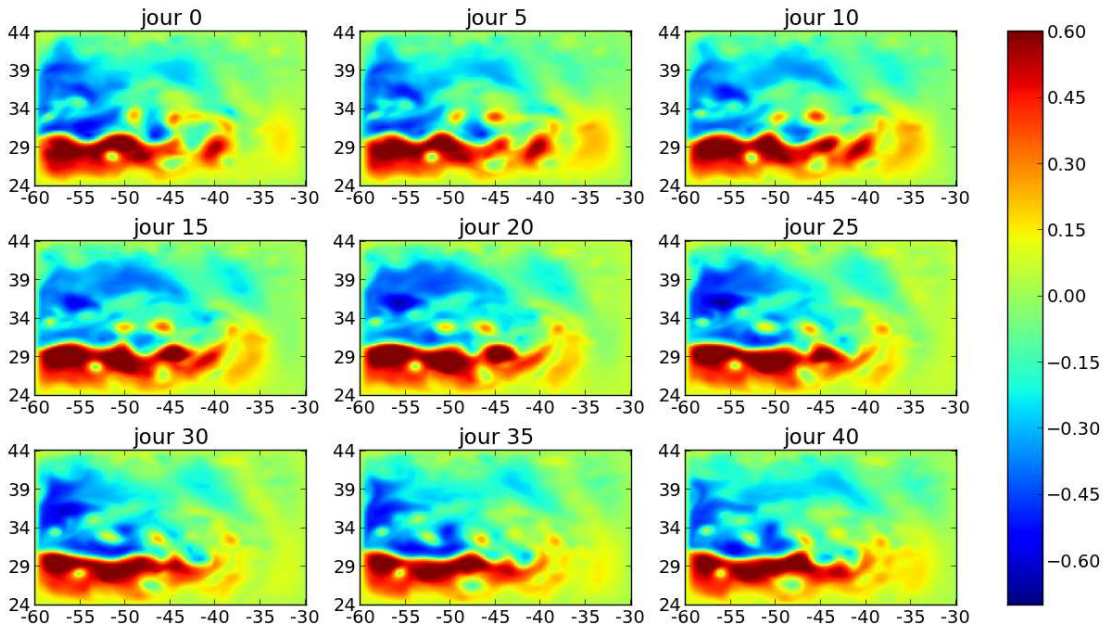

Figure 4.6 – *Idem Figure 4.5 pour SEABASS12*

Si l'on s'intéresse maintenant à l'évolution de la température, nous pouvons voir sur la Figure 4.7 qu'elle présente une légère dérive temporelle :

- La moyenne de la température sur les 2 premiers niveaux verticaux a tendance à diminuer légèrement au cours du temps
- Plus en profondeur, elle augmente au contraire légèrement

Ce comportement est tout à fait logique. En effet, dans cette configuration, il n'y aucun forçage thermique. Ainsi, la température, pour une période suffisamment longue, sous l'effet du mélange

vertical, va avoir tendance à s'uniformiser sur le domaine. Cette uniformisation apparaît plus clairement sur les couches supérieures, $\overrightarrow{\nabla_z T}$ étant plus fort en surface, comme nous l'avons montré sur la Figure 4.2. Cependant, la durée des expériences d'assimilation que nous allons réaliser dans cette étude étant de l'ordre d'un an au maximum, ce phénomène est négligeable, à la condition de prendre une ébauche suffisamment proche de l'état vrai pour que les deux trajectoires soient cohérentes du point de vue thermique.

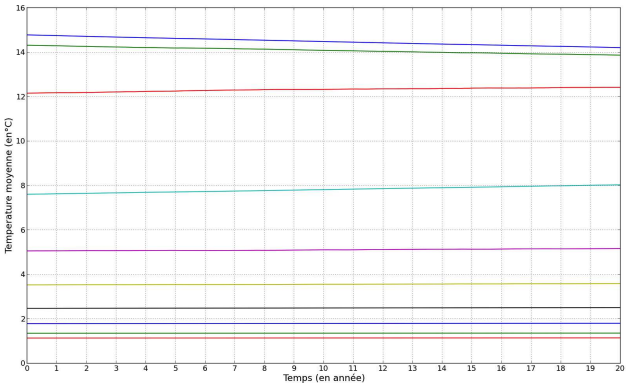

Figure 4.7 – *Évolution temporelle de la moyenne spatiale de température (en °C), pour 20 ans de simulation de SEABASS (après 80 ans de spinup), sur 10 niveaux horizontaux. Le premier niveau présente la température la plus élevée, tandis que le dernier niveau présente la plus basse.*

4.2.2 Statistiques de l'écoulement dans SEABASS

Écoulement moyen

La Figure 4.8 illustre bien les deux cellules de circulations pour les deux résolutions horizontales présentées (1/4° et 1/12°), celle au nord, présentant une hauteur d'eau négative et celle au sud présentant une hauteur d'eau positive. Le décollement du jet au sud à plus haute résolution est ici encore clairement illustré. Au 1/12°, le jet apparaît bien déstabilisé comparativement au 1/4°, comme ses oscillations en témoignent.

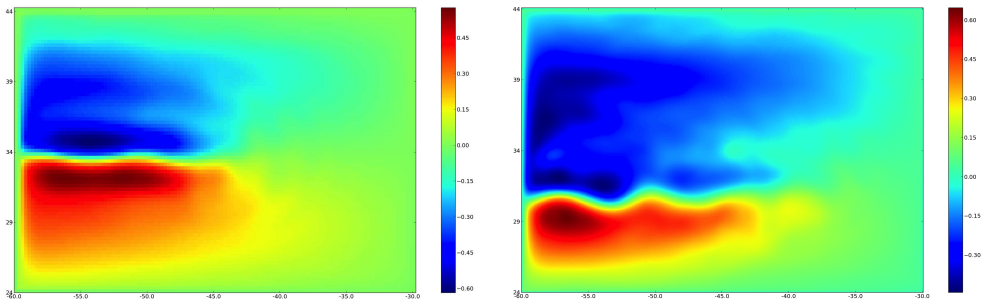

Figure 4.8 – *Moyenne temporelle du champ de SSH (en m) pour 2 ans d'intégration libre de SEABASS, au 1/4° à gauche, au 1/12° à droite*

Les Figures 4.9 et 4.10 montrent deux zones localisées présentant une grande énergie cinétique :

- La frontière ouest du domaine, où la vitesse méridienne est forte (de l'ordre de $1m.s^{-1}$)
- Le jet zonal, où la vitesse zonale est élevée (de l'ordre de $1m.s^{-1}$)

Si l'on regarde dans les eaux profondes, ce contraste s'atténue et une plus large zone énergétique entourant le jet apparaît.

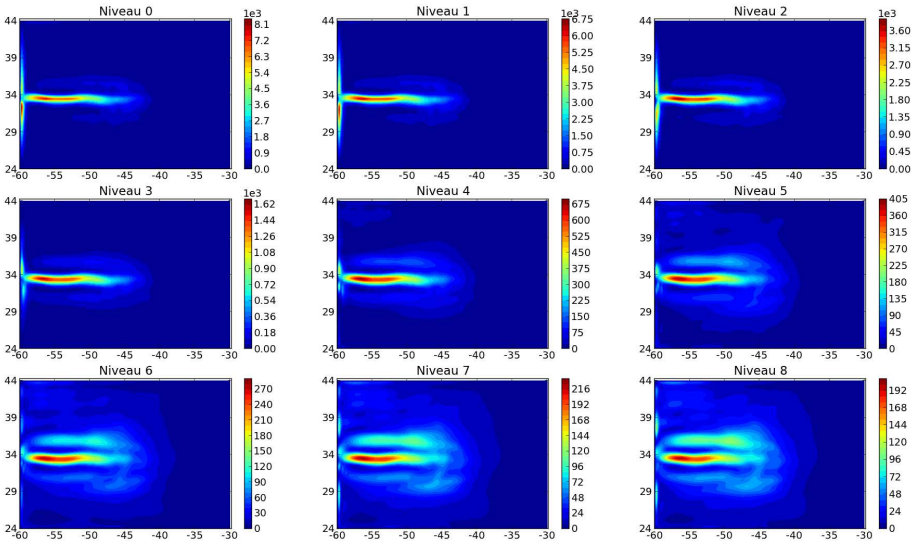

Figure 4.9 – *Énergie cinétique ($cm^2.s^{-2}$) moyennée sur 1 an pour SEABASS4*

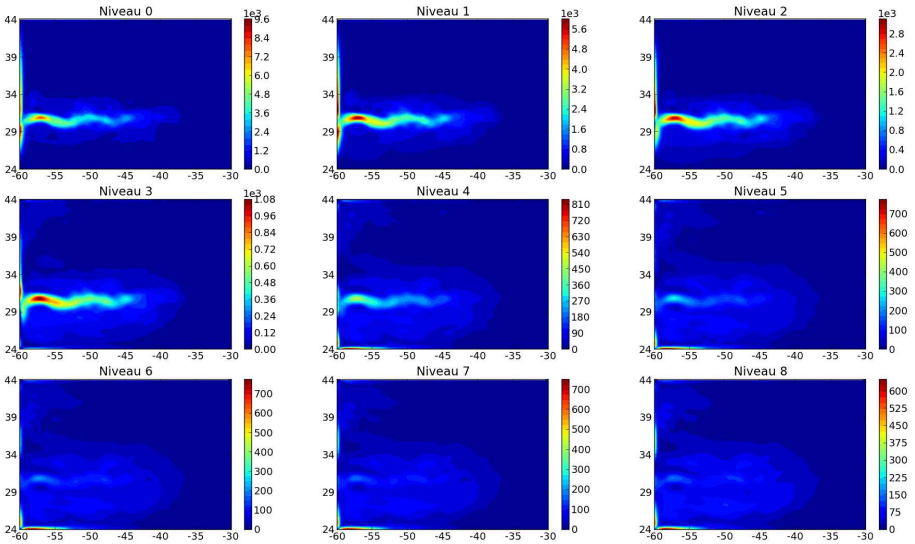

Figure 4.10 – *Idem Figure 4.9 pour SEABASS12*

Variabilité de l'écoulement

Si l'on s'intéresse aux écarts types de SSH (*cf.* Figure 4.11), il apparaît clairement que le jet est la région présentant la plus forte variabilité du domaine en terme de dynamique. Ces fortes variations dans cette région traduisent les structures tourbillonnaires méso-échelles se détachant régulièrement du jet, structures non-linéaires recirculant vers l'ouest, comme nous l'avons mentionné précédemment.

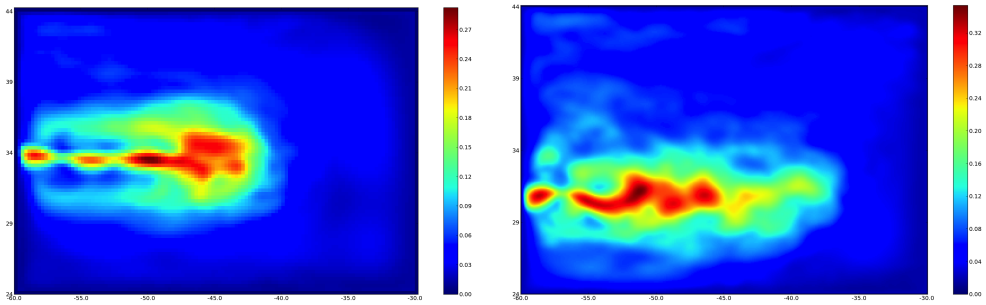

Figure 4.11 – *Écart-type de SSH au 1/4° (à gauche) et au 1/12° (à droite)*

Cette dynamique est également illustrée par les Figures 4.12 et 4.13, présentant l'énergie cinétique turbulente moyenne, pour les deux résolutions étudiées. En effet, nous voyons que celle-ci est maximum dans la région du jet. Nous pouvons remarquer également qu'au décollement du jet zonal, ces deux indicateurs montrent une variabilité s'étendant sur plusieurs degrés de latitude. En effet, la zone de décollement varie dans le temps, le jet étant déstabilisé par les structures non-linéaires.

Les niveaux d'énergie cinétique turbulente sont comparables à ceux observés dans la région du Gulf Stream ou du Kuroshio (voir, par exemple, Ducet et Le Traon (2001)).

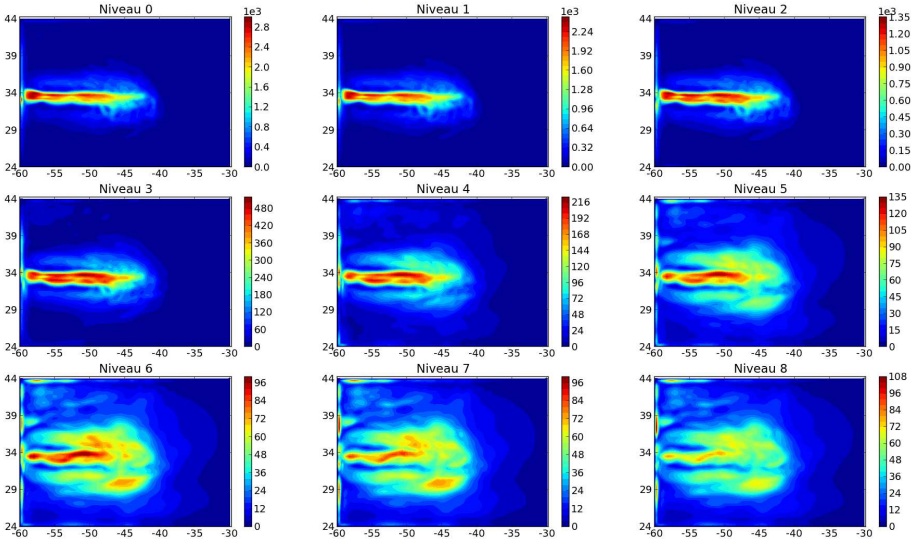

Figure 4.12 – *Énergie cinétique turbulente ($cm^2.s^{-2}$) moyennée sur 1 an pour SEABASS4*

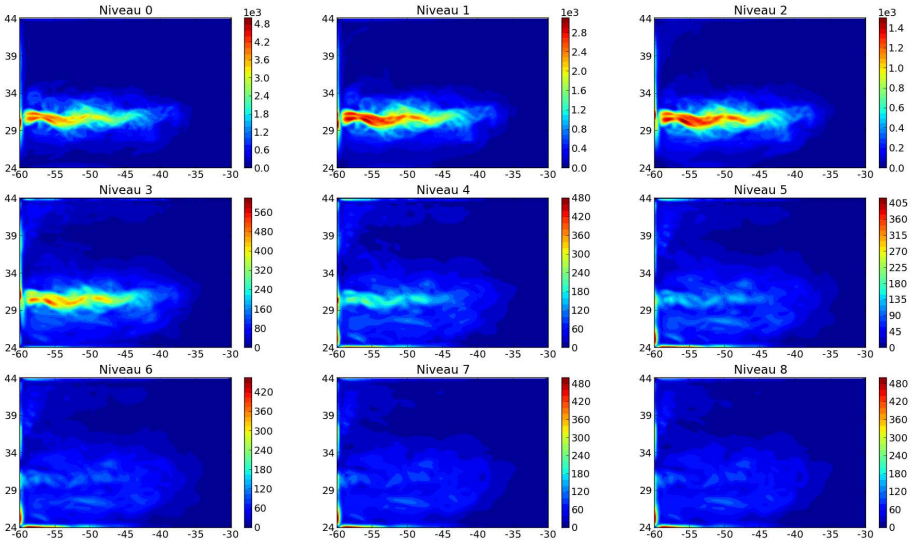

Figure 4.13 – *Idem que la Figure 4.12 pour SEABASS12*

Nous pouvons noter tout de même des différences sur la variabilité dynamique de SEABASS en fonction de la résolution horizontale considérée. En effet, au 1/4°, la variabilité est concentrée sur l'extrémité est du jet, tandis que pour SEABASS12, elle présente une répartition plus uniforme le long de l'extension du jet. Les plus grosses structures tourbillonnaires s'échappent majoritairement de l'extrémité est du jet. De plus fines structures évoluent quant à elles partout dans cette région, contribuant de manière significative à l'évolution de la forme et de la position du jet. À plus basse résolution, les échelles de ces fines structures ne sont pas résolues, induisant un jet plus stable. Seules les grosses structures non-linéaires sont représentées.

Ces dynamiques restent malgré tout très complexes et dépendent de réglages fins des paramètres du modèle.

Décomposition EOFs de l'écoulement

Les EOFs présentées ici ont été calculées à partir de sorties instantanées journalières de SSH, sur une période de simulation de deux ans. La Figure 4.14 montre que les 27 premières EOFs contiennent environ 90% de l'information de la variabilité naturelle du modèle. Il apparaît qu'il faut plus d'EOFs (32) pour SEABASS12 pour atteindre ce score, ce qui est cohérent avec le fait que la variabilité de l'écoulement est plus riche à plus haute résolution.

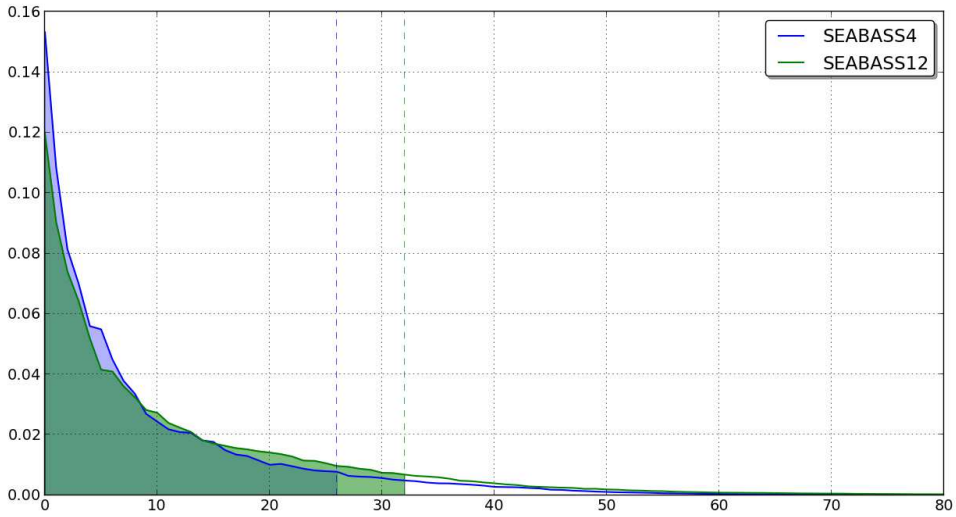

Figure 4.14 – *Variance expliquée (en pourcentages) par les EOFs calculées pour SEABASS4 (en bleu) et SEABASS12 (en vert). Les parties colorées correspondent à une variance totale expliquée de 90%.*

Comme nous l'avons également vu plus haut, la variabilité est principalement concentrée dans le jet zonal. Ainsi, les structures dominantes des premières EOFs, présentées sur les Figures 4.15 et 4.16 se situent bien dans cette région. Si l'on s'intéresse à la première EOF pour les deux résolutions, traduisant le principal mode de variabilité de l'écoulement, nous observons dans SEABASS4 deux structures distinctes au nord et au sud du jet. En revanche, pour SEABASS12, nous observons des structures circulaires tout au long du jet. Ceci confirme la plus grande stabilité horizontale du jet à la plus basse résolution.

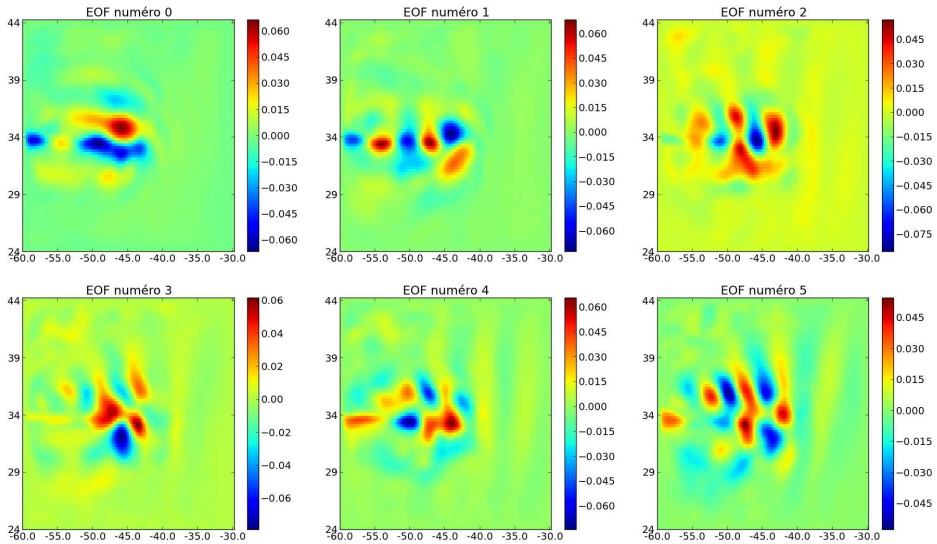

Figure 4.15 – *Structures de SSH présentées par les 6 premières EOFs (en terme de variance expliquée) calculées sur l'écoulement de SEABASS4*

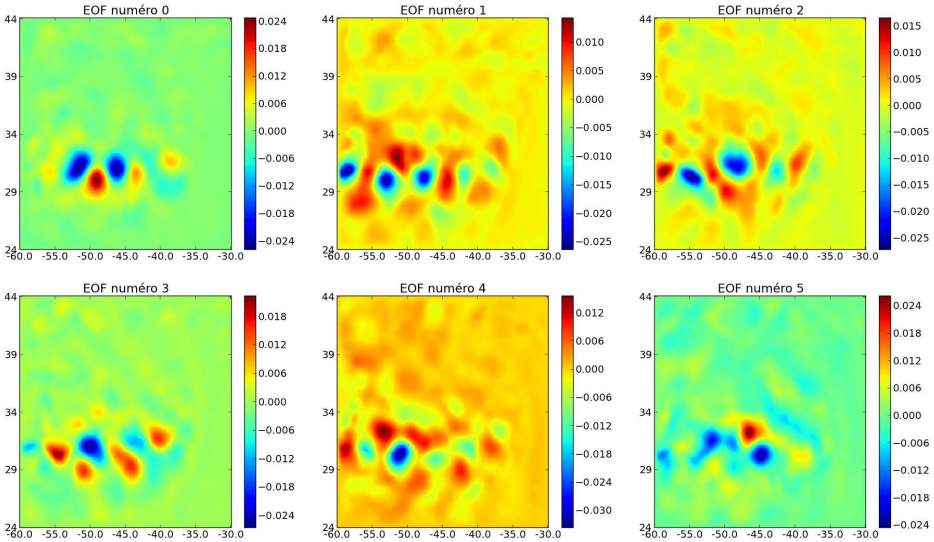

Figure 4.16 – *Idem Figure 4.15 pour SEABASS12*

Sur ces mêmes figures, il apparaît également que les échelles des structures spatiales présentées par les EOFs décroissent avec la part de variance expliquée. En s'intéressant aux EOFs d'ordre plus élevé (voir Figures 4.17 et 4.18), ces échelles diminuent clairement et ne sont plus seulement dans la région du jet mais s'étendent dans le bassin.

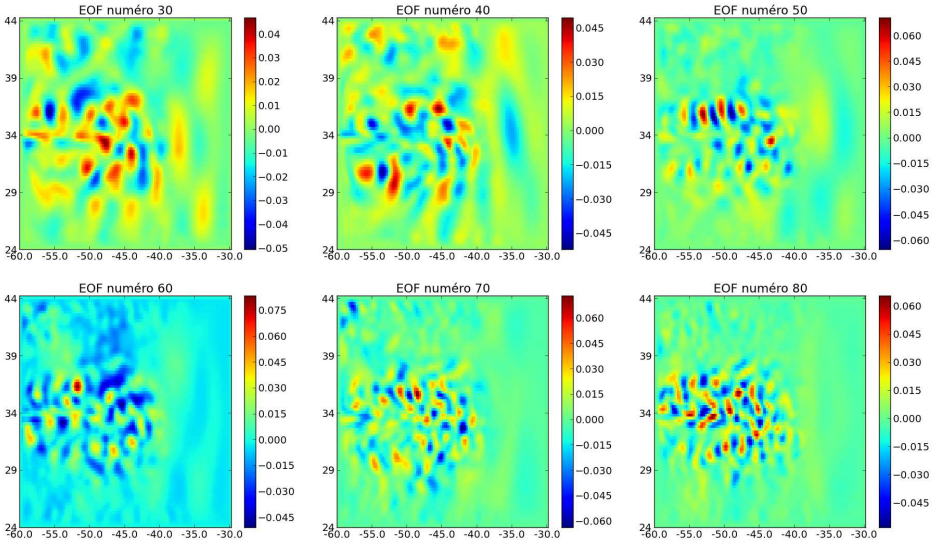

Figure 4.17 – *Structures de SSH présentées par les EOFs n°30 à 80 (en terme de variance expliquée) calculées sur l'écoulement de SEABASS4*

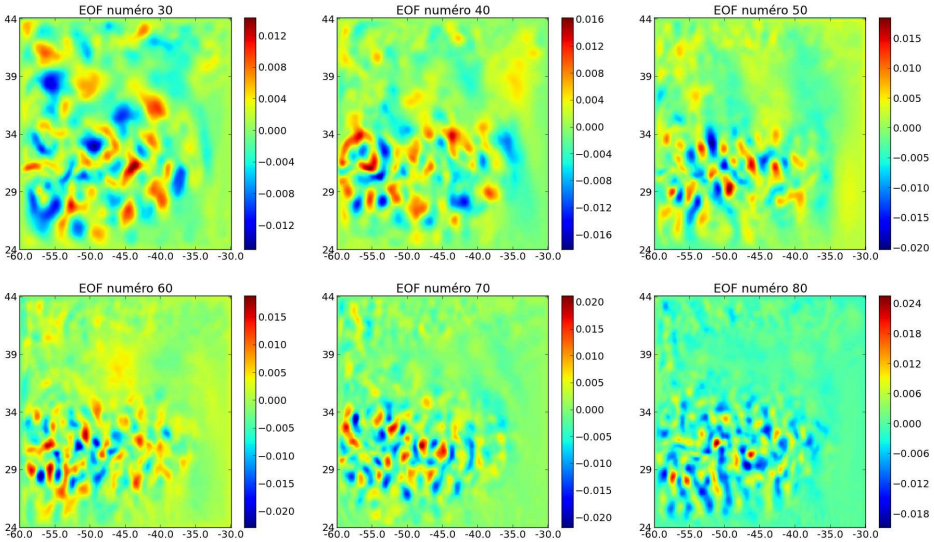

Figure 4.18 – *Idem Figure 4.17 pour SEABASS12*

Ces EOFs nous serviront par la suite à caractériser les échelles des structures d'erreurs produites en sortie du système d'assimilation.

4.2.3 Indicateurs de non-linéarité

Temps de décorrélation

La corrélation temporelle permet de quantifier la durée pendant laquelle l'état simulé de l'écoulement est relié statistiquement à l'état initial. Il est intéressant, dans le cadre de l'AD variationnelle d'en avoir un aperçu. En effet, les performances de cette catégories d'algorithmes d'AD dépend en partie de la longueur de la fenêtre temporelle durant laquelle les données seront assimilées.

Soit $\mathbf{x}_m(t_i) = \int_\Omega \mathbf{x}(t_i, x, y, z) d\Omega$ l'intégrale spatiale d'une variable d'état de notre modèle. Nous rappelons ici le calcul du facteur de corrélation temporelle $\rho(T)$ (montré dans l'équation 1.2, basé sur le principe de l'autocorrélation :

$$\rho(T) = \frac{\sum_{i=1}^{N} (\mathbf{x}_m(t_i) - \bar{\mathbf{x}}_m)(\mathbf{x}_m(t_i + T) - \bar{\mathbf{x}}_m)}{N \sqrt{\sum_{i=1}^{N} (\mathbf{x}_m(t_i + T) - \bar{\mathbf{x}}_m)^2} \sqrt{\sum_{i=1}^{N} (\mathbf{x}_m(t_i) - \bar{\mathbf{x}}_m)^2}} \tag{4.4}$$

où $\bar{\mathbf{x}}_m = \frac{1}{N} \sum_{i=1}^{N} x_m(t_i)$.

Nous allons regarder ce facteur pour différentes échelles spatiales de l'écoulement. Pour ce faire, nous avons choisi d'utiliser la décomposition en EOF. Comme nous l'avons vu plus haut,

nous pouvons décomposer la variable d'état considérée sur les EOFs du modèle :

$$\mathbf{x}(x,y,z,t) = \bar{\mathbf{x}}(x,y,z) + \sum_{i=1}^{N} c_i(t) EOF_i(x,y,z) \tag{4.5}$$

À l'aide de cette décomposition, nous avons choisi de distinguer la partie « méso-échelle », appelée $\mathbf{x}_{me}$, et la partie « subméso-échelle », appelées $\mathbf{x}_{sub}$ de la variable d'état considérée. Pour ce faire, nous avons tronqué arbitrairement la décomposition de la variable aux 10 premières EOFs pour $\mathbf{x}_{me}$, considérant le reste des EOFs pour $\mathbf{X}_{sub}$:

$$\mathbf{x}_{me}(x,y,z,t) = \bar{\mathbf{x}}(x,y,z) + \sum_{i=1}^{10} c_i(t) EOF_i(x,y,z) \tag{4.6}$$

$$\mathbf{x}_{sub}(x,y,z,t) = \bar{\mathbf{x}}(x,y,z) + \sum_{i=11}^{N} c_i(t) EOF_i(x,y,z) \tag{4.7}$$

Ainsi, nous calculons $\rho_{me}(T)$ et $\rho_{sub}(T)$ à partir de ces deux grandeurs. C'est ce qui est présenté dans la Figure 4.19.

Le niveau de corrélation est jugé empiriquement. On remarque qu'il décroît assez vite au cours des 15 premiers jours pour SEABASS4 puis réaugmente pour enfin osciller autour de 0 après $T = 60$ jours pour SEABASS4, moment à partir duquel nous pouvons considérer l'écoulement totalement décorrélé de la condition initiale. Le même phénomène est observé pour SEABASS12, mais sur de plus courtes périodes, où le facteur de corrélation oscille autour de 0 à partir de 20 jours. L'augmentation des phénomènes non-linéaires, sources de la décorrélation à la condition initiale, est ainsi traduite.

Quelle que soit la résolution horizontale considérée, la partie « subméso-échelle » de l'écoulement se décorrèlera plus rapidement.

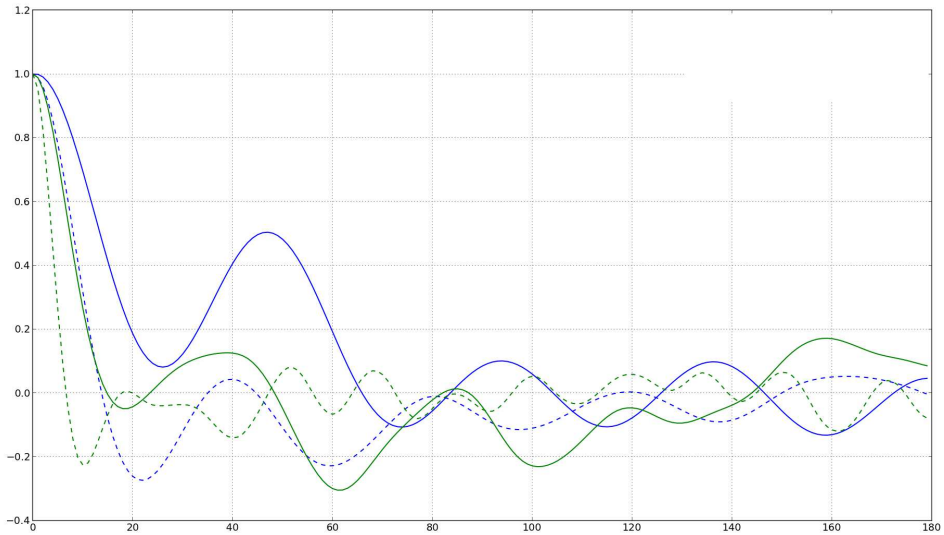

Figure 4.19 – *Évolution du facteur de décorrélation temporelle $\rho(T)$ pour la méso-échelle de l'écoulement (trait continu) et la subméso-échelle (en pointillés), pour SEABASS4 (en bleu) et SEABASS12 (en vert)*

4.3 Conception des expériences d'assimilation de données avec SEABASS

À partir des caractéristiques de SEABASS, nous avons mis en place des jeux d'expériences jumelles d'AD, afin d'étudier l'impact des non-linéarités, présentées dans la Section 4.2, sur l'algorithme du 4D-VAR incrémental. Les différentes composantes de la conception de ces jeux d'expériences sont présentées ici.

Nous avons mis en place un jeu d'expériences se basant sur des caractéristiques communes, fondées sur la configuration SEABASS au 1/4° présentée dans la Section 4.1.2. Nous détaillerons tout d'abord les caractéristiques de l'ébauche et de l'état vrai, puis nous justifierons les paramétrisations diverses du système d'AD que nous avons effectuées.

4.3.1 Ébauche et état vrai

Toutes les expériences réalisées dans nos travaux ont pour but de contrôler la condition initiale. Dans le cadre d'expériences jumelles, définir l'état vrai et l'ébauche revient à prendre deux conditions initiales différentes. Les expériences jumelles consistent à simuler une trajectoire de référence (dite aussi *vraie*), à partir d'une condition initiale donnée. Cette condition initiale est choisie après suffisamment de temps d'intégration libre du modèle pour qu'il soit statistiquement stationnaire comme montré dans la Section 4.2. Cette trajectoire de référence sera appelée par la suite trajectoire *vraie*. Les observations seront générées à partir de cette trajectoire.

Nous avons choisi ici pour état initial vrai $\mathbf{x}_0^t$ celui obtenu après une intégration libre de 80 ans de la configuration SEABASS au 1/4°. Comme signalé dans la section 4.2, la moyenne horizontale de température présente une légère dérive temporelle. Pour ne pas avoir à prendre en compte cette tendance, nous avons choisi un état initial d'ébauche $\mathbf{x}_0^b$ peu éloigné en temps de l'état vrai : il est donc pris au bout de 80 ans et 3 mois d'intégration libre du modèle. Ces trois mois de décalage assure tout de même une décorrélation temporelle entre $\mathbf{x}_0^t$ et $\mathbf{x}_0^b$, comme nous l'avons également mentionné dans la Section 4.2. Ainsi, notre ébauche est un état initial suffisamment éloigné dans le temps de l'état initial vrai pour que l'erreur entre les deux trajectoires soit du même ordre de grandeur que la variabilité naturelle du modèle. Pour caractériser ceci, nous avons calculé des cartes d'erreur de SSH entre la trajectoire vraie (d'où sont tirées les observations) et la trajectoire d'ébauche (décalée de 3 mois dans le temps). Ceci est présenté sur la Figure 4.20, qui, comparativement à la Figure 4.11, montre les mêmes structures et les mêmes ordres de grandeur.

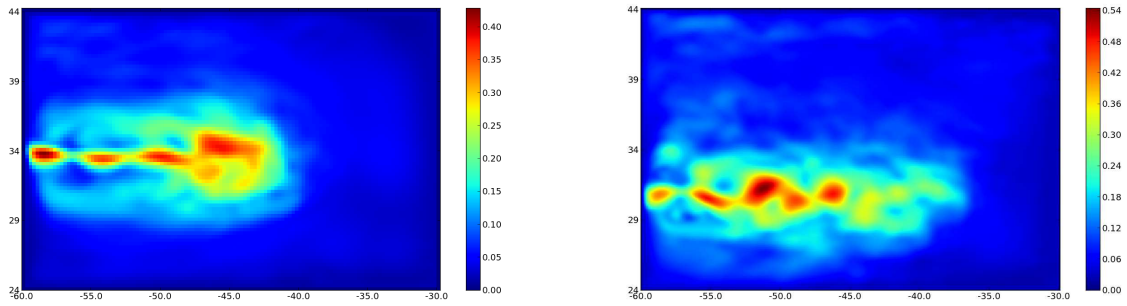

Figure 4.20 – *Écart type des cartes d'erreur d'ébauche de SSH sur deux ans d'intégration, pour SEABASS4 (à gauche) et SEABASS12 (à droite)*

Comme c'est classiquement le cas en assimilation de données pour des expériences jumelles, le modèle est ici considéré comme parfait. L'intérêt des expériences jumelles est ici de contrôler typiquement l'état vrai en réduisant au minimum les sources d'erreurs (erreur modèle, erreur de représentativité) qui pourraient « parasiter » notre compréhension des performances du système d'AD.

4.3.2 Génération des observations

Pour construire nos jeux d'observations, nous avons choisi de simuler des traces altimétriques. Nous justifions ce choix par plusieurs arguments :

- Depuis l'avènement des satellites altimétriques, la mesure de la hauteur de l'eau est une source d'information majeure pour l'océanographie. Elle est aujourd'hui systématiquement utilisée en océanographie opérationnelle. La considérer dans notre étude permet d'établir des conclusions pour un réseau d'observation « réaliste », dans ce sens.
- La variable observée (SSH) porte sur la partie dynamique de notre modèle, qui nous intéresse dans cette étude.
- La couverture temporelle et spatiale est suffisamment dense pour observer les phénomènes non-linéaires méso-échelles.

Nous observerons donc la SSH dans notre cas. Pour générer ces observations, nous avons besoin de deux informations :

1. L'empreinte spatiale et temporelle des traces altimétriques sur le domaine de SEABASS
2. Les champs de SSH de la trajectoire de référence

Pour le premier point, nous avons mis au point un programme simulant la trajectoire au sol d'un satellite selon des caractéristiques fournies en entrée. Ainsi, nous pouvons simuler des traces telles que celles produites par les satellites de type SARAL/AltiKA ou Jason-1.

En ce qui concerne le deuxième point, pour des raisons algorithmiques, de coûts de calcul et de stockage, nous ne pouvons pas obtenir les champs de SSH à chaque instant de prise de mesure faite par le satellite. Qui plus est, les sorties de notre modèle sont sur une grille discrétisée, et ne coincïdent que rarement avec la position exacte des observations satellitaires. Ainsi, nous devons faire une interpolation temporelle d'une part et spatiale d'autre part.

Pour la dimension temporelle, nous procédons comme suit. Nous disposons de champs complets de SSH (issus de notre trajectoire de référence), à intervalles de 12 heures. À partir de la trajectoire simulée du satellite entre $t - 6h$ et $t + 6h$, nous avons les positions des observations dans cet intervalle. Alors, la valeur des observations simulées sont interpolées à partir du champ issu du modèle à l'instant t par un opérateur d'interpolation bilinéaire. Une illustration de traces simulées selon ce procédé est présentée dans la Figure 4.21.

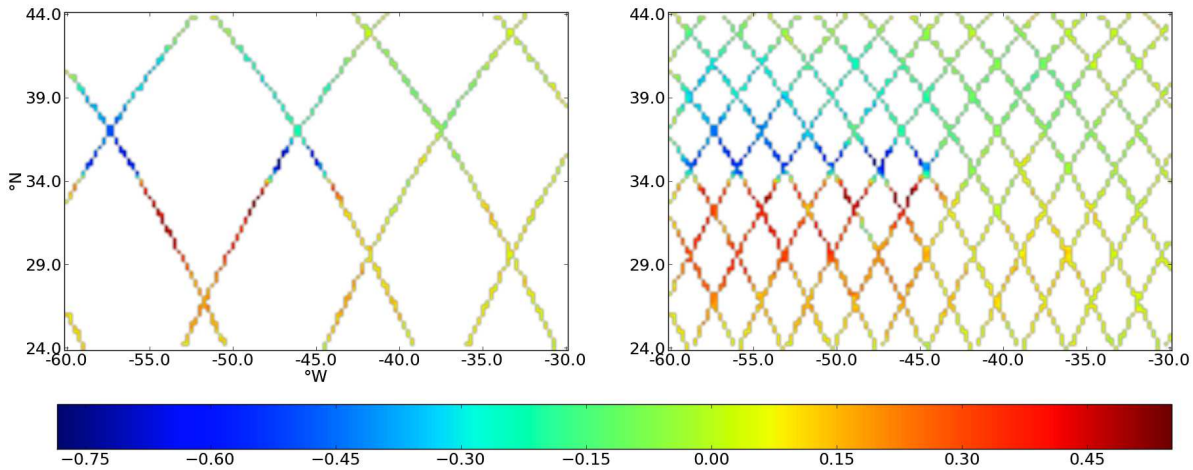

Figure 4.21 – *Exemples de mesures altimétriques simulées (JASON-1) après 2 jours (à gauche) et 10 jours (à droite)*

Il est à noter que nous ajoutons un bruit blanc d'écart-type que nous prescrivons à 0.03 m, quel que soit l'altimètre simulé. Les observations ainsi obtenues présentent une erreur décorrelée en temps et en espace.

