



HAL
open science

Vers une prise en compte des erreurs-modèle en assimilation de données 4D-variationnelle. Application à un modèle réaliste d'océan

Arthur Vidard

► To cite this version:

Arthur Vidard. Vers une prise en compte des erreurs-modèle en assimilation de données 4D-variationnelle. Application à un modèle réaliste d'océan. Mathématiques [math]. Université Joseph-Fourier - Grenoble I, 2001. Français. NNT: . tel-00325349

HAL Id: tel-00325349

<https://theses.hal.science/tel-00325349v1>

Submitted on 29 Sep 2008

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.

Vers une prise en compte des erreurs modèle
en assimilation de données 4D variationnelle.
Application à un modèle réaliste d'océan.

Patrick, Arthur VIDARD

5 octobre 2001

Table des matières

Introduction	7
I Contexte général	11
1 Les méthodes classiques d’assimilation de données	13
1.1 Les différents intervenants	14
1.1.1 Définitions	14
1.1.2 Hypothèses et simplifications généralement faites	19
1.2 Les méthodes stochastiques	20
1.2.1 Le filtre de Kalman et ses amis	20
1.2.2 L’interpolation Optimale	22
1.2.3 Le Nudging, la SMC (Successive Corrections Method) et l’ACS (Analysis Correction Scheme)	23
1.3 Les méthodes variationnelles	24
1.3.1 le 3D-Var	24
1.3.2 Le 4D-Var	26
1.3.3 Les méthodes duales	31
2 Les différents modèles utilisés	33
2.1 Les modèles “jouets”	34
2.1.1 L’équation de Burger	34
2.1.2 Le modèle Shallow Water	35
2.2 Le modèle aux équations primitives OPA	38
3 Techniques algorithmiques	47
3.1 Les opérateurs adjoints	47
3.2 Méthodes de minimisation	49
3.3 Conditionnement et adimensionalisation	51
3.4 Les matrices de covariance d’erreur en assimilation variationnelle	54
3.4.1 Construction de l’opérateur B	54
3.4.2 Les effets de l’opérateur d’équilibre sur l’assimilation variationnelle	59
3.5 Diagnostics	61
3.5.1 Les diagnostics d’optimisation :	61
3.5.2 Les diagnostics statistiques :	62

3.5.3	Les diagnostics ImA : Information Minus Analysis.	63
3.5.4	Les données indépendantes :	63
4	Le coupleur PALM	65
4.1	PALM quesaco ?	65
4.1.1	Principe et définitions	65
4.1.2	Les commandes	67
4.1.3	Trucs et astuces	68
4.2	PALM par l'exemple	69
II La Prise en compte des erreurs modèles en assimilation variationnelle - Études préliminaires		81
5	Contrôle de l'erreur modèle : méthode complète	83
5.1	La méthode complète	84
5.1.1	Premiers résultats numériques	85
5.2	Réduction de la taille de l'espace de contrôle	90
5.2.1	Principe	90
5.2.2	Application à notre problème	91
5.2.3	Résultats numériques : Partie II	92
5.3	Conclusion	93
6	Le Nudging Optimal	95
6.1	Motivations	96
6.2	La Méthode	97
6.3	Équivalence avec le filtre de Kalman	100
6.4	Expériences préliminaires	102
6.5	Expérimentation avec le modèle Shallow-Water	104
6.5.1	Le Modèle	104
6.5.2	L'assimilation	105
6.5.3	Expérience numérique	106
6.6	Conclusion	108
7	4D-Var avec contrôle de l'erreur systématique du modèle	111
7.1	Un exemple scalaire	111
7.2	La méthode	113
7.3	Expérimentation avec le modèle Shallow-Water	114
7.3.1	Préconditionnement et minimisation	114
7.3.2	Résultats	115
7.3.3	Des observations moins fournies	118
7.4	Conclusion	124

III Application au modèle de circulation générale océanique	
OPA	129
8 Comment OPA réagit-il?	131
8.1 Influence des différentes sources d'erreur	131
8.1.1 L'erreur sur la condition initiale (expérience CT1)	131
8.1.2 L'erreur dans la physique du modèle (expérience CT2)	134
8.1.3 L'erreur sur les forçages (expérience CT3)	136
8.1.4 Conclusion	136
8.2 Les résultats du 4D-Var incrémental sur OPA	138
9 Contrôle de l'erreur systématique : Expériences jumelles	143
9.1 Configuration favorable	145
9.1.1 Description de l'expérience	145
9.1.2 Résultats	145
9.1.3 Premières conclusions	147
9.2 Comme en "vrai"	148
9.2.1 Description de l'expérience	148
9.2.2 Résultats	148
9.3 Conclusion des expériences jumelles	152
10 Contrôle de l'erreur systématique : Expériences réalistes	155
10.1 Configuration	155
10.2 Résultats	156
10.2.1 Une qualité conservée	156
10.2.2 Comparaison aux données indépendantes	158
10.2.3 Un biais réduit	160
10.2.4 Une assimilation "en douceur"	164
10.2.5 La salinité	166
10.3 Pas toujours plus efficace	168
10.4 Conclusion des expériences réalistes	170
10.5 Assimiler des données altimétriques	170
Conclusions et perspectives	173
A Résultats sur le 3D-Var et le 4D-Var incrémental.	179

Introduction

On peut sans hésiter affirmer que sans l'océan la vie ne serait pas possible sur notre planète. L'océan recouvre près de 70% de la surface terrestre, il transporte et échange d'importantes quantités de matière (CO₂, vapeur d'eau) et d'énergie (chaleur et quantité de mouvement) avec l'atmosphère, et de ce fait son influence sur le climat est considérable. Connaître et comprendre les océans est donc un enjeu majeur de la recherche environnementale. Les objectifs principaux sont les prévisions à diverses échéances (semaine, mois, saison) de l'évolution des multiples composantes de l'océan (température, salinité, courants marins, biologie et chimie de l'océan, vagues et marées, ...). Les enjeux de ces prévisions sont multiples

- enjeux de sécurité publique (des populations) : outre l'impact direct sur la sécurité des personnes (tempêtes), l'océan par son action sur le climat peut être source de graves perturbations. El niño est sans doute le phénomène le plus célèbre : les déplacements de masses d'eau chaude dans l'océan pacifique tropical qu'il représente engendrent des dérèglements climatiques catastrophiques pour les pays d'Amérique latine avoisinants.
- enjeux environnementaux : à l'heure où de nombreuses interrogations se posent quant à l'augmentation du taux de CO₂ dans l'atmosphère, savoir comment le formidable réservoir à gaz carbonique qu'est l'océan va se comporter est d'une importance capitale : va-t-il en absorber d'avantage ? Va-t-il nous le rendre dans quelques décennies alors que la situation sera peut-être plus délicate qu'actuellement ?
- enjeux économiques : une bonne connaissance de l'état de l'océan et de son évolution permet de bien choisir les zones de pêche, elle est aussi d'une grande portée pour les forages pétroliers off-shore ainsi que pour les activités maritimes civiles et militaires. Plus indirectement, son action sur le climat revêt une importance particulière pour l'agriculture.

Convaincue de l'importance de ces enjeux la communauté internationale a fourni ces dernières années un important effort pour améliorer la connaissance de l'océan, cela s'est traduit par la création de nombreux programmes internationaux tels que WOCE¹, TOGA², etc. Ces programmes visent à la préservation et à la gestion des ressources marines, des besoins en matière de protection et de gestion des environnements côtiers. Ils visent à étudier l'impact de l'océan sur le climat global de notre planète, à l'approfondissement

¹World Ocean Circulation Experiment

²Tropical Ocean Global Atmosphere

des connaissances scientifiques sur les phénomènes associés au changement climatique et surtout à l'impact de l'activité humaine. Cela a abouti à la multiplication des moyens d'observation de l'océan. En plus des données *in situ* (bouées dérivantes ou ancrées, navires marchands volontaires, missions océanographiques, tomographies, etc.), de nombreuses missions d'observations satellites ont été lancées parmi lesquelles on peut citer ERS 1 (1991-2000), TOPEX/POSEIDON (1992-fin 2002 ?), ERS 2 (1995-...) et bientôt relayées par Jason 1 (lancement prévu février 2002) et ENVISAT (lancement prévu novembre 2001). Ces coûteux programmes démontrent bien la prise de conscience de l'importance capitale du rôle de l'océan dans notre environnement.

Parallèlement à cela, la puissance sans cesse grandissante des moyens de calcul a permis l'utilisation de modèles de plus en plus sophistiqués et performants, à des résolutions de plus en plus fine. Malgré ces améliorations les modèles sont certes réalistes mais on connaît mal les conditions initiales et les conditions aux bords. De plus des simplifications persistent dans la physique des modèles et certains processus ne peuvent être pris en compte du fait de la résolution adoptée (effets sous-maille).

De leur cotés les observations sont de grande précision mais inégalement réparties en temps et/ou en espace : elles peuvent être constituées d'informations fournies sur la surface de l'océan mais d'aucune information sur les profondeurs comme c'est cas pour des observations satellitales ou *a contrario* d'une bonne information sur la verticale mais peu nombreuses par rapport à l'immensité de l'océan (données issues de bouées).

Pour utiliser efficacement ces deux sources d'information (modèle et observations) de façon complémentaire on a introduit les méthodes dites "d'assimilation de données". Schématiquement le problème consiste à déterminer, sur une séquence de temps donnée, la trajectoire discrète du modèle qui approche au mieux les observations sur la période considérée.

Le projet MERCATOR est la contribution française au projet international GODAE³, il est cofinancé par le CNES, le CNRS/INSU, l'IFREMER, l'IRD, Meteo-France et le SHOM. Sa mission est de fournir un système de prévision océanographique opérationnel. Les premiers bulletins de prévision ont été produits au début de l'année 2001

La partie "Assimilation" de la chaîne opérationnelle MERCATOR est gérée par le coupleur d'assimilation PALM. Ce logiciel, développé par l'équipe de modélisation du climat et de son changement global du CERFACS, a pour but de gérer de façon la plus modulaire possible cette chaîne d'assimilation, de sorte que la manipulation et l'évolution de cette chaîne soit la plus aisée possible.

L'utilisation de PALM fut le point de départ de cette thèse. Avant d'être utilisé de façon opérationnelle et alors que le prototype était encore en cours de développement il convenait de s'assurer de son ergonomie et de la pertinence de ses choix algorithmiques dans un contexte de recherche. Dans le cadre d'une collaboration entre le CERFACS et le projet IDOPT commun à l'INRIA Rhone-Alpes et au LMC, le commencement de cette thèse fut donc un test "grandeur nature" de PALM. Après un début fastidieux (il a bien fallu essayer les plâtres), PALM s'est rapidement révélé comme étant un outil précieux dans la conception et l'utilisation de chaînes d'assimilation de données.

³Global Ocean Data Assimilation Experiment

Une des chaînes d'assimilation les plus répandues en mode opérationnel est le 4D-Var⁴; sa principale limitation est le fait que le modèle est considéré comme étant une exacte représentation de la réalité (ce qui en pratique n'est jamais le cas). Après avoir dompté PALM et le modèle réaliste d'océan OPA, le principal objectif de cette thèse fut donc d'introduire la prise en compte de l'erreur modèle dans l'assimilation 4D-Var en gardant toujours à l'esprit les contraintes opérationnelles en terme de temps de calculs et de facilité d'implémentation.

Plan du mémoire de thèse

Dans cette thèse on a tenté de conserver, quand cela était possible, une approche intuitive dans la présentation des méthodes et des résultats afin de favoriser la compréhension des mécanismes internes à chaque méthode. Le contexte dans lequel s'est déroulée cette thèse a favorisé l'aspect "application" jusqu'à obtenir une méthode pré-opérationnelle.

La première partie de ce mémoire sera consacrée à la présentation non exhaustive des différentes méthodes d'assimilation de données les plus couramment utilisées, en insistant principalement sur le 4D-Var et ses variantes, puis à la présentation des différents modèles utilisés au cours de cette thèse pour valider les méthodes d'assimilation. Cette partie se terminera par une discussion sur certains aspects techniques de l'assimilation de données et par la présentation du coupleur d'assimilation PALM.

La deuxième partie sera consacrée aux études préliminaires sur le contrôle de l'erreur modèle, appliquées à des modèles simplifiés. En plus de la méthode complète, trop coûteuse en temps de calcul, deux variantes seront testées : Le nudging Optimal et le contrôle de la part systématique de l'erreur modèle, et elles seront comparées au 4D-Var appliqué à la même configuration.

La dernière partie verra l'application au modèle réaliste d'océan OPA de l'une de ces variantes : le contrôle de la part systématique de l'erreur modèle. Les résultats seront comparés à la méthode classique afin de mettre en évidence les améliorations qu'elle apporte.

⁴opérationnelle, entre autres, à Météo France et au Centre Européen de Prévisions Météorologiques à Moyen Terme (Reading, GB)

Première partie
Contexte général

Chapitre 1

Les méthodes classiques d'assimilation de données

Le terme “assimilation de données” désigne une classe de méthodes mathématiques très usitées notamment en météorologie et en océanographie. Ces méthodes permettent de recomposer “le mieux possible” l'état d'un système en utilisant les informations fournies par la prédiction d'un modèle et celles apportées sous forme directe ou statistique par des observations de ce système. L'expression “le mieux possible” n'a pas forcément toujours le même sens suivant les méthodes d'assimilation employées, les problèmes considérés, etc ...

En océanographie et en météorologie ces méthodes ont deux buts principaux :

- Le “lissage” : l'assimilation permet de fournir, pour une période de temps donnée, une estimation de l'évolution du système (trajectoire) qui s'ajuste optimalement aux observations et au modèle utilisé en un sens à définir. Elle permet une analyse historique du comportement du système, de repérer des comportements spécifiques, etc... C'est aussi un outil pertinent pour le test d'hypothèses et pour des études de sensibilité.
- Le “filtrage” : l'assimilation permet d'obtenir une estimation du meilleur état présent du système à partir de toutes les informations passées et présentes. Seul l'état présent est optimal (toujours dans un sens à définir). Cette deuxième approche est le type même des méthodes opérationnelles de prévision en météorologie qui consistent à corriger, à partir d'un champ d'observation sur une période donnée, la trajectoire du modèle durant son intégration afin d'obtenir de bonnes conditions initiales (ou aux limites) pour la prévision de l'état du système pour la période suivante. Bien que la prévision ne soit pas de l'assimilation de données au sens strict du terme, elle est le prolongement direct du filtrage le plus récent.

On distingue généralement deux classes de méthodes d'assimilation de données : les méthodes variationnelles d'une part, et les méthodes stochastiques d'autre part. Les premières sont fondées sur la théorie du contrôle optimal et sont actuellement opérationnelles en météorologie (4D-var). La variable de contrôle est (une partie de) la condition initiale de la fenêtre temporelle sur laquelle on va assimiler les données, et le modèle est considéré comme étant exact. L'optimisation est réalisée par une méthode de gradient, la valeur de

celui-ci étant obtenue grâce aux méthodes adjointes (LE DIMET, TALAGRAND 1986 [72]). Plusieurs centaines d'itérations (donc autant d'intégrations du modèle) sont généralement nécessaires pour que la méthode converge. Tout cela engendre un coût en temps de calcul élevé. Il faut de plus tenir compte de l'écriture de l'adjoint du modèle direct (ce qui n'est pas une mince affaire).

L'approche stochastique est basée, quant à elle, sur la théorie de l'estimation statistique optimale, la méthode la plus célèbre étant le filtre de Kalman. Ce filtre requiert, dans sa version complète, la manipulation de matrices (multiplication, résolution de systèmes linéaires...) de taille $N \times N$ où N est la taille du vecteur d'état (typiquement $10^6 - 10^7$ pour des applications réalistes), et ce pour chaque pas de temps. Ceci suppose des capacités de stockage et de calcul importantes, et même hors de portée pour des applications réelles en météorologie ou en océanographie. On est donc amené le plus souvent à faire des hypothèses simplificatrices. De par leur construction ces méthodes s'appliquent plus particulièrement aux problèmes de filtrage (d'où leur nom) bien qu'il en existe quelques variantes adaptées au lissage. Dans le cas linéaire ce filtre est optimal (ce qui n'est pas forcément le cas des variantes adaptées aux cas non-linéaires), c'est pourquoi il sert de référence pour l'ensemble des méthodes d'assimilation de données dans ce cas là.

Le chapitre qui suit présente un panorama non exhaustif des méthodes d'assimilation de données (il en pousse de nouvelles régulièrement); une attention particulière sera portée sur le filtre de Kalman et sur le 4D-Var.

1.1 Les différents intervenants de l'assimilation de données

Avant d'introduire les différentes méthodes ou classes de méthodes d'assimilation, il convient de décrire les différents opérateurs mathématiques intervenant dans les processus d'assimilation. On remarquera, tout au long de ce chapitre, que quelle que soit la méthode utilisée elle peut être décrite à l'aide d'un nombre réduit de ces opérateurs et que ceux ci sont communs aux différents algorithmes. Le chapitre suivant présentera le coupleur PALM qui tire partie de cette constatation.

1.1.1 Définitions

Remarques préliminaires

Il est tout d'abord nécessaire d'introduire les notations qui seront utilisées dans la suite. Le formalisme retenu dans cet ouvrage sera principalement celui introduit par IDE *et al.* en 1997 [63]. Sans le reformuler entièrement, on peut juste préciser que la casse et la typographie auront leur importance. En effet si les minuscules italiques a, b, c, \dots représentent des valeurs scalaires, les majuscules M, H, \dots seront utilisées pour désigner les opérateurs non linéaires. Les caractères écrits en gras représenteront les vecteurs $\mathbf{x}, \mathbf{y}, \dots$ (pour les minuscules) et les matrices ou les opérateurs linéaires $\mathbf{M}, \mathbf{H}, \dots$ (en majuscules).

Le modèle fournit une prévision de l'état du système : tout ce qui se rapporte directement à cette prévision (état prédit, erreur sur cet état, matrice de covariance de cette erreur, etc...) est noté avec l'exposant f (pour *forecast*). De la même façon les exposants t , b , a et o représenteront respectivement l'état vrai (l'équivalent de la réalité dans l'espace des variables du modèle), l'ébauche (ou *first guess* ou *background* : c'est la condition initiale du système sur la fenêtre temporelle d'assimilation avant l'assimilation), l'état analysé (produit par l'assimilation) et les observations. On rajoutera l'exposant c qui ne sera utilisé que dans la première partie de ce chapitre et représentera l'état *continu* du système par opposition aux états discrétisés sur une grille représentant l'espace.

Dans cette première partie, on prendra soin de distinguer les processus *déterministes* qui peuvent être entièrement décrits par des équations aux dérivées partielles et les processus *stochastiques* dont on ne connaît que les statistiques et qui pourront être estimés au moyen de celles-ci.

Le vecteur d'état \mathbf{x}

Le vecteur d'état représente l'ensemble des variables pronostiques du modèle, c'est à dire les variables dont la dérivée par rapport au temps est prise en compte explicitement dans les équations du modèle. Les autres variables du système (appelées variables diagnostiques) peuvent être entièrement déterminées à partir du vecteur d'état.

Les composantes du système océanique peuvent être décrites de façon générale au moyen d'un système d'équations aux dérivées partielles (edp) non linéaires (cf chapitre sur les modèles). Notons $\mathbf{x}_{t_i}^c$ le vecteur pronostique de ce système d'edp (c pour continu) au temps t_i . En termes purement mathématiques, les éléments de ce vecteur sont des fonctions et on peut poser $\mathbf{x}_{t_i}^c \in \mathcal{U}$ où \mathcal{U} est un espace de fonctions. En supposant que le problème est bien posé au sens de Hadamard (HADAMARD [58]) il existe un unique opérateur $M_{(t_i, t_{i+1})}^c$ qui permet d'obtenir la solution $\mathbf{x}_{t_{i+1}}^c$ à partir de la solution $\mathbf{x}_{t_i}^c$

$$\mathbf{x}_{t_{i+1}}^c = M_{(t_i, t_{i+1})}^c(\mathbf{x}_{t_i}^c) \quad (1.1)$$

Par souci de clarté, on supposera, dans un premier temps, que tous les paramètres, les conditions aux limites et le forçage de M^c sont fixés et qu'aucune approximation n'a été faite dans le système d'edp. Ceci revient à dire que l'opérateur M^c est déterministe et que (1.1) est une représentation exacte et complète de la réalité. De par la complexité des systèmes océaniques et le nombre d'interactions avec l'extérieur pouvant les influencer, il est évident que ces hypothèses sont irréalistes, mais elles permettent une grande simplification dans l'analyse du problème. En pratique les sources d'erreur sont multiples, elles peuvent provenir de la discrétisation, de la physique manquante, des forçages, etc. Dans toute cette thèse on ne les différenciera pas, mais pour un meilleur traitement, il faudra sans doute adapter notre démarche au type d'erreur que l'on cherche à corriger.

Le modèle de prévision M

En pratique nous ne connaissons pas $\mathbf{x}_{t_i}^c$ (et en particulier $\mathbf{x}_{t_0}^c$) ni M^c , mais seulement une approximation discrète de ceux-ci :

$$\mathbf{x}_{t_{i+1}} = M_{(t_i, t_{i+1})}(\mathbf{x}_{t_i}) \quad (1.2)$$

où \mathbf{x}_{t_i} est un vecteur de taille n (on notera \mathcal{U}^n l'espace auquel il appartient) approximant l'état continu $\mathbf{x}_{t_i}^c$ dont l'évolution est donnée par (1.1) et M est le modèle discret. *Ce dernier décrit sous forme d'équations mathématiques discrétisées les connaissances qu'on a a priori sur les processus physiques régissant le système étudié (voir le chapitre 2 les décrivant de façon plus détaillée).* Le vecteur d'état discret quant à lui peut être composé d'un simple vecteur de \mathbb{R}^n comme c'est le cas pour l'équation de Burger, il peut aussi être l'assemblage de plusieurs champs 3D dans des modèles plus conséquents comme les équations primitives (avec, dans ce cas précis les vitesses de courant, la température et la salinité)

Nous pouvons maintenant introduire l'état vrai \mathbf{x}^t qui représentera l'état à estimer au moyen d'un processus d'assimilation. Bien que l'état continu \mathbf{x}^c soit le candidat le plus naturel (c'est la "réalité vraie") on lui préférera sa représentation dans l'espace de l'état discret. Ce choix permettra de comparer directement \mathbf{x}^t avec l'état approximé donné par (1.2). Pour cela définissons un opérateur linéaire $\mathbf{\Pi}$ de projection de \mathcal{U} dans \mathcal{U}^N d'une manière appropriée à la discrétisation (1.2) (par exemple si le domaine a été discrétisé en volumes finis et si les N composantes du vecteur \mathbf{x}_{t_i} représentent des moyennes sur ces volumes, alors \mathcal{U}^N sera un assemblage de fonctions constantes et les N composantes de $\mathbf{\Pi}\mathbf{x}_{t_i}^c$ seront les moyennes de $\mathbf{x}_{t_i}^c$ sur chaque volume). En résumé, on définira l'état vrai comme représentation dans \mathcal{U}^n de l'état continu $\mathbf{x}_{t_i}^c$ par :

$$\mathbf{x}_{t_i}^t = \mathbf{\Pi}\mathbf{x}_{t_i}^c \quad (1.3)$$

L'état vrai $\mathbf{x}_{t_i}^t$ est inconnu, mais en appliquant l'opérateur $\mathbf{\Pi}$ à l'équation (1.1) et en posant

$$\delta_i = \mathbf{\Pi}M_{(t_i, t_{i+1})}^c(\mathbf{x}_{t_i}^c) - M_{(t_i, t_{i+1})}(\mathbf{\Pi}\mathbf{x}_{t_i}^c) \quad (1.4)$$

on peut écrire

$$\mathbf{x}_{t_{i+1}}^t = M_{(t_i, t_{i+1})}(\mathbf{x}_{t_i}^t) + \delta_i \quad (1.5)$$

Par sa définition (1.4) δ_i représente l'erreur de modèle, on remarque que cette erreur dépend généralement de l'état continu $\mathbf{x}_{t_i}^c$. On peut également remarquer que du fait de l'impossibilité de connaître exactement l'état continu, cette dépendance rend l'erreur modèle impossible à estimer d'un point de vue déterministe. Cependant, en supposant que $M_{(t_i, t_{i+1})}$ est une bonne approximation de $M_{(t_i, t_{i+1})}^c$ on peut dire que δ_i est petit. C'est pourquoi on peut représenter (et c'est bien pratique) ces erreurs modèles comme une perturbation stochastique à l'équation (1.5).

La matrice de covariance d'erreur de modèle \mathbf{Q}

Si on suppose que δ_i est un processus stochastique dépendant de $\mathbf{x}_{t_i}^c$ on peut reformuler

$$\delta_i = G_i(\mathbf{x}_{t_i}^c)\eta_i \quad (1.6)$$

Dans cette représentation G_i est un opérateur dépendant de l'état continu $\mathbf{x}_{t_i}^c$, reflétant la dépendance de δ_i par rapport à l'état et η_i est un vecteur de taille m représentant un

processus stochastique de moyenne

$$\bar{\eta}_i = E[\eta_i] \quad (1.7)$$

où $E[.]$ représente l'espérance mathématique, et de matrice de covariance

$$\mathbf{Q}_i = E[(\eta_i - \bar{\eta}_i)(\eta_i - \bar{\eta}_i)^T] \quad (1.8)$$

(1.5) peut donc être reformulé

$$\mathbf{x}_{t_{i+1}}^t = M_{(t_i, t_{i+1})}(\mathbf{x}_{t_i}^t) + G_i(\mathbf{x}_{t_i}^c)\eta_i \quad (1.9)$$

L'équation (1.9) qui décrit l'évolution de l'état vrai, laisse apparaître que si $\eta_{t_{i+1}}$ est un processus stochastique, il en sera de même pour $\mathbf{x}_{t_{i+1}}^t$

Les observations \mathbf{y}^o

Le vecteur d'observation rassemble les données qui sont susceptibles d'être utilisées (assimilées) par les méthodes présentées plus loin. Les sources éventuelles d'informations sont multiples, elles peuvent être in situ comme les bouées TAO ou les profileurs XBT, ou provenant des satellites d'observation, etc...

L'opérateur d'observation discret H

Les quantités observées n'étant pas nécessairement localisées aux points de grille ou aux pas de temps du modèle, ni même forcément de même nature ou métrique que les variables du modèle, il faut définir un opérateur dit d'observation. Celui ci "traduit" les variables \mathbf{x} de l'espace du modèle dans l'espace des observations.

Afin de relier les observations à l'état \mathbf{x}^t que nous voulons estimer, supposons que l'on dispose, à chaque temps $t_{i,k}$, $k=1, 2, 3, \dots$, d'un nombre p_k d'observations (i.e. $\mathbf{y}^o \in \mathbb{R}^{p_k}$).

Ce sont des observations de l'état *continu* \mathbf{x}^c , on peut donc écrire

$$\mathbf{y}_{t_i}^o = H_{t_i}^c(\mathbf{x}_{t_i}^c) + \varepsilon_{t_i}^m \quad (1.10)$$

où ε^m est l'erreur de mesure dépendant de l'instrument de mesure utilisé et $H^c \in \mathcal{C}(\mathcal{U}, \mathbb{R}^{p_k})$ est l'opérateur d'observation continu. Les erreurs de mesure sont indépendantes du modèle utilisé, elles proviennent principalement de l'imprécision ou du dérèglement des appareils de mesure¹. De plus les données peuvent nécessiter un pré-traitement qui lui-même pourra induire certaines erreurs.

On considérera que ε^m est stochastique, indépendant de l'état \mathbf{x}^c et de moyenne

$$\bar{\varepsilon}_i = \langle \varepsilon_i \rangle$$

que l'on supposera connue ainsi que sa matrice de covariance

$$\mathbf{R}_i^m = E[(\varepsilon_i^m - \bar{\varepsilon}_i^m)(\varepsilon_i^m - \bar{\varepsilon}_i^m)^T]$$

¹Ce point peut être discuté car la façon de quantifier l'erreur de mesure peut être dépendante du contexte, et donc du modèle

L'équation (1.5) est formulée au moyen de l'état vrai *discret* $\mathbf{x}_{t_i}^t = \mathbf{\Pi}\mathbf{x}_{t_i}^c$; cela nécessite donc l'introduction de H_{t_i} l'opérateur d'observation discret qui agit sur $\mathbf{x}_{t_i}^t$ au lieu de $\mathbf{x}_{t_i}^c$. On peut réécrire (1.10)

$$\mathbf{y}_{t_i}^o = H_{t_i}(\mathbf{x}_{t_i}^t) + \varepsilon_{t_i} \quad (1.11)$$

où

$$\varepsilon_{t_i} = \varepsilon_{t_i}^m + \varepsilon_{t_i}^r$$

est l'erreur d'observation, et

$$\varepsilon_{t_i}^r = H_{t_i}^c(\mathbf{x}_{t_i}^c) - H_{t_i}(\mathbf{\Pi}\mathbf{x}_{t_i}^c)$$

est généralement appelée erreur de représentativité (cf LORENC 1986 [77]), c'est une conséquence des difficultés qu'on a à représenter la réalité dans l'espace des variables du modèle (la différence entre la réalité et \mathbf{x}^t). Cette définition englobe tous les processus appartenant à la physique manquante du modèle comme par exemple, les processus non représentables numériquement sur la grille modèle ou certains phénomènes physiques non résolus par les équations du modèle. On voit bien, par définition, qu'elle dépend de l'état *continu* du système. Pour des champs à haute variabilité, cette erreur de représentativité peut être supérieure à l'erreur de mesure (COHN 1997 [26]).

La matrice de covariance d'erreur d'observation \mathbf{R}

On définit la matrice de covariance d'erreur d'observation au temps t_i \mathbf{R}_i comme :

$$\mathbf{R}_i = E[\varepsilon_{t_i}\varepsilon_{t_i}^T] \quad (1.12)$$

Le modèle linéaire tangent \mathbf{M} et son adjoint \mathbf{M}^T

Pour la plupart des méthodes d'assimilation il sera nécessaire de disposer de *modèle linéaire tangent au modèle direct*

$$\mathbf{M}_{(t_i, t_{i+1})} = \left. \frac{\partial M_{(t_i, t_{i+1})}}{\partial \mathbf{x}} \right|_{\mathbf{x}_{t_i}}$$

ainsi que de l'adjoint \mathbf{M}^T de celui ci. Dans ce cas on supposera l'hypothèse linéaire tangente (TLH) vérifiée c'est à dire que les variations de la prévision du modèle au voisinage de l'état analysé dépendront linéairement de l'état initial, en d'autres termes pour \mathbf{x}_i assez proche de \mathbf{x}_i^a alors

$$M_{(t_i, t_{i+1})}(\mathbf{x}_i) - M_{(t_i, t_{i+1})}(\mathbf{x}_i^a) = \mathbf{M}_{(t_i, t_{i+1})}[\mathbf{x}_i - \mathbf{x}_i^a] \quad (1.13)$$

L'opérateur d'observation linéaire tangent et son adjoint

Il sera de même nécessaire de disposer de *l'approximation linéaire tangente de l'opérateur d'observation et de son adjoint.*

De la même façon que pour le modèle linéaire tangent, on pose l'hypothèse linéaire tangente pour \mathbf{x}_i assez proche de \mathbf{x}_i^a alors

$$H_{t_i}(\mathbf{x}_i) - H_{t_i}(\mathbf{x}_i^a) = \mathbf{H}_{t_i}[\mathbf{x}_i - \mathbf{x}_i^a] \quad (1.14)$$

La matrice de covariance d'erreur de l'ébauche \mathbf{B}

On a dit plus haut que l'état pouvait être considéré comme une variable aléatoire, on peut supposer qu'il en est de même pour la condition initiale $\mathbf{x}_{t_0}^t$. On introduit \mathbf{x}^b l'ébauche de notre problème qui est la condition initiale que l'on a *a priori* de notre période d'assimilation. De la même façon qu'on a défini l'erreur modèle, on écrit l'erreur de l'ébauche $\varepsilon^b = \mathbf{x}^b - \mathbf{x}_{t_0}^t$

$$\mathbf{B} = E[(\mathbf{x}^b - \mathbf{x}^t)(\mathbf{x}^b - \mathbf{x}^t)^T]$$

La matrice de covariance d'erreur de prévision \mathbf{P}_i^f et la matrice de covariance d'erreur d'analyse \mathbf{P}_i^a

On définit enfin l'erreur de prévision $\mathbf{x}_i^f - \mathbf{x}_i^t$ et sa matrice de covariance

$$\mathbf{P}_i^f = E[(\mathbf{x}_i^f - \mathbf{x}_i^t)(\mathbf{x}_i^f - \mathbf{x}_i^t)^T].$$

De la même façon on définit l'erreur d'analyse $\mathbf{x}_i^a - \mathbf{x}_i^t$ et sa matrice de covariance

$$\mathbf{P}_i^a = E[(\mathbf{x}_i^a - \mathbf{x}_i^t)(\mathbf{x}_i^a - \mathbf{x}_i^t)^T].$$

1.1.2 Hypothèses et simplifications généralement faites

- Comme c'est habituellement le cas en assimilation de données, les statistiques d'erreur sont supposées Gaussiennes, centrées, et identiquement distribuées dans tout cet ouvrage, cependant une formulation plus générale et l'utilisation de statistiques non Gaussiennes peuvent être trouvées dans la thèse de LAGARDE [70].
- Dans la suite de ce chapitre on supposera connus G_i , \mathbf{Q}_i et $\bar{\eta}_i$. On posera aussi $G_i = \mathbf{G}_i = Id$, c'est à dire que l'on considérera que l'erreur modèle ne dépend pas de l'état du système. C'est une hypothèse forte mais qui clarifie énormément les notations et simplifie les calculs.
- On supposera par la suite que les erreurs d'observation sont non corrélées entre elles spatialement (les \mathbf{R}_i sont des matrices diagonales) et non biaisées ($E[\varepsilon] = 0$). Ces hypothèses sont très discutables lorsque les observations utilisées sont fournies par un satellite; en effet l'instrument de mesure étant toujours le même les erreurs de mesures sont vraisemblablement très corrélées spatialement. De plus la méconnaissance relative de la surface de référence (le géoïde) induit clairement un biais dans ces observations. Dans le cas de l'utilisation de données *in situ* ces hypothèses peuvent être considérées comme valides.
- Une hypothèse supplémentaire est nécessaire : la non corrélation des erreurs de modèle et de celles d'observation

$$E[\eta_i, \varepsilon_i^T] = 0.$$

Les sources de ces deux erreurs étant très différentes on peut supposer que cette hypothèse sera toujours respectée.

- Les erreurs modèle sont de plus supposées Gaussiennes. Comme pour les erreurs d'observation, celles de prévision sont souvent considérées comme de moyenne nulle

$$E[\eta_i] = 0$$

C'est même une hypothèse nécessaire à la validation théorique des méthodes qui seront présentées dans ce chapitre. Cependant la présence de biais dans les modèles opérationnels a souvent été mise en évidence. La prise en compte de ce biais modèle est l'un des objectifs de cette thèse, cet aspect sera donc discuté dans la deuxième partie.

1.2 Les méthodes stochastiques

Envisagées dès les années 70 par les centres de prévision météo, elles se caractérisent par la recherche du B.L.U.E. (Best Linear Unbiased Estimator). Plus précisément le principe est d'ajuster la trajectoire du modèle afin qu'elle coïncide le mieux possible avec les champs d'observations disponibles. Chaque étape d'analyse détermine un état optimal compte tenu des observations, du modèle et des hypothèses faites sur les erreurs de ceux-ci. Cet état analysé servira ensuite de condition initiale pour la prévision.

1.2.1 Le filtre de Kalman et ses amis

En 1960, R. E. Kalman et R. Bucy ont décrit une solution récursive pour des problèmes de filtrage linéaire de données discrètes. Depuis cette date, grâce en particulier aux avancées faites dans le domaine des moyens de calcul, le filtre de Kalman (KF) a été l'objet de nombreuses recherches et d'applications dans de nombreux domaines (par exemple pour le guidage d'engins spatiaux). L'introduction de cette méthode en météorologie revient à Jones en 1965 [65]. En océanographie et en météorologie la taille des problèmes empêche toute application du filtre de Kalman dans sa version complète mais quelques variantes nettement moins coûteuses ont été proposées et donnent des résultats très encourageants.

Le but de ce filtre est de minimiser la variance de l'erreur de l'état analysé c'est à dire qu'on va corriger la trajectoire du modèle en fonction des observations de manière à minimiser l'écart entre l'état vrai et l'état analysé. Une des caractéristiques de cette méthode est qu'elle fournit directement une estimation de la matrice de covariance associée à cette erreur, en tenant compte de sa propagation par la dynamique du modèle. Nous présenterons ici la version du filtre étendue aux dynamiques non linéaires (EKF) [64]. En partant des équations (1.9) et (1.11)

$$\begin{aligned} \mathbf{x}_{i+1}^t &= M_{(t_{i+1}, t_i)}(\mathbf{x}_i^t) + \eta_{i+1} \\ \mathbf{y}_i^o &= H_{t_i}(\mathbf{x}_i^t) + \varepsilon_i \end{aligned}$$

le principe du filtre de Kalman est d'appliquer l'algorithme suivant par récurrence sur les temps d'observation t_i Il se décompose en deux phases : une phase de prédiction qui consiste en l'intégration du modèle à partir de l'état d'analyse de l'itération précédente.

$$\mathbf{x}_{i+1}^f = M_{(t_{i+1}, t_i)}(\mathbf{x}_i^a) \tag{1.15}$$

On peut alors calculer la matrice \mathbf{P}_{i+1}^f de covariance d'erreur de prévision en soustrayant \mathbf{x}_{i+1}^t à l'équation précédente et en utilisant l'hypothèse linéaire tangente (1.13)

$$\mathbf{x}_{i+1}^f - \mathbf{x}_{i+1}^t = \mathbf{M}_{(t_{i+1}, t_i)}[\mathbf{x}_i^a - \mathbf{x}_i^t] + [M_{(t_{i+1}, t_i)}(\mathbf{x}_i^t) - \mathbf{x}_{i+1}^t]$$

puis multipliant à droite cette équation par sa transposée et prenant son espérance, on obtient

$$\mathbf{P}_{i+1}^f = \mathbf{M}_{(t_{i+1}, t_i)}^T \mathbf{P}_i^a \mathbf{M}_{(t_{i+1}, t_i)} + \mathbf{Q}_i \quad (1.16)$$

les termes de corrélation croisée étant nuls (voir partie 1.1.2).

La deuxième phase est dite de correction : on obtient l'état analysé en corrigeant l'état prédit en fonction des observations par

$$\mathbf{x}_{i+1}^a = \mathbf{x}_{i+1}^f + \mathbf{K}_{i+1}[\mathbf{y}_{i+1}^o - H_{i+1}(\mathbf{x}_{i+1})] \quad (1.17)$$

où \mathbf{K} est une matrice de gain qui minimise la variance de l'erreur d'analyse. De la même manière que pour \mathbf{P}^f on peut calculer \mathbf{P}^a grâce à l'équation précédente et on obtient:

$$\mathbf{P}_{i+1}^a = [I - \mathbf{K}_{i+1} \mathbf{H}_{i+1}] \mathbf{P}_{i+1}^f [I - \mathbf{K}_{i+1} \mathbf{H}_{i+1}] + \mathbf{K}_{i+1} \mathbf{R}_{i+1} \mathbf{K}_{i+1}^T$$

De façon à minimiser la trace de \mathbf{P}_{i+1}^a on détermine un gain \mathbf{K}_{i+1} optimal

$$\mathbf{K}_{i+1} = \mathbf{P}_{i+1}^f \mathbf{H}_{i+1}^T [\mathbf{H}_{i+1} \mathbf{P}_{i+1}^f \mathbf{H}_{i+1}^T + \mathbf{R}_{i+1}]^{-1} \quad (1.18)$$

et pour ce gain, la matrice de covariance d'erreur d'analyse se simplifie et s'écrit

$$\mathbf{P}_{i+1}^a = \mathbf{P}_{i+1}^f - \mathbf{K}_{i+1} \mathbf{H}_{i+1} \mathbf{P}_{i+1}^f \quad (1.19)$$

Les équations de (1.15) à (1.19) représentent l'algorithme du filtre de Kalman étendu.

Dans le cas d'une dynamique linéaire il s'agit d'un filtre stochastique optimal, ce qui explique les efforts développés dans les autres méthodes pour obtenir une équivalence avec ce filtre dans le cas linéaire.

Cependant dans le cas non linéaire le filtre n'est plus nécessairement optimal. BUDGELL [19] l'a appliqué pour un modèle simplifié mais tout de même non linéaire et a mis en évidence les difficultés qui peuvent être rencontrées lors des brusques changements de comportements. Dans des configurations réalistes en météorologie ou en océanographie, la taille N des systèmes rencontrés (typiquement $N = 10^6 - 10^8$) rend l'application de la version complète du filtre inaccessible malgré les moyens de calcul actuels. En effet l'étape (1.16) de propagation des erreurs par le modèle va nécessiter $2N$ intégration de ce modèle et cela demandera de stocker des objets de taille N^2 . De plus les covariances d'erreurs \mathbf{R} et surtout \mathbf{Q} sont très mal connues et leur rôle est déterminant pour la qualité de l'analyse. Pour pallier ces difficultés, de nombreuses variantes ont été proposées. Sans prétendre être exhaustif on peut citer

- Le filtre de Kalman d'ensemble, proposé par EVENSEN [46] : on effectue une série de tirages aléatoires sur la fonction de probabilité de l'état du modèle, et on fait évoluer chacun d'eux en intégrant le modèle pour en déduire les covariances d'erreur. Cette variante si elle résout l'aspect covariances mal connues, aggrave les difficultés liées au coût du filtre.

- Les filtres de Kalman adaptatifs peuvent eux aussi pallier le manque d'information *a priori* sur les erreurs. Il s'agit de déterminer, en plus de l'état optimal, certains paramètres des matrices de covariance d'erreur (MYERS ET TAPLEY [86], DEE [38]). BLANCHET et al [12] ont mis en oeuvre une telle méthode pour mieux définir les covariances d'erreur du modèle, cependant cela implique une hausse non négligeable du coût en temps de calcul.
- Afin de réduire le coût du EKF, on peut essayer de propager les matrices de covariance d'erreur seulement sur un espace réduit. La base de cet espace est généralement déterminée par une analyse en composantes principales dont on ne garde que les premières (CANE et al [20]).
- On peut aussi considérer que les erreurs ne sont corrélées que dans un voisinage restreint, les matrices de covariance associées sont donc creuses ce qui réduit le coût (PARRISH ET COHN [92]).

Parmi les variantes les plus célèbres du filtre de Kalman on peut citer le filtre SEEK (Singular Evolutive Extended Kalman) proposé par PHAM et al [95] qui propage les matrices de covariance d'erreurs sur une base réduite déterminée par une analyse EOF (Empirical Orthogonal Function) et cette base peut évoluer pendant l'assimilation. De nombreuses variantes du SEEK ont été expérimentées sur des cas réalistes (Ballabrera et al [4], et les Thèses de CARME [21], PARENT [91], TESTUT [114] et HOTEIT [61])

1.2.2 L'interpolation Optimale

L'interpolation optimale (OI) peut être considérée comme une version simplifiée du filtre de Kalman ou *a contrario* on peut voir le KF comme une extension temporelle de l'OI (comme le 4D-Var est une extension temporelle du 3D-Var). En raison de son faible coût en calcul, elle a très tôt été appliquée en météorologie (GANDIN [50] 1963) et est encore une méthodes largement utilisée de façon opérationnelle en océanographie en raison de son faible coût, notamment avec SOFA (SubOptimal Filtering and Assimilation, DE MEY [35]) dans le cadre de MERCATOR. Ici la matrice de covariance d'erreur de l'ébauche est considérée comme étant le produit d'une matrice diagonale, ayant les variances d'erreur sur la diagonale et d'une matrice décrivant les corrélations spatiales. La méthode ne fait plus évoluer la matrice complète mais seulement les variances. Le choix des structures de corrélation est donc déterminant pour l'efficacité de cette méthode. En raison de la taille des systèmes en océanographie et en météorologie de nombreuses approximations doivent être faites dans la construction de ces structures. De plus le fait que l'aspect temporel ne soit pas traité directement par la méthode limite son efficacité.

1.2.3 Le Nudging, la SMC (Successive Corrections Method) et l'ACS (Analysis Correction Scheme)

Nudging

Une des plus simples et plus intuitives méthodes d'assimilation de données consiste à effectuer un rappel Newtonien vers les observations. Ceci est fait en ajoutant un terme de rappel vers les observations dans les équations du modèle

$$\mathbf{x}_{t_i}^a = \mathbf{x}_{t_i}^f + \mathbf{K}[\mathbf{y}_{t_i}^o - H_{t_i}(\mathbf{x}_{t_i}^f)]$$

L'amplitude \mathbf{K} du rappel est définie de façon empirique et permet de délocaliser l'attracteur du système original. Cette approche n'a, à ma connaissance, pas été utilisée pour des applications météorologiques mais elle a fréquemment été employée en océanographie [60] [121] [15]. Une des principales limitations de cette méthode est qu'elle engendre des distorsions entre les variables qui sont observées et celles qui ne le sont pas, ce qui peut générer des ondes de gravité [59] et de ce fait altérer les prévisions. Bien que le nudging ne soit pas une méthode stochastique à part entière, elle trouve sa place dans le présent paragraphe car elle peut être vue comme un filtre de Kalman avec une matrice de gain \mathbf{K} fixe.