4.3.3 Paramétrisation de NEMOVAR

Comme nous l'avons vu dans la Section 3.4, il est nécessaire de paramétrer NEMOVAR pour spécifier les matrices de covariances d'erreurs. À partir des statistiques du modèles présentées dans la Section 4.2, nous détaillons ici les paramètres choisis pour nos expériences.

Pour l'algorithme de minimisation de la fonction coût quadratique, nous utilisons celui détaillé dans la Section 3.4.3. La matrice de covariances d'erreurs d'ébauche nécessite plusieurs paramétrisations, dépendantes de la connaissance que l'on a du modèle, comme expliqué dans la Section 3.4.2 :

1. L'écart-type de l'erreur sur la partie non-équilibrée de chaque variable (et totale sur T).
2. Les longueurs de corrélations

Pour les premières, nous nous sommes basées sur la variabilité du modèle telle que présentée dans la Section 4.2, pour une résolution horizontale donnée. Les valeurs spécifiés pour les deux résolutions utilisées sont présentées dans le tableau 4.6.

Résolution	T_s	T_{ml}	T_{do}	SSH	U	V
1/4°	0.2	0.5	0.05	0.28	0.43	0.48
1/12°	0.2	0.7	0.1	0.39	0.58	0.9

Table 4.6 – *Écarts-types des erreurs d'ébauches, à différentes résolutions, pour les différentes variables du modèle (dans leurs unités respectives). Pour la température, ils sont spécifiés pour la surface (s), la couche de mélange (ml) et l'océan profond (do)*

En ce qui concerne les longueurs de corrélation, nous les avons calculées directement sur les sorties de notre modèle. Les valeurs spécifiés pour les deux résolutions utilisées sont présentées dans le tableau 4.7.

Résolution	T	SSH	U	V
1/4°	1.	1.	1.	1.
1/12°	0.25	0.25	0.25	0.25

Table 4.7 – *Longueurs de corrélation des erreurs d'ébauches, en degré, à différentes résolutions, pour les différentes variables du modèle*

Pour la matrice de covariances d'erreur d'observations, dans notre version de NEMOVAR, seule sa diagonale est à spécifier. Autrement dit, il faut renseigner l'écart-type de l'erreur associé à chaque type d'observation que l'on assimile. Dans notre cas, cette valeur est celle de l'amplitude du bruit blanc rajouté aux observations altimétriques simulées, soit 0.03 m dans nos expériences [3].

3. Ce qui correspond au niveau moyen d'erreur des observations issues de l'altimètre Jason-1

Troisième partie

Étude qualitative et quantitative de l'algorithme 4D-VAR incrémental appliqué à la modélisation océanographique

Chapitre 5

Applications du 4D-VAR incrémental à SEABASS au 1/4°

Dans ce chapitre, nous ferons l'analyse de différentes expériences d'AD variationnelle autour de la configuration SEABASS considérée à une résolution horizontale d'un quart de degré. Les objectifs principaux sont, d'une part, de décrire en détail la réponse du système d'assimilation 4D-VAR incrémental dans ce contexte de modèle de bassin océanique *eddy-permitting*, et d'autre part de qualifier et quantifier l'influence de la non-linéarité au sein de ces expériences.

Nous analyserons divers aspects des réponses du système d'AD face à la non-linéarité à l'aide de différents diagnostics physiques ou méthodologiques.

5.1 Analyse d'une expérience d'assimilation « bien réglée »

Le propos que nous développerons dans cette section est d'analyser sous divers angles une expérience d'assimilation variationnelle de données que nous considérons comme « bien réglée » avec la configuration SEABASS au 1/4°. Cette expérience, hormis la paramétrisation du système NEMOVAR présentée en 4.3.3, présente les caractéristiques suivantes :

- La période totale d'assimilation est d'un an
- Elle est découpée en 6 cycles de 2 mois
- Au sein de chaque cycle, 5 boucles externes et 10 boucles internes seront effectuées.

Nous ferons tout d'abord une étude de convergence de l'algorithme du 4DVAR incrémental puis nous nous intéresserons à l'aide de différents diagnostics, aux caractéristiques des erreurs d'analyse et de prévision.

5.1.1 Évolution des fonctions coûts

Pour qualifier la convergence lors d'une expérience de minimisation, le premier diagnostic est de regarder l'évolution de la fonction coût au fil des itérations, globalement puis composante par composante.

Comme nous l'avons vu dans la Section 2.2.2, au sein d'une boucle externe, à chaque itération de boucle interne (appelée par la suite *itération interne*), via l'exploration de l'espace de contrôle, les termes J_b et J_o sont modifiés. Le premier terme, traduisant l'écart entre l'état initial d'ébauche et celui recherché par la minimisation, augmentera tandis que celui de la fonction coût d'obser-

vation diminuera sauf problème, puisque l'algorithme cherche en premier lieu à rapprocher la trajectoire du modèle des observations,

Fonction coût totale

Régimes transitoire et permanent Comme expliqué par Desroziers et Ivanov (2001), un diagnostic statistique permettant de juger de la convergence d'un système bien calibré consiste à vérifier que $E(J) = p/2$ à l'optimum, où p est le nombre d'observations au cours du cycle d'assimilation. Sur la figure 5.1, présentant l'évolution de la fonction coût totale, nous voyons que nous atteignons ce seuil au bout du troisième cycle d'assimilation, pour s'y fixer par la suite. Lors des premiers cycles, la fonction coût décroît assez fortement, mais l'ébauche est encore trop éloignée de l'état vrai pour prétendre à atteindre ce seuil.

Ici, un régime transitoire puis un régime permanent peuvent être identifiés. Le premier se caractérise par une ébauche éloignée de l'état vrai, ce qui donne lieu à des incréments de grande amplitude, montrés sur la figure 5.2, mettant à mal tout d'abord l'hypothèse linéaire tangente mais aussi contrôlant moins efficacement les différentes échelles spatiales et temporelles, comme nous le verrons plus loin dans ce chapitre. À la suite, le régime permanent se définit par la réduction de l'écart entre l'ébauche et l'état vrai au fur et à mesure des cycles, produisant des incréments de plus petite amplitude (*cf.* figure 5.2).

Enchaînement des cycles d'assimilation D'un cycle d'assimilation à l'autre, nous constatons un saut de la fonction coût entre chaque cycle, plus marqué dans le régime transitoire. En effet, à l'issue d'un cycle d'assimilation, l'état final corrigé devient l'ébauche du cycle suivant. Cet état final corrigé est produit par l'algorithme qui n'a pas "vu" les observations du cycle suivant. Ainsi, la fonction coût d'observation J_o n'a pas de raison d'être équivalente au début du cycle suivant. L'écart aux nouvelles observations n'a pas été pris en compte lors du cycle précédent et voit sa valeur logiquement augmentée à l'interface des deux cycles.

Cependant, dans le cadre de cette expérience, le modèle est parfait et les observations sont issues d'une intégration de ce modèle, seulement entâchées d'un bruit blanc (décorrélé en espace et en temps). Au fur et à mesure des cycles, le contrôle de la trajectoire va progressivement réduire les écarts sur les différentes échelles spatiales et temporelles entre l'ébauche et l'état vrai, expliquant les sauts inter-cycles beaucoup moins marqués au fur et à mesure de l'expérience.

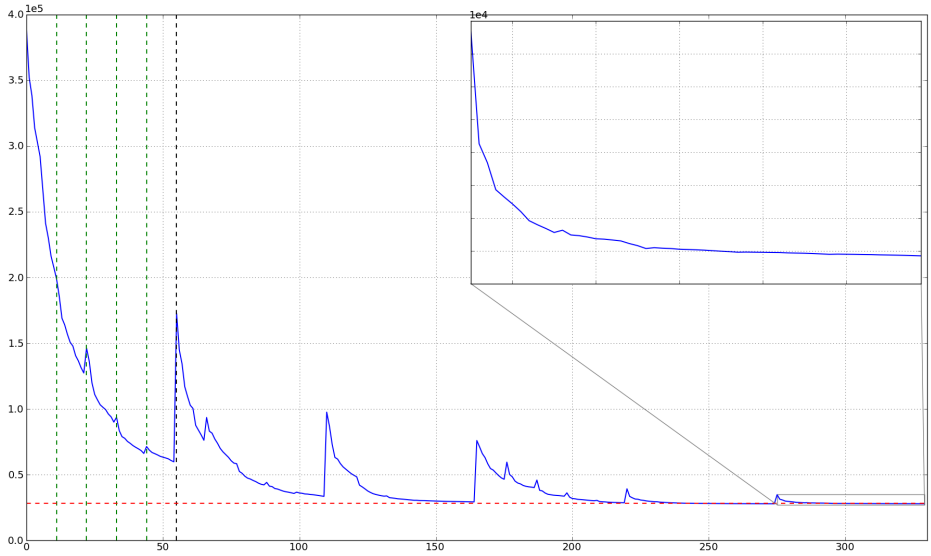

Figure 5.1 – *Évolution de la valeur de la fonction coût totale J au fil des cycles (les deux premiers séparés par une ligne noire verticale), des boucles externes (les cinq premières séparées par une ligne verte verticale) et des itérations de minimisation (dont le nombre figure en abscisse). La ligne horizontale rouge représente la valeur $p/2$ où p est le nombre moyen d'observations par cycle, vers laquelle la fonction J doit théoriquement converger en moyenne.*

Enchaînement des boucles externes Au sein d'un cycle d'assimilation, la valeur de fonction coût augmente parfois entre la fin d'une boucle externe et au début de la suivante, notamment en régime transitoire. Partant de l'équation (2.29), la nouvelle boucle externe construit les opérateurs linéarisés $\mathbf{M}^{(k+1)}$ et $\mathbf{H}^{(k+1)}$ autour de la trajectoire du modèle issue de l'incrément trouvé à la boucle externe précédente. J_o n'a donc pas de raison d'être identique à la fin de la boucle externe k et au début de la boucle externe $k+1$.

De même, le nouveau vecteur d'innovation $\mathbf{d}_i^{(k+1)}$ n'est plus le même : il est recalculé en tant qu'écart de la trajectoire non-linéaire du modèle aux observations à partir de l'incrément précédemment trouvé. Ces changements, induits par la nouvelle approximation quadratique de la fonction coût, expliquent les sauts observés.

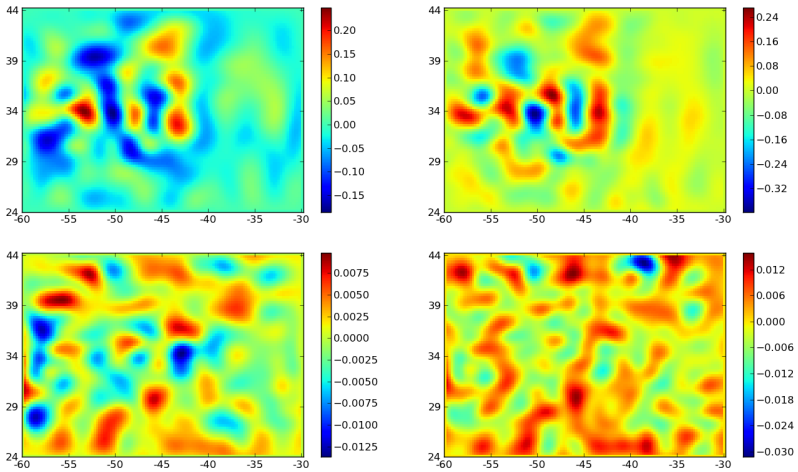

Figure 5.2 – *Incréments de SSH (en m) obtenus lors du premier cycle d'assimilation (en haut) et du dernier (en bas), à la fin de la première(à gauche) et de la cinquième et dernière boucle externe (à droite). Les incréments présentés sont cumulatifs : ce sont les incréments réellement appliqués au début de chaque boucle externe et donc la somme des incréments obtenus lors des boucles externes précédentes. Il faut bien noter la différence d'amplitude des incréments présentés ici.*

Composantes de la fonction coût

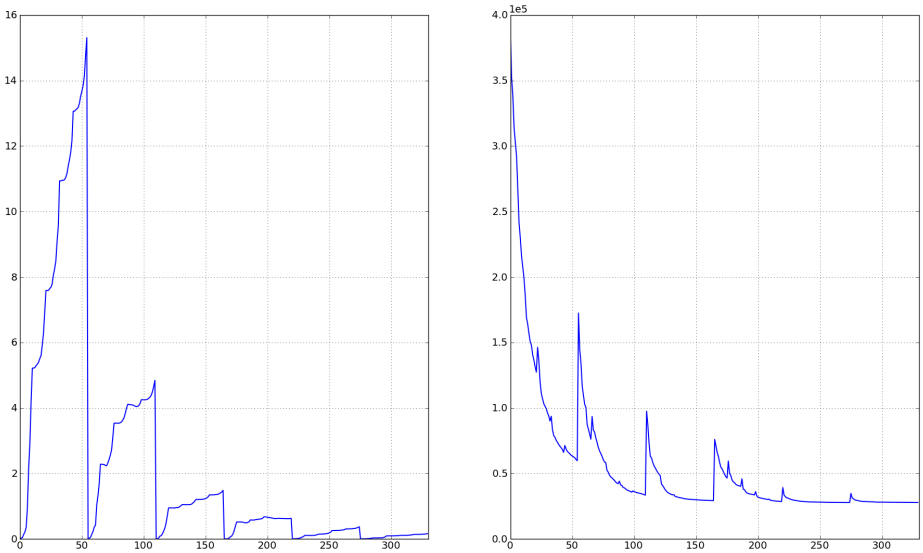

Figure 5.3 – *Évolution de la valeur des contributions J_b et J_o (respectivement à gauche et à droite) à la fonction coût totale, en fonction des itérations de minimisation.*

Sur la figure 5.3, les contributions de J_b et J_o à l'évolution de J sont représentées. Plusieurs remarques viennent.

Au début de chaque cycle (*i.e.* à la première itération de minimisation de la première boucle externe), J_b est effectivement nulle, ce qui est cohérent avec le fait que $\delta\mathbf{x_0}$ est nul à cette étape

de l'algorithme. De fait, la fonction coût d'ébauche va augmenter quand l'algorithme produira un incrément pour se rapprocher des observations. Dans le régime transitoire d'assimilation, cette augmentation est plus forte : ce phénomène est cohérent avec le fait que l'ébauche est éloignée de l'état corrigé, et comme expliqué précédemment, les incréments obtenus par la minimisation sont de forte amplitude (*cf.* figure 5.2), traduisant un écart plus grand avec l'incrément d'ébauche.

La contribution de J_b est négligeable devant celle de J_o. En effet, la paramétrisation de la diagonale de $\mathbf{B}$ est faite selon la variabilité naturelle du modèle sur chacune des variables, détaillée au §4.2. Le modèle étant turbulent, cette variabilité n'est donc pas négligeable (0.1 m pour la SSH, par exemple). En revanche, pour la matrice $\mathbf{R}$, l'écart-type spécifié pour les erreurs d'observations est inspiré des niveaux d'erreurs des instruments précis de mesure altimétrique ($0.03\ m$ pour les mesures de SSH simulées ici). La confiance accordée aux observations dans la fonction coût est donc plus importante que celle accordée à l'ébauche.

5.1.2 Etude de l'erreur d'analyse

L'étude de l'évolution de la fonction coût réalisée ci-dessus nous donne une idée de la performance de l'algorithme de minimisation. Cependant, physiquement, elle ne nous renseigne pas sur la qualité de l'analyse durant cette expérience. Pour qualifier et quantifier cette erreur d'analyse, nous présenterons tout d'abord les erreurs quadratiques moyennes, spatiales et temporelles, puis une série de diagnostics, ayant pour socle la projection des cartes d'erreurs sur les EOFs du modèle, présentées dans la Section 4.2.

Erreurs quadratiques moyennes

Les diagnostics utilisés ci-après sont les erreurs quadratiques, pour chaque variable d'état, à niveau horizontal fixé :

- Moyennées en temps (appelées RMSTE dans la suite, *cf.* équation (5.1)), elles permettent de distinguer les structures spatiales d'erreurs moyennes et leurs amplitudes.
- Moyennées en espace (appelées RMSSE plus loin, *cf.* équation (5.2)), elles permettent d'identifier l'évolution de l'erreur au cours de la période d'assimilation.

Dans les deux cas, ces diagnostics ne sont pas normalisés et sont donc dans l'unité de la variable observée.

$$RMSTE[v](x,y,z) = \sqrt{\frac{1}{T}\int_0^T v^2(x,y,z,t)dt} \tag{5.1}$$

$$RMSSE[v](z,t) = \sqrt{\frac{1}{|\Omega_{x,y}|}\int_{\Omega_{x,y}} v^2(x,y,z,t)d\Omega_{x,y}} \tag{5.2}$$

La figure 5.4, présentant l'évolution temporelle de RMSSE pour l'erreur d'ébauche et l'erreur d'analyse de SSH (la variable observée), confirme la tendance montrée au §5.1.1 : la convergence s'améliorant au fil des cycles, l'erreur moyenne d'analyse décroît également au cours de l'expérience, se stabilisant sur les derniers cycles.

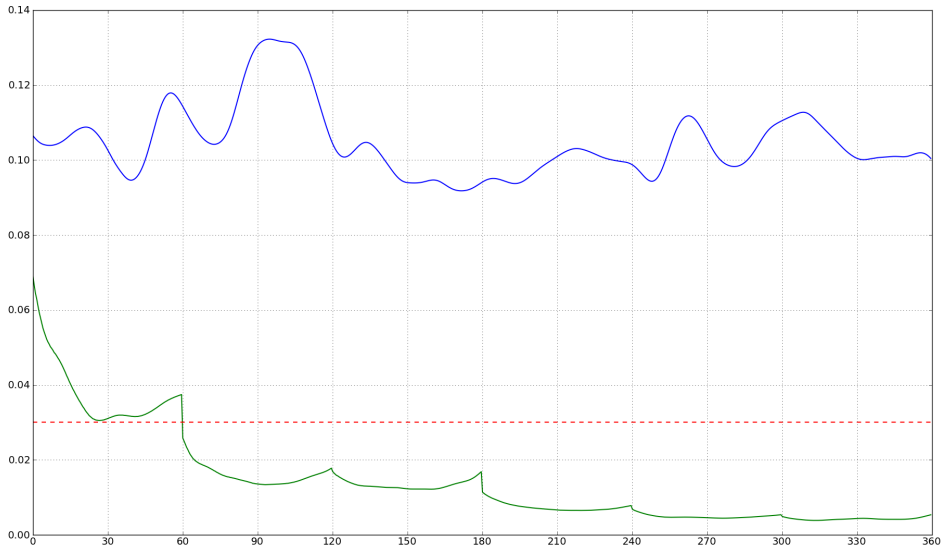

Figure 5.4 – *Évolutions des erreurs quadratiques d'ébauche (en bleu) et d'analyse (en vert), moyennées spatialement, non normalisées (RMSSE), pour la SSH (en m), en fonction du temps (en jours). La ligne horizontale rouge représente l'erreur d'observation.*

Nous pouvons voir que l'erreur décroît très vite par rapport à l'erreur d'ébauche alors que le plateau de convergence de la minimisation n'était pas atteint. En effet, la RMSSE de SSH est en moyenne en deça de l'erreur d'observation dès le premier cycle, alors que nous avons vu au §5.1 que le plateau de convergence était mal approché. Une première conclusion est donc que l'algorithme permet de réduire déjà efficacement l'erreur d'analyse, en deça de l'erreur d'observation prescrite (3 cm) sans que la convergence de l'algorithme ne soit numériquement atteinte. Il est à noter qu'en fin d'expérience, l'erreur d'analyse est environ 10 fois plus petite que l'erreur d'ébauche.

Sur les figures 5.5, 5.6 et 5.7, représentant la même quantité que ci-dessus mais pour les variables non-observées, la tendance de la réduction de l'erreur initiale est du même ordre que pour la SSH, en surface ou dans l'océan profond. L'information portée par les observations de surface pénètre bien verticalement, pour des cycles de deux mois d'assimilation. Cette information de SSH est propagée sur les variables non-observées, à la fois en horizontal et en vertical, à travers trois composantes du système d'assimilation : le modèle, le modèle adjoint et la matrice B.

Pour les premiers cycles, même si l'état initial contrôlé n'est pas l'optimum, l'état final analysé est de qualité suffisante pour constituer une bonne ébauche d'état initial pour le cycle suivant. C'est ainsi que le régime permanent d'assimilation s'installe.

Au sein de chaque cycle, de façon plus marquée pour ceux en régime transitoire, nous observons des courbes de RMS en forme de « u », traduisant une meilleure correction de l'état au milieu de la fenêtre d'assimilation. Remarqué notamment par Derber (1987) et Pires et al. (1996), ce phénomène traduit la nécessité pour corriger la trajectoire analysée d'avoir des observations passées et futures, les premières permettant de contrôler les modes stables, déterminant la précision de l'analyse de l'état présent ; les secondes permettant de contrôler les modes instables, réduisant l'erreur de prédiction. Au centre de la fenêtre temporelle, où les deux sources d'informations sont disponibles, l'erreur est minimale.

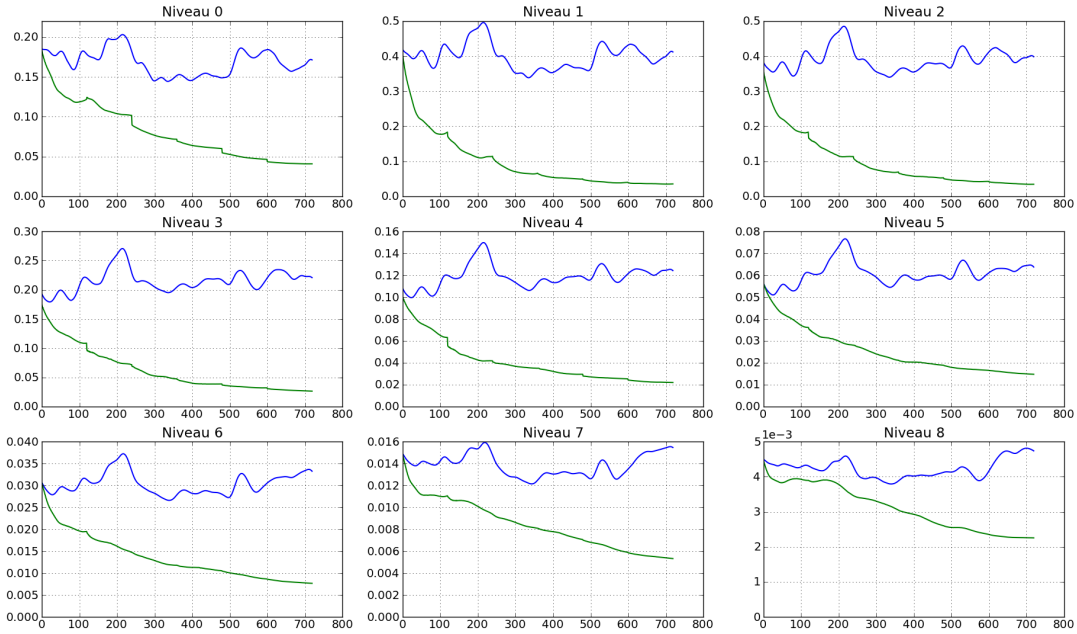

FIGURE 5.5 – *Évolutions des erreurs quadratiques d'ébauche (en bleu) et d'analyse (en vert), moyennées spatialement, non normalisées, pour la température, en °C ($RMSSE[T]$), en fonction du temps (en demi-jours), du premier au neuvième niveau vertical.*

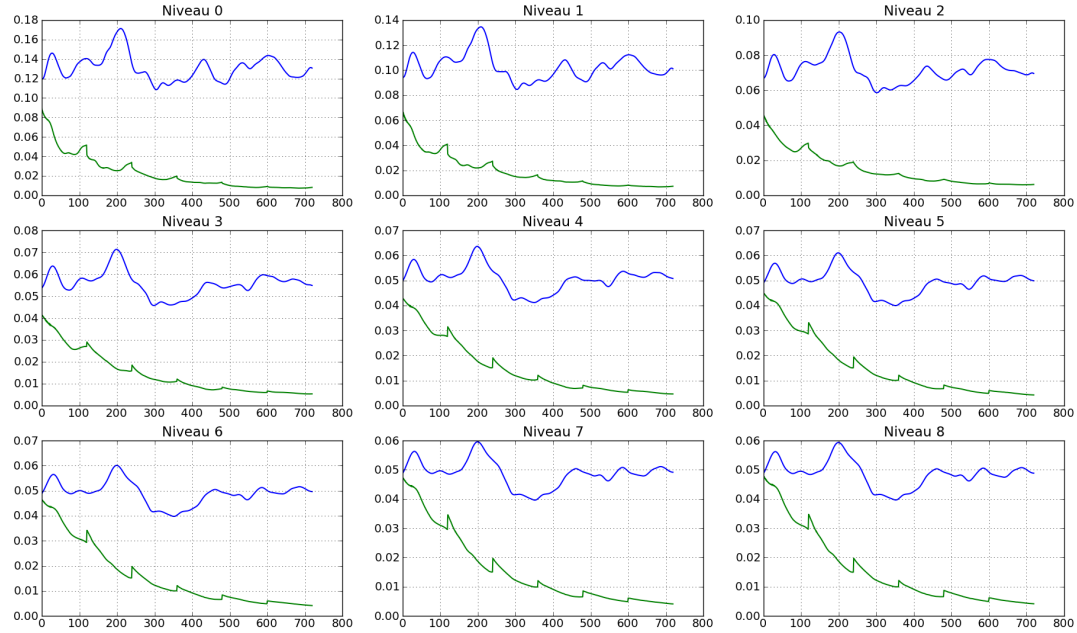

FIGURE 5.6 – *Idem figure 5.5 pour la vitesse zonale (en $m.s^{-1}$)*

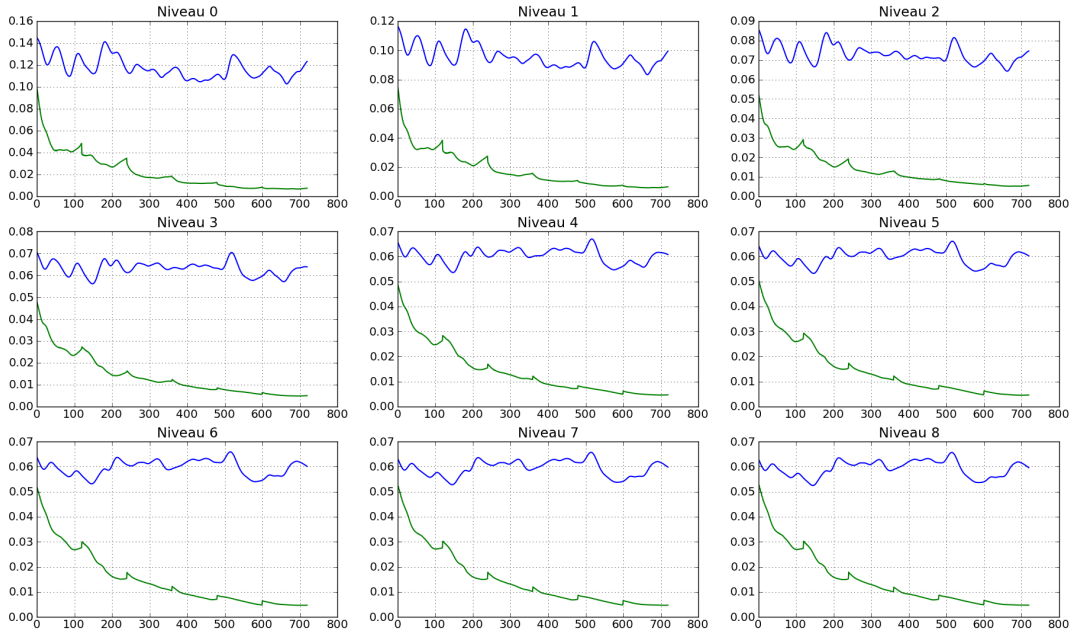

Figure 5.7 – *Idem figure 5.5 pour la vitesse méridienne (en $m.s^{-1}$)*

En regardant le champ de RMSTE de SSH de la figure 5.8, les structures d'erreurs d'ébauche principales apparaissent concentrées au niveau du jet :

- Au niveau de son décollement, dont la variabilité naturelle présente dans le modèle, évoquée au §4.2, est du même ordre de grandeur que celle observée sur l'erreur d'ébauche ;
- Au niveau de son extrémité *est*, d'où la plupart des tourbillons s'échappent du jet.

Si l'on regarde l'erreur d'analyse, ces principales structures d'erreurs sont bien réduites et ne sont plus dominantes sur l'ensemble du domaine. Le 4D-VAR incrémental réussit à recaler correctement la position du jet au fil des itérations. En revanche, nous voyons toujours, à l'extrémité *est* du jet, des structures plus fines qui restent présentes, indiquant que l'algorithme ne contrôle pas au mieux les phénomènes physiques de petites échelles dans cette zone géographique présentant une forte variabilité.

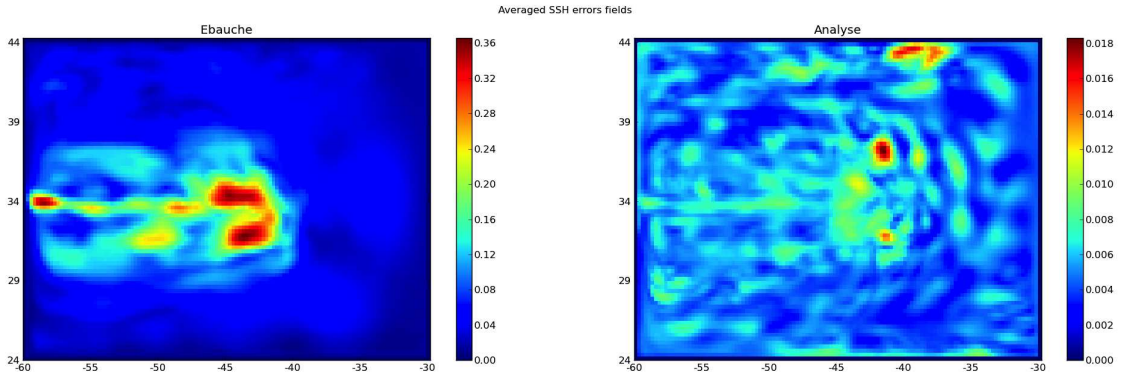

Figure 5.8 – *Champ spatial de l'erreur quadratique de SSH moyennée temporellement (RMSTE), sur les six derniers mois de l'expérience d'assimilation. L'erreur d'ébauche est à gauche et l'erreur d'analyse à droite. Les plages de couleurs sont différentes afin d'identifier les structures d'erreurs dans les deux cas. Il faut noter que l'amplitude de l'erreur d'analyse est vingt fois inférieure à celle de l'ébauche.*

Projection sur les EOFs

Pour faire écho à l'étude des structures d'erreurs spatiales, nous présentons ici un diagnostic basé sur la projection des cartes d'erreurs sur les EOFs du modèle, c'est à dire ses modes principaux de variabilité naturelle. Pour une variable donnée, à partir d'une série temporelle de cartes d'erreurs, nous projetons chacune de ces cartes sur les EOFs principales du modèle. Par exemple, pour une variable d'état donnée x, à l'instant j, nous pouvons écrire :

$$x_j^a - x_j^t = \sum_{i=1}^{N} c_{i,j} EOF_i$$

Nous obtenons, pour un échantillon de cartes d'erreurs une suite de coefficients $c_{i,j}$, j étant associé à une carte d'erreur donnée et i à une EOF donnée. Dans les figures suivantes, nous définissons les C_i comme la moyenne quadratique des coefficients de projection associée à chaque EOF i (voir équation (5.3)). La répartition de ces coefficients permet d'étudier les échelles spatiales des structures d'erreurs.

$$C_i = \sqrt{\frac{1}{T} \sum_{j=1}^{T} c_{i,j}^2} \tag{5.3}$$

Afin d'obtenir un échantillon de cartes d'erreurs statistiquement cohérent, nous présentons les résultats obtenus dans les 6 derniers mois de l'expérience d'AD, pour la SSH, les autres variables se comportant de façon comparable. Nous nous restreignons à l'étude de la projection sur les 100 premières EOFs (sur 730 calculées), classées par ordre décroissant de variance expliquée. Ainsi, comme montré au §4.2, les premières EOFs représentent des phénomènes grandes échelles, puis ces échelles décroissent avec le rang d'EOFs. Les 100 EOFs retenues représentent une variance expliquée totale de 96%.

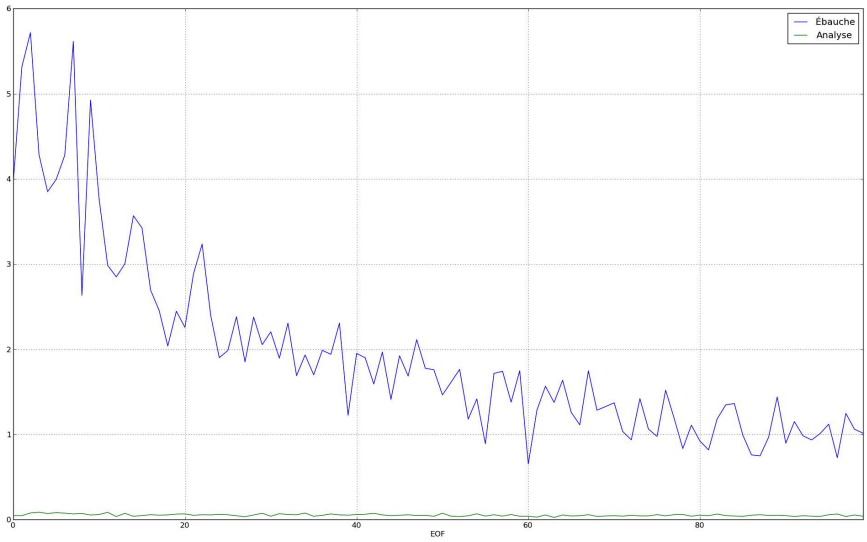

Figure 5.9 – *Moyenne quadratique des coefficients de projection des cartes d'erreurs de SSH sur les 100 premières EOFs du modèle, dont le numéro est en abscisse. Les EOFs sont classées par ordre décroissant de variance expliquée. En bleu, l'erreur d'ébauche et en vert, l'erreur d'analyse.*

Si l'on considère des cartes d'erreurs non normalisées, en comparant erreur d'ébauche et erreur d'analyse, nous voyons sur la figure 5.9 que toutes les échelles sont corrigées à l'issue de l'expérience.

Les c_i pour l'erreur d'ébauche sont répartis gloablement de façon décroissante en fonction du rang d'EOF. Ceci est cohérent avec le fait que l'ébauche suit la variabilité naturelle du modèle.

Il est cependant difficile de voir sur cette courbe comment se répartissent les échelles des structures spatiales d'erreurs d'analyse.

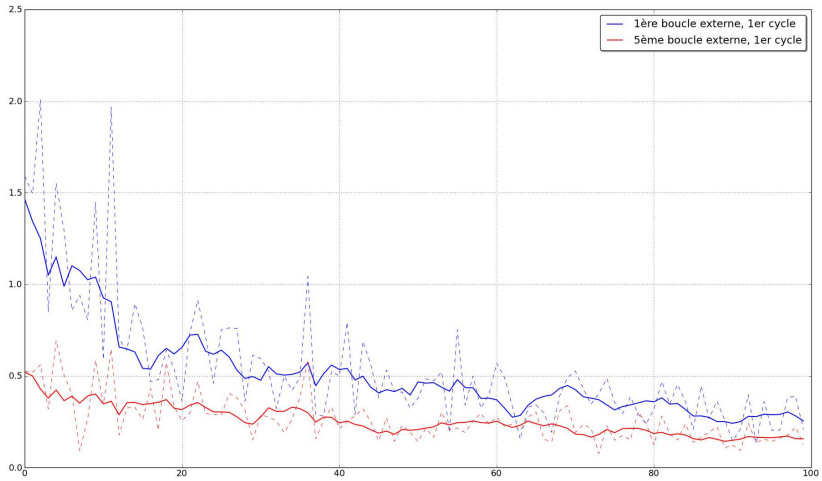

Figure 5.10 – *Idem figure 5.9 mais pour l'erreur d'analyse obtenue à la fin de la première boucle externe du premier cycle (en bleu) et celle à la fin de la cinquième boucle externe du même cycle (en rouge). Les lignes continues sont les moyennes glissantes des lignes en pointillés.*

La figure 5.10 montre les coefficients de projection correspondant aux cartes d'erreurs d'analyse à l'issue de la première boucle externe du premier cycle du régime d'assimilation permanent (correspondant aux mois 7 et 8 de l'expérience) et à l'issue de la dernière boucle externe du même cycle. Les erreurs ont été réduites sur toutes les échelles. Cependant, si l'on regarde l'écart entre les deux courbes, il apparaît que les coefficients associés aux grandes échelles sont plus fortement réduits à l'issue de la dernière boucle externe, la réduction sur les moyennes et petites échelles étant moindre. L'algorithme corrige de façon plus importante les grandes échelles.

La figure 5.11 est focalisée sur les dernières boucles externes du premier et du dernier cycle du régime d'assimilation permanent (séparés donc de 6 mois). Il vient, cette fois, que toutes les échelles ont été corrigées significativement, au fil des cycles d'assimilation.

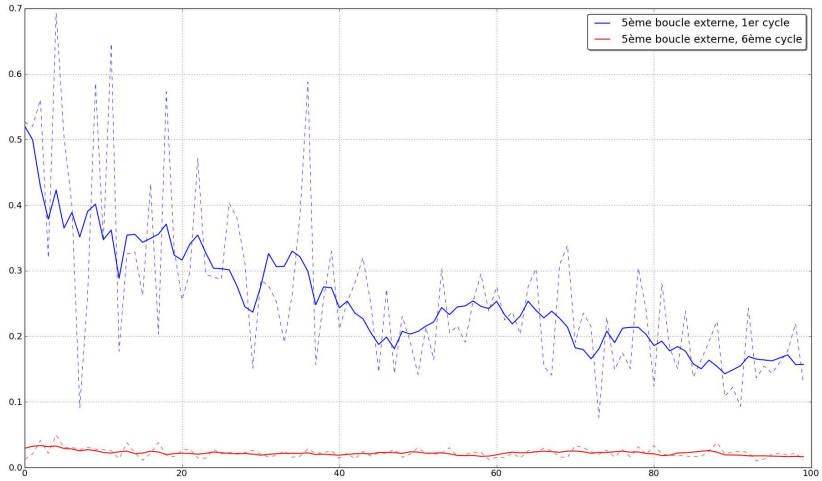

Figure 5.11 – *Idem figure 5.10 mais pour l'erreur d'analyse obtenue à la dernière boucle externe du premier cycle (en bleu) et du dernier cycle (en rouge)*

Densité spectrale de puissance de l'erreur d'analyse

La densité spectrale de puissance [1] représente la puissance d'un signal x suivant les fréquences qui le composent. Soit $\hat{f}(u,v)$ la transformée de Fourier d'un champ 2D $f(x,y)$ qui se calcule comme suit :

$$\hat{f}(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\cdot \mathrm{e}^{-i(ux+vy)}\mathrm{d}x\mathrm{d}y \tag{5.4}$$

où u et v représentent les fréquences spatiales dans les deux directions. Dans le cas d'un signal 2D continu, la PSD vaut :

$$PSD(f) = |\hat{f}(u,v)|^2 \tag{5.5}$$

Comme nous ne disposons que de signaux discrets, nous appliquons une FFT [2] à ces signaux. La FFT applique une transformée de Fourier à un ensemble de fréquences discrètes, à partir d'un signal du domaine temporel échantillonné de manière discrète sur un intervalle de temps fini. La FFT suppose que l'ensemble fini de données est une période d'un signal périodique. C'est pourquoi la nature finie du signal échantillonné peut donner un signal tronqué ayant des caractéristiques spectrales différentes du signal continu d'origine.

Appliquée à un champ spatial de SSH, la PSD permet entre autre de caractériser l'importance de chaque échelle spatiale relativement aux autres. En 2D, nous obtenons une carte des amplitudes associées à chaque fréquence spatiale, dans les deux directions (latitude et longitude). Nous réalisons ensuite une succession de moyennes radiales, moyennes effectuées sur des anneaux d'épaisseur équivalente à une maille de grille. Le résultat est alors un spectre 1D, fonction de la longueur d'onde ou de la fréquence spatiale.

Ainsi, appliqué aux champs d'erreurs d'analyse, ce diagnostic nous renseigne sur les caractéristiques des échelles spatiales des erreurs. Nous nous sommes intéressés à l'application de ce diagnostic au cours des cycles d'assimilation.

1. *Power Spectral Density* en anglais, abrégé *PSD*
2. *Fast Fourier Transform*, Transformée de Fourier Rapide

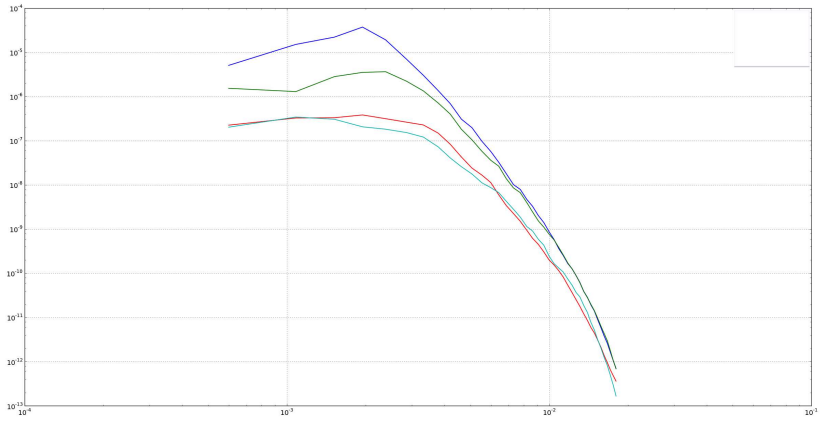

Figure 5.12 – *PSD de l'erreur d'analyse de SSH pour l'erreur d'ébauche (en bleu), l'erreur d'analyse durant le 1er cycle (en vert), le 3ème cycle (en rouge) et le sixième cycle (en turquoise). Abscisse en km^{-1}*

La Figure 5.12 montre que le 4DVAR incrémental corrige tout d'abord les plus grandes échelles, entre 1000 et 500 km. Ce sont en effet celles-ci qui sont les plus réduites à l'issue du premier cycle d'assimilation. En revanche, une bosse aux moyennes échelles dénote la concentration des structures d'erreurs entre 500 km et 200 km. Au troisième cycle, l'amplitude de l'erreur est réduite à toutes les échelles (ce qui se traduit par un spectre inférieur en tout point à celui du cycle précédent). De plus, la « bosse » s'est aplanie : les erreurs ne sont plus spécialement concentrées aux moyennes échelles. Enfin, durant les cycles suivantes, seules les erreurs aux moyennes échelles sont réduites.

5.1.3 Etude de l'erreur de prévision

Comme il est mentionné au Chapitre 2, la capacité à fournir une bonne condition initiale pour la prévision est une application principale de l'assimilation de données en océanographie. L'erreur de prévision est la différence entre la trajectoire vraie, connue dans nos expériences jumelles, et la trajectoire issue de l'initialisation du modèle par le dernier état analysé :

$$x_t^f = \mathcal{M}_{T_f \to T_f + t}(x^a(T_f))$$

où T_f est le temps final de l'analyse. Ceci émule le principe de la prévision, période durant laquelle les observations nous sont inaccessibles et où l'on ne peut compter que sur l'évolution du modèle pour simuler la réalité à venir, partant d'un état le plus en phase possible avec la réalité connue.

Les figures 5.13, 5.14 et 5.15 représentent l'erreur de prévision, respectivement de la SSH, de T et de u par rapport aux erreurs d'ébauche. Nous prenons l'amplitude du bruit blanc rajouté aux observations, d'une amplitude de $0.03\ m$, comme un ordre de grandeur du niveau d'erreur acceptable pour la prévision. Le niveau d'erreur de SSH reste en dessous de ce seuil 90 jours. Si l'on regarde les autres variables, cette période coïncide avec le maintien d'une erreur de prévision faible. Au delà de cette durée, l'erreur d'analyse dérive rapidement et rejoint l'erreur d'ébauche au bout de 6 mois.

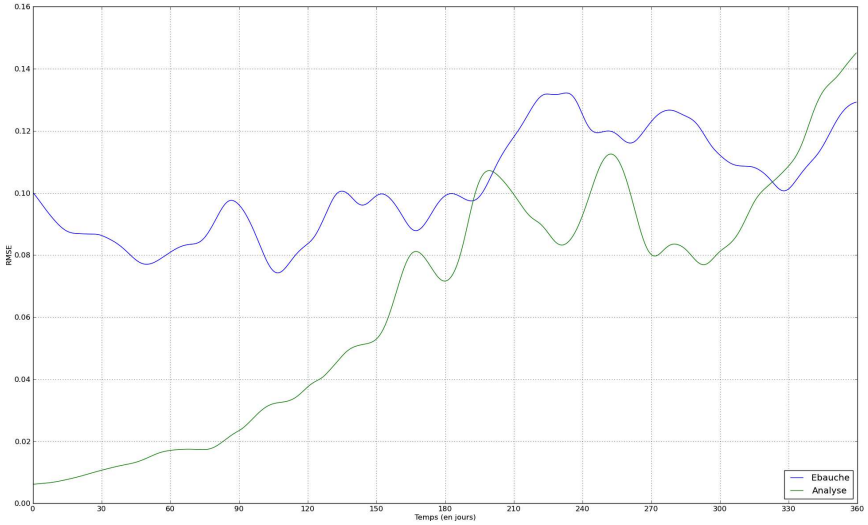

Figure 5.13 – *Évolutions des erreurs quadratiques d'ébauche et de prévision, moyennées spatialement, non normalisées (RMSSE), pour la SSH (en m), en fonction du temps (en jours). En bleu, la courbe de l'erreur d'ébauche, en vert l'erreur de prévision.*

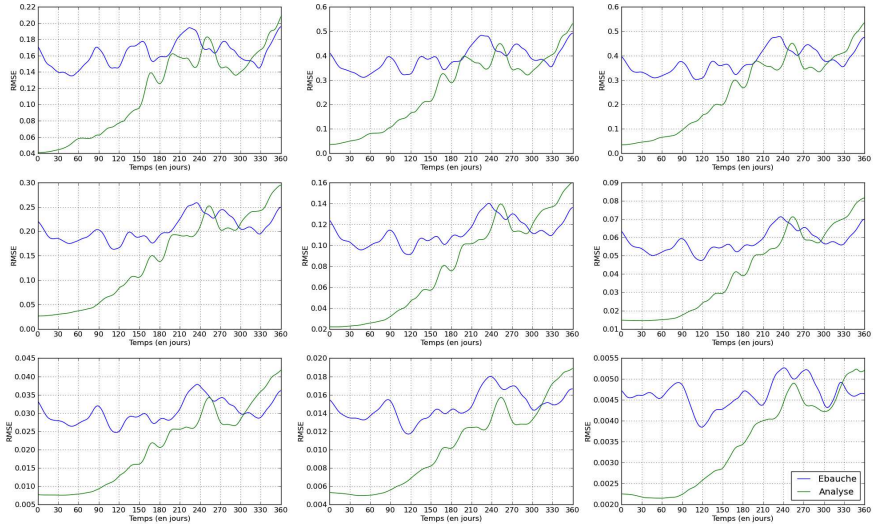

Figure 5.14 – *Évolutions des erreurs quadratiques d'ébauche et de prévision, moyennées spatialement, non normalisées (RMSSE), pour la température (en °C), en fonction du temps (en jours), du premier au neuvième niveau vertical. En bleu, la courbe de l'erreur d'ébauche, en vert l'erreur de prévision.*

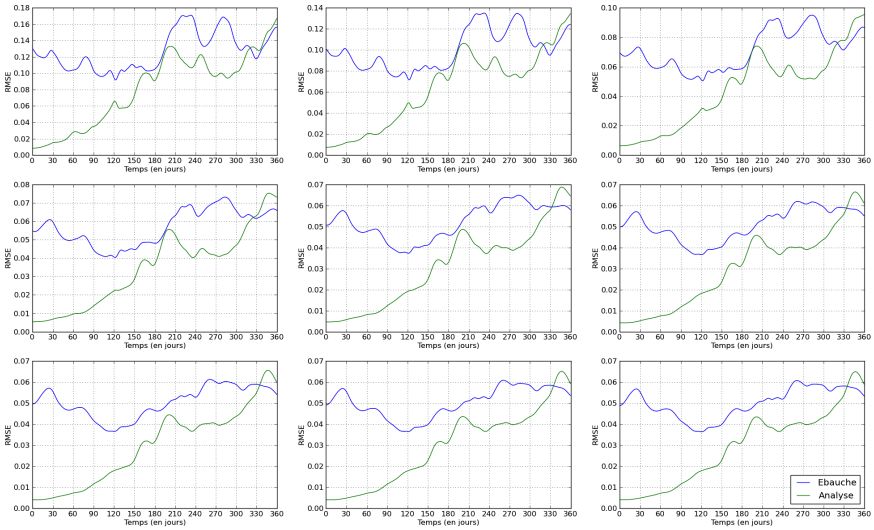

Figure 5.15 – *Idem figure 5.14, pour la vitesse zonale (en $m.s^{-1}$)*

L'algorithme a suffisamment bien convergé pour conserver une erreur de prévision faible sur plusieurs semaines.

5.1.4 Comparaison 3D-FGAT/4DVAR

Il est intéressant de comparer l'état analysé à l'issue du 4DVAR incrémental présenté ci-dessus et celui obtenu à l'issue de l'algorithme 3D-FGAT. En effet, ce dernier est aujourd'hui mis en place dans les systèmes d'AD basés sur les méthodes variationnelles dans certains centres d'océanographie opérationnels (UK-MetOffice, ECMWF), et présente l'avantage d'un faible coût numérique.

Pour ce faire, nous avons mis en place trois expériences basées sur l'algorithme 3D-FGAT testant 3 fenêtres temporelles d'AD différentes, pour une période totale d'un an : 1 jour (360 cycles), 5 jours (72 cycles), 10 jours (36 cycles). Les matrices de covariance d'erreur sont spécifiées de la même façon que décrit dans § 4.3.3. Chaque incrément est apliqué selon une méthode d'IAU uniforme, comme décrit au § 2.2.2. Pour cette raison, à chaque cycle d'assimilation, nous effectuons une seule boucle externe. En effet, comparer l'écart aux observations sur une trajectoire où l'incrément correspondant à la condition initiale est ajouté tout au long de l'intégration n'a guère de sens. C'est pourtant ce qu'il faudrait faire si l'on voulait enchaîner deux boucles externes. L'IAU contraint donc à ne réaliser qu'une seule boucle externe, ce qui est effectivement pratiqué dans les centres opérationnels utilisant cet algorithme. Au sein de chaque boucle externe, 10 itérations internes sont effectuées.

Sur la Figure 5.16, il apparaît que les fonctions sont effectivement minimisées, et la convergence de l'algorithme de minimisation se fait dès les premiers cycles.

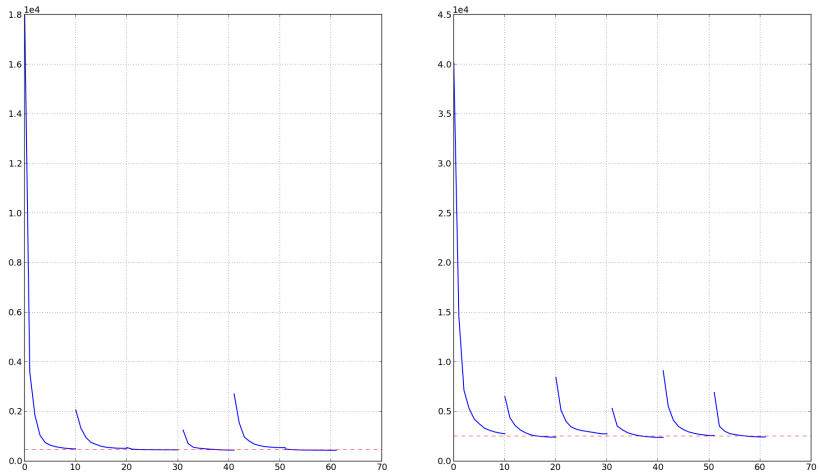

Figure 5.16 – *Évolution des fonctions coûts, en 3DFGAT, pour les 6 premiers cycles d'assimilation, pour des fenêtres de 1 jour (à gauche) et 5 jours (à gauche)*

Cependant, si l'on s'intéresse à la réduction de l'erreur d'analyse, malgré la convergence, l'algorithme paraît peu performant. Ainsi, la Figure 5.17 montre clairement que l'erreur d'analyse est bien moins réduite avec le 3DFGAT qu'avec le 4DVAR incrémental (pour une fenêtre de deux mois). La Figure 5.18 confirme cette constatation, avec des niveaux et structures d'erreurs se rapprochant de l'erreur d'ébauche (voire équivalents pour les fenêtres d'un jour en 3D-FGAT).

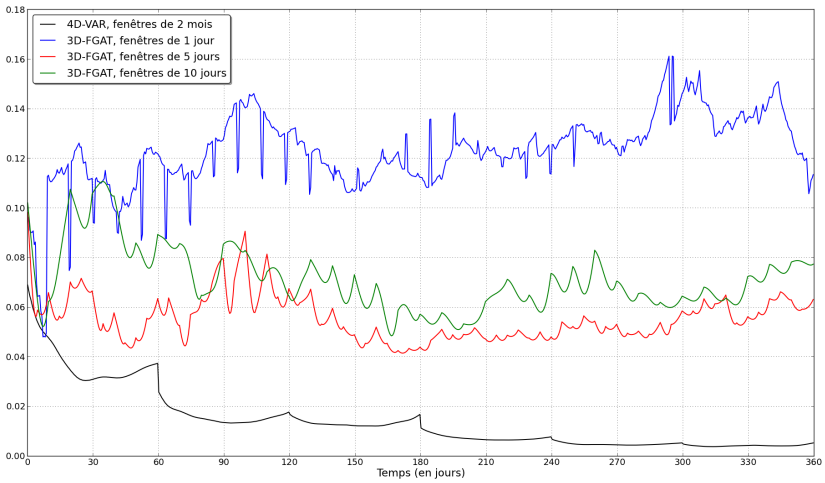

Figure 5.17 – *Comparaison de l'évolution temporelle (abscisse en jours) de la RMSSE d'analyse, pour la SSH, en 3DFGAT (fenêtres de 1 jour en bleu, 5 jours en vert, 10 jours en rouge) et en 4DVAR, pour des fenêtres d'AD de 2 mois (en noir).*

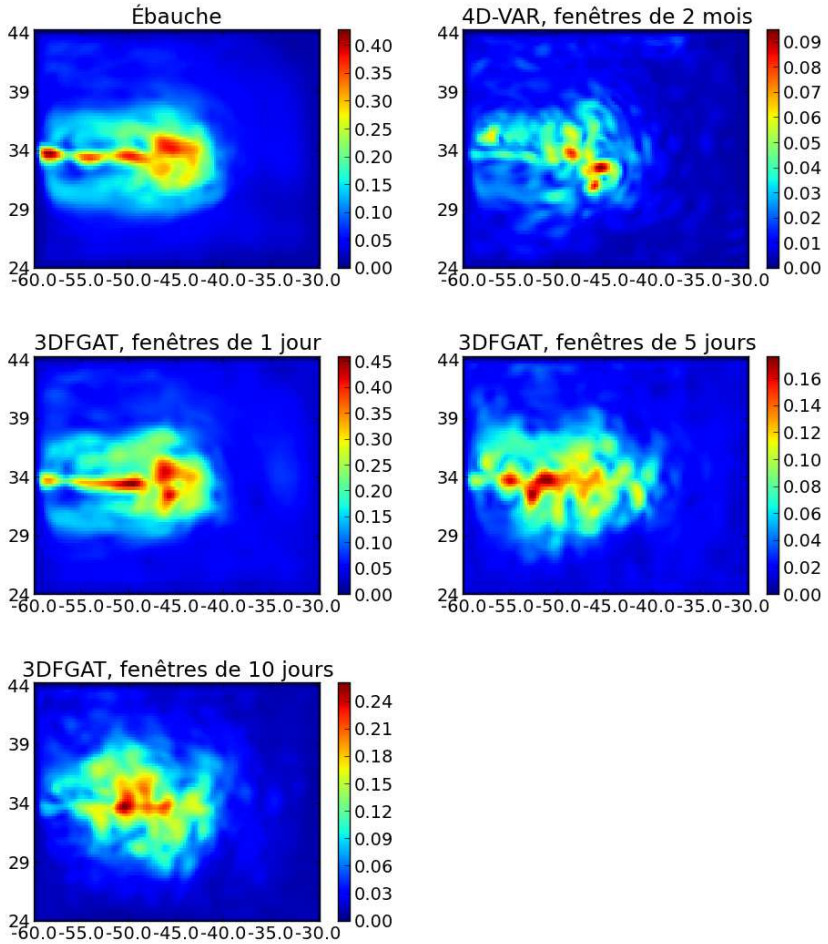

Figure 5.18 – *Comparaison de la RMSTE de SSH, de l'ébauche (en haut à gauche), de l'analyse en 3DFGAT (fenêtres de 1, 5 et 10 jours) et en 4DVAR (fenêtres de 2 mois), respectivement de gauche à droite et de haut en bas*

Il vient aussi que la fenêtre la plus pertinente pour le 3DFGAT est celle de cycles de 5 jours. Dans notre configuration, les observations ne sont pas nombreuses, ce phénomène s'accentuant évidemment avec des fenêtres plus courtes. Ceci explique la faible correction de l'erreur pour des fenêtres de 1 jour. On observe également des pics réguliers, espacés de 10 jours, période correspondant à celle de répétitivité du satellite simulé ici.

Si l'on regarde la Figure 5.19, il est clair que la correction en profondeur par un incrément obtenu via l'algorithme du 3D-FGAT n'est pas satisfaisante. Ne disposant que des informations sur la surface, le 3DFGAT ne propage pas correctement l'information sur la dimension verticale, au contraire du 4DVAR. Ainsi, pour ce dernier, l'information de surface portée par les observations altimétriques permet de réduire également l'erreur en profondeur, y compris sur les variables non-observées.

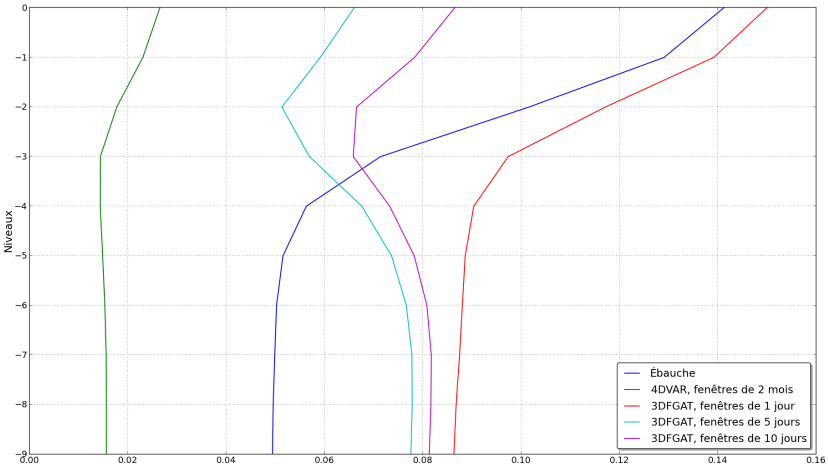

Figure 5.19 – *Comparaison des profils verticaux de RMSTE de vitesse zonale, moyennée sur chaque couche, de l'ébauche (en bleu), de l'analyse en 3DFGAT et en 4DVAR*

Pour des fenêtres temporelles trop longues, comme celles de 10 jours, l'approximation des modèles linéaires tangent et adjoint par l'identité est mise à mal. Ainsi, la réduction de l'erreur va décroître avec l'augmentation de la fenêtre d'AD.

La fenêtre de 5 jours représente ainsi un compromis entre le nombre d'observations disponibles et la validité de l'approximation faite par l'algorithme du 3D-FGAT.

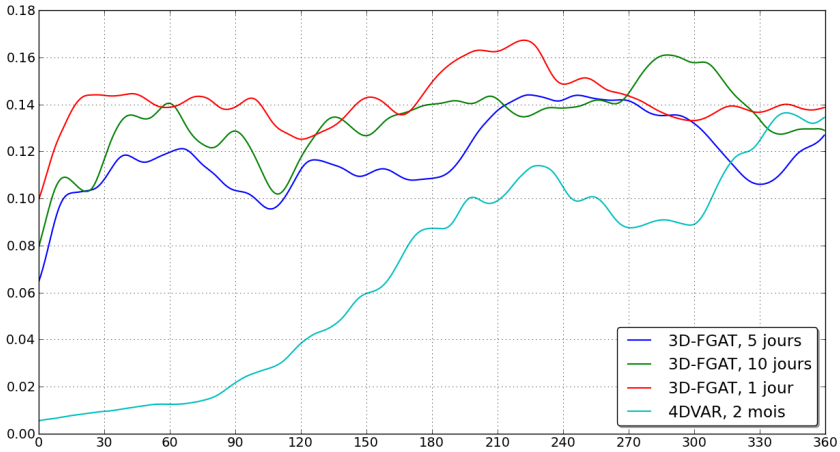

Figure 5.20 – *Comparaison de l'évolution temporelle (abscisse en jours) de la RMSSE de prévision, pour la SSH, en 3DFGAT et en 4DVAR*

Les mêmes conclusions peuvent être tirés pour l'erreur de prévision, comme l'illustre la Figure 5.20.

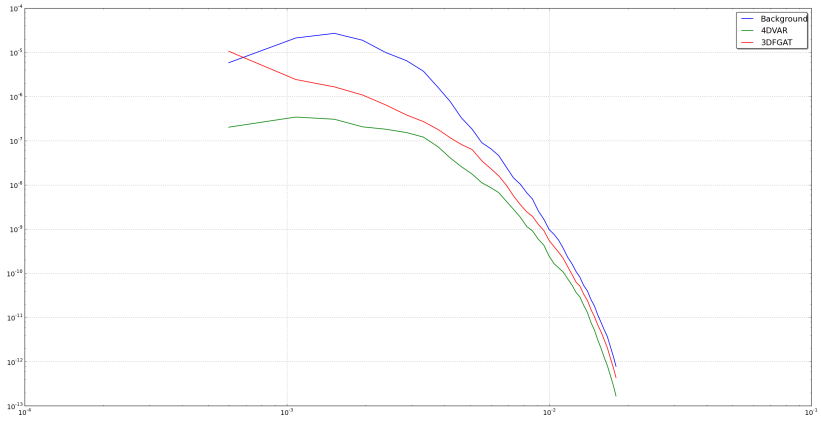

Figure 5.21 – *PSD de l'erreur d'ébauche (en bleu) de SSH et d'analyse de SSH (abscisse en cycle/km), pour le 3DFGAT (en vert) et le 4DVAR (en rouge), pour le dernier mois d'assimilation*

Sur la Figure 5.21, reprenant le diagnostic basé sur la PSD exposé plus haut, il est clair que le 4DVAR présente des erreurs plus faibles à toutes les échelles et de façon légèrement plus marquée aux grandes échelles.

Synthèse L'algorithme du 3D-FGAT, aujourd'hui utilisé dans certains centres opérationnels d'océanographie paraît nettement moins performant que le 4DVAR incrémental, dans le contexte de nos expériences. Dans le meilleur des cas, l'erreur d'analyse est une dizaine de fois supérieure à celle obtenue en 4D-VAR, et l'erreur de prévision rejoint la variabilité du modèle au bout de quelques jours seulement.

Le coût de calcul du 4D-VAR est de l'ordre de dix fois supérieur à celui du 3D-FGAT, mais il apparaît qu'il apporte une réelle plus-value sur la qualité des résultats à l'issue du processus d'AD.

5.2 Sensibilité des performances du système d'AD aux différents paramètres de minimisation en contexte non-linéaire

Les performances du 4D-VAR incrémental dans un contexte non-linéaire dépendent évidemment de la paramétrisation détaillée en 4.3.3. Dans le cadre d'expériences jumelles, nous pouvons supposer que cette paramétrisation dépend essentiellement de la connaissance que l'on a du modèle, auquel on a évidemment ici accès et que l'on suppose parfait. Cependant, d'autres paramètres de l'algorithme qui vont nous intéresser ici sont plus difficiles à spécifier a priori.

En premier lieu, la longueur de la fenêtre d'assimilation joue un rôle primordial sur la capacité de l'algorithme à contrôler efficacement la condition initiale. Comme étudié par Luong et al. (1998), cette longueur doit être suffisante pour permettre de laisser se développer les différents phénomènes physiques captés par les observations. Cependant, cette longueur est contrainte par la validité de l'hypothèse linéaire tangente. Cette hypothèse peut être mise à mal par deux contraintes :

- Sur de longues périodes d'intégration, la propagation de la perturbation par le modèle linéaire tangent s'éloignera de celle par le modèle non-linéaire ;
- Si $\mathbf{x}_0^b$ est éloigné de $\mathbf{x}_0^t$, le développement de Taylor au premier ordre qui la sous-tend ne sera plus valable.

Ainsi, dans un premier temps, à paramètres de minimisation équivalents (même paramétrisation de NEMOVAR, même nombre de boucles externes et d'itérations internes), sur une même période totale d'assimilation (*i.e.* 1 an), nous comparerons cinq expériences avec différentes longueurs de cycle d'assimilation : 15 jours, 1 mois, 2 mois, 3 mois et 4 mois.

Le coût de calcul d'une expérience d'assimilation variationnelle étant tout à la fois élevé et critique pour une application réaliste, nous avons fixé les nombres de boucles externes et d'itérations internes, invariables tout au long des expériences. Nous verrons que la performance du système d'AD est clairement dépendante du nombre d'itérations de minimisation dans le cas où l'ébauche est éloignée de l'état vrai (*i.e.* au début des expériences), à coût de calcul constant. Ainsi, selon la confiance que l'on a dans l'ébauche, ce paramètre devra être ajuster. Cependant, la convergence numérique de l'algorithme n'étant pas envisageable (à précision machine), nous fixons le critère du coût de calcul.

5.2.1 Longueur de la fenêtre temporelle d'assimilation

Avec les mêmes diagnostics qu'au §5.1, nous étudions ici, sous différents angles, la réaction du système d'AD à la longueur temporelle de fenêtre d'assimilation.

Longueur de fenêtre et convergence de l'algorithme La première constatation tirée de la figure 5.22 est la dépendance entre la vitesse de convergence et la longueur de la fenêtre d'assimilation. Pour des fenêtres courtes (15 jours et 1 mois), le régime transitoire d'assimilation évoqué au §5.1.1 est plus court, le plateau de convergence est atteint plus rapidement en terme de temps d'intégration du modèle. Plusieurs points favorisent une convergence plus rapide pour des périodes d'assimilation courtes :

- Dans le cas où le modèle et/ou l'opérateur d'observation sont non-linéaires, l'allongement de la fenêtre temporelle dégrade la fonction coût non-linéaire en terme de convexité et du

nombre de minimums locaux, comme illustré au §2.3.3

- La propagation d'une perturbation (*i.e.* un incrément) via le modèle linéaire tangent s'écarte de celle via le modèle non-linéaire avec l'allongement de la période d'assimilation

Quand nous explorons de plus longues fenêtres (3 mois et 4 mois), la convergence est beaucoup plus lente et n'est pas satisfaisante pour les paramètres de minimisation que nous nous sommes fixés.

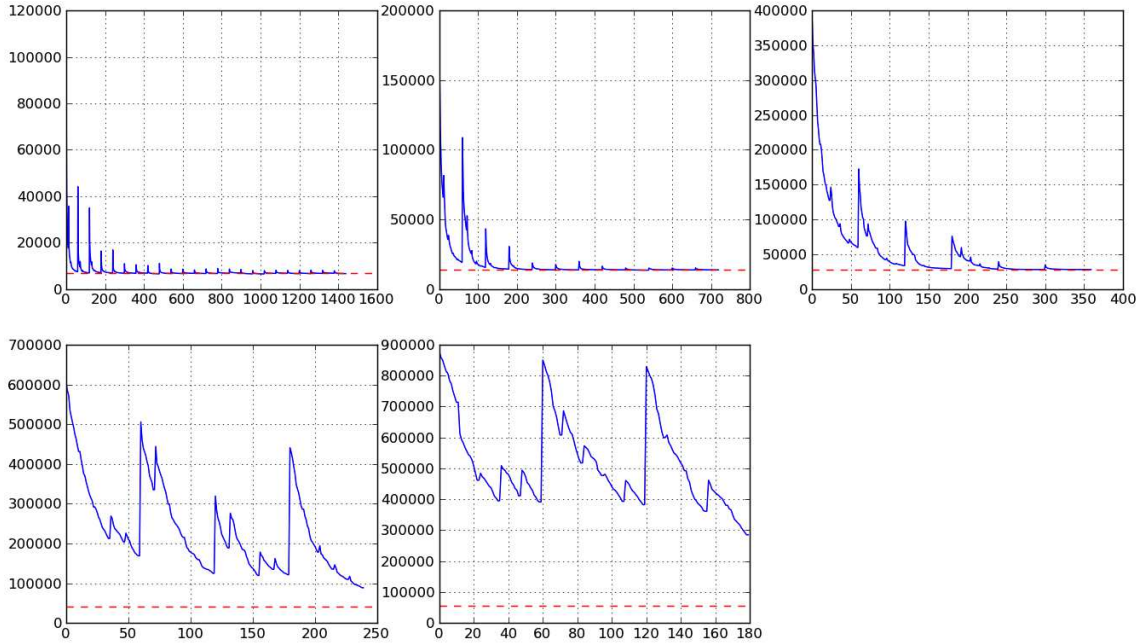

Figure 5.22 – *Évolution des fonctions coûts pour des expériences avec différentes longueurs de fenêtres, respectivement de gauche à droite et de haut en bas : 15 jours, 1 mois, 2 mois, 3 mois et 4 mois. En abscisse, le nombre d'itérations cumulées. En pointillés rouges, nous indiquons $p/2$ où p est le nombre moyen d'observations disponibles pour chaque cycle.*

Erreur d'analyse Si l'on s'intéresse à l'erreur d'analyse, à travers les figures 5.23 à 5.26, on voit que l'algorithme ne réduit pas efficacement cette erreur pour des périodes d'assimilation trop grandes. Ceci est consistant avec la divergence observée plus haut sur la minimisation des fonctions coûts pour ces mêmes expériences.

A contrario, les expériences aux fenêtres plus courtes réduisent bien les niveaux d'erreur. Notons malgré tout les différences notables entre les trois expériences (15 jours, 1 mois et 2 mois) aux niveaux d'erreurs comparables :

- Si l'on s'intéresse aux niveaux d'erreur en profondeur sur les variables (non-observées) de vitesse et de température, ceux-ci sont plus faibles pour les fenêtres plus longues (2 mois). L'information portée sur la surface par la SSH et transportée par le modèle direct, adjoint et la matrice B pénètre mieux à la verticale sur une plus longue période. Ainsi si la différence entre l'expérience d'1 mois et de 2 mois n'est pas flagrante sur la variable observée, en surface, elle est plus nette sur les variables non-observées dans l'océan profond.
- A niveaux d'erreurs similaires, en comparant les expériences de 15 jours, 1 mois et 2 mois, sur les figures 5.23 à 5.26, la trajectoire analysée via des cycles de longues périodes d'assimilation est mieux filtrée, évitant mécaniquement la trop grande répétition des sauts inter-cycles. Qui plus est, en prenant mieux en compte les différents processus à diffé-

rentes échelles temporelles, ces sauts sont également moins marqués. Par contre, le régime transitoire d'assimilation s'étale sur une plus longue période de temps, comme le montrent les figures référencées ci-dessus ainsi que la figure 5.22.

– La convergence de l'algorithme de minimisation (*cf.* Figure 5.22) n'est pas l'unique indicateur de la performance de l'algorithme. En effet, nous voyons que sur toutes les variables, observées ou non, la fenêtre de 2 mois assure des niveaux d'erreur au pire équivalents, au mieux meilleurs, sur chaque couche horizontale. Mais si l'on revient à l'évolution des fonctions coûts, les plus courtes fenêtres accéléraient l'algorithme de minimisation.

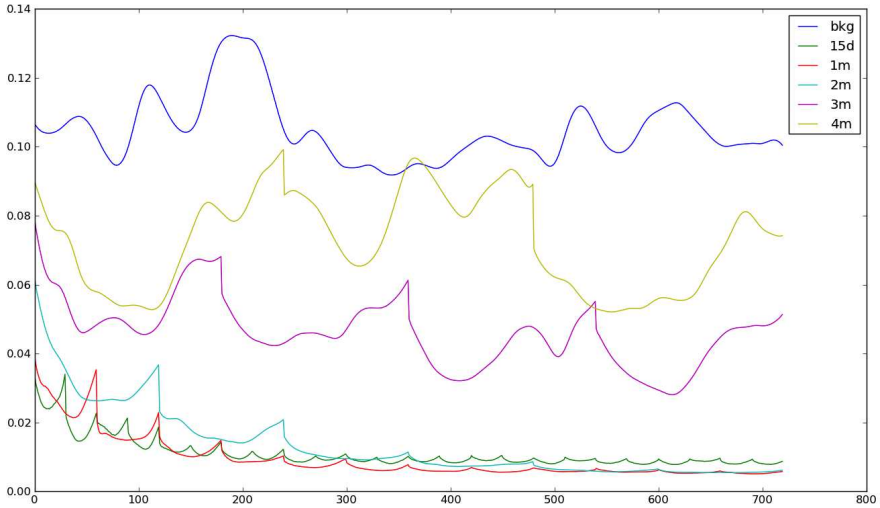

Figure 5.23 – *Évolutions des erreurs quadratiques d'ébauche et d'analyse, moyennées spatialement, non normalisées ($RMSSE(h)$), pour la SSH (en m), en fonction du temps (en jours).*

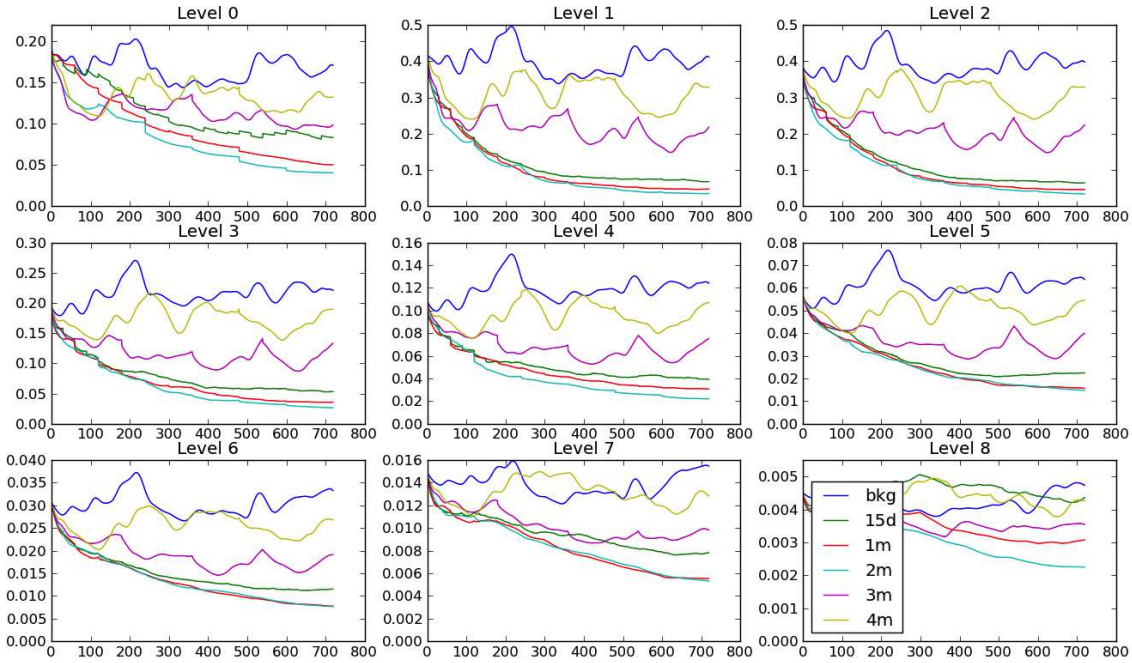

Figure 5.24 – *Idem figure 5.23 pour la température (en °C) ($RMSSE(T)$)*

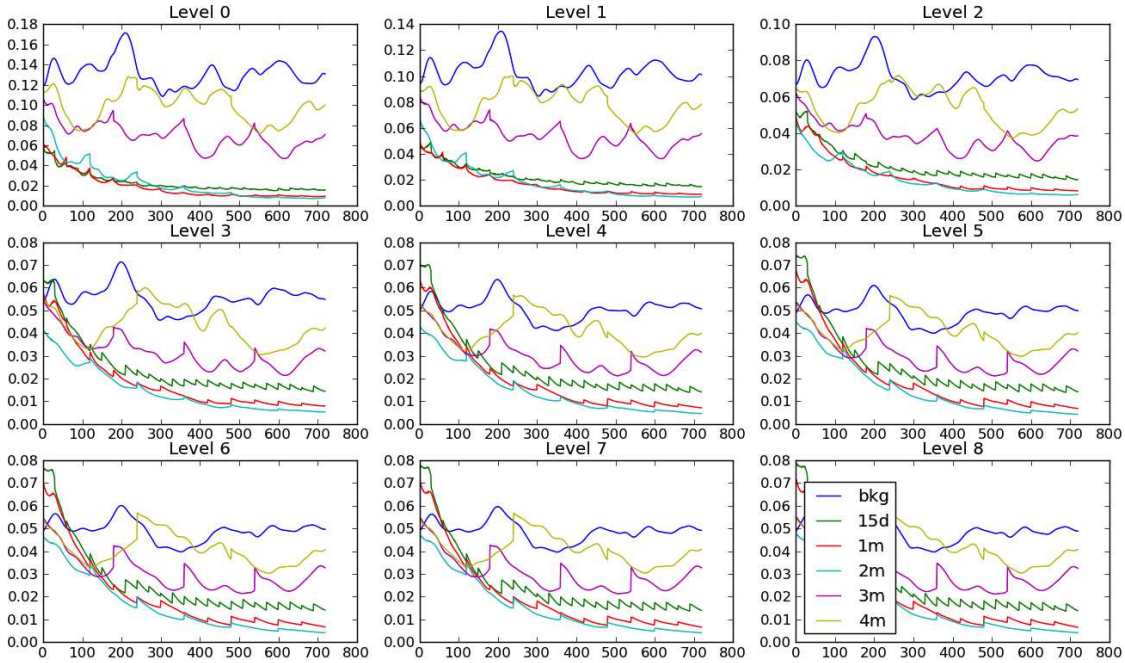

FIGURE 5.25 – *Idem figure 5.23 pour la vitesse zonale (en $m.s^{-1}$) ($RMSSE(u)$)*

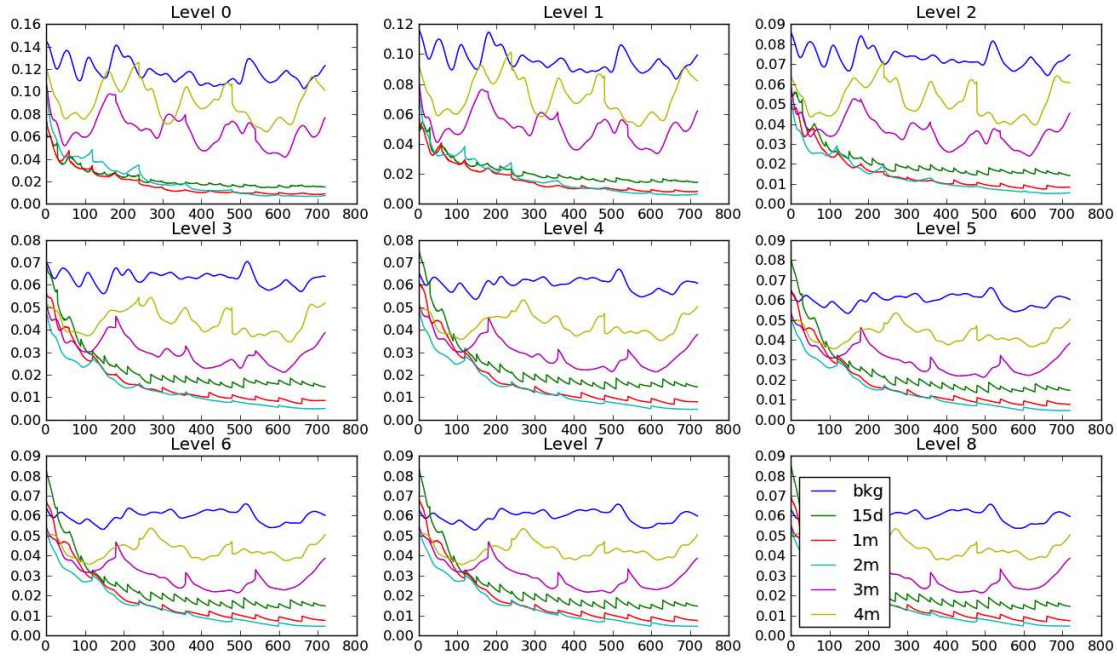

FIGURE 5.26 – *Idem figure 5.23 pour la vitesse méridienne (en $m.s^{-1}$) ($RMSSE(v)$)*

Les champs de $RMSTE(h)$ de la figure 5.27 confirment la faible diminution de l'erreur d'analyse pour les trop longues fenêtres d'assimilation. Les structures d'erreurs ressemblent à celles observées sur l'erreur d'ébauche. En revanche, pour les courtes fenêtres, les structures d'erreurs comportent essentiellement des petites échelles. Sur le champ de l'expérience de 15 jours, l'empreinte des traces satellites simulées est clairement visible. Puis elle disparaît pour les fenêtres

plus longues. Ceci traduit encore la diffusion plus efficace de l'information portée par ces observations pour de longues périodes.