SCM (Successive Corrections Method)

La SCM est une variante du nudging où l'on itère la relaxation Newtonienne (le rappel aux observations). Cette méthode a été l'une des premières à être appliquée à des problèmes pratiques. BERGTHORSSON et DOOS [11] ont été les premiers à l'introduire en 1955, rapidement suivis par CRESSMAN [32]. L'algorithme de la SCM peut s'écrire :

$$(\mathbf{x}_{t_i}^a)^{j+1} = (\mathbf{x}_{t_i}^a)^j + \mathbf{K}[\mathbf{y}_{t_i}^o - H_{t_i}((\mathbf{x}_{t_i}^a)^j)]$$

avec $(\mathbf{x}_{t_i}^a)^0 = \mathbf{x}_{t_i}^f$ est l'état prédit au temps t_i (avant assimilation), $(\mathbf{x}_{t_i}^a)^k$ est l'état analysé toujours au temps t_i après k corrections et \mathbf{K} est une matrice de gain fixée. Le point crucial de cette méthode est le choix du gain \mathbf{K} . Il était tout d'abord choisi de façon empirique, puis des améliorations successives ont abouti par CRESSMAN [32] à préconiser que les poids soient choisis de façon à lisser l'influence des observations dans l'analyse, évitant ainsi de brusques sauts dans l'état analysé. DALEY [33] a montré que si la matrice \mathbf{K} est choisie de telle manière que la méthode converge, alors celle-ci converge toujours vers les observations à mesure que les itérations augmentent. Pour cette raison on se restreint à seulement quelques itérations pour éviter une analyse trop bruitée [83].

ACS (Analysis Correction Scheme)

L'ACS est un dérivé de la SCM. Ici les corrections sont faites aussi bien sur les observations que sur l'état prédit, ceci pour prendre en compte les erreurs d'observation. BRATSETH [18] proposa d'utiliser une méthode d'OI comme référence pour obtenir une matrice \mathbf{K} optimale. LORENC, BELL et MACPHERSON [78] ont aussi montré que cette méthode converge vers le résultat de l'OI à mesure que le nombre d'itérations augmente. Dans le

cas linéaire l'algorithme de l'ACS peut s'écrire :

$$\begin{cases} (\mathbf{x}_{t_i}^a)^{j+1} = (\mathbf{x}_{t_i}^a)^j + \mathbf{W}_i \mathbf{Q}_i [(\mathbf{y}_{t_i}^o)^j - H_{t_i}((\mathbf{x}_{t_i}^a)^j)] \\ (\mathbf{y}_{t_i}^o)^{j+1} = (\mathbf{y}_{t_i}^o)^j - \mathbf{Q}[\mathbf{y}_{t_i}^o - H_{t_i}((\mathbf{x}_{t_i}^a)^j)] \end{cases} \quad (1.20)$$

où $(\mathbf{y}_{t_i}^o)^0$ est le vecteur d'observation et $\mathbf{W}_i = \mathbf{B}_i \mathbf{H}_{t_i}^T \mathbf{R}_i^{-1}$, $\mathbf{Q}_i = (\mathbf{H}_{t_i} \mathbf{W}_i + \mathbf{I})^{-1}$. L'état analysé est la limite quand k tend vers l'infini des $(\mathbf{x}_{t_i}^a)^k, (\mathbf{y}_{t_i}^o)^k$. Des variantes de cette méthode sont encore utilisées au UKMO² avec le modèle d'océan FOAM.

1.3 Les méthodes variationnelles

Introduites dès 1955 par SASAKI [103][104] les méthodes variationnelles sont devenues, depuis le début des années 90, celles utilisées de façon opérationnelle dans les centres de prévision météorologique (au NMC en 1991 [93], au ECMWF en 1996 [31] [45] et Météo-France en 1997).

Cette autre approche de l'assimilation de données n'est plus basée sur des théories statistiques, mais sur la théorie de l'optimisation. Le principe de ces méthodes est la minimisation d'une fonctionnelle (la fonction objective) mesurant les caractéristiques indésirables de la solution du modèle. Ces caractéristiques peuvent être l'écart aux observations, la présence d'ondes de gravité, le non respect de certains équilibres, etc... Dans la suite on ne s'intéressera qu'à l'écart aux observations et à l'éloignement à l'ébauche de la condition initiale.

Si les statistiques ne sont plus les bases de ces méthodes, elles restent largement utilisées pour les calibrer et définir la fonction à minimiser.

1.3.1 le 3D-Var

Le 3D-Var (pour 3 Dimensional VARIational assimilation) proposé par SASAKI [103] consiste à chercher l'état le plus vraisemblable à partir des connaissances que l'on a des lois de probabilité des erreurs d'observation et de celles de l'ébauche.

Comme son nom l'indique le 3D-Var traite de problèmes tridimensionnels, par abus de langage on utilisera également cette appellation pour des problèmes à une ou deux dimensions (pas de 1D ou 2D-Var) pour éviter des risques de confusion avec l'extension temporelle de cette méthode³.

Comme pour le filtre de Kalman, le 3D-Var consiste à minimiser la distance (au sens des moindres carrés cette fois ci) entre l'état estimé et les différentes sources d'informations (la prévision précédente et les observations), le nouvel état (état analysé) étant pris comme point de départ de la prévision suivante.

L'état analysé est l'état \mathbf{x} minimisant la fonction objective

$$J(\mathbf{x}_i) = \underbrace{\frac{1}{2} [H_{t_i}(\mathbf{x}_i) - \mathbf{y}_i^o]^T \mathbf{R}_i^{-1} [H_{t_i}(\mathbf{x}_i) - \mathbf{y}_i^o]}_{J_o} + \underbrace{\frac{1}{2} [\mathbf{x}_i - \mathbf{x}_i^b]^T \mathbf{B}^{-1} [\mathbf{x}_i - \mathbf{x}_i^b]}_{J_b} \quad (1.21)$$

²U.K. met. office : centre de météorologie du Royaume Uni.

³le 4D-Var qui deviendra 2D-Var appliqué à des modèles 1D et 3D-Var pour des modèles 2D

\mathbf{x}^b étant issu de l'intégration par le modèle de l'état assimilé à l'étape précédente. L'équilibre entre le terme d'écart aux observations (J_o) et celui de l'ébauche ou de régularisation (J_b) est effectué grâce aux inverses des matrices de covariance d'erreur d'observation et de l'ébauche, c'est à dire à la confiance qu'on a dans les observations et l'ébauche respectivement.

La minimisation se fait grâce à une méthode de gradient, celui ci étant facilement calculé depuis (1.21).

$$\nabla J = \mathbf{H}_{t_i}^T \mathbf{R}_i^{-1} [H_{t_i}(\mathbf{x}_i) - \mathbf{y}_i^o] + \mathbf{B}^{-1} [\mathbf{x}_i - \mathbf{x}_i^b]$$

Si les variables du modèle sont reliées linéairement aux observations c'est à dire si l'opérateur d'observation H est linéaire, la solution de ce problème est équivalente à celle obtenue par l'OI (LORENC [77])

Si avec des observations réparties régulièrement dans le temps on peut effectuer une étape d'assimilation 3D-Var à chaque pas de temps où celles ci sont présentes, en pratique les sources d'observation sont multiples et inégalement réparties. Afin de ne pas multiplier les phases d'assimilation on est amené à décider arbitrairement de la taille de chaque fenêtre d'assimilation (par exemple on décide d'assimiler toutes les six heures les observations obtenues durant ce laps de temps). Sans être exhaustif, on peut citer deux stratégies pour utiliser ces observations

- 3D-var Classique : on regroupe (moyenne) toutes les observations de la fenêtre d'assimilation à l'instant où l'on fait l'assimilation (voir fig (1.1)) en les ramenant éventuellement sur la grille du modèle

$$J(\mathbf{x}_{t_0}) = \frac{1}{2} [\bar{H}(\mathbf{x}_{t_0}) - \bar{\mathbf{y}}^o]^T \bar{\mathbf{R}}^{-1} [\bar{H}(\mathbf{x}_{t_0}) - \bar{\mathbf{y}}^o] + \frac{1}{2} [\mathbf{x}_{t_0} - \mathbf{x}_{t_0}^b]^T \mathbf{B}^{-1} [\mathbf{x}_{t_0} - \mathbf{x}_{t_0}^b] \quad (1.22)$$

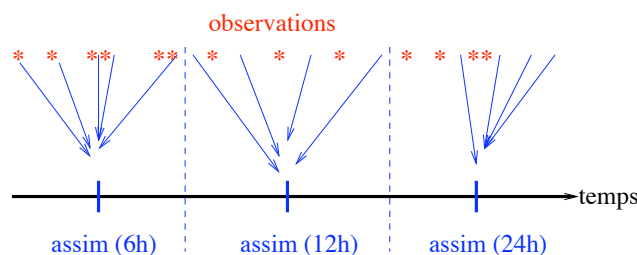


FIG. 1.1: Répartition des observations pour le 3D-Var classique

- 3D-FGAT (pour First Guess at Appropriate Time) : les données sont comparées à leur équivalent modèle au même instant alors que la correction se fait toujours à un instant donné de la fenêtre. Pour ce faire on utilise une formulation incrémentale c'est à dire qu'on estime seulement l'incrément $\delta \mathbf{x}_{t_0} = \mathbf{x}_{t_0} - \mathbf{x}_{t_0}^b$ nécessaire sur notre

état (et plus l'état complet) ce qui permet d'utiliser l'hypothèse linéaire tangente (1.14)(voir partie suivante pour plus de détails)

$$J(\delta \mathbf{x}_{t_0}) = \sum_{t_i=t_0-3h}^{t_0+3h} \frac{1}{2} [\mathbf{H}_{t_i} \delta \mathbf{x}_{t_0} - \mathbf{d}_{t_i}]^T \mathbf{R}_i^{-1} [\mathbf{H}_{t_i} \delta \mathbf{x}_{t_0} - \mathbf{d}_{t_i}] + \frac{1}{2} [\delta \mathbf{x}_{t_0}]^T \mathbf{B}^{-1} [\delta \mathbf{x}_{t_0}] \quad (1.23)$$

avec $\mathbf{d}_{t_i} = H(\mathbf{x}_{t_i}) - \mathbf{y}_{t_i}^o$

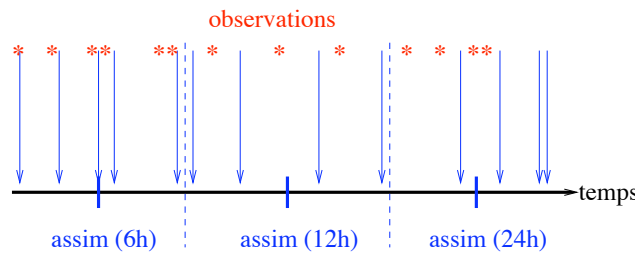


FIG. 1.2: Répartition des observations pour le 3D-FGAT

Pour éviter des problèmes de “choc” dans le modèle au moment de l'introduction de la correction de l'état on peut ne pas appliquer cette correction à un instant donné de la fenêtre mais la répartir sur l'ensemble de celle ci.

1.3.2 Le 4D-Var

Le 4D-Var est l'extension temporelle du 3D-Var; elle ne vise plus à obtenir l'état optimal à un instant donné, mais la trajectoire optimale sur une fenêtre de temps donnée. Pour une assimilation les observations sont maintenant distribuées, à la fois dans l'espace, mais aussi dans le temps sur toute la fenêtre d'assimilation. Seul l'état initial de cette fenêtre est modifié et l'état initial optimal obtenu doit permettre au modèle d'avoir la trajectoire la plus proche possible de l'ensemble des observations utilisées.

Cette amélioration du 3D-Var (LE DIMET, 1980 [71]) permet d'ajouter la connaissance de l'évolution du système (le modèle) comme information pour l'assimilation.

De nombreuses applications à des modèles réalistes météorologiques [115] [133] et océanographiques [84] [106] [80] [53] ont été effectuées.

L'amélioration ainsi apportée conjuguée au fort développement des moyens de calculs fait que le 4D-Var est venu remplacer le 3D-Var dans les systèmes de prévision opérationnels de l'atmosphère de l'ECMWF en 1997 et de Météo France en 2000.

Le 4D-Var classique

Comme dans le cas du 3D-Var, la méthode est représentée au moyen d'une fonction objective à minimiser

$$J(\mathbf{x}_0) = \overbrace{\frac{1}{2} \sum_{i=1}^n [H_{t_i}(\mathbf{x}_i) - \mathbf{y}_i^o]^T \mathbf{R}_i^{-1} [H_{t_i}(\mathbf{x}_i) - \mathbf{y}_i^o]}^{J_o} + \overbrace{\frac{1}{2} [\mathbf{x}_0 - \mathbf{x}^b]^T \mathbf{B}^{-1} [\mathbf{x}_0 - \mathbf{x}^b]}^{J_b} \quad (1.24)$$

Du fait de la dépendance très indirecte du terme J_o au vecteur de contrôle \mathbf{x}_0 , calculer le gradient de cette fonctionnelle directement est très ardu voire impossible. L'utilisation des méthodes adjointes rend ce calcul possible à un coût raisonnable (LE DIMET, 1980 [71], LE DIMET ET TALAGRAND, 1986 [72]). Celles ci peuvent se décrire de la façon suivante :

Si on dispose du modèle décrivant l'évolution temporelle de l'état du système :

$$\begin{cases} \mathbf{x}_{i+1} = M_{(t_i, t_{i+1})}(\mathbf{x}_i) \\ \mathbf{x}_0 = \mathbf{x}_0 \end{cases} \quad (1.25)$$

Et si on applique une petite perturbation $\mathbf{h}_{\mathbf{x}_0}$ à \mathbf{x}_0 , la dérivée de Gâteaux $\hat{\mathbf{x}}$ suit l'évolution :

$$\begin{cases} \hat{\mathbf{x}}_{i+1} = \mathbf{M}_{(t_i, t_{i+1})} \hat{\mathbf{x}}_i \\ \hat{\mathbf{x}}_0 = \mathbf{h}_{\mathbf{x}_0} \end{cases} \quad (1.26)$$

De même la dérivée directionnelle de la fonction coût s'écrit

$$\hat{J}(\mathbf{x}_0, \mathbf{h}_{\mathbf{x}_0}) = \sum_{i=1}^n [\mathbf{H}_{t_i} \hat{\mathbf{x}}_i]^T \mathbf{R}_i^{-1} [H_{t_i}(\mathbf{x}_i) - \mathbf{y}_i^o] + \mathbf{h}_{\mathbf{x}_0}^T \mathbf{B}^{-1} [\mathbf{x}_0 - \mathbf{x}^b] \quad (1.27)$$

d'autre part, par définition :

$$\hat{J}(\mathbf{x}_0, \mathbf{h}_{\mathbf{x}_0}) = (\nabla_{\mathbf{x}_0} J, \mathbf{h}_{\mathbf{x}_0}) \quad (1.28)$$

en introduisant \mathbf{x}^* l'état adjoint, on peut écrire l'égalité (simple réarrangement) :

$$\sum_{i=0}^n [\mathbf{x}_i^*]^T \hat{\mathbf{x}}_i = \sum_{i=0}^{n-1} [\mathbf{x}_{i+1}^*]^T [\mathbf{M}_{(t_i, t_{i+1})} \hat{\mathbf{x}}_i + [\mathbf{x}_0^*]^T \hat{\mathbf{x}}_0] \quad (1.29)$$

$$= \sum_{i=0}^{n-1} [\mathbf{M}_{(t_i, t_{i+1})}^T \mathbf{x}_{i+1}^*]^T \hat{\mathbf{x}}_i + [\mathbf{x}_0^*]^T \hat{\mathbf{x}}_0 \quad (1.30)$$

si on définit dans (1.30) l'état adjoint par

$$\begin{cases} \mathbf{x}_i^* = \mathbf{M}_{(t_i, t_{i+1})}^T \mathbf{x}_{i+1}^* + \mathbf{H}_{t_i}^T \mathbf{R}_i^{-1} (H_{t_i}(\mathbf{x}_i) - \mathbf{y}_i^o) \\ \mathbf{x}_n^* = 0 \end{cases} \quad (1.31)$$

par identification avec (1.27) et par (1.28), on obtient :

$$\nabla_{\mathbf{x}_0} J = \mathbf{x}_0^* + \mathbf{B}^{-1} [\mathbf{x}_0 - \mathbf{x}^b] \quad (1.32)$$

Par la suite on préférera formuler le vecteur de contrôle en tant qu'incrément sur la condition initiale et non comme la condition initiale en sa totalité. Les deux problèmes sont rigoureusement identiques mais cela permet de simplifier un peu les notations. Cette formulation habituellement réservée à la variante incrémentale (voir plus loin) a été préférée car elle simplifie l'écriture du terme J_b de *background* de la fonction objective. Celle ci s'écrit maintenant

$$J(\delta \mathbf{x}_0) = \overbrace{\frac{1}{2} \sum_{i=1}^n [H_{t_i}(\mathbf{x}_i) - \mathbf{y}_i^o]^T \mathbf{R}_i^{-1} [H_{t_i}(\mathbf{x}_i) - \mathbf{y}_i^o]}^{J_o} + \overbrace{\frac{1}{2} [\delta \mathbf{x}_0]^T \mathbf{B}^{-1} \delta \mathbf{x}_0}^{J_b} \quad (1.33)$$

où $\delta \mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}^b$ est l'incrément ajouté à la condition initiale. Dans cette formulation l'expression du gradient devient

$$\nabla_{\delta \mathbf{x}_0} J = \mathbf{x}^*_0 + \mathbf{B}^{-1} \delta \mathbf{x}_0 \quad (1.34)$$

Dans le cas linéaire

Dans le cas d'un modèle et d'un opérateur d'observation linéaire on peut introduire \mathbf{G} l'opérateur d'observation généralisé qui contient l'opérateur d'observation et le modèle.

Si on note

$$\mathbf{y}^o = \begin{pmatrix} \mathbf{y}_1^o \\ \mathbf{y}_2^o \\ \dots \\ \mathbf{y}_n^o \end{pmatrix}; \mathbf{M} = \begin{pmatrix} \mathbf{M}_{t_0, t_1} \\ \mathbf{M}_{t_0, t_2} \\ \dots \\ \mathbf{M}_{t_0, t_n} \end{pmatrix}; \mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \dots \\ \mathbf{x}_n \end{pmatrix} = \mathbf{M} \mathbf{x}_0; \mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & 0 & \dots & 0 \\ 0 & \mathbf{R}_2 & & \dots \\ \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & \mathbf{R}_n \end{pmatrix}$$

$$; \mathbf{H} = \begin{pmatrix} \mathbf{H}_1 & 0 & \dots & 0 \\ 0 & \mathbf{H}_2 & & \dots \\ \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & \mathbf{H}_n \end{pmatrix} \text{ et } \mathbf{G} = \mathbf{H} \mathbf{M}$$

alors la fonction coût (1.33) devient

$$J(\delta \mathbf{x}_0) = \frac{1}{2} [\mathbf{G} \mathbf{x}_0 - \mathbf{y}^o]^T \mathbf{R}^{-1} [\mathbf{G} \mathbf{x}_0 - \mathbf{y}^o] + \frac{1}{2} [\delta \mathbf{x}_0]^T \mathbf{B}^{-1} \delta \mathbf{x}_0 \quad (1.35)$$

Une telle formulation permet de considérer formellement le 4D-Var comme un 3D-Var avec l'opérateur d'observation généralisé remplaçant l'opérateur d'observation.

Remarque :

Si les données ne sont pas uniformément réparties dans l'espace et dans le temps (ce qui est le cas en pratique) la méthode d'assimilation devra être capable de "diffuser" l'information fournie sur l'ensemble du domaine et de la fenêtre temporelle. Dans toutes les variantes du 4D-Var l'information apportée par les données se propage grâce au modèle adjoint pour les aspects temporels et spatiaux et grâce à la matrice de covariance d'erreur de *background* pour les aspects spatiaux. Pour s'en convaincre regardons les équations du 4D-Var écrites sous la forme (1.35). En dérivant formellement la fonction objective (1.33) et en utilisant

le fait que à l'optimum $\nabla J(\delta \mathbf{x}^a) = 0$ on obtient $\delta \mathbf{x}^a$

$$\begin{aligned} J(\delta \mathbf{x}_0) &= \frac{1}{2}[\mathbf{H}\mathbf{M}\delta \mathbf{x}_0 - \mathbf{d}]^T \mathbf{R}^{-1}[\mathbf{H}\mathbf{M}\delta \mathbf{x}_0 - \mathbf{d}] + \frac{1}{2}[\delta \mathbf{x}_0]^T \mathbf{B}^{-1} \delta \mathbf{x}_0 \\ \nabla J &= \mathbf{M}^T \mathbf{H}^T \mathbf{R}^{-1}[\mathbf{H}\mathbf{M}\delta \mathbf{x}_0 - \mathbf{d}] + \mathbf{B}^{-1} \delta \mathbf{x}_0 \\ \delta \mathbf{x}^a &= [\mathbf{B}^{-1} + \mathbf{M}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\mathbf{M}]^{-1} \mathbf{M}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d} \end{aligned} \quad (1.36)$$

En utilisant une formule classique du calcul matriciel⁴ on peut réécrire (1.36)

$$\delta \mathbf{x}^a = \overbrace{\mathbf{B}\mathbf{M}^T \mathbf{H}^T}^{(1)} \overbrace{[\mathbf{R}^{-1} + \mathbf{H}\mathbf{M}\mathbf{B}\mathbf{M}^T \mathbf{H}^T]^{-1}}^{(2)} \mathbf{d} \quad (1.37)$$

Le terme (2) de l'équation (1.37) est clairement dans l'espace des observations, c'est à dire que si il n'y avait que ce terme l'incrément ne corrigerait l'état qu'aux endroits effectivement observés. C'est le terme (1) qui permet de corriger l'ensemble du vecteur d'état. Comme il a été dit plus haut l'intégration rétrograde du modèle adjoint va faire remonter l'information apportée par \mathbf{d} depuis la fin de la fenêtre d'assimilation jusqu'à la condition initiale tout en diffusant cette information grâce à la physique du modèle vers les points et les variables non observées. Cette diffusion de l'information dans l'espace et entre les variables sera parachevée grâce aux structures de corrélation de la matrice \mathbf{B} .

On comprend bien, au regard de cette approche intuitive des fonctionnements de l'assimilation de données, l'importance d'avoir de bonnes structures de corrélation spatiales et entre variables dans la matrice \mathbf{B} si on a peu d'observations. Le rôle de \mathbf{B} est encore plus crucial pour le 3D-Var pour lequel l'intégration du modèle adjoint n'a pas lieu.

En résumé si on a peu d'informations par les observations, on a besoin d'en apporter par les matrices de covariances. La nature de ces informations est toutefois différente : dans le cas des matrices de covariances d'erreur il s'agit plus de restreindre la recherche de l'incrément à ceux satisfaisant à ces covariances (on a réduit l'ensemble admissible).

Variantes incrémentales

La précédente approche, peut avoir deux principales limitations :

- La taille du problème peut rendre le volume de calcul et le stockage de données nécessaires trop importants pour des applications opérationnelles de longue durée (intégration des modèles, optimisation)
- Les non linéarités du modèle qui impliqueront que la fonction objective ne sera pas quadratique et ce qui pourra rendre la minimisation moins efficace.

COURTIER et al [30] ont proposé la variante incrémentale du 4D-Var. Elle a pour but de fournir une solution approchée du 4D-Var à moindre coût.

Elle utilise l'hypothèse linéaire tangente (1.13) ainsi que son équivalent sur l'opérateur d'observation $H_{t_i}(\mathbf{x} + \delta \mathbf{x}) = H_{t_i}(\mathbf{x}) + \mathbf{H}_{t_i} \delta \mathbf{x}$ pour ne s'intéresser qu'à l'évolution d'une

⁴ $(\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{C}^{-1} = \mathbf{A} \mathbf{B}^T (\mathbf{C} + \mathbf{B} \mathbf{A} \mathbf{B}^T)^{-1}$ si les inverses existent (voir WUNSCH [130], 1996 pp 99)

petite variation $\delta \mathbf{x}_0$ de la condition initiale, celle ci se faisant suivant le modèle linéaire tangent \mathbf{M} associé au modèle M (linéarisé autour de la trajectoire issue de l'ébauche \mathbf{x}^b). On peut réécrire la fonction objective du 4D-Var :

$$\begin{aligned} J(\delta \mathbf{x}_0) &= \frac{1}{2} \sum_{i=1}^n [H_{t_i}(M_{t_i, t_0}(\mathbf{x}_0)) - \mathbf{y}_i^o]^T \mathbf{R}_i^{-1} [H_{t_i}(M_{t_i, t_0}(\mathbf{x}_0)) - \mathbf{y}_i^o] + \frac{1}{2} [\delta \mathbf{x}_0]^T \mathbf{B}^{-1} \delta \mathbf{x}_0 \\ &= \frac{1}{2} \sum_{i=1}^n [\mathbf{H}_{t_i} \mathbf{M}_{t_i, t_0} \delta \mathbf{x}_0 - \mathbf{d}_i]^T \mathbf{R}_i^{-1} [\mathbf{H}_{t_i} \mathbf{M}_{t_i, t_0} \delta \mathbf{x}_0 - \mathbf{d}_i] + \frac{1}{2} [\delta \mathbf{x}_0]^T \mathbf{B}^{-1} \delta \mathbf{x}_0 \end{aligned} \quad (1.38)$$

où $\mathbf{d}_i = \mathbf{y}_i^o - H_{t_i}(M_{t_i, t_0}(\mathbf{x}^b))$. La fonction objective ainsi obtenue (1.38) est quadratique ce qui permettra une minimisation plus efficace. On obtient notre condition initiale optimale de la même façon que dans le 4D-Var classique par l'addition de l'incrément optimal au terme de background.

$$\mathbf{x}_0^a = \mathbf{x}^b + \delta \mathbf{x}_0^a$$

Une telle formulation permet également d'utiliser, dans le but de réduire le coût de la minimisation, un modèle simplifié (résolution plus faible et/ou physique plus simple). Si on note S et \mathbf{S} l'opérateur de simplification et son linéaire tangent, le vecteur d'état simplifié s'écrit :

$$\mathbf{x}_i^s = S(\mathbf{x}_i)$$

et l'incrément associé :

$$\delta \mathbf{x}_i^s = \mathbf{S} \delta \mathbf{x}_i$$

On pose \mathbf{S}^{-I} l'opérateur inverse généralisé de \mathbf{S}^5 , alors l'évolution de l'incrément $\delta \mathbf{x}_i^s$ pourra s'écrire

$$\delta \mathbf{x}_i^s = \mathbf{S} \mathbf{M}_{t_i, t_0} \mathbf{S}^{-I} \delta \mathbf{x}_0^s$$

et notre problème pourra se ramener à la minimisation dans l'espace simplifié de la fonction objective :

$$J^s(\delta \mathbf{x}_0^s) = \frac{1}{2} (\delta \mathbf{x}_0^s)^T (\mathbf{S}^{-I})^T \mathbf{B}^{-1} \mathbf{S}^{-I} \delta \mathbf{x}_0^s + \frac{1}{2} [\mathbf{H} \mathbf{S}^{-I} \delta \mathbf{x}_0^s - \mathbf{d}_i]^T \mathbf{R}_i^{-1} [\mathbf{H} \mathbf{S}^{-I} \delta \mathbf{x}_0^s - \mathbf{d}_i]$$

L'état analysé du problème de départ se retrouve alors aisément

$$\mathbf{x}_0^a = \mathbf{x}^b + \mathbf{S}^{-I} (\delta \mathbf{x}_0^s)^a$$

Pour tenir compte des non linéarités du modèle et de l'opérateur d'observation, il est nécessaire de recalculer la trajectoire de référence (autour de laquelle sont linéarisés les opérateurs) on parle alors de boucle externe (par opposition aux boucles internes que représentent les itérations de la minimisation). Dans la méthode opérationnelle à Météo-France, la résolution des opérateurs simplifiés évolue au fil de la boucle externe, on parle alors de technique multi-incrémentale (VEERSÉ et THÉPAUT 1998 [120]).

⁵ \mathbf{S} n'est pas forcément inversible, si $\mathbf{y} = \mathbf{S}\mathbf{x}$ alors on définit alors son inverse généralisée \mathbf{S}^{-I} telle que $\mathbf{x} = \mathbf{S}^{-I}\mathbf{y}$ où \mathbf{x} est obtenu comme solution du problème des moindres carrés par minimisation de $J(\mathbf{x}) = \|\mathbf{y} - \mathbf{S}\mathbf{x}\|^2$

La variante “modèle imparfait”

Que ce soit pour le 4D-Var classique ou pour sa variante incrémentale, on a considéré que le modèle décrivait exactement le comportement du système. Cependant, malgré l’utilisation de modèles extrêmement sophistiqués, ceux-ci comportent encore des erreurs qui ne peuvent être négligées (voir WERGEN, 1992 [129]) et qui ne pourront sans doute jamais l’être. Il est en effet peu probable que l’on soit en mesure de pouvoir un jour modéliser de façon exacte des systèmes aussi chaotiques que l’océan ou l’atmosphère.

Deux solutions possibles semblent se dégager : d’une part une analyse approfondie des erreurs modèles permettant une correction *a priori* des équations, et d’autre part l’ajout d’un terme correctif directement dans les équations du modèle, celui ci étant déterminé par contrôle optimal.

Pour prendre en compte ces inexactitudes, on peut rajouter dans la variable de contrôle un terme (η) destiné à corriger les erreurs du modèle. Le modèle corrigé devient

$$\begin{cases} \mathbf{x}_{i+1} = M_{(t_i, t_{i+1})}(\mathbf{x}_i) + \eta_{i+1} \\ \mathbf{x}_0 = \mathbf{x}^b + \delta \mathbf{x}_0 \end{cases} \quad (1.39)$$

Et on cherche à minimiser la fonction objective

$$J(\delta \mathbf{x}_0, \eta_1, \dots, \eta_m) = \frac{1}{2} \sum_{i=1}^n [H(\mathbf{x}_i) - \mathbf{y}_i^o]^T \mathbf{R}_i^{-1} [H(\mathbf{x}_i) - \mathbf{y}_i^o] + \frac{1}{2} [\delta \mathbf{x}_0]^T \mathbf{B}^{-1} \delta \mathbf{x}_0 + \frac{1}{2} \sum_{i=1}^n \eta_i^T \mathbf{Q}_i^{-1} \eta_i \quad (1.40)$$

En utilisant la même démarche que dans le cas du 4D-Var classique et en introduisant le même modèle adjoint, on peut obtenir l’expression des différentes composantes du gradient :

$$\begin{cases} \nabla_{\delta \mathbf{x}_0} J = \mathbf{x}_0^* + \mathbf{B}^{-1} \delta \mathbf{x}_0 \\ \nabla_{\eta_i} J = \mathbf{x}_i^* + \mathbf{Q}_i^{-1} \eta_i \end{cases} \quad (1.41)$$

Pour cette variante, les problèmes de coûts supplémentaires induits et les éventuelles solutions seront discutés dans le chapitre 5.

1.3.3 Les méthodes duales

Introduites par Bennett en 1992 [7] (méthode des représentants) et par Amodei en 1995 [2], ces méthodes souvent appelées 3D- et 4D-PSAS (pour Physical-Space Statistical Analysis System) permettent de minimiser uniquement dans l’espace dual des observations, et par là même de diminuer grandement la taille du vecteur de contrôle. En optimisation, on sait que le problème associé au 4D-Var (ou au 3D-Var) dans le cas linéaire admet une formulation duale (LIONS 1968 [75]). De plus dans le cas d’un nombre fini m d’observations, la solution peut s’exprimer comme combinaison linéaire des noyaux associés à chaque opérateur d’observation élémentaire (les représentants de BENNETT). Alors que BENNETT propose de calculer explicitement ces représentants (ou une partie d’entre eux), AMODEI propose une méthode permettant d’obtenir le produit Λ -vecteur (Λ étant la base des représentants) sans avoir à calculer explicitement cette base. Cette méthode n’a eu des applications pratiques que très récemment (LOUVEL 1999 [79] et LAGARDE 2000 [70])

Présentons la méthode duale du problème du 4D-Var incrémental avec modèle imparfait. On pose $\mathbf{m}_i = \mathbf{R}_i^{-1}[\mathbf{H}_i \delta \mathbf{x}_i - \mathbf{d}_i]$.

En utilisant des notations semblables à (1.35), la fonction objective à minimiser est

$$J(\mathbf{m}) = \frac{1}{2} \mathbf{m}^T (\lambda + \mathbf{R}) \mathbf{m} - \mathbf{m}^T \mathbf{d}$$

et son gradient

$$\nabla J = (\lambda + \mathbf{R}) \mathbf{m} - \mathbf{d}$$

En intégrant l'équation

$$\begin{cases} \mathbf{x}_i^* = \mathbf{M}_{(t_i, t_{i+1})}^T \mathbf{x}_{i+1}^* + \mathbf{H}_i^T \mathbf{m}_i \\ \mathbf{x}_n^* = 0 \end{cases} \quad (1.42)$$

on peut calculer les incréments

$$\begin{cases} \delta \mathbf{x}_{i+1} = \mathbf{M}_{(t_i, t_{i+1})} \delta \mathbf{x}_i + \mathbf{Q}_i \mathbf{x}_i^* \\ \delta \mathbf{x}_0 = \mathbf{B} \mathbf{x}_0^* \end{cases} \quad (1.43)$$

Λ n'est jamais calculé explicitement, mais seulement le produit $\Lambda \mathbf{m}_i = \mathbf{H}_i \delta \mathbf{x}_i$.

Chapitre 2

Les différents modèles utilisés

Dans le présent chapitre on introduira les différents modèles utilisés afin de valider les méthodes introduites dans le présent ouvrage. Ils seront au nombre de trois, allant du plus sommaire au plus réaliste. Chacun d’entre eux est un compromis différent entre réalisme et facilité d’utilisation.

La vocation première d’un modèle de circulation générale de l’océan (OGCM) est la représentation spatio-temporelle des écoulements océaniques de grande échelle. Un tel modèle est construit à l’aide d’un ensemble d’équations issues des lois phénoménologiques de la mécanique des fluides (équations de Navier-Stokes) appliquées aux écoulements géophysiques planétaires et qui décrivent le comportement des variables d’état de l’océan.

Un modèle réaliste de circulation océanique est généralement très compliqué et difficile à manipuler, avec nombre de paramètres pouvant influencer sur son efficacité. C’est pourquoi, afin de valider les différentes hypothèses que nous serons amenés à faire, nous utiliserons deux modèles simplifiés mais pertinents pour la modélisation océanique : l’équation de Burger et un modèle Shallow-water. Ces applications préliminaires nous assureront de la viabilité de certains choix et nous permettront d’en éliminer d’autres. Elles permettront également d’utiliser des configurations qui auraient été trop coûteuses dans des applications réalistes et d’en comparer les résultats avec les approximations envisagées pour permettre les applications opérationnelles.

Le modèle opérationnel que nous avons sélectionné est le modèle OPA dans sa version 8.1 (MADEC ET AL. [82]). Ce choix nous a été dicté par sa large diffusion dans la communauté océanographique française et par le fait qu’il a été choisi dans des projets nationaux d’envergure comme CLIPPER ou MERCATOR. Ce modèle représente efficacement la dynamique non linéaire et les processus thermodynamiques de l’océan. Dans le présent ouvrage nous utiliserons une version caractérisée par une représentation verticale en coordonnées géopotentielles (z) et une approximation “toit rigide” de la surface océanique. La partie de ce chapitre consacrée à OPA reprendra les caractéristiques de ce modèle décrites dans le manuel de référence [82].

2.1 Les modèles “jouets”

2.1.1 L'équation de Burger

C'est le plus simple des modèles que nous allons utiliser. Il peut être vu comme une approximation 1D des équations de Navier-Stokes. Son très faible coût numérique nous permettra d'effectuer un test préliminaire sur toutes les hypothèses que nous pourrions élaborer. Cependant son unique dimension et la faible non linéarité de ses équations ne sont évidemment pas représentatives des problèmes que l'on pourra rencontrer dans une application réaliste.

L'équation de Burger s'écrit :

$$\begin{cases} \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \frac{\partial \mathbf{u}}{\partial x} - \nu \frac{\partial^2 \mathbf{u}}{\partial x^2} = \mathbf{g} & x \in]0, 1[, t > 0 \\ \mathbf{u}(0, t) = \mathbf{u}(1, t) = 0 \\ \mathbf{u}(x, 0) = \mathbf{u}_0(x) \end{cases} \quad (2.1)$$

Si on choisit $\mathbf{u}_0(x) = \sin \pi x$ et $\mathbf{g} = -e^{-t} \sin \pi x (1 + \nu + e^{-t} \cos \pi x)$, alors $\mathbf{u}(x, t) = e^{-t} \sin \pi x$ est solution de (2.1). La connaissance de la solution exacte nous permettra de maîtriser tous les paramètres de l'expérience (erreurs introduites dans le modèle, dans les observations, qualité du résultat, etc ...).

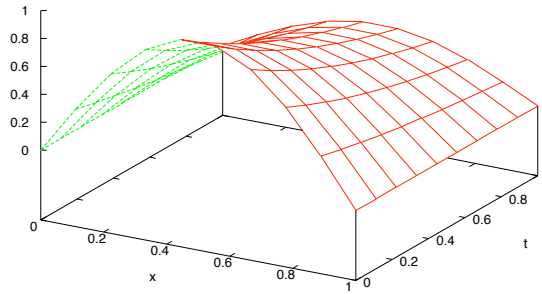


FIG. 2.1: la solution de l'équation de Burger pour $\mathbf{g}(x, t) = -e^{-t} \sin \pi x (1 + \nu + e^{-t} \cos \pi x)$

Pour discrétiser ce modèle on utilisera un schéma centré d'ordre deux en espace et un schéma d'Euler explicite en temps. Ce choix impose des critères de stabilité assez contraignants sur les pas de temps Δt et de discrétisation Δx : $\nu \frac{\Delta t}{\Delta x^2} \leq \frac{1}{2}$ et $\Delta t \leq 2\nu$ (on utilisera $N = 5000$ pas de temps pour $I = 20$ points de grille) mais a toujours le mérite de la simplicité.

On pose $i = 1, \dots, I - 1$ et $n = 0, \dots, N - 1$ et le modèle discrétisé s’écrit :

$$\begin{cases} \frac{1}{\Delta t}(u_i^{n+1} - u_i^n) + \frac{u}{\Delta x}(u_{i+1}^n - u_{i-1}^n) - \frac{\nu}{(\Delta x)^2}(u_{i+1}^n - 2u_i^n + u_{i-1}^n) = g(i\Delta x, n\Delta t) \\ u_i^0 = \sin i\pi\Delta x \\ u_0^n = 0 = u_I^n \end{cases} \quad (2.2)$$

2.1.2 Le modèle Shallow Water

C’est le deuxième modèle dont on disposera. L’appellation anglophone Shallow Water (eau peu profonde) sera préféré à la plus française “équation de Saint Venant” car plus représentative des approximations qui y sont faites.

Comme pour l’équation de Burger, le modèle Shallow Water présente l’intérêt d’un coût en calcul relativement faible et d’une grande facilité de manipulation. Cependant, même si la configuration utilisée est loin d’être réaliste, son comportement est proche de celui qu’on pourrait obtenir avec un “vrai” modèle d’océan. Une superposition de modèles Shallow-Water peut même constituer un modèle opérationnel, comme c’est le cas de MICOM (Miami Isopycnic Coordinate Model) (R. BLECK et D.B. BOUDRA 1986 [16]). Ses équations 2D et ses non-linéarités plus importantes en font un bon cas test pour l’introduction de nouvelles méthodes d’assimilation de données. La version qui sera utilisée dans le présent ouvrage, ainsi que son linéaire tangent et son adjoint ont été développés en collaboration avec Sophie DURBIANO dans le cadre de son travail de thèse.

Les équations :

Le point de départ de tous les modèles océaniques est l’équation de Navier-Stokes (2.3) qui traduit la conservation de la quantité de mouvement, à laquelle on adjoint l’équation de continuité (2.4) représentant la conservation de la masse.

$$\rho \frac{D\mathbf{u}}{Dt} - \mu \Delta \mathbf{u} = -\nabla \mathbf{p} + G + C \quad (2.3)$$

$$\frac{1}{\rho} \frac{D\rho}{Dt} = -\operatorname{div} \mathbf{u} \quad (2.4)$$

avec :

\mathbf{u}, \mathbf{p} et ρ la vitesse (dans les trois dimensions), la pression et la densité du fluide respectivement ;

D la dérivé lagrangienne est définie par : $\frac{D\mathbf{u}}{Dt} = \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u}$;

le terme des forces de gravité $G = -\rho g \mathbf{k}$;

le terme des forces de Coriolis $C = -2\boldsymbol{\Omega} \wedge \mathbf{u}$;

le coefficient de friction μ .

Dans les équations du modèle Shallow-Water, l'océan est vu comme une mince couche d'eau à la surface de la terre (d'où le nom). Cela se traduit par une intégration verticale des équations (2.3) et (2.4). On fait de plus l'hypothèse que le fluide est incompressible (pas de termes en $\frac{\partial \rho}{\partial t}$ est constant), homogène et hydrostatique (c'est à dire que $\partial \mathbf{p} / \partial z = -\rho g$). Ceci nous permet de ne conserver que trois variables, les deux composantes de la vitesse (zonale \mathbf{u} et méridionale \mathbf{v}) ainsi que la hauteur d'eau $\mathbf{h} = -\mathbf{p} / \rho g$. Cette formulation est dite à *gravité réduite* car la suppression de la composante verticale de la vitesse est compensée par une variation du champ gravitationnel.

Les équations obtenues dans un repère cartésien après application de ces hypothèses caractérisent le très classique modèle Shallow Water (e.g. ADCROFT et MARSHALL 1998 [1]) :

$$\begin{cases} \partial_t \mathbf{u} - (f + \zeta) \mathbf{v} + \partial_x B = \frac{\tau}{\rho_0 \mathbf{h}} - r \mathbf{u} + \nu \Delta \mathbf{u} \\ \partial_t \mathbf{v} - (f + \zeta) \mathbf{u} + \partial_y B = \frac{\tau}{\rho_0 \mathbf{h}} - r \mathbf{v} + \nu \Delta \mathbf{v} \\ \partial_t \mathbf{h} + \partial_x (\mathbf{h} \mathbf{u}) + \partial_y (\mathbf{h} \mathbf{v}) = 0 \end{cases} \quad (2.5)$$

où $\zeta = \partial_x \mathbf{v} - \partial_y \mathbf{u}$ est la vorticité relative, $B = g^* \mathbf{h} + \frac{1}{2}(\mathbf{u}^2 + \mathbf{v}^2)$ le potentiel de Bernoulli, g^* le potentiel de gravité réduite, r le coefficient de friction linéaire, ν le coefficient de viscosité, $f = f_0 + \beta y$ le facteur de Coriolis et y la coordonnée méridionale (ordonnée) du repère.

Les dérivées partielles du système ont été approximées au moyen d'un schéma aux différences finies centré d'ordre deux sur une grille définie avec $\Delta i = \Delta j = 25 \text{ km}$. Avec ce type de grille, si l'on évalue les valeurs de \mathbf{u} , \mathbf{v} , et \mathbf{h} au même point on constate numériquement un phénomène de perte de masse. Pour pallier cela, on utilise des grilles identiques pour \mathbf{u} , \mathbf{v} et \mathbf{h} mais décalées les unes par rapport aux autres suivant les axes. Cette grille se nomme grille C d'Arakawa (voir fig 2.5).