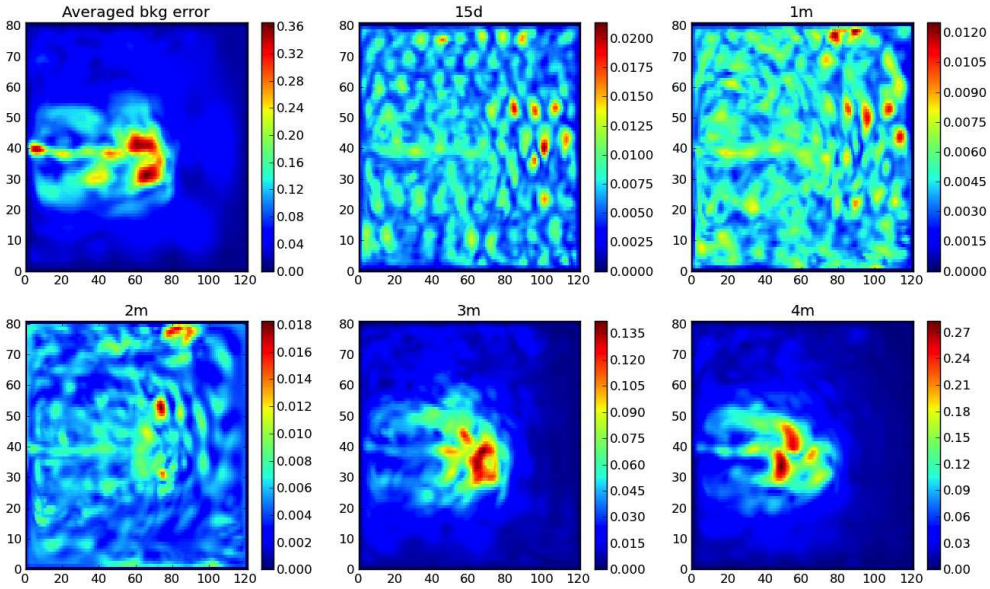

Figure 5.27 – *Champ spatial de l'erreur quadratique de SSH (en m) moyennée temporellement ($RMSTE(h)$), sur les six derniers mois de l'expérience d'assimilation. Les plages de couleurs sont différentes afin d'identifier les structures d'erreurs dans tous les cas.*

Erreur de prévision En ce qui concerne l'erreur de prévision, les résultats sont comparables à ceux évoqués pour l'erreur d'analyse. La figure 5.28 confirme les performances des fenêtres de 15 jours, 1 mois et 2 mois par rapport aux deux plus longues fenêtres. Là encore, nous pouvons remarquer que la fenêtre de 2 mois maintient une plus faible erreur de prévision, de l'ordre de l'erreur d'observation, 2 mois de plus que les fenêtres de 15 jours et 1 mois. Une plus longue période permet de contrôler plus efficacement les modes instables à partir desquels se développent les erreurs de prévisions (Pires et al., 1996).

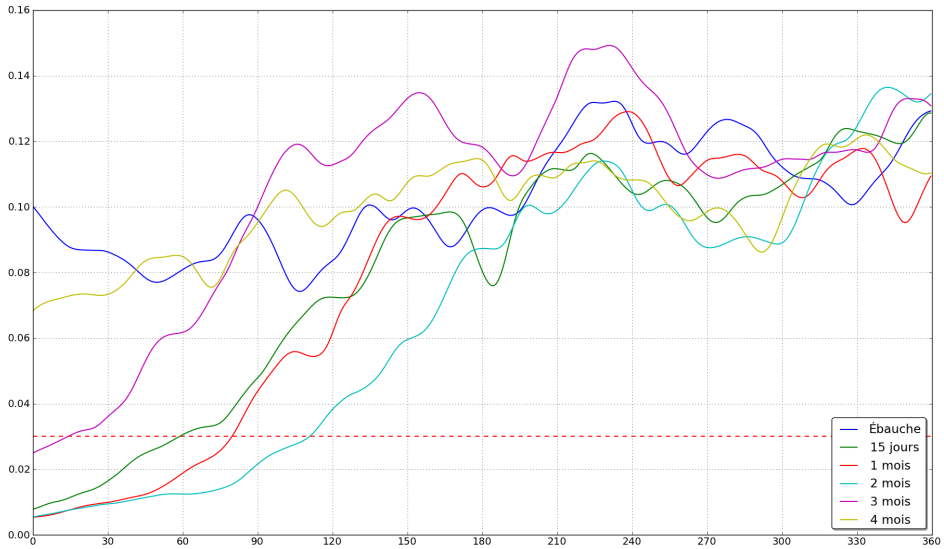

Figure 5.28 – *Évolutions des erreurs quadratiques d'ébauche et de prévision, moyennées spatialement, non normalisées, pour la SSH (en m), en fonction du temps (en jours). En pointillés rouges, l'amplitude du bruit blanc rajouté aux observations.*

Synthèse Dans le cas d'un modèle non-linéaire, la longueur temporelle de la fenêtre d'assimilation doit être le résultat d'un compromis entre plusieurs besoins contradictoires. Elle doit être suffisamment longue pour :

- Permettre à l'information portée par les observations de se propager sur les variables non-observées, y compris sur la verticale dans le cas présent, via le modèle direct et adjoint et la matrice $\mathbf{B}$.
- Contrôler efficacement les modes instables afin de réduire l'erreur de prévision…
- … ce qui peut se voir comme un contrôle des interactions des différents processus se développant dans le modèle, à des échelles temporelles différentes.

Mais elle doit être suffisamment courte pour :

- Éviter la multiplication des minimums locaux de la fonction coût
- Ne pas trop dégrader le conditionnement de la fonction coût…
- … afin d'obtenir un niveau de convergence acceptable de l'algorithme de minimisation à coût de calcul donné.

La spécification de cette période est liée, en premier lieu, au « degré » de non-linéarité du modèle considéré.

5.2.2 Influence du nombre d'itérations de boucle interne

Motivation Une étape nécessaire dans la conception d'une expérience d'AD variationnelle de type 4DVAR incrémental, appliquée à un modèle non-linéaire, est de spécifier le nombre d'itérations de boucles externes et d'itérations internes. En effet, partant d'une ébauche éloignée de l'état vrai, il n'y a pas de garantie a priori de convergence vers le minimum global de la fonction coût non-quadratique. En pratique, non seulement nous n'avons pas cette garantie, mais à cela se rajoute la difficulté de la convergence numérique de la minimisation des fonctions coût quadratiques. Pour des problèmes de grande taille, le coût de calcul est prohibitif si l'on veut converger

à la précision numérique, étant donné le mauvais conditionnement de la fonction coût. Pour spécifier correctement ces paramètres, il convient d'étudier la sensibilité du système à ceux-ci.

Étude de convergence Dans le cas de notre système d'assimilation de données, en partant d'une ébauche éloignée de l'état vrai, s'il est vrai que le nombre de boucles externes n'affecte la convergence qu'en la tronquant, et éventuellement en allongeant le régime transitoire d'assimilation, il n'en est pas de même pour le nombre d'itérations de boucle interne. Nous avons repris les expériences d'AD de périodes 15 jours, 1 mois et 2 mois étudiées au §5.2, en prenant cette fois-ci 30 itérations de boucle interne au lieu de 10.

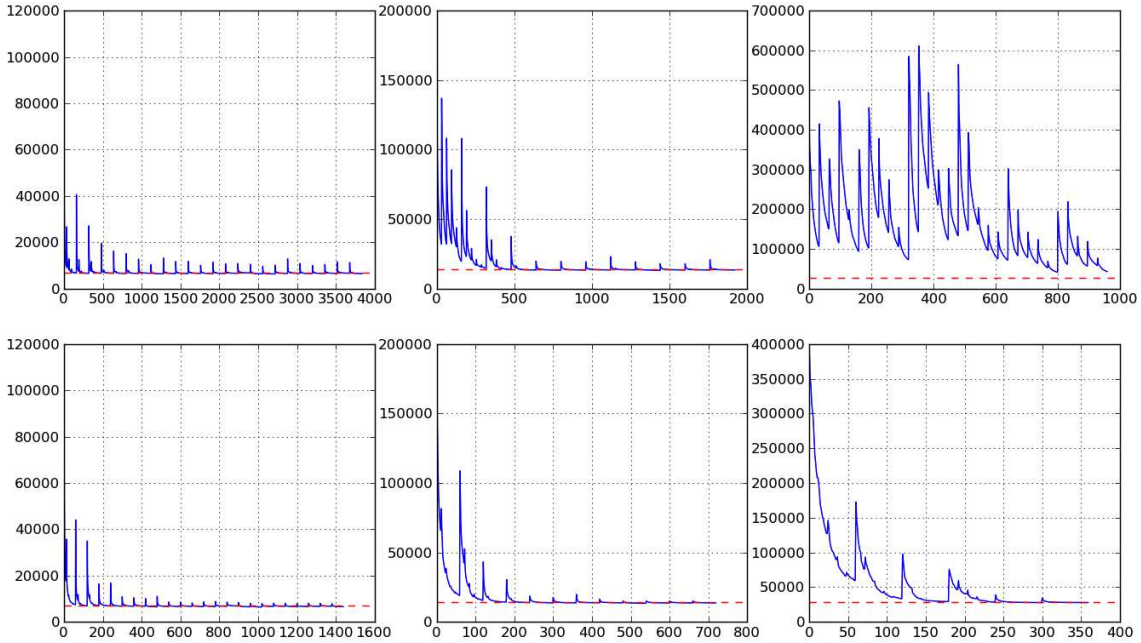

Figure 5.29 – *Évolution des fonctions coûts pour des expériences avec différentes paramétrisations. En haut : chaque cycle comporte 5 boucles externes et 30 itérations de boucles internes. En bas, 5 boucles externes et 10 itérations de boucles internes. De gauche à droite, la fenêtre d'assimilation vaut respectivement 15 jours, 1 mois et deux mois.*

Sur la figure 5.29, il apparaît que pour les fenêtres de 15 jours et d'un mois, la minimisation se fait de façon comparable quel que soit le nombre d'itérations internes. Cependant, il est à noter que le régime transitoire d'assimilation, de façon plus visible sur la fenêtre d'un mois, présente dans le cas de 30 itérations internes des sauts importants de la fonction coût à l'interface des boucles externes sur le premier cycle, sauts se propageant également au début des deux cycles suivants, pour enfin disparaître et converger de façon semblable à l'expérience de référence.

En revanche, pour la fenêtre de 2 mois, ces sauts ne s'atténuent pas de cycles en cycles et rendent la minimisation très difficile sur la période totale d'assimilation, ce qui n'est pas le cas en prenant un nombre restreint d'itérations internes.

Une interprétation de ce phénomène a été proposée par Trémolet (2007a). Plus le nombre de boucles internes est élevé, plus la propagation linéaire de la perturbation contrôlée va prendre de l'importance. En pratique, lorsque l'ébauche est éloignée de l'état vrai, l'incrément voit son amplitude augmenter au fur et à mesure des boucles internes. Ceci, couplé à une longue période d'intégration, va donc faire croître l'écart entre la propagation de la perturbation via le modèle

linéaire tangent et via le modèle non-linéaire. Au passage à la boucle externe suivante, ceci aura pour effet d'écarter significativement la nouvelle trajectoire des observations. Et la boucle interne compensera alors ce nouvel écart en produisant un incrément de grande amplitude, ce qui peut éventuellement faire diverger l'algorithme.

Sur les fenêtres plus courtes, deux phénomènes contiennent la divergence entre le modèle linéaire tangent et le modèle non-linéaire :

- La courte durée de la fenêtre va limiter la divergence de la propagation de la perturbation
- le nombre d'observations ainsi réduit va mécaniquement diminuer l'amplitude de la pertubation contrôlée

5.3 Conclusion

Nous avons vu au cours de ce chapitre que le 4DVAR incrémental, appliqué à un bassin océanique au 1/4°, donne de bons résultats en terme de réduction d'erreur d'analyse et de prévision pour des fenêtres longues d'un et deux mois.

Il est apparu qu'au cours d'une période d'AD, le 4DVAR incrémental corrige en premier lieu les grandes échelles de l'écoulement (supérieures à 500 km) puis se concentre sur les moyennes échelles (inférieures à 500km).

Le niveau d'erreur obtenu avec le 4DVAR est entre dix et quinze fois inférieur à celui obtenu en 3DFGAT.

Nous avons également mis en évidence la sensibilité du système d'AD à la paramétrisation de l'algorithme de minimisation et à la longueur de la fenêtre d'AD.

Chapitre 6

Influences de l'intensification des non-linéarités sur le 4D-VAR incrémental

Dans ce chapitre, nous explorons le comportement des méthodes variationnelles d'AD de façon analogue au chapitre précédent, mais appliquées à SEABASS au 1/12°. Notre objectif est ainsi de mettre en évidence les implications de l'augmentation de la résolution spatiale et donc des non-linéarités, qualitativement et quantitativement sur les algorithmes d'AD variationnels. Notre but est alors d'exposer les contraintes dues aux non-linéarités et de proposer des pistes d'amélioration du système d'AD pour y faire face.

Dans une première partie, notre objectif est d'analyser sous différents contextes les performances du système d'AD variationnelle lorsque l'on augmente la résolution du modèle. Nous comparerons tout d'abord une expérience d'AD appliquée à SEABASS12 par rapport aux expériences réalisées sur SEABASS4, afin de dégager les principaux effets de l'augmentation de la non-linéarité sur le système d'AD. Nous verrons ensuite la sensibilité de ses performances en fonction de différentes caractéristiques, telles que la longueur de la fenêtre d'assimilation ou le nombre d'itérations de minimisation. Une comparaison entre le 4DVAR et le 3DFGAT sera également présentée.

Dans un second temps, nous explorerons les pistes permettant d'améliorer les performances de ce système d'AD variationnelle pour la haute résolution, tout d'abord en s'intéressant à l'influence de la qualité de l'ébauche par rapport à l'état vrai puis en considérant différents échantillonnages spatio-temporels du réseau d'observation. La stratégie d'assimilation progressive, pouvant être considérée comme une amélioration méthodologique, sera enfin présentée.

6.1 Analyses d'expériences pour SEABASS au 1/12°

6.1.1 Comparaison SEABASS4 vs SEABASS12

Après avoir vu au chapitre précédent la bonne qualité de la trajectoire analysée produite par le 4DVAR incrémental avec le modèle SEABASS au 1/4°, nous voulons maintenant voir comment ces performances se préservent ou se dégradent lorsque la résolution, et donc la non-linéarité, du modèle augmente. Dans un premier temps, nous considérons une expérience basée sur SEABASS12, décrite par les caractéristiques suivantes :

- Les observations sont des traces altimétriques simulées, de type Jason-1, auxquelles est rajouté un bruit blanc d'une amplitude de 3 cm.

- Les fenêtres temporelles d'assimilation sont d'un mois.
- Chaque cycle d'assimilation comporte 6 boucles externes et 10 itérations internes de minimisation

Nous comparons cette expérience à celle similaire réalisée au 1/4°, présentée dans le chapitre précédent.

Si l'on s'intéresse d'abord à la convergence, il apparaît sur la Figure 6.1 que d'ores et déjà l'algorithme est beaucoup moins efficace au 1/12°. Alors qu'au 1/4°, la fonction coût atteignait la valeur théorique de convergence au bout du 4ème cycle, elle n'est jamais approchée de façon consistante au cours de l'assimilation.

De même, si les sauts de valeur de la fonction coût à l'interface des cycles d'assimilation diminuent de façon monotone au 1/4°, il n'en est rien pour le 1/12°. Cependant, l'algorithme semble effectivement améliorer la convergence de cycles en cycles.

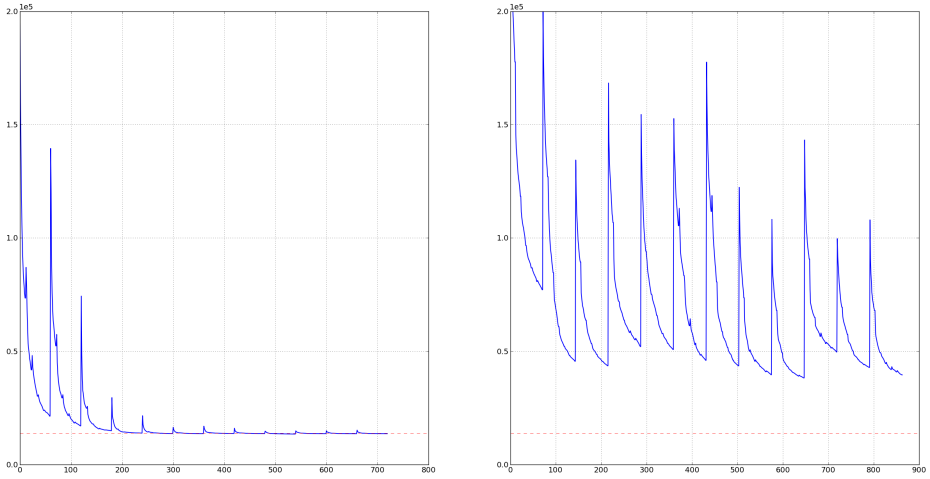

Figure 6.1 – *Évolution des fonctions coûts au cours de la période d'assimilation (itérations de minimisation en abscisse) pour deux expériences similaires (traces altimétriques simulées de type Jason-1, fenêtres d'assimilation de 1 mois) au 1/4° (à gauche) et au 1/12° (à droite). En pointillés rouges, la valeur moyenne théorique de convergence* $p/2$.

Si l'on s'intéresse à l'évolution temporelle d'erreur d'analyse de SSH, moyennée spatialement, cet écart de qualité de convergence entre les deux résolutions apparaît également. Sur la Figure 6.2, l'erreur d'analyse présente un niveau moyen de 1 cm au 1/4° pour un niveau moyen de 5 cm au 1/12°. L'amélioration par rapport à l'ébauche reste toutefois très nette, puisque l'erreur d'ébauche moyenne était de l'ordre de 14 cm.

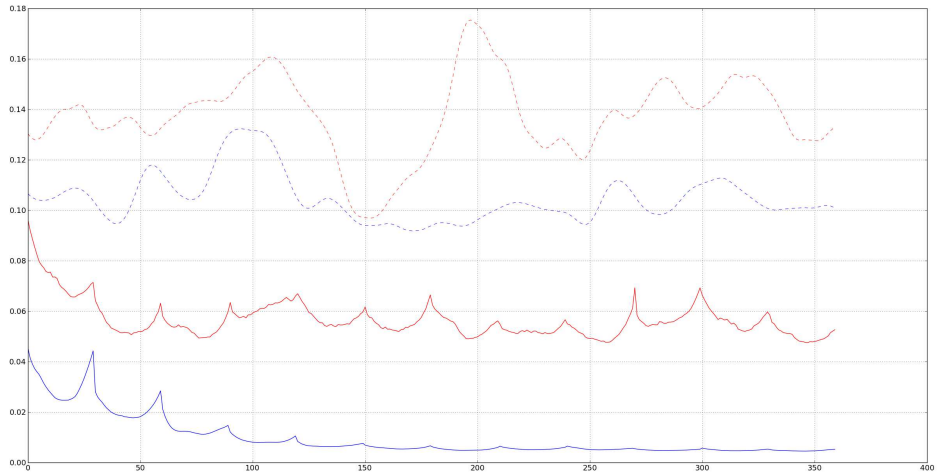

Figure 6.2 – *Évolution temporelle (en jours) de la RMS d'analyse de SSH (en trait plein) et d'ébauche (en pointillés), au 1/4° (en bleu) et au 1/12° (en rouge)*

Enfin, en regardant la stucture spectrale de l'erreur d'analyse au fil des cycles sur la Figure 6.3, toujours en comparant l'expérience au 1/4° et celle au 1/12°, elle se réduit de manière similaire au fur et à mesure des cycles d'assimilation : les structures d'erreurs de plus grandes échelles sont celles corrigées le plus tôt dans la séquence d'AD, puis l'algorithme s'attache à réduire les structures de plus petites échelles. Seule l'amplitude moindre de la réduction au fil des cycles d'AD au 1/12° indique que l'algorithme rencontre plus de difficultés à cette résolution.

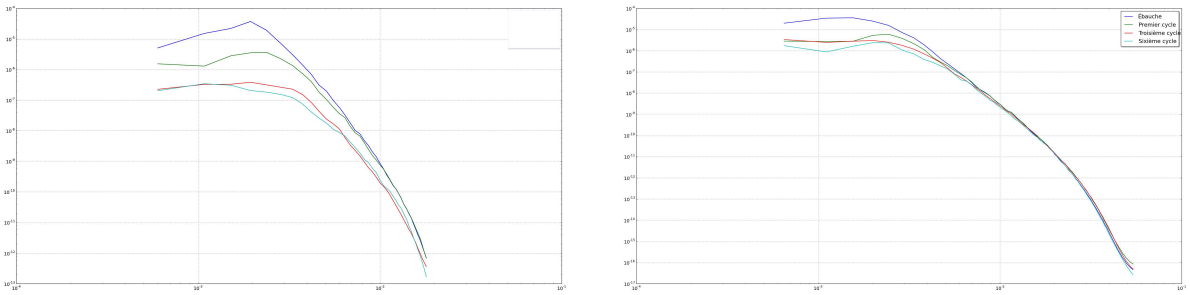

Figure 6.3 – *PSD de l'erreur de SSH d'ébauche (en bleu) et d'analyse durant le 1er cycle (en vert), le 3ème cycle (en rouge) et le sixième cycle (en turquoise), pour une expérience d'AD avec SEABASS4 (à gauche) et SEABASS12 (à droite)*

Ainsi, l'augmentation de la résolution du modèle a un effet clairement négatif sur la qualité de convergence de l'algorithme de minimisation. Cependant, malgré cela, la trajectoire analysée est quand même contrôlée avec un niveau d'erreur moyen acceptable. Nous allons voir dans la suite quels sont les impacts des paramètres du système d'AD sur ses performances au 1/12°.

6.1.2 Rôle de la longueur de fenêtre d'AD

Dans un premier temps, nous étudions la qualité de convergence de l'algorithme pour différentes longueurs de fenêtres d'assimilation. Pour ce faire, nous avons réalisé trois expériences jumelles, basées sur des observations altimétriques simulées, d'échantillonage de type *Jason-1*,

avec trois longueurs de fenêtres différentes : 5, 10 et 30 jours. Chaque cycle d'assimilation, pour ces trois expériences, est composée de 6 boucles externes chacune comportant 10 itérations internes.

Sur la Figure 6.4, il apparaît que les fenêtres les plus courtes semblent donner les moins bons résultats de convergence. Pour la fenêtre de 5 jours, la fonction coût semble diverger, *i.e.* s'éloigner de cycle en cycle de la valeur théorique de convergence (indiquée par la ligne rouge en pointillés). Pour la fenêtre de 10 jours, ce phénomène est moins présent, mais au sein de chaque cycle, le palier de convergence n'est jamais atteint. Et la fonction coût présente d'importants sauts entre chaque cycle, comparables à ceux observés en tout début de minimisation au 1/4°. Il faut noter que la plus longue fenêtre d'assimilation présente la meilleure qualité de convergence au fil de la minimisation, sans toutefois converger totalement.

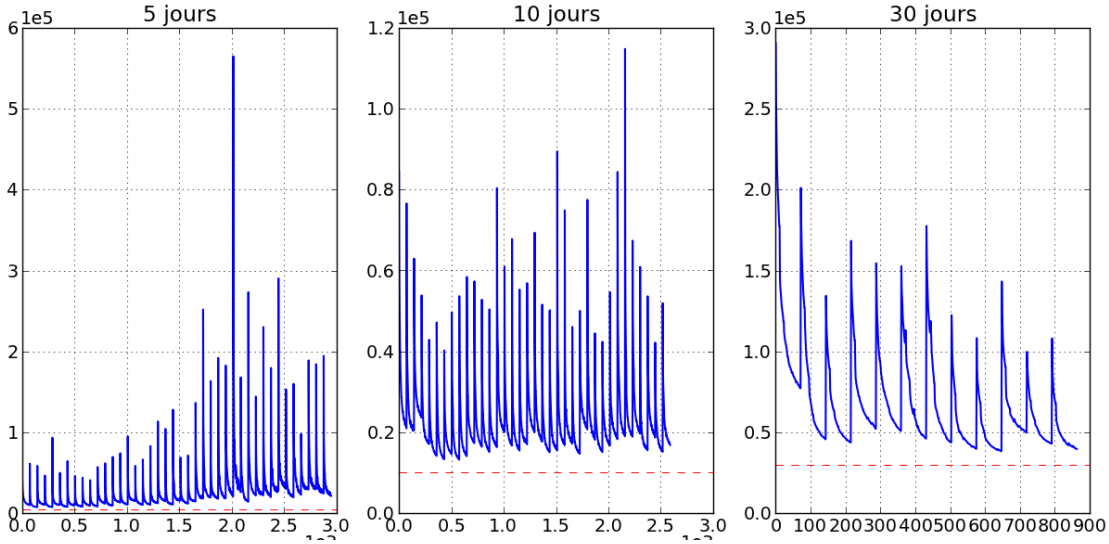

Figure 6.4 – *Évolution des fonctions coûts pour trois expériences d'un an d'analyse de 4DVAR incrémental avec SEABASS 1/12°, en fonction des itérations de minimisation, selon trois longueurs temporelles de fenêtres d'assimilation. De gauche à droite, 72 cycles de 5 jours, 36 cycles de 10 jours et 12 cycles de 30 jours. En pointillés rouges, p/2, où p est le nombre d'observations, valeur moyenne théorique de convergence de la minimisation.*

Ce comportement peut s'expliquer par différents facteurs. Tout d'abord, nous voyons qu'au sein des cycles, y compris pour les courtes fenêtres, l'algorithme ne converge pas (*i.e.* le gradient de la fonction est encore fort). Ceci se traduit généralement en pratique par l'obtention d'un incrément de forte amplitude. Appliqué au modèle, celui-ci va déstabiliser l'état initial d'ébauche. Ainsi, pour une analyse de 1 mois, avec de courtes fenêtres, le modèle va être déstabilisé 6 fois (fenêtres de 5 jours) ou 3 fois (fenêtres de 10 jours) contre 1 fois pour la fenêtre de 1 mois. Ceci impacte non seulement l'analyse, mais également le calcul des nouvelles fonctions coûts quadratiques entre chaque boucle externe.

Pour un cadre expérimental équivalent, nous voyons ici que l'augmentation du degré de non-linéarité du modèle a un impact fort sur la convergence de l'algorithme. Si celui-ci minimise bien la fonction coût au sein de chaque cycle, il est clair qu'il ne converge pas à la fois pour chaque cycle mais également au long de la période totale d'assimilation.

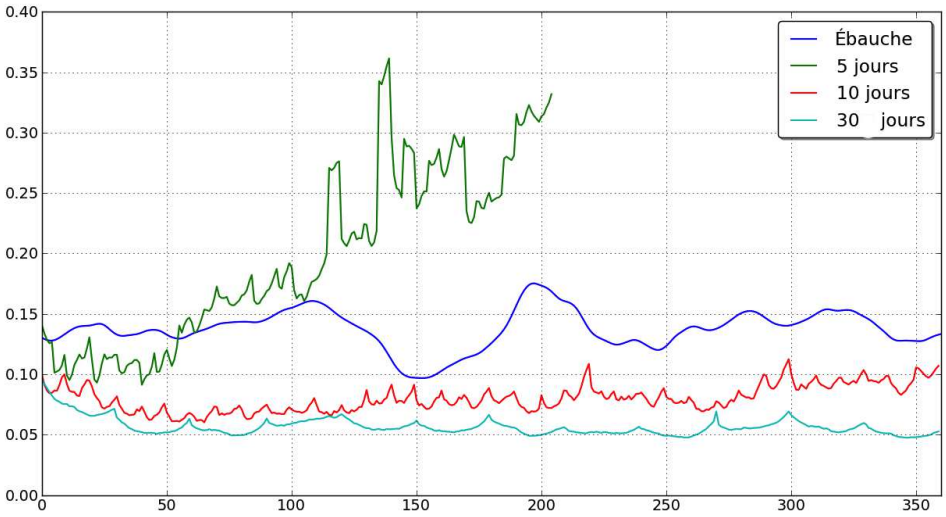

Figure 6.5 – *Évolution temporelle, en jours, de l'erreur moyenne quadratique d'ébauche (en bleu) et d'analyse, sur la SSH, pour différentes longueurs de fenêtres d'assimilation : 5 jours en vert, 10 jours en rouge et 30 jours en turquoise.*

Sur la Figure 6.5, cette analyse se confirme. L'application fréquente d'incréments non-équilibrés physiquement et de fortes amplitudes fait diverger la trajectoire analysée. Ainsi, plus la fenêtre s'étend, meilleure est la qualité de la trajectoire analysée, en terme de niveau moyen d'erreur et également de stabilité. Et malgré la faible convergence de la minimisation montrée précédemment, le niveau d'erreur d'analyse, pour la fenêtre d'un mois, est acceptable : 5 cm, en moyenne.

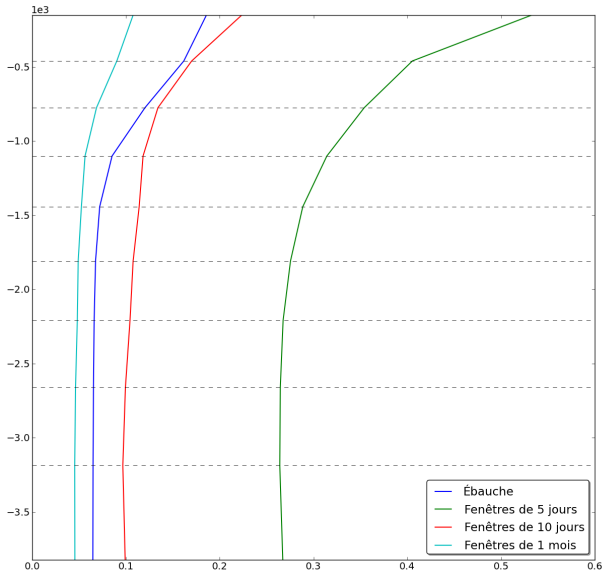

Figure 6.6 – *Moyenne spatiale et temporelle, pour chaque niveau vertical de l'erreur d'ébauche (en bleu) de la vitesse zonale, et de l'erreur d'analyse pour différentes longueurs de fenêtres d'assimilation : 5 jours en vert, 10 jours en rouge et 30 jours en turquoise.*

Sur la Figure 6.6, les deux expériences avec de courtes fenêtres montrent que l'erreur d'analyse est dégradée par rapport à l'erreur d'ébauche. Pour des fenêtres de 5 jours, nous avons pré-

cédemment remarqué que cette expérience faisait diverger la trajectoire analysée de l'état vrai. En revanche, pour la fenêtre de 10 jours, nous arrivions, sur la variable observée, à contrôler la trajectoire analysée. Ici, nous voyons, et notamment en profondeur où le phénomène est plus flagrant, que l'information disponible sur une fenêtre de 10 jours, tant en terme d'observations que d'évolution du modèle, n'est pas assez pertinente pour contrôler efficacement les variables non-observées, notamment sur la verticale.

6.1.3 Impact des paramètres de minimisation

Au vu des résultats ci-dessus, nous nous sommes posés la question de savoir si le nombre de boucles externes et d'itérations internes choisis étaient appropriés pour un modèle à plus haute résolution. En effet, ceux que nous avons prescrits au 1/12° sont équivalents à ceux du 1/4°. Pour espérer diminuer le gradient la fonction coût plus avant au sein de chaque cycle, nous avons donc augmenté le nombre de boucles externes et internes. Nous avons réalisé une expérience avec des fenêtres d'un mois, mais en considérant 20 boucles externes et 25 itérations internes. Nous n'avons pu réaliser que 4 cycles d'assimilation pour de tels paramètres, qui augmentent le coût de calcul, déjà élevé, d'un facteur 10.

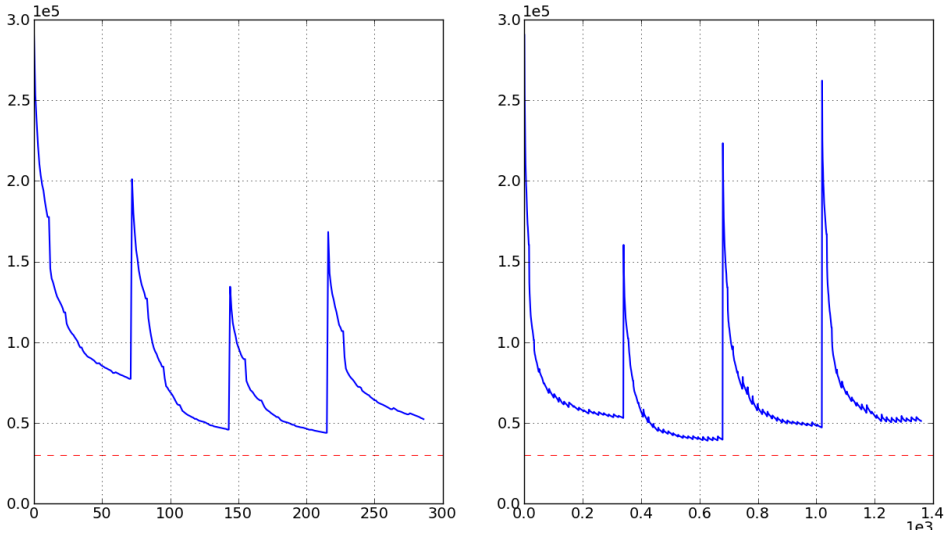

Figure 6.7 – *Évolution des fonctions coûts pour deux expériences enchaînant 4 cycles de 1 mois, avec 6 boucles externes et 10 itérations internes (à gauche) et 20 boucles externes pour 25 itérations internes (à droite). En abscisse sont figurées les itérations de minimisation.*

La Figure 6.7 montre deux choses :

- Au sein de chaque cycle, la fonction coût semble un peu plus diminuer, présentant un plateau de convergence de façon plus probante
- D'importants sauts intercycles sont présents, indiquant que l'incrément trouvé lors du cycle précédent est loin de l'incrément optimal du système. Ceci indique clairement que le bassin de convergence s'effectue autour d'un minimum local de la fonction coût.

Ainsi, l'effet de l'augmentation drastique du nombre de boucles externes et d'itérations internes n'est pas spectaculaire.

La Figure 6.8 montre qu'en réalité, en poussant la minimisation de la sorte, la trajectoire analysée n'est pas spécialement de meilleure qualité (courbe rouge par rapport à la courbe verte) :

si le niveau d'erreur semble à peu près équivalent, de nombreuses discontinuités dans la trajectoire d'analyse peuvent être observées. Ainsi, pour un coût de calcul jusqu'à dix fois supérieur, la solution obtenue est plutôt de moins bonne qualité.

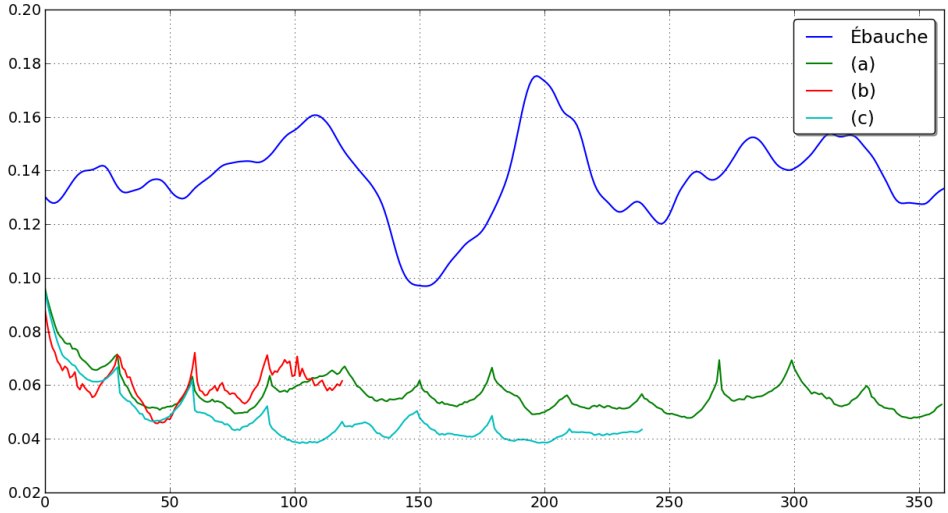

Figure 6.8 – *Évolution de l'erreur moyenne quadratique d'analyse de SSH (erreur d'ébauche en bleu), en fonction du temps (en jours) pour différentes expériences : (a) expérience standard en vert (observations simulées de type Jason-1, fenêtres de 1 mois, 6 boucles externes, 10 itérations internes), (b) expérience forçant la minimisation en rouge (observations simulées de type Jason-1, fenêtres de 1 mois, 20 boucles externes, 25 itérations internes), et (c), en turquoise expérience avec un réseau d'observation enrichi (observations simulées de type Jason-1 et SARAL/AltiKA)*

6.1.4 4DVAR vs. 3DFGAT

De la même manière qu'au chapitre précédent, nous nous sommes intéressés à comparer les performances du 3DFGAT et du 4DVAR pour SEABASS12. L'expérience en 3DFGAT considérée ici est en tout point semblable à celle du 4DVAR à deux différences près : les fenêtres d'assimilation durent 5 jours et non 1 mois et, vu que nous appliquons l'IAU, chaque cycle ne comporte qu'une seule boucle externe.

Sur la Figure 6.9, on voit que :

- Au 1/12°, l'erreur d'analyse issue du 3DFGAT est de l'ordre de l'erreur d'ébauche (qui est de l'ordre de 14 cm)
- Le 4DVAR améliore à nouveau nettement cette erreur d'analyse même si c'est de façon moins spectaculaire qu'au 1/4°.

Figure 6.9 – *Évolution temporelle (en jours) de la RMS d'analyse de SSH après 3DFGAT (en trait plein) et après 4DVAR (en pointillés), au 1/4° (en bleu) et au 1/12° (en rouge)*

Si l'on s'intéresse aux spectres de l'erreur d'analyse pour les deux algorithmes aux deux résolutions, nous voyons que le 4DVAR réduit plus fortement que le 3DFGAT l'énergie portée par l'erreur d'analyse à toutes les échelles, de façon tout de même plus efficace au 1/4° pour les moyennes et petites échelles.

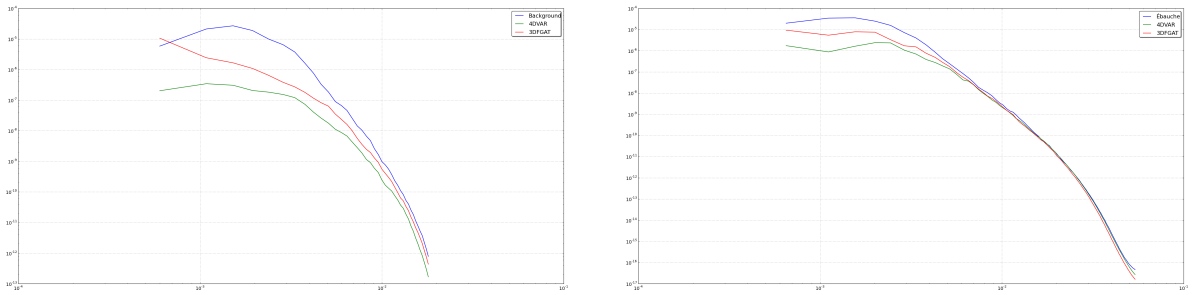

Figure 6.10 – *PSD de l'erreur de SSH d'ébauche (en bleu) et d'analyse en 3DFGAT (en rouge) et en 4DVAR (en vert) avec SEABASS4 (à gauche) et SEABASS12 (à droite)*

En résumé, de façon semblable au 1/4° quoique moins efficace, le 4DVAR appliqué à SEABASS12 permet de réduire significativement l'erreur d'analyse comparativement au 3DFGAT, et ce, à toutes les échelles.

6.1.5 Synthèse

Il est clair que l'augmentation de la résolution du modèle impacte fortement la qualité de convergence du 4DVAR incrémental. Les sauts de valeurs de la fonction coût à l'interface de chaque cycle, pour les expériences réalisées, indiquent que l'algorithme ne trouve jamais l'état (quasi-)optimal, ni ne s'en rapproche clairement, contrairement au 1/4°, où ces sauts diminuent drastiquement, de façon monotone, au cours de la période d'AD (pour les « bonnes » expériences).

Malgré cette faible convergence, l'erreur d'analyse de la SSH est d'un niveau moyen de 5 cm, pour des fenêtres d'AD de 1 mois, ce qui est un résultat tout à fait acceptable. Si l'on s'intéresse

aux vitesses, nous voyons que l'algorithme permet également de faire pénétrer l'information sur la verticale. Mis à part un problème sur la température dû aux réglages des écarts-type d'erreur d'ébauche, le 4DVAR incrémental permet donc de contrôler relativement bien la condition initiale pour un modèle dynamique océanique au 1/12°.

Cependant, nous voyons qu'en comparant au 1/4°, nous pouvons peut-être espérer de meilleures performances, en terme de qualité de convergence et conséquemment, de réduction des erreurs d'analyse. Pour ce faire, nous pouvons proposer plusieurs raisons méthodologiques permettant expliquer ce comportement, raisons liées aux non-linéarités présentes dans le modèle :

- L'hypothèse linéaire tangente est elle aussi impactée par l'augmentation des non-linéarités du modèles. La corrélation entre la perturbation propagée par le modèle linéaire tangent et celle propagée par le modèle directe diminue plus rapidement qu'avec le modèle au 1/4°. Ceci est illustré sur la Figure 4.19.
- La fonction coût, comme exposé dans la section 2.3.3, va présenter un plus grand nombre de minima locaux, à cause des non-linéarités. En partant d'une ébauche trop éloignée de l'état vrai, l'algorithme peut alors converger vers un minimum local et rester dans un bassin de convergence sous-optimal sans arriver à en sortir. Cependant, ce phénomène est difficile à diagnostiquer, sachant que nous n'avons pas accès à la forme analytique de la fonction coût.
- Le rapport entre la quantité d'information portée par les observations et le nombre de degrés de liberté effectif du modèle diminue quand la résolution augmente. Nous explorerons cet aspect du système d'AD dans la section suivante.

Le premier point étant inhérent à la méthode 4DVAR incrémental, il est vain d'essayer de le surmonter directement. En revanche les deux autres points sont intéressants à diagnostiquer dans notre cas, puisque les expérimentateurs peuvent avoir, dans une certaine mesure, la main sur ces aspects physiques et numériques du système d'AD, que sont l'ébauche et le réseau d'observation utilisé.

6.2 Améliorations des performances du 4DVAR pour la haute résolution

6.2.1 Diminuer la distance entre l'ébauche et l'état vrai

Dans toutes les expériences présentées précédemment, nous avons considéré une ébauche éloignée de l'état vrai par 3 mois de simulation. Au vu de la variabilité du modèle, les deux conditions initiales sont alors décorrélées. Si la fonction coût présente plusieurs bassins d'attraction vers des minima locaux, l'algorithme de minimisation va, dans ces conditions, plus probablement converger vers un de ceux-ci plutôt que vers l'état optimal. Pour vérifier ce point et juger de l'impact de la qualité de l'ébauche sur les performances du 4DVAR incrémental, nous avons réalisé deux expériences :

- Une première expérience où l'ébauche est identique à l'état vrai.
- Une deuxième expérience où l'ébauche est éloignée temporellement de 2 jours de l'état vrai. Ceci correspond à un écart RMS de 2 cm (Figure 6.11). Notons qu'il suffit d'une dizaine de jours pour que cet écart se stabilise à une dizaine de centimètres.

Dans ces expériences, les observations sont toujours bruitées, par un bruit blanc de 3 cm d'amplitude, suivant des traces altimétriques de type Jason-1.

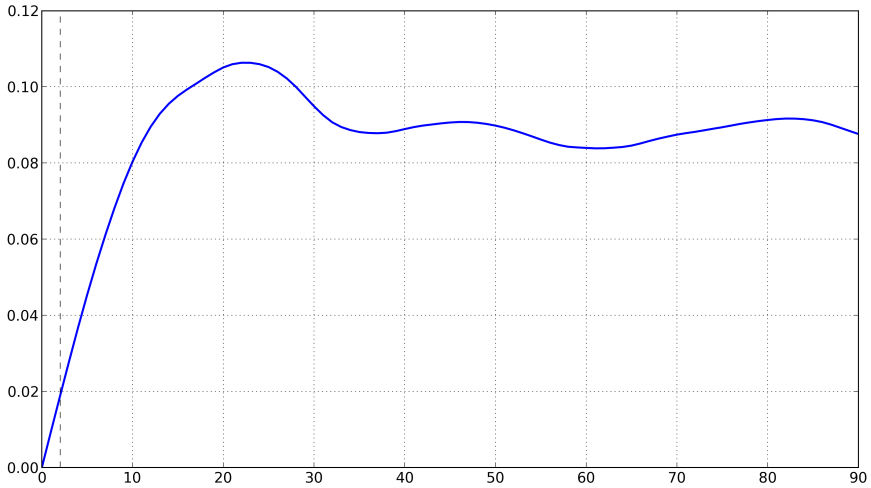

Figure 6.11 – *Écart RMS de SSH (en m) entre l'état initial vrai et les états simulés suivants à partir de cette condition initiale. Le trait vertical en pointillés indique l'écart au bout de 2 jours de simulation.*

Si l'on regarde l'évolution des fonctions coûts pour ces deux expériences comparativement à l'expérience de référence (ébauche prise 3 mois après l'état vrai) sur la Figure 6.12, nous pouvons noter que l'algorithme converge rapidement pour ces deux expériences pour les premiers cycles. En revanche, on note une dégradation de la qualité de convergence au fur et à mesure que l'on enchaîne les cycles d'assimilation, fonction coût qui rejoint peu à peu celle de l'expérience de référence. Ceci se fait logiquement plus rapidement pour lorsque l'ébauche est légèrement différente de l'état vrai au départ.

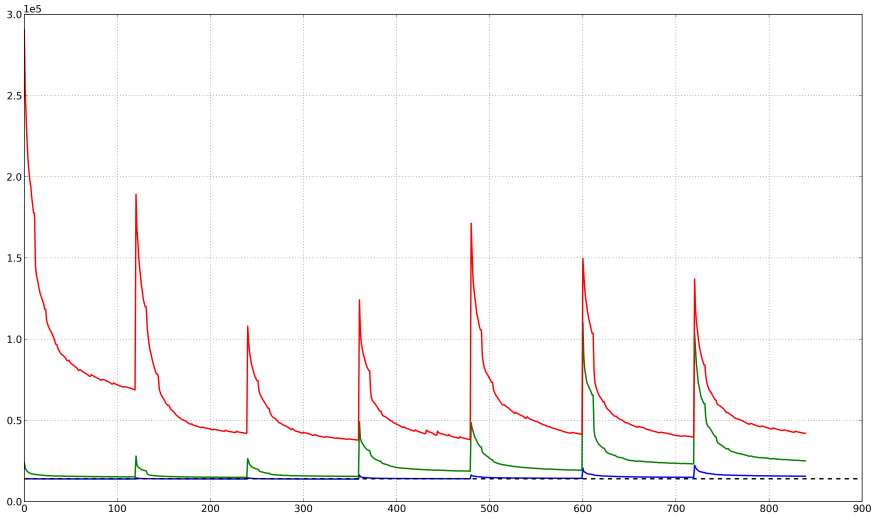

Figure 6.12 – *Évolution des fonctions coûts, en fonction des itérations de minimisation, à partir d'une ébauche identique à l'état vrai (en bleu), d'une ébauche prise 2 jours après l'état vrai (en vert) et d'une ébauche prise 3 mois après l'état vrai (en rouge). En pointillés noirs, la valeur moyenne théorique de convergence $p/2$.*

Si l'on s'intéresse à l'erreur d'analyse sur la dynamique (Figure 6.13), elle reste très faible lors des premiers cycles, pour les deux expériences, ce qui confirme la bonne convergence de

l'algorithme lors de ceux-ci. En revanche, pour les derniers cycles d'assimilation, cette erreur a significativement augmenté, restant tout de même en dessous de celle de l'expérience de référence. Les plus faibles erreurs sont obtenues, comme attendu, en partant d'une ébauche identique à l'état vrai.

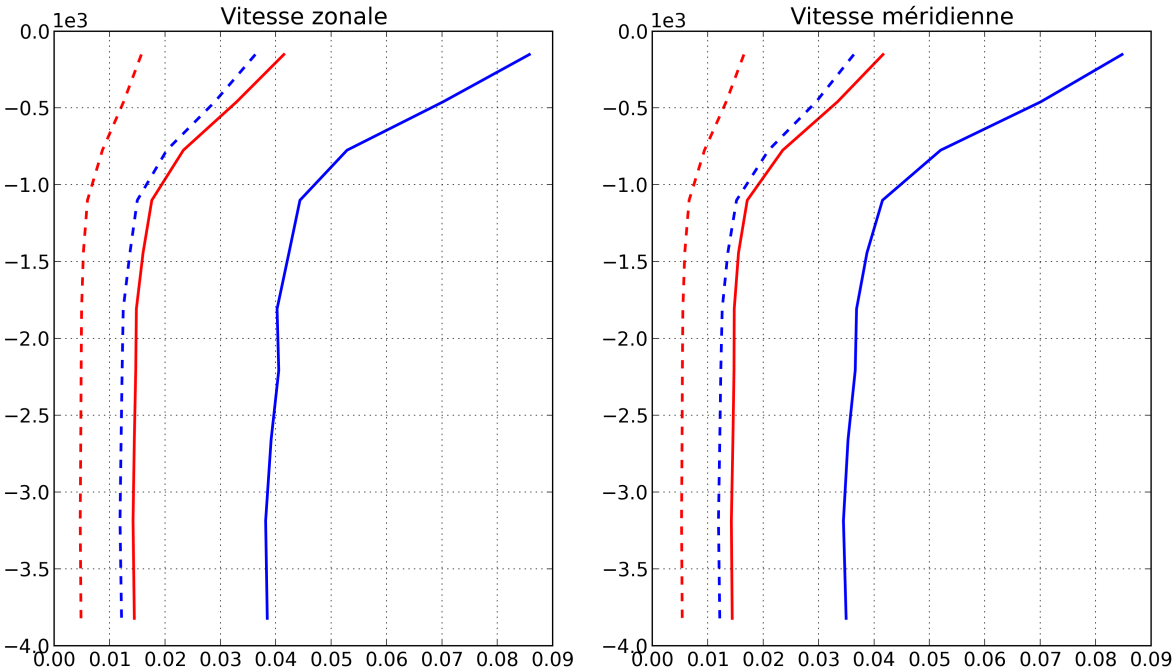

Figure 6.13 – *RMS d'analyse moyenne par niveaux verticaux de vitesse zonale (à gauche) et méridienne (à droite), pour une ébauche identique à l'état vrai (en rouge) et une ébauche prise 2 jours après l'état vrai (en bleu). En pointillés l'erreur réalisée au cours des trois premiers cycles d'assimilation, en trait plein lors des trois derniers cycles d'assimilation.*

En regardant l'énergie de l'erreur d'analyse de SSH contenue à différentes échelles, présentée par la Figure 6.14, ce phénomène se confirme également.

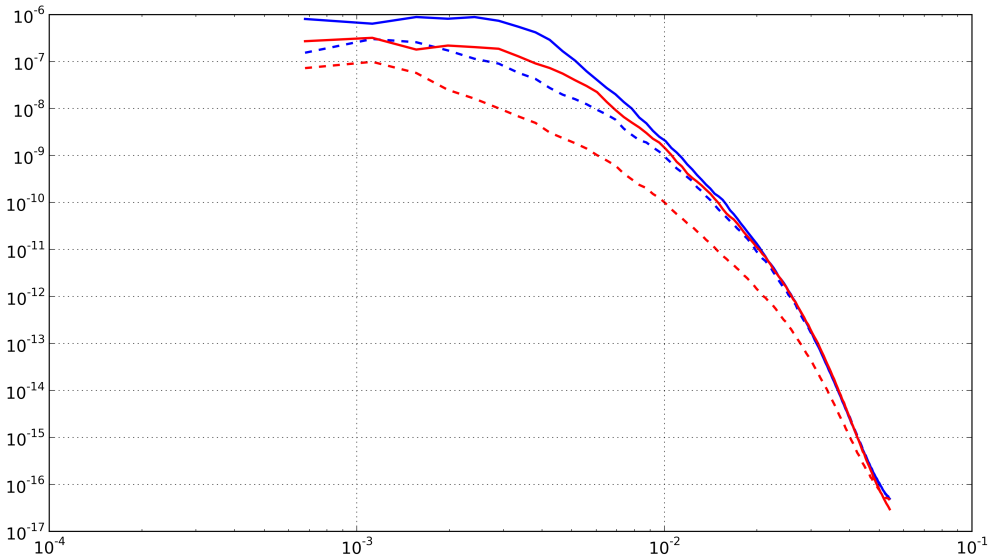

Figure 6.14 – *Densité spectrale de puissance de SSH des cartes d'erreurs d'analyse, pour une ébauche identique à l'état vrai (en rouge) et une ébauche prise 2 jours après l'état vrai (en bleu). En pointillés l'erreur réalisée au cours des trois premiers cycles d'assimilation, en trait plein lors des trois derniers cycles d'assimilation.*

Qualité de l'ébauche vs. performances du 4DVAR Ainsi, nous voyons qu'en partant d'une ébauche proche de l'état vrai, les problèmes de convergence constatés auparavant disparaissent, au cours des premiers cycles d'AD. On contrôle ainsi efficacement l'écoulement. L'enjeu est donc de rapprocher le plus rapidement possible l'ébauche de l'état vrai pour éviter que l'algorithme ne reste dans les bassins d'attractions de minima locaux éloignés de l'état optimal recherché.

Intérêt de l'évolutivité de B Cependant, au fur et à mesure des cycles d'AD, nous remarquons une dégradation de la qualité de convergence. Celle-ci provient de deux facteurs conjugués :

- La présence de l'erreur d'observation.
- Dû aux écarts-types des erreurs spécifiés dans notre fonction coût, le terme J_o est plus important de 3 ordres de grandeurs que J_b.

Ainsi, la faible confiance que nous accordons à l'ébauche va diriger l'algorithme a essayer de plutôt contraindre l'écoulement en fonction des observations et donc de leurs erreurs. De cycles en cycles, cette surreprésentation de l'erreur d'observation va mener l'algorithme à s'écarter de la trajectoire optimale, de façon plus ou moins rapide en fonction de la qualité initiale de l'ébauche. Ce phénomène est décrit mathématiquement dans Moodey et al. (2013).

Si l'on considère la situation inverse, *i.e.* en partant d'une ébauche de qualité médiocre, si nous convergeons à un moment vers l'état vrai, ce problème apparaîtra et le système d'AD divergera. Une solution est donc bien sûr d'adapter la confiance portée à l'ébauche au cours de la minimisation, comme cela est le cas par exemple dans toutes les méthodes de filtrage évolutif. En première instance, cela pourrait se faire simplement en ajoutant un poids devant le terme J_b de la fonction coût. En analysant la qualité de convergence au fur et à mesure des cycles, selon que l'on converge ou non, *via* l'amplitude de l'incrément courant par exemple, ce poids pourrait être alors augmenté ou diminué afin d'éviter la surreprésentation des erreurs d'observations. Toutefois, cette possibilité n'existe pas à l'heure actuelle dans notre système d'AD.

6.2.2 Accorder le réseau d'observation au système d'AD

Dans cette section, nous étudions l'impact de l'échantillonnage spatial et temporel du réseau d'observation altimétriques sur les performances du système variationnel d'AD. Notre objectif est de regarder comment l'échantillonnage spatial et temporel du réseau d'observation altimétrique impacte les performances du 4DVAR incrémental, et dans quelle mesure nous pouvons améliorer ces dernières à l'aide d'un réseau approprié aux caractéristiques du modèle.

Les expériences ont consisté à considérer différents réseaux d'observations simulés selon les trajectoires des altimètres SARAL/AltiKA et Jason-1, présentant des échantillonnages spatio-temporels différents, illustrés sur les Figures 6.15 et 6.16.

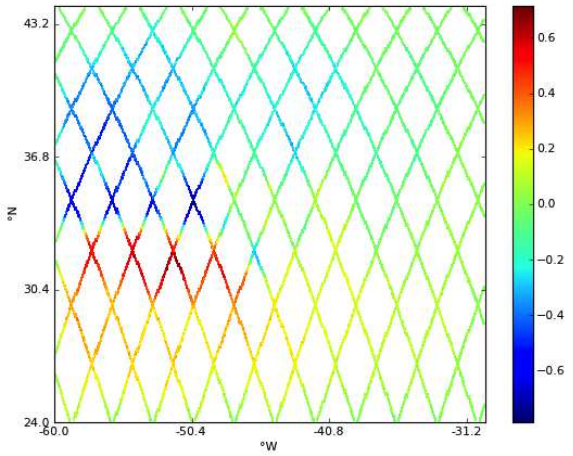

Figure 6.15 – *Exemple de traces altimétriques suivant la trajectoire de Jason-1 après 10 jours*

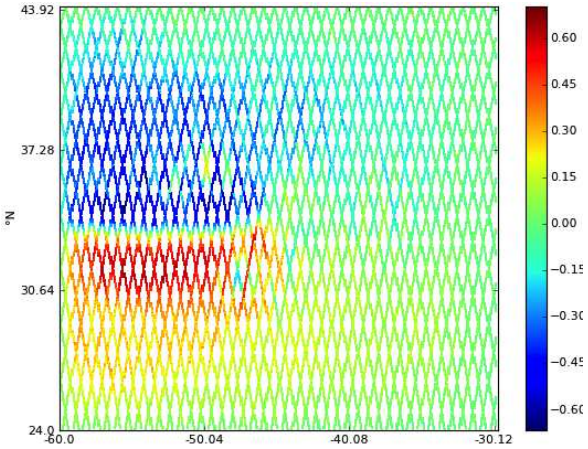

Figure 6.16 – *Exemple de traces altimétriques suivant la trajectoire de SARAL/AltiKA après 30 jours*

Nous allons détailler dans cette section les résultats obtenus pour trois expériences jumelles qui diffèrent sur les observations retenues pour chacune :

- **Expérience J1** : Un jeu d'observations issu de traces simulées selon la trajectoire de Jason-1 (répétivité de 10 jours, inter-traces à l'équateur de 350 km)
- **Expérience SA** : Un jeu d'observations issu de traces simulées selon la trajectoire de SARAL/AltiKA (répétivité de 35 jours, inter-traces à l'équateur de 70 km)
- **Expérience JS**, combinant les 2 jeux d'observations.

Sur une même fenêtre temporelle, les traces que nous simulons selon les trajectoires de Jason-1 et SARAL/ALtiKA fournissent environ le même nombre de mesures. Le but de cette comparaison est d'identifier l'impact de l'échantillonnage sur notre système d'AD variationnel pour SEABASS au 1/4° puis 1/12°.

Échelles spatiales représentées dans les observations vs. modèle

Convergence au 1/12° Nous nous intéressons en premier lieu à la convergence de l'algorithme pour les trois expériences, illustrée par la Figure 6.17. L'ébauche est située 3 mois après l'état vrai. L'expérience Jason-1+SARAL/AltiKA converge mieux que l'expérience J1. Ici, nous voyons que l'expérience SA converge également mieux que l'expérience J1. Ainsi, pour un même nombre moyen d'observation, une répartition spatio-temporelle différente mène à différents résultats de convergence.

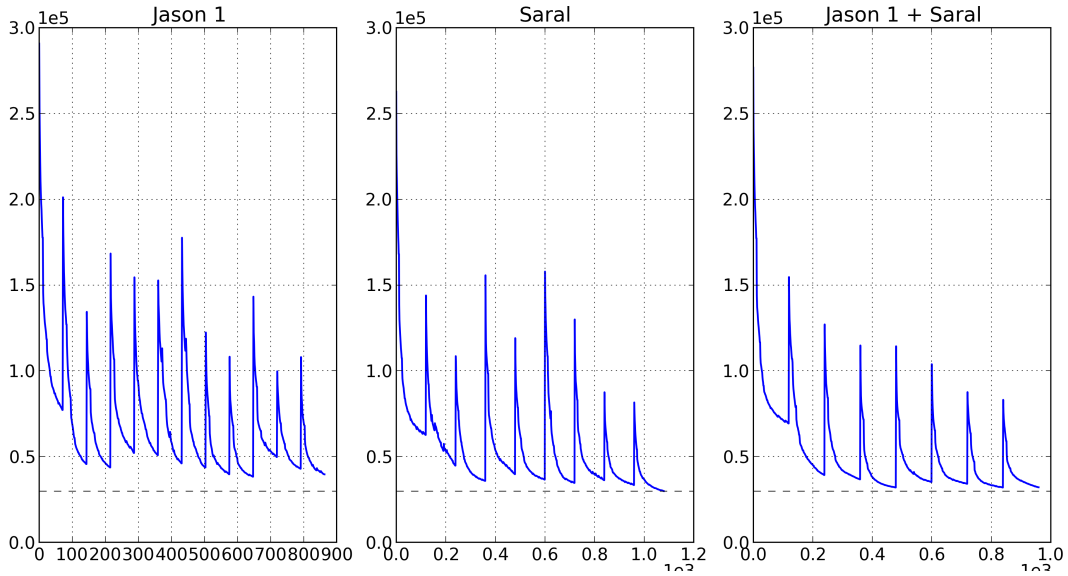

Figure 6.17 – *Évolution des fonctions coûts pour les trois expériences $\mathcal{J}1$ (à gauche), SA (au centre) et $\mathcal{J}S$ (à droite), en fonction des itérations de minimisation. En pointillés noirs, le niveau moyen théorique de convergence de la minimisation.*

Erreur d'analyse La Figure 6.18 montre l'évolution de l'erreur d'analyse pour les trois expériences sur la variable prognostique observée, SSH. Là encore, un échantillonnage de type SARAL/AltiKA permet de diminuer le niveau moyen d'erreur sur les 120 derniers jours de l'analyse (5.8 cm pour J1, 0.51 pour SA et 0.45 pour JS).

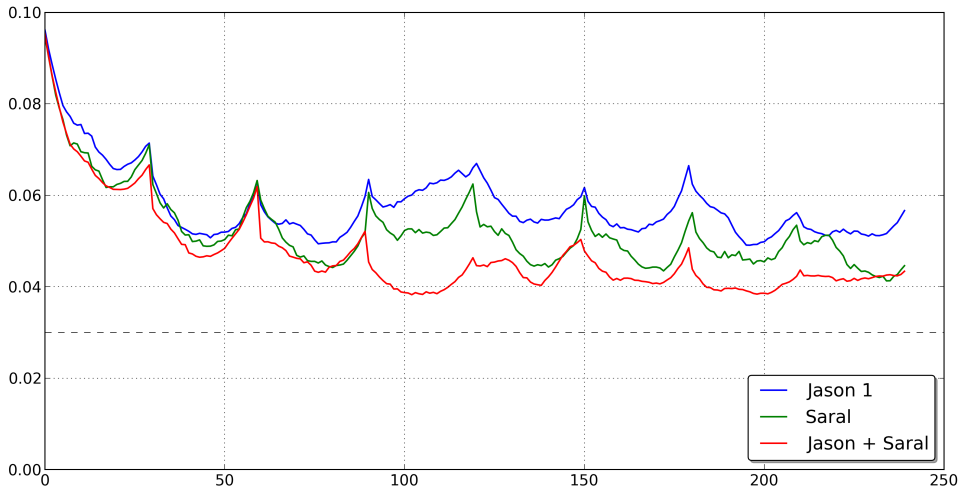

Figure 6.18 – *Évolution temporelle (en jours) de l'erreur moyenne quadratique d'analyse de SSH pour les trois expériences $\mathcal{J}1$ (bleu), SA (en vert) et $\mathcal{J}S$ (en rouge)*

En revanche, sur une variable non-observée, comme la vitesse zonale, montrée sur la Figure 6.19, l'erreur d'analyse est du même ordre que l'on utilise Jason-1 ou SARAL/AltiKA.

Il est cependant clair sur ces deux figures que les deux satellites combinés offrent une meilleure réduction de l'erreur d'analyse, sur la variable observée et celles non-observées, en surface comme en profondeur.

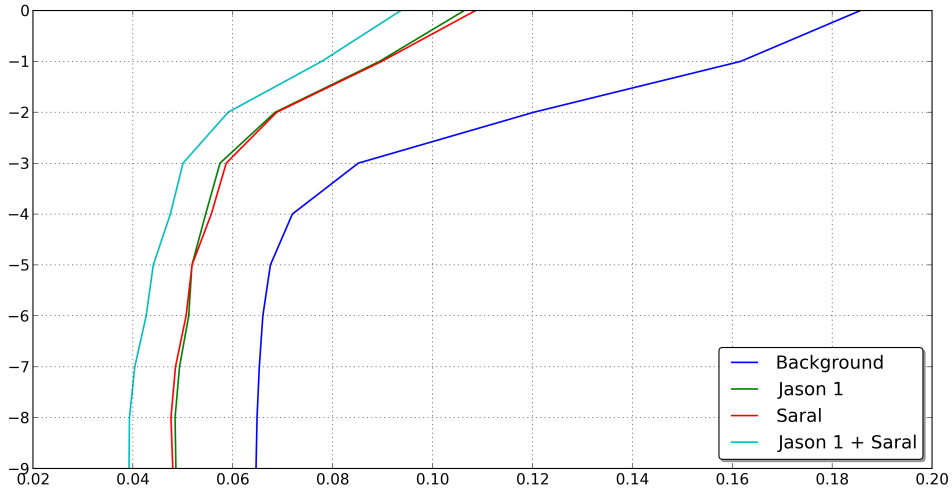

Figure 6.19 – *Idem Figure 6.18 mais pour la vitesse zonale (en m.s^{-1}), sur chaque couche, moyennée sur les 120 derniers jours de la trajectoire analysée*

Si l'on s'intéresse à l'énergie contenue dans les structures d'erreurs à différentes échelles, présentée dans la Figure 6.20, le scénario JS permet de la réduire à toutes les échelles. En revanche, le scénario SARAL/AltiKA seul réduit mieux l'erreur aux échelles inférieures à 300 km par rapport à Jason-1 seul.

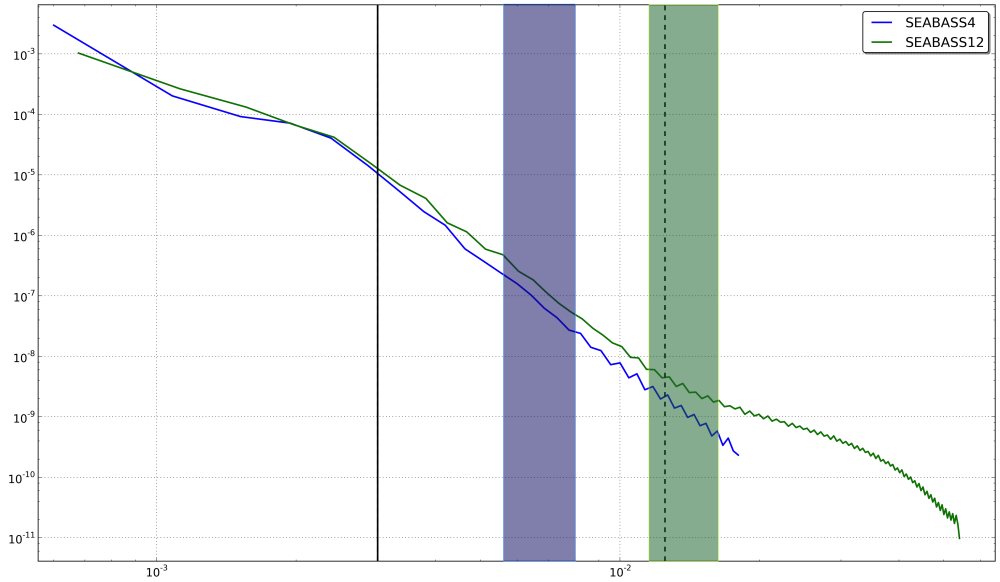

Figure 6.20 – *PSD des cartes d'erreurs d'analyse de SSH pour les différents scénarios altimétriques (SARAL en bleu, Jason-1 en vert et les deux combinés en rouge). Au centre, un zoom sur les plus grandes échelles du spectre (de 3000 km à 100 km). Le trait vertical en pointillés correspond à des échelles de 300 km.*

Interprétations D'après Tai (2004), pour une grille de traces altimétriques données, en la supposant complète (*i.e.* la fenêtre d'assimilation plus longue que la période de répétitivité du satellite) et régulière (*i.e.* sur une zone moyennement étendue), la plus petite échelle captable par les observations est celle de l'intertrace, peu importe l'échantillonnage temporel pour SEABASS12.

Ainsi, le flux d'information porté par les observations va être en deça de celui représentable par le modèle. On peut en effet estimer la résolution effective du modèle (voir Skamarock (2004), *i.e.* les échelles correctement résolues étant donné les choix numériques à environ 80 km, alors que l'intertrace de Jason-1 est d'approximativement 300 km et celle de SARAL/AltiKA de 80 km (*cf.* Figure 6.21). Ainsi, capter les échelles de l'ordre de 100 km, représentées dans le modèle, va réellement permettre de contraindre la dynamique à ces échelles lors du processus d'assimilation. Ici, au 1/12°, un réseau de type Jason-1 seul est en quelque sorte sous-dimensionné pour contraindre convenablement la trajectoire analysée. Un réseau plus dense en espace, tel que celui de SARAL/AltiKA, est alors plus indiqué. L'information apportée conjointement par les deux réseaux améliore encore plus les performances du système d'AD.

Dans le cas des deux altimètres combinés, les échelles spatiales contenues dans les observations seront encore inférieures à celles de SARAL/AltiKA seul, même si cette fois-ci l'indicateur n'est plus constant sur l'ensemble du domaine. L'information ainsi apportée permet donc au système d'AD de contrôler plus efficacement l'écoulement analysé.

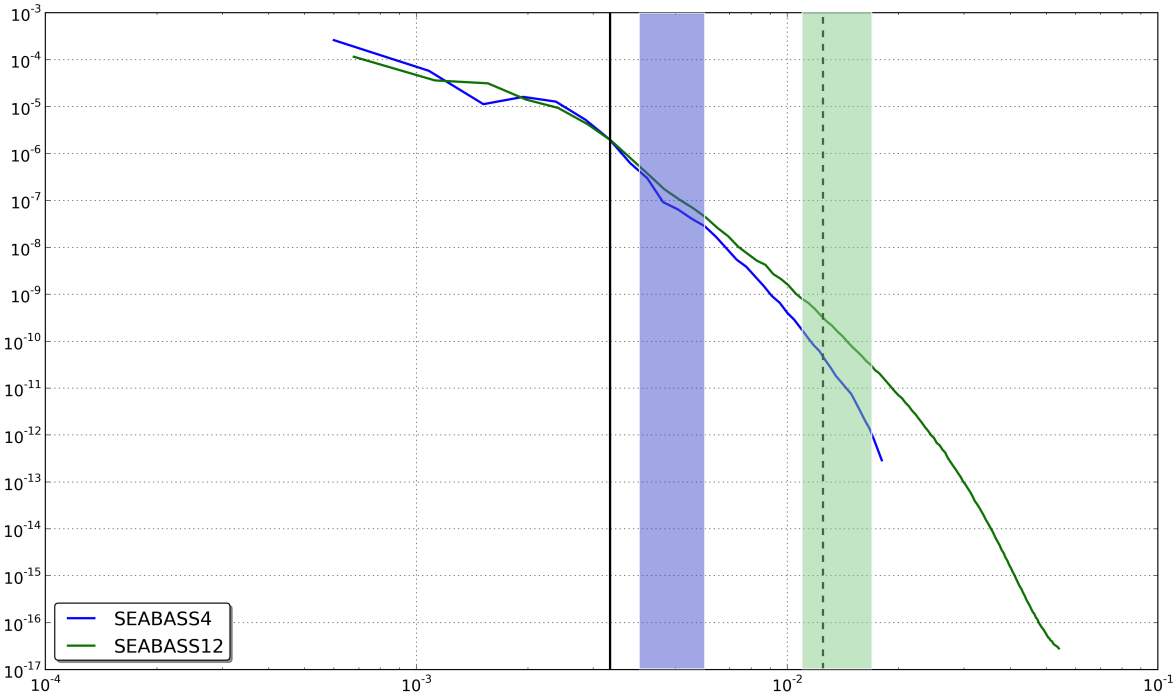

Figure 6.21 – *PSD de SSH pour SEABASS au 1/4° (en bleu) et au 1/12° (en vert). Le trait vertical en pointillés correspond à la borne inférieure des échelles captées SARAL/AltiKA et en trait plein celles par Jason-1. La zone bleutée correspond à l'ordre de grandeur de la résolution spatiale effective attendue du modèle au 1/4° et la zone verte celle au 1/12°.*

Un corollaire de l'interpétation que nous venons d'exposer serait que, pour un modèle au 1/4°, l'information supplémentaire apportée par SARAL/AltiKA par rapport à celle apportée par Jason-1 n'est pas réellement utilisable par le modèle, qui ne résout plus ces échelles supplémentaires (*cf.* Figure 6.21). Afin de tester cette hypothèse, nous allons donc regarder le comportement du système d'AD pour ces différents scénarios au 1/4°.

Dégradation du conditionnement de la fonction coût

Erreur d'analyse au 1/4° L'évolution de l'erreur d'analyse de SSH, montré par les Figures 6.22 et 6.23 nous informe un peu plus sur la qualité du contrôle de la trajectoire analysé rencontré au cours des différentes expériences. Pour les fenêtres de 15 jours, le scénario SARAL/AltiKA est clairement moins bon, présentant, en régime permanent d'assimilation, un niveau moyen d'erreur 3 fois supérieur aux autres expériences. Pour les fenêtres de 1 et 2 mois, les trois scénarios présentent des niveaux d'erreurs d'analyse équivalents.

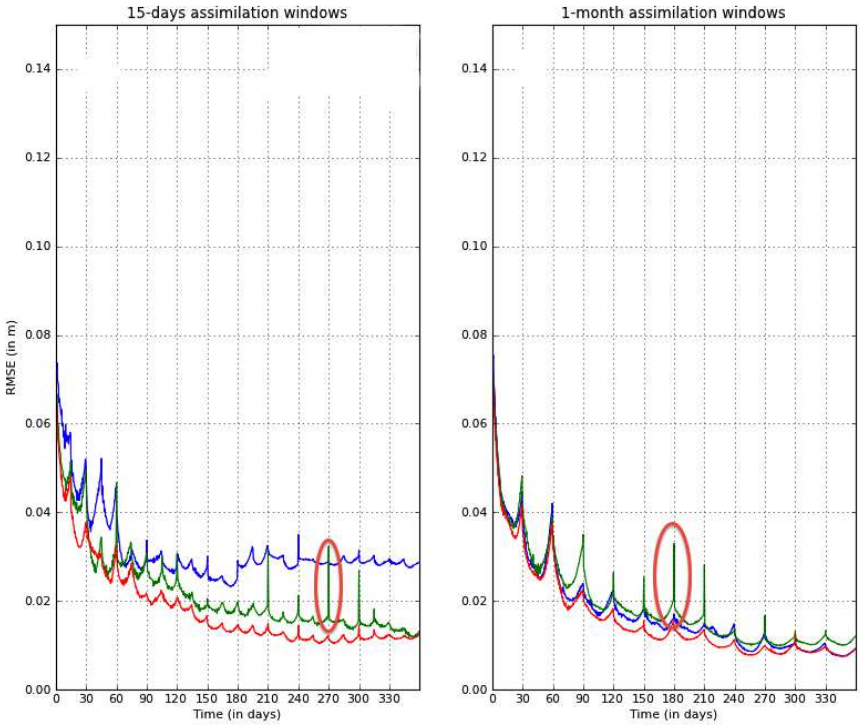

Figure 6.22 – *Évolution de la RMSE d'analyse de SSH, pour différents scénarios altimétriques (SARAL/AltiKA en bleu, Jason-1 en vert, les deux combinés en rouge) pour différentes longueurs de fenêtres d'AD : 15 jours à gauche 15 jours et 1 mois (à droite).*

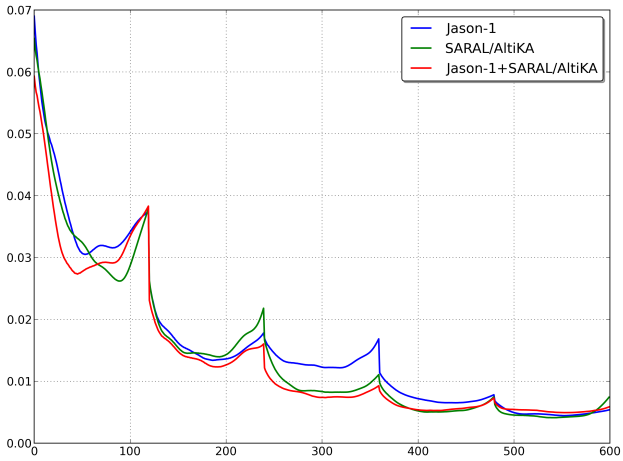

Figure 6.23 – *Idem Figure 6.22 pour une fenêtre d'AD de 2 mois*

Sur la Figure 6.24, des structures d'erreurs légèrement plus persistantes et étendues spatialement sont remarquables sur le scénario Jason-1 seulement. Ces structures d'erreurs ont une taille comparable à celle de l'intertrace de Jason-1.