Temporellement, un schéma leap-frog (2.6) est utilisé avec $\Delta t = 30 \text{ min}$: si on écrit le modèle comme

$$\frac{d\mathbf{u}}{dt} = F(\mathbf{u}, t)$$

la discrétisation temporelle donnera

$$\mathbf{u}^{t+\Delta t} = \mathbf{u}^{t-\Delta t} + 2\Delta t F(\mathbf{u}_t, t) \quad (2.6)$$

Ce schéma utilisé seul entraîne une dissociation des modes numériques pairs et impairs; c'est pourquoi on lui adjoint un filtre d'Asselin (2.7) afin d'éviter cette dissociation :

$$\mathbf{u}_f^t = \mathbf{u}^t + \gamma(\mathbf{u}_f^{t-\Delta t} - 2\mathbf{u}^t + \mathbf{u}^{t+\Delta t}) \quad (2.7)$$

(l'indice f indique la valeur filtrée)

Notre cas test :

Le forçage du vent τ est fixé constant en temps, sa composante méridionale est nulle et sa composante zonale est sinusoïdale :

$$\tau = \tau_0 \cos\left(\frac{2\pi y}{L}\right) \vec{i} \quad (2.8)$$

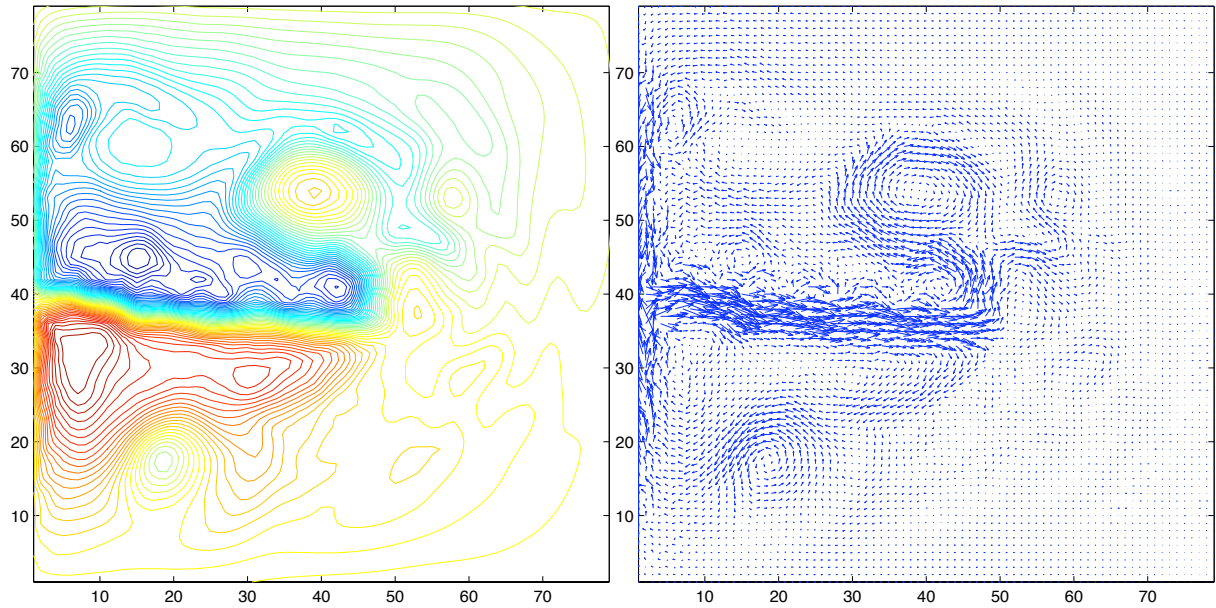


FIG. 2.2: Champs de hauteur d'eau (à gauche) et de vitesse (à droite) pour le modèle *Shallow Water*

L étant la longueur du bassin, dans notre cas, on se contentera d'un bassin de 2000 km de coté avec un fond plat (une grande, très grande, piscine). Pour calibrer le modèle, nous avons fixé :

$$L = 2000 \text{ km}, f_0 = 0.7 \times 10^{-4} \text{ s}^{-1}, \beta = 2 \times 10^{-11} \text{ m}^{-1} \text{ s}^{-1}, \nu = 15 \text{ m}^2 \text{ s}^{-1}, r = 1 \times 10^{-7} \text{ s}^{-1}, \rho_0 = 10^3 \text{ kg m}^{-3}, g^* = 0.02 \text{ m s}^{-2}, \text{ et } \tau_0 = 0.015 \text{ N m}^{-2}.$$

Ces choix confèrent au modèle un comportement proche de celui de l'océan atlantique dans les latitudes moyennes avec un fort courant traversant le bassin (notre gulf-stream à nous), créant deux cellules tourbillonnaires (voir fig(2.2))

Expériences jumelles

Avec ces modèles “jouets” on ne se place pas dans un contexte réaliste mais on cherche un bon cadre d'études pour les méthodes d'assimilation de données. Pour cela il est commode de connaître exactement notre système, c'est à dire de connaître l'état vrai \mathbf{x}^t , l'erreur sur les observations ε^o , sur la prévision η , etc ... Cette connaissance complète nous permettra d'étudier le comportement de chacune des méthodes d'assimilation que nous allons tester. Si ceci n'est évidemment pas possible dans le cadre d'expériences réalistes, cela le devient dans le cadre d'*expériences jumelles*. Pour cela on définit deux modèles : le modèle *vrai* et le modèle *de prévision*. le deuxième sera issu du premier dans lequel on aura effectué quelques modifications de manière à simuler le cas d'un modèle imparfait.

Dans notre cas test, si les paramètres décrits dans la partie précédente représentent le

modèle *de prévision*, le modèle *vrai* est défini par :

$$\nu = 0.9 \times 15m^2s^{-1}, r = 0.9 \times 10^{-7}s^{-1} \text{ et } \tau_{vrai} = \tau_{prev}(1 + 0.8 \times \sin(\frac{2\pi t}{\Delta t \times 480}))$$

les autres paramètres étant identiques.

Le modèle vrai va engendrer la trajectoire vraie \mathbf{x}^t du système. Les observations simulées seront obtenues par une application de l'opérateur d'observation H et l'ajout de l'erreur d'observation qu'on aura défini au préalable :

$$\mathbf{y}^o = H(\mathbf{x}^t) + \varepsilon$$

Cette approche permet, grâce à la connaissance de l'état vrai \mathbf{x}^t , de quantifier avec précision l'impact de l'assimilation de données.

2.2 Le modèle aux équations primitives OPA

Les modèles aux équations primitives sont les plus réalistes mais aussi les plus coûteux en temps de calcul. OPA (Océan PARallélisé) est un modèle aux équations primitives de la circulation océanique régionale et globale. Il a été (et est encore) développé au LODYC (Laboratoire d'Océanographie DYnamique et de Climatologie). Les variables pronostiques sont le champ tri-dimensionnel de vitesse et les caractéristiques thermohalines de l'océan (la température et la salinité).

Les équations primitives découlent de celles de Navier-Stokes (2.3) auxquelles on adjoint une équation d'état liant la température et la salinité aux vitesses de courants.

Dans OPA, les hypothèses suivantes ont été ajoutées :

- l'approximation "terre sphérique" : la surface géopotentielle est supposée sphérique ce qui implique que la gravité est parallèle au rayon de la terre.
- l'approximation "couche mince" : la profondeur de l'océan est supposée négligeable par rapport au rayon de la terre.
- l'hypothèse "fermeture turbulente" : les effets sous-maillages sont paramétrés par un terme prenant en compte les processus d'échelle inférieure à celle de la discrétisation des variables.
- l'approximation de Boussinesq : les variations de densité sont supposées négligeables (sauf pour leurs contributions à la force de flottabilité).
- l'hypothèse hydrostatique : l'équation des moments verticaux est réduite à un équilibre entre le gradient de pression vertical et la force de flottabilité.
- hypothèse incompressible : la divergence de la vitesse est supposée nulle.

Les équations

Sans entrer dans les détails (pour cela voir le manuel de référence d'OPA [82]) les équations résolues par ce modèle sont, dans un repère $(\mathbf{i}, \mathbf{j}, \mathbf{k})$:

(1) l'équilibre des moments (conservation de la quantité de mouvement):

$$\frac{\partial \mathbf{u}_h}{\partial t} = - \left[(\nabla \times \mathbf{u}) \times \mathbf{u} + \frac{1}{2} \nabla (\mathbf{u}^2) \right]_h - f \mathbf{k} \times \mathbf{u}_h - \frac{1}{\rho_0} \nabla_h \mathbf{p} + \mathbf{D}^u \quad (2.9)$$

(2) l'équilibre hydrostatique :

$$\frac{\partial \mathbf{p}}{\partial z} = -\rho g \quad (2.10)$$

(3) l'incompressibilité et conservation de la masse :

$$\nabla \cdot \mathbf{u} = 0 \quad (2.11)$$

(4) la conservation de la chaleur :

$$\frac{\partial \mathbf{T}}{\partial t} = -\nabla \cdot (\mathbf{t}\mathbf{u}) + D^t \quad (2.12)$$

(5) la conservation du sel :

$$\frac{\partial \mathbf{s}}{\partial t} = -\nabla \cdot (\mathbf{s}\mathbf{u}) + D^s \quad (2.13)$$

(6) l'équation d'état :

$$\rho = \rho(\mathbf{T}, \mathbf{s}, \mathbf{p}) \quad (2.14)$$

avec

\mathbf{u} le vecteur vitesse 3D ($\mathbf{u} = \mathbf{u}_h + w\mathbf{k}$, l'indice $_h$ désignant la partie horizontale du vecteur, i.e sur le plan (\mathbf{i}, \mathbf{j}))

\mathbf{T} la température

\mathbf{s} la salinité

t le temps

z la coordonnée verticale

ρ la densité in-situ, ρ_0 une densité de référence

\mathbf{p} la pression

f l'accélération de Coriolis ($f = 2\Omega \cdot \mathbf{k}$, où Ω est la vitesse angulaire de rotation de la Terre)

g la gravité

\mathbf{D}^u , D^t et D^s sont des paramètres permettant la prise en compte des effets "sous-mailles" (cf plus loin)

Les conditions aux bords

Les frontières d'un modèle d'océan sont de multiples natures : les côtes, la topographie du fond de l'océan et les interfaces avec l'air ou la glace sur la surface. À travers chacune de ces frontières l'océan peut échanger des flux de chaleur, d'eau douce, de sel ou de moments avec la croûte terrestre, les continents, la glace ou l'atmosphère. Les caractéristiques de ces échanges peuvent être différents suivant le type de l'interface :

- Interface côte - océan : Le flux principal entre la surface des continents et l'océan est l'apport d'eau douce par les fleuves et rivières (runoff). Cet apport joue un rôle important, notamment à l'embouchure des grands fleuves, en modifiant la salinité de l'eau. Même si il peut être négligé pour les intégrations du modèle à court terme, il est impératif de le prendre en compte pour des prévision à long terme du fait de son influence sur les caractéristiques des masses d'eau qui se formeront.
- Interface fond - océan : à part à quelques endroits très localisés, les flux de chaleur et de sel avec le fond de la mer sont négligeables. Dans les équations d'OPA ils seront toujours négligés. Il n'en sera pas de même pour les moments. On considère qu'il n'y a pas d'échange d'eau avec le fond de l'océan; en d'autres termes, la vitesse au fond est considéré comme étant parallèle à la frontière. Dans les équations du modèle cela se traduit par

$$w = \mathbf{u}_h \cdot \nabla_h(H) \quad (2.15)$$

où w est la composante verticale de la vitesse et H la profondeur du fond de l'océan. Il faut de plus considérer les forces de friction avec le fond de l'océan; celles-ci doivent être considérées en tant que flux turbulent à travers les conditions aux limites de fond ou de bord. Différentes expressions sont possibles; ces paramétrisations sont incluses dans \mathbf{D}^u (cf [82]).

- Interface atmosphère - océan : Les conditions cinématiques de surface plus les flux d'eau douce $P - E$ (Précipitations - Évaporation) donnent la condition :

$$w = \frac{\partial \eta}{\partial t} + \mathbf{u}_h|_{z=\eta} \cdot \nabla_h(\eta) + P - E \quad (2.16)$$

où η est la hauteur du niveau de la mer (pour bien voir la différence entre η et H on peut se référer au dessin 2.3).

En négligeant la tension de surface, les conditions limites dynamiques entraînent la continuité de la pression sur l'interface $z = \eta$. L'atmosphère et l'océan échangent aussi des moments horizontaux (du fait du vent) et de la chaleur. Ces échanges sont simulés par des forçages et/ou par le couplage avec un modèle d'atmosphère.

- Interface glace - océan : Même si elle n'interviendra pas dans notre configuration (on comprendra pourquoi par la suite), l'influence de la glace ne peut être négligée dans le cas général. L'océan et la banquise échangent de l'eau douce, de la chaleur, de la salinité et des moments. La surface de l'eau doit être à température de solidification

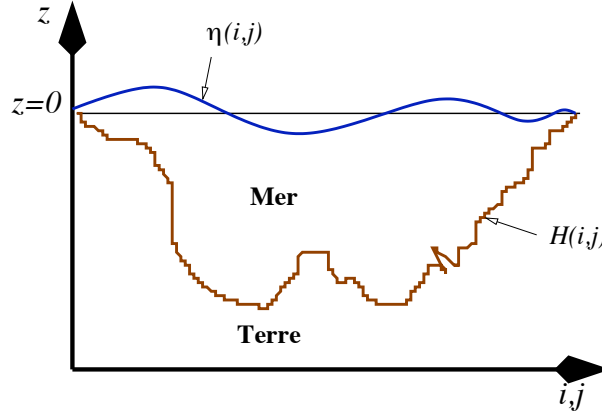


FIG. 2.3: L'océan est délimité par deux surfaces, $z = -H(i, j)$ et $z = \eta(i, j)$, H est la profondeur de l'océan et η est la hauteur de la surface de l'eau (par rapport à $z=0$).

à l'interface. Du fait de la très faible salinité de la glace (environ 4 p.s.u. par rapport aux quelques 34 p.s.u. de l'océan), le cycle glaciation - fonte entraîne d'importants flux d'eau douce et de sel.

- frontières ouvertes : Quand on ne s'intéresse qu'à une zone délimitée de l'océan il est nécessaire de définir les échanges avec l'océan extérieur à cette zone. Pour plus de détails se reporter à la description de la configuration utilisée dans cette thèse.

Le gradient de pression horizontal

La pression à une profondeur donnée z est composée de la pression de surface (à $z = 0$) \mathbf{p}_s plus la pression hydrostatique \mathbf{p}_h .

$$\mathbf{p}(i, j, z, t) = \mathbf{p}_s(i, j, t) + \mathbf{p}_h(i, j, z, t) \quad (2.17)$$

Si \mathbf{p}_h peut être calculée en intégrant (2.10) sur la verticale (et en supposant que dans (2.14) la pression peut être approximée linéairement par rapport à la profondeur)

$$\mathbf{p}_h(i, j, z, t) = \int_{\zeta=z}^{\zeta=0} g\rho(\mathbf{T}, \mathbf{s}, \zeta) d\zeta \quad (2.18)$$

la pression de surface nécessite un traitement spécifique. Deux options sont alors possibles. Dans la première, dite de "surface libre", on introduit η comme variable pronostique du modèle (en rajoutant une équation). Utiliser cette stratégie consiste en une excitation des ondes de gravité externes. L'écoulement est alors barotrope (moyenné sur la verticale) et la surface oscille de bas en haut avec la gravité comme force de rappel. Cependant la vitesse de phase de telles ondes est élevée (quelques centaines de mètres par seconde) ce qui impose, si elle sont présentes dans le modèle, des pas de temps très réduits. La deuxième option, l'approximation "toit rigide", élimine ces ondes en imposant $\eta = 0$ (la surface de l'océan est la surface $z = 0$). D'un autre côté cette approximation augmente la vitesse des ondes de surface à l'infini et altère, entre autres, les ondes de Rossby barotropes.

Au commencement des travaux contenus dans le présent rapport, seule la deuxième approche était codée dans OPA. Pour cette raison c'est celle ci qui sera utilisée. On suppose donc que la surface de l'océan est un toit rigide sur lequel la pression \mathbf{p}_s s'exerce. Il en découle de façon évidente que la composante verticale de la vitesse du courant en surface est nulle. De plus, suivant l'équation de continuité (2.11) et de la condition au fond (2.15) on peut montrer que la moyenne verticale de l'écoulement est de divergence nulle.

$$\operatorname{div}(H\bar{\mathbf{u}}_h) = 0 \quad (2.19)$$

Il conviendra, dans les différents schémas d'assimilation de prendre garde à conserver cette divergence nulle.

En utilisant ceci et en moyennant (2.9), on obtient une expression de la pression de surface (cf [82]) ce qui permet de retrouver la hauteur du niveau de la mer :

$$H = \frac{\mathbf{p}_s}{g\rho_0} \quad (2.20)$$

C'est cette hauteur du niveau de la mer qui pourra être comparée aux mesures altimétriques fournis par les satellites.

Discrétisation

OPA peut fonctionner en utilisant les coordonnées z (figure (2.4)) ou les coordonnées s (l'altitude est relative au fond de l'océan). Comme dans le cas du modèle Shallow-Water, les schémas numériques utilisés pour résoudre les équations numériques sont fondés sur un schéma classique aux différences finies, centré d'ordre 2. On utilise la généralisation tri-dimensionnelle de la grille C de la classification d'Arakawa : l'océan est décomposé en cellules centrées sur les points sur lesquels sont définis \mathbf{T} , \mathbf{s} , \mathbf{p} , ρ et χ et avec \mathbf{u} , \mathbf{v} et \mathbf{w} définis au centre de chaque face (fig (2.5)).

Pour la discrétisation temporelle un schéma "saute-mouton" (leap-frog : équation (2.6)) est utilisé auquel on rajoute comme dans le modèle Shallow-Water, un filtre d'Asselin (2.7).

Les échelles spatiales caractéristiques des nombreux processus composant les écoulements océaniques sont très variées allant des distances propres à la diffusion moléculaire à des longueurs d'onde de l'ordre du millier de kilomètres. La discrétisation spatiale ne permettant pas de représenter toutes ces échelles, on est amené à paramétrer les effets des processus "sous-mailles". Ces paramétrisations sont des descriptions empiriques qui adoptent des formalismes de type "diffusion" (les termes D^x dans les équations primitives).

Notre configuration : fenêtre pacifique tropicale (TDH)

Pour la validation des méthodes introduites dans cette thèse, nous nous intéresserons particulièrement à la restriction d'OPA au domaine couvrant l'océan Pacifique entre les deux tropiques. Cette fenêtre s'étend approximativement entre 120° Est et 70° Ouest et entre 30° Nord et 30° Sud. Les équations du modèle sont résolues sur une grille horizontale isotrope. La résolution méridienne est de 1/2° à l'équateur et décroît en latitude, pour atteindre 2° aux frontières Nord et Sud. La résolution zonale est de 1° sur l'ensemble du bassin. Sur la

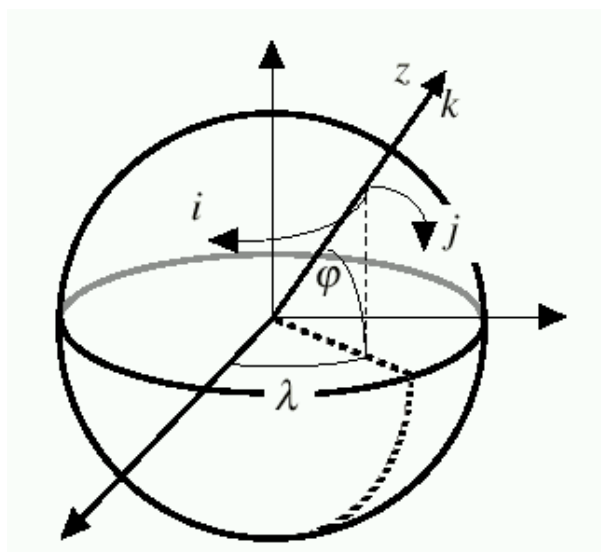


FIG. 2.4: le système de coordonnées polaires (λ, ϕ, z) et son équivalent en coordonnées z (i, j, k)

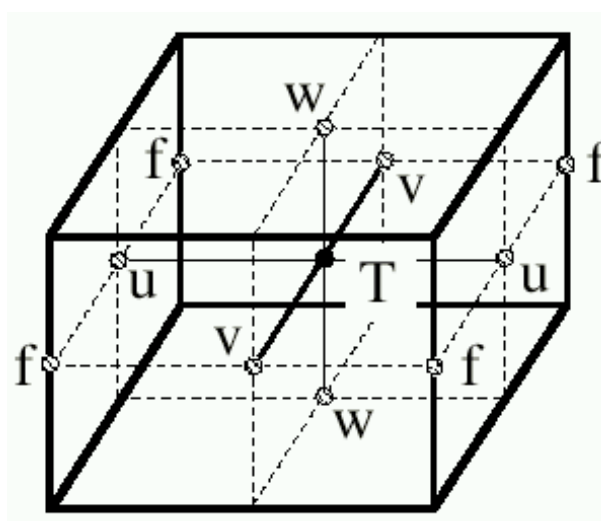


FIG. 2.5: Grille C d'Arakawa : T indique les points où sont définis la température, la salinité, la densité la pression et la divergence horizontale, f indique les points où les vorticités relative et planétaires sont définies, u, v, w sont explicites

verticale le modèle est décomposé en 25 niveaux avec une résolution allant de 10 m pour les couches supérieures à 150 m dans les profondeurs. La bathymétrie a été interpolée du masque des données de l'atlas de LEVITUS. Le pas de temps est d'une heure et demie. Au delà de 20° Nord et 20° Sud existent des zones "tampon" dans lesquelles les champs de salinité et de température sont rappelés vers leurs valeurs climatologiques (résolution des frontières ouvertes). Des conditions de non glissement et d'absence de flux de chaleur et de

sel sont appliquées au fond de l'océan et le long des côtes. Un rappel à la SST (Sea Surface Temperature : température de surface de l'océan) est appliqué sur l'ensemble du bassin. Un schéma pronostique TKE (Turbulent Kinetic Energy : énergie cinétique turbulente) est utilisé pour résoudre le mélange vertical. Les conditions aux limites de ce modèle incluent les flux de moments, de chaleur et de salinité à la surface. Dans notre cas-test, les flux de moments sont calculés à partir des tensions de vent fournies par une combinaison des données des satellites ERS 1 et 2. Les flux de chaleurs sont estimés au moyen des moyennes saisonnières de SST de Reynolds et des données climatologiques journalières de SST issues des réanalyses ECMWF (VIALARD et al , 2001 [122]). Un terme de correction constant est appliqué à l'apport d'eau douce (précipitation moins évaporation, pas de *runoff*). Ce terme est calculé de façon à éviter que la SSS (Sea Surface Salinity : salinité de surface de l'océan) ne dérive hors de la barre d'erreur fixée a priori.

Ce choix a été dicté par la nécessité de disposer d'un modèle à la fois réaliste et pas trop difficile à manipuler (l'océan global est trop coûteux en calcul); de plus la dynamique dans le pacifique tropical est moins non linéaire que dans d'autres parties de l'océan ce qui amoindrira les difficultés pour l'assimilation; et pour finir il existe déjà de nombreuses expériences d'assimilation sur cette zone et ce modèle ce qui permettra de multiples comparaisons avec les méthodes introduites dans cette thèse.

Les données utilisées : Le réseau TAO et les profileurs XBT

- TAO :

Le réseau de bouées TAO (pour Tropical Atmosphere Ocean) est composé de deux types de bouées : 64 ATLAS (Autonomous Temperature Line Acquisition System) et 5 PROTEUS (Profile Telemetry of Upper Ocean Currents). Ces deux types fournissent en (quasi) temps réel des mesures du vent de surface, de la température et de l'humidité relative de l'air , de la température de surface et de la température de sub-surface jusqu'à 500m. Les bouées PROTEUS fournissent en plus la mesure des courants océaniques dans les premiers 300m. Les données de surface sont retransmises via un satellite (3 ou 4 fois par jour) tandis que celles de sub-surface sont transmises comme moyenne journalière de la veille. La zone couverte par les bouées TAO s'étend de 130° Est à 90° Ouest et de 10° Nord à 10° Sud (cf Fig. 2.7).

Pour les expériences d'assimilation seules les données de température seront utilisées. L'incertitude sur de telles données est de l'ordre de 0,1° K

- XBT :

L'autre composante de notre système d'observation est le réseau XBT (eXpandable BathyThermographs). Ces sondes sont larguées à partir de bateaux commerciaux volontaires. Elles fournissent elles aussi un profil de la température. La zone couverte par ces mesures est nettement plus étendue que la zone TAO (voir 2.7) mais ces données sont aussi beaucoup plus asynoptiques (distribuées de façon irrégulière).

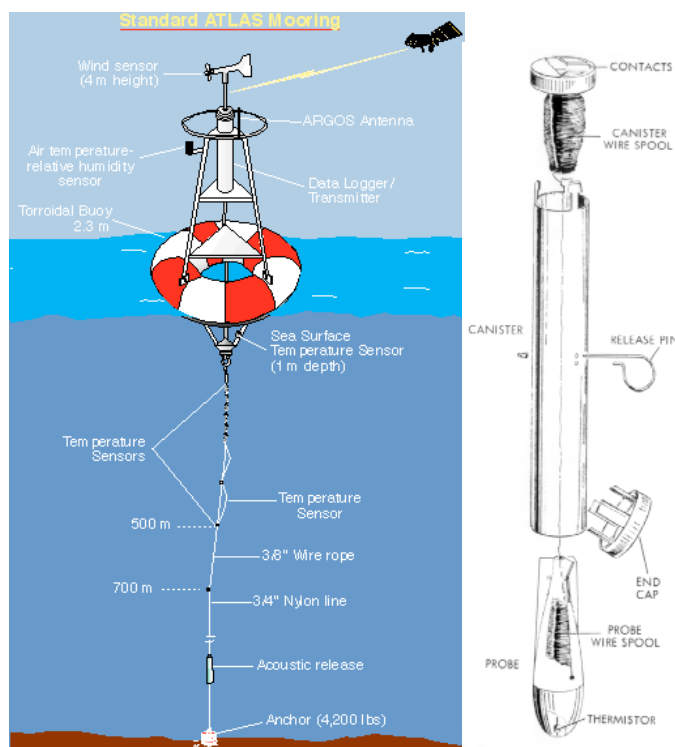


FIG. 2.6: Bouée de mesure TAO (à gauche) et sonde XBT (à droite)

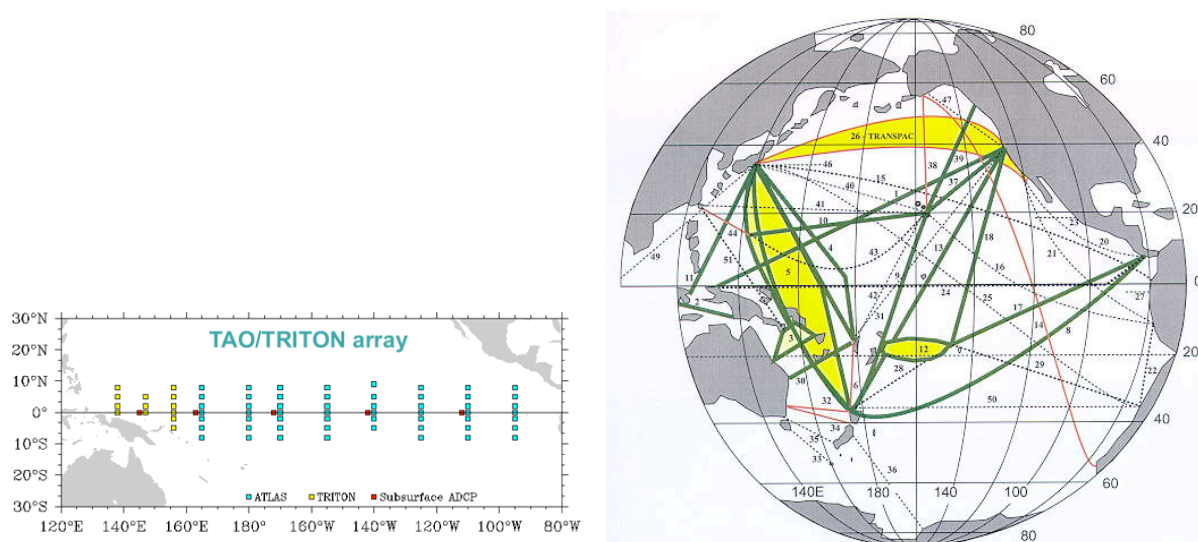


FIG. 2.7: Réseau TAO (à gauche) et XBT (à droite) d'observation dans le pacifique

Des données indépendantes pour valider l'assimilation

Comme on l'a vu dans le premier chapitre nous n'avons pas de connaissance exacte de l'état réel de l'océan. Le but de l'assimilation est de recomposer de façon acceptable cet

état réel. Il se pose alors, du fait de cette méconnaissance, le problème d'estimation de la qualité de l'état analysé produit par l'assimilation de données. Si un simple regard sur la physionomie de ces champs permet d'éliminer les résultats les plus farfelus ce ne peut être considéré comme suffisant. La comparaison avec les données assimilées peut s'avérer utile mais on comprend bien en considérant le cas extrême d'un seul thermomètre perdu dans l'océan que l'assimilation de cette donnée pourra fournir un état très éloigné de l'état réel mais avec une température au point du thermomètre très proche de celle observée. D'une façon plus générale la comparaison aux données assimilées est biaisée car on a tout mis en œuvre dans la méthode d'assimilation pour que le modèle se rapproche de ses valeurs.

C'est pourquoi, en cas d'expérience avec des données réelles il est nécessaire de disposer d'un jeu de données indépendantes de celles utilisées pour l'assimilation, et de préférence apportant des informations sur des variables du modèle non observées.

Dans notre configuration on utilisera les données issues des courantomètres des bouées TAO ainsi que les mesures des anomalies de hauteur de la mer fournies par le satellite TOPEX/POSEIDON. Celles ci sont fournies sous la forme de cartes surfaciques tous les 10 jours. Typiquement l'erreur sur de telles données est de l'ordre de 2 cm.

Chapitre 3

Techniques algorithmiques

Dans ce chapitre différents aspects de l'assimilation variationnelle seront traités du très technique codage des opérateurs adjoints au plus qualitatif problème de préconditionnement de la minimisation. Ces aspects ont une place prépondérante dans la mise en œuvre de l'assimilation de données et afin de ne pas les “noyer” dans les descriptions plus théoriques des algorithmes d'assimilation, ils ont été regroupés dans ce chapitre.

3.1 Les opérateurs adjoints

En ordonnant de façon adéquate les opérations algébriques on peut réduire la taille et le nombre des produits de matrices, ce qui permet de diminuer le nombre de calculs et donc le coût numérique. En particulier dans les problèmes de minimisation, quand il faut évaluer la dérivée d'une fonction scalaire (par exemple le terme J_o du 4D-Var) par rapport à un vecteur de taille importante, il est avantageux d'aborder le problème de façon *rétrograde*, c'est à dire depuis la fonction scalaire vers le vecteur de contrôle. Algébriquement cela correspond à remplacer un ensemble d'opérateurs par leurs adjoints (d'où le nom *méthodes adjointes*).

Par définition, si on a \mathbf{A} un opérateur linéaire d'un espace \mathcal{E} dans un espace \mathcal{F} et les produits scalaires associés $(\cdot, \cdot)_{\mathcal{E}}$ et $(\cdot, \cdot)_{\mathcal{F}}$, l'adjoint de \mathbf{A} est l'opérateur linéaire \mathbf{A}^* tel que

$$(\mathbf{A}\mathbf{x}, \mathbf{y})_{\mathcal{F}} = (\mathbf{x}, \mathbf{A}^*\mathbf{y})_{\mathcal{E}} \quad (3.1)$$

Comme on se place dans le cas d'espaces de dimension finie, $\mathcal{E}, (\cdot, \cdot)_{\mathcal{E}}$ et $\mathcal{F}, (\cdot, \cdot)_{\mathcal{F}}$ sont des espaces de Hilbert, le théorème de Riesz nous donne alors l'existence et l'unicité de \mathbf{A}^* . De plus si $(\cdot, \cdot)_{\mathcal{E}}$ et $(\cdot, \cdot)_{\mathcal{F}}$ sont les produits scalaires euclidiens sur \mathcal{E} et \mathcal{F} alors $\mathbf{A}^* = \mathbf{A}^T$. Dans cette thèse on se placera toujours dans ce cas pour la dérivation des codes.

Codage

La principale difficulté de l'utilisation des méthodes adjointes réside dans le codage des opérateurs adjoints, celui-ci peut être obtenu directement à partir de l'opérateur direct (CHAO et CHANG [23]). La première étape est, si l'opérateur direct est non linéaire, de coder ses dérivées de Gateaux. Par exemple si on a dans le code direct

$$\mathbf{w} = a\mathbf{w} + b\mathbf{x}\cdot\mathbf{y}$$

avec a et b des constantes et \mathbf{w} , \mathbf{x} , \mathbf{y} des variables, alors la ligne correspondante dans le code linéaire tangent sera

$$\delta\mathbf{w} = a\delta\mathbf{w} + b(\delta\mathbf{x}\cdot\mathbf{y} + \mathbf{x}\cdot\delta\mathbf{y})$$

où $\delta\mathbf{w}$, $\delta\mathbf{x}$ et $\delta\mathbf{y}$ représentent les variables tangentées (i.e les perturbations appliquées aux variables directes), et \mathbf{x} et \mathbf{y} sont les variables directes enregistrées lors de l'application de l'opérateur direct.

Comme il a été dit plus haut coder l'adjoint d'un opérateur correspond à effectuer la transposition de cet opérateur. En raison de la taille des problèmes qui nous intéressent ici il n'est pas concevable de stocker et manipuler les matrices représentant les différents opérateurs. En océanographie ou en météorologie l'approche la plus commune pour obtenir le code de l'adjoint est d'appliquer la transposition à chacune des étapes élémentaires du code direct. La plupart du temps il y a un morceau de code adjoint pour chaque instruction du code direct, celles-ci étant considérées comme des opérateurs linéaires élémentaires. Cette approche permet de travailler localement dans le code, et l'adjoint d'une séquence d'opération est la séquence des transposées de chaque opération dans l'ordre inverse (BOUTTIER ET COURTIER, 1999[17]).

Typiquement une étape élémentaire d'un code linéaire tangent peut s'écrire

$$\mathbf{w} = a\mathbf{x} + b\mathbf{y} + c\mathbf{z} \quad (3.2)$$

où \mathbf{w} , \mathbf{x} , \mathbf{y} et \mathbf{z} sont des variables (perturbations) et a , b et c sont soit des constantes soit des variables dépendantes du résultat de l'application de l'opérateur direct. Cette étape peut être considérée comme une opération linéaire de la forme $\mathbf{u} = \mathbf{M}\mathbf{v}$

$$\begin{pmatrix} \mathbf{w} \\ \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} a & b & c \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix}$$

l'adjoint de cette étape peut donc être obtenu par une transposition formelle :

$$\begin{pmatrix} \mathbf{x}^* \\ \mathbf{y}^* \\ \mathbf{z}^* \end{pmatrix} = \begin{pmatrix} a & 1 & 0 & 0 \\ b & 0 & 1 & 0 \\ c & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{w}^* \\ \mathbf{x}^* \\ \mathbf{y}^* \\ \mathbf{z}^* \end{pmatrix}$$

on a donc l'étape adjointe de (3.2)

$$\begin{cases} \mathbf{x}^* = a\mathbf{w}^* + \mathbf{x}^* \\ \mathbf{y}^* = b\mathbf{w}^* + \mathbf{y}^* \\ \mathbf{z}^* = c\mathbf{w}^* + \mathbf{z}^* \end{cases}$$

où \mathbf{w}^* , \mathbf{x}^* , \mathbf{y}^* et \mathbf{z}^* sont les gradients de l'opérateur direct par rapport aux variables directes \mathbf{w} , \mathbf{x} , \mathbf{y} et \mathbf{z} respectivement.

Afin de faciliter la tâche de l'assimilateur, des dérivateurs automatiques de codes adjoints ont été développés comme par exemple Odyssee à l'INRIA ([47]).

Dans le cas où l'opérateur direct est le modèle et que celui-ci est non linéaire, il sera nécessaire de stocker la trajectoire directe dite trajectoire de référence. En effet celle-ci apparaît dans les équations des modèles tangents et adjoints. De plus le modèle adjoint étant intégré de façon rétrograde (du pas de temps final au pas de temps initial) il est nécessaire de conserver la totalité de cette trajectoire.

3.2 Méthodes de minimisation

Les méthodes variationnelles reposent sur la minimisation d'une fonctionnelle. Pour la minimiser on utilise généralement un algorithme itératif de descente.

On considère qu'on est dans le cas d'un problème de minimisation sans contrainte de la fonction objective $J(\mathbf{x})$ de \mathbb{R}^n dans \mathbb{R} . On suppose que pour tout \mathbf{x}^k on est capable de déterminer $\nabla_{\mathbf{x}}J(\mathbf{x}^k)$ le gradient de la fonction objective par rapport au vecteur de contrôle au point \mathbf{x}^k .

Pour déterminer un minimum \mathbf{x}^* une méthode de descente procédera par descentes successives de la forme

$$\mathbf{u}^{k+1} = \mathbf{x}^k - \rho^k G_k \mathbf{d}^k$$

où $\mathbf{d}^k \in \mathbb{R}^n$ est la direction de descente obtenue à partir du gradient $\nabla_{\mathbf{x}}J(\mathbf{x}^k)$ et $\rho^k \in \mathbb{R}$ est le pas de descente.

Le coût de l'assimilation de données est proportionnel au nombre d'évaluations de la fonction objective et de son gradient (ce nombre est appelé nombre de *simulations*). Quand le vecteur de contrôle est réactualisé, on parle alors d'une *itération*. Chaque *itération* peut requérir une ou plusieurs *simulations* suivant la méthode de minimisation utilisée. On appelle *simulateur* l'opérateur qui au vecteur de contrôle \mathbf{x} associe la valeur de la fonction objective en ce point et son gradient

$$\mathbf{x} \mapsto (J(\mathbf{x}), \nabla J(\mathbf{x}))$$

Dans les méthodes variationnelles présentées au premier chapitre la fonction objective est une fonction scalaire d'un vecteur réel dans un espace euclidien; dans la plupart des cas elle est quadratique et il n'y a pas de contrainte sur le vecteur de contrôle \mathbf{x} . Il existe de nombreux algorithmes "prêts à l'emploi" capables d'effectuer ce genre de minimisation (on parle alors de *minimiseurs*). Parmi ceux-ci on peut citer le très simple mais généralement peu efficace algorithme de plus grande pente qui vise à actualiser \mathbf{x} en ajoutant une correction proportionnelle à $-\nabla J(\mathbf{x})$ ou les plus populaires Gradients conjugués et méthodes de quasi-Newton (qui seront utilisés dans cette thèse).

Gradient conjugué

Le principe de cette méthode est de choisir une $k^{\text{ème}}$ direction de descente \mathbf{d}^k qui soit une projection du gradient $\nabla_{\mathbf{x}}J(\mathbf{x}^k)$ sur le sous-espace de \mathbb{R}^n orthogonal aux \mathbf{d}^i pour $i = 0, 1, \dots, k-1$.

Cet algorithme s'adresse particulièrement aux problèmes de minimisation sans contrainte de fonction objective quadratique,

$$J = \frac{1}{2}(\mathbf{x}, \mathbf{Ax}) + (\mathbf{b}, \mathbf{x})$$

où \mathbf{A} est une matrice $n \times n$ définie positive et $\mathbf{b} \in \mathbb{R}$. La méthode calcule le pas de descente ρ^k pour chaque direction ce qui implique que pour une fonction J quadratique la méthode converge théoriquement en au plus n itérations.

Une itération du gradient conjugué est donnée par :

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \rho^k \mathbf{d}^k \mathbf{d}^{k+1} = -\nabla_{\mathbf{x}} J(\mathbf{x}^{k+1}) + \beta^k \mathbf{d}^k \quad (3.3)$$

avec

$$\rho^k = \frac{\langle \nabla_{\mathbf{x}} J(\mathbf{x}^k), \mathbf{d}^k \rangle}{\langle \mathbf{d}^k, \mathbf{Ad}^k \rangle}, \quad \beta^k = \frac{\langle \nabla_{\mathbf{x}} J(\mathbf{x}^{k+1}), \mathbf{Ad}^k \rangle}{\langle \mathbf{d}^k, \mathbf{Ad}^k \rangle}$$

et

$$\mathbf{d}^0 = \nabla_{\mathbf{x}} J(\mathbf{x}^0)$$

Dans les applications qui nous intéressent, le calcul de ρ_k est très coûteux car il demande plusieurs intégrations du modèle.

Quasi-Newton

Une étape d'un algorithme Quasi-Newton appliqué à notre problème s'écrit :

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \rho^k G_k \nabla_{\mathbf{x}} J(\mathbf{x}^k)$$

où G_k approxime $\mathcal{H}_{\mathbf{x}^k}^{-1}$, l'inverse du Hessien au point \mathbf{u}^k .

Nous avons utilisé le minimiseur M1QN3 développé à l'INRIA. Celui-ci utilise une méthode Quasi-Newton dans laquelle G_k est calculé suivant une formule BFGS à mémoire limitée (GILBERT et LEMARÉCHAL [52])

La principale caractéristique de J qui peut affecter l'efficacité de la minimisation (si on suppose que J est quadratique ou presque) est le conditionnement de son Hessien. Cette quantité mesure l'ellipticité des iso-surfaces de la fonction objective et cela décrit la difficulté due au fait que le gradient ne pointe pas exactement dans la direction de l'optimum. Dans ce cas le minimiseur a des difficultés à converger. Ce phénomène est appelé effet *narrow valley* (vallée étroite) (CIARLET, 1982 [25])

Le conditionnement de la minimisation est défini par le rapport entre la plus grande et la plus petite valeur propre du Hessien de J . Plus ce nombre est grand plus le problème est mal conditionné.

Dans le cas extrême d'un conditionnement égal à 1 (le Hessien est proportionnel à l'identité) la fonction objective est dite *sphérique* et l'optimum est trouvé en seulement une itération ($-\nabla J$ pointe directement vers le minimum).

Quand J est elliptique il est possible, au moyen d'un changement de variable (appelé *préconditionnement*) de réduire le conditionnement (voir section suivante).

3.3 Conditionnement et adimensionalisation

Pour que la minimisation des fonctions objectives de l'assimilation de données variationnelle soit efficace, il convient de prendre en compte les problèmes d'unité des différentes variables. Pour ne pas alourdir les notations plaçons nous dans le cas linéaire (modèle et opérateur d'observation). Si on reprend la formulation (1.35) :

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x})^T \mathbf{B}^{-1} \delta\mathbf{x} + \frac{1}{2}[\mathbf{y}^o - \mathbf{G}\mathbf{x}]^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{G}\mathbf{x}]$$

et si on exprime les composantes du vecteur de contrôle dans les unités $(u_1, \dots, u_n) = U$, alors la matrice \mathbf{B} est exprimée en unité $U^T U$ et le gradient

$$\nabla J(\delta\mathbf{x}) = \underbrace{\mathbf{B}^{-1}}_{\text{unité}=U^{-1}U^{-T}} \underbrace{\delta\mathbf{x}}_{\text{unité}=U} + \underbrace{\mathbf{G}^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{G}]}_{\text{unité}=U^{-1}U^{-T}} \underbrace{\delta\mathbf{x}}_{\text{unité}=U} \quad (3.4)$$

est donc exprimé en U^{-1} .

Lors de la minimisation on est donc amené à effectuer des opérations du type

$$\underbrace{\delta\mathbf{x}_{i+1}}_{\text{unité}=U} = \underbrace{\delta\mathbf{x}_i}_{\text{unité}=U} + \lambda \underbrace{\nabla J(\delta\mathbf{x}_i)}_{\text{unité}=U^{-1}} \quad (3.5)$$

où \mathbf{x}_i est le $i^{\text{ème}}$ itéré de la minimisation et λ un réel déterminé par le minimiseur. L'équation (3.5) met en évidence une inconsistance au niveau des unités qui est intrinsèque à ces méthodes. Sans aboutir à des résultats faux, cette inconsistance rend le calcul du gradient et donc l'efficacité de la minimisation dépendante du système d'unité utilisé pour représenter $\delta\mathbf{x}$. Une manière de s'affranchir de cette inconsistance est d'effectuer un préconditionnement, c'est à dire un changement de variable dans l'espace de minimisation, permettant d'obtenir un vecteur de contrôle sans dimension.

D'autre part, si $\delta\mathbf{x}^{\text{opt}}$ reste petit, on peut écrire par un développement de Taylor au 1^{er} ordre de ∇J :

$$\nabla J(\delta\mathbf{x}^{\text{opt}}) \approx \nabla J(0) + \mathcal{H} \delta\mathbf{x}^{\text{opt}} \quad (3.6)$$

où \mathcal{H} est la hessienne de J par rapport à $\delta\mathbf{x}$. Comme de plus $\nabla J(\delta\mathbf{x}^{\text{opt}}) = 0$, on a

$$\delta\mathbf{x}^{\text{opt}} \approx -\mathcal{H}^{-1} \nabla J(0) \quad (3.7)$$

L'équation (3.7) nous montre que la minimisation peut se traduire par l'inversion de la hessienne. Comme dans la section précédente cela montre que pour plus d'efficacité, il est nécessaire que la hessienne de la fonction objective soit bien conditionnée.