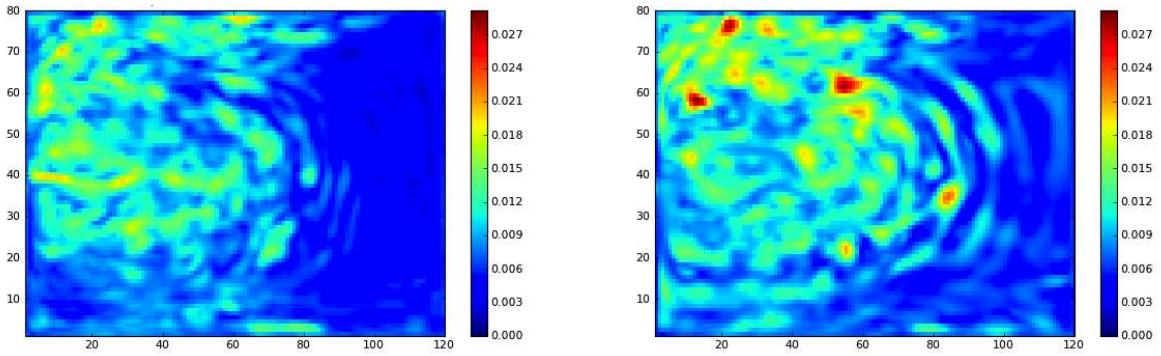

Figure 6.24 – *Champs moyens de RMS d'erreur d'analyse de SSH, pour les 6 derniers mois d'assimilation pour le scénario SARAL/AltiKA à gauche et Jason-1 à droite (fenêtre d'assimilation de 1 mois)*

Pour la fenêtre de 15 jours, l'échantillonnage de type SARAL/AltiKA ne couvre pas l'ensemble de notre domaine, ce qui explique la plus forte erreur d'analyse pour cette expérience.

Il vient qu'au 1/4°, pour des fenêtres de moins d'un mois, le type de réseau, pour peu qu'il soit suffisamment régulier en espace, n'influe pas ou peu sur la convergence et la qualité de l'erreur d'analyse. S'il présente de grosses zones spatiales non-couvertes, telles que les traces de SARAL/AltiKA pour une fenêtre de 15 jours, le 4DVAR incrémental ne permettra pas de contrôler efficacement l'erreur d'analyse.

Pour les fenêtres plus longues, les deux réseaux donnent également une erreur d'analyse comparable. En effet, Jason-1 permet déjà de représenter presque toutes les échelles spatiales réoslues par le modèle. *A contrario*, les observations de SARAL/AltiKA peuvent introduire une erreur de représentativité [1]. Cette fois-ci, logiquement, la combinaison des deux altimètres n'apportent plus d'informations utiles au système d'AD pour contraindre l'écoulement. En revanche, il faut garder en tête que lorsqu'il y a trop d'observations, le conditionnement de la fonction coût peut être dégradée, comme montré dans la section 2.3.2 et expliqué dans Andersson et al. (2000). Cependant, dans les expériences présentées ici, ceci ne se traduit pas sur les niveaux d'erreurs d'analyse.

Convergence au 1/4° L'évolution de la fonction coût, présentée sur la Figure 6.25 révèle deux informations :

- Pour les plus courtes fenêtres (celles de 15 jours et d'un mois), quelle que soit la couverture spatiale du réseau d'observation, la qualité de convergence est très similaire quelque soit le scénario, même si la courbe rouge donne légèrement une meilleure convergence
- Pour la plus longue fenêtre, celle où le régime transitoire d'assimilation est plus long, une couverture spatiale de type SARAL/AltiKA permet clairement une meilleure convergence,

1. Information contenue dans les observations que le modèle n'est pas capable de représenté (*i.e.* phénomènes non-modélisés, processus sous-maille, aliasing)

se traduisant par des paliers de convergence intracycles plus marqués, et une diminution monotone des sauts de fonctions coûts inter-cycles.

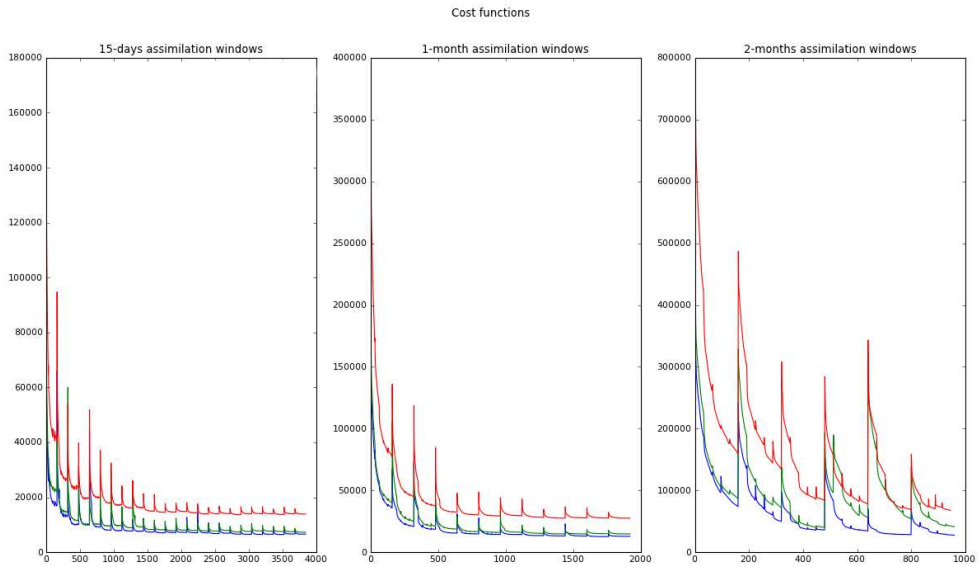

Figure 6.25 – *Évolution des fonctions coûts pour les trois expériences $\mathcal{J}1$ (en vert), SA (en bleu) et $\mathcal{J}S$ (en rouge), en fonction des itérations de minimisation, pour différentes longueurs de fenêtres d'assimilation (de gauche à droite, 15 jours, 1 mois et 2 mois). Les fonctions coûts sont mises à l'échelle pour pouvoir être comparées.*

Synthèse

L'échantillonnage spatial des observations à considérer, pour une résolution et une fenêtre d'AD données, doit être capable de capter toutes les échelles représentables par le modèle, en évitant d'en surreprésenter. Ajouter plus d'information peut revenir à surcontraindre le système d'AD et à dégrader le conditionnement de la fonction coût, rendant la minimisation plus ardue. Ceci introduit également une erreur de représentativité qui n'est pas prise en compte dans notre système d'AD.

Un travail sur l'adéquation entre le réseau d'observation et le modèle permet donc au 4DVAR incrémental de gagner en performance lors du processus d'assimilation indépendamment des caractéristiques de l'ébauche.

6.2.3 Considérations méthodologiques : l'assimilation progressive ou quasi-statique

Si nous avons vu quelles sont les pistes d'amélioration de l'ébauche et du réseau d'observations pour augmenter les performances du 4DVAR incrémental lorsque l'on augmente la résolution du modèle, restent à considérer les pistes d'amélioration méthodologiques en touchant au coeur de l'algorithme de minimisation. Nous allons éprouver une stratégie auparavant testée avec succès par Luong et al. (1998), Pires et al. (1996) ou encore Titaud et al. (2010), nommée *assimilation progressive* ou *assimilation quasi-statique*.

Principe de la méthode

L'assimilation progressive consiste à démarrer un cycle de minimisation sur une fenêtre temporelle courte que l'on étend progressivement jusqu'à la longueur de fenêtre d'AD désirée T. Si on décrit étape par étape cet algorithme, en s'aidant de la Figure 6.26, ceci donne :

1. Commencer par assimiler les données sur un court intervalle temporel $[0, T_1]$
2. Utiliser le vecteur initial identifié à l'étape précédente comme nouveau *first guess* pour assimiler des données sur un intervalle temporel plus grand $[0, T_2]$
3. On reboucle à l'étape 2 en utilisant des intervalles de plus en plus grands jusqu'à couvrir la totalité de la fenêtre d'assimilation désirée.

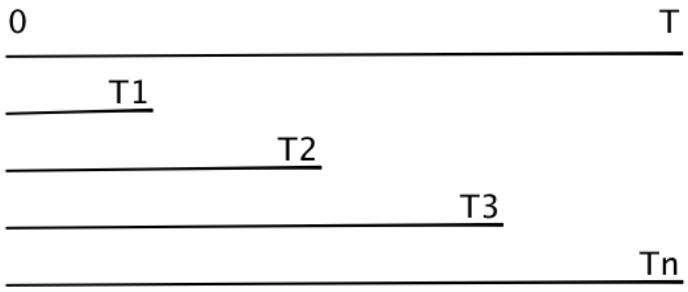

Figure 6.26 – *Représentation schématique de l'augmentation de la longueur temporelle de la fenêtre d'AD au cours d'un cycle d'assimilation progressive.*

L'idée principale sous-jacente de l'assimilation progressive est la suivante : sur de courtes fenêtres temporelles l'approximation de la fonction coût non-linéaire par la fonction coût quadratique est meilleure. Ainsi, le risque de se trouver dans un bassin d'attraction d'un minimum local, et donnant un état sous-optimal, est réduit.

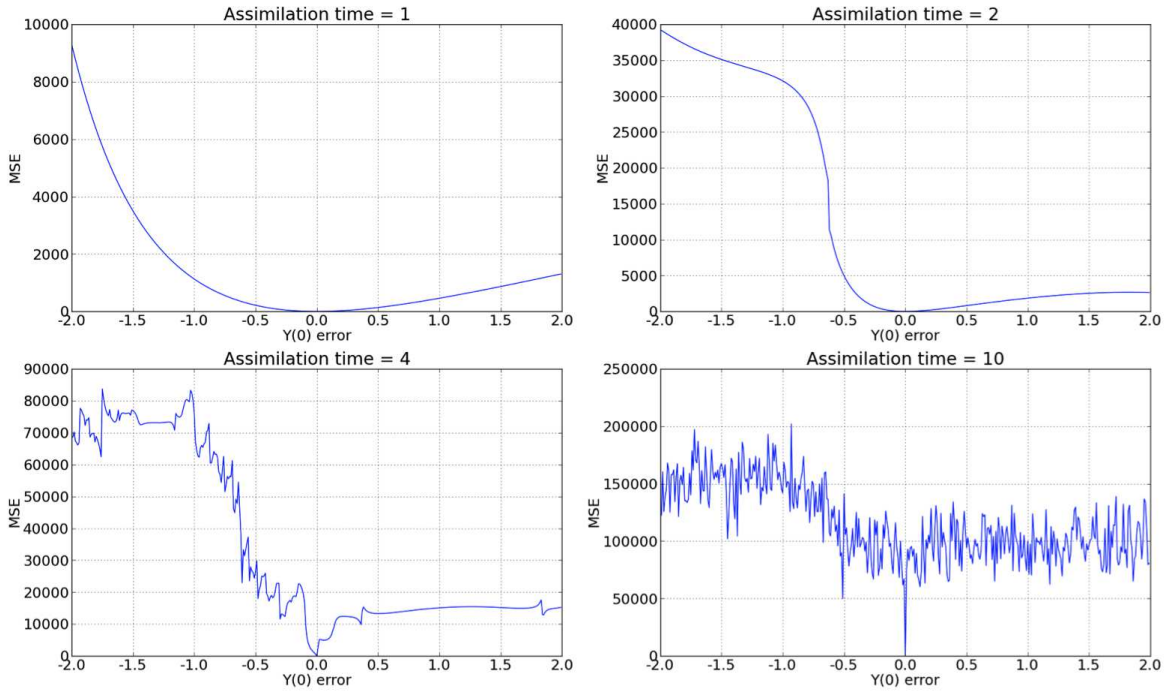

Figure 6.27 – *Exemples de fonctions coût, basées sur le modèle de Lorenz à 3 variables, pour différentes longueurs de fenêtres temporelles. Plus la fenêtre d'agrandit, plus le nombre de minima locaux croît, rendant la minimisation de plus en plus ardue.*

Si l'on reprend l'exemple de la fonction coût basée sur le modèle de Lorenz à 3 variables (*cf.* chapitre 2), illustré par la Figure 6.27, pour la plus courte fenêtre, la fonction coût non-linéaire est convexe et présente un seul minimum global. Faire une première minimisation sur cette fenêtre nous permet donc d'espérer se rapprocher de l'état optimal recherché.

Cependant, en reprenant cet exemple, si l'on ne converge pas vers l'état optimal, nous voyons que pour les très longues fenêtres d'AD, rien ne garantit que l'algorithme ne converge pas vers un minimum local. L'assimilation progressive est une stratégie empirique et aucun résultat théorique ne garantit la convergence lors de son application.

Ensuite, en pratique, contrairement à l'exemple précédent, aucun argument ne garantit que la solution du problème d'optimisation (*i.e.* l'état optimal recherché) soit identique pour différentes longueurs de fenêtres. De plus, comme nous n'avons pas accès aux formes des différentes fonctions coûts dans notre application, le choix de la taille des fenêtres d'AD successives est choisi empiriquement.

Malgré tout, le succès de l'application de cette méthode dans les études citées au-desus, dont celle de Luong et al. (1998) dans un cadre similaire (4DVAR incrémental appliqué à un bassin océanique turbulent idéalisé, basé toutefois sur les équations de la théorie quasi-géostrophique) nous pousse à l'appliquer dans notre contexte. Ceci nous permettra de voir si cette stratégie améliore les performances de notre système en terme de contrôle de l'écoulement analysé.

Assimilation progressive au 1/12°

Nous avons effectué une expérience, enchaînant 4 cycles d'assimilation progressive sur le modèle au 1/12°. Chacun de ces cycles comporte :

- Une première fenêtre d'assimilation de 15 jours, avec 6 boucles externes et 10 itérations de minimisation pour chaque
- Puis une fenêtre d'assimilation de 1 mois, dont l'ébauche est l'état initial analysé précédemment (mêmes paramètres de minimisation)

Dans ce cas, l'assimilation progressive coûte 1.5 temps CPU de plus qu'une assimilation séquentielle classique enchaînant 4 fenêtres de 1 mois chacune, à paramètres de minimisation identiques. Nous comparons dans la suite l'expérience de 4DVAR classique (ou purement séquentielle) et l'expérience d'assimilation progressive décrite ci-dessus.

Sur la Figure 6.28, l'assimilation progressive améliore effectivement l'erreur d'analyse, dès le premier cycle d'assimilation. Le niveau d'erreur moyen lors du dernier est de 4 cm en assimilation progressive contre 6 cm pour le 4DVAR classique.

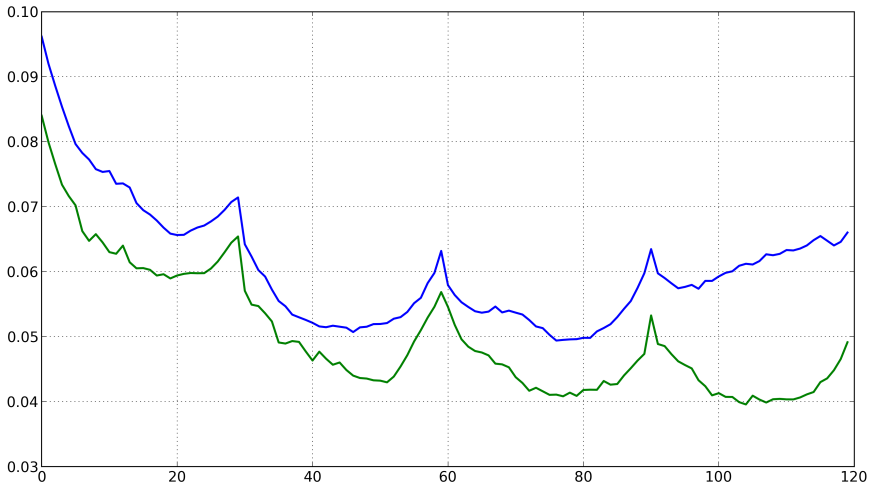

Figure 6.28 – *Évolution temporelle (en jours) de la RMS d'analyse moyenne de SSH pour l'expérience de 4DVAR classique (en bleu) et l'expérience d'assimilation progressive (en vert)*

Si l'on s'intéresse à l'énergie contenue dans l'erreur d'analyse aux différents échelles spatiales, avec la Figure 6.29, celle-ci est déjà moindre pour l'assimilation progressive dès le premier cycle. Les échelles spatiales mieux contrôlées sont comprises entre 200 et 500 km. Au 4ème cycle, toutes les échelles sont mieux résolues, avec une nette amélioration sur les plus grandes échelles (1000 km).

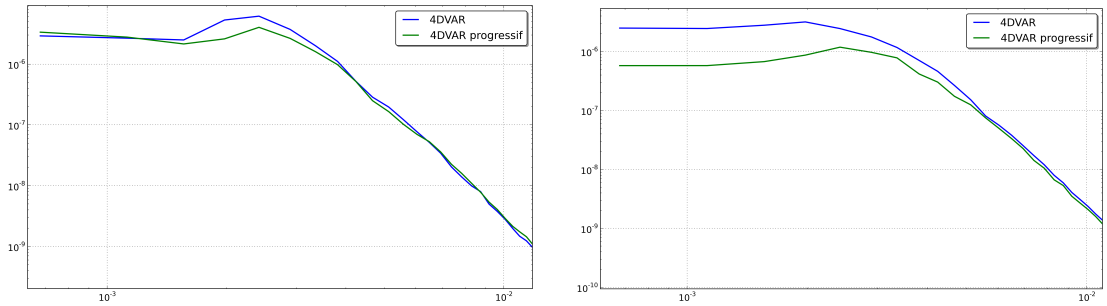

Figure 6.29 – *PSD de l'erreur d'analyse de SSH, pour le 4DVAR classique (en bleu) et le 4DVAR progressif (en vert), au cours du premier cycle d'assimilation (à gauche) et du 4ème cycle (à droite)*

Ainsi, sur la variable observée, l'assimilation progressive permet d'améliorer la réduction de l'erreur d'analyse. En revanche, sur la Figure 6.30, nous constatons une dégradation de l'erreur d'analyse sur les vitesses en profondeur par rapport au 4DVAR classique. À ce stade de notre travail, nous ne pouvons expliquer ce phénomène.

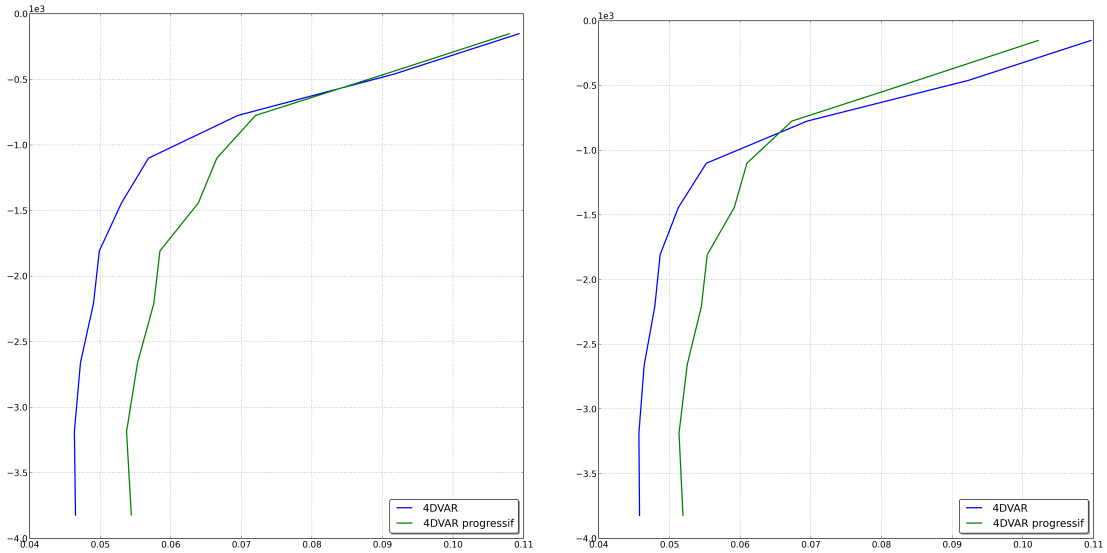

Figure 6.30 – *Profil d'erreur d'analyse de vitesse (zonale à gauche, méridienne à droite) moyennée temporellement au cours du 4ème cycle d'assimilation, pour le 4DVAR classique (en bleu) et le 4DVAR progressif (en vert)*

Assimilation progressive au 1/4°

Pour compléter les tests précédents, nous allons maintenant voir si de telles améliorations peuvent être constatées pour SEABASS au 1/4°, où le système d'AD réduit déjà très correctement l'erreur d'analyse.

Nous avons donc réalisé une assimilaiton progressive au 1/4° enchaînant successivement des fenêtres de 15 jours, 1 mois, 2 mois, 3 mois et 4 mois.

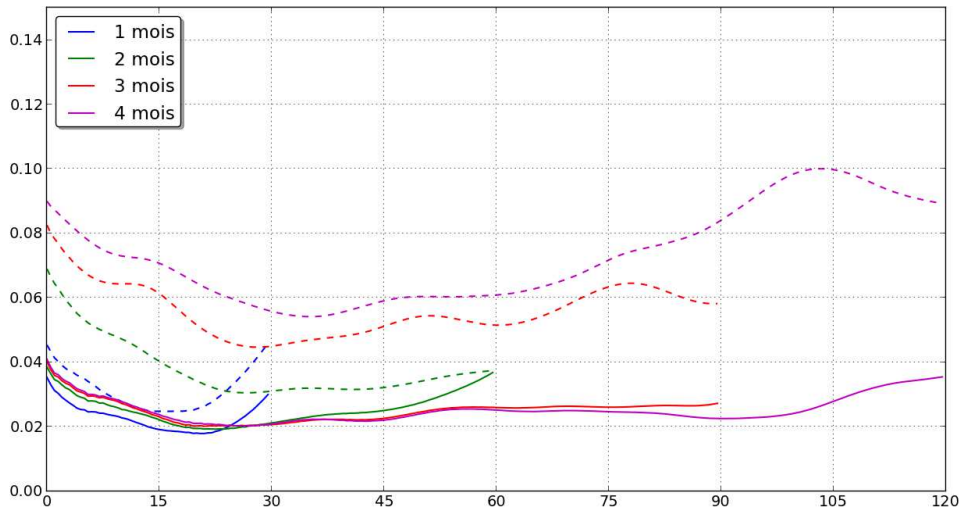

Figure 6.31 – *Évolution temporelle (en jours) de la RMS d'analyse moyenne de SSH pour différentes fenêtres d'assimilation de type 4DVAR classique (en pointillés) et leurs équivalents au cours de la phase d'assimilation progressive (évolution des fenêtres de 15 jours, 1 mois, 2 mois, 3 mois, 4 mois)*

Sur la Figure 6.31, l'erreur d'analyse sur la SSH est également bien réduite par l'assimilation progressive, pour différentes longueurs de fenêtres d'AD. Elle permet également d'étendre la fenêtre d'assimilation à 4 mois, en conservant une erreur d'analyse sur la SSH acceptable.

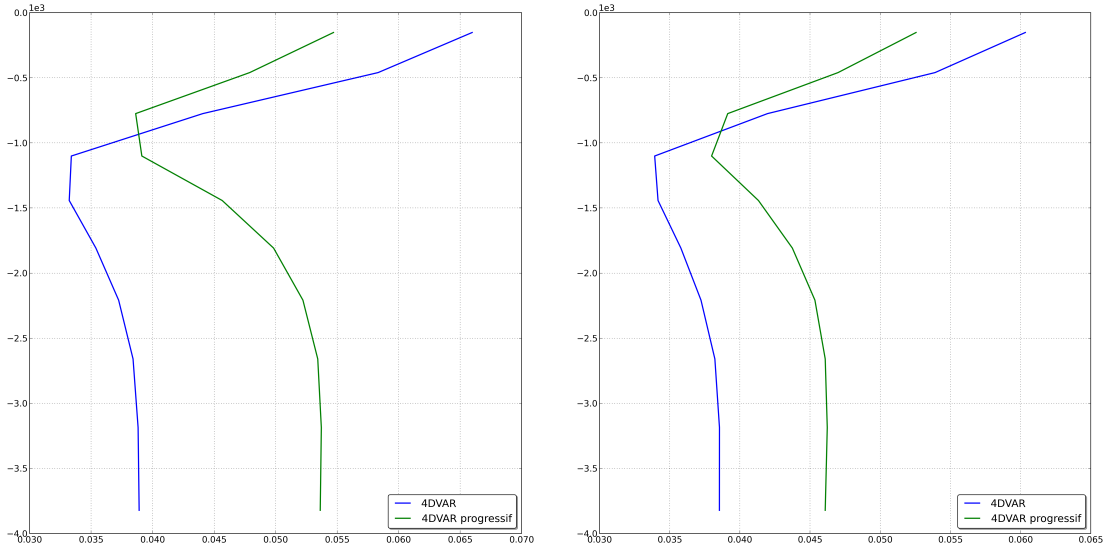

Figure 6.32 – *Profil d'erreur d'analyse de vitesse (zonale à gauche, méridienne à droite) moyennée temporellement au cours du 1er cycle d'assimilation de 2 mois, pour le 4DVAR classique (en bleu) et le 4DVAR progressif (en vert)*

En revanche, on constate sur la Figure 6.32 que le niveau d'erreur d'analyse sur les vitesses en profondeur est dégradé par rapport au 4DVAR classique, de la même manière qu'au 1/12°.

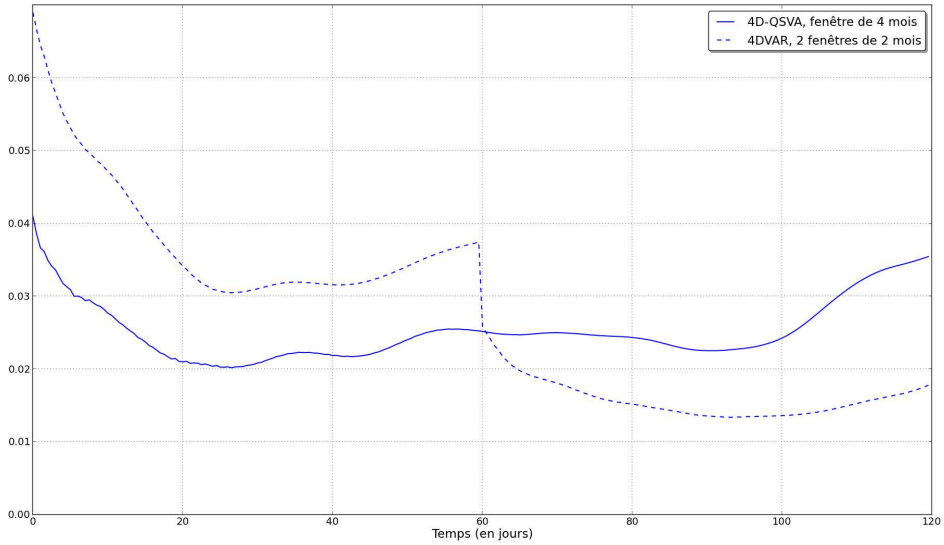

Figure 6.33 – *Évolution temporelle (en jours) de la RMS d'analyse moyenne de SSH pour 2 fenêtres de 2 mois (en pointillés) et d'une fenêtre de 4 mois après assimilation progressive (en trait plein)*

La Figure 6.33 nous permet de noter qu'il n'est pas forcément avantageux de préférer un fenêtre de 4 mois issue de l'assimilation progressive plutôt qu'enchaîner deux cycles de 2 mois en assimilation classiques. En effet, nous voyons que le niveau d'erreur d'analyse de SSH est inférieur dès le début du second cycle dans le deuxième cas.

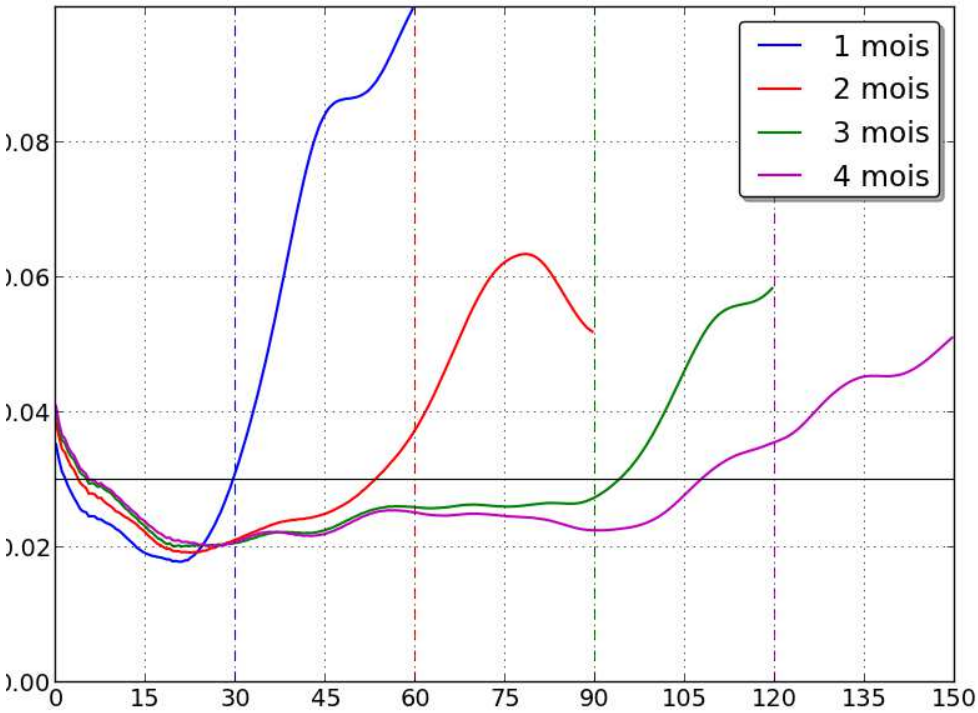

Figure 6.34 – *Évolution temporelle (en jours) de la RMS d'analyse et de prévision moyenne de SSH les différentes fenêtres durant l'assimilation progressive. À gauche des traits verticaux en pointillés correspond la phase d'analyse de la fenêtre correspondante (même couleur), à droite la phase de prévision.*

Enfin, si l'on s'intéresse à l'erreur de prévision sur la SSH, avec la Figure 6.34, au cours du processus d'assimilation progressive, celle-ci reste plus basse sur une plus longue période temporelle à mesure que l'on augmente la taille de la fenêtre d'AD.

Synthèse

Si le système d'AD, avec le 4DVAR classique, enchaînant les cycles de façon séquentielle, donne déjà de bonnes performances et converge convenablement, comme au 1/4° ici, l'assimilation progressive n'est pas forcément indiquée. Elle ne diminuera que peu, ou pas, le niveau d'erreur d'analyse sur plusieurs cycles d'assimilation pour un coût numérique significativement supérieur.

En revanche, au 1/12°, où le système d'AD présente des difficultés à contrôler l'écoulement, l'assimilation progressive se révèle efficace pour améliorer significativement l'erreur d'analyse, sur la variable observée, dès le premier cycle d'assimilation.

Cependant, nous avons remarqué que la stratégie d'AD progressive, quelle que soit la résolution, dégradait la qualité de l'erreur d'analyse sur les variables non-observées en profondeur. Nous n'avons pas d'interprétations encore à ce sujet. Un travail plus avancé est nécessaire pour esquisser les causes de ce phénomène. De la même manière, explorer les différentes stratégies sur la taille des intervalles successifs lors du cycle d'AD progressif peut être une piste à explorer pour optimiser l'application de cette méthode.

6.3 Conclusion

Au cours de ce chapitre, il apparaît que l'augmentation de la résolution spatiale du modèle commence à poser des difficultés à l'application du 4DVAR incrémental, même si l'erreur d'analyse est encore acceptable et le contrôle de la trajectoire effectif.

Au 1/12°, la convergence de l'algorithme de minimisation est plus lente, comparativement au 1/4°. Les fenêtres courtes (5 et 10 jours) ne permettent pas de contrôler effectivement la trajectoire analysée. Une fenêtre temporelle d'AD d'un mois produit encore les meilleures performances, donnant un niveau d'erreur d'analyse moyen de 5 cm sur la SSH. Augmenter drastiquement le nombre de boucles externes et d'itérations internes n'améliore pas sensiblement ces résultats, au prix d'un coût de calcul 10 fois supérieur dans notre cas. Toutefois, il est important de noter que, de la même manière qu'au 1/4°, le 4DVAR est encore beaucoup plus performant que le 3DFGAT qui ne permet pas de réduire convenablement l'erreur d'ébauche.

Nous avons également vu qu'en considérant une ébauche plus proche de l'état vrai, le système d'AD ne présente pas de régime transitoire d'assimilation. En revanche, au fur et à mesure des cycles, la convergence se dégrade, dû au fait que $\mathbf{B}$ n'évolue pas dans notre système.

L'importance de l'échantillonnage des observations pour améliorer les performances a été mise en évidence. Il apparaît que les échelles spatiales représentées dans le modèle d'une part et dans les observations d'autre part doivent être en adéquation pour optimiser les performances de notre système d'AD variationnelle. En revanche, il faut garder en tête qu'une trop grande densité d'observations peut dégrader le conditionnement de la fonction coût et rendre la minimisation plus lente .

Enfin, méthodologiquement, l'assimilation progressive permet d'améliorer la qualité de l'analyse, sur la variable observée, au 1/12°. Si au 1/4°, le 4DVAR incrémental est suffisamment efficace pour ne pas recourir à la régularisation de la fonction coût proposée par l'assimilation progressive, au 1/12°, en revanche, cette dernière permet d'améliorer plus rapidement la qualité de l'ébauche et, ce faisant, accélérer la convergence du 4DVAR. Il reste tout de même à explorer la dégradation de l'erreur d'analyse sur les variables non-observées en profondeur lors de l'application de cette stratégie.

Ainsi, améliorer les performances du 4DVAR incrémental dans ce contexte peut se faire sur deux axes :

- La mise en adéquation du réseau d'observations par rapport aux échelles représentées dans le modèle
- L'amélioration de la qualité de l'ébauche, par une stratégie d'assimilation progressive par exemple

D'autres pistes méthodologiques, telles que l'amélioration des préconditionneurs (voir Tshimanga et al. (2008)) ou les méthodes multi-échelles (voir Neveu (2011)) pourraient également permettre d'améliorer les performances du 4DVAR en contexte fortement non-linéaire.

Chapitre 7

Vers la très haute résolution : exploration des difficultés de l'assimilation de données

Les systèmes de prévisions océaniques et de réanalyse du futur opéreront à très haute résolution. Ces systèmes traiteront alors des échelles et des niveaux de turbulence totalement inaccessibles à l'heure actuelle pour des raisons de coût numérique très élevé. Ce chapitre se place dans cette optique et vise à discuter quelques-unes des problématiques liées à ces futurs systèmes. Pour ce faire, nous avons mis en place la configuration SEABASS précédemment décrite au 1/100° (soit $\Delta x \approx 1$ km).

Nous présenterons dans un premier temps cette configuration, ses caractéristiques numériques et physiques, comparativement aux autres résolutions horizontales précédemment étudiées. À partir de ces caractéristiques, nous extrapolerons une partie des résultats obtenus aux chapitres précédents sur les performances du système d'AD variationnelle dans ce contexte.

Nous aborderons également les pistes méthodologiques et expérimentales pour prolonger le travail décrit dans ce manuscrit pour un contexte fortement turbulent.

7.1 Considérations générales sur la très haute résolution

Comme nous l'avons évoqué à plusieurs reprises auparavant, la tendance de l'océanographie physique est de représenter de plus en plus les dynamiques méso- et subméso-échelles océaniques, élargissant ainsi le spectre des échelles spatiales et temporelles simulées pour comprendre et représenter le plus précisément possible la circulation océanique générale. Plusieurs travaux (voir notamment Capet et al. (2008) ou Klein et al. (2008), par exemple) indiquent l'importance d'une telle évolution. Une étude physique comparative, beaucoup plus exhaustive que celle proposée ici, d'une configuration de bassin océanique similaire, avec NEMO à différentes résolutions horizontales (1°, 1/9°, 1/27° et 1/54°) a été réalisée par Lévy et al. (2010), présentant bien l'influence de la dynamique méso- et subméso-échelle sur la circulation générale. Quelques éléments d'intérêt de la représentation de la subméso-échelle peuvent être notés :

- La représentation des échelles spatiales entre 100 km et 1 km permettent d'affiner les prévisions à ces mêmes échelles.
- La dynamique méso- et subméso-échelle intervient de manière importante dans la cir-

culation océanique générale et représente un élément clé dans le système de transferts énergétiques de l'écoulement océanique
- La bio-géochimie océanique, domaine d'intérêt croissant dans la communauté océanique, est concernée en premier lieu par ces dynamiques. Les processus biogéochimiques sont fortement impactés par les circulations à très petites échelles.

Cette évolution a des impacts directs sur la conception des réseaux d'observations et des modèles numériques de circulation océanique. Pour les premiers, un exemple flagrant est le programme d'altimétrie satellitaire SWOT : avec la prise de mesure novatrice à large fauchée, la couverture spatiale doit permettre d'observer ces échelles. Ceci implique un flux de données toujours plus important en terme de densité spatiale des observations.

En ce qui concerne la modélisation, la représentation des dynamiques subméso-échelles implique l'augmentation de la résolution spatiale, et conséquemment temporelle (pour des raisons de stabilité des schémas numériques), des modèles. En augmentant ainsi la résolution, le niveau de turbulence effectivement modélisé s'accroît. En d'autres termes, l'aspect chaotique des modèles augmente, les rendant ainsi plus sensibles aux perturbations de leurs paramètres d'entrée tels que la condition initiale, les erreurs numériques (*e.g.* erreurs d'arrondis), le choix des schémas d'advection, les conditions limites (*e.g.* forçages, bathymétrie). Le coût de calcul augmente également drastiquement, tant en terme de stockage, en mémoire vive et morte, que de temps de CPU utilisé.

Comme nous l'avons déjà évoqué auparavant, ces directions prises par l'océanographie physique apportent de sérieuses contraintes aux méthodes d'AD :

1. De manière générale, la densité croissante des réseaux d'observations rend de plus en plus caduque l'hypothèse aujourd'hui généralement faite de la décorrélation spatiale et temporelle des erreurs d'observation.
2. La chaoticité grandissante des modèles rend inévitable la prise en compte systématique de l'erreur modèle, ce dernier pouvant présenter des trajectoires significativement différentes pour de très petites variations de ses paramètres d'entrée.
3. Enfin, même si la résolution spatiale et temporelle des réseaux d'observation d'une part et des modèles d'autre part croissent, elles ne sont pas forcément en phase. Ainsi, les observations peuvent contenir des phénomènes que le modèle n'est pas en mesure de représenter, c'est ce que que l'on appelle l'erreur de représentativité (*cf.* chapitre 6).

Pour les méthodes variationnelles d'AD plus spécifiquement, la très forte non-linéarité du modèle a deux conséquences importantes, déjà mises en valeur lors du passage du 1/4° au 1/12°, présenté au chapitre 6, mais qui risquent d'être encore plus contraignantes à très haute résolution :
- La fonction coût non-linéaire présentera de très nombreux minima locaux, vers lesquels l'algorithme de minimisation peut converger.
- L'hypothèse linéaire tangente, dans le cadre du 4DVAR incrémental, ne sera valide que sur de très courtes fenêtres d'intégrations du modèle.

7.2 Caractéristiques physiques et numériques de SEABASS au 1/100°

7.2.1 Description succinte de la configuration

La configuration SEABASS au 1/100° reprend les caractéristiques des configurations précédemment étudiées : même domaine géographique, même discrétisation verticale, même forçage par un vent sinusoïdal en espace et stationnaire en temps, et salinité constante.

En revanche, la grille discrétisée comporte 2001 points sur la latitude et 3001 points sur la longitude. Les paramètres de diffusion horizontale (pour la température) et de viscosité sont baissés à $-1.28.10^8$ m^4.s^{-1}. Et pour des raisons de stabilité des schémas numériques, le pas de temps est réduit à 36 s.

7.2.2 Quelques considérations de coût numérique

Pour des résolutions de cet ordre de grandeur, il faut souligner le coût numérique comparativement aux résolutions plus basses. Nous avons réalisé une phase de spinup de 28 ans afin de faire ensuite une simulation de référence sur une période de 30 jours (soit 72000 pas de temps). Ces calculs ont été effectués sur la machine *froggy* du mésocentre de calcul HPC grenoblois CIMENT [1]. Cette machine est un supercalculateur Bull DLC, composée de nœuds de calculs disposant de 24 Gb de mémoire RAM. Chaque nœud est muni de 2 processeurs Intel SandyBridge, chacun composé de 8 coeurs de fréquence d'horloge de 2.6 GHz. Le tout est relié par un réseau de type InfiniBand non-bloquant. Notre simulation s'est effectuée sur 1024 coeurs. Dans ces conditions, son coût numérique est le suivant :

- Le temps CPU total est de 410h par année de simulation (il faudrait 410 h de calcul pour un seul CPU, pour simuler un an d'écoulement)
- La mémoire vive utilisée est de 1Gb par noeud
- Le coût de stockage d'un champ complet du modèle est de 1 Gb.

Extrapolation à une expérience d'assimilation En extrapolant ces caractéristiques à un cycle d'assimilation de 4DVAR incrémental, pour une fenêtre temporelle d'un mois, comportant 10 boucles externes et 10 itérations internes pour chacune (expérience équivalente à un cycle au 1/12°), en devant stocker la trajectoire du modèle non-linéaire à chaque boucle externe, le coût numérique de cette seule et petite expérience est alors à peu près le suivant :

- Le temps CPU total est composé, pour une boucle externe, de l'intégration du modèle non-linéaire, puis des intégrations des modèles linéaire tangent et adjoint (approximativement 4x plus chère que celle du modèle non-linéaire) pour chaque itération interne. Au total, pour chaque boucle externe, le temps CPU est de 4100 h/an environ. Pour le cycle entier, le temps CPU total est alors de 41000 h/an.
- La mémoire vive utilisée au sein d'une itération interne oscille entre 3 et 4 Gb par noeud.
- En supposant stocker la trajectoire du modèle non-linéaire tous les jours, il faudrait constamment gérer 30Gb en mémoire morte au minimum.

Autrement dit, le coût de calcul, en valeur absolue commence à devenir prohibitif, surtout lorsque les besoins d'une étude comme celle-ci comprennent de nombreuses expériences, elles-mêmes

1. http ://ciment.ujf-grenoble.fr/

composées de nombreux cycles d'assimilation. Il faut aussi tenir compte du post-traitement des sorties du système d'assimilation, le calcul des erreurs commençant à requérir également un certain niveau de parallèlisme. Et ces calculs considèrent que nous disposons de 1024 processeurs (ou plus) en permanence. À titre de comparaison, les expériences menées au 1/4° ont été réalisées sur 16 processeurs et au 1/12° sur 64 processeurs au maximum ! Dans ces conditions, une étude exhaustive du système d'AD pour une telle résolution demande de très importants efforts en ressources de calcul.

Extension à une configuration réaliste et globale Le propre de notre configuration est qu'elle présente un domaine géographique de type régional et idéalisé. Pour passer à une configuration globale et réaliste, cela implique un surcoût numérique colossal. En se basant sur les caractéristiques de la configuration ORCA025 (grille globale au 1/4° comportant 46 niveaux), une idée approximative du coût numérique d'une simulation globale au 1/100° pourrait être la suivante, sur 1024 processeurs, pour 1 mois de simulation, en supposant une scalabilité [2] optimale :

- Temps CPU total de 3 125 000 heures par année de simulation
- Mémoire vive totale utilisée de 60Gb
- Stockage d'un champ instantané de 30Gb

Ainsi, sans parler des soucis de mise en place d'une telle configuration, rien que le coût d'une petite intégration est déjà très important. L'assimilation de données, à l'aide de méthodes requiérant approximativement 100 fois le coût du modèle direct (que ce soit via les méthodes variationnelles ou d'ensemble) paraissent peu applicables de façon opérationnelle et encore moins dans un contexte plus prospectif, nécessitant plusieurs phases d'essais/erreurs.

Configurations	Temps CPU (en h/an)	RAM (totale, en Gb)	Stockage (en Gb)
SEABASS12 sans assim.	6	1.5	0.15
SEABASS12 avec assim.	600	6	0.15
SEABASS100 sans assim.	410	16	1
SEABASS100 avec assim.	41 000	64	1
ORCA4 sans assim.	5000	2.4	2
ORCA4 avec assim.	500 000	10	2
ORCA100 sans assim.	3 125 000	60	30
ORCA100 avec assim.	312 500 000	240	30

Table 7.1 – *Approximations de coûts numériques du modèle direct avec et sans assimilation pour SEABASS (configuration régionale) et ORCA (configuration globale) à différentes résolutions. Le stockage est donné pour un champ de sortie du modèle (direct ou linéaire tangent). Pour l'assimilation, nous considérons un cycle de 10 boucles externes et 10 itérations internes pour chaque.*

Qui plus est, en terme de calcul, nous avons supposé que cette configuration présentait une très bonne scalabilité. Si l'on passe à une configuration réaliste, en prenant en compte les conditions aux limites périodiques, le repli de la grille de discrétisation au nord, le temps de calcul ne va pas forcément diminuer de façon linéaire lorsque l'on augmente le nombre de processeurs,

2. Dans le contexte du calcul numérique, la scalabilité est l'aptitude d'un système à maintenir ses performances lorsque les ressources augmentent. Dans notre configuration, doubler le nombre de processeurs diminue par 2 le temps de calcul : elle présente quasiment une scalabilité optimale

les échanges d'information entre les processeurs et les noeuds devenant alors de plus en plus importants dans le temps effectif de calcul.

Ainsi, au vu des coûts numériques, la perspective d'une application réaliste des méthodes variationnelles ou d'ensemble à de telles résolutions paraît peu probable à court et moyen terme. Cependant, la puissance de calcul augmentant [3], il est envisageable de commencer à explorer l'efficacité et les caractéristiques de ces méthodes à l'aide de petites applications académiques, dans un futur proche.

7.2.3 Variabilité et aspects énergétiques de SEABASS100

Nous allons commenter ici les caractéristiques physiques de la configuration SEABASS au 1/100°, relativement au 1/4°, au 1/12° et au 1/24°. Sur la Figure 7.1, nous voyons qu'à cette résolution des structures très fines se développent partout dans le domaine. On observe également une recirculation beaucoup plus marquée qu'au 1/4° et au 1/12° de structures tourbillonnaires de l'est vers l'ouest autour du jet.

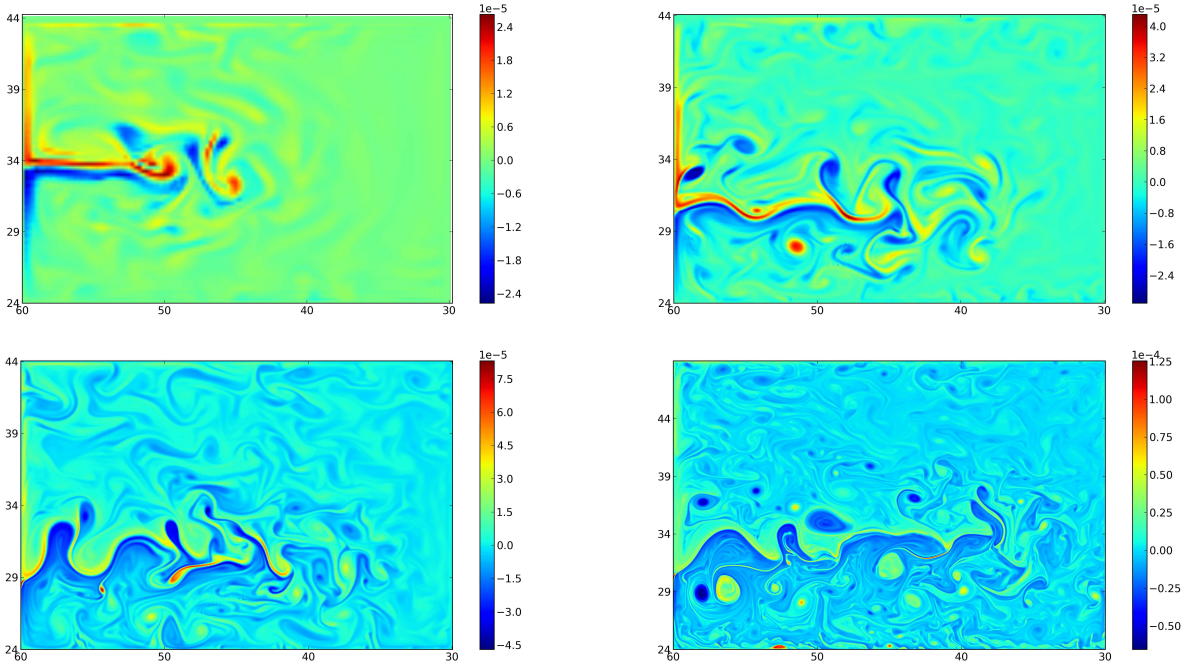

Figure 7.1 – *Vorticité relative instantanée (en s^{-1}), pour 4 résolutions horizontales différentes : 1/4°, 1/12° 1/24° et 1/100° (respectivement de gauche à droite et de haut en bas)*

Le point de décollement du jet, qui s'était décalé vers le sud entre le 1/4° et le 1/12° reste inchangé lorsque l'on passe au 1/24° et au 1/100°. En revanche, l'extension du jet ne cesse de s'accroître lorsque la résolution augmente. Au 1/100°, il s'étend au delà des 2/3 de la longueur du domaine.

Le problème sur lequel nous venons de discuter est celui de la séparation des courants de

3. Voir la loi empirique de Moore : *le nombre de transistors des microprocesseurs sur une puce de silicium double tous les deux ans.*

bord-ouest et de sa représentation physique et numérique. À l'heure actuelle, aucun modèle global ne la représente correctement sur l'ensemble des courants de bord-ouest. Ce problème est l'objet d'une littérature abondante, discutant des différentes causes physiques et de leurs représentations numériques, de ce biais, important pour l'océanographie dynamique :

- L'influence des conditions limites dans un contexte non-linéaire. Voir par exemple Verron et Le Provost (1991), Chassignet et Gent (1991), Verron et Blayo (1996) ou Adcroft et Marshall (1998).
- La représentation de la vorticité potentielle, discutée, entre autre, par Cessi et al. (1987) ou Kiss (2002)

D'autres origines physiques peuvent être également notées, telles que, par exemple, les variations de la topographie de fond ou la rencontre entre deux courants de bord-ouest. Pour un discussion exhaustive sur ce sujet, voir, par exemple, l'article de revue de Chassignet et Marshall (2008).

Enfin, de multiples états d'équilibres, accentuant la problématique de la représentation de la séparation du courant de bord-ouest, ont également été observés lors de simulations numériques, dont les transitions sont la conséquence de perturbations physiques et/ou numériques (conditions limites, erreurs d'arrondis). La mise en évidence de ces multiples états d'équilibres dans un modèle quasi-géostrophique a été faite dans les travaux, par exemple, de Jiang et al. (1995), ou plus récemment Nauw et al. (2004) et Deremble et al. (2012). Un travail en collaboration avec Bruno Deremble pour prolonger ce résultat dans un modèle aux équations primitives, à l'aide de SEABASS, a été initié au cours de cette thèse [4].

Sur la Figure 7.2 [5], nous pouvons remarquer qu'avec l'augmentation de la résolution, le jet est clairement de plus en plus déstabilisé, comme le montrent ses méandres dans l'EKE de surface. Si, au 1/4°, les structures tourbillonnaires s'échappent essentiellement de l'extrémité est du jet, à plus haute résolution, ces recirculations prennent plus d'importance tout le long du jet. On constate également une croissance des niveaux EKE dans la zone du jet avec l'augmentation de la résolution, d'un facteur 2 entre le 1/4° e le 1/24°. Ceci traduit l'apport énergétique de la dynamique méso-échelle à la circulation générale.

4. Non abouti à ce jour

5. N'ayant, pour des raisons de coût numérique, qu'un mois de sorties journalières pour SEABASS100, nous ne présentons pas l'EKE pour cette résolution

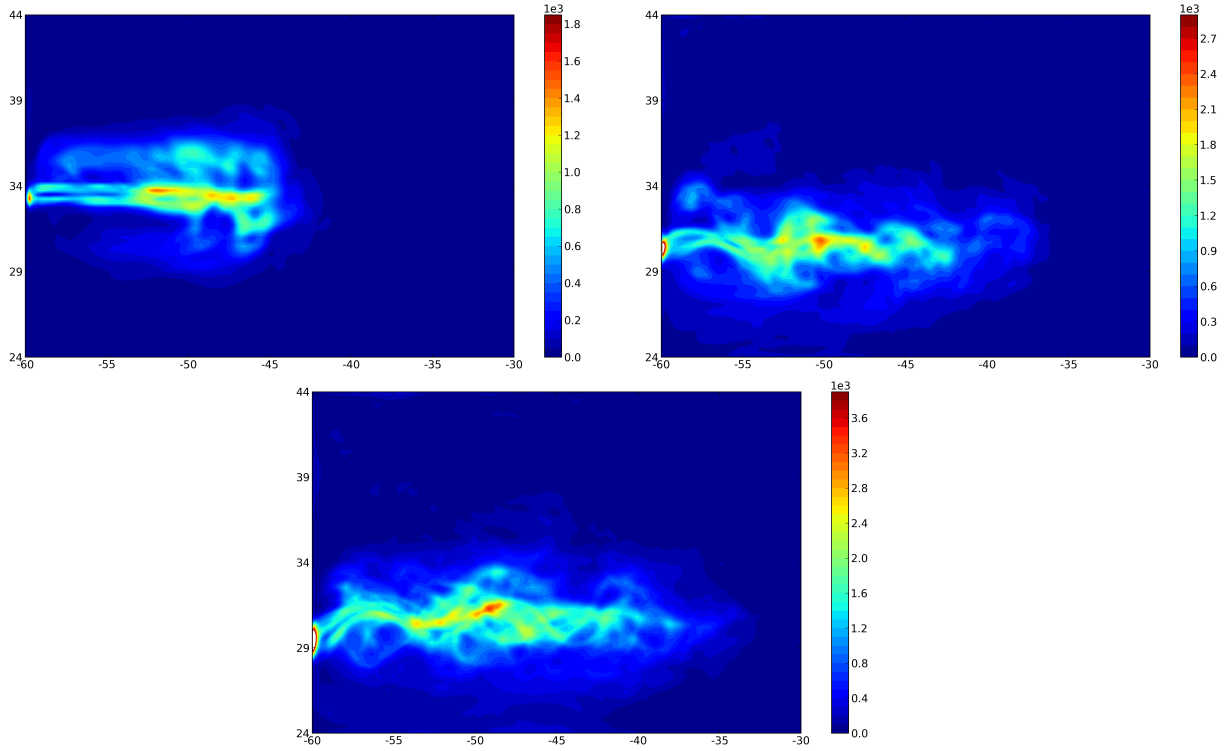

Figure 7.2 – *Champs d'énergie cinétique turbulente de surface (en $m^2.s^{-2}$), moyennés sur un an à partir de champs instantanés journaliers, pour 3 résolutions horizontales différentes : 1/4°, 1/12°, 1/24° (respectivement de gauche à droite et de haut en bas)*

Comme le montre la Figure 7.3, l'énergie contenue aux plus petites échelles, de l'ordre du km, est beaucoup plus importante. Il apparaît, de façon naturelle, qu'à mesure où la résolution augmente, l'énergie contenue dans les petites échelles augmente également. Si l'on s'intéresse aux pentes de ces spectres, nous remarquons qu'elles sont égales à k^{-5}, indiquant que nos simulations présentent une dynamique Quasi-Géostrophique et non SQG (pente attendue de k^{-4}). Ceci peut être expliqué par deux facteurs :

- La résolution verticale vers la surface du bassin est trop grossière
- La salinité étant constante

Ces deux points font que notre configuration ne semble pas représenter correctement les anomalies de densité de surface, prises en compte dans la théorie SQG.

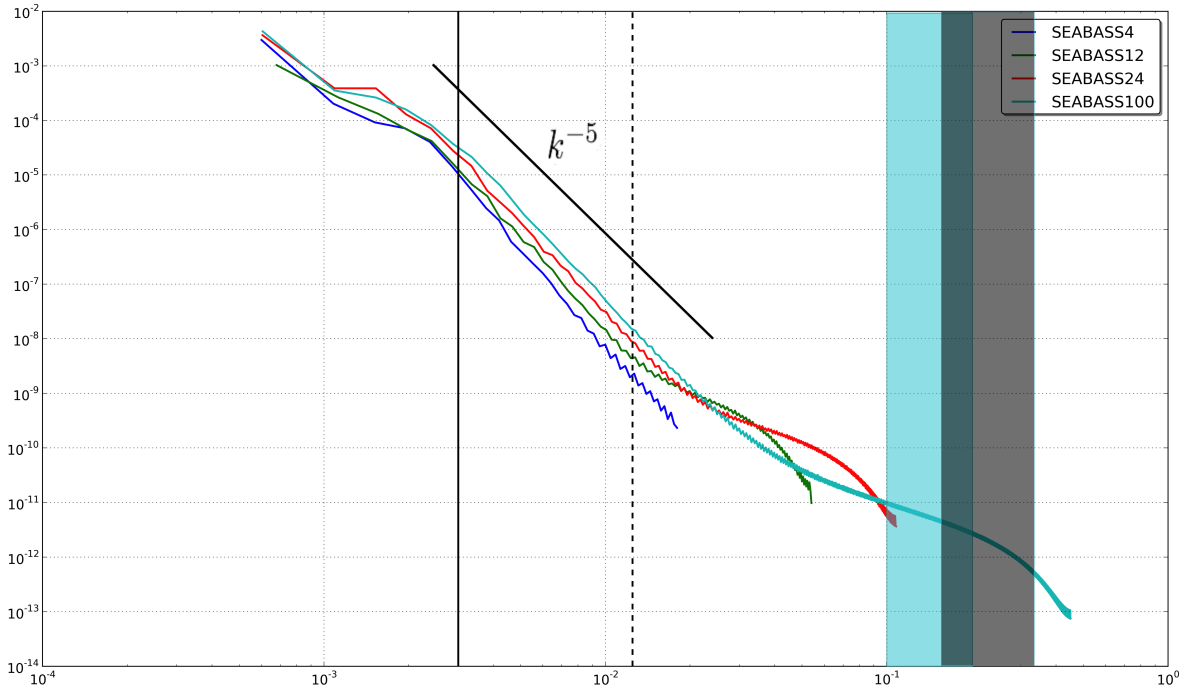

Figure 7.3 – *Densité de puissance spectrale de la SSH, pour 4 résolutions horizontales différentes (abscisse en cycle/km). Les traits verticaux correspondent aux plus petites échelles représentées dans les observations (trait plein pour les traces Jason-1 et pointillés pour les traces SARAL/AltiKA). La zone turquoise correspond à la résolution effective de SEABASS100. La zone grisée correspond aux plus petites échelles potentiellement représentées dans les observations issues du futur altimètre SWOT.*

Cette pente en k^{-5} indique que, pour la méso-échelle, il y a bien cascade énergétique inverse et aux plus petites échelles, une cascade énergétique directe. Ceci est cohérent avec les résultats obtenus auparavant : la circulation turbulente à méso-échelle influe bien sur l'état moyen du jet à plus grande échelle (ilustré par le fait que l'état moyen du jet est différent selon que l'on résoud ou non les dynamiques méso- et subméso-échelles).

7.3 Extrapolation des résultats obtenus pour la très haute résolution

À partir des résultats obtenus aux chapitres précédents, et compte-tenu des caractéristiques de SEABASS au 1/100°, nous allons ici essayer d'imaginer quelles sont les principales contraintes qui se poseraient pour les méthodes d'AD variationnelles dans un tel contexte.

7.3.1 Méthodes variationnelles d'AD et chaoticité

Nous venons de voir qu'au 1/100°, une part non-négligeable de l'énergie contenue dans la circulation océanique est à méso- et subméso-échelle. Or les structures se développant à ces échelles sont en grande partie gouvernées par le terme non-linéaire d'advection dans le système des équations primitives. Dans ce contexte, en augmentant la résolution, la chaoticité du modèle s'accroît à travers ces dynamiques. Autrement dit, une perturbation physique ou numérique

minime peut changer drastiquement la circulation océanique simulée. Ceci est déjà visible à partir du 1/12°.

Sensibilité aux perturbations numériques Grâce à un travail réalisé en collaboration avec Jean-Marc Molines [6], nous avons réalisé une série d'expériences sur SEABASS au 1/12°, mettant en évidence la sensibilité de la circulation simulée à une perturbation introduite dans le système. Pour paralléliser le calcul de la circulation dans les modèles océaniques, la stratégie majoritairement utilisée est la décomposition de domaine. Cette dernière consiste à découper le domaine horizontal en sous-domaines rectangulaires et à assigner à chaque processeur un sous-domaine donné. Les échanges d'informations entre les sous-domaines, nécessaires pour les schémas d'advection et de diffusion, sont réalisés à travers des halos de recouvrement. Ceux-ci consistent à inclure dans chaque sous-domaine quelques lignes et colonnes des sous-domaines voisins, dont le nombre dépend de l'ordre des schémas utilisés.

Nous avons alors réalisé une suite de simulations de type spin-up, différant uniquement sur la décomposition en sous-domaine considérée. En réalité, d'une décomposition à l'autre, nous introduisons de manière numérique une perturbation : les échanges entre sous-domaines ne se font plus nécessairement de façon identique d'une simulation à l'autre, les erreurs numériques d'arrondis viennent alors modifier légèrement les calculs.

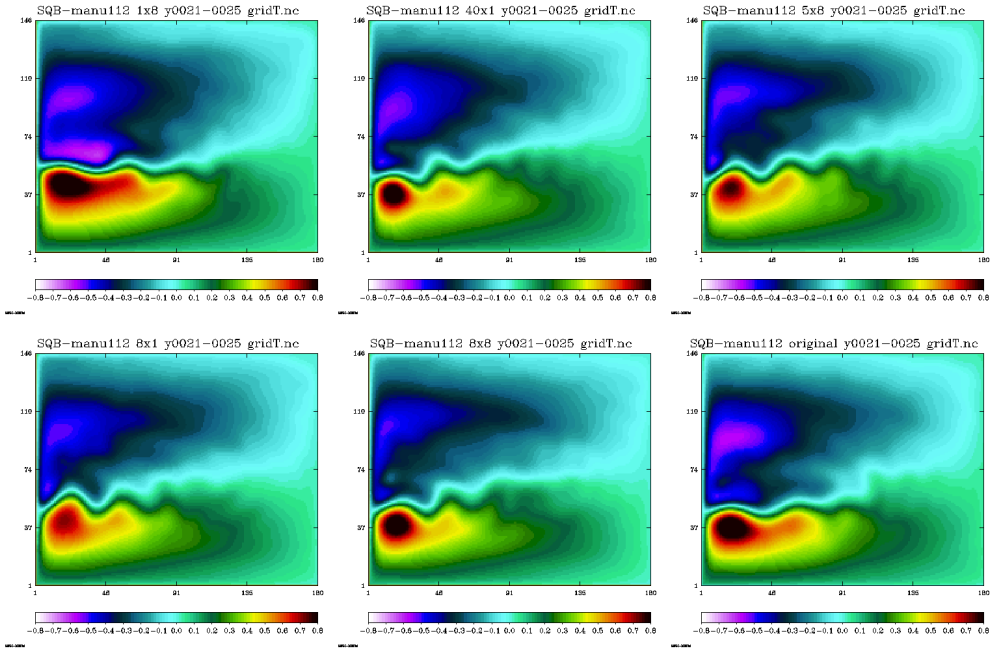

Figure 7.4 – *Champs moyens de SSH entre les années 21 et 25 d'une phase d'intégration libre du modèle, pour différentes décomposition de domaines au 1/12° : respectivement de gauche à droite et de haut en bas 1x8, 40x1, 5x8, 8x1, 8x8, 1x1. Communication de Jean-Marc Molines*

Il apparaît alors, sur la Figure 7.4, que l'état moyen du modèle est en réalité modifié dû à ces perturbations « artificielles » à la fois aux grandes comme aux petites échelles :

- Des différences notables dans le gradient de SSH sont remarquées dans les deux gyres nord et sud

6. *IR CNRS, équipe MEOM, LGGE*

– Les oscillations dans le jet zonal apparaissent ou s'accroissent dans certaines configurations, oscillations dont la longueur d'onde semble reliée au nombre de sous-domaines dans la direction longitudinale.

Une autre expérience confirmant la sensibilité du modèle aux perturbations numériques fut réalisée : entre deux simulations différentes, nous avons uniquement recompilé le code numérique en modifiant les paramètres d'optimisation du compilateur, toutes choses égales par ailleurs. Des différences significatives sur un état moyen peuvent alors être observées sur la Figure 7.5

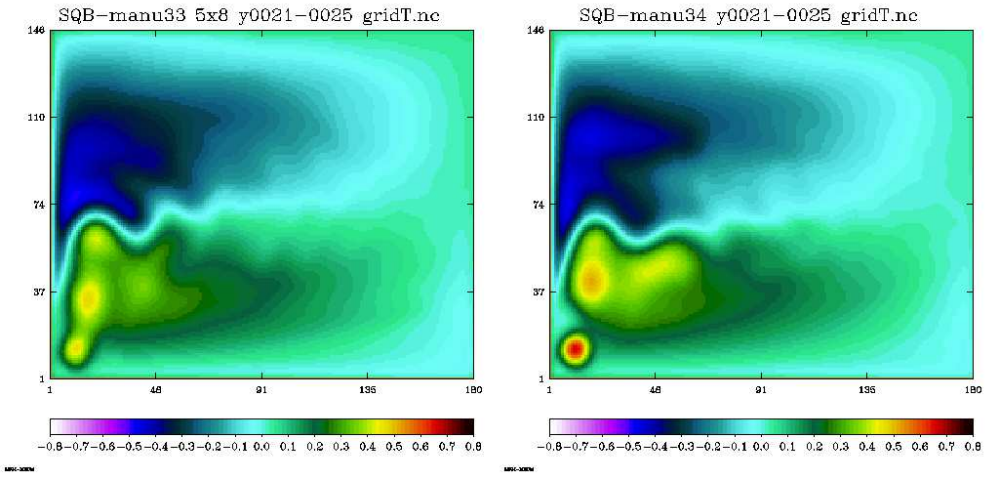

Figure 7.5 – *Champs moyens de SSH entre les années 21 et 25 d'une phase d'intégration libre du modèle, pour deux options d'optimisations de compilation différentes : -O3 à gauche, -O2 à droite, avec le compilateur Intel Fortran 11. Communication de Jean-Marc Molines*

Ainsi, nous voyons que le modèle peut donner différents états moyens à cause de perturbations numériques minimes et ce, dès le 1/12°. En augmentant la résolution, ce phénomène ne peut que s'accentuer.

Corrélation à la condition initiale Un autre effet de l'augmentation de l'aspect chaotique du modèle est que la trajectoire du modèle s'éloigne plus rapidement d'un état initial donné. Comme illustré sur la Figure 7.6, nous voyons que l'augmentation de la résolution accentue ce phénomène.

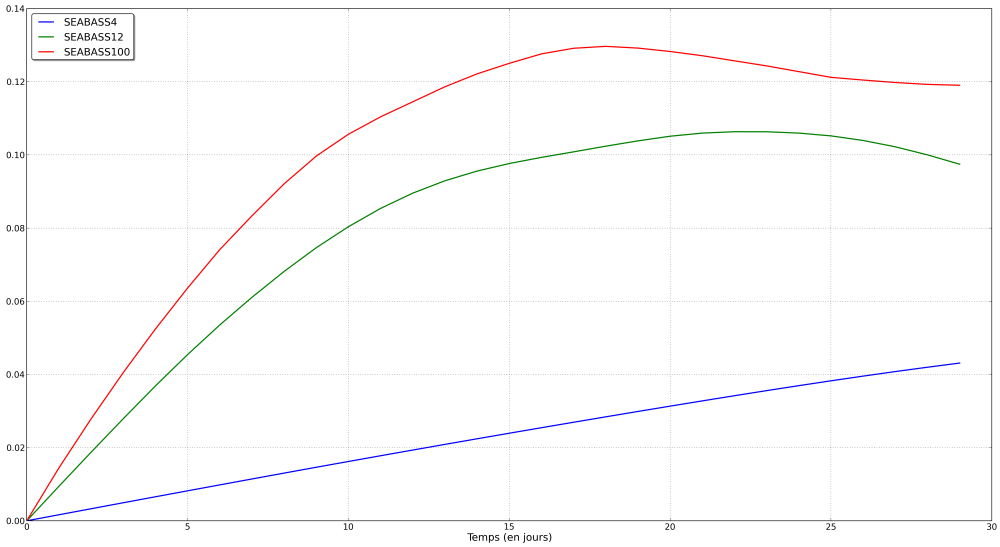

Figure 7.6 – *Évolution temporelle (en jours) de la moyenne spatiale de SSH de l'écart entre une condition initiale fixée et les états suivants, à différentes résolutions : 1/4° (en bleu), 1/12° (en vert) et 1/100° (en rouge)*

Alors qu'il faut 30 jours, au 1/4°, pour que cet écart atteigne un niveau de 4 cm, celui-ci est atteint dès le 3ème jour au 1/100°. À cette résolution, au bout de 10 jours, l'écart ainsi calculée est de 10 cm.

Conséquences pour les méthodes variationnelles de données Nous avons vu aux chapitres précédents que l'augmentation des non-linéarités du modèle posaient deux problèmes majeurs pour l'AD variationnelle :

- L'hypothèse linéaire tangente voit sa validité restreinte sur de plus courtes périodes temporelles ou pour de plus petites perturbations. Ainsi, la taille des fenêtres temporelles d'AD va être amenée à diminuer à mesure que la résolution augmente, pour éviter une trop grande divergence entre le modèle LT et le modèle non-linéaire. Des calculs sont en cours pour illustrer cet aspect.
- La fonction coût voit son nombre de minima locaux augmenter.

L'algorithme de minimisation n'est alors plus assuré de converger convenablement. Au chapitre précédent, nous avons vu qu'en prenant une ébauche « proche » de l'état vrai, il est possible d'accélérer la convergence de l'algorithme, en se situant dès le début à proximité du bassin d'attraction du minimum recherché, évitant ainsi de converger vers un minimum local trop éloigné. Les indications précédentes nous montrent qu'en augmentant la résolution, il devient de plus en plus difficile d'assurer la proximité entre l'ébauche et l'état vrai, la trajectoire du modèle se décorrélant rapidement d'un état initial donné.

Une ébauche de mauvaise qualité implique alors deux choses :

- La forte probabilité de converger vers un minimum local éloigné de l'état vrai
- La production d'incréments (*i.e.* de perturbations) de forte amplitude dont la propagation par le modèle linéaire tangent ne sera plus consistante avec celle réalisée par le modèle non-linéaire.

Dans ces conditions, il paraît indispensable de développer des stratégies pour rapprocher rapidement l'ébauche de l'état vrai, même si celui-ci est inconnu en expérience réaliste. Pour se faire, les stratégies telles que l'assimilation progressive peuvent alors apparaître comme incon-

tournables. Cependant, en gardant en tête le coût numérique à de telles résolutions, les algorithmes de type multi-échelle (multi-incrémental, multi-grilles) sont certainement à considérer, pour réduire le temps de calcul.

7.3.2 Très haute résolution et flux d'information

L'autre ingrédient du processus d'AD sont les observations. Au 1/100°, en supposant que le modèle peut résoudre les échelles spatiales jusqu'à $8\Delta x$, il peut représenter des structures de longueur d'onde de 8 km. On voit sur la Figure 7.3 que le réseau d'observations de type SARAL/AltiKA est capable de capter la majorité des échelles représentées dans les configurations au 1/4° et au 1/12°. En revanche, au delà, une partie significative de l'énergie contenue dans le modèle n'est plus représentée dans les observations. Il devient donc très délicat de contraindre les dynamiques de plus petites échelles par les observations, dans le cadre des méthodes d'AD. Avec le futur programme SWOT, le programme satellitaire altimétrique à haute résolution, la méso- et subméso-echelle, représentée dans SEABASS100 pourra être représentée dans les observations.

Nous avons évoqué plus haut que l'accroissement de la résolution des modèles implique une réduction de la fenêtre temporelle d'AD. Ceci pose une problématique fortement en lien avec les observations disponibles, à deux niveaux. Tout d'abord, si les fenêtres ne font pas la taille suffisante pour couvrir une période de répététivité du satellite, le réseau d'observation comportera alors des lacunes dans la couverture spatiale du domaine. Ces « trous » vont alors décroître le niveau de représentativité des observations en terme d'échelles spatiales. Il sera d'autant plus difficile alors de contrôler ces échelles, comme nous l'avons vu au chapitre précédent pour des fenêtres de 15 jours avec le satellite SARAL/AltiKA (d'une période de répétitivité de 30 jours). Ensuite, vient la question de la pénétration verticale de l'information lors du processus d'assimilation. La source majeure, en termes quantitatifs, des observations se fait, aujourd'hui et dans le futur proche, via l'altimétrie satellitaire, circonscrite à la surface de l'océan. Si, dans nos expériences, les fenêtres étaient suffisamment longues pour contrôler la circulation dynamique en profondeur à partir d'observations de la surface, ceci est moins évident pour des fenêtres plus courtes, ne couvrant pas suffisamment la gamme des longueurs temporelles en jeu.

Enfin, l'augmentation de la densité spatiale et temporelle des réseaux d'observations posent la question des erreurs d'observations. Dans bon nombre d'applications des méthodes d'AD, l'hypothèse de la décorrélation spatiale et temporelle des erreurs d'observation est faite. Celle-ci ne sera plus forcément justifiée pour des observations à haute résolution, captées par le même instrument de mesure. Ceci dit, certains travaux, comme ceux réalisés récemment dans le code NEMOVAR par Olivier Titaud et Anthony Weaver, commencent prendre en compte la corrélation spatiale des erreurs d'observations pour l'assimilation variationnelle. On peut citer également, pour l'assimilation séquentielle stochastique, les travaux de Brankart et al. (2009), par exemple.

7.4 Synthèse

L'augmentation drastique de la résolution spatiale des modèles pose de nombreuses problématiques aux systèmes d'AD en général, et variationnels en particulier. L'aspect de plus en plus chaotique des modèles rendent la minimisation de la fonction coût du 4DVAR incrémental de

plus en plus difficile :

- En mettant à mal l'hypothèse linéaire tangente et réduisant ainsi la taille acceptable des fenêtres temporelles d'AD
- En dégradant le conditionnement de la fonction coût et en augmentant le nombre de minima locaux.

Il devient alors de plus en plus ardu de converger vers l'état optimal recherché. À ceci, s'ajoute un coût numérique considérable, actuellement peu adapté à l'exploration méthodologique entreprise dans cette étude.

De la même manière, se pose la question de la résolution des systèmes d'observations. Aujourd'hui insuffisante pour faire face convenablement à des modèles au 1/100°, le programme SWOT (voir chapitre 1) fournira des produits permettant en théorie de contrôler jusqu'à la dynamique subméso-échelle. Cependant, pour des modèles à plus faibles résolutions, il conviendra de filtrer les produits altimétriques qui en sont issus pour prendre en compte l'erreur de représentativité entre les observations et les modèles. Enfin, à cause de la réduction probable des fenêtres d'assimilation d'AD, ces observations de surface, aussi denses soient-elles, ne pourraient répondre que partiellement à la problématique de propagation verticale de l'information dans le processus d'AD.

Pour répondre à toutes ces contraintes, tenant d'une part aux caractéristiques propres des modèles et des observations mais également à leur imbrication lors de l'AD, des techniques telles que l'assimilation progressive ou les méthodes multi-échelles sont à considérer sérieusement. D'une part pour se rapprocher le plus rapidement possible de l'état vrai, autour duquel la convergence vers l'état optimal est plus rapide, et d'autre part pour réduire le coût numérique, inhérent à ces résolutions spatiales.

Enfin, SEABASS apparaît ici comme un outil intéressant pour le développement de méthodes d'AD en océanographie, en permettant de suivre simplement, sur certains aspects, les tendances actuelles futures de l'océanographie physique. Il est à noter que cette configuration est utilisé dans la stratégie de benchmarks mise en place dans le projet européen SANGOMA, ayant comme objectif, entre autre, de comparer sur différents niveau de réalisme l'efficacité de diverses méthodes d'AD stochastiques séquentielles.

Conclusion

Rappels du contexte et des objectifs

L'océan est le terrain de jeu de dynamiques évoluant sur un large spectre d'échelles, spatiales et temporelles. La circulation générale de l'océan résulte de l'évolution et de l'interaction de ces dynamiques. Il convient alors, pour la comprendre et la prédire le plus précisément possible, d'accéder le plus largement possible à ces échelles spatiales et temporelles. Pour ce faire, que ce soit dans l'observation de l'océan ou dans sa modélisation, ceci passe par une augmentation accrue de la résolution spatiale et temporelle.

L'assimilation de données a pour objet de combiner de façon optimale les informations portées par les modèles, les observations et les erreurs qui leur sont associées afin d'estimer le plus précisément possible l'état de l'océan, en contrôlant la condition initale ou les paramètres des modèles, par exemple. Plusieurs grandes familles de méthodes, partageant de nombreux points communs permettent de réaliser cet objectif. Parmi elles, les méthodes variationnelles résument le problème ainsi posé par la minimisation d'une fonction coût intégrant l'écart entre l'océan simulé et les observations. Tout l'art de cette classe de méthodes d'AD revient alors à réaliser effectivement la minimisation cette fonction coût en fonction du vecteur de contrôle recherché, dans notre cas la condition initiale.

Dans le contexte de la dynamique de l'océan, augmenter la résolution des modèles et des observations revient à représenter la dynamique méso- (de l'ordre de 100 km) et submeso-échelle (de l'ordre de 10 à 1 km). À ces échelles, la circulation est essentiellement turbulente, *i.e.* non-linéaire. Ceci pose plusieurs problèmes aux méthodes d'AD, variationnelles en particulier :

- La fonction coût définie plus haut n'est pas quadratique et peut présenter de très nombreux minima locaux. La minimisation ne garantit plus de converger vers l'état optimal recherché.
- Dans le cadre du 4DVAR incrémental, la fonction coût totale est approchée par une suite de fonctions coûts quadratiques, à l'aide du modèle linéaire tangent. L'hypothèse linéaire tangente voit sa validité circonscrite à de plus courtes fenêtres temporelles à mesure que la résolution augmente.
- Enfin, l'augmentation de la résolution augmente *de facto* la dimension du problème d'AD. Ceci se traduit directement par un coût numérique croissant.

Dans le cadre de ce travail, nous nous intéressons aux difficultés posées par les non-linéarités de la circulation océanique sur les méthodes variationnelles. Pour ce faire, nous avons mis en place un cadre expérimental, en mettant en place une configuration océanique idéalisée présentant une dynamique turbulente caractéristique des séparations des courants de bord-ouest, comme celle du *Gulf Stream* au large du cap Hatteras, par exemple. Nous avons également choisi de nous

intéresser aux observations issues de l'altimétrie satellitaire (citons les programmes Jason-1, SARAL/AltiKA, ou le futur programme SWOT), permettant aujourd'hui de capter la circulation méso-échelle et dans un futur proche la circulation subméso-échelle.

Ce travail s'inscrit logiquement dans le contexte géographique Grenoblois, articulé autour des équipes MEOM (anciennement appartenant au LEGI [7] et, depuis 2013 au LGGE [8]) et MOISE (LJK [9]). Dans ce contexte, l'historique de l'étude des méthodes d'AD variationnelles, appliquées à l'océanographie, s'est étoffé sur plusieurs niveaux, méthodologiques et applicatifs :

- Des études méthodologiques pour l'océanographie ont été réalisées, dans des contextes idéalisés par Ngodock (1996), Luong et al. (1998) ou dans des contextes réalistes par Vidard (2001) ou Rémy et al. (2002).
- Durbiano (2001) ou Robert et al. (2005a) ont proposé des développements des méthodes variationnelles combinés avec d'autres approches méthodologiques
- Faugeras (2002) utilisent les méthodes variationnelles pour l'estimation de paramètres dans un modèle biogéochimique

À ceci s'ajoute les nombreux travaux portant sur les méthodes d'AD séquentielles stochastiques, portant sur des aspects méthodologiques comme dans, par exemple, Pham et al. (1998b) ou Cosme et al. (2010) ou applicatifs, réalisés par exemple dans Carmillet et al. (2001) ou Brankart (2003). Il est apparu, fort de cet historique de recherche en AD appliquée à l'océanographie, qu'il était pertinent de mener une étude méthodologique comparable à celle de Luong et al. (1998) prenant en compte l'évolution des modèles et des observations. C'est dans le cadre du projet ANR VODA (2009-2012), visant à contribuer au développement du système NEMOVAR et finançant mes premières années de thèse, que cette étude a été initiée.

Pour atteindre nos objectifs, nous avons réalisé plusieurs expériences jumelles d'AD, à plusieurs résolutions, afin de mettre en évidence les problèmes posés par les non-linéarités de la circulation océanique. Notre étude permet de les identifier et de les caractériser, sous différents angles de vues : méthodologiques, physiques et statistiques. Nous avons également proposé quelques pistes d'amélioration des systèmes variationnels d'AD pour répondre à ces enjeux.