Or, la hessienne de la fonction objective de l'assimilation variationnelle est égale à l'inverse de la matrice de covariance d'erreur d'analyse

$$\mathcal{H}^{-1} = \mathbf{P}^a = [\mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}]^{-1} \quad (3.8)$$

Démonstration :

La hessienne est obtenu en différenciant deux fois J par rapport à la variable de contrôle \mathbf{x}

$$\begin{aligned} J(\delta\mathbf{x}) &= \frac{1}{2}(\delta\mathbf{x})^T \mathbf{B}^{-1} \delta\mathbf{x} + \frac{1}{2}[\mathbf{y}^o - \mathbf{G}\mathbf{x}]^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{G}\mathbf{x}] \\ \nabla J &= \mathbf{B}^{-1} \delta\mathbf{x} + \mathbf{G}^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{G}\mathbf{x}] \\ \mathcal{H} &= \mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \end{aligned} \quad (3.9)$$

En utilisant à nouveau le fait que $\nabla J(\delta\mathbf{x}^a) = 0$, en insérant l'état vrai \mathbf{x}^t et en écrivant $\delta\mathbf{x}^a = \mathbf{x}^a - \mathbf{x}^t$ dans les équations :

$$\begin{aligned} 0 &= \mathbf{B}^{-1} \delta\mathbf{x}^a - \mathbf{G}^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{G}\mathbf{x}^a] \\ &= \mathbf{B}^{-1} [\mathbf{x}^a - \mathbf{x}^t + \mathbf{x}^t - \mathbf{x}^b] - \mathbf{G}^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{G}\mathbf{x}^t + \mathbf{G}[\mathbf{x}^t - \mathbf{x}^a]] \\ &= \mathbf{B}^{-1} [\mathbf{x}^a - \mathbf{x}^t] - \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} [\mathbf{x}^t - \mathbf{x}^a] - \mathbf{B}^{-1} [\mathbf{x}^b - \mathbf{x}^t] - \mathbf{G}^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{G}\mathbf{x}^t] \end{aligned} \quad (3.10)$$

Ainsi

$$[\mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}] [\mathbf{x}^a - \mathbf{x}^t] = \mathbf{B}^{-1} [\mathbf{x}^b - \mathbf{x}^t] + \mathbf{G}^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{G}\mathbf{x}^t] \quad (3.11)$$

On multiplie à droite 3.11 par sa transposée et on prend la moyenne du résultat. Le terme de droite contient alors deux termes facteurs de $(\mathbf{x}^b - \mathbf{x}^t)(\mathbf{y}^o - \mathbf{G}\mathbf{x}^t)^T$ qui est nul car on a supposé que les erreurs de *background* et d'observation étaient non corrélées. Les termes restants donnent :

$$\begin{aligned} [\mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}] \mathbf{P}^a [\mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}] &= \mathbf{B}^{-1} \mathbf{B} \mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{R} \mathbf{R}^{-1} \mathbf{G} \\ &= \mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \end{aligned}$$

On a donc bien :

$$\mathbf{P}^a = [\mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}]^{-1} = \mathcal{H}^{-1} \quad (3.12)$$

■

Le préconditionnement idéal étant l'inverse de la racine carrée de la hessienne, on aimerait pouvoir utiliser cette équation pour estimer $\mathcal{H}^{-1/2}$; cependant le terme $\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}$ est difficile à estimer et à manipuler. En pratique on fait l'hypothèse un peu brutale $\mathcal{H} = \mathbf{B}^{-1}$ et afin d'obtenir un meilleur conditionnement de la hessienne et par la même occasion d'adimensionnaliser le problème on effectue le préconditionnement :

$$\mathbf{v} = \mathbf{B}^{-\frac{1}{2}} \delta\mathbf{x} \quad (3.13)$$

où \mathbf{v} , la nouvelle variable de contrôle est clairement sans unité. En pratique c'est l'inverse du changement de variable (3.13) qui est effectué (on a besoin de $\delta\mathbf{x}$ pour intégrer le modèle) et on retrouve facilement le gradient de la fonction objective par rapport à $\delta\mathbf{x}$ par

$$\nabla J_{\delta\mathbf{x}} = \mathbf{B}^T \nabla J_{\mathbf{v}} \quad (3.14)$$

Par ce changement de variable on obtient la hessienne

$$\mathcal{H}_{\mathbf{v}} = \mathbf{B}^{\frac{T}{2}} \mathcal{H}_{\delta\mathbf{x}} \mathbf{B}^{\frac{1}{2}} = Id + \mathbf{B}^{\frac{T}{2}} \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \mathbf{B}^{\frac{1}{2}}$$

avec $\mathbf{B}^{\frac{T}{2}} \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \mathbf{B}^{\frac{1}{2}}$ symétrique définie positive. Toutes les valeurs propres sont augmentées de 1 et le rapport entre la plus haute et la plus basse valeur propre est sensiblement réduit (COURTIER, 1997 [28]).

Pour illustrer l'importance de la hessienne dans l'efficacité et la qualité de l'analyse, plaçons nous dans le cas d'un problème unidimensionnel. Dans ce cas la hessienne est la dérivée seconde (c'est à dire la convexité) de la fonction objective de l'assimilation de données variationnelle. La partie supérieure de la figure 3.1 représente deux exemples de fonction objective : la première (à gauche) avec une forte convexité et la seconde (à droite) avec un plus faible convexité. On définit $p(\mathbf{x})$ comme étant la probabilité que l'état analysé obtenu par l'assimilation soit le bon. La hessienne est une mesure de l'écart type de la fonction de densité de $p(\mathbf{x})$ (partie inférieure de la figure 3.1). Plus la fonction de probabilité de l'analyse est "étroite", plus la probabilité que l'état estimé est le bon est haute c'est à dire plus l'analyse est fine (BOUTTIER ET COURTIER, 1999[17]).

Remarque

On ne peut effectuer les raisonnements de cette partie que si l'opérateur d'observation H et le modèle M sont différentiables. Dans ce cas la fonction objective J est quadratique et \mathcal{H} ne dépend pas de l'état analysé (ni du *background* et des observations). \mathbf{P}^a peut alors être défini à partir de J avant même que l'analyse soit effectuée. Si la linéarisation n'est pas exacte, \mathcal{H} n'est pas constant et peut dépendre fortement du vecteur de contrôle. Par exemple si J est continuellement différentiable mais pas strictement convexe, il y aura des points pour lesquels $\mathcal{H} = 0$. Si J n'est pas quadratique et si ∇J n'est pas continue, il y aura même des points pour lesquels \mathcal{H} ne sera pas définie. Il sera alors impossible de calculer \mathbf{P}^a à partir de la hessienne. En pratique J est modifié de façon à utiliser les opérateurs linéaires tangents de H et M valables dans un proche voisinage de $\delta\mathbf{x}^a$ (c'est la variante 4D-Var incrémental).

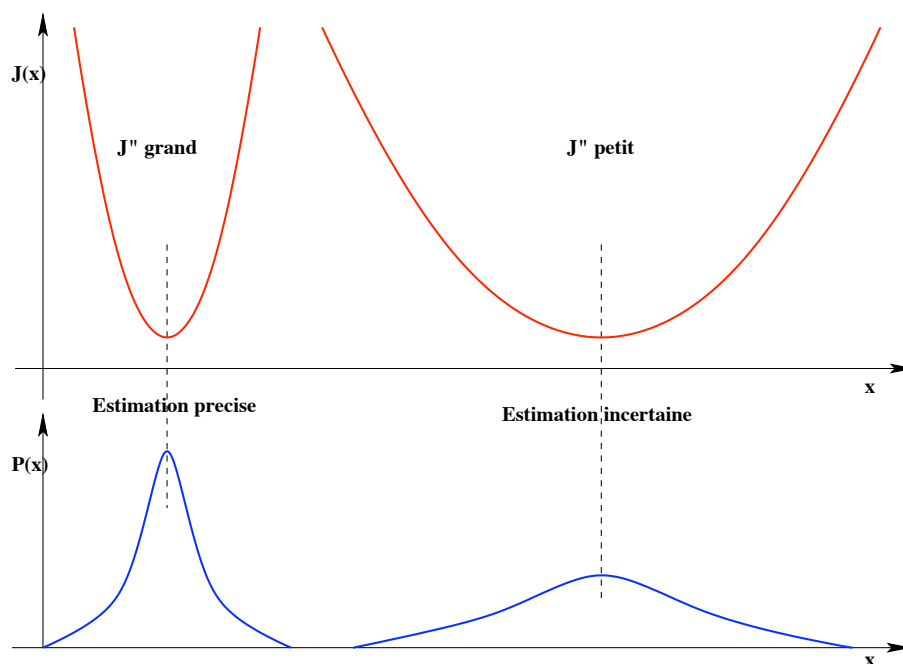


FIG. 3.1: Relation entre la hessienne et la qualité de l'analyse

3.4 Les matrices de covariance d'erreur en assimilation variationnelle

Si l'estimation, le stockage et la manipulation de la matrice \mathbf{R} de covariance d'erreur d'observation ne posent généralement aucune difficulté en raison de sa forme habituellement supposée diagonale, les autres matrices de covariance d'erreur ne peuvent, dans des configurations réalistes offrir de telles facilités et nécessitent un traitement particulier. En effet en raison de la taille du système (typiquement $10^6 - 10^7$ en océanographie) il paraît impossible de stocker et manipuler des matrices de covariance d'erreur complètes; de plus nous n'avons qu'une connaissance très limitée des covariances d'erreur réelles et pouvons donc de ce fait difficilement calculer cette matrice.

Dans cette partie nous décrirons plus précisément une construction de la matrice \mathbf{B} de covariance d'erreur d'*ébauche*, mais de telles méthodes peuvent être appliquées pour approximer les autres matrices de covariance d'erreur.

En pratique et du fait de l'utilisation d'une méthode de gradient ce n'est pas la matrice \mathbf{B} qui nous intéresse mais plutôt l'opérateur qui au vecteur de contrôle \mathbf{x} associe le produit de ce vecteur par \mathbf{B}

$$\mathbf{x} \mapsto \mathbf{B}\mathbf{x} \quad (3.15)$$

Par abus de notation, nous appellerons \mathbf{B} cet opérateur

3.4.1 Construction de l'opérateur \mathbf{B}

Si on se place dans le cas du modèle Shallow-Water on a le vecteur d'état \mathbf{x} composé de trois variables \mathbf{u} , \mathbf{v} , \mathbf{h} . On peut décomposer l'opérateur \mathbf{B} en 9 blocs

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{uu} & \mathbf{B}_{uv} & \mathbf{B}_{uh} \\ \mathbf{B}_{vu} & \mathbf{B}_{vv} & \mathbf{B}_{vh} \\ \mathbf{B}_{hu} & \mathbf{B}_{hv} & \mathbf{B}_{hh} \end{pmatrix} \quad (3.16)$$

où les blocs diagonaux \mathbf{B}_{uu} , \mathbf{B}_{vv} , \mathbf{B}_{hh} représentent les covariances spatiales de chaque variable par rapport à elle même et les autres termes représentent les covariances entre les différentes variables.

Les blocs diagonaux

Dans un premier temps on ne s'intéresse qu'aux blocs diagonaux : pour chaque variable on peut estimer et stocker les variances d'erreur mais du fait de la taille du problème le calcul et le stockage des covariances spatiales reste hors de portée. La première étape sera donc de modéliser les corrélations spatiales pour chaque variable du vecteur d'état. Pour satisfaire aux hypothèses du chapitre 1 on cherche à obtenir des fonctions de corrélation homogènes et isotropes.

Comme ce que l'on cherche à modéliser c'est l'opérateur \mathbf{B} , cela pourra être obtenu par la résolution d'une équation de diffusion (WEAVER et COURTIER, 2000 [127]) avec le vecteur d'entrée comme condition initiale

$$\frac{\partial \eta}{\partial t} - \kappa \nabla^2 \eta = 0$$

avec ∇^2 le Laplacien sur la sphère et κ le coefficient de diffusion. En appliquant une transformation de Fourier pour intégrer l'équation, on obtient pour solution une fonction, homogène et isotrope approximant une gaussienne ce qui définira une fonction de corrélation valide (définie positive) modulo une renormalisation.

Ce filtre Laplacien (qu'on appellera \mathbf{L}) diffuse l'information sur la sphère (voir la figure 3.2) et on multiplie le résultat par une matrice Λ de normalisation qui a pour but de reconvertir le résultat du filtre en matrice de corrélation (avoir 1 sur les diagonales). Les coefficients de Λ sont calculés en appliquant \mathbf{L} à un dirac sur chaque point de grille.

On peut donc imaginer pour quantifier l'influence de chaque point de grille sur les autres, au dessus de ce point de grille une fonction Gaussienne ayant pour hauteur la variance du point et pour écart type le coefficient κ de diffusion pour quantifier le degré de corrélation des points entre eux.

On a ainsi obtenu un opérateur \mathbf{C} de corrélation sur la grille

$$\mathbf{C} = \Lambda \mathbf{L} \Lambda$$

On effectue la même opération sur chacune des variables du modèle (les caractéristiques de l'opérateur \mathbf{L} changeant quelque peu) pour obtenir trois matrices de corrélation \mathbf{C}_u , \mathbf{C}_v et \mathbf{C}_h . En multipliant à gauche et à droite les matrices de corrélation par les matrices diagonales des écarts types, on obtient les blocs diagonaux de l'opérateur de covariance

$$\mathbf{B}_{uvh} = \Sigma_{uvh} \mathbf{C}_{uvh} \Sigma_{uvh}$$

les corrélations croisées (blocs non diagonaux de \mathbf{B}) n'étant pas encore modélisées car on n'a introduit que des corrélations spatiales pour chaque variable séparément. On dit alors que l'on dispose d'un opérateur \mathbf{B} *monovarié*.

Les corrélation croisées

Pour obtenir la matrice \mathbf{B} de covariance d'erreur de *background* complète (ou *multivariée*) on peut écrire :

$$\mathbf{B} = \mathbf{K}_b \hat{\mathbf{B}} \mathbf{K}_b^T \quad (3.17)$$

où

$$\hat{\mathbf{B}} = \begin{pmatrix} \mathbf{B}_{uu} & 0 & 0 \\ 0 & \mathbf{B}_{vv} & 0 \\ 0 & 0 & \mathbf{B}_{hh} \end{pmatrix}$$

et \mathbf{K} est un "opérateur d'équilibre" permettant de relier les variables du modèle entre elles.

Dans cette formulation (DERBER et BOUTTIER 1999 [39]) on a d'abord séparé les variables du modèle en deux composantes : équilibrée et non équilibrée sauf pour une variable prise dans sa totalité (que l'on considérera comme référence de l'équilibre c.a.d pour laquelle il n'y aura pas de distinction entre partie équilibrée et partie non équilibrée). Cette variable de référence sera alors utilisée pour établir un équilibre linéaire avec le reste des variables. Dans notre cas nous sommes en présence de trois variables : les courants \mathbf{u}

et \mathbf{v} et la hauteur d'eau \mathbf{h} . La hauteur d'eau sera notre variable de référence, le vecteur d'état total s'écrira donc

$$\begin{cases} \mathbf{h} = \mathbf{h}' \\ \mathbf{u} = \mathbf{u}' + \mathbf{u}^{eqh} \\ \mathbf{v} = \mathbf{v}' + \mathbf{v}^{eqh} \end{cases} \quad (3.18)$$

où \mathbf{h}' , \mathbf{u}' , \mathbf{v}' représentent les parties non équilibrées de \mathbf{h} , \mathbf{u} , \mathbf{v} respectivement et \mathbf{u}^{eqh} et \mathbf{v}^{eqh} les parties de \mathbf{u} et \mathbf{v} en équilibre avec la hauteur d'eau \mathbf{h} .

Pour notre cas-test on utilise l'équilibre géostrophique comme relation linéaire (notée \mathbf{K}'_b entre \mathbf{h} d'une part et \mathbf{u} et \mathbf{v} d'autre part.

L'équilibre géostrophique est une approximation de l'équation des moments (2.9). Elle est basée sur le fait que, pour des écoulements fluides qui évoluent sur des échelles de temps de plus d'une journée et sur des échelles spatiales de plus de 10 km, et situés assez loin de l'équateur, les termes prépondérants sont les termes de gradient de pression et de force de Coriolis. Dans le cas général, cet équilibre s'écrit (en coordonnées cartésiennes) :

$$\rho f \mathbf{v}^{eqh} = \frac{\partial \mathbf{p}}{\partial x} \quad (3.19)$$

$$\rho f \mathbf{u}^{eqh} = -\frac{\partial \mathbf{p}}{\partial y} \quad (3.20)$$

Compte tenu des approximations faites dans le modèle Shallow-Water (cf section 2.1.2) et de l'équilibre hydrostatique (2.10) on obtient la pression en fonction de la hauteur d'eau $\mathbf{p} = \rho g \mathbf{h}$. Pour le modèle Shallow-Water, l'équilibre géostrophique peut donc s'écrire

$$\mathbf{v}^{eqh} = \frac{g}{f} \frac{\partial \mathbf{h}}{\partial x} \quad (3.21)$$

$$\mathbf{u}^{eqh} = -\frac{g}{f} \frac{\partial \mathbf{h}}{\partial y} \quad (3.22)$$

Le vecteur d'état du modèle étant défini comme la somme de ses composantes équilibrées et non équilibrées notre opérateur \mathbf{K}_b complet s'écrit :

$$\mathbf{K}_b = \mathbf{K}'_b + \mathbf{I} = \begin{pmatrix} \mathbf{I} & 0 & -\frac{g}{f} \frac{\partial}{\partial y} \\ 0 & \mathbf{I} & \frac{g}{f} \frac{\partial}{\partial x} \\ 0 & 0 & \mathbf{I} \end{pmatrix}$$

Le choix de cette relation d'équilibre n'est pas du au hasard ni le résultat d'un infâme bricolage, l'équilibre géostrophique est une approximation de l'attracteur de ce système. L'utilisation de l'opérateur \mathbf{K} permet de repositionner la solution de l'assimilation de données sur (ou non loin de) l'attracteur la rendant plus cohérente avec la physique du système.

Comme on l'a vu précédemment (section (3.3)) ce n'est pas l'opérateur \mathbf{B} qui nous intéresse, mais plutôt ses racines carrées $\mathbf{B}^{\frac{1}{2}}$ et $\mathbf{B}^{\frac{T}{2}}$. Les écritures (3.17) et (3.4.1) permettent d'écrire $\mathbf{B}^{\frac{1}{2}} = \mathbf{K}_b \Sigma \mathbf{C}^{\frac{1}{2}}$. avec

$$C_{uvh}^{\frac{1}{2}} = \Lambda L^{\frac{1}{2}}$$

L'opérateur \mathbf{L} étant auto-adjoint, sa racine carrée est obtenue en n'intégrant les équations (3.4.1) que sur la moitié du temps

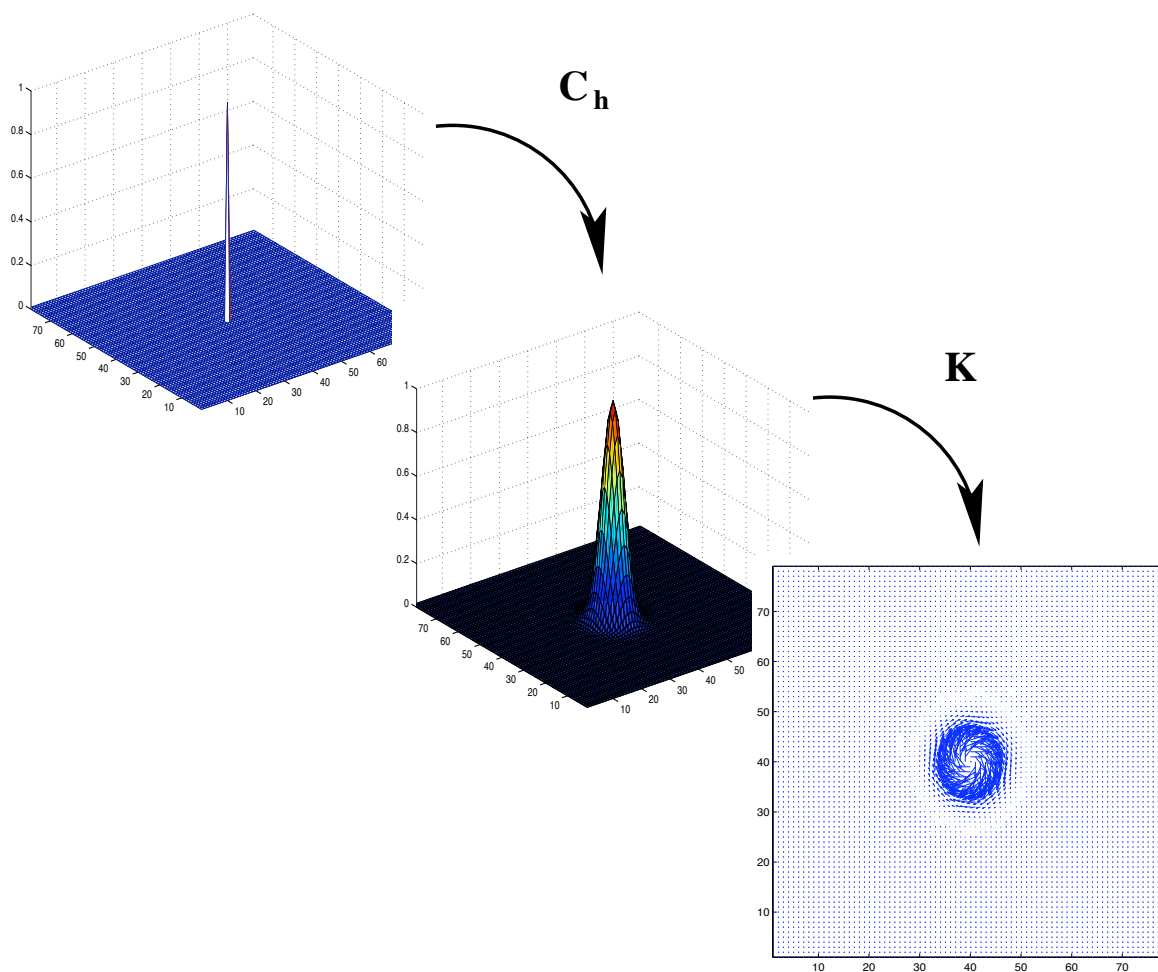


FIG. 3.2: Effet sur les hauteur d'eau (à gauche) et sur le champ de vitesse de l'action de l'opérateur \mathbf{B} multivarié sur un dirac sur les hauteur d'eau

Etape	Résultat	ce que l'on obtient
1- Vecteur de départ	\mathbf{x}	c'est le vecteur que l'on veut multiplier par $\mathbf{B}^{\frac{1}{2}}$
2- Résolution de l'équation de diffusion	$\mathbf{L}^{\frac{1}{2}}\mathbf{x}$	diffusion de l'information sur le domaine
3- Normalisation	$\Lambda\mathbf{L}^{\frac{1}{2}}\mathbf{x}$	on obtient des corrélations spatiales (opérateur monovarié)
4- Multiplication par la racine carrée des écarts types	$\Sigma\Lambda\mathbf{L}^{\frac{1}{2}}\mathbf{x}$	on obtient des covariances spatiales (opérateur monovarié)
5- Application de l'opérateur d'équilibre	$\mathbf{K}_b\Sigma\Lambda\mathbf{L}^{\frac{1}{2}}\mathbf{x}$	on obtient des covariances spatiales (opérateur multivarié)

TAB. 3.1: *Résumé des opérations pour aboutir à la création de l'opérateur de multiplication par $\mathbf{B}^{\frac{1}{2}}$*

3.4.2 Les effets de l'opérateur d'équilibre sur l'assimilation variationnelle

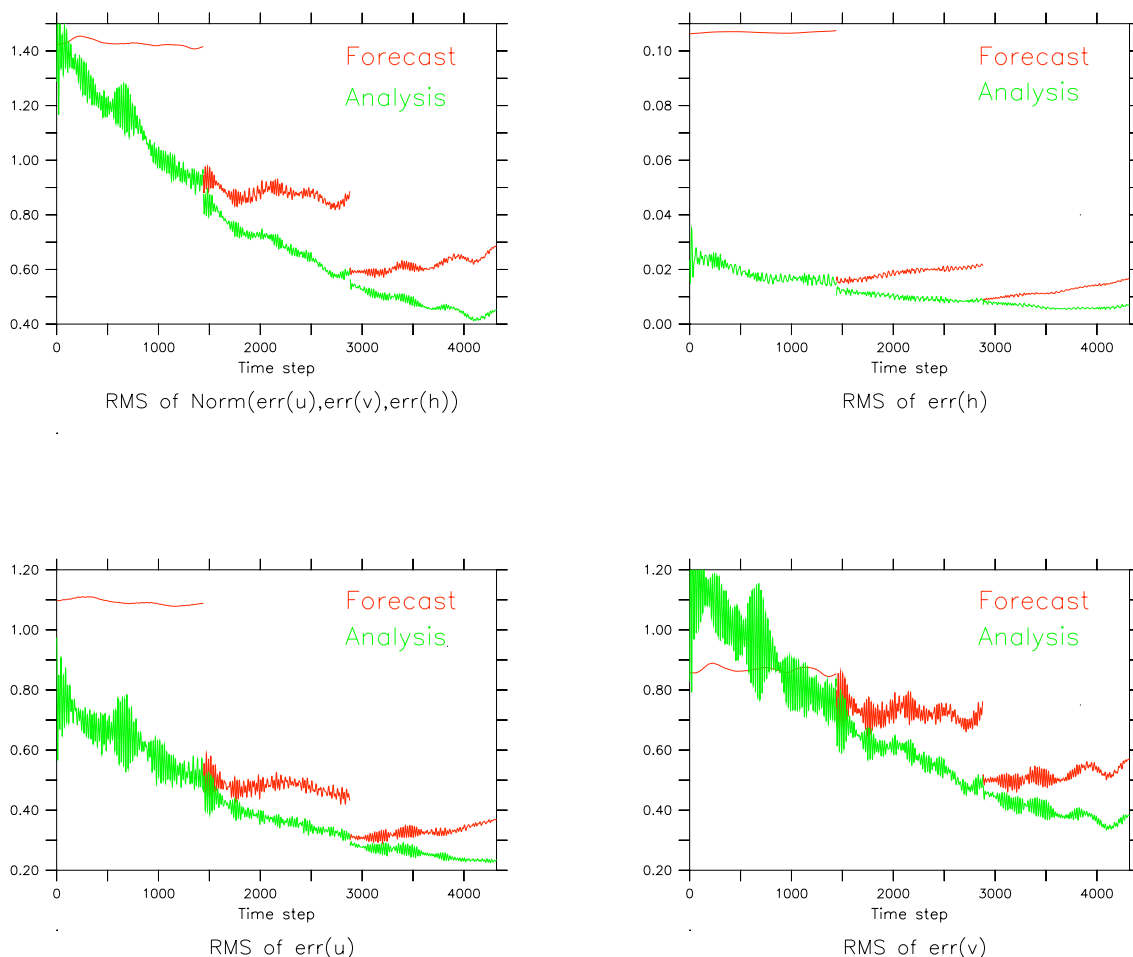


FIG. 3.3: norme rms de l'erreur sur le vecteur d'état en fonction du temps après un 4D-Var avec un opérateur B monovarié

Si dans la représentation de la matrice de covariance d'erreur d'ébauche on néglige les termes de corrélations croisées (on fixe $\mathbf{K} = \mathbf{I}$ dans la formulation (3.17)), les corrections induites par l'assimilation de données vont être principalement appliquées aux variables observées. Dans le cas d'un 3D-Var les autres variables ne seront pas corrigées. Le 4D-Var diffusera les informations sur les autres variables à l'aide de l'intégration rétrograde du modèle adjoint mais cela ne suffira généralement pas en pratique à obtenir une correction qui soit cohérente avec la physique du système.

Pour illustrer l'influence de cet opérateur d'équilibre on a appliqué au modèle Shallow-Water un 4D-Var avec dans le premier cas un opérateur \mathbf{B} monovarié (bloc-diagonal) puis en lui adjoignant l'opérateur de balance \mathbf{K} (matrice \mathbf{B} multivariée).

On a travaillé sur trois fenêtres d'assimilation successives de 1 mois chacune. Sur les graphes (3.3) et (3.4) on a tracé la norme de l'écart à l'état vrai : $\|\mathbf{x} - \mathbf{x}^t\|$ (qui

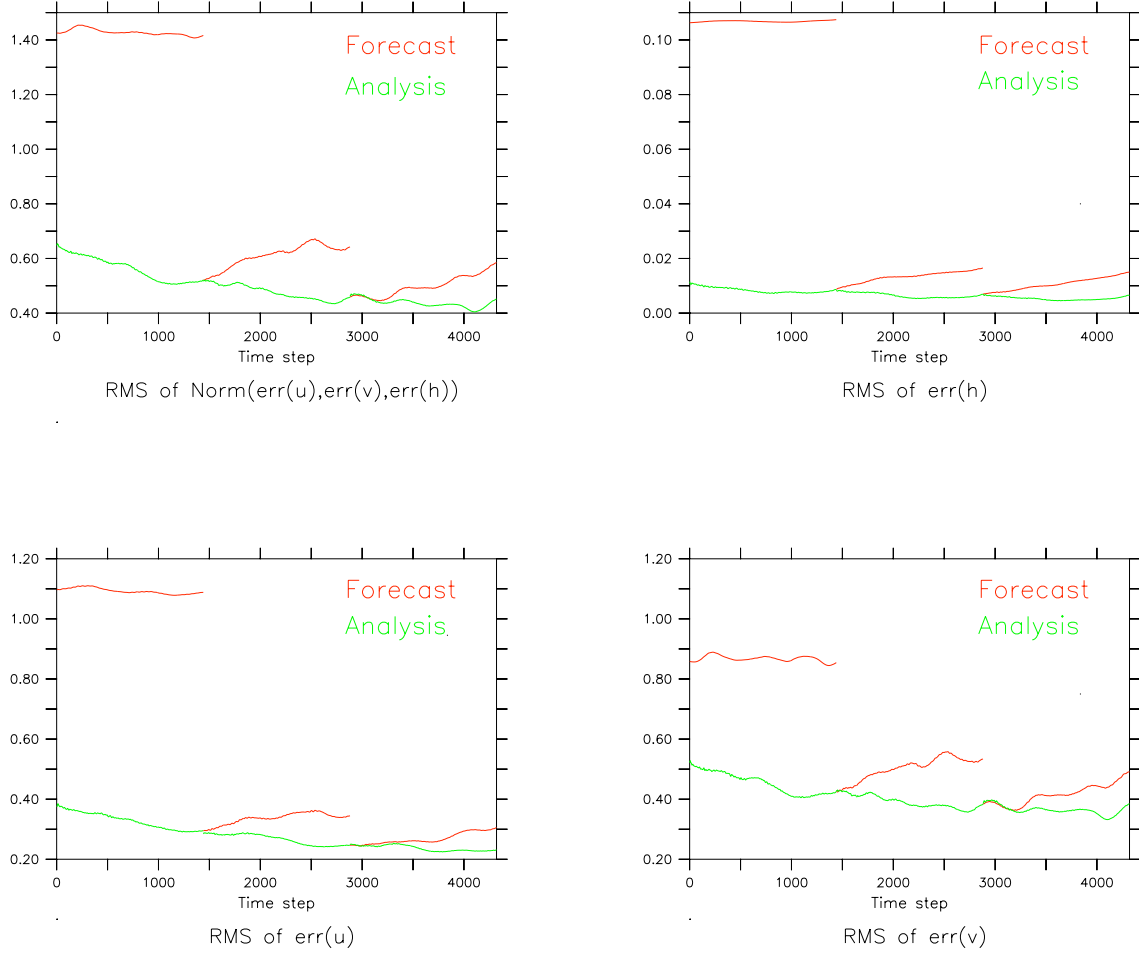


FIG. 3.4: norme rms de l'erreur sur le vecteur d'état en fonction du temps après un 4D-Var avec un opérateur B multivarié

est disponible car on se place dans le cadre d'expérience jumelles). Les observations sont composées d'un sous-échantillonnage spatial et temporel de la hauteur d'eau \mathbf{h} . Les courbes rouges (ou foncées pour ceux qui n'ont pas la chance de disposer de la couleur) représentent l'erreur de la prévision faite par le modèle à partir de l'ébauche $\mathbf{x}_{t_0} = \mathbf{x}^b$ dans la première partie puis à partir de l'état analysé au temps final de la fenêtre d'assimilation précédente. Les courbes vertes (claires) représentent l'évolution temporelle de l'erreur à partir de l'état analysé $\mathbf{x}_{t_0} = \mathbf{x}^b + \delta\mathbf{x}^a$. Sur ces deux figures le schéma en haut à gauche représente l'erreur globale sur les trois variables alors que les trois autres graphiques représentent l'erreur particulière à un des champs (\mathbf{h} , en haut à droite, \mathbf{u} en bas à gauche et \mathbf{v} en bas à droite).

On remarque aisément l'effet d'une correction non équilibrée sur la condition initiale sur les figures (3.3). Comme on n'observe que la hauteur d'eau les corrections de l'état initial sont principalement effectuées sur cette partie du vecteur d'état. Si ce comportement reste pertinent d'un point de vue mathématique, il en résulte que les courants ne sont plus cohérents avec la hauteur d'eau du point de vue de la physique du modèle. Ce dernier tente donc de retrouver son équilibre ce qui induit de fortes ondes de gravité (les fortes

oscillations de la figure (3.3) qui s'atténuent avec le temps (le modèle retourne doucement à l'équilibre). On a, par la correction apporté à la condition initiale, éloigné l'état du système de son attracteur.

L'utilisation de l'opérateur d'équilibre géostrophique \mathbf{K} décrit dans la section précédente permet d'obtenir un opérateur \mathbf{B} multivarié et donc de faire disparaître ces ondes disgracieuses (figure 3.4) qui sont en contradiction avec les particularité de l'attracteur du système (caractérisé par la faiblesse de ses ondes de gravité) et qui rendaient inutilisable la prévision découlant de cette analyse. De plus la correction sur les vitesses de courant est plus importante (la partie équilibrée vient s'ajouter à la partie non équilibrée des courants) et permet d'obtenir sur ces variables une analyse de bien meilleure qualité.

Ce petit exemple met en évidence l'importance de disposer d'une matrice de covariance d'erreur d'*ébauche* multivariée si les observations ne portent pas sur l'ensemble des variables contrôlées par l'assimilation. Dans cet exemple l'opérateur d'équilibre fournit l'information manquante du fait de la non observation des vitesses de courant.

Dans cette thèse, pour l'application de l'assimilation de données variationnelle au modèle réaliste OPA, nous ne disposons pas d'un opérateur \mathbf{B} multivarié, sa construction étant nettement plus complexe que le simple équilibre géostrophique du Shallow-Water. Nous verrons par la suite que ce manque induit de sérieux problèmes et ne permet pas d'obtenir une analyse de qualité sur les variables qui ne sont pas observées.

3.5 Diagnostics

L'évaluation de l'efficacité des méthode d'assimilation est toujours délicate. En effet à part dans le cadre d'expériences jumelles on ne dispose pas de l'état vrai, il est donc difficile de quantifier l'erreur d'analyse. Si les diagnostics sur les aspects physiques permettent d'éliminer des résultats très improbables, il n'en est pas de même pour les petites anomalies : comment savoir s'il s'agit alors d'une erreur de l'analyse ou si cette légère anomalie était présente dans la réalité. Pour compléter les aspects physiques on dispose d'une batterie de diagnostics mathématiques et empiriques afin d'évaluer la qualité et la performance de l'analyse :

3.5.1 Les diagnostics d'optimisation :

D'un point de vue uniquement numérique il est intéressant de disposer des valeurs à chaque simulation successive de la fonction objective J et de ses différentes composantes J_o , J_b , etc,...

Une décroissance de la fonction objective irrégulière (plus d'une simulation par itération) constatée *a posteriori* peut révéler un mauvais conditionnement du problème qui rend difficile l'estimation du nouvel itéré au minimiseur.

La norme du gradient $\|\nabla J\|$ et de ses différentes composantes, $\|\nabla J_o\|$, $\|\nabla J_b\|$, ... à chaque itération donne aussi de précieuses informations sur l'efficacité de l'optimisation.

3.5.2 Les diagnostics statistiques :

- Si en assimilation de données variationnelle on cherche à minimiser la fonction objective

$$J(\delta\mathbf{x}) = \frac{1}{2}[\mathbf{G}\delta\mathbf{x} - \mathbf{d}]^T \mathbf{R}^{-1}[\mathbf{G}\delta\mathbf{x} - \mathbf{d}] + \frac{1}{2}\delta\mathbf{x}^T \mathbf{B}^{-1}\delta\mathbf{x} \quad (3.23)$$

on obtient l'état optimal $\delta\mathbf{x}^a$ de la même façon que pour l'équation (1.37).

$$\delta\mathbf{x}^a = \mathbf{K}\mathbf{d} \quad (3.24)$$

où $\mathbf{K} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1}$.

On a immédiatement :

$$\mathbf{G}\delta\mathbf{x}^a - \mathbf{d} = [\mathbf{G}\mathbf{K} - \mathbf{I}]\mathbf{d} \quad (3.25)$$

On obtient ainsi

$$J(\delta\mathbf{x}^a) = \frac{1}{2}\mathbf{d}^T[(\mathbf{G}\mathbf{K} - \mathbf{I})^T \mathbf{R}^{-1}(\mathbf{G}\mathbf{K} - \mathbf{I}) + \mathbf{K}^T \mathbf{B}^{-1}\mathbf{K}]\mathbf{d} \quad (3.26)$$

Comme $\mathbf{I} - \mathbf{G}\mathbf{K} = \mathbf{R}[\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}]^{-1}$ et $\mathbf{K}^T \mathbf{B}^{-1}\mathbf{K} = [\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}]^{-1}\mathbf{G}\mathbf{B}\mathbf{G}^T[\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}]^{-1}$ on obtient :

$$J(\delta\mathbf{x}^a) = \frac{1}{2}\mathbf{d}^T[\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}]^{-1}\mathbf{d} \quad (3.27)$$

$2 \times J(\delta\mathbf{x}^a)$ est donc la norme au carré du vecteur innovation \mathbf{d} associée à sa propre matrice de covariance et, si de plus les erreurs sont supposées Gaussiennes, son espérance est égale à la taille du vecteur innovation (TALAGRAN, 1998 [112]) c'est à dire au nombre p d'observations :

$$E[J(\delta\mathbf{x}^a)] = \frac{p}{2} \quad (3.28)$$

et

$$var[J(\delta\mathbf{x}^a)] = \frac{p}{2} \quad (3.29)$$

Dans les applications océaniques ou météorologiques les fluctuations de $J(\delta\mathbf{x}^a)$ sont négligeables par rapport à son espérance $E[J(\delta\mathbf{x}^a)]$ (TALAGRAN, 1998 [112]). Le diagnostic $J(\delta\mathbf{x}^a) \approx \frac{p}{2}$ peut donc se révéler pertinent. En effet $J(\delta\mathbf{x}^a) < \frac{p}{2}$ pourra signifier que les variances d'erreur d'observation ou de background spécifiées sont trop grandes. *A contrario* $J(\delta\mathbf{x}^a) > \frac{p}{2}$ signifiera que les variances d'erreur d'observation ou de background spécifiées sont trop petites ou que le vecteur innovation contient des observations erronées. Ce diagnostic est toutefois à prendre avec précaution; en effet imaginons que nos deux matrices \mathbf{B} et \mathbf{R} sont mal évaluées à un même facteur α près. alors on a :

$$J_\alpha(\delta\mathbf{x}_\alpha^a) = \frac{1}{2}\mathbf{d}^T[\alpha\mathbf{G}\mathbf{B}\mathbf{G}^T + \alpha\mathbf{R}]^{-1}\mathbf{d} = \frac{1}{\alpha}J(\delta\mathbf{x}^a) \quad (3.30)$$

alors que l'état optimal reste inchangé.

$$\delta \mathbf{x}_\alpha^a = \alpha \mathbf{B} \mathbf{H}^T [\alpha \mathbf{H} \mathbf{B} \mathbf{H}^T + \alpha \mathbf{R}]^{-1} \mathbf{d} = \delta \mathbf{x}^a \quad (3.31)$$

Ce diagnostic n'est donc pertinent que dans le cas où $E[J(\delta \mathbf{x}^a)] \neq \frac{\rho}{2}$.

- Sur une longue suite de cycles d'assimilation on peut calculer la moyenne des incréments à la condition initiale. Théoriquement si le modèle ne comporte pas de biais cette moyenne doit être nulle. Dans le cas contraire il deviendra nécessaire de revoir l'hypothèse d'erreur modèle de moyenne nulle.

3.5.3 Les diagnostics ImA : Information Minus Analysis.

Les informations provenant de $\mathbf{y}^o - H(\mathbf{x}^b)$ et $\mathbf{y}^o - H(\mathbf{x}^a)$ peuvent s'avérer très pertinentes dans le cadre de comparaisons de méthodes car contenant des informations sur la physique manquante du modèle. Ces informations peuvent prendre diverses formes :

- moyenne globale : $\frac{1}{n} \sum_{i=1}^n \mathbf{y}_i^o - H(\mathbf{x}_i^{b,a})$
- rms (root mean square) : $\sqrt{\frac{1}{n} \sum_{i=1}^n [\mathbf{y}_i^o - H(\mathbf{x}_i^{b,a})]^2}$
- à chaque instant la variance estimée : $\sum_{i=1}^n [\mathbf{y}_i^o - H(\mathbf{x}_i^{b,a}) - \text{Moyenne globale}]^2$

où n est le nombre d'observations.

3.5.4 Les données indépendantes :

Comme il a été dit à la fin du chapitre 2 la comparaison avec les données observées n'est pas suffisante. Il est nécessaire de comparer le résultat de l'assimilation avec des données indépendantes de l'analyse et de préférence portant sur des variables qui ne sont pas observées.

Chapitre 4

Le coupleur PALM

Le projet français d'océanographie opérationnelle MERCATOR a pour mission de mettre en place un système de simulation et de prévision de l'océan global à haute résolution. Ce projet nécessite, entre autres, la mise en place de chaînes performantes d'assimilation de données in-situ et satellitaires. Ces techniques étant en permanente évolution les contraintes sur les chaînes d'assimilation deviennent doubles :

- Une flexibilité accrue pour pouvoir suivre l'évolution de ces techniques.
- Des performances et une fiabilité dignes d'un système opérationnel.

Le souci de flexibilité impose une complète modularité afin de pouvoir créer et tester de nouvelles configurations de la chaîne sans efforts d'implémentation. C'est ce besoin qui a nécessité le développement du logiciel PALM effectué dans l'équipe "Modélisation du Climat et de son Changement Global" du CERFACS.

Comme on l'a remarqué précédemment les méthodes d'assimilation peuvent être décrites à l'aide d'un nombre restreint d'opérateurs communs qui ont été listés au début du chapitre 1 (les différents intervenants). C'est cette constatation qui a abouti à la création du logiciel PALM.

Dans le cadre de cette thèse a vu le jour la première utilisation "grandeur nature" du coupleur PALM. Ce chapitre sera consacré à une brève description des fonctionnalités de PALM et un exemple d'une "palmisation" d'une méthode d'assimilation sera donné.

4.1 PALM quesaco ?

Outre la gestion des aspects système, tels que le rapatriement des données, la sauvegarde des *restart* en cas de panne informatique, le stockage et l'envoi des résultats, etc..., le logiciel PALM doit donc gérer les différentes étapes de l'algorithme d'assimilation sélectionné.

4.1.1 Principe et définitions

On peut voir PALM comme un coupleur entre les différentes composantes d'une chaîne d'assimilation (opérateur d'observation, modèle, minimiseur, etc...). La philosophie de PALM est de découper toute méthode d'assimilation de données en une suite d'étapes

fondamentales et de séparer celles qui relèvent du problème physique de celles qui relèvent de l'algèbre. Toute étape liée au problème physique (modèle, opérateur d'observation, matrices de variance/covariance, etc...) est à voir comme une boîte fonctionnelle (que l'on pourra appeler "module" ou "unité") qui se branche à la structure de lancement et communication PALM et qui peut être remplacée par n'importe quelle autre boîte qui assure la même fonctionnalité. PALM se charge ensuite de gérer la partie purement algébrique du problème d'assimilation en effectuant l'enchaînement adéquat des unités appropriées. En d'autres termes, partant du principe que toutes les méthodes d'assimilation peuvent être décrites à l'aide d'un nombre restreint d'opérateurs communs, le logiciel PALM permet de construire et de faire évoluer les chaînes d'assimilation sans avoir à modifier les différentes unités, ou *a contrario*, permet de remplacer les unités par leur équivalent plus performant ou plus adapté au problème traité sans pour autant modifier la chaîne d'assimilation.

Avant d'entrer dans une description plus détaillée du coupleur PALM, introduisons les quelques notions de base qui y sont associées :

- Les unités : morceaux de code fortran indépendants, elles décrivent les différentes étapes du processus d'assimilation. Elles doivent être vues comme des modules indépendants les uns des autres et de la méthode d'assimilation choisie. Typiquement les unités peuvent représenter le modèle, l'opérateur d'observation ou les opérateurs approximant les matrices de covariances.
- Les objets : ils représentent l'information consommée, produite et/ou échangée par les différentes unités. On peut citer en exemple d'objet l'ébauche \mathbf{x}^b , entrée de l'unité "modèle", ou encore le vecteur innovation \mathbf{d} calculé à partir de l'opérateur d'observation et permettant le calcul de la fonction objective (en assimilation variationnelle).
- Les branches : groupes d'unités à exécuter en séquence, les chaînes d'assimilation de données peuvent être organisées en branches qui fonctionnent en parallèle. Elles sont dépendantes de la méthode choisie. L'exemple le plus classique est, en assimilation variationnelle 4D, une branche qui gère l'intégration du modèle tandis qu'une autre, récupère les sorties de celle ci pour calculer le vecteur innovation au fur et à mesure. Elles sont générées par l'interface PréPALM (voir plus loin).
- Les communications : les objets doivent transiter entre les différentes unités (d'une même branche ou d'une branche différente)
- Le buffer : c'est un espace mémoire dans lequel les objets peuvent être stockés soit parce qu'il sont en cours d'élaboration, soit parce qu'il est nécessaire de les conserver après leur utilisation par certaines unités. Un certain nombre d'opérations simples peuvent être faites sur les objets contenus dans le buffer (addition, soustraction, remise à zéro, etc...), elles seront décrites plus précisément dans la suite du chapitre.
- PréPALM : Si les unités sont des programmes fortran indépendants codés séparément, les branches et les communications entre les unités qui constituent la méthode d'assimilation doivent être décrites pour permettre au coupleur PALM de les gérer. PréPALM est une interface graphique développée spécialement à cet effet (c'est un

logiciel X-Windows réalisé en Tcl/Tk). Dans la partie “*PALM par l'exemple*” du présent chapitre la description de la mise sous PALM d'une chaîne d'assimilation sera illustrée par des copies d'écran de cette interface.

4.1.2 Les commandes

Le but de PALM n'étant pas de recréer un langage abscons, un effort a été fait pour réduire la “sur-couche” PALM à un nombre minimal d'interfaces. Les unités sont des programmes FORTRAN dans lesquels on rajoute les interfaces PALM. Ces dernières définissent quels objets (variables) sont nécessaires à l'unité en entrée et fournis en sortie par celle ci.

Dans la liste des interfaces PALM qui suit, en règle générale, seules les deux premières (PALM_Put/PALM_Get) sont incluses dans les unités, les autres étant plutôt réservées aux codes de branche.

- PALM_Put/PALM_Get(cl_space,cl_name,il_time,il_usrtag,OBJET,il_err) : principale interface PALM; par celle-ci l'unité envoie (ou reçoit) l'objet appelé cl_name dans la liste de communication (et qui appartient à l'espace cl_space). À cet objet est associée la variable fortran OBJET, la communication est définie au temps PALM il_time (et éventuellement avec le signet il_usrtag).
- PALM_Step(il_step,il_err) : Étape de synchronisation : elle peut avoir plusieurs fonctions. En premier lieu, elle permet de synchroniser plusieurs branches. Cela évitera qu'une branche commence à faire des calculs en utilisant des objets obsolètes. L'autre rôle des PALM_Step est de changer le statut des objets contenus dans le buffer. Ces changements de statut décrits au moyen d'un fichier adjoint à PréPALM (dont on épargnera la syntaxe au lecteur) sont de quatre ordres :

PL_ACT_READY : Préciser qu'un objet est prêt dans le Buffer : Un objet peut avoir été construit petit à petit en ajoutant (ou soustrayant) d'autres objets (typiquement le vecteur innovation est construit à l'aide des écarts aux observations à chaque pas de temps), les unités nécessitant cet objet ne pourront le récupérer (au moyen d'un PALM_Get) que lorsque ce PALM_Step aura été effectué .

PL_ACT_NO_READY : Une fois qu'il a été récupéré, un objet peut nécessiter quelques modifications pour être utilisé par d'autres unités. Il est alors nécessaire de préciser qu'un objet du Buffer n'est pas (plus) prêt; il ne pourra alors être réutilisé que lorsqu'un nouveau PL_ACT_READY sera effectué.

PL_ACT_DELETE : Lorsque l'objet ne sera plus réutilisé, pour économiser de la place mémoire, on peut effacer un objet du Buffer

PL_ACT_NULLIFY : Il peut aussi être utile de réinitialiser un objet (le remettre à zéro) dans le Buffer (typiquement à la fin d'une boucle).

Ces étapes de synchronisation peuvent aussi être déclenchés non plus par cette commande mais par une communication précise (cela est aussi défini dans PréPALM).

- `PALM_Launch(cl_unit,il_err)` : Commande servant à une branche pour lancer l'unité `cl_unit`.
- `PALM_Start(cl_branch,il_err)` : Commande servant à une branche pour lancer une autre branche (nommée `cl_branch` dans PréPALM).
- `PALM_Finalize(il_err)` : Fin d'une branche.
- `PALM_Track_on(il_level)` : Commande permettant d'activer le mode trace de PALM, i.e de fournir, dans un fichier, des informations sur les communications effectuées lors de l'exécution. Cette option est utile lors du débogage de la chaîne d'assimilation. Le paramètre `il_level` décrit le niveau de verbosité (la quantité d'information) désiré.
- `PALM_Track_off` : désactive le mode trace.

4.1.3 Trucs et astuces

- La "Palmisation" d'un pan de code fortran pourra toujours se traduire par la création d'une routine englobant toute l'unité. Celle ci peut être décrite de la sorte :

SUBROUTINE UNIT

CALL INIT_UNIT : Initialisation de l'unité, soit par lecture sur fichier, soit par `PALM_Get` d'objets provenant d'une unité chargée des diverses initialisations. Typiquement l'appel de cette routine ne doit être fait que lors du premier lancement de cette unité. (étape 1)

CALL INPUT_UNIT : Routine comprenant les divers `PALM_Get` correspondant aux entrées de l'unité. (étape 2)

CALL ROUTINE : Appel à la routine d'origine allégée éventuellement des appels aux routines appartenant à d'autres unités. Il peut aussi être nécessaire de rajouter des entrée/sorties (`PalM_Get/PALM_put`) si ceux ci ne peuvent être effectuer hors de la routine (cf exemple plus loin). (étape 3)

CALL OUTPUT_UNIT : `PALM_Put` des sorties de l'unité. (étape 4)

END UNIT

On verra un exemple précis de création de routine de ce type dans la section suivante.

- Afin d'assurer une modularité complète il convient de décrire, au moyen de `PALM_Put` et/ou de `PALM_Get`, l'ensemble des entrées/sorties possibles de chaque unité. Comme elles ne seront pas forcément utilisées par toutes les méthodes d'assimilation il suffira, le cas échéant, de ne pas décrire dans PréPALM les communications inutiles, le driver PALM ignorera alors ces `PALM_Put/PALM_Get` optionnels.

4.2 PALM par l'exemple

Nous allons étudier la création d'une assimilation 4D-Var incrémental (cf 1.3.2) sur le modèle aux équations primitives OPA décrit dans la partie 2.2. Ceci est la mise sous PALM de la méthode développée par Anthony WEAVER.

Afin d'être le plus clair possible cette mise sous PALM sera décrite étape par étape et illustré par les copies d'écrans PréPALM d'une part et le parallèle entre la version d'origine et la version finale d'autre part.

Ce code, ainsi que les variantes introduites dans la suite seront appelés OPALM

L'initialisation

La lecture des fichiers d'initialisation du modèle et le calcul des différents paramètres de l'expérience sont effectués de manière identique que l'on utilise PALM ou pas. Dans la version "palmissée" les étapes de cette phase ont simplement été regroupés dans l'unité OPA_INIT

Le Modèle

Code d'origine	Avec PALM	description
...l'initialisation est stockée dans les COMMON	<code>PALM_Get(bathy, grille, ...)</code>	...Récupération de l'initialisation
<code>x_init=</code>	<code>CALL PALM_Get(x_init)</code>	...condition initiales
<code>t_deb=</code>	<code>CALL PALM_Get(t_deb)</code>	... temps de début de l'intégration
<code>t_fin=</code>	<code>CALL PALM_Get(t_fin)</code>	... temps de fin de l'intégration
<code>CALL(step)</code>	<code>CALL(step)</code>	Intégration temporelle du modèle
... écriture des sorties du modèle dans un fichier	<code>CALL PALM_Put(sortie)</code>	Sorties du modèle

TAB. 4.1: découpage grossier de l'intégration temporelle du modèle

La création de la boîte modèle sera étudiée très précisément afin de bien comprendre les étapes nécessaires à la palmisation. Par la suite les unités seront décrites plus succinctement. On appelle cette unité “Modèle” par abus de notation puisqu’il s’agit en fait de l’intégration temporelle du modèle

Comme nous l’avons vu précédemment le modèle OPA, donné ici en exemple, décrit l’évolution de quatre variables d’état : les deux composantes du champ de vitesse \mathbf{u} et \mathbf{v} , la température \mathbf{T} et la salinité s . Ces quatre champs seront donc les entrées de notre unité “modèle”, on peut y ajouter le pas de temps de départ (et la date associée) et celui d’arrivée.

L’unité modèle doit également être capable de fournir l’état du système à chaque pas de temps et de recevoir un terme de forçage (une correction d’erreur modèle par exemple).

La palmisation de ce modèle pourra se traduire par la création d’une routine supplémentaire faisant les `PALM_Get` correspondant aux entrées précitées suivi de l’appel à la routine modèle d’origine et enfin les `PALM_Put` des sorties. Pratiquement cela a abouti à la création de la routine `M_DIR` (modèle direct) qui, après un appel à une routine chargée de récupérer, au moyen de `PALM_get` les différents paramètres d’initialisation du modèle, récupère les entrées du modèle.

Dans le tableau 4.1, \mathbf{x} représente le vecteur d’état (\mathbf{u} , \mathbf{v} , \mathbf{T} , s). Dans la configuration originale, après une phase d’initialisation, nécessaire à tous les opérateurs (calculer la grille de discrétisation, les zones de terre, etc...) qui a été regroupée dans l’unité `OPA_INIT`, l’intégration du modèle et le calcul du vecteur innovation $\mathbf{d}_i = H_i(\mathbf{x}_i) - \mathbf{y}_i^o$ étaient regroupés dans la routine d’intégration du modèle. Les variables d’initialisation et d’entrée du modèle résidaient en mémoire (dans les `COMMON` : variables partagées par toutes les routines). Si la transmission de l’information par `COMMON` à été conservée à l’intérieur de notre nouvelle unité, il n’en est pas de même pour la transmission aux autres parties du code.

Notre nouvelle unité `M_DIR` commencera donc naturellement par une série de `PALM_Get` d’objet provenant de l’initialisation (comme par exemple les masques définissant les zones de terre, la bathymétrie, etc...). Cette phase d’initialisation (étape 1 de paragraphe 4.1.3) n’aura lieu que lors de la première intégration du modèle. Une fois l’initialisation effectuée, pour pouvoir intégrer le modèle, il est nécessaire de récupérer l’état initial du modèle ainsi que les pas de temps de début et de fin d’intégration (étape 2 ou tableau 4.1). L’étape d’intégration du modèle en elle même peut donc enfin avoir lieu. Comme on peut le voir dans le tableau 4.2, elle se traduira par l’appel de la routine d’origine (étape 3) dans laquelle la partie calcul du vecteur d’innovation et la partie de sauvegarde de la trajectoire de référence nécessaire pour les modèles linéaires tangent et adjoint (routine `trjwri`) ont été remplacée par les `PALM_Put` des différentes composantes du vecteur d’état (PALM gèrera ensuite la transmission de cette information aux unités concernées). Des `PALM_Get` de termes correctifs du vecteur d’état ont aussi été ajoutés, ils ne serviront pas dans les cas exposés ici, mais seront utilisés dans les variantes du 4D-Var présentées dans la deuxième

partie de cette thèse. La dernière étape, voit l'envoi de l'état du système à la fin de l'intégration ainsi que de la date calculée au pas de temps final.

On peut voir sur la figure 4.1 le résultat de la description dans PréPALM des différentes communications entre les unités OPA_INIT et M_DIR ici utilisées pour la génération des observations pour les expériences jumelles (d'où le nom MODELOBS). Les 52 entrées de la boîte modèle sont constituées pour la plupart des différents paramètres d'initialisation auxquels s'ajoutent les composantes de la condition initiale (\mathbf{u} , \mathbf{v} , \mathbf{t} , \mathbf{s}) et les bornes temporelles (t_{deb} et t_{fin}) de la fenêtre d'intégration. Les six sorties sont composées de l'état final (4 champs), et des tendances de la fonction de courant barotrope (2 champs). Les autres éventuels PALM_Get/PALM_Put ne sont pas décrits dans PréPALM, ils ne seront donc pas effectués lors du déroulement du code (communications optionnelles).

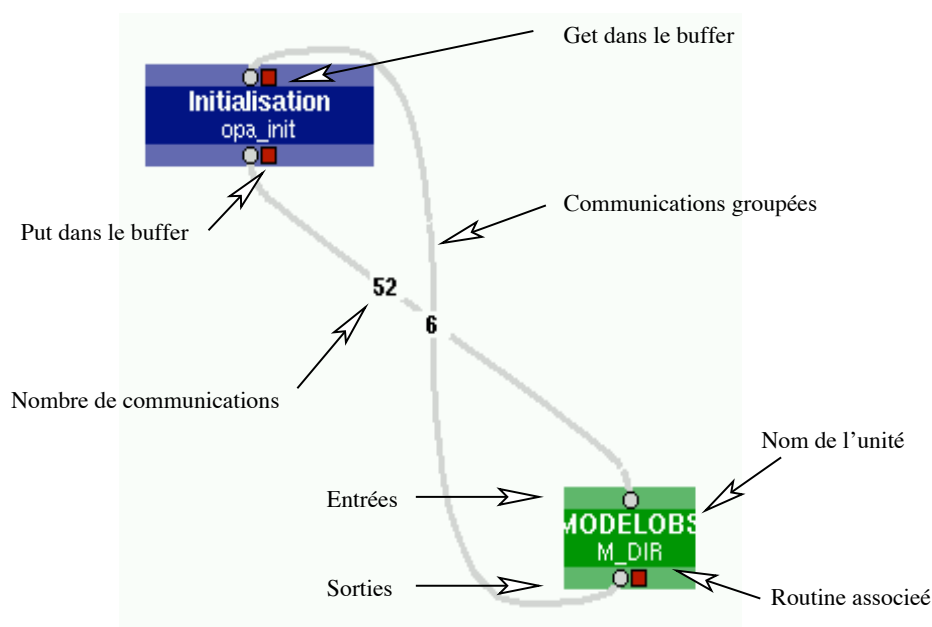


FIG. 4.1: Premiers pas

L'opérateur d'observation

Le graphique 4.1 représente la partie modèle de la génération des observations dans le cadre des expériences jumelles. Pour compléter cette fabrication d'observations synthétiques il convient de "récupérer" les sorties du modèle à chaque pas de temps et de les traduire dans l'espace des observations au moyen de l'opérateur d'observation. Cette opération, $\mathbf{y}_i^o = H(\mathbf{x}_i^t)$, à l'origine incluse dans la routine `tammis` appelée par la routine `step` d'intégration

Code d'origine	Avec PALM	description
FOR stp=t_deb TO t_fin	FOR stp=t_deb TO t_fin	boucle temporelle
DO	DO	
...	...	début de l'intégration du modèle
CALL trjwri(stp)	CALL PALM_Put(x(stp))	sauvegarde de la trajectoire de référence (qui servira aussi au calcul du vecteur innovation avec PALM)
CALL tammis(stp)	délocalisé	calcul du vecteur innovation
...	...	suite de l'intégration temporelle du modèle
END DO	END DO	Fin de la boucle

TAB. 4.2: découpage fin de l'intégration temporelle du modèle (dans la routine `step`)

du modèle, est effectuée par l'unité `H_OBS` (cf tableau 4.3). Elle reçoit en entrée le vecteur d'état (en provenance du modèle), le pas de temps correspondant ainsi qu'une indication sur le mode de fonctionnement souhaité (ici génération des observations) en provenance de la branche qui gère les observations. En sortie elle place dans le buffer la partie y_i^o du vecteur observation y^o calculée au temps t_i (communication symbolisée par le petit carré en bas de la boîte `H_OBS` sur la figure 4.2). La branche des observations appelle l'opérateur dans une boucle sur les pas de temps (voir la description de la branche tableau 4.6)

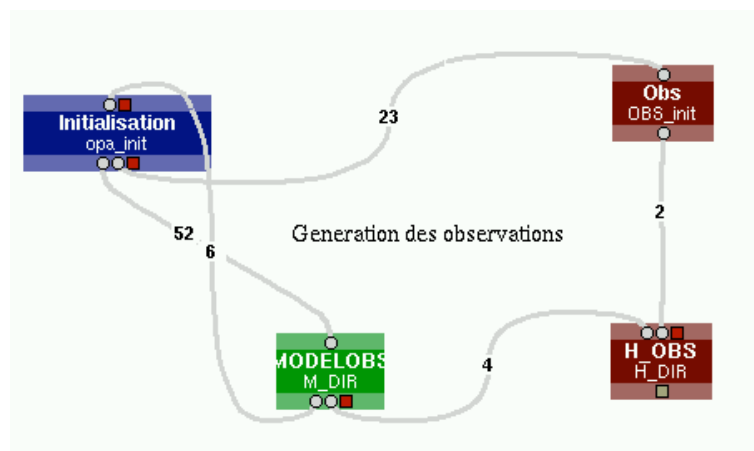


FIG. 4.2: Génération des observations

Code d'origine	Avec PALM	description
dans la routine <code>tammis(stp)</code> de <code>step</code>	dans une <u>unité</u> indépendante H	
le vecteur d'état \mathbf{x} est dans les COMMON	<code>Palm_Get(x)</code>	récupération du vecteur d'état
	<code>PALM_Get(stp)</code>	récupération du pas de temps courant
<code>y(stp)=H(x(stp))</code>	<code>y(stp)=H(x(stp))</code>	calcul de l'équivalent à l'état du modèle dans l'espace des observations au temps <code>stp</code>
<code>y</code> est stocké dans les COMMON	<code>CALL PALM_Put(y(stp))</code>	on garde l'équivalent à l'état du modèle dans l'espace des observations

TAB. 4.3: Opérateur d'observation

Pour nous échauffer nous avons construit un ensemble PALM capable de générer des observations synthétiques issues du modèle OPA. On peut maintenant entrer dans le vif du sujet et décrire la chaîne d'assimilation 4D-Var incrémental.

La méthode d'assimilation

La méthode d'assimilation choisie est une variante incrémentale du 4D-Var (voir la partie 1.3.2) à laquelle on applique un changement de variable composé d'une part du préconditionnement par la racine carrée de la matrice de covariance d'erreur d'ébauche et d'autre part de la projection des vitesses de courants dans l'espace des flux barotropes non divergents (i.e. la divergence de la moyenne des flux sur la hauteur est nulle). Cette projection est nécessaire afin que l'assimilation respecte l'égalité (2.19) imposée par le choix "toit rigide".

Si on note Π cette projection, alors la méthode se ramène à effectuer la minimisation de la fonctionnelle J définie par

$$\mathbf{v} = \mathbf{B}^{\frac{-1}{2}} \Pi^{-I} \delta \mathbf{x}_0 \quad (4.1)$$

$$J(\mathbf{v}) = \frac{1}{2} \sum_{i=1}^n [\mathbf{H}_i \mathbf{M}_{t_i, t_0} \delta \mathbf{x}_0 - \mathbf{d}_i]^T \mathbf{R}_i^{-1} [\mathbf{H}_i \mathbf{M}_{t_i, t_0} \delta \mathbf{x}_0 - \mathbf{d}_i] + \frac{1}{2} [\mathbf{v}]^T \mathbf{v} \quad (4.2)$$

où $\mathbf{d}_i = \mathbf{y}_i - H_i(M_{t_i, t_0}(\mathbf{x}^b))$

La boucle externe

Pour commencer il est nécessaire de calculer la trajectoire \mathbf{x} de référence (utilisée pour calculer les modèles linéaire tangent à cette trajectoire et son adjoint) ainsi que son équivalence

$H(\mathbf{x})$ dans l'espace des observations (pour calculer l'innovation $\mathbf{d} = \mathbf{y}^o - H(\mathbf{x})$). On voit au signet (0) de la figure 4.3 que la routine utilisée pour générer la trajectoire de référence est la même que celle utilisée pour les observations synthétiques; cependant le nom de l'unité a changé car les objets échangés ne sont pas les mêmes (il faut donc différencier les deux unités dans PréPALM même si la routine associée est la même). On remarque notamment que les communications afférentes à l'initialisation ont disparu, ces informations étant toutes stockées dans la mémoire du processeur (dans les COMMON). Si elles sont toujours décrites dans la routine, leur absence de la description PréPALM fait qu'elles seront ignorées.

À chaque pas de temps, H stocke le vecteur innovation \mathbf{d}_i dans le buffer ((1) de la figure 4.3) en y construisant un unique vecteur \mathbf{d} regroupant toutes les observations. Pour pouvoir être utilisé ce vecteur devra être marqué comme étant "prêt" au moyen d'une étape de synchronisation (un PL_ACT_READY). Cette étape ne sera pas déclenchée par un PALM_Step mais par l'envoi par H du \mathbf{d}_i au dernier pas de temps. Une unité Loop est aussi créée, son rôle est de gérer les boucles (loop en anglais) de minimisation et le minimiseur lui-même ((2) de la figure 4.3).

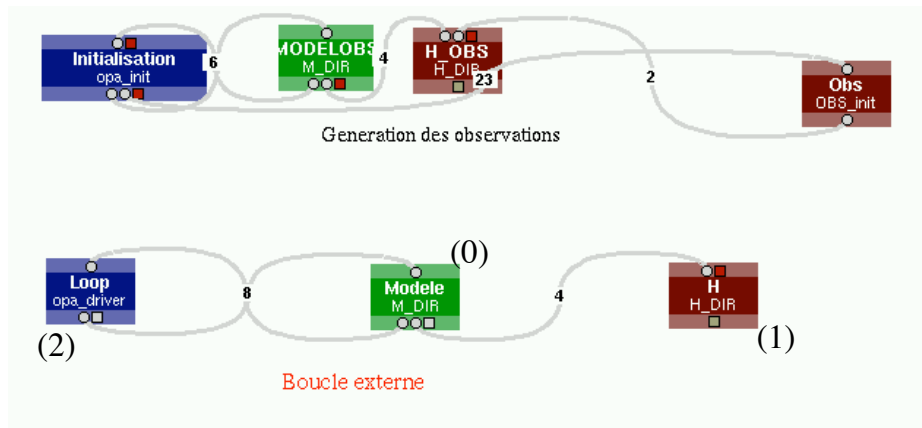


FIG. 4.3: Boucle externe

Les modèles tangent et adjoint et l'opérateur de changement de variable

De la même façon que pour les autres unités, on peut maintenant créer la boîte B (qui, par abus de notation regroupe les multiplications par $\mathbf{B}^{\frac{1}{2}}$ et Π). La construction de B s'apparente à celle de H (tableau 4.3). Cette unité reçoit du minimiseur le vecteur de contrôle \mathbf{v} ((4) de la figure 4.4), effectue le changement de variable et envoie l'incrément (sous la forme de quatre champs) au modèle linéaire tangent (5). Ce dernier peut donc produire les $\delta\mathbf{x}_i$ qu'il envoie chaque pas de temps à l'opérateur linéaire tangent de l'opérateur d'observation (6). Celui ci génère $\mathbf{H}\delta\mathbf{x}$ en additionnant les contributions de chaque pas de temps dans le buffer (7) (là encore une étape de synchronisation en PL_ACT_READY sera

nécessaire). L'écart au vecteur innovation $\mathbf{d} - \mathbf{H}\delta\mathbf{x}$ peut ainsi être calculé (dans l'unité d'algèbre TMP_ALG (8)) puis pondérée par l'inverse de la matrice \mathbf{R} de covariance d'erreur d'observation (9). Ceci permettra de déterminer (à nouveau dans l'unité d'algèbre) la valeur de la fonction objective et de l'envoyer au minimiseur Loop (10). Toutes ces étapes peuvent être suivies sur la figure (4.4).

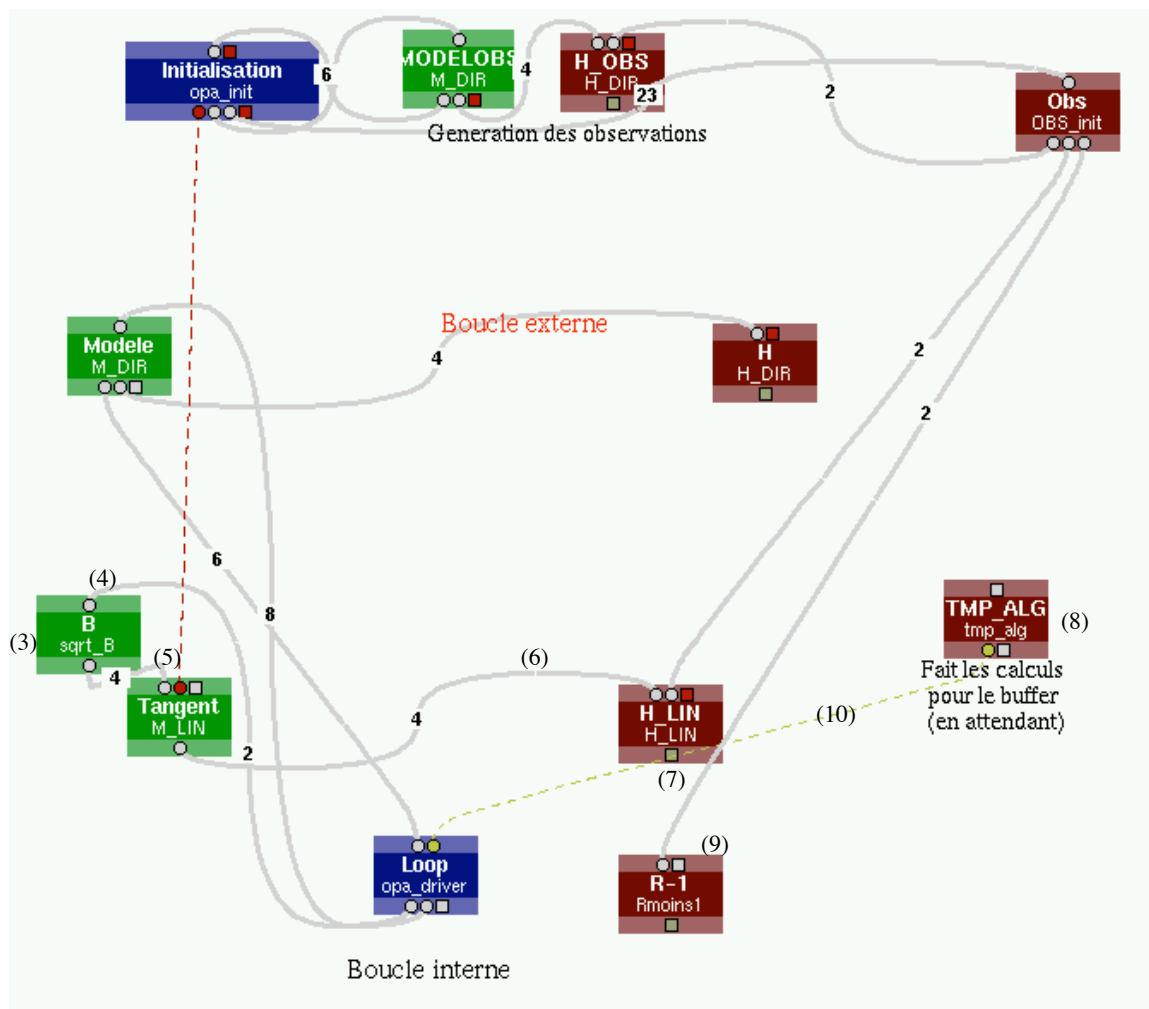


FIG. 4.4: Calcul de la fonction objective

Pour parachever la construction de l'algorithme, il ne reste plus qu'à effectuer le calcul du gradient. L'écart à l'innovation pondéré par les covariances d'erreur d'observation est récupéré par l'adjoint de l'opérateur d'observation (signet (11) de la figure 4.5) qui calcule ainsi le terme de forçage du modèle adjoint (12). A l'issue de l'intégration rétrograde de ce dernier, l'adjoint de l'opérateur de changement de variable (B_ADJ, (13)) s'applique à \mathbf{x}_0^* pour fournir le gradient de J_0 au minimiseur ce qui termine notre boucle interne (14).

Les branches

Dans ce qui précède on a décrit la création des unités et des différentes communications. Ces unités sont indépendantes de la méthode d'assimilation utilisée, et c'est leur agencement qui définira l'algorithme d'assimilation. Cet agencement est défini dans les branches générées avec l'aide de PréPALM. Pour la méthode d'assimilation décrite ici, elle sont au nombre de trois :

- La branche principale : elle gère l'ensemble de l'algorithme et envoyant divers signaux aux autres branches et effectue la minimisation (voir tableau 4.4).
- La branche modèle : elle effectue, suivant les ordres envoyés par la branche principale, l'intégration des modèles direct, tangent ou adjoint, ou elle applique l'opérateur réalisant l'inverse du changement de variable (ou son adjoint) (voir tableau 4.5).
- La branche des observations : elle applique les différents opérateurs d'observations, calcule la fonction coût et le terme de forçage du modèle adjoint (voir tableau 4.6).

Dans ces trois tableaux sont résumées les actions des branches, plusieurs "petites" communications ont été omises (les bornes temporelles et quelques *flags*).

Ces trois branches peuvent fonctionner en parallèle en échangeant des informations (objets) à certains moments. Pour éviter qu'une branche en avance sur les autres ne recommence une itération en utilisant des objets générés par une autre branche pour l'itération précédente, il est parfois nécessaire de synchroniser les différentes branches. Ceci est fait au moyen de `PALM_Step` insérés dans les branches. Chaque branche exécutant un `PALM_Step` reste bloquée tant que les autres ne sont pas arrivées au `PALM_Step` correspondant.

La mise au formalisme PALM d'une méthode d'assimilation alourdit quelque peu le codage de chaque unité mais celles ci ne nécessiteront plus d'être modifiées au gré des méthodes d'assimilation de données utilisées. Seuls les codes de branches devront changer et cela se fait aisément grâce à l'interface PréPALM. En d'autres termes construire les différentes unités PALM représente un surcroît de travail mais fournit un important gain de temps et une facilité accrue dans la manipulation et la construction des différentes méthodes.

Cet apport a permis de construire toute une série de méthodes d'assimilation (3D-VAR, 3D-FGAT, 3D-PSAS, 4D-VAR, 4D-VAR incrémental, 4D-VAR modèle imparfait) avec le modèle Shallow-Water et ce, en un temps record.

Code de branche	description
CALL PALMLaunch(Initialisation)	lancement de l'unité initialisation qui fera les calculs et les PALM_Put correspondants
indic='geneobs'	la prochaine étape sera la génération des observations
CALL PALM_Put(indic)	envoie l'information aux autres branches
FOR noutit = 1 TO nrefup DO	Boucle externe
indic='traj de ref'	la prochaine étape sera le calcul de la trajectoire de référence et du vecteur innovation
CALL PALM_Put(indic)	envoie l'information aux autres branches
CALL PALM_Put(x(t_deb))	envoi de la condition initiale de référence
FOR nit = 1 TO nitmax DO	boucle interne
indic = 'simulator'	la prochaine étape sera le calcul de la fonction objective et du gradient
CALL PALM_Put(indic)	envoie l'information aux autres branches
CALL PALM_Put(v)	envoi du vecteur de contrôle
CALL PALM_Get(Jo)	récupération de la valeur du terme J_o de la fonction objective depuis la branche observation
J=Jo+v ^T v	Calcul de la fonction objective
CALL PALM_Get(∇Jo)	récupération du gradient de J_o depuis la branche modèle
∇J=∇Jo+v	calcul du gradient
CALL PALM_Put(J, ∇J)	Envoi du gradient et de la fonction objective
CALL PALMLaunch(minim)	calcul du nouveau vecteur de contrôle
ENDDO	fin de la boucle interne
ENDDO	fin de la boucle externe
indic = 'fin'	on a terminé l'assimilation
CALL PALM_Put(indic)	envoie l'information aux autres branches

TAB. 4.4: Branche principale

Code de branche	description
WHILE indic .NE. 'fin'	tant que indic n'est pas égale à "fin"
CALL PALM_get(indic)	que doit faire cette branche ?
SELECT CASE(indic)	les différents cas
CASE('geneobs')	la génération des observations
CALL PALM_Launch('MODELOBS')	lancement de l'unité MODELOBS
CASE('traj de ref')	Calcul de la trajectoire de référence
CALL PALM_Launch(Modele)	lancement de l'unité modèle
CASE('simulator')	Calcul de la fonction objective et de son gradient
CALL PALM_Launch(B)	inverse du changement de variable
CALL PALM_Launch(Tangent)	lancement du modèle tangent
CALL PALM_Launch(Adjoint)	lancement du modèle adjoint
CALL PALM_Launch(B_ADJ)	adjoint de l'inverse du changement de variable
END SELECT CASE	

TAB. 4.5: Branche modèle

Code de branche	description
WHILE indic .NE. 'fin'	tant que indic n'est pas égale à "fin"
CALL PALM_get(indic)	que doit faire cette branche ?
SELECT CASE(indic)	les différents cas
CASE('geneobs')	la génération des observations
DO ib_i = t_deb, t_fin	boucle sur le temps
CALL PALM.Launch('H_OBS')	lancement de l'unité H_OBS
ENDDO	
CASE('traj de ref')	Calcul de la trajectoire de référence
DO ib_i = t_deb, t_fin	boucle sur le temps
CALL PALM.Launch(H)	lancement de l'opérateur d'observation qui récupérera directement l'état de l'unité modèle : calcul des \mathbf{d}_i
ENDDO	
CASE('simulator')	Calcul de la fonction objective et de son gradient
DO ib_i = t_deb, t_fin	boucle sur le temps
CALL PALM.Launch(H.LIN)	linéaire tangent de l'opérateur d'observation : calcul de $\mathbf{H}_i \delta \mathbf{x}_i$
ENDDO	
CALL PALM.Launch(R-1)	multiplications des $[\mathbf{H}_i \delta \mathbf{x}_i - \mathbf{d}_i]$ par la matrice \mathbf{R}^{-1} , envoi de J_o
DO ib_i = t_deb, t_fin	boucle sur le temps
CALL PALM.Launch(H.ADJ)	adjoint de l'inverse du changement de variable : envoi de $\mathbf{H}^T \mathbf{R}^{-1} [\mathbf{H}_i \delta \mathbf{x}_i - \mathbf{d}_i]$ au modèle adjoint
ENDDO	
END SELECT CASE	

TAB. 4.6: Branche observation

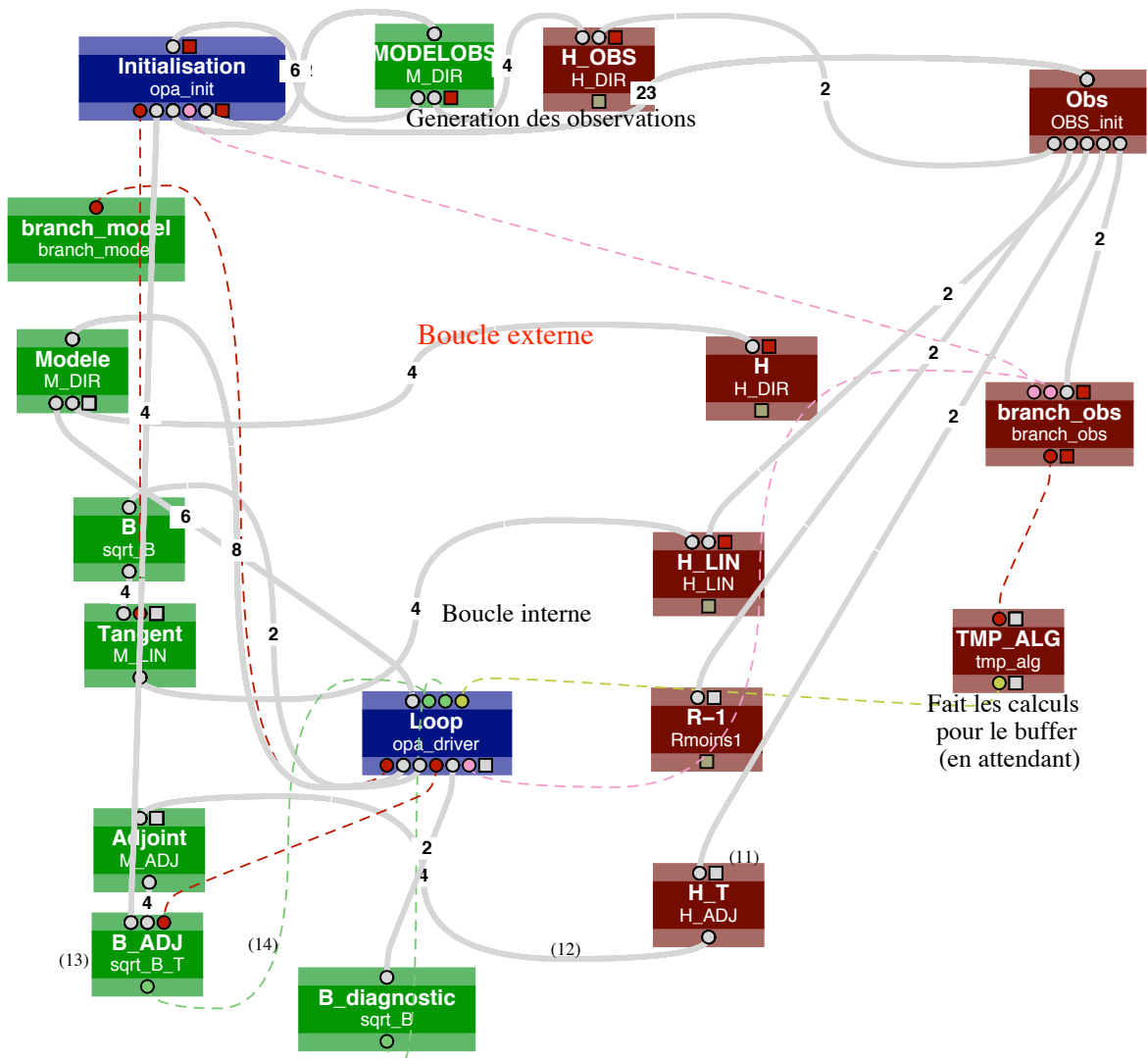


FIG. 4.5: La méthode complète

Deuxième partie

La Prise en compte des erreurs modèles en assimilation variationnelle - Études préliminaires

Chapitre 5

Contrôle de l'erreur modèle : méthode complète

Ce chapitre reprend quasiment intégralement l'article *4D-variational data analysis with imperfect model. Reduction of the size of control* [124], des incohérences avec les notations des précédents chapitres peuvent donc apparaître.

(Lorenc, 1986 [77]) a montré que pour un modèle linéaire décrivant exactement le comportement du système et un opérateur d'observation également linéaire, le 4D-Var et le filtre de Kalman donnent le même résultat à la fin de la fenêtre d'assimilation. Cependant dans le cas plus réaliste d'un modèle imparfait, le filtre de Kalman prend en compte les erreurs du modèle ce qui n'est pas le cas pour le 4D-Var.

L'étape suivante dans le développement des méthodes d'assimilation de données est la prise en compte des erreurs de modèle. Comme nous le montrerons dans la suite la méthode de contrôle d'erreur modèle dans le 4D-Var est beaucoup trop coûteuse dans sa forme complète pour des applications réalistes opérationnelles. C'est pourquoi on est amené à ne s'intéresser qu'à une partie seulement de l'erreur à corriger.

Dans le présent chapitre, nous appliquerons une technique classique pour prendre en compte cette erreur modèle (voir [JAZWINSKY [64], 1970] et [COHN [26], 1997]). et nous évoquerons une approche permettant de réduire de façon drastique la taille du contrôle et par là même le coût de la minimisation.

Le principe de cette méthode est de, partant du 4D-Var classique avec modèle imparfait, ne corriger les erreurs modèles que suivant certaines directions bien choisies. Nous appliquerons cette méthode dans le cas très simple des équations de Burger linéaires et nous verrons que celle ci fournit des résultats de bien meilleure qualité que le 4D-Var classique et ce pour un coût très raisonnable.

Cette approche n'est pas la seule pour prendre en compte de façon raisonnable les imperfections du modèle : nous nous intéresserons dans les deux chapitres suivants à d'autres alternatives proposées par GRIFFITH et al. [56], DEE et DA SILVA [37], D'ANDREA et VAUTARD 1999 [34] dans le cadre d'un contrôle de la part systématique de l'erreur (biais) seulement, ou par ZOU, NAVON et LE DIMET, 1992 [132] utilisant le résidu d'observation (nudging optimal).

5.1 La méthode complète

Comme il a été dit plus haut, la méthode opérationnelle pour les prévisions météorologiques est le contrôle optimal de l'ébauche¹. Malgré des résultats probants, cette approche comporte plusieurs limitations importantes. L'une d'entre elles est le fait que le modèle soit une contrainte forte. Ceci implique l'utilisation de modèles extrêmement sophistiqués, très proches du comportement réel du système, d'où un coût en calcul imposant (de plus le modèle n'est jamais rigoureusement exact). Le but principal du présent rapport sera donc d'étudier une méthode permettant de relâcher la contrainte et ainsi de permettre plus d'approximation dans le modèle. Pour ce faire, nous allons rajouter un terme de correction d'erreur au modèle, destiné à rattraper à chaque instant les incertitudes de celui-ci. Ce terme deviendra alors notre variable de contrôle et nous tenterons, par une méthode variationnelle, de minimiser l'écart entre la solution de ce modèle "faible" et les observations.

Rappel de la méthode complète

Le principe est très simple : on ajoute dans les équations le terme $B\eta$ représentant la correction de l'erreur du modèle.

$$\begin{cases} \frac{d\mathbf{x}}{dt} = M(\mathbf{x}) + B\eta \\ \mathbf{x}(0) = \mathbf{x}^b + \delta\mathbf{x}_0 \end{cases} \quad (5.1)$$

avec $\eta \in [\mathcal{C}(0, T)]^m$.

Maintenant le but de l'assimilation est de déterminer, en plus de la condition initiale, une estimation de ce terme correctif. Notre problème se traduit donc par la minimisation de la fonctionnelle $J(\delta\mathbf{x}_0, \eta)$ qui est la norme de la différence entre la prévision du modèle et les observations plus un terme de régularisation qui interdit au terme de contrôle d'être trop important.

$$J(\eta) = \frac{1}{2} (\| H(\mathbf{x}) - \mathbf{y}^o \|^2 + (N\eta, \eta))$$

N est utilisé pour pondérer l'importance du terme de régularisation dans la fonction objective. Dans le cas linéaire, à la condition que la matrice \mathbf{H} soit semi définie positive et que \mathbf{N} est symétrique définie positive, J est α -convexe sur un espace lui-même convexe. Il admet donc un minimum unique η^* (cf [75, LIONS]).

Considérons $\hat{\mathbf{x}}$ la dérivée directionnelle de \mathbf{x} dans la direction h_η , et \hat{J} la dérivée directionnelle de J . En modifiant les calculs du paragraphe 1.2, on obtient :

$$\begin{cases} \frac{d\hat{\mathbf{x}}}{dt} = \mathbf{M}_{|\mathbf{x}} \cdot \hat{\mathbf{x}} + Bh_\eta \\ \hat{\mathbf{x}}(0) = 0 \end{cases} \quad (5.2)$$

$$\hat{J}(\eta, h_\eta) = \int_0^T (\mathbf{H}^T (H(\mathbf{x}) - \mathbf{y}^o), \hat{\mathbf{x}}) + (N\eta, h_\eta) dt = (\nabla J, h_\eta) \quad (5.3)$$

¹condition initiale

On introduit \mathbf{x}^* l'état adjoint qui est le même que dans le cas précédent :

$$\begin{cases} -\frac{d\mathbf{x}^*}{dt} = \mathbf{M}_{|\mathbf{x}}^T \mathbf{x}^* + \mathbf{H}^T(H(\mathbf{x}) - \mathbf{y}^o) \\ \mathbf{x}^*(T) = 0 \end{cases} \quad (5.4)$$

De la même façon que précédemment, on obtient : $\nabla J = B^T \mathbf{x}^* + N\eta$.

C'est à dire : η est optimal pour $\eta = -N^{-1}B^T \mathbf{x}^*$

Ainsi on obtient le système d'optimalité :

$$\begin{cases} \frac{d\mathbf{x}}{dt} = M(\mathbf{x}) - BN^{-1}B^T \mathbf{x}^* + g(t) \\ \mathbf{x}(0) = U \\ -\frac{d\mathbf{x}^*}{dt} = \mathbf{M}_{|\mathbf{x}}^T \mathbf{x}^* + \mathbf{H}^T(H(\mathbf{x}) - \mathbf{y}^o) \\ \mathbf{x}^*(T) = 0 \end{cases} \quad (5.5)$$

Dans le cas d'un modèle linéaire, une comparaison entre le contrôle de l'erreur modèle et le filtre de Kalman (cf [123]) a montré que les choix adéquats pour N et B sont, respectivement, \mathbf{R} (la matrice de covariance d'erreur d'observation) et \mathbf{H}^T .

Dans cette section nous comparerons cette méthode et le 4D-Var en utilisant le modèle très simple des équations de Burger avec des conditions aux bord de Dirichlet homogènes.

On s'intéresse au modèle : $x \in]0, 1[, t > 0, \nu = 1.1, \alpha = 0.9$

$$\begin{cases} \frac{\partial u}{\partial t} + \alpha \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = g \\ u(0, t) = u(1, t) = 0 \\ u(x, 0) = \sin \pi x \end{cases} \quad (5.6)$$

pour lequel on connaît une solution exacte (cf 2.1).

5.1.1 Premiers résultats numériques

Afin de comparer l'efficacité de cette méthode on pourra agir, dans un premier temps, sur deux paramètres : en bruitant le terme de forçage g (soit par un bruit corrélé : un facteur multiplicatif, soit par un bruit blanc) et en sous échantillonnant les observations. Les résultats obtenus par le contrôle de l'erreur modèle seront comparés :

- Avec les résultats obtenus sans contrôle ("run libre").
- Avec les résultats obtenus en contrôlant seulement la condition initiale.
- Avec la solution exacte $e^{-t} \sin(\pi x)$.

Pour cette expérience nous utiliserons 5000 pas de temps ($\Delta t = \frac{1}{5000}$) et 20 pas de discrétisation en espace ($\Delta x = \frac{1}{20}$).

La méthode de minimisation utilisée ici est un gradient conjugué “maison”, la matrice hessienne de la fonction objective étant utilisé comme préconditionnement. Le produit Hessien-vecteur est calculé grâce à l'adjoint du second ordre (cf Le Dimet et Charpentier [24]). Le critère d'arrêt est $\frac{\|\nabla J\|}{\|\nabla J_0\|} < 10^{-5}$.

Observations non bruitées

Au lieu d'initialiser le modèle avec la condition initiale exacte $\sin \pi x$ et d'utiliser le terme de forçage correspondant à la solution exacte, on ajoute sur ces deux termes un bruit blanc de 10% d'amplitude

Ici le poids sur la fonction objective à été fixé à $N = 0$ (les observations étant exactes il n'y a pas de nécessité a restreindre les possibilité de variation du vecteur de contrôle) et $B = \text{Identité}$.

De plus on a des observations uniquement un pas de temps sur 100 et un pas de discrétisation en espace sur 2. Sous ces conditions l'algorithme de contrôle de l'erreur modèle converge en 25 itérations vers un résultat très satisfaisant même aux pas de temps qui n'ont pas été observés. Dans les mêmes conditions la méthode classique converge plus rapidement mais vers un résultat définitivement moins intéressant.

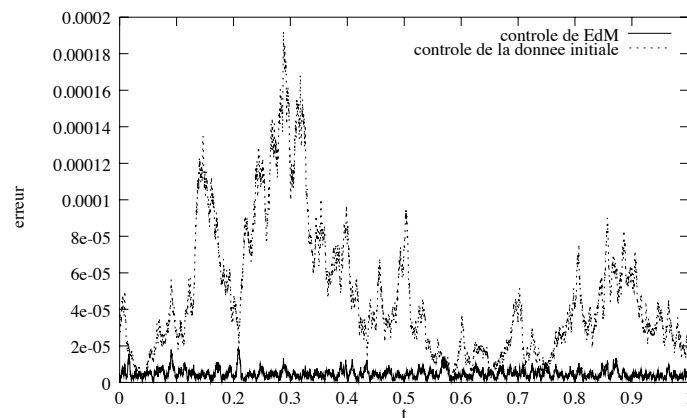


FIG. 5.1: rms de l'erreur en fonction du temps pour les deux méthodes

Premier bilan

Cette première expérience nous montre que dans le cas d'un modèle inexact, la méthode du contrôle de l'EdM permet de recomposer fidèlement l'évolution du système, ce qui n'est pas le cas de la méthode classique de contrôle de l'ébauche.

Il est vrai que les résultats obtenus sont plus qu'encourageants. Toutefois jusqu'à présent, les observations étaient **exactes** ce qui n'est évidemment jamais le cas en pratique et pose certains problèmes à notre méthode.

C'est ce que nous allons développer maintenant.

Prise en compte des bruits d'observation

On a vu dans les paragraphes précédents que l'algorithme de contrôle de l'erreur de modèle donnait de bons résultats sous couvert de perturbations acceptables du second membre ou de l'état initial. Le principe même du rajout d'un terme de correction d'erreur de modèle incite celui-ci à "coller" le plus possible aux observations. Un des problèmes de cette méthode vient donc du fait que, en pratique, les observations ne peuvent, en général pas être considérées comme exactes. Ceci implique que si l'on n'y prend garde, le contrôle de l'EdM suivra scrupuleusement l'évolution des observations. On obtiendra donc un résultat de très mauvaise qualité. Ceci est illustré par la figure 5.2 où l'on a assimilé des observations comportant un bruit blanc d'amplitude 1% sans prendre en compte le terme de régularisation. La solution contrôlée tend à rejoindre les observations d'où un résultat calamiteux.

Il faut donc éviter que la norme du terme de correction V ne devienne trop grande, c'est à dire en pratique augmenter le poids du terme de régularisation (NV, V) dans la fonctionnelle à minimiser. Le choix de N peut-être fait en utilisant l'analogie contrôle optimal-filtre de Kalman introduite au chapitre 2 i.e $N \longleftrightarrow R$ où R est la matrice de covariance d'erreur d'observation. Le bruit appliqué ici aux observations étant ici de 1%, une application directe de $N \longleftrightarrow R$ donnerait $N = 10^{-4} * I$. En pratique, on obtient de meilleurs résultats avec un poids plus important.

Voici deux exemples (figure 5.2) en ne bruitant que le second membre (bruit blanc de 10%) avec $N = 0$ et $N = 4.10^{-2} * I$. Alors que dans le premier cas la solution trouvée est impressionnante par sa mauvaise qualité, un choix plus judicieux pour la matrice N permet d'obtenir un résultat acceptable (l'erreur culmine à 5.10^{-5}).

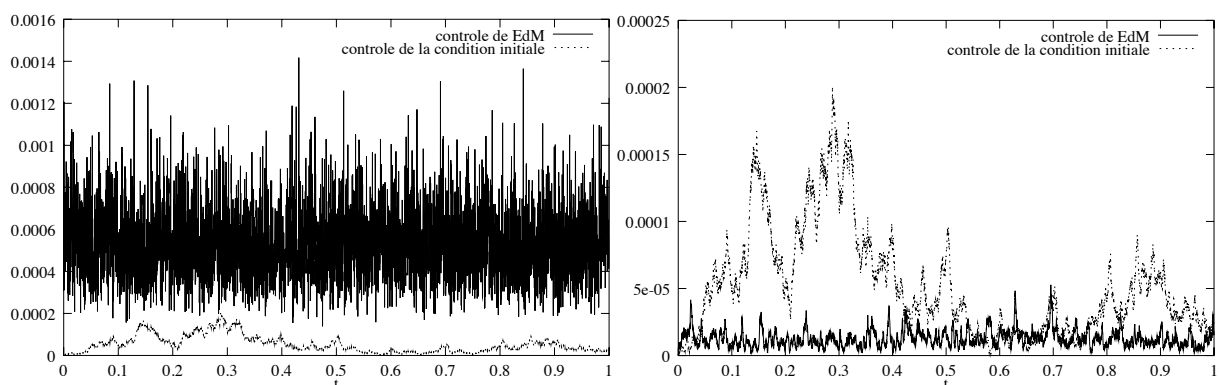


FIG. 5.2: Norme rms de l'erreur en fct du temps pour $N = 0$ et pour $N = 4.10^{-2} * Identité$

Dans le cas précédent on observait la totalité de la trajectoire; si on ne l'observe que partiellement (cf fig 5.3) la méthode donne des résultats catastrophiques. Il est donc temps de se souvenir des analogies avec le filtre de Kalman : en prenant $B = C^T$. Les résultats

alors obtenus sont redevenus acceptables (fig 5.4 et 5.5). On remarque cependant que, en fin de période, les résultats ne sont pas satisfaisants. Cela vient sans doute du fait que, à cet endroit, les erreurs sans contrôle sont déjà très petites (de l'ordre de 10^{-5}) et la méthode "hésite" entre minimiser l'écart aux observations et le terme de régularisation.

En faisant varier N on peut obtenir des résultats différents sur les différentes périodes. Une meilleure connaissance de l'erreur d'observation permettrait sans doute d'affiner le choix de N .

Prendre $B = C^T$ a de plus l'intérêt de réduire la taille du vecteur de contrôle. En effet celui-ci devient de la taille du vecteur d'observation.

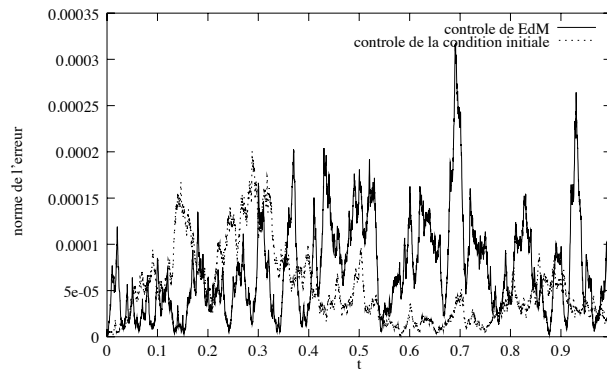


FIG. 5.3: Norme rms de l'erreur en fct du temps pour $N = 10^{-3} * \text{Identité}$ en n'observant seulement 1 pas de temps sur 50, avec $B = \text{Identité}$

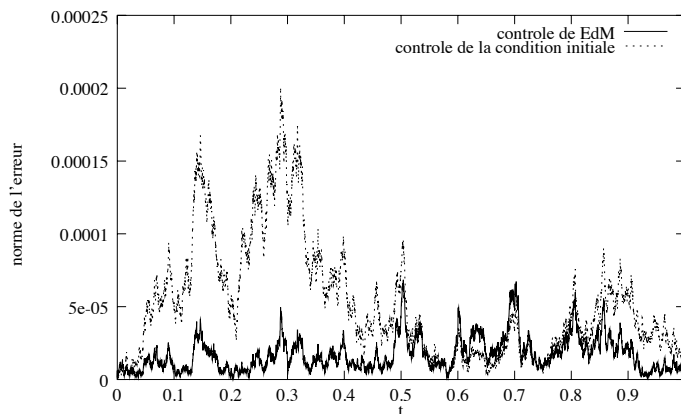


FIG. 5.4: Norme rms de l'erreur en fct du temps pour $N = 10^{-3} * \text{Identité}$ en observant seulement 1 pas de temps sur 50, avec $B = \mathbf{H}^T$

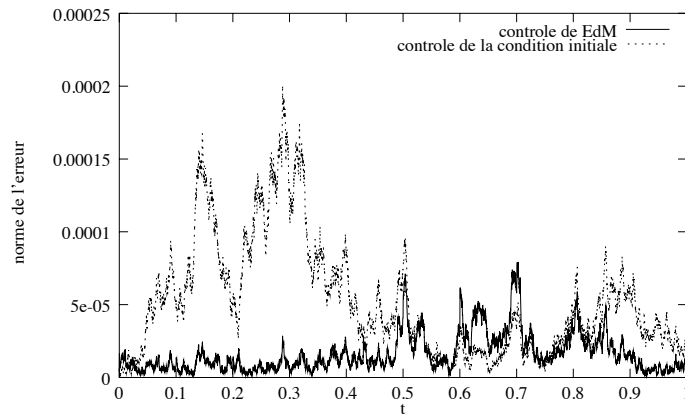


FIG. 5.5: Norme rms de l'erreur en fct du temps pour $N = 5.10^{-4} * \text{Identité}$ en observant seulement 1 pas de temps sur 50, avec $B = \mathbf{H}^T$

En présence d'erreur modèle il est évident que le contrôle de la condition initiale seule est insuffisant, cependant la méthode utilisée ici n'est pas applicable pour des modèles réalistes. En effet le vecteur de contrôle est multiplié par le nombre de pas de temps de la période d'assimilation. Compte tenu de la taille des problèmes océaniques ou atmosphériques le stockage et la manipulation d'un tel vecteur est hors de portée des moyens de calculs actuels. Dans la partie suivante on va introduire une approche visant à réduire la taille du vecteur de contrôle et l'illustrer sur la petite configuration utilisée dans ce chapitre.

5.2 Réduction de la taille de l'espace de contrôle

Jusqu'à présent on avait dans notre cas un vecteur de contrôle V de taille $20 \times 5000 = 100000$. En considérant la simplicité de notre modèle, on peut imaginer ce que ça donnerait pour des modèles de météorologie ou d'océanographie (typiquement $10^8 - 10^{10}$). Il est donc nécessaire pour des raisons évidentes de taille mémoire et de coût de calcul de réduire la taille de ce vecteur.

5.2.1 Principe

On s'intéresse toujours au problème

$$\begin{cases} \frac{d\mathbf{x}}{dt} = M(\mathbf{x}) + B\eta \\ \mathbf{x}(0) = \mathbf{x}_0 + \delta\mathbf{x}_0 \end{cases} \quad (5.7)$$

Mais on va essayer de restreindre la liberté de l'erreur de modèle en ne contrôlant plus que suivant certains vecteurs et en ne prenant que les premiers termes du développement en série de Fourier du terme d'erreur.

Le choix des directions suivant lesquelles on va contrôler est directement guidé par le modèle. En effet si \mathbf{x}_1 est la solution du modèle dans laquelle on a rajouté une petite perturbation h à $t = t_i$, c'est à dire:

$$\begin{cases} \frac{d\mathbf{x}_1}{dt} = M(\mathbf{x}_1) + B\eta \\ \mathbf{x}_1(t_i) = \mathbf{x}_i + h \end{cases} \quad (5.8)$$

alors l'erreur $\hat{\mathbf{x}} = \mathbf{x}_1 - \mathbf{x}$ sera solution de l'équation :

$$\begin{cases} \frac{d\hat{\mathbf{x}}}{dt} = \mathbf{M}_{|\mathbf{x}}\hat{\mathbf{x}} \\ \hat{\mathbf{x}}(t_i) = h \end{cases} \quad (5.9)$$

On voit aisément que les directions h suivant lesquelles les erreurs grandiront le plus rapidement seront celle qui maximisent le facteur d'amplification $A(h)$

$$A(h) = \frac{\langle \mathbf{M}h, \mathbf{M}h \rangle}{\langle h, h \rangle} = \frac{\langle \mathbf{M}^T \mathbf{M}h, h \rangle}{\langle h, h \rangle}$$

L'opérateur $\mathbf{M}^T \mathbf{M}$ étant symétrique les directions suivant lesquelles les perturbations vont le plus se développer seront les vecteurs propres de $\mathbf{M}^T \mathbf{M}$ associés aux plus grandes valeurs propres (LACARRA et TALAGRAND, 1987 [69]).

Ainsi si on décompose l'erreur dans la base engendrée par ces vecteurs propres, les composantes associées aux vecteurs propres de valeurs propres supérieures à 1 seront amplifiées par le modèle au cours du temps pendant que les autres auront tendance à disparaître.

Cette approche est dans le cas général valable seulement localement, mais notre cas test étant linéaire, elle reste valable globalement.

5.2.2 Application à notre problème

On peut utiliser ce résultat de la manière suivante :

- On choisit les m vecteurs propres χ_i de $\mathbf{M}_{|\mathbf{x}}^T \mathbf{M}_{|\mathbf{x}}$ correspondant aux m plus grandes valeurs propres de cette matrice. (au maximum, on se contente de prendre ceux dont les valeurs propres associées sont positives, i.e. dont l'exponentielle est plus grande que 1). Par exemple, dans notre cas les valeurs propres de $\mathbf{M}_{|\mathbf{x}}^T \mathbf{M}_{|\mathbf{x}}$ pour 20 pas de discrétisation en espace :

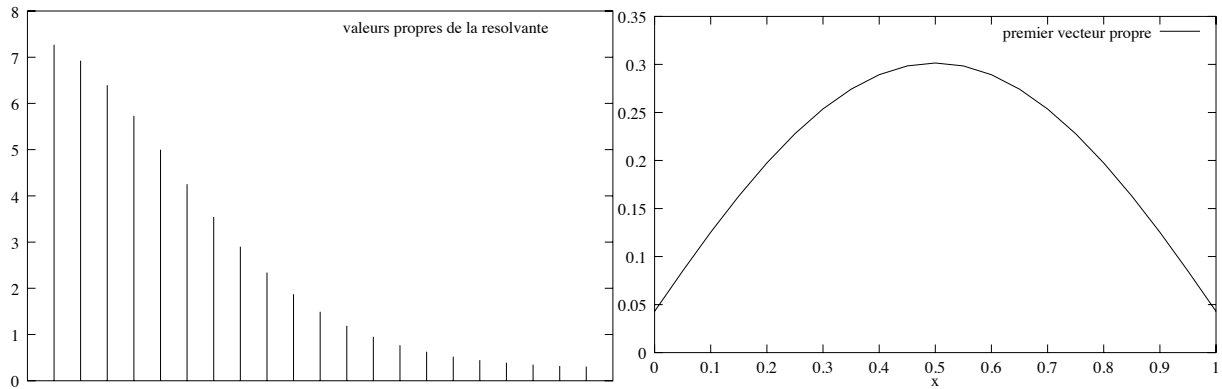


FIG. 5.6: Valeurs propres de $\mathbf{M}_{|\mathbf{x}}^T \mathbf{M}_{|\mathbf{x}}$ et premier vecteur propre

Les vecteurs propres et les valeurs propres associées ont été calculés grâce à une réduction de Householder et un algorithme QL.

- On choisit une base de fonction dépendant du temps $\lambda_j(t)$, par exemple les l premiers termes d'un développement en série de Fourier.

$$\lambda_j(t) = \sin(2\pi jt) \text{ ou } \cos(2\pi jt)$$

Et on cherche une erreur de modèle de la forme

$$\eta(x, t) = \sum_{i=1}^m \sum_{j=1}^l \alpha_{ij} \lambda_j(t) \chi_i(x)$$

Pour simplifier on notera $\eta = \Lambda(t)\alpha$ où α est le vecteur colonne composé des α_{ij} et Λ la matrice dont les colonnes sont les produits $\lambda_j \chi_i$.

On ne contrôle plus que sur α_{ij} . La fonction coût devient alors :

$$J(\alpha) = \frac{1}{2} \left(\int_0^T \| \mathbf{H}\mathbf{x} - \mathbf{x}_{obs} \|^2 + (N\Lambda\alpha, \Lambda\alpha) dt \right)$$

Calcul du gradient

De la même façon que pour la méthode complète et grâce au même modèle adjoint, on obtient le gradient :

$$\nabla J = \int_0^T (\Lambda^T \mathbf{x}^* + \Lambda^T N \Lambda \alpha) dt$$

5.2.3 Résultats numériques : Partie II

On se place dans le cas de l'expérience 5.2, c'est à dire qu'un bruit blanc non corrélé de 10% est appliqué sur le second membre et de 1% sur les observations en observant la totalité de la trajectoire, sans se préoccuper, dans un premier temps de la matrice N . Le choix du nombre de vecteurs propres et de fonctions de temps n'est pas évident car le plus n'est pas forcément le mieux : un petit nombre permet d'éviter de donner trop de liberté au contrôle mais un trop petit l'empêche de corriger suffisamment. Empiriquement : à la lecture du tableau 5.1 il apparaît qu'un bon compromis se situe à 2-3 vecteurs propres et une quinzaine de fonctions de temps. On a donc ramené la taille du vecteur de contrôle de 100000 à 30 !! On note (cf fig 5.7) qu'avec ce choix on obtient des résultats légèrement meilleurs que la méthode sans réduction !

Ce résultat peut paraître surprenant à première vue, mais il peut être expliqué : la décomposition sur une base réduite peut filtrer les erreurs d'observation et donc permettre d'obtenir de meilleurs résultats.

Nb de fonctions de temps	Nombre de vecteurs propres							
	2		3		4		5	
5	24	$1.54e^{-5}$	36	$1.50e^{-5}$	46	$1.51e^{-5}$	57	$1.53e^{-5}$
10	28	$1.4e^{-5}$	44	$1.41e^{-5}$	59	$1.45e^{-5}$	74	$1.49e^{-5}$
15	39	$1.33e^{-5}$	53	$1.35e^{-5}$	63	$1.40e^{-5}$	84	$1.45e^{-5}$
20	47	$1.35e^{-5}$	57	$1.4e^{-5}$	82	$1.46e^{-5}$	87	$1.53e^{-5}$

TAB. 5.1: Nombre d'itérations et Norme rms de l'erreur pour quelques valeurs du nombre de vecteurs propres et de la dimension de la base temporelle

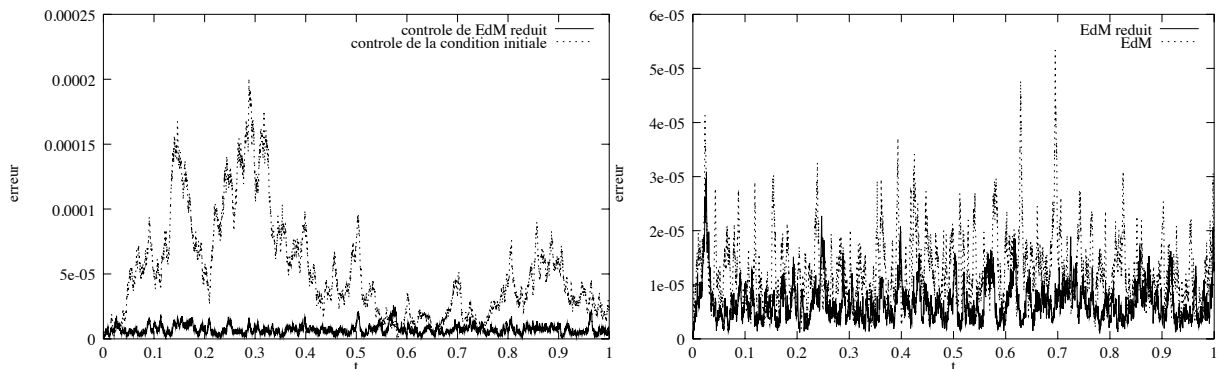


FIG. 5.7: Norme rms de l'erreur en fct du temps pour 15 fonctions de base en temps et pour 2 vecteurs propres

Comme c'était le cas précédemment, la méthode supporte le sous-échantillonnage combiné au bruitage des observations. Il est important de noter que le choix de "bonnes" matrices \mathbf{N} et \mathbf{B} donne quasiment **exactement** le même résultat que dans le chapitre 4 (fig 5.8) pour moitié moins d'itérations.

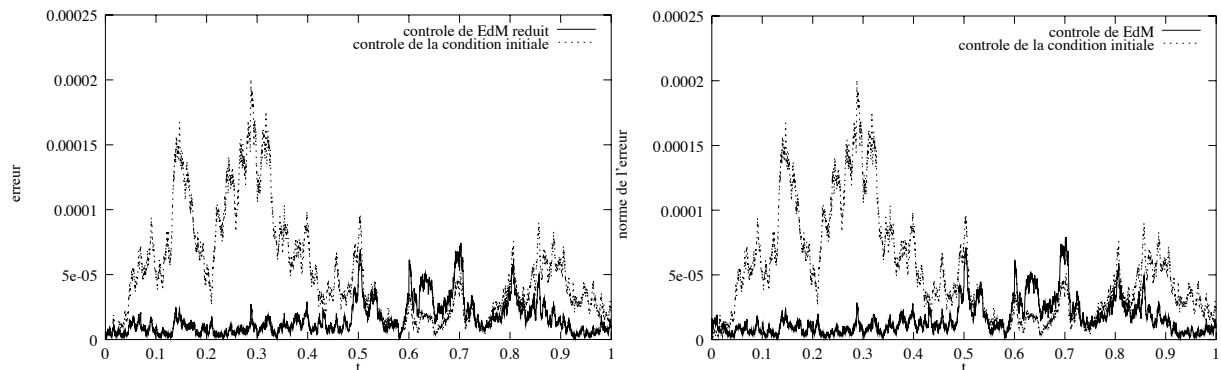


FIG. 5.8: Norme rms de l'erreur en fonction du temps avec des observations sous échantillonnées et bruitées. Comparaisons entre les méthodes de contrôle de l'erreur modèle entière et réduite

5.3 Conclusion

Cette étude préliminaire met en évidence la nécessité de prendre en compte, quand celle-ci est présente, l'erreur modèle dans les méthodes variationnelles d'assimilation de données. En effet la méthode classique de contrôle de la condition initiale ne permet pas de retrouver correctement l'état vrai. Dans des applications réalistes l'erreur modèle est moins importante que dans l'exemple proposé ici et donc la différence entre les deux approches est moins significative. Cependant, bien que moindre, l'erreur modèle est et sera toujours présente. La prise en compte de l'erreur modèle devra permettre d'obtenir une analyse de meilleure qualité et éventuellement de diagnostiquer certaines faiblesses du modèle.

Le surcoût numérique de l'approche complète décrite au début de ce chapitre la rend inutilisable pour des applications réalistes. Pour pouvoir prendre en compte l'erreur modèle il est donc nécessaire de réduire la taille de contrôle. Bien que prometteuse l'approche présentée en deuxième partie de ce chapitre et consistant à ne corriger l'erreur modèle que suivant certaines directions bien choisies ne sera pas développée dans la suite, elle représente une partie du travail de thèse de Sophie DURBIANO [42] qui a testé d'autres familles de directions privilégiées plus adaptées à des problèmes non linéaires (vecteurs singulier, vecteurs lyapunov, de breeding, etc.). Une des pistes explorées dans la suite sera de considérer qu'une information importante sur l'erreur modèle est contenue dans le vecteur innovation $\mathbf{d} = \mathbf{x}^a - \mathbf{y}^o$ et d'étudier comment cette information peut être utilisée pour corriger l'analyse. Cette approche sera traitée dans le chapitre suivant. Une autre approche consistera à ne conserver dans notre vecteur de contrôle que la première approximation d'un développement de l'erreur, c'est à dire un terme constant de la taille du

vecteur d'état. Cette méthode sera étudiée au chapitre "Contrôle de l'erreur systématique du modèle" et appliquée au modèle de circulation océanique générale OPA.

Chapitre 6

Le Nudging Optimal

Comme le chapitre précédent, celui ci reprend l'article *Determination of optimal nudging coefficients* [125]; cependant celui ci étant plus récent les notations doivent y être en accord avec le reste de cette thèse.

Le principe de la méthode complète de contrôle de l'erreur présentée dans le chapitre précédent est d'ajouter dans le vecteur de contrôle un terme de correction de l'erreur résiduelle qui est ajouté aux équations du modèle à chaque pas de temps. En raison de la taille de ce problème (la dimension de la variable d'état (typiquement 10^6 - 10^7) fois le nombre de pas de temps), cette approche est hors d'atteinte des moyens de calculs actuels. Afin de réduire le coût du contrôle de l'erreur modèle on a proposé de ne contrôler cette erreur que suivant certaines directions bien choisie (voir VIDARD *et al* , 2001 [124] et DURBIANO, 2001 [42]), ou seulement la part systématique de l'erreur (biais) comme l'ont proposé NICHOLS *et al*, 2000 [57].

Dans ce chapitre on va introduire une autre technique pour traiter des erreurs modèle, basée sur une idée de ZOU, NAVON et LE DIMET, 1992 [132]. Cette méthode est une extension de la méthode du *nudging* (ou relaxation newtonienne) définie au paragraphe 1.2.3 qui fut la première chaîne d'assimilation de données opérationnelle en océanographie (LYNE *et al*, 1982 [81], KRISH *et al*, 1991 [68], LORENC *et al*, 1991 [78]). Cela consiste en effectuer un rappel de l'état du modèle vers les observations pendant la période d'assimilation en ajoutant un terme de '*nudging*' aux équations du modèle.

Le terme de *nudging* est défini par la différence entre les observations et la solution du modèle, multiplié par un coefficient de *nudging*.

Le *nudging* peut être vu comme une approximation du filtre de Kalman (KF), les coefficients de *nudging* étant ceux qu'on obtiendrait avec KF dans le cas linéaire (cf LORENC, 1986 [77] et LORENC *et al*, 1991 [78]). Cependant, en pratique le KF est très coûteux (stockage et manipulation de matrices $N \times N$ où N est la taille du vecteur d'état) en comparaison du très peu coûteux *nudging*. De plus, de façon à être appliqué, le KF requiert d'effectuer des simplifications qui font que la période d'assimilation ne sera pas prise en compte dans son ensemble.

Dans ce chapitre nous utiliserons une méthode de contrôle optimal pour estimer à la fois la condition initiale (comme dans le 4D-Var) et les coefficients de *nudging* afin de corriger les erreurs modèle de façon plus efficace (d'un point de vue calcul).

Dans un premier temps nous rappellerons les équations du 4D-Var et introduirons le

nudging optimal, et nous montrerons que cette nouvelle méthode et le filtre de Kalman sont équivalents dans un sens qui reste à définir. La section 2 présentera une assimilation de données 4D-Var *nudging* optimal appliqué au “petit” modèle des équations de Burger. Trois formes de matrice de *nudging* seront testées afin de déterminer quel compromis entre la taille du contrôle et la qualité du résultat sera optimal.

Finalement cette méthode sera appliquée sur le modèle Shallow-Water et comparée à une assimilation 4D-Var classique sur la même configuration. On montrera alors qu’au prix d’un petit surcoût cette méthode fournit une amélioration sensible au 4D-Var.

6.1 Motivations

Le but de l’approche classique de l’assimilation de données 4D-Var et de tenter de réduire la différence entre les observations et l’état du modèle sur une période de temps donnée en contrôlant la condition initiale de cette période de temps.

Si on a le modèle discrétisé M , soit $\mathbf{x} \in \mathcal{C} \subset \mathbb{R}^N$, \mathbf{x}^b l’ébauche, ou *first guess* de la minimisation, l’évolution de l’état du système peut être décrit par :

$$\begin{cases} \mathbf{x}_0 = \mathbf{x}^b + \delta\mathbf{x}_0 \\ \mathbf{x}_{i+1} = M_{(t_i, t_{i+1})}(\mathbf{x}_i) \end{cases} \quad (6.1)$$

Le but de la méthode est de rechercher $\delta\mathbf{x}_0$ qui minimise la fonction objective :

$$J(\delta\mathbf{x}_0) = \overbrace{\frac{1}{2} \sum_{i=0}^n \langle \mathbf{R}^{-1}(H_i(\mathbf{x}_i) - \mathbf{y}_i^o), H_i(\mathbf{x}_i) - \mathbf{y}_i^o \rangle}^{J_o} + \overbrace{\frac{1}{2} \langle \mathbf{B}^{-1} \delta\mathbf{x}_0, \delta\mathbf{x}_0 \rangle}^{J_b}$$

\mathbf{R} est la matrice $m \times m$ de covariance d’erreur d’observation et \mathbf{B} la matrice $N \times N$ de covariance d’erreur d’ébauche, $\mathbf{y}_i^o \in \mathbb{R}^m$ est le vecteur d’observation au temps t_i et H_i est l’opérateur d’observation qui calcule l’équivalent dans l’espace des observations de l’état du modèle au temps t_i .

Le premier terme de la fonction objective (appelé J_o) représente l’écart aux observations que nous désirons minimiser dans un cas idéal de modèle décrivant exactement l’évolution du système et d’observations exactes. Cependant en raison de la présence d’erreur sur les observations et dans les équations du modèle ce terme ne peut pas, et ne doit pas être réduit à 0. C’est pourquoi on introduit le second terme (J_b). Celui ci empêche la solution de trop s’éloigner d’un état physiquement cohérent ce qui générerait des oscillations afin de retrouver un équilibre physique. L’équilibre entre les deux termes de la fonction objective est fourni par les inverses des matrices de covariance d’erreur \mathbf{B}^{-1} et \mathbf{R}^{-1} qui représentent la confiance qu’on a dans les termes d’ébauche et d’observation respectivement.

En raison de la présence du terme J_b et des erreurs modèle, le résidu d’innovation $\mathbf{d}_k = H_k(\mathbf{x}_k^a) - \mathbf{y}_k^o$ reste non nul après une assimilation 4D-Var.

Les sources d’erreur peuvent être écrites comme :

- Erreur de prévision :

$$\varepsilon_k^f = \mathbf{x}_k^f - \mathbf{x}_k^t$$

- Erreur d'observation :

$$\varepsilon_k^o = \mathbf{y}_k^o - H_k(\mathbf{x}_k^t)$$

où \mathbf{x}_k^f représente l'état du modèle au temps t_k et \mathbf{x}_k^t l'état vrai.

On remarque que le résidu d'innovation peut être écrit :

$$\begin{aligned} \mathbf{d}_k &= H_k(\mathbf{x}_k^f) - \mathbf{y}_k^o \\ &\approx \mathbf{H}_k \varepsilon_k^f - \varepsilon_k^o \end{aligned} \quad (6.2)$$

où $\mathbf{H}_k = \frac{\partial H_k}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{x}_k^f}$ est l'approximation linéaire tangente de l'opérateur d'observation, et ε^f est supposé négligeable comparé à \mathbf{x}^f

L'équation (6.2) montre qu'une importante information sur les erreurs de modèle et d'observation est contenue dans le résidu d'innovation. Dans la suite de ce chapitre on essaiera d'utiliser cette information pour améliorer les résultats du schéma d'assimilation 4D-Var.

Formellement, dans le cas d'un état complètement et directement observé, une approche intuitive serait de modifier la prévision en utilisant un terme de correction fonction du résidu d'innovation. Dans ce cas, l'inverse de l'opérateur d'observation \mathbf{H}^{-1} existe et on peut écrire :

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{H}_k^{-1}(\mathbf{d}_k - \varepsilon^o)$$

Le KF (KALMAN, 1960 [66]) produit un état analysé utilisant le vecteur innovation afin de corriger la prévision par

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k \mathbf{d}_k$$

où \mathbf{K} est appelée matrice de gain et calculée par :

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

avec \mathbf{P}^f la matrice de covariance d'erreur de prévision.

On peut souligner que dans le filtre de Kalman, si les observations sont supposées exactes ($\mathbf{R} = 0$), on retrouve la matrice de gain par : $\mathbf{K}_n = \mathbf{H}_n^{-1}$ et $\varepsilon_n^f = \mathbf{H}_n^{-1} \mathbf{d}_k$

Cependant, en raison du coût trop élevé du filtre de Kalman, on va, dans la suite de ce chapitre, tenter d'estimer une matrice de gain "optimale" en utilisant une méthode d'assimilation de données variationnelle (4D-Var) et en étendant la méthode de nudging optimal introduit par ZOU, NAVON et LE DIMET en 1992 [132].

6.2 La Méthode

Dans cette section on introduira le point principal de ce chapitre : le schéma d'assimilation de données 4D-Var nudging optimal (4D-ON). Les calculs du gradient et de la fonction coût ne requièrent pas d'outil supplémentaire par rapport au 4D-Var classique mais seulement quelques calculs et stockages supplémentaires.

Le principe des méthodes de nudging est d'effectuer un rappel de l'état du modèle vers les observations en ajoutant un terme de nudging dans les équations. Ce terme est l'écart entre les observations et l'état du modèle $\mathbf{y}_{i+1}^o - H_{i+1}(M_{(t_{i+1}, t_i)}(\mathbf{x}_i))$ pondéré par l'opérateur de nudging \mathbf{K}_i . Dans cette partie on considérera le cas d'une méthode de nudging "complète" c'est à dire que \mathbf{K}_i est vue comme une matrice $N \times m$ où N est la dimension du vecteur d'état et m la dimension du vecteur d'observation. Le problème devient : comment estimer \mathbf{K}_i .

Soit M le modèle discret non linéaire décrivant l'évolution de l'océan ou de l'atmosphère. On s'intéresse alors au problème :

$$\begin{cases} \mathbf{x}_0 = \mathbf{x}^b + \delta \mathbf{x}_0 \\ \mathbf{x}_{i+1} = M_{(t_{i+1}, t_i)}(\mathbf{x}_i) + \mathbf{K}_{i+1}(\mathbf{y}_{i+1}^o - H_{i+1}(M_{(t_{i+1}, t_i)}(\mathbf{x}_i))) \end{cases} \quad (6.3)$$

Et on désire minimiser la fonction objective :

$$\begin{aligned} J(\delta \mathbf{x}_0, \mathbf{K}) = & \frac{1}{2} \sum_{i=0}^n \langle \mathbf{R}^{-1} H_i(\mathbf{x}_i) - \mathbf{y}_i^o, H_i(\mathbf{x}_i) - \mathbf{y}_i^o \rangle + \frac{1}{2} \langle \mathbf{B}^{-1} \delta \mathbf{x}_0, \delta \mathbf{x}_0 \rangle \\ & \underbrace{+ \frac{1}{2} \sum_{i=1}^n \langle \mathbf{Q}_i^{-1} \mathbf{K}_i(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))), \mathbf{K}_i(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) \rangle}_{J_{nudz}} \end{aligned} \quad (6.4)$$

Basiquement, on considérera une fonction objective du 4D-Var avec \mathbf{R} la matrice de covariance d'erreur d'observation et \mathbf{B} la matrice de covariance d'erreur d'ébauche, à laquelle on ajoute un terme de régularisation J_{nudz} avec \mathbf{Q} , la matrice de covariance d'erreur de prévision.

Le terme de régularisation provient de la formulation classique de la fonction objective du 4D-Var avec modèle imparfait (cf JAZWINSKY [64], 1970 et COHN [26], 1997). Dans cette formulation la fonction objective est :

$$\begin{aligned} J(\delta \mathbf{x}_0, \eta_1, \dots, \eta_n) = & \frac{1}{2} \sum_{i=0}^n \langle \mathbf{R}^{-1} H_i(\mathbf{x}_i) - \mathbf{y}_i^o, H_i(\mathbf{x}_i) - \mathbf{y}_i^o \rangle \\ & + \frac{1}{2} \sum_{i=1}^n \langle \mathbf{Q}_i^{-1} \eta_i, \eta_i \rangle + \frac{1}{2} \langle \mathbf{B}^{-1} \delta \mathbf{x}_0, \delta \mathbf{x}_0 \rangle \end{aligned} \quad (6.5)$$

où η_i est le terme de correction du modèle au temps t_i . Ainsi η_i à simplement été remplacé par le terme de correction par nudging.

Afin de calculer le gradient de la fonction objective on calcul le Lagrangien sous contrainte de (6.4). Soit L le Lagrangien définit par :

$$L(\mathbf{K}, \delta \mathbf{x}_0; \mathbf{x}^*) = J(\mathbf{K}, \delta \mathbf{x}_0) + \sum_{i=1}^n \langle \mathbf{x}_i^*, \mathbf{x}_i - M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}) - \mathbf{K}_i(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) \rangle$$

\mathbf{x}^* étant le multiplicateur de Lagrange. En calculant les dérivées partielles

$$\frac{\partial L}{\partial \mathbf{x}^*}(\mathbf{K}, \mathbf{v}; h_{\mathbf{x}}^*) = \mathbf{x}_i - M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}) - \mathbf{K}_i^T(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))$$

On peut tout d'abord retrouver :

$$\frac{\partial L}{\partial \mathbf{x}^*} = 0 \implies (6.3)$$

De plus,

$$\begin{aligned} \frac{\partial L}{\partial \delta \mathbf{x}_0}(\mathbf{K}, h_{\delta \mathbf{x}_0}, \mathbf{x}^*) &= \sum_{i=0}^n \langle \mathbf{H}_i^T (H_i(\mathbf{x}_i) - \mathbf{y}_i^o), \hat{\mathbf{x}}_i \rangle + \langle \mathbf{B}^{-1} \delta \mathbf{x}_0, h_{\delta \mathbf{x}_0} \rangle \\ &- \sum_{i=1}^n \langle \mathbf{M}_{(t_i, t_{i-1})}^T \mathbf{H}_i^T \mathbf{K}_i^T \mathbf{Q}^{-1} \mathbf{K}_i (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))), \hat{\mathbf{x}}_{i-1} \rangle \\ &- \sum_{i=1}^n \langle \mathbf{x}_i^*, \hat{\mathbf{x}}_i \rangle + \langle \mathbf{M}_{(t_i, t_{i-1})}^T \mathbf{x}_i^*, \hat{\mathbf{x}}_{i-1} \rangle - \langle \mathbf{M}_{(t_i, t_{i-1})}^T \mathbf{H}_i^T \mathbf{K}_i^T \mathbf{x}_i^*, \hat{\mathbf{x}}_{i-1} \rangle \\ &- \langle \mathbf{x}_0^*, h_{\delta \mathbf{x}_0} \rangle \end{aligned}$$

Ainsi, si l'état adjoint \mathbf{x}^* est défini par :

$$\begin{cases} \mathbf{x}_n^* = \mathbf{H}_n^T \mathbf{R}^{-1} (H_n(\mathbf{x}_n) - \mathbf{y}_n^o) \\ \mathbf{x}_i^* = (\mathbf{M}_{(t_{i+1}, t_i)} - \mathbf{K}_{i+1} \mathbf{H}_{i+1} \mathbf{M}_{(t_{i+1}, t_i)})^T \mathbf{x}_{i+1}^* + \mathbf{H}_i^T \mathbf{R}^{-1} H_i(\mathbf{x}_i) - \mathbf{y}_i^o \\ - \mathbf{M}_{(t_{i+1}, t_i)}^T \mathbf{H}_{i+1}^T \mathbf{K}_{i+1}^T \mathbf{Q}^{-1} \mathbf{K}_{i+1} (\mathbf{y}_{i+1}^o - H_{i+1}(M_{(t_{i+1}, t_i)}(\mathbf{x}_i))) \end{cases} \quad (6.6)$$

On obtient $\frac{\partial L}{\partial \delta \mathbf{x}_0}(\mathbf{K}, h_{\delta \mathbf{x}_0}, \mathbf{x}^*) = \langle \nabla_{\delta \mathbf{x}_0} J, h_{\delta \mathbf{x}_0} \rangle$ c'est à dire :

$$\nabla_{\delta \mathbf{x}_0} J = -\mathbf{x}^*(0) + \mathbf{B}^{-1} \delta \mathbf{x}_0 \quad (6.7)$$

Et finalement :

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{K}_i}(\mathbf{v}, h_{\mathbf{K}_i}, \mathbf{x}_i^*) &= \langle \mathbf{Q}^{-1} \mathbf{K}_i (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))), h_{\mathbf{K}_i} (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) \rangle \\ &- \langle \mathbf{x}_i^*, h_{\mathbf{K}_i} (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) \rangle \\ &= \langle \nabla_{\mathbf{K}_i} J, h_{\mathbf{K}_i} \rangle \end{aligned} \quad (6.8)$$

Ainsi le gradient de la fonction objective par rapport à \mathbf{K}_i peut être écrit :

$$\begin{aligned} \nabla_{\mathbf{K}_i} J &= \mathbf{Q}^{-1} \mathbf{K}_i (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))^T \\ &- \mathbf{x}^*(t) (\mathbf{y}^o(t_i) - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))^T \end{aligned} \quad (6.9)$$

A condition que la contribution de la condition initiale au gradient (équation (6.2)) ait été calculée (grâce à une intégration du modèle direct et de son adjoint) le calcul de la contribution du terme de nudging ne requiert pas de calculs "lourds" supplémentaires. En effet $\nabla J_{\mathbf{K}_i}$ est obtenu par le produit de vecteurs utilisés pour le calcul de $\nabla J \delta \mathbf{x}_0$ (soit $\mathbf{K}_i (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))$, $(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))^T$ et \mathbf{x}^*).

6.3 Équivalence avec le filtre de Kalman

Dans cette partie on mettra en évidence que, dans un cas linéaire, le filtre de Kalman et la méthode de nudging optimal introduite dans la section précédente fourniront les mêmes résultats. Cette équivalence est fondée sur une démonstration fournie par LI et NAVON ([131]) pour un plus classique 4D-Var.

Considérons un schéma de nudging optimal pour un modèle linéaire discret où les \mathbf{K}_i sont considérées comme étant des matrices $n \times m$

$$\begin{cases} \mathbf{x}_0 = \mathbf{x}^b + \delta \mathbf{x}_0 \\ \mathbf{x}_{i+1} = \mathbf{M}_{(t_i, t_{i+1})} \mathbf{x}_i + \mathbf{K}_{i+1} (\mathbf{y}_{i+1}^o - \mathbf{H}_{i+1} (\mathbf{M}_{(t_i, t_{i+1})} \mathbf{x}_i)) \end{cases} \quad (6.10)$$

Par souci de simplicité on écrit $\mathbf{d}_i = \mathbf{y}_i^o - \mathbf{H}_i (\mathbf{M}_{(t_i, t_i)} \mathbf{x}_{i-1})$ et $\mathbf{c} = (\delta \mathbf{x}_0, \mathbf{K}_1, \dots, \mathbf{K}_n)$. On désire minimiser la fonction objective :

$$J(\mathbf{c}) = \frac{1}{2} \sum_{i=0}^n \langle \mathbf{R}^{-1} (\mathbf{H}_i (\mathbf{x}_i) - \mathbf{y}_i^o), \mathbf{H}_i (\mathbf{x}_i) - \mathbf{y}_i^o \rangle + \frac{1}{2} \sum_{i=0}^n \langle \mathbf{Q}_i^{-1} \mathbf{K}_i \mathbf{d}_i, \mathbf{K}_i \mathbf{d}_i \rangle + \langle \mathbf{B}^{-1} \delta \mathbf{x}_0, \delta \mathbf{x}_0 \rangle$$

où $\langle \cdot, \cdot \rangle$ représente le produit scalaire Euclidien ‘‘classique’’, et (\cdot, \cdot) le produit scalaire de Frobenius entre deux matrices : $(A, B) = \sum_j \sum_i a_{ij} b_{ij}$.

Tout d’abord, on peut noter qu’il existe une relation entre ces deux produits scalaires. Si \mathbf{X} est un vecteur de taille n , \mathbf{Y} un vecteur de taille m et \mathbf{B} une matrice $n \times m$, alors

$$\langle \mathbf{X}, \mathbf{B}\mathbf{Y} \rangle = (\mathbf{X}, \mathbf{Y}^T, \mathbf{B})$$

Si les erreurs modèle sont considérées comme étant Gaussiennes, LI et NAVON ([131]) ont démontré la propriété d’additivité du 4D-Var, c’est à dire que résoudre le problème d’optimisation sur l’ensemble de la période d’assimilation de données fournit le même résultat que si l’on découpe notre période d’assimilation en deux sous espaces $1..n_1$ et $n_1 + 1..n$ et si on définit deux fonctions objectives :

$$J_1(\mathbf{c}) = \frac{1}{2} \langle \mathbf{B}^{-1} \delta \mathbf{x}_0, \delta \mathbf{x}_0 \rangle + \frac{1}{2} \sum_{i=0}^{n_1} (\dots)$$

et

$$J_2(\mathbf{c}) = \frac{1}{2} \langle (\mathbf{P}_{\mathbf{x}_{n_1}}^a)^{-1} (\mathbf{x}_{n_1} - \hat{\mathbf{x}}_{n_1}), \mathbf{x}_{n_1} - \hat{\mathbf{x}}_{n_1} \rangle + \frac{1}{2} \sum_{i=n_1+1}^n (\dots)$$

où $\mathbf{P}_{\mathbf{x}_{n_1}}^a$ est la matrice de covariance d’erreur d’analyse issue de la première sous-période d’assimilation (qui peut être obtenue par le calcul du Hessian de J_1 (RABIER *et al* [98])) et $\hat{\mathbf{x}}_{n_1}$ est l’état analysé (Le résultat de cette assimilation).

Afin de pouvoir comparer le schéma de nudging optimal au filtre de Kalman on suppose que l’assimilation sur un premier sous espace d’observation $\{\mathbf{y}_1, \dots, \mathbf{y}_{k-1}\}$ a déjà été effectuée, et ainsi on connaît $\mathbf{P}_{\mathbf{x}_{n_1}}^a$ et $\hat{\mathbf{x}}_{n_1}$.

Notre problème se ramène donc à minimiser la fonction objective :

$$\begin{aligned}
J_k(\mathbf{c}_k) &= \frac{1}{2} \left\langle (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} (\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}), \mathbf{x}_{n_1} - \hat{\mathbf{x}}_{n_1} \right\rangle \\
&+ \frac{1}{2} \left\langle \mathbf{R}^{-1} (\mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k^o), \mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k^o \right\rangle \\
&+ \left\langle \mathbf{Q}_k^{-1} \mathbf{K}_k \mathbf{d}_k, \mathbf{K}_k \mathbf{d}_k \right\rangle
\end{aligned} \tag{6.11}$$

En utilisant

$$\mathbf{x}_{k-1} = \mathbf{M}_{k,k-1}^{-1} (\mathbf{x}_k - \mathbf{K}_k \mathbf{d}_k)$$

La fonction objective peut être réécrite comme :

$$\begin{aligned}
J_k &= \frac{1}{2} \left\langle (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} (\mathbf{M}_{k,k-1}^{-1} (\mathbf{x}_k - \mathbf{K}_k \mathbf{d}_k) - \hat{\mathbf{x}}_{k-1}), \mathbf{M}_{k,k-1}^{-1} (\mathbf{x}_k - \mathbf{K}_k \mathbf{d}_k) - \hat{\mathbf{x}}_{k-1} \right\rangle \\
&+ \frac{1}{2} \left\langle \mathbf{R}^{-1} \mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k^o, \mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k^o \right\rangle \\
&+ \frac{1}{2} \left\langle \mathbf{Q}_k^{-1} \mathbf{K}_k \mathbf{d}_k, \mathbf{K}_k \mathbf{d}_k \right\rangle
\end{aligned} \tag{6.12}$$

Ainsi notre problème devient : trouver $(\hat{\mathbf{x}}_k, \hat{\mathbf{K}}_k)$ tel que

$$\begin{cases} \frac{\partial J_k}{\partial \mathbf{x}_k} = 0 \\ \frac{\partial J_k}{\partial \mathbf{K}_k} = 0 \end{cases} \tag{6.13}$$

En utilisant (6.3), (6.13) devient :

$$\begin{cases} 0 = \mathbf{H}_k^T \mathbf{R}^{-1} (\mathbf{H}_k(\hat{\mathbf{x}}_k) - \mathbf{y}_k^o) + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} [\mathbf{M}_{k,k-1}^{-1} (\hat{\mathbf{x}}_k - \hat{\mathbf{K}}_k \mathbf{d}_k) - \hat{\mathbf{x}}_{k-1}] \\ 0 = \mathbf{Q}_k^{-1} \hat{\mathbf{K}}_k \mathbf{d}_k \mathbf{d}_k^T + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} \mathbf{M}_{k,k-1}^{-1} \hat{\mathbf{K}}_k \mathbf{d}_k \mathbf{d}_k^T \\ + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} [\hat{\mathbf{x}}_{k-1} - \mathbf{M}_{k,k-1}^{-1} \hat{\mathbf{x}}_k] \end{cases} \tag{6.14}$$

Grâce à (6.14) on obtient

$$\hat{\mathbf{K}}_k \mathbf{d}_k = \mathcal{H}_{\mathbf{Q}_k}^{-1} \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} [\hat{\mathbf{x}}_{k-1} - \mathbf{M}_{k,k-1}^{-1} \hat{\mathbf{x}}_k] \tag{6.15}$$

où $\mathcal{H}_{\mathbf{Q}_k} = \mathbf{Q}_k^{-1} + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} \mathbf{M}_{k,k-1}^{-1}$

Utiliser (6.15) dans (6.14) amène à

$$\begin{aligned}
0 &= \mathbf{H}_k^T \mathbf{R}^{-1} (\mathbf{y}_k^o - \mathbf{H}_k(\hat{\mathbf{x}}_k)) \\
&+ \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} (I + \mathbf{M}_{k,k-1}^{-1} \mathcal{H}_{\mathbf{Q}_k}^{-1} \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1}) [\hat{\mathbf{x}}_{k-1} - \mathbf{M}_{k,k-1}^{-1} \hat{\mathbf{x}}_k]
\end{aligned}$$

On peut le réécrire

$$\begin{aligned}
& [\mathbf{H}_k^T \mathbf{R}^{-1} \mathbf{H}_k + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} \mathbf{M}_{k,k-1}^{-1} \\
& + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} \mathbf{M}_{k,k-1}^{-1} \mathcal{H}_{\mathbf{Q}_k}^{-1} \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} \mathbf{M}_{k,k-1}^{-1}] [\mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1} - \hat{\mathbf{x}}_k] \\
& = \mathbf{H}_k^T \mathbf{R}^{-1} [\mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1} - \mathbf{y}_k^o] \tag{6.16}
\end{aligned}$$

Si on pose

$$\mathbf{P}_k^f = \mathbf{M}_{k,k-1} \mathbf{P}_{\mathbf{x}_{k-1}}^a \mathbf{M}_{k,k-1}^T + \mathbf{Q}_k \tag{6.17}$$

En utilisant un lemme d'inversion matricielle¹ on a :

$$\begin{aligned}
\mathcal{H}_{\mathbf{Q}_k}^{-1} & = \mathbf{Q}_k - \mathbf{Q}_k (\mathbf{P}_k^f)^{-1} \mathbf{Q}_k \\
& = \mathbf{M}_{k,k-1} \mathbf{P}_{\mathbf{x}_{k-1}}^a \mathbf{M}_{k,k-1}^T (\mathbf{P}_k^f)^{-1} \mathbf{Q}_k \tag{6.18}
\end{aligned}$$

En utilisant (6.18) dans (6.16) et (6.17) $\times (\mathbf{P}_k^f)^{-1}$

$$[\mathbf{H}_k^T \mathbf{R}^{-1} \mathbf{H}_k + (\mathbf{P}_k^f)^{-1}] [\mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1} - \hat{\mathbf{x}}_k] = \mathbf{H}_k^T \mathbf{R}^{-1} [\mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1} - \mathbf{y}_k^o] \tag{6.19}$$

on peut obtenir en utilisant une autre formule classique de calcul matriciel² :

$$\hat{\mathbf{x}}_k = \mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1} + \mathbf{P}_k^f \mathbf{H}_k^T [\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}]^{-1} [\mathbf{y}_k^o - \mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1}]$$

Ce qui est exactement l'expression de $\hat{\mathbf{x}}_k$ dans les équations du filtre de Kalman. Cela nous montre que, en théorie et quand le modèle est linéaire, le nudging optimal donne les mêmes résultats que le filtre de Kalman qui est optimal.

6.4 Expériences préliminaires

Dans cette section nous allons appliquer le schéma d'assimilation de données nudging optimal sur l'équation unidimensionnelle de Burger. Ce pourra être une première validation simple de ce schéma et permettra de faire un choix approprié quant à la forme des \mathbf{K}_i

Il est à noter que, comme c'était le cas dans le chapitre précédent, une solution analytique exacte est associée à un terme de forçage f fixé

Pour cette expérience on fixe $f(x, t) = e^{-t}(\pi u \cos \pi x + (-1 + \nu \pi^2) \sin \pi x)$, et ainsi $u = e^{-t} \sin \pi x$ sera une solution exacte de (2.1).

Le 4D-ON est défini en ajoutant le terme de nudging directement dans les équations du modèle :

$$\begin{cases} \frac{\partial u}{\partial t} + \alpha \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = f + g(u_{obs} - h(u)) \\ u(0, t) = u(1, t) = 0 \\ u(x, 0) = \sin \pi x \end{cases} \tag{6.20}$$

¹ $(\mathbf{A}^T \mathbf{B}^{-1} \mathbf{A} + \mathbf{C}^{-1})^{-1} = \mathbf{C} - \mathbf{C} \mathbf{A}^T (\mathbf{A} \mathbf{C} \mathbf{A}^T + \mathbf{B})^{-1} \mathbf{A} \mathbf{C}$

² $(\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{C}^{-1} = \mathbf{A} \mathbf{B}^T (\mathbf{C} + \mathbf{B} \mathbf{A} \mathbf{B}^T)^{-1}$ si les inverses existent (cf Wunsch [130], 1996 pp 99)

On s'intéressera à une fenêtre temporelle $[0, T]$ discrétisée en $N + 1$ pas de temps. De même pour le domaine spatial $[0, 1]$ discrétisé en E points.

Durant cette séquence temporelle on définit M observations aux points (x_m, t_m) $m = 1..M$.

Ainsi on peut réécrire nos équations :

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = f + \sum_{m=1}^M g_m (u_m^{obs} - u_m) \delta(x - x_m) \delta(t - t_m)$$

en formulation discrète :

$$\frac{d\mathbf{u}}{dt}(t) + M(\mathbf{u})(t) = F(t) + \sum_{k=1}^K \mathbf{K}_k (\mathbf{u}_k^{obs} - H(\mathbf{u}_k)) \delta(t - t_k)$$

où H est l'opérateur d'observation.

Pour cette expérience la minimisation sera effectuée par une méthode de gradient conjuguée "maison" utilisant le Hessien comme préconditionneur. Le produit Hessien-vecteur est calculé à l'aide de l'adjoint du modèle au second ordre (voir [LE DIMET *et al*, [24], 1998])

On va appliquer simultanément deux types d'erreur modèle sur le terme de forçage f : un bruit blanc de 5% et un bruit corrélé (un biais) de 10%. Les observations sont obtenues à partir de la solution analytique, seulement un point sur trois en espace et un pas de temps sur 50 sont retenus comme des observations et on ajoute un bruit de 5% sur ces observations.

Le but de cette expérience est d'analyser le comportement de cette méthode selon trois types de matrices de coefficients de nudging différents :

1. \mathbf{K}_k est une matrice pleine $E \times M$, c'est à dire qu'à chaque point le terme de correction est une combinaison linéaire de tous les écarts prévision/observation au même temps. c'est le cas idéal, car chaque composante du vecteur d'état sera corrigée et pas seulement celles qui seront observées. Même si cette forme de \mathbf{K}_k n'est applicable qu'à des problèmes de très petite taille, cette expérience permet de comparer le résultat de ce cas idéal avec de plus simples formes de la matrice de nudging et de mettre en évidence que le plus n'est pas forcément le mieux.
2. \mathbf{K}_k est une matrice diagonale. Dans ce cas les équations du modèle ne sont corrigées qu'aux points d'observation et les coefficients sont différents pour chaque observation.
3. \mathbf{K}_k est un coefficient scalaire (variant seulement avec le temps). C'est le même cas que précédemment, mais pour un temps fixé t_k le coefficient de correction est le même à tous les points d'observation.

Utiliser une de ces trois formes de \mathbf{K}_k augmente la taille du vecteur de contrôle (comparé au 4D-Var classique dont la taille de contrôle est E) respectivement par $E \times M \times N$, $M \times N$ et N .

Dans le tableau (6.1) sont montrés le niveau d'erreur rms et le coût en calcul pour ces trois différentes formes de 4D-ON. La première colonne présente les résultats bruts de l'assimilation et la seconde est obtenue en utilisant une interpolation linéaire en temps du terme de correction (après la minimisation) (équation [6.21]).

$$\mathbf{u}(t) = \mathbf{u}(t) + \mathbf{K}_k(\mathbf{u}_k^{obs} - H(\mathbf{u}(t))) \text{ if } t = t_k$$

$$\mathbf{u}(t) = \mathbf{u}(t) + \left(\frac{t-t_k}{t_{k+1}-t_k} \mathbf{K}_{k+1} + \frac{t_{k+1}-t}{t_{k+1}-t_k} \mathbf{K}_k \right) \left(\frac{t-t_k}{t_{k+1}-t_k} \mathbf{u}_{k+1}^{obs} + \frac{t_{k+1}-t}{t_{k+1}-t_k} \mathbf{u}_k^{obs} - H(\mathbf{u}(t)) \right) \text{ if } t_k < t < t_{k+1} \quad (6.21)$$

Forme de \mathbf{K}	Taille de contrôle	Brut		Interpolée	
		rms	CPU	rms	CPU
\mathbf{K} est une matrice pleine	$E \times M \times N$	15,97	203,96s	10,64	205,29s
\mathbf{K} est une matrice diagonale	$M \times N$	21,68	5,92s	7,19	5,94s
\mathbf{K} est un coefficient scalaire	N	20,78	2,83s	6,67	2,84s
Sans assimilation		90,98	0,1s		
4D-Var		72,26	1,78s		

TAB. 6.1: Norme de l'erreur et temps CPU pour les deux méthodes selon la forme de \mathbf{K}

Après cette première expérience on peut noter que le 4D-ON “complet” (\mathbf{K}_k matrice pleine) ne fournit pas réellement de meilleurs résultats que les autres variantes. Même si dans le cas de la méthode “brute” la matrice pleine est un peu plus efficace que les formes plus simple de \mathbf{K}_k , la différence ne justifie pas l'important surcoût en calcul induit (le contrôle d'une matrice pleine est plus de 70 fois plus coûteux que le contrôle de coefficients scalaires !). De plus dans les résultats de la méthode interpolée (smoothed 4D-ON) la matrice diagonale et les coefficients scalaires donnent de meilleurs résultats. Cela peut s'expliquer par le trop grand nombre de degrés de liberté de la méthode complète. On constate ici que plus le vecteur de contrôle est simple, plus le conditionnement du problème d'optimisation est bon, ce qui implique que la minimisation est plus efficace. De plus on peut noter que la matrice diagonale et les coefficients scalaires donnent à peu près les mêmes résultats ce qui pourrait indiquer que le vecteur innovation contient des informations sur les corrélations spatiales, c'est à dire que le vecteur innovation est lui-même bien équilibré sur la grille entière (les coefficients de nudging optimal sont presque les mêmes pour tous les points observés au même temps donné).

6.5 Expérimentation avec le modèle Shallow-Water

6.5.1 Le Modèle

Pour cette deuxième expérience nous allons utiliser le modèle Shallow-Water introduit dans la section 2.1.2.

La condition initiale (first guess) de notre fenêtre d'assimilation de données est issue de six ans de spin-up du modèle de prévision alors que l'état initial "vrai" et les observations sont simulés à partir d'un spin-up du modèle "vrai".

6.5.2 L'assimilation

La fonction objective est définie par (6.4) mais afin d'améliorer l'efficacité de la minimisation on opère un changement de variable,

$$\mathbf{v} = \mathbf{B}^{-\frac{1}{2}} \delta \mathbf{x}_0 \quad (6.22)$$

où $\mathbf{B}^{-\frac{1}{2}}$ est défini par $A = A^{\frac{1}{2}} A^{\frac{T}{2}}$ et construit suivant la méthode décrite au paragraphe 3.4.

On construit \mathbf{Q} de la même façon, en changeant les écarts types ($\Sigma_{\mathbf{Q}}$) et éventuellement les paramètres de l'équation de diffusion(\mathbf{C}).

De la même façon, pour ce problème plus réaliste on ne peut pas réellement contrôler une matrice \mathbf{G} entière mais seulement quelques coefficients. Cela implique que les corrections seront appliquées seulement aux points observés. Afin de pallier ce défaut on peut prémultiplier le terme de correction par $\mathbf{Q}^{\frac{T}{2}}$ ce qui lissera la correction et simplifiera le calcul de la fonction objective.

On peut réécrire l'équation (6.10)

$$\begin{cases} \mathbf{x}_0 = \mathbf{x}^b + \mathbf{B}^{\frac{T}{2}} \mathbf{v} \\ \mathbf{x}_{i+1} = M_{(t_{i+1}, t_i)}(\mathbf{x}_i) + \mathbf{Q}^{\frac{T}{2}} \mathbf{G}_{i+1}(\mathbf{y}_{i+1}^o - H_{i+1}(M_{(t_{i+1}, t_i)}(\mathbf{x}_i))) \end{cases} \quad (6.23)$$

Et la fonction objective

$$\begin{aligned} J(\mathbf{v}, \mathbf{G}) &= \frac{1}{2} \sum_{i=0}^n \langle \mathbf{R}^{-1} H_i(\mathbf{x}_i) - \mathbf{y}_i^o, H_i(\mathbf{x}_i) - \mathbf{y}_i^o \rangle \\ &+ \frac{1}{2} \sum_{i=1}^n \langle \mathbf{G}_i(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))), \mathbf{G}_i(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) \rangle \\ &+ \frac{1}{2} \langle \mathbf{v}, \mathbf{v} \rangle \end{aligned} \quad (6.24)$$

Ce changement de variable modifie les équations adjointes

$$\begin{cases} \mathbf{x}_n^* = \mathbf{H}_n^T \mathbf{R}^{-1} (H_n(\mathbf{x}_n) - \mathbf{y}_n^o) \\ \mathbf{x}_i^* = (\mathbf{M}_{(t_{i+1}, t_i)} - \mathbf{Q}^{\frac{T}{2}} \mathbf{G}_{i+1} \mathbf{H}_{i+1} \mathbf{M}_{(t_{i+1}, t_i)})^T \mathbf{x}_{i+1}^* + \mathbf{H}_i^T \mathbf{R}^{-1} H_i(\mathbf{x}_i) - \mathbf{y}_i^o \\ \quad - \mathbf{M}_{(t_{i+1}, t_i)}^T \mathbf{H}_{i+1}^T \mathbf{G}_{i+1}^T \mathbf{G}_{i+1} (\mathbf{y}_{i+1}^o - H_{i+1}(M_{(t_{i+1}, t_i)}(\mathbf{x}_i))) \end{cases} \quad (6.25)$$

Le termes du gradient deviennent

$$\begin{cases} \nabla_{\mathbf{v}} J = \mathbf{B}^{\frac{1}{2}} \mathbf{x}_0^* + \mathbf{v} \\ \nabla_{\mathbf{G}_i} J = \mathbf{Q}^{\frac{1}{2}} \mathbf{x}_i^* (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))^T \\ \quad + \mathbf{G}_i (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))^T \end{cases} \quad (6.26)$$

Ce deuxième changement de variable ne peut réellement être considéré comme un bon préconditionnement car $\mathbf{Q}^{\frac{T}{2}}$ n'est pas la racine carrée de l'inverse de la hessienne de $J_{nudging}$. Ainsi l'efficacité de la minimisation peut être améliorée en considérant $J_{nudging} = \frac{1}{2} \sum_{i=1}^n \langle \tilde{\mathbf{Q}}^{-1} \mathbf{G}_i, \mathbf{G}_i \rangle$ (avant le changement de variable) où \mathbf{G}_i a la taille du vecteur d'état.

Bien que le problème d'optimisation soit modifié et l'équivalence avec Kalman soit perdue, l'assimilation donne à peu près la même qualité de résultat comparé à la précédente forme du terme $J_{nudging}$ et le nombre d'itération est réduit de façon drastique (à peu près 1/10).

La méthode de minimisation employée dans la suite est un quasi-Newton à mémoire limitée N1QN3 développé à l'INRIA par *Gilbert and Lemaréchal* 1993.

6.5.3 Expérience numérique

Dans cette section on va comparer les résultats du 4D-ON avec ceux du 4D-Var classique. L'assimilation est appliquée sur des périodes de 30 jours. Les observations sont simulées comme étant des traces satellites (voir fig[6.1]) avec une trace tous les trois jours. Ce n'est pas un cas réellement favorable pour le nudging en raison du manque d'information fournie par les observations. En effet seulement 500 points de grille sont observés alors que la taille du vecteur de contrôle est d'à peu près 20000.

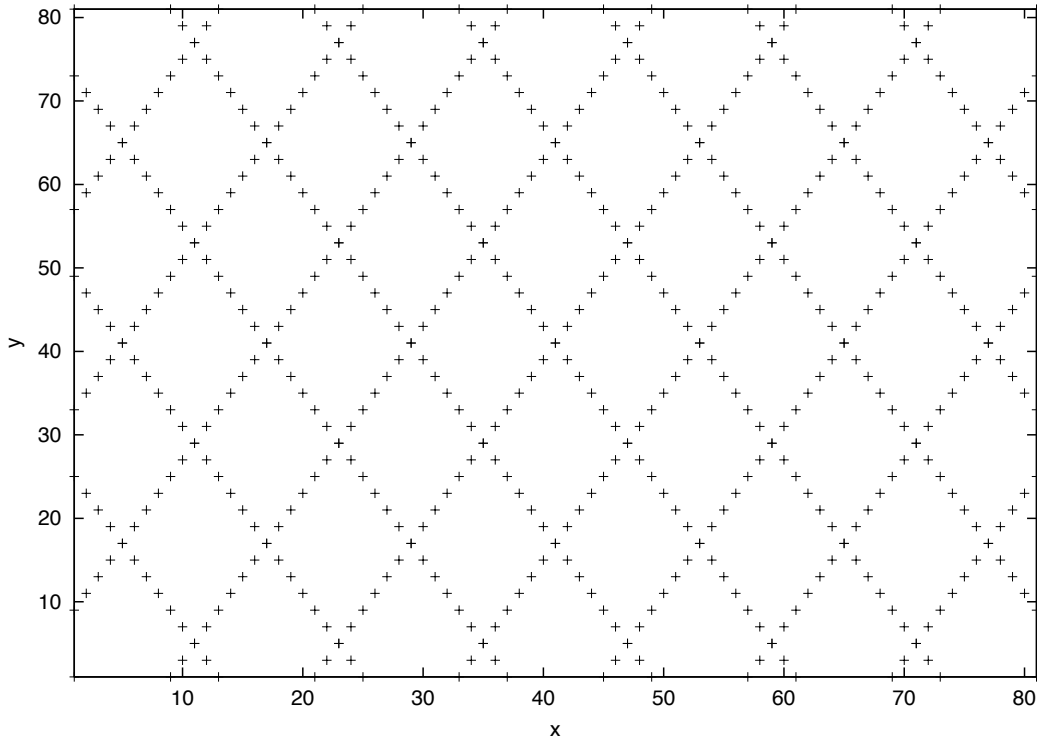


FIG. 6.1: *Positions des observations (traces satellites)*

Ainsi la version précédente du 4D-ON (ligne pointillés longs) peut donner de très mauvais résultats en comparaisons avec le 4D-Var (ligne pleine) et le smoothed 4D-ON (avec

changement de variable)(ligne pointillé courts, voir fig[6.2]). Le 4D-ON “brut” ne corrige le modèle qu’aux points de grille observés. Ainsi ces corrections ne sont pas consistantes avec les autres coordonnées. Du fait de cette inconsistance le modèle peut engendrer des ondes de gravité afin de retourner à un état d’équilibre (oscillations déplaisantes de la figure 6.2). De plus le terme J_o est seulement calculé à partir des composantes du vecteur d’état corrigé et ne représente pas réellement l’écart entre la réalité et l’état du modèle. En effet l’état analysé complet peut très bien être un champ très lisse avec quelques Dirac sur les points observés et le terme J_o sera très petit. On peut corriger en partie ce défaut, en restreignant la liberté des coefficients de nudging (en donnant plus de poids au terme J_{nudg}) mais alors on limitera l’apport du 4D-ON comparé au 4D-Var.

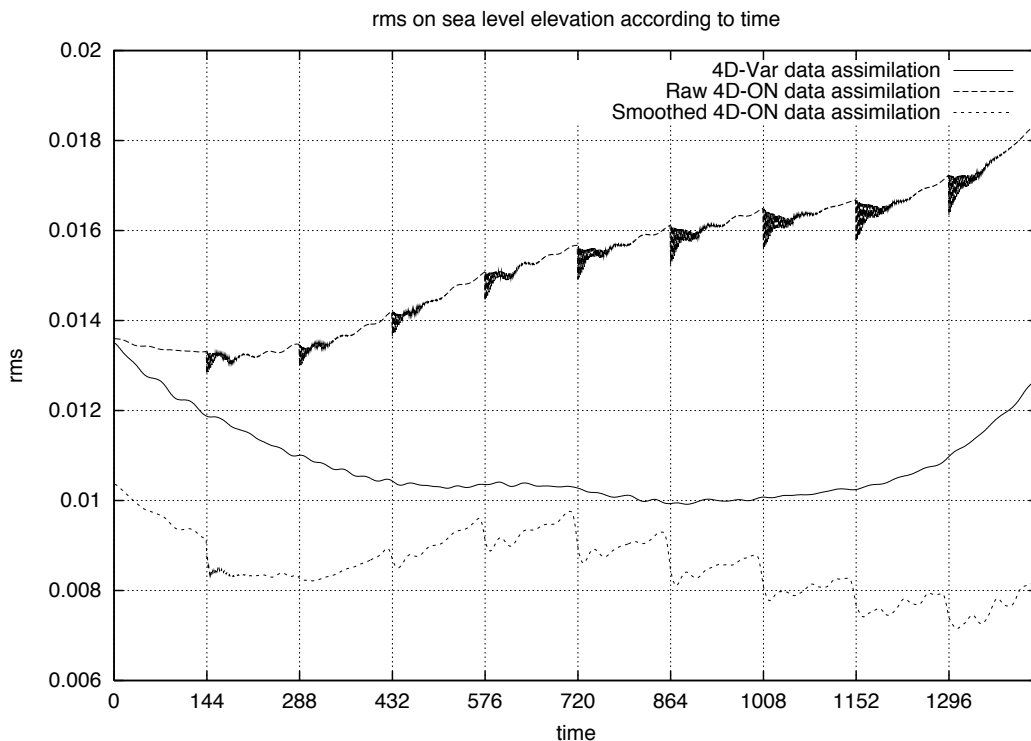


FIG. 6.2: Norme rms de l’erreur sur la SSH pour la version lissée du 4D-ON comparée au 4D-Var et au 4D-ON “brut”

La version “lissée” du 4D-ON permet le sous échantillonnage des observations en remplaçant une correction de style “Dirac” par une correction Gaussienne (avec l’aide de $\mathbf{Q}^{\frac{1}{2}}$). En plus du lissage spatial, un lissage temporel est appliqué au terme de correction en utilisant un coefficient décroissant en temps. Tous ces développements ont permis au 4D-ON d’améliorer sensiblement les résultats du 4D-Var pour le même nombre d’itérations. Sur la composante observée (élévation du niveau de la mer) l’amélioration apportée par le 4D-ON est très significative (jusqu’à 50% d’erreur rms en moins, voir figure 6.2). Sur les autres composantes (vitesse de courant, figure 6.3), grâce à l’opérateur de balance inclus dans \mathbf{Q} , le gain en rms reste de 10%.

L’introduction du lissage implique, pour ce modèle Shallow-Water, que le coût en calcul d’une itération augmente de 10% par rapport au 4D-Var classique. En comparaison du

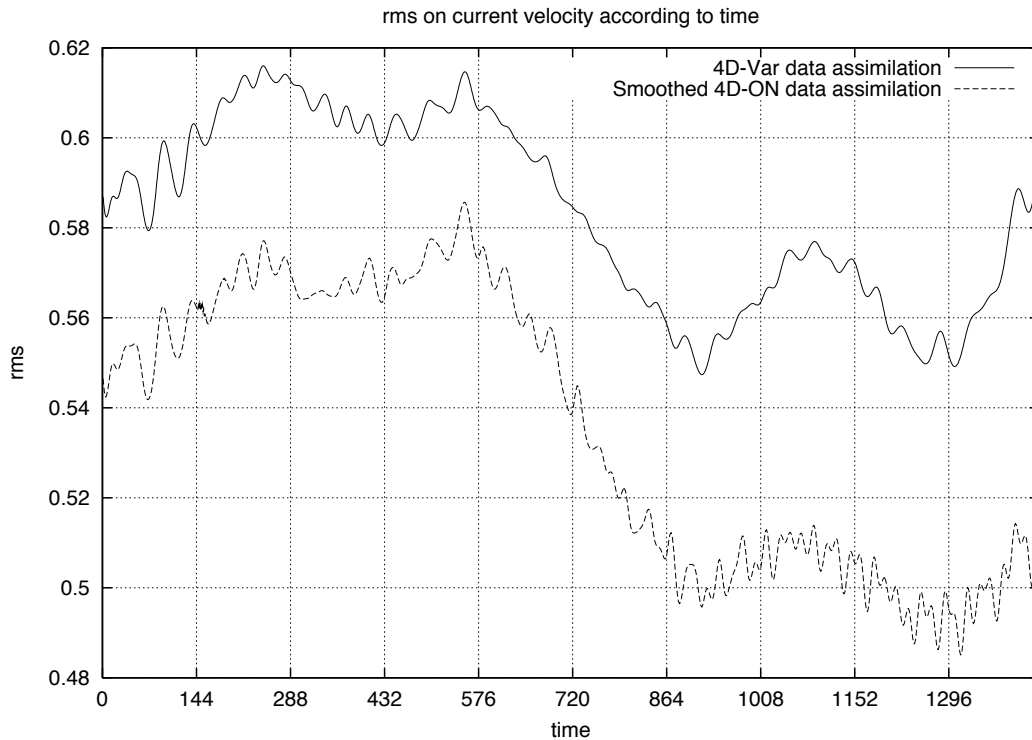


FIG. 6.3: Norme rms de l'erreur sur les vitesses de courant pour la version lissée du 4D-ON comparée au 4D-Var

gain fournit par le 4D-ON ce surcoût n'est pas une limitation.

En raison des chocs induits par les corrections de nudging, on pourrait craindre que l'état analysé à la fin de la période d'assimilation 4D-ON ne soit pas consistant avec la dynamique du modèle et ainsi ne pourra pas fournir une prévision de qualité. Les figures 6.5 et 6.4 montrent que l'on peut être rassuré : après la période d'assimilation (sur la partie gauche des figures) on effectue un mois supplémentaire de prévision en partant de l'état analysé à la fin de la période d'assimilation (partie droite). En raison des erreurs modèle la norme rms de l'erreur augmente petit à petit, mais la dérive est à peu près la même pour le 4D-Var classique que pour le smoothed 4D-ON. L'état analysé obtenu par nudging optimal est donc aussi équilibré que celui obtenu par un 4D-Var classique (mais plus près de l'état vrai).

6.6 Conclusion

Dans ce chapitre on a présenté une nouvelle méthode d'assimilation de donnée dérivée du 4D-Var et du nudging. On a montré que sous condition d'effectuer le changement de variable introduit dans la section 6.5.2, le 4D-ON fournit des résultats très encourageants. En effet grâce aux lissages spatial et temporel le 4D-ON donne des résultats très sensiblement meilleurs que le 4D-VAR. De plus cette méthode est très facile à obtenir en partant d'un algorithme 4D-Var existant (peu de modifications algorithmiques et tous les nouveaux termes sont déjà calculés) même si l'aspect théorique est quelque peu compliqué.

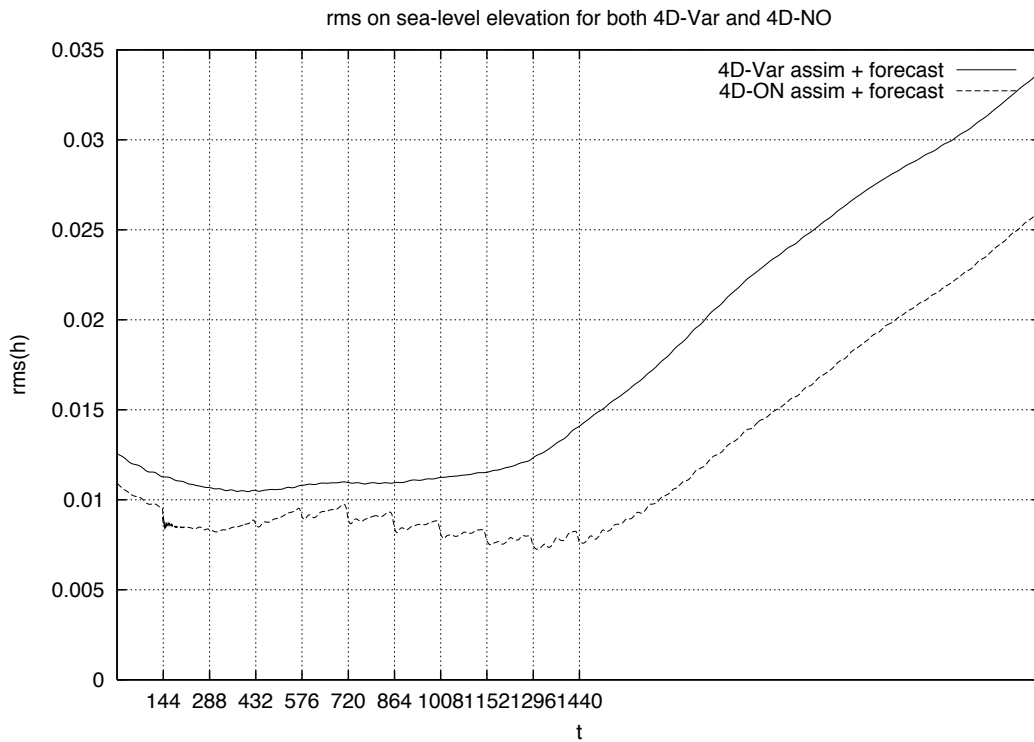


FIG. 6.4: Norme rms de l'erreur sur la SSH pour la version lissée du 4D-ON comparée au 4D-Var pour un mois d'assimilation et un mois de prévision

L'expérience sur l'équation de Burger a montré que le contrôle de la matrice \mathbf{K} complète n'est pas nécessaire et peut même être néfaste. Ainsi le contrôle peut être restreint à quelques coefficients scalaires sans perte de qualité.

Grâce à ces simplifications et en raison de la croissance actuelle de la puissance des moyens de calcul, le surcoût induit par le contrôle du terme de nudging (à peu près 10%) n'est pas une réelle limitation pour des applications réalistes et le 4D-ON pourra efficacement être appliqué dans un futur proche à des problèmes océanographiques ou météorologiques.

En parallèle à ces applications à des modèles réalistes quelques développements peuvent être effectués : dans les cas étudiés dans ce chapitre, \mathbf{C}_Q l'opérateur de diffusion inclus dans l'opérateur \mathbf{Q} est le même que celui de l'opérateur \mathbf{B} bien qu'il n'y ait aucune raison à cela, les erreurs modèle et celles de l'ébauche n'ont pas nécessairement les mêmes structures de corrélation. Trouver un équilibre adéquat entre ces deux opérateurs pourrait être un développement futur naturel de cette méthode.

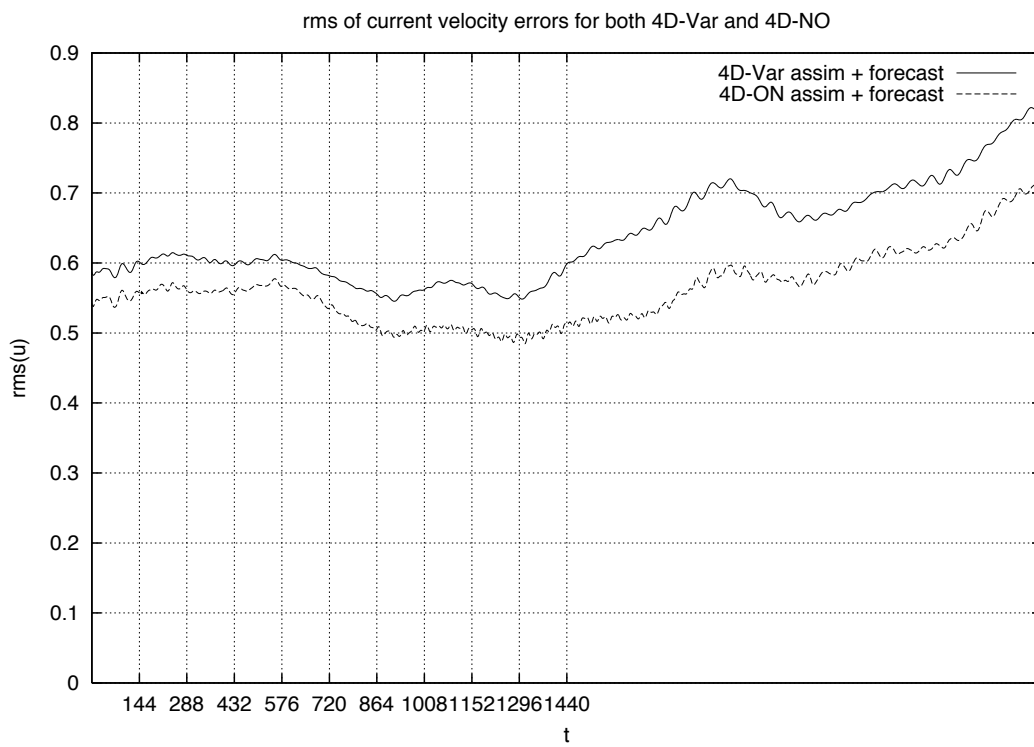


FIG. 6.5: Norme rms de l'erreur sur les vitesses de courant pour la version lissée du 4D-ON comparée au 4D-Var pour un mois d'assimilation et un mois de prévision

Chapitre 7

4D-Var avec contrôle de l'erreur systématique du modèle

Supposons, comme d'habitude, que \mathbf{x}_i^t représente l'état vrai au temps t_i et \mathbf{x}_i^f l'état prédit, alors l'erreur peut s'écrire

$$\eta_i = \mathbf{x}_i^t - \mathbf{x}_i^f$$

d'un autre coté l'erreur d'observation s'écrit si on combine l'erreur de mesure et celle de représentation :

$$\varepsilon_i = \mathbf{y}_i^o - H_i(\mathbf{x}_i^t)$$

Jusqu'à maintenant les erreurs étaient considérées comme étant non-biaisées c'est à dire $E[\eta_i] = E[\varepsilon_i] = 0$. De nombreux travaux ont démontré dans la pratique l'existence d'une erreur systématique sur l'état provenant de celle incluse dans le modèle. On verra même dans la suite de ce chapitre et surtout dans la partie suivante l'importance de son influence dans les problèmes d'assimilation de données.

D'un point de vue théorique, l'optimalité du filtre de Kalman n'est réalisée que sous les hypothèses de bruits η_i et ε_i non biaisés et non corrélés entre eux. Il est donc nécessaire, pour que les comparaisons Kalman/variationnel aient un sens de nous ramener à ce cas.

7.1 Un exemple scalaire

Pour mettre en évidence la nécessité de considérer cette erreur systématique séparément des autres erreurs analysons le comportement de l'assimilation variationnelle sur un exemple scalaire simple. On se place dans le cadre d'un 3D-Var afin de simplifier au maximum les notations, mais des résultats identiques peuvent être obtenus dans le cadre du 4D-Var (il s'agit du même problème d'optimisation).

En reprenant l'exemple donné par DEE et DA SILVA [37] pour illustrer le rôle du biais dans le filtre de Kalman et en l'appliquant à l'assimilation variationnelle, supposons que x^b et y^o sont des scalaires avec $E[(\varepsilon)^2] = E[(\eta)^2] = \sigma^2$, $E[\varepsilon] = b^o = 0$ et $E[\eta] = b$

En reprenant la fonction objective du 3D-Var

$$J(x) = \frac{1}{2}(x - y^o)^2\sigma^{-1} + \frac{1}{2}(x - x^b)\sigma^{-1}$$

on obtient facilement l'état analysé (qui minimise la fonction objective)

$$x^a = \frac{1}{2}(y^o + x^b)$$

ainsi que l'erreur d'analyse :

$$\eta^a = \frac{1}{2}(\varepsilon + \eta)$$

dont la moyenne est :

$$b^a = \frac{1}{2}b$$

et la covariance d'erreur d'analyse

$$\langle (\eta^a)^2 \rangle = \frac{1}{4}b^2 + \frac{1}{2}\sigma^2 \quad (7.1)$$

Dans ce cas la moyenne de l'erreur de l'incrément d'analyse est égale à $b/2$ et sous-estime le biais d'un facteur 2.

On remarque ici que l'assimilation permet de réduire le biais dans l'état analysé mais ne l'élimine pas complètement. On peut néanmoins tenter de le réduire davantage en donnant plus de poids au terme J^o d'écart aux observations dans la fonction objective : si on fait l'hypothèse (fausse) dans l'écriture de la fonction objective que $E[(\varepsilon)^2] = 2\sigma^2$, on obtient de la même façon que précédemment :

$$b^a = \frac{1}{3}b \quad \langle (\eta^a)^2 \rangle = \frac{1}{9}b^2 + \frac{5}{9}\sigma^2 \quad (7.2)$$

Le biais dans l'analyse est moins important et la variance totale est moindre mais, si augmenter encore le poids du terme d'écart aux observations permet de réduire le biais, cela augmentera la variance de la composante aléatoire de l'erreur.

La figure (7.1) résume cet exemple, elle représente la dépendance des différentes composantes de l'erreur d'analyse au coefficient α si on écrit

$$J(x) = \frac{\alpha}{2}(x - y^o)^2\sigma^{-1} + \frac{1}{2}(x - x^b)\sigma^{-1}$$

Cet exemple illustre le fait que si on ne prend pas en compte l'erreur systématique *explicitement* dans la méthode d'assimilation, le biais ne pourra être réduit que moyennant une augmentation de la part aléatoire de l'erreur d'analyse. De plus cela impose de spécifier des covariances d'erreur intentionnellement inexactes.

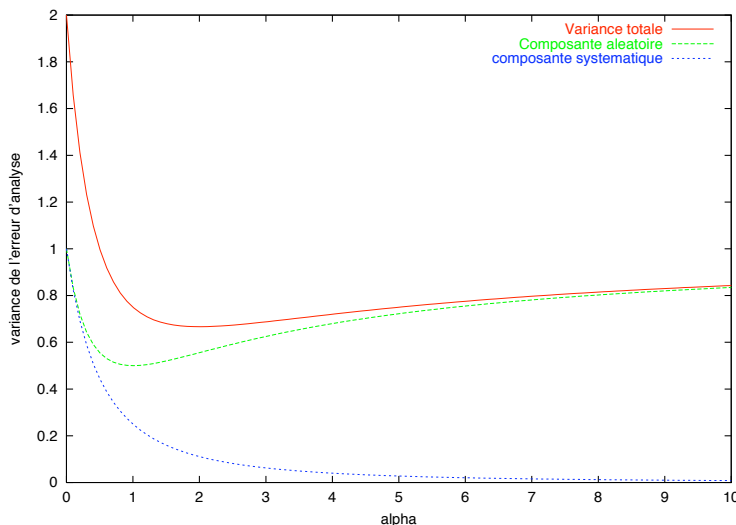


FIG. 7.1: Variance de l'erreur d'analyse en fonction du poids donné au terme d'écart aux observations de la fonction objective

7.2 La méthode

Nous allons introduire ici la méthode de contrôle de l'erreur systématique (4D-SEC), elle dérive directement du cas très classique de contrôle de la condition initiale (le 4D-Var). Comme dans toute méthode d'assimilation de données l'objectif de cette méthode est de trouver un bon "compromis" entre l'évolution d'un système fournie par un modèle et la trace de cette évolution obtenue par les observations. Pour ce faire nous allons utiliser une méthode de contrôle optimal visant à minimiser une fonction objective décrivant le "compromis" en question. Le vecteur de contrôle ne sera plus uniquement la condition initiale comme c'était le cas pour le 4D-Var, mais on contrôlera aussi un terme de correction de l'erreur systématique.

Soit \mathbf{x} l'état de l'océan ou de l'atmosphère et $M_{(t_i, t_{i+1})}$ le modèle décrivant l'évolution de cet état entre les temps t_i et t_{i+1} . Si on reprend la formulation du 4D-Var avec modèle imparfait décrite dans la partie 1.3.2 du chapitre 1, le modèle corrigé s'écrit :

$$\begin{cases} \mathbf{x}_{i+1} = M_{(t_i, t_{i+1})}(\mathbf{x}_i) + \eta_{i+1} \\ \mathbf{x}_0 = \mathbf{x}^b + \delta \mathbf{x}_0 \end{cases} \quad (7.3)$$

et on cherche à minimiser la fonction objective

$$J(\delta \mathbf{x}_0, \eta_1, \dots, \eta_n) = \frac{1}{2} \sum_{i=1}^n [H(\mathbf{x}_i) - \mathbf{y}_i^o]^T \mathbf{R}_i^{-1} [H(\mathbf{x}_i) - \mathbf{y}_i^o] + \frac{1}{2} [\delta \mathbf{x}_0]^T \mathbf{B}^{-1} \delta \mathbf{x}_0 + \frac{1}{2} \sum_{i=1}^n \eta_i^T \mathbf{Q}_i^{-1} \eta_i \quad (7.4)$$

Grâce aux méthodes adjointes, on obtient l'écriture du gradient :

$$\begin{cases} \nabla_{\delta \mathbf{x}_0} J = \mathbf{x}_0^* + \mathbf{B}^{-1} \delta \mathbf{x}_0 \\ \nabla_{\eta_i} J = \mathbf{x}_i^* + \mathbf{Q}_i^{-1} \eta_i \end{cases} \quad (7.5)$$

Ici on va tenter de modéliser le terme d'erreur par

$$\eta_i = \Phi_{(t_1, t_i)}(\eta_1) \quad (7.6)$$

où Φ modélise l'évolution de l'erreur et η_1

Dans ce cas on remplace dans la formulation précédente η_i par $\Phi_{(t_1, t_i)}(\eta_1)$. Le modèle corrigé devient

$$\begin{cases} \mathbf{x}_0 = \mathbf{x}^b + \delta \mathbf{x}_0 \\ \mathbf{x}_{i+1} = M_{(t_i, t_{i+1})}(\mathbf{x}_i) + \Phi_{(t_i, t_{i+1})}(\eta_1) \end{cases} \quad (7.7)$$

Si on ne s'intéresse qu'à la part systématique de l'erreur, c'est à dire qu'on fixe $\Phi = \mathbf{Id}$ et que l'on pose $\mathbf{Q}_i = \mathbf{Q}$ quelque soit i , la fonction objective que nous voulons minimiser devient :

$$J(\mathbf{x}_0, \eta_1) = \frac{1}{2} \sum_{i=0}^n (H_i(\mathbf{x}_i) - \mathbf{y}_i^o)^T \mathbf{R}^{-1} (H_i(\mathbf{x}_i) - \mathbf{y}_i^o) + \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}^b) + \frac{n}{2} \eta_1^T \mathbf{Q}^{-1} \eta_1$$

Et on obtient les expressions des différentes composantes du gradient :

$$\begin{cases} \nabla_{\mathbf{x}_0} J = \mathbf{x}_0^* + \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}^b) \\ \nabla_{\eta_1} J = \sum_{i=1}^n \mathbf{x}_i^* + n \mathbf{Q}^{-1} \eta_1 \end{cases} \quad (7.8)$$

7.3 Expérimentation avec le modèle Shallow-Water

On va tester cette méthode sur notre cas académique de modèle Shallow-Water. Les conditions de l'expérience sont les mêmes que dans le chapitre sur le nudging optimal : une configuration d'expériences jumelles avec un modèle décrivant la réalité et le même modèle entaché d'erreur servant de modèle de prévision.

Le champ 7.2 représente la moyenne de l'erreur de modèle sur la hauteur d'eau \mathbf{h} durant cette période. C'est cette moyenne que nous allons tenter d'estimer par la méthode introduite dans ce chapitre.

7.3.1 Préconditionnement et minimisation

Si on introduit un terme de correction de l'erreur systématique dans les équations du modèle, et que l'on contrôle ce terme et la condition initiale, le gradient de la fonction coût sera obtenu par le système (7.8). Cependant la grande disparité entre les dimensions des composantes du vecteur d'état (\mathbf{u} , \mathbf{v} sont de l'ordre de 1 m s^{-1} et \mathbf{h} de 500 m) et les valeurs des composantes $\nabla_{\mathbf{x}_0} J$, $\nabla_{\eta_1} J$ du gradient (la première est à peu près n fois moins importante que la deuxième avec n : le nombre de pas de temps de la période d'assimilation) vont impliquer un très mauvais conditionnement pour notre minimisation. C'est pourquoi il est important d'effectuer un preconditionnement de la variable de contrôle. Pour ceci on utilisera, les racines carrées des matrices \mathbf{B} et \mathbf{Q} (au sens $\mathbf{B} = \mathbf{B}^{\frac{1}{2}} \mathbf{B}^{\frac{T}{2}}$) comme cela est

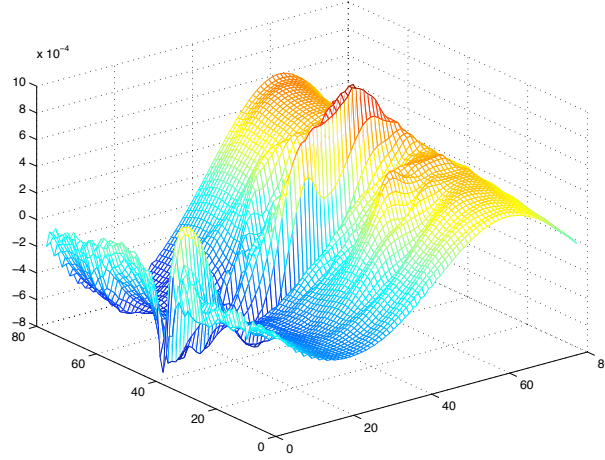


FIG. 7.2: Part systématique de l'erreur sur la SSH sur l'ensemble du domaine

décrit au chapitre *techniques algorithmiques*. \mathbf{B} et \mathbf{Q} seront également construites suivant la méthode introduite dans ledit chapitre.

Ce préconditionnement a deux principaux avantages : il permet d'adimensionaliser les composantes de la variable de contrôle et il améliore le conditionnement (voir chapitre 3, partie 3.3).

Ainsi notre nouveau vecteur de contrôle sera :

$$\mathbf{v} = \begin{bmatrix} \mathbf{B}^{-\frac{1}{2}} & 0 \\ 0 & \sqrt{n}\mathbf{Q}^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 - \mathbf{x}^b \\ \eta_1 \end{bmatrix}$$

Et le problème initial deviendra le problème équivalent :

$$J(\mathbf{v}) = \underbrace{\frac{1}{2} \sum_{i=0}^n (H_i(\mathbf{x}_i) - \mathbf{y}_i^o)^T \mathbf{R}_i^{-1} (H_i(\mathbf{x}_i) - \mathbf{y}_i^o)}_{J_o} + \overbrace{\mathbf{v}^T \mathbf{v}}^{J_b}$$

et

$$\nabla_{\mathbf{v}} J = \mathbf{v} + \begin{bmatrix} \mathbf{B}^{\frac{T}{2}} & 0 \\ 0 & \frac{1}{\sqrt{n}}\mathbf{Q}^{\frac{T}{2}} \end{bmatrix} \begin{bmatrix} \nabla_{\mathbf{x}_0} J \\ \nabla_{\eta_1} J \end{bmatrix}$$

7.3.2 Résultats

La période d'assimilation est de 1 mois, le champ entier l'élévation de surface \mathbf{h} est observé une fois tous les trois jours. Pour l'assimilation on a effectué 100 itérations du minimiseur Quasi Newton à mémoire limitée N1QN3 de l'INRIA (*Gilbert et Lemaréchal* [52]).

On a obtenu par cette assimilation un état optimal (ou état analysé) sur l'ensemble de la période d'assimilation. On utilise l'état analysé à la fin de la fenêtre d'assimilation comme condition initiale pour intégrer le modèle de prévision sur une période de un mois supplémentaire (période qu'on appellera par la suite période de prévision).

Sur la figure 7.3 est représentée l'évolution temporelle de la norme rms de l'erreur sur \mathbf{h} sur l'ensemble du domaine, pour la période d'assimilation (partie gauche du graphe) et

pour la période de prévision (partie droite). La courbe verte (trait pointillé pour la version en noir et blanc) représente le résultat après une assimilation 4D-Var. On reconnaît sur la période d'assimilation la forme en "u" caractéristique de l'évolution de la norme d'erreur sur l'ensemble de la période d'assimilation du 4D-Var quand le modèle n'est pas parfait. En effet, dans ce cas, la méthode doit s'éloigner de la véritable condition initiale afin d'obtenir un résultat plus proche des observations sur l'ensemble de la période d'assimilation. Le champ de gauche de la figure 7.4 illustre bien ce phénomène : elle représente l'erreur sur l'incrément de la condition initiale, c'est à dire la différence entre l'incrément estimé par la méthode d'assimilation et l'incrément qu'on devrait avoir pour recomposer exactement la condition initiale. Ce champ représente donc l'erreur introduite par le 4D-Var sur la condition initiale. La ressemblance avec l'erreur systématique représentée sur la figure 7.2 est frappante : pour tenter de corriger l'erreur modèle le 4D-Var est obligé d'accroître l'erreur sur la condition initiale. En général l'erreur est minimale au milieu de la période d'assimilation car c'est ici que l'on dispose du plus grand nombre d'information dans le passé et dans le futur. Une fois la période d'assimilation terminée, le modèle libre de toute assimilation dérive lentement en raison des erreurs introduites dans ses équations. L'introduction du contrôle de l'erreur systématique permet de faire disparaître cette forme en "u" (courbe rouge ou trait plein de la figure 7.3). La courbe obtenue est quasiment horizontale ce qui peut signifier qu'on a réussi à éliminer une bonne part de l'erreur. L'introduction du contrôle de l'erreur systématique a permis de réduire très sensiblement l'erreur sur l'ensemble de la fenêtre d'assimilation. De plus la condition initiale peut être estimée indépendamment de la correction d'erreur modèle. Ceci est mis en évidence par la partie droite de la figure 7.4 représentant l'erreur sur l'incrément de la condition initiale si on contrôle aussi l'erreur systématique. La structure ressemblant à l'erreur présente dans le modèle à disparu, laissant place à un bruit de faible amplitude.

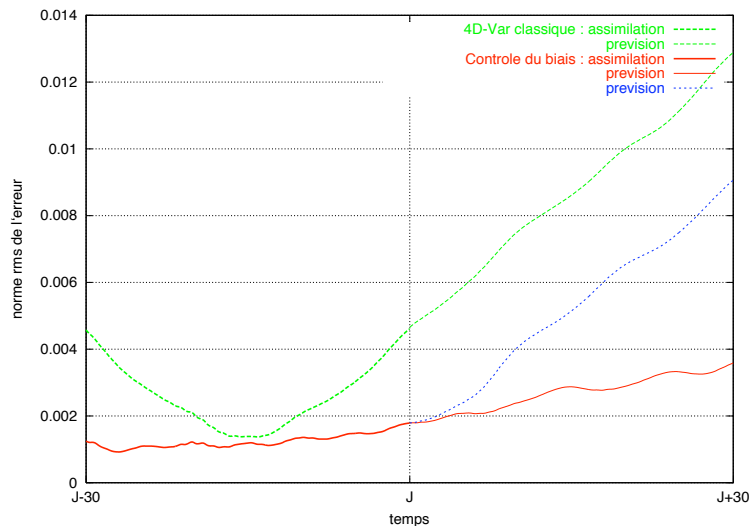


FIG. 7.3: Évolution en temps de la r.m.s de l'erreur sur l'élévation de surface de la mer pour la période d'assimilation puis celle de prévision

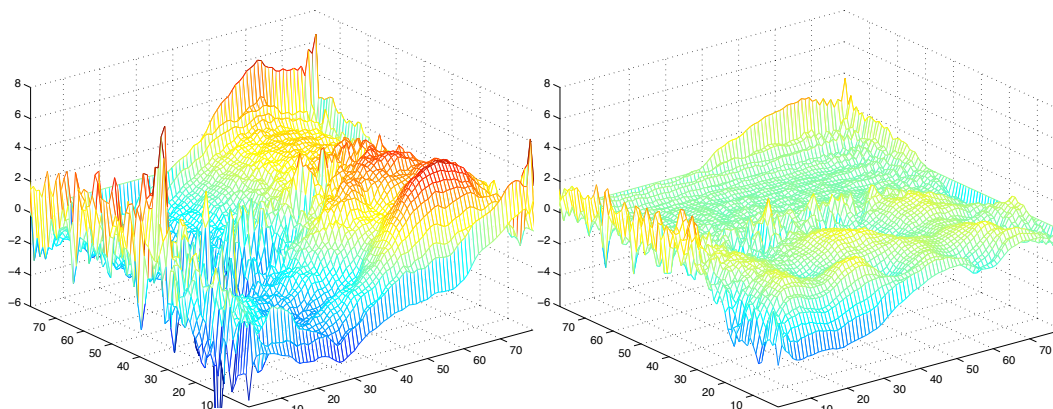


FIG. 7.4: Erreur sur l'incrément sur la condition initiale pour le 4D-Var (à gauche) et le 4D-SEC (à droite)

Pour la période prévision deux expériences ont été menées avec le contrôle de l'erreur systématique. Pour la courbe rouge (ou trait plein) de la figure 7.3 on a, pour la période de prévision, conservé dans les équations du modèle le terme \mathbf{b} de correction de l'erreur. La courbe bleue (ou pointillé gras) représente l'évolution de la norme de l'erreur pour la période de prévision si on ne conserve pas le terme η_1 dans les équations du modèle. Si on conserve le terme de correction dans les équations la dérive de la prévision est nettement moins importante ce qui montre que l'erreur estimée est cohérente avec l'erreur présente dans le modèle. Pour se convaincre un peu plus de la qualité de cette estimation, on peut regarder la figure 7.5 : celle ci représente l'erreur estimée par le 4D-SEC. Si on compare cette figure au biais effectivement présent dans la modèle (figure 7.2) de très fortes similitudes apparaissent notamment la présence des deux phases caractéristiques. Ceci est un indice supplémentaire pour nous permettre de conclure que l'erreur est bien estimée.

Les figures 7.6 représentent l'équivalent de la figure 7.3 pour les vitesses de courant \mathbf{u} et \mathbf{v} . Les résultats du 4D-SEC sur les composantes non observées du vecteur d'état restent de très bonne qualité. Le niveau d'erreur durant la période d'assimilation reste plus bas que pour le 4D-Var et la bonne estimation de l'erreur sur ces composantes permet de réduire, là encore, la dérive du modèle. Ces bons résultats sont obtenus en partie grâce au fait que l'on dispose d'un opérateur de multiplication par \mathbf{Q} multivarié (obtenu à partir de l'équilibre géostrophique).

L'introduction du contrôle de l'erreur systématique dans le 4D-Var ne requiert pas de lourdes modifications de l'algorithme. Il suffit d'ajouter les termes de correction dans les équations du modèle et, pour obtenir le gradient, d'ajouter l'état adjoint à chaque pas de temps. Le modèle adjoint devant de toute façon être intégré cela ne demande presque

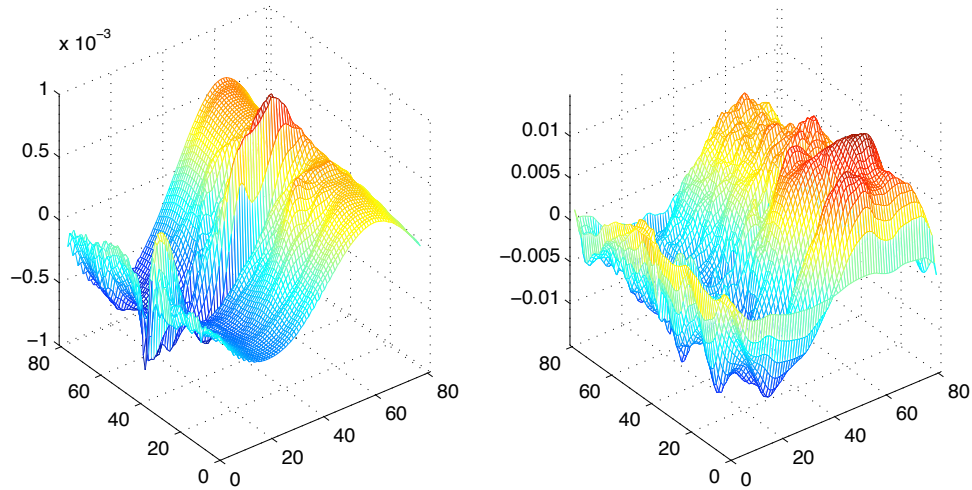


FIG. 7.5: Erreur présente dans le modèle (à gauche) et erreur estimée par le 4D-SEC avec la totalité du champ de SSH observée tous les trois jours (à droite)

pas de calcul supplémentaire. Une itération ne coûte donc quasiment pas plus cher en temps de calcul que pour le 4D-Var classique. Cependant la taille du vecteur de contrôle étant doublée on pourrait craindre que la minimisation converge moins vite et que l'on ait besoin de plus d'itérations pour obtenir un résultat satisfaisant. La figure 7.7 montre que si la méthode converge effectivement plus doucement, le nombre d'itérations ne devra pas forcément être augmenté. Cette figure représente l'évolution du logarithme du terme d'écart aux observations de la fonction objective en fonction du nombre d'itérations pour les deux méthodes. Dans le pire des cas (un petit nombre d'itérations) les deux méthodes sont aussi proches l'une que l'autre des observations. Si on continue la minimisation, assez vite le 4D-Var a fait le maximum que lui autorise le contrôle de la seule condition initiale, alors que l'introduction du contrôle de l'erreur systématique permet de continuer à se rapprocher des observations.

7.3.3 Des observations moins fournies

Dans la partie précédente les observations étaient composées du champ total de SSH tout les trois jours ce qui pourrait correspondre, dans la réalité, à une carte surfacique issue de l'interpolation des traces satellites. Un des avantages des méthodes variationnelles 4D est de pouvoir utiliser les observations directement au moment où celles ci sont produites, on est donc tenté de vouloir assimiler non plus des cartes, mais les observations le long des traces satellites (fig 6.1 du précédent chapitre).

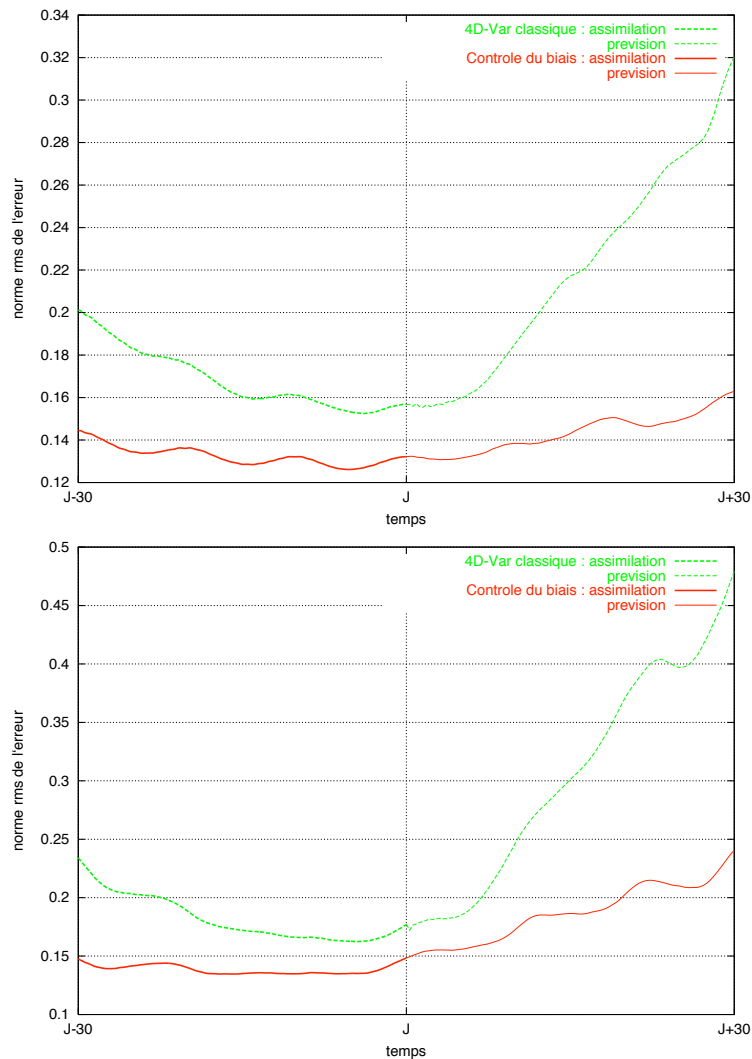


FIG. 7.6: Évolution en temps de la *r.m.s* de l'erreur sur les vitesses de courant en fonction du temps pour la période d'assimilation puis celle de prévision

Si dans l'expérience précédente le problème de minimisation était surdéterminé, c'est à dire qu'on avait plus d'observation que de degrés de liberté (20000 degrés de liberté pour 67000 observations), ce n'est plus le cas en observant la SSH suivant les traces satellites (plus que 5000 observations) et cela pourrait entraîner quelques difficultés.

Les figures 7.8 et 7.9 représentent toujours la norme rms de l'erreur pour les trois composantes du vecteur d'état respectivement, mais cette fois plusieurs cycles d'assimilation successifs sont tracés (traits épais) ainsi que la norme de l'erreur sur la période de prévision associée (traits fins). On remarque que pour les périodes d'assimilation l'introduction du contrôle de l'erreur systématique a permis de réduire le niveau global de l'erreur. Cependant les choses peuvent encore être améliorées car la forme en "u" de la courbe d'erreur persiste, ce qui pourrait signifier que la présence de l'erreur de modèle persiste. De plus les résultats pour la période de prévision ne sont pas ou peu (suivant le cycle d'assimilation) améliorés par rapport à la méthode classique.

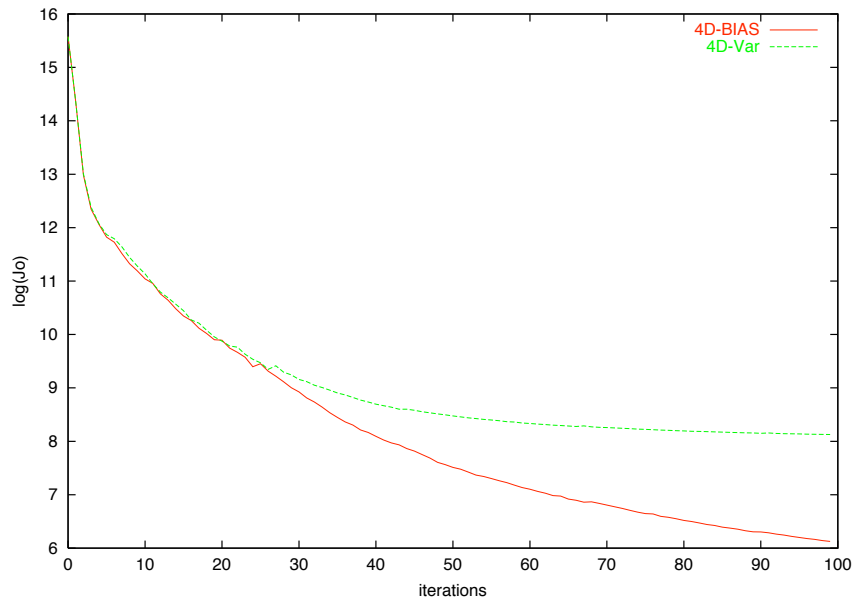


FIG. 7.7: Comparaison des logarithmes de termes J_o des fonctions coûts

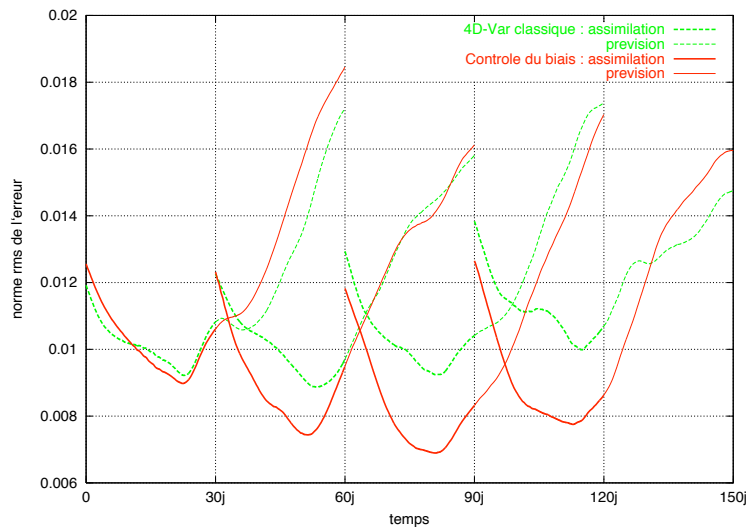


FIG. 7.8: Évolution en temps de la r.m.s de l'erreur sur l'élévation de surface de la mer pour la période d'assimilation puis celle de prévision

Dans le terme estimé de correction d'erreur (figure 7.10) les deux phases visibles sur le champ 7.2 ne sont plus dominantes (même si elles existent encore). On ne peut donc plus affirmer que le terme de correction d'erreur a été bien estimé (ou alors il faut un certain aplomb). Le nombre d'observations n'étant plus suffisant la méthode ne dispose plus d'assez d'informations pour recomposer de façon assez fidèle l'erreur systématique du modèle.

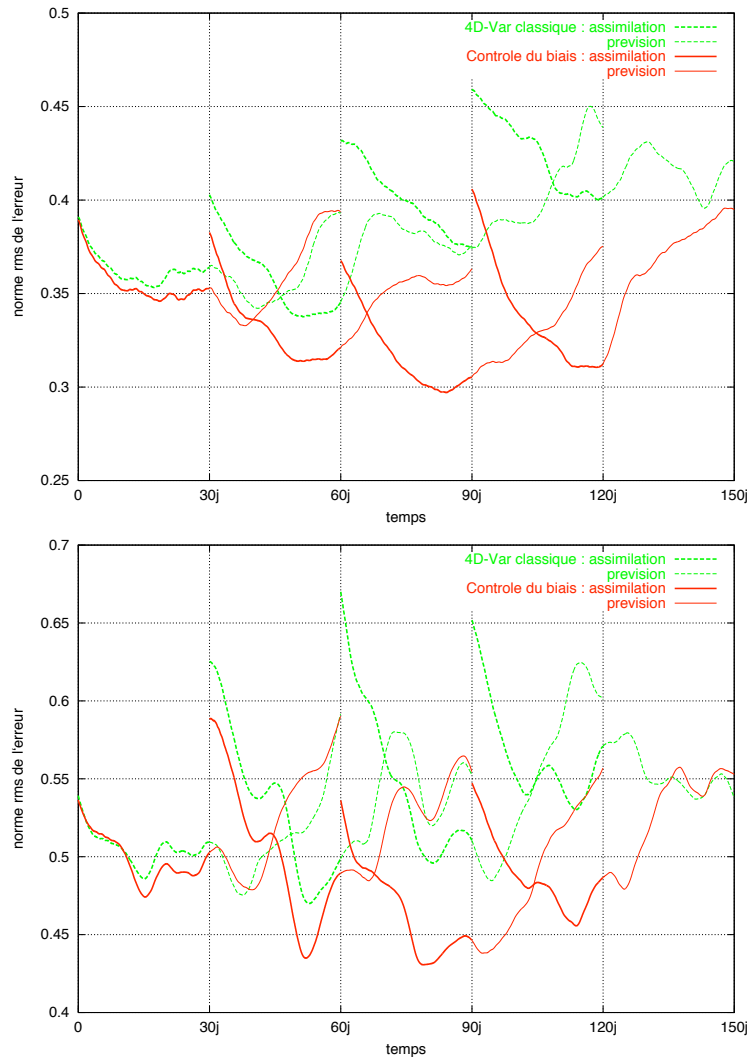


FIG. 7.9: Évolution en temps de la r.m.s de l'erreur sur les vitesses de courant en fonction du temps pour la période d'assimilation puis celle de prévision

L'apport du contrôle de l'erreur systématique permet tout de même de réduire l'écart aux observations (et même l'erreur globale sur la période d'assimilation). Cependant rien ne dit que ce terme de correction d'erreur est consistant l'erreur systématique effectivement présente dans le modèle, il peut très bien inclure des informations sur d'autres types d'erreur (la part non systématique de l'erreur ou sur la part résiduelle de l'erreur sur la condition initiale) ou même être simplement un terme permettant de réduire l'écart aux observations et ne correspondre à rien aux endroits qui ne sont pas observés (phénomènes d'aliasing).

Il convient de combler ce manque d'information par l'introduction d'une nouvelle source d'information. Une des pistes possible est de modifier le terme d'ébauche de l'erreur systématique cela, plus précisément, grâce à la matrice \mathbf{Q} . Comme cela a été évoqué lors de la présentation de la méthode de construction des opérateurs/matrices de covariance

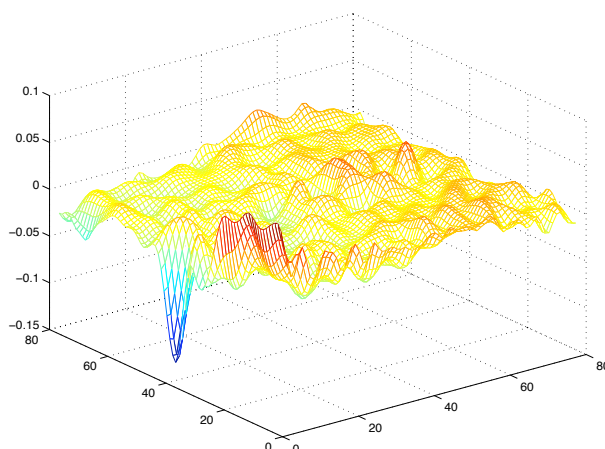


FIG. 7.10: Erreur estimée par le 4D-SEC en observant suivant les traces satellites

d'erreur, cela revient à diminuer l'ensemble des états admissibles pour la minimisation afin d'obtenir un état cohérent avec la physique.

Pour illustrer cette approche, on va focaliser notre attention sur le 3^{ème} cycle des figures précédentes (voir figures 7.11 et 7.12).

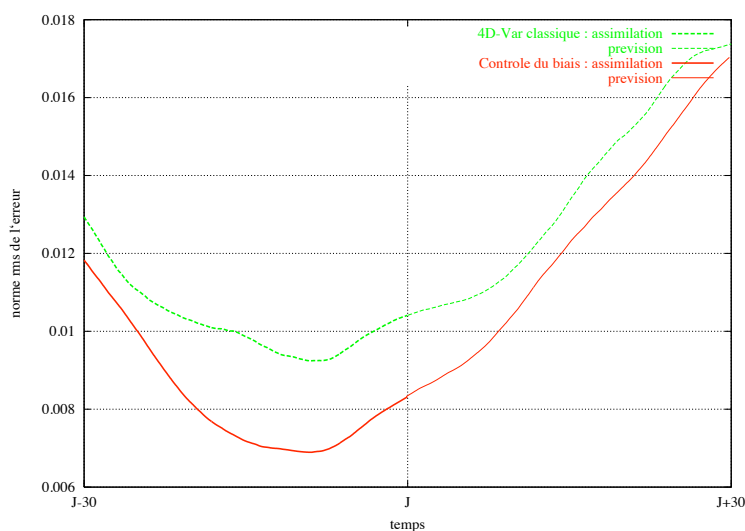


FIG. 7.11: Évolution en temps de la r.m.s de l'erreur sur l'élévation de surface de la mer pour la période d'assimilation puis celle de prévision

L'information supplémentaire que l'on peut apporter est la connaissance *a priori* que l'on a sur les erreurs. Dans le cas présent on va introduire dans la matrice \mathbf{Q} des contraintes sur la structure en deux phases principales qui apparaissent dans l'erreur "théorique" et

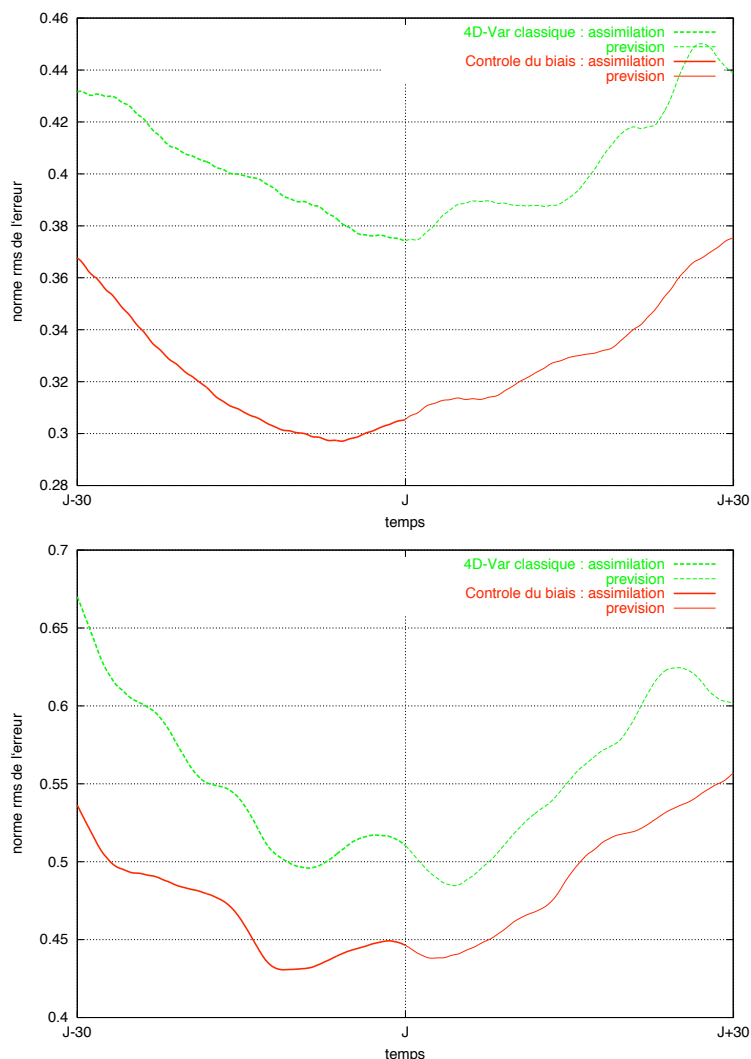


FIG. 7.12: Évolution en temps de la r.m.s de l'erreur sur les vitesses de courant en fonction du temps pour la période d'assimilation puis celle de prévision

laisser plus de liberté au niveau du jet. En d'autres termes on va imposer à la méthode de ne corriger que les grandes échelles de l'erreur (ces deux phases) car on ne dispose pas de suffisamment d'information pour corriger efficacement les petites échelles.

Techniquement cela pourra se traduire par l'ajout d'un opérateur \mathbf{S} dans la construction de l'opérateur de multiplication par la racine carrée de la matrice de covariance d'erreur de modèle (équation 7.9).

$$\eta_1 = \mathbf{K}\hat{\mathbf{Q}}^{\frac{1}{2}}\mathbf{S}\alpha \quad (7.9)$$

$$\nabla_{\alpha}J = \mathbf{S}^T\hat{\mathbf{Q}}^{\frac{T}{2}}\mathbf{K}^T\nabla_{\eta_1}J \quad (7.10)$$

où $\hat{\mathbf{Q}}$ est l'opérateur de covariance d'erreur de modèle non équilibré et \mathbf{K} est l'opérateur d'équilibre.

Ici l'opérateur \mathbf{S} peut être construit en utilisant la forme "sinusoïdale" de l'erreur systématique à estimer de façon très simple par l'équation 7.3.3

$$(\mathbf{S}\alpha)(i, j) = \alpha(j) \sin(-\pi i/I) \quad i = 1..I \quad j = 1..J \quad (7.11)$$

Cette approche a de plus l'intérêt de réduire de façon drastique la taille de contrôle de la partie erreur systématique (on est passé de $I \times J$ à I).

La focalisation sur les seules grandes échelles permet de retrouver un terme de correction d'erreur cohérent avec l'erreur effectivement présente dans le modèle et ceci se traduit immédiatement par une prévision de bien meilleure qualité (figures 7.13 et 7.14). Sur la période d'assimilation le gain apporté par la variante "grandes échelles" n'est pas évident à première vue, mais le fait que les courbes d'erreur se rapprochent d'une droite horizontale est un indice supplémentaire pour affirmer que le terme de correction d'erreur est mieux estimé.

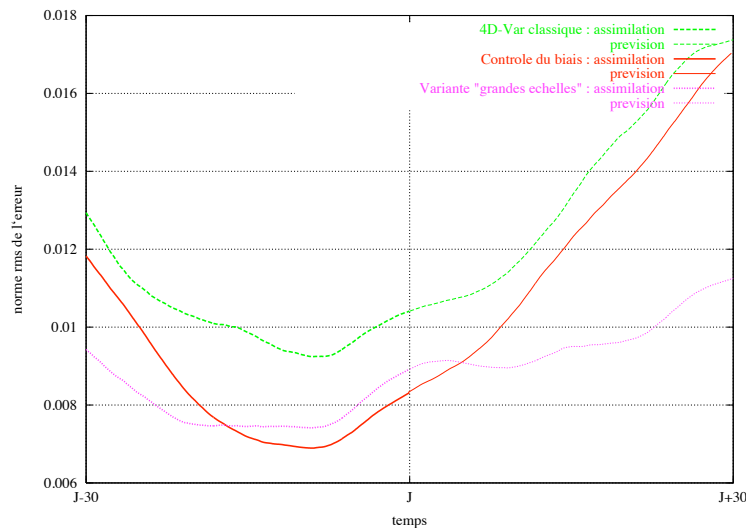


FIG. 7.13: Évolution en temps de la r.m.s de l'erreur sur l'élévation de surface de la mer pour la période d'assimilation puis celle de prévision

7.4 Conclusion

Contrôler un terme de correction d'erreur systématique au lieu d'un terme d'erreur aléatoire présente deux principaux avantages. Premièrement le surcoût en calcul est presque inexistant alors que les résultats sont bien meilleurs qu'avec le 4D-Var. En effet le terme d'écart aux observations est dès le début de la minimisation plus petit dans le cas du 4D-SEC par rapport à celui du 4D-Var. De plus le surcoût d'une itération est négligeable.

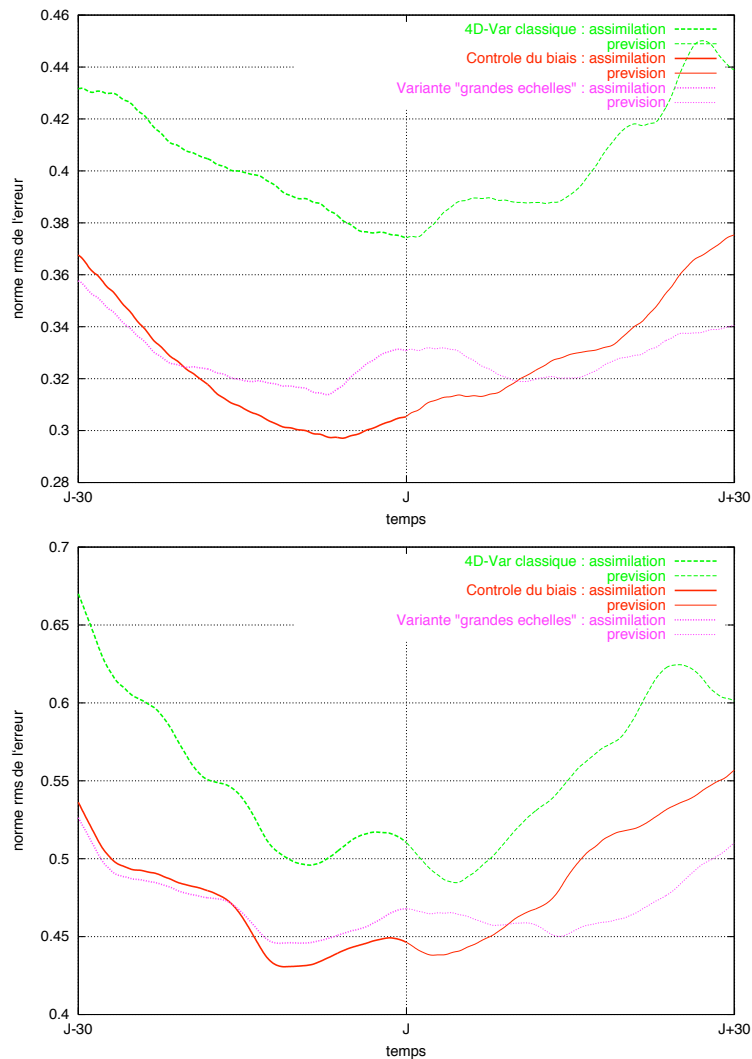


FIG. 7.14: Évolution en temps de la r.m.s de l'erreur sur les vitesses de courant en fonction du temps pour la période d'assimilation puis celle de prévision

Un autre intérêt (et non des moindres) de cette approche est de pouvoir conserver le terme de correction d'erreur optimal trouvé, dans les équations du modèle pour la prévision, ce qui améliore de façon très sensible celle-ci.

Cependant il convient de prendre garde lorsque les observations sont peu nombreuses. L'introduction du contrôle de l'erreur systématique du modèle dans le 4D-Var double le nombre de degrés de liberté de la minimisation, cela peut induire un manque d'information. Ce manque peut être comblé en améliorant les matrices de covariances d'erreur modèle ou à défaut, en favorisant fortement les structures grandes échelles dans celles-ci. Autrement dit, si on manque d'information pour estimer notre vecteur de contrôle, soit on en apporte, soit on restreint notre recherche. L'approche utilisée en fin de ce chapitre est difficilement applicable directement pour des applications réalistes car on ne connaît pas, par définition, la structure de l'erreur présente dans le modèle. Il faudra alors utiliser d'autres sources d'information provenant par exemple de la connaissance *a priori* qu'on a de l'erreur (on

sait quels sont les faiblesses du modèle, quelles approximations y sont faites, etc,...), de la moyenne des incréments du 4D-Var sur une longue période (si il n'y avait pas d'erreur systématique celle ci serait nulle), etc...

On aurait pu aussi aborder le problème en amont et se poser la question “ce modèle est-il le mieux adapté à la série d'observations dont on dispose ?”. Supposons que l'on possède plusieurs modèles de “finesse” différentes (en raison de la résolution adoptée ou des différences de paramétrisation ou d'approximation) pour simuler le comportement d'un même système et que l'on dispose d'un jeu d'observations donné, utiliser le modèle le plus fin ne sera pas nécessairement le meilleur choix pour faire de l'assimilation de données. En effet si celui ci modélise des phénomènes qui ne sont pas représentés par les observations ceux ci ne pourront être contrôlés par l'assimilation ce qui risque de nuire à la qualité de l'analyse. D'un autre coté, utiliser un modèle trop grossier ne permettra pas une analyse performante. On peut supposer qu'il existe un bon compromis dans la finesse du modèle qui correspond au réseau d'observation utilisé pour l'assimilation de données.

La question de savoir si il est utile, voir prudent, d'utiliser un modèle très sophistiqué si on ne dispose que de peu d'information ou, *a contrario* d'exiger des observations très précises si le modèle est grossier, n'a, à notre connaissance, jamais été étudié en détail bien qu'il s'agisse d'une question de fond importante.

Pour aller plus loin dans la modélisation de l'erreur modèle on peut considérer que le terme d'erreur évolue dans le temps suivant une évolution fixée *a priori* (par exemple en fonction des saisons).

$$\eta_i = \mathcal{A}_{(t_i, t_1)} \eta