Un travail d'ingénierie

Pour réaliser cette étude, j'ai dû prendre en main, utiliser et parfois développer (ou participer au développement) plusieurs outils logiciels communautaires. Le premier outil au centre de notre étude est le logiciel de modélisation océanique NEMO, basé sur les équations primitives de l'océan. Au sein de ce modèle, nous avons mis en place la configuration de bassin océanique idéalisée SEABASS, à différentes résolutions (1/4°, 1/12°, 1/24° et 1/100°). Cette configuration permet à la fois de représenter une activité tourbillonnaire statistiquement réaliste tout en restant simple à prendre en main, maintenir le tout pour un coût numérique raisonnable, permettant d'explorer la très haute résolution. SEABASS répond à un besoin de point de convergence pour les développements méthodologiques en AD autour de NEMO. C'est à ce titre qu'elle est inclue comme cas-test dans la stratégie de benchmarks proposée dans le projet européen SANGOMA.

7. Laboratoire des Écoulements Géophysiques et Industriels
8. Laboratoire de Glaciologie et de Géophysique de l'Environnement
9. Laboratoire Jean Kuntzmann

Il m'a également fallu acquérir les notions et problématiques de l'altimétrie satellitaire, ne serait-ce que pour simuler des traces selon des altimètres existants, mais également pour en retirer des conclusions réalistes pour les liens entre AD et échantillonnage altimétrique.

Pour mettre en place un système d'AD, certaines briques numériques sont indispensabes : l'opérateur d'observation, l'applicateur d'incrément (ou de perturbation) et, dans le cas variationnel, du modèle linéaire tangent et adjoint de NEMO. Suite à l'initiative du groupe NEMO-ASSIM, ces briques sont désormais disponibles dans le cœur numérique de NEMO (modules OBS, ASM, TAM). C'est dans ce cadre qu'au cours de cette thèse, j'ai été recruté en tant qu'IR CNRS pour assurer le développement, la maintenance et la documentation de NEMOTAM ainsi que participer aux activités techniques et de recherche portées par NEMO-ASSIM. Mentionnons ici quelques applications réelles de ces différents modules :

- L'opérateur d'observation OBS peut servir aux océanographes pour colocaliser les champs simulés par le modèle aux temps et lieux des observations, afin de tirer des informations statistiques comparables entre simulations et observations, à des buts d'analyse physique ou de calibration/validation des modèles. Il est bien sûr possible d'utiliser ce module pour toute expérience d'AD d'ensemble.
- TAM permet également des applications d'analyse de sensibilité et d'identification de paramètres.

C'est également dans le cadre de NEMO-ASSIM que l'inclusion de SEABASS dans le système NEMO comme configuration de référence pour l'AD.

Enfin, le système d'AD est basé sur NEMOVAR, un logiciel comprenant les briques essentielles à l'AD variationnelle :

- La modélisation des matrices de covariances d'erreur
- Différents préconditionnements de la fonction coût
- Les algorithmes de minimisation proprement dit.

Au travers de tous ces outils, nous apercevons que cette étude comporte une forte dimension technique. Il a fallu en effet prendre en main chaque logiciel, les comprendre et les modifier pour arriver à mettre en place le cadre expérimental que nous désirions. Enfin, il faut souligner la forte dimension communautaire, techniquement et scientifiquement, de tous ces éléments.

Résultats

Performances du 4DVAR incrémental en contexte non-linéaire Notre travail a d'abord consisté à caractériser les performances de notre système d'AD variationnelle appliqué à une circulation turbulente. Partant de la configuration SEABASS au 1/4°, le 4DVAR donne de bons résultats en terme de réduction d'erreur d'analyse et de prévision pour des fenêtres longues d'un et deux mois. Le niveau d'erreur d'analyse de SSH après assimilation est d'un centimètre, alors que le niveau d'erreur d'observation prescrit est de 3 cm. Il est apparu qu'au cours d'une période d'AD, le 4DVAR incrémental corrige en premier lieu les grandes échelles de l'écoulement (supérieures à 500 km) puis se concentre sur les moyennes échelles (inférieures à 500km). L'information portée par les observations de surface permet également de contrôler efficacement les variables non-observées en profondeur.

Lorsque l'on augmente la résolution de SEABASS au 1/12°, le système d'AD fait alors face à des difficultés de minimisation : l'algorithme converge beaucoup plus lentement qu'au 1/4°, voire diverge sous certaines conditions. Il est à noter que l'augmentation du nombre d'itérations de minimisation n'améliore pas la qualité de convergence de l'algorithme.

Cependant, malgré ces difficultés de convergence, l'algorithme permet un contrôle effectif de l'écoulement, pour des fenêtres d'AD d'un mois. Pour les expériences favorables, le niveau moyen de l'erreur d'analyse de SSH est d'environ 4 cm, sachant que l'erreur d'ébauche est d'un niveau moyen de 15 cm. Là encore, l'information portée par les observations de surface pénètre verticalement de façon convenable au cours de l'assimilation.

Pour tenter d'améliorer les perfomances du 4DVAR au 1/12°, nous avons mis en place la stratégie dite d'assimilation progressive, étudiée par Pires et al. (1996) en métérologie et Luong et al. (1998) en océanographie. Dans notre cas, une telle stratégie améliore significativement l'erreur d'analyse sur la variable observée. En revanche, sur les variables non-observées en profondeur, elle induit une légère dégradation de l'erreur d'analyse, contrairement à ce qu'avait pu analyser Luong et al. (1998). Nous n'avons pas encore eu le temps d'explorer plus avant ce phénomène.

Qualité de l'ébauche vs. qualité de convergence Au 1/12°, les non-linéarités qui se développent dans le modèle mènent à une fonction coût présentant potentiellement de très nombreux minima locaux. Notre algorithme peut alors converger vers des états sous-optimaux, qui plus est lorsque l'ébauche est éloignée de l'état vrai. Nous avons montré que nous convergeons très bien lorsque l'ébauche est justement proche de l'état vrai. Ainsi, la qualité de l'ébauche est déterminante quant aux performances du 4DVAR.

Cependant, dans notre système, la confiance accordée au terme d'ébauche dans la fonction coût n'évolue pas au cours des cycles d'AD. Ainsi, même si nous convergeons vers l'état optimal, les erreurs d'observations vont alors être surreprésentées et faire diverger l'algorithme de cet état optimal. Une solution pour pallier ce problème est de mettre un poids devant le terme d'ébauche de la fonction coût, et de le faire varier au cours de l'assimilation, selon que l'état courant est proche ou non de l'état optimal. Une mesure de cette caractéristique peut être faite *via* l'amplitude des incréments obtenus au cours de la minimisation.

Impact de l'échantillonnage altimétrique sur le 4DVAR Un autre aspect du système d'AD peut influer sur ses performances : l'adéquation entre les échelles captées par les observations et celles représentables par le modèle. Nous avons mis en évidence que lorsque les observations contenaient les mêmes échelles que celles résolues par le modèle, l'erreur d'analyse était significativement réduite. À l'inverse, si les observations ne résolvent pas les échelles représentées par le modèle, le système a plus de difficultés à contrôler l'écoulement. En revanche, si les observations contiennent une part non-négligeable des échelles non-résolues par le modèle, ceci peut mener à une dégradation du conditionnement de la fonction coût et ralentir la convergence de l'algorithme de minimisation.

À l'aide de la résolution effective du modèle et de celle des observations, il faut donc autant que faire se peut mettre l'échantillonnage des observations en phase avec les échelles représentées dans le modèle.

3DFGAT vs. 4DVAR Il est apparu au cours de nos expériences que le 4DVAR présentait de bien meilleures performances que le 3DFGAT. Au 1/4°, si le 3DFGAT permet de contrôler l'écou-

lement, le niveau moyen d'erreur d'analyse reste 10 à 15 fois supérieur à celui obtenu avec le 4DVAR. Au 1/12°, ceci se vérifie également. Cet aspect n'est pas anodin, sachant que l'algorithme 3DFGAT est aujourd'hui celui majoritairement utilisé par les centres d'océanographie opérationnelle pratiquant les méthodes d'AD variationnelles.

Les limites de notre étude Il faut souligner ici les limites que comporte notre étude. Tout d'abord, la configuration à laquelle nous nous sommes intéressé est encore loin d'une configuration océanique réaliste, sous de nombreux aspects :

- Elle ne présente pas de bathymétrie, ni de topographie côtière
- Le forçage est constant et analytique
- La dimension verticale de notre dynamique est loin de pouvoir représenter tous les processus à l'œuvre dans l'océan
- ...

Ainsi, nous ne prenons pas en compte des aspects importants régissant l'écoulement océanique.

Aussi, même si les observations altimétriques regardent une quantité intégrée sur la verticale (la SSH) de la dynamique de l'océan, elles restent des mesures de surface. Il serait utile de compléter notre étude par des profils verticaux, afin de voir l'impact de l'observation verticale de l'océan sur les performances du système d'AD.

Il faut également noter que nous avons réalisé des expériences jumelles, en supposant le modèle parfait, c'est à dire en supposant l'erreur modèle nulle. Les méthodes variationnelles à contraintes fortes (comme celles utilisées dans cette thèse) ne prennent pas en compte l'erreur modèle. De nombreux travaux s'intéressent à cette question : nous pouvons citer en AD variationnelle Trémolet (2007b), pour la météorologie, par exemple ou Brankart (2013) pour les méthodes d'ensemble, en océanographie. En ce qui concerne les méthodes variationnelles, l'algorithme du 4DVAR à contrainte faible prend explicitement en compte l'erreur modèle.

Cette étude comporte ainsi certains aspects exploratoires encore ouverts. Nous pouvons citer par exemple :

- Compréhension plus fine des mécanismes de l'assimilation progressive
- Influence de l'évolutivité de $\mathbf{B}$
- Application du système d'AD à la très haute résolution (1/100°)

Enfin, de façon plus générale, notre travail est, et ne peut être, qu'essentiellement expérimental. Même si cet état de fait est inhérent à l'application des méthodes d'AD aux systèmes géophysiques, présentant un fort caractère non-linéaire, ceci limite la portée généraliste des conclusions tirées à l'issue de cette étude.

Les limites de ce travail ont le mérite de poser de nombreuses questions et d'ouvrir de larges perspectives que nous évoquerons par la suite.

Les grands défis présents et futurs de l'AD

Les résolutions étudiées ici sont celles actuellement ou très prochainement utilisées en océanographie physique, que ce soit pour la recherche ou par les centres opérationnels. Celles-ci sont amenées à croître dans un futur proche, de façon à représenter correctement les dynamiques

méso- et subméso-échelles. Cet enjeu majeur de l'océanographie apporte avec lui plusieurs problématiques pour les méthodes d'AD dont nous listons ici deux d'entre elles :

1. Augmenter la résolution spatiale et temporelle a un coût scientifique et numérique qui n'est pas sans implication sur l'océanographie physique et l'AD.
2. La très haute résolution amène avec elle une circulation océanique de plus en plus turbulente, chaotique. Pour l'AD, ceci amène des contraintes fortes : les méthodes d'AD aujourd'hui applicables à l'océanographie posent de nombreuses hypothèses qui sont invalidées dans ce nouveau contexte, comme la gaussianité des erreurs ou (quasi-)linéarité des modèles.

Le premier point soulève de nombreuses questions. La première est bien évidemment celle du coût numérique. Si la loi de Moore peut nous faire techniquement espérer l'investigation de la très haute résolution, elle n'est qu'empirique. Ensuite, l'augmentation des ressources numériques se fait également par des changements de paradigmes d'architectures matérielle et logicielle. Ainsi, accéder à la très haute résolution requiert un travail d'ingénierie très amont de la science océanographique elle-même.

Dans un second temps, le volume croissant de données disponibles pose la question de leur traitement. Rien que dans cette étude, le volume de données dépasse plusieurs Téra-octets. Outre leur manipulation, qui peut s'avérer chronophage et gourmande en terme numérique, nous ne pouvons prétendre en avoir tiré toutes les informations utiles à l'AD en océanographie. Ainsi comment s'assurer de tirer le plus d'information possible de données de plus en plus volumineuses sans se perdre en diagnostics ? Hormis le travail communautaire sur un même jeu de données, nous avons essayé ici d'utiliser des diagnostics synthétiques (EOF, analyse spectrale) résumant l'information de façon claire pour un coût raisonnable. Un enjeu du traitement de données est de mettre au point (ou d'utiliser) des stratégies de *data mining* efficaces et appropriées à nos données.

Au delà des diagnostics *a posteriori*, le volume croissant des données peut devenir une contrainte forte dans leur manipulation au sein de l'AD elle-même. En effet, l'altimétrie fournit (et fournira) un nombre colossal de données, dont l'assimilation n'est pas évidente, en terme de coût numérique et de paramérisations de leurs erreurs. Ainsi, plutôt que de considérer chaque mesure en tant que telle, plongée dans l'espace du modèle, il peut être (et sera sûrement) judicieux de traiter également ces données de façon structurée. L'assimilation d'images, consistant à tirer une information physique pertinente d'une image et à l'assimiler dans le modèle, est une voie en ce sens, qui pourra être utilisé pour des satellites comme SWOT. À ce sujet, nous pouvons citer les travaux récents de Gaultier et al. (2013).

En ce qui concerne le second point, nous avons vu dans le dernier chapitre les conséquences potentielles de la très haute résolution sur l'AD variationnelle. Premièrement, l'accroissement du nombre de minima locaux dû aux fortes non-linéarités du modèle est un vrai défi pour les algorithmes de minimisation basés sur l'information portée par le gradient. Ensuite, le coût numérique rend l'AD, de façon opérationnelle, encore inaccessible à de telles résolutions aujourd'hui. Des problématiques équivalentes sont rencontrées par les méthodes séquentielles. Nous verrons dans le paragraphe suivant les perspectives permettant d'affronter ces problématiques dans le cas des méthodes variationnelles.

Perspectives

Comme nous venons de le mentionner, l'AD variationnelle doit, aujourd'hui et demain, faire face à l'augmentation drastique du coût numérique, ceci couplé à des difficultés de convergence. Pour répondre à ces contraintes, plusieurs voies sont envisageables.

L'assimilation progressive est une voie prometteuse pour l'amélioration de la convergence à haute résolution. Même si aucun résultat théorique ne garantit son efficacité et que son coût numérique est plus élevé que l'AD purement séquentielle, elle permet tout de même d'améliorer les performances de nos systèmes d'AD.

La prolongation directe de ce travail à plus haute résolution est une première voie. Ceci passe par la définition d'expériences jusqu'au 1/100°, basé sur des observations comme le programme SWOT pourra en fournir à l'avenir. Dans ce cadre, une réflexion approfondie sur les diagnostics à employer doit être envisagée. Cet axe de travail resterait toutefois assez centré sur les aspects méthodologiques.

À moyen terme, une autre voie prometteuse est celle de l'AD multi-échelles. Citons l'algorithme du 4DVAR multi-incrémental, par exemple. Son principe est de réaliser la boucle externe du 4DVAR incrémental à la plus haute résolution et de réaliser la minimisation des fonctions coûts quadratiques à une résolution dégradée. On peut noter tout d'abord que cette méthode réduit le coût de calcul comparativement à une expérience intégralement à haute résolution. Ensuite, nous avons vu qu'à plus basse résolution, la convergence de l'algorithme de minimisation était plus rapide. Ainsi, une telle méthode, malgré son manque d'assises théoriques, permet de répondre aux deux enjeux forts cités plus haut. Bien connue en météorologie, aucune publication, à notre connaissance, n'explore cette méthode en océanographie. D'autres méthodes, telles que l'assimilation multi-grille, explorée dans Neveu (2011) vont également dans ce sens.

À moyen terme, les méthodes d'AD hybrides, combinant les principes de la minimisation d'une fonction coût et des méthodes stochastiques, présentent également quelques intérêts dans ce contexte. Des travaux passés, notamment réalisés par Krysta et al. (2008), conjuguaient les caractéristiques des méthodes séquentielles de filtrage et des méthodes variationnelles, permettant ainsi l'évolutivité consistante de $\mathbf{B}$ au sein du 4DVAR. Actuellement, de nouvelles hybridations sont en pleine exploration et nous pouvons citer deux façons de voir ces méthodes hybrides :

- Faire un ensemble d'expériences d'AD de type 4DVAR. Cette stratégie se nomme *En-4DVAR*. Cette méthode permet d'accéder aux incertitudes sur l'analyse issue de l'AD. Si elle ne répond pas spécialement aux problématiques qui nous intéressent, le gain en terme d'information sur l'AD est important, au prix d'un effort numérique substantiel.
- Faire un expérience de type 4DVAR en approchant le gradient de la fonction coût par une méthode de type monte-carlo, basé sur une simulation d'ensemble. Cette méthode réduit effectivement le coût numérique du 4DVAR classique, en se passant de l'intégration des modèles linéaire tangent et adjoint. Cependant, les difficultés de convergence ne peuvent que s'accroître si l'on se passe du gradient exact de la fonction coût. Cet argument est développé plus avant dans Laloyaux (2012) et Gratton et al. (2013).

De façon beaucoup plus prospective, dépendante en premier lieu des ressources numériques aujourd'hui loin d'être envisageables, l'optimisation stochastique peut apparaître comme une voie prometteuse. En prenant comme exemples rudimentaires un algorithme de type recuit si-

mulé, ou des algorithmes génétiques, cette classe de méthodes d'optimisation permettrait de s'affranchir des problèmes liés à l'irrégularité de la fonction coût et permettrait également d'envisager de poser le problème d'optimisation d'une nouvelle manière (multi-critères, multi-objectifs). Plus largement, si cette approche devient un jour envisageable, elle permettrait de reconsidérer l'architecture théorique des méthodes variationnelles.

Enfin, plus concrètement, je m'inscris de façon effective dans deux contextes, liés entre eux.

En tant qu'IR NEMO-ASSIM, il me paraît important d'inscrire ces perspectives dans le contexte communautaire de l'AD en océanographie avec NEMO. Notamment, la poursuite des efforts de communication et de développement d'outils facilitant la discussion et le développement à la fois méthodologique et applicatifs des différentes perspectives d'AD me paraît important. Des initiatives telles que présenter SEABASS comme configuration de référence pour l'AD avec NEM0 ou mettre à disposition des chercheurs et ingénieurs les éléments essentiels à l'AD, entre autre, fait parti de mes priorités.

Ensuite, toujours en tant qu'IR, mes activités de recherche s'inscrivent au sein de l'équipe MEOM à laquelle je suis rattachée. Cette équipe est composée d'océanographes et d'assimilateurs, travaillant aussi bien sur les aspects amont de l'océanographie physique (modélisation, développement méthodologiques en AD) qu'aval (*e.g.* réalisation de simulations, distributions de configurations NEMO). Cette équipe s'inscrit dans des projets d'études, physiques et méthodologiques, réalistes des circulations océaniques, que ce soit en terme d'observation ou de modélisation. Dans ce contexte, l'extension de notre étude à une application réaliste paraît naturelle. À ceci s'ajoute l'historique des travaux de MEOM en AD pour l'océanographie développée plus haut. Ainsi, La mise en place du système d'AD variationnel appliqué à une configuration globale ou régionale à haute résolution (*e.g.* 1/4° ou 1/12°), associée à l'utilisation d'observations réelles produites par les programmes Jason-1-2 ou SARAL/AltiKA ou simulées de type SWOT, permettrait de compléter de nombreux aspects de notre étude : effets des phénomènes physiques et processus que nous n'avons pas pris en compte (*e.g.* forçages, bathymétrie), influence effective de la circulation turbulente sur le système de modélisation. Du point de vue de l'AD, ceci permettrait au sein de l'équipe, de confronter les caractéristiques de différents systèmes d'AD sur une configuration réaliste. Qui plus est, un tel travail se met réellement au service de l'océanographie physique, tant du point de vue de la recherche que de l'application opérationnelle.

Bibliographie

Adcroft, A. et D. Marshall, 1998 : How slippery are piecewise-constant coastlines in numerical ocean models ? *Tellus A*, **50**, 95–108.

Alvera-Azcárate, A., A. Barth, M. Rixen, et J. Beckers, 2005 : Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions : application to the Adriatic Sea surface temperature. *Ocean Modelling*, **9**, 325–346.

Andersson, E., M. Fisher, R. Munro, et A. McNally, 2000 : Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, adn the explanation of a case of poor convergence. *Quarterly Journal of the Royal Meteorological Society*, **126**, 1455–1472.

Aumont, O. et L. Bopp, 2006 : Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, **20**.

Auroux, D. et J. Blum, 2005 : Back and forth nudging algorithm for data assimilation problems. *C.R. Acad. Sci. Paris*, **I**, 873–878.

Balmaseda, M. A., K. Mogensen, et A. T. Weaver, 2013 : Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, 1132–1161.

Blayo, E., J. Verron, et J.-M. Molines, 1994 : Assimilation of TOPEX/poseidon altimeter data into a circulation model of the north atlantic. *Journal of Geophysical Research*, **24**, 691–705.

Bloom, S. C., L. L. Takacs, A. M. da Silva, et D. Ledvina, 1996 : Data Assimilation Using Incremental Analysis Updates. *Monthly Weather Review*, **124**, 1256–1271.

Blum, J., B. Luong, et J. Verron, 1998 : Variational assimilation of altimeter data into a non-linear ocean model : temporal strategies. *ESAIM : Proceedings*, **4**, 21–57.

Brankart, J.-M., 2003 : Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models : Application to a 1993–1996 hindcast of the North Atlantic Ocean circulation. *Journal of Geophysical Research*, **108**, 3074, doi :10.1029/2001JC001198.
URL `http://doi.wiley.com/10.1029/2001JC001198`

— 2013 : Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modelling. *Ocean Modelling*, **66**, 64–76.

Brankart, J.-M., E. Cosme, C.-E. Testut, et J. Verron, 2010 : Efficient Adaptive Error Parameterizations for Square Root or Ensemble Kalman Filters : Application to the Control of Ocean Mesoscale Signals. *Monthly Weather Review*, **138**, 932–950.

Brankart, J.-M., C. Ubelmann, C.-E. Testut, E. Cosme, P. Brasseur, et J. Verron, 2009 : Efficient Parameterization of the Observation Error Covariance Matrix for Square Root or Ensemble Kalman Filters : Application to Ocean Altimetry. *Monthly Weather Review*, **137**, 1908–1927.

Brasseur, P., 2006 : Ocean data assimilation using sequential methods based on the Kalman Filter. *Ocean weather forecasting :An integrated view of oceanography.*, J. Verron et E. P. Chassignet, eds., Springer, chapter Ocean weat.

Brasseur, P. et J. Verron, 2006 : The SEEK filter method for data assimilation in oceanography : a synthesis. *Ocean Dynamics*, **56**, 650–661.

Buizza, R., 1995 : Optimal perturbation time evolution and sensitivity of ensemble prediction to perturbation amplitude. *Quarterly Journal of the Royal Meteorological Society*, **121**, 1705–1738.

Cao, Y., J. Zhu, I. M. Navon, et Z. Luo, 2007 : A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids*, **53**, 1571–1583.

Capet, X., J. C. McWilliams, M. J. Molemaker, et a. F. Shchepetkin, 2008 : Mesoscale to Submesoscale Transition in the California Current System. Part I : Flow Structure, Eddy Flux, and Observational Tests. *Journal of Physical Oceanography*, **38**, 29–43.

Carmillet, V., J.-M. Brankart, P. Brasseur, H. Drange, G. Evensen, et J. Verron, 2001 : A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical–biochemical model of the North Atlantic ocean. *Ocean Modelling*, **3**, 167–192.
URL `http://www.sciencedirect.com/science/article/pii/S1463500301000075`

Cessi, P., G. Ierley, et W. Young, 1987 : A Model of the Inertial Recirculation Driven by Potential Vorticity Anomalies. *Journal of Physical Oceanography*, **17**, 1640–1652.

Chassignet, E. et D. Marshall, 2008 : Gulf Stream separation in numerical ocean models. *Ocean Modeling in an Eddying Regime*, 39–62.

Chassignet, E. P. et P. R. Gent, 1991 : The Influence of Boundary Conditions on Midlatitude Jet Separation in Ocean Numerical Models. *Journal of Physical Oceanography*, **21**, 1290–1299.

Chassignet, E. P. et J. Verron, 1998 : *Ocean Modeling and Parameterization*. Kluwer.

— 2006 : *Ocean Weather Forecasting*. Springer.

Colling, A., 2004 : *Ocean Circulation*. Butterworth - Heinemann.

Cosme, E., J.-M. Brankart, J. Verron, P. Brasseur, et M. Krysta, 2010 : Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation. *Ocean Modelling*, **33**, 87–100.

Courtier, P., J. Thépaut, et A. Hollingsworth, 1994 : A strategy for operational implementation of 4D-Var , using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, **120**, 1367–1387.

Daget, N., A. T. Weaver, et M. A. Balmaseda, 2009a : Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Quarterly Journal of the Royal Meteorological Society*, **135**, 1071–1094.

— 2009b : Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Quarterly Journal of the Royal Meteorological Society*, **135**, 1071–1094.

Deltel, C., 2002 : *Estimation de la circulation dans l'océan Atlantique Sud par assimilation variationnelle de données in situ. Impact du contrôle optimal des forçages et de l'hydrologie aux frontières ouvertes. 185 pp.*. Ph.D. thesis, Univ. de Bretagne Occidentale.

Derber, J. et F. Bouttier, 1999 : A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus A*, **51**, 195–221.

Derber, J. C., 1987 : Variational four-dimensional analysis using quasi-geostrophic constraints. *Monthly weather review*, **115**, 998–1008.

Deremble, B., E. Simonnet, et M. Ghil, 2012 : Multiple equilibria and oscillatory modes in a mid-latitude ocean-forced atmospheric model. *Nonlinear Processes in Geophysics*, **19**, 479–499.

Desroziers, G. et S. Ivanov, 2001 : Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, **127**, 1433–1452.

Di Lorenzo, E., A. M. Moore, H. G. Arango, B. D. Cornuelle, A. J. Miller, B. Powell, B. S. Chua, et A. F. Bennett, 2007 : Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS) : Development and application for a baroclinic coastal upwelling system. *Ocean Modelling*, **16**, 160–187.

Dickey, T., M. Lewis, et G. Chang, 2006 : Optical oceanography : Recent advances and future directions using global remote sensing and in situ observations. *Reviews of Geophysics*, **44**, RG1001.

Dobricic, S. et N. Pinardi, 2008 : An oceanographic three-dimensional variational data assimilation scheme. *Ocean Modelling*, **22**, 89–105.

Ducet, N. et P. Le Traon, 2001 : A comparison of surface eddy kinetic energy and Reynolds stresses in the Gulf Stream and the Kuroshio Current systems from merged TOPEX/Poseidon and ERS-1/2 altimetric data. *Journal of geophysical research*, **106**, 603–622.

Durbiano, S., 2001 : *Vecteurs caractéristiques de modèles océaniques pour la réduction d'ordre en assimilation de données*. Ph.D. thesis, Université de Grenoble.

Evensen, G., 1994 : Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**, 10143–10162.

— 2003 : The Ensemble Kalman Filter : theoretical formulation and practical implementation. *Journal of Geophysical Research*, **53**.

Faugeras, B., 2002 : *Assimilation variationnelle de données dans un modèle couplé océan-biogéochimie*. Ph.D. thesis, Université de Grenoble.

Ferron, B., 2011 : A 4D-variational approach applied to an eddy-permitting North Atlantic configuration : Synthetic and real data assimilation of altimeter observations. *Ocean Modelling*, **39**, 370–385.

Fichefet, T. et M. M. Maqueda, 1997 : Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *Journal of Geophysical Research*, **102**, 12609–12646.

Fisher, M., 1998 : Minimization algorithms for variational data assimilation. *Recent Developments in Numerical Methods for Atmospheric Modelling*, ECMWF, 364–385.

Forget, G., B. Ferron, et H. Mercier, 2008 : Combining argo profiles with a general circulation model in the north atlantic. part1 : estimation of hydrographic and circulation anomalies from synthetic profiles, over a year. *Ocean Modelling*.

Gaultier, L., J. Verron, J.-M. Brankart, O. Titaud, et P. Brasseur, 2013 : On the inversion of submesoscale tracer fields to estimate the surface ocean circulation. *Journal of Marine Systems*, **126**, 33–42.

Gauthier, P., C. Charette, L. Fillion, P. Koclas, et S. Laroche, 1999 : Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I : The global analysis. *Atmosphere-Ocean*, **37**, 103–156.

Giering, R. et T. Kaminski, 1998 : Recipes for adjoint code construction. *ACM Trans. Math. Softw.*, **24**, 437–474.

Gill, A., J. Green, et A. Simmons, 1974 : Energy partition in the large-scale ocean circulation and the production of mid-ocean eddies. *Deep Sea Research and Oceanographic Abstracts*, **21**, 499–528.

Gilmour, I., 2001 : Linear regime duration : Is 25 hours a long time in synoptic weather forecasting ? *American Meteorological Society*, **58**, 3525–3539.

Gouretski, V. et F. Reseghetti, 2010 : On depth and temperature biases in bathythermograph data : Development of a new correction scheme based on analysis of a global ocean database.

Gratton, S., P. Laloyaux, et A. Sartenaer, 2013 : Derivative-free optimization for large-scale nonlinear data assimilation problems. *Quarterly Journal of the Royal Meteorological Society*, doi :10.1002/qj.2169.

Hascoët, L. et V. Pascual, 2004 : Tapenade 2.1 user's guide. Technical Report 0300, INRIA.

Ide, K., P. Courtier, M. Ghil, et A. C. Lorenc, 1997 : Unified Notation for Data Assimilation : Operationnal, Sequential and Variational. *J. Met. Soc. Japan*, 1–20.

Ishikawa, Y., T. Awaji, et N. Komori, 2001 : Dynamical Initialization for the Numerical Forecasting of Ocean Surface Circulations Using a Variational Assimilation System. *Journal of Physical Oceanography*, **31**, 75–93.

Jazwinski, A. H., 1970 : *Stochastic processes and filtering theory.*

Jiang, S., F.-f. Jin, et M. Ghil, 1995 : Multiple Equilibria, Periodic, and Aperiodic Solutions in a Wind-Driven, Double-Gyre, Shallow-Water Model. *Journal of Physical Oceanography*, **25**, 764–786.

Kiss, A. E., 2002 : Potential vorticity "crises", adverse pressure gradients, and western boundary current separation. *Journal of Marine Research*, **60**, 779–803.

Klein, P., B. L. Hua, G. Lapeyre, X. Capet, S. Le Gentil, et H. Sasaki, 2008 : Upper Ocean Turbulence from High-Resolution 3D Simulations. *Journal of Physical Oceanography*, **38**, 1748–1763.

Korres, G., N. Pinardi, et A. Lascaratos, 2000 : The Ocean Response to Low-Frequency Interannual Atmospheric Variability in the Mediterranean Sea. Part II : Empirical Orthogonal Functions Analysis. *Journal of Climate*, **13**, 732–745.

Krysta, M., E. Blayo, E. Cosme, J. Verron, et A. Vidard, 2008 : A 4D-Var-SEEK smoother hybrid. Towards applications in oceanography. *WWRP/THORPEX workshop on 4D-VAR and Ensemble Kalman filter inter-comparisons.*

Laloyaux, P., 2012 : *Recent developments in optimization methods for data assimilation in oceanography.* Ph.D. thesis, Université de Namur.

Lawless, A., N. Nichols, et S. Ballard, 2003 : A comparison of two methods for developping the linearization of a shallow-water model. *Quarterly Journal of the Royal Meteorological Society*, **129**, 1237–1254.

Le Dimet, F.-X. et O. Talagrand, 1986 : Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects. *Tellus A*, **38A**, 97–110.

Le Traon, P. Y., P. Klein, B. L. Hua, et G. Dibarboure, 2008 : Do Altimeter Wavenumber Spectra Agree with the Interior or Surface Quasigeostrophic Theory ? *Journal of Physical Oceanography*, **38**, 1137–1142.

Lesieur, M., 1984 : *Turbulence in fluids.* Springer.

Lévy, M., P. Klein, A.-M. Tréguier, D. Iovino, G. Madec, S. Masson, et K. Takahashi, 2010 : Modifications of gyre circulation by sub-mesoscale physics. *Ocean Modelling*, **34**, 1–15.

Lions, J. L., 1968 : *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles.* Paris.

Lorenc, A. C., S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, et F. W. Saunders, 2000 : The Met. Office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, **126**, 2991–3012.

Luong, B., J. Blum, et J. Verron, 1998 : A variational method for the resolution of a data assimilation problem in oceanography. *Inverse Problems*, **14**, 979–997.

Madec, G., 2008 : NEMO Ocean Engine. Technical report, Note du Pole de Modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27 ISSN No 1288-1619.

Madec, G., P. Delecluse, M. Imbard, et C. Lévy, 1998 : OPA 8.1 Ocean General Circulation Model reference manual. Technical report, Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No11, 91pp.

McWilliams, J., 1996 : Modeling the oceanic general circulation. *Annual Review of Fluid Mechanics*.

Miller, R., M. Ghil, et F. Gauthiez, 1994 : Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the atmospheric sciences*, **51**, 1037–1056.

Mirouze, I., 2010 : *Régularisation de problèmes inverses à l'aide de l'équation de diffusion, avec applictaion à l'assimilation variationnelle de données océaniques*. Ph.D. thesis, Université Toulouse 3 Paul Sabatier.

Mirouze, I. et A. Weaver, 2010 : Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1421–1443.

Moodey, A. J. F., A. S. Lawless, R. W. E. Potthast, et P. J. van Leeuwen, 2013 : Nonlinear error dynamics for cycled data assimilation methods. *Inverse Problems*, **29**, 025002.

Moore, A., H. Arango, E. Di Lorenzo, B. Cornuelle, A. Miller, et D. Neilson, 2004 : A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model. *Ocean Modelling*, **7**, 227–258.

Moore, A., J. Vialard, A. Weaver, D. Anderson, R. Kleeman, et J. Johnson, 2003 : The role of air-sea interaction in controlling the optimal perturbations of low-frequency tropical coupled ocean-atmosphere modes. *Journal of Climate*, **16**, 951–968.

Moore, A. M., 1991 : Data Assimilation in a Quasi-geostrophic Open-Ocean Model of the Gulf Stream Region Using the Adjoint Method. *Journal of Physical Oceanography*, **21**, 398–427.

Moore, A. M., H. G. Arango, G. Broquet, B. S. Powell, A. T. Weaver, et J. Zavala-Garay, 2011 : The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems. *Progress in Oceanography*, **91**, 34–49.

Nauw, J. J., H. A. Dijkstra, et E. P. Chassignet, 2004 : Frictionally Induced Asymmetries in Wind-Driven Flows. *Journal of Physical Oceanography*, **34**, 2057–2072.

Neveu, E., 2011 : *Application des méthodes multigrilles à l'assimilation variationnelle de donneées en geéophysique*. Ph.D. thesis, Université de Grenoble.

Ngodock, H. E., 1996 : *Assimilation de données et analyse de sensibilité : une application à la circulation océanique*. Ph.D. thesis, Université de Grenoble.

Pascual, A., Y. Faugère, G. Larnicol, et P.-Y. Le Traon, 2006 : Improved description of the ocean mesoscale variability by combining four satellite altimeters. *Geophysical Research Letters*, **33**, L02611.

Pedlosky, J., 1987 : *Geophysical Fluid Dynamics*. Springer.

Pham, D. T., J. Verron, et M. Christine Roubaud, 1998a : A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, **16**, 323–340.

Pham, D. T., J. Verron, et M. C. Roubaud, 1998b : A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, **16**, 323–340.

Pires, C., R. Vautard, et O. Talagrand, 1996 : On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus A*, **48A**, 96–121.

Powell, B., H. Arango, A. Moore, E. Di Lorenzo, R. Milliff, et D. Foley, 2008 : 4DVAR data assimilation in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). *Ocean Modelling*, **25**, 173–188.

Preisendorfer, R. W., 1988 : *Principal component analysis in meteorology and oceanography*.

Rémy, E., 1999 : *Assimilation variationnelle de données tomographiques simulées dans des modèles de circulation océanique*. Ph.D. thesis, Univ. Paris 6.

Rémy, E., F. Gaillard, et J. Verron, 2002 : Variational assimilation of ocean tomographic data : Twin experiments in a quasi-geostrophic model. *Quarterly Journal of the Royal Meteorological Society*, **128**, 1739–1458.

Ricci, S., A. T. Weaver, J. Vialard, et P. Rogel, 2005 : Incorporating State-Dependent Temperature–Salinity Constraints in the Background Error Covariance of Variational Ocean Data Assimilation. *Monthly Weather Review*, **133**, 317–338.

Robert, C., S. Durbiano, E. Blayo, et J. Verron, 2005a : A reduced-order strategy for 4D-Var data assimilation. *Journal of Marine Systems*, **57**, 70–82.

— 2005b : A reduced-order strategy for 4D-Var data assimilation. *Journal of Marine* . . . .

Sévellec, F., T. Huck, M. BenJelloul, N. Grima, J. Vialard, et A. Weaver, 2008 : Optimal surface salinity perturbations of the meridional overturning and heat transport in a global ocean general circulation model. *Journal of Physical Oceanography*, **38**, 2739–2754.

Sévellec, F., T. Huck, M. BenJelloul, et J. Vialard, 2009 : Non-normal multidecadal response of the thermohaline circulation induced by optimal surface salinity perturbations. *Journal of Physical Oceanography*, **39**, 852–872.

Skamarock, W., 2004 : Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly weather review*, **132**, 3019–3032.

Stammer, D., C. Wunsch, I. Fukumori, et J. Marshall, 2002 : State estimation improves prospects for ocean research. *Eos, Transactions American Geophysical Union*, **83**, 289.

Tai, C.-K., 2004 : The Resolving Power of à Single Exact-Repeat Altimetric Satellite Or à Coordinated Constellation of Satellites. *Journal of Atmospheric ans Oceanic Technology*, **21**, 810–818.

Tber, M. H., L. Hascoet, A. Vidard, et B. Dauvergne, 2007 : Building the Tangent and Adjoint codes of the Ocean General Circulation Model OPA with the Automatic Differentiation tool TAPENADE. Research Report RR-6372, INRIA.

Titaud, O., A. Vidard, I. Souopgui, et F. a.-X. Le Dimet, 2010 : Assimilation of image sequences in numerical models. *Tellus A*, **62**, 30–47.

Trémolet, Y., 2007a : Incremental 4D-Var convergence study. *Tellus A*, **59**, 706–718.

— 2007b : Model-error estimation in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **133**, 1267–1280, doi :10.1002/qj.94.

Tshimanga, J., S. Gratton, A. Weaver, et A. Sartenaer, 2008 : Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **134**, 751–769.

Verron, J., 1992 : Nudging satellite data into quasi-geostrophic ocean models. *Journal of Geophysical Research*, **97**, 7479–7491.

Verron, J. et E. Blayo, 1996 : The No-Slip Condition and Separation of Western Boundary Currents. *Journal of Physical Oceanography*, **26**, 1938–1951.

Verron, J. et W. Holland, 1989 : Impacts de données d'altimétrie satellitaire sur les simulations numériques des circulations océaniques aux latitudes moyennes. *Annales Geophysicae*, **71**, 31–46.

Verron, J. et C. Le Provost, 1991 : Response of eddy-resolved general circulation numerical models to asymmetrical wind forcing. *Dynamics of Atmospheres and Oceans*, **15**, 505–533.

Vidard, A., 2001 : *Vers une prise en compte de l'erreur modèle en assimilation de données 4D-variationnelle.*. Ph.D. thesis, Université de Grenoble.

Vossepoel, F. C., A. Weaver, J. Vialard, et P. Delecluse, 2003 : Adjustment of near-equatorial wind stress with 4d-var data assimilation in a model of the pacific ocean. *Monthly Weather Review*, **132**, 2070–2083.

Weaver, A. et P. Courtier, 2001 : Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, **127**, 1815–1846.

Weaver, a. T., C. Deltel, E. Machu, S. Ricci, et N. Daget, 2005 : A multivariate balance operator for variational ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **131**, 3605–3625.

Weaver, A. T., J. Vialard, et D. L. T. Anderson, 2003 : Three- and Four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I : Formulation, internal diagnostics, and consistency. *Monthly Weather Review*, **131**, 1360–1378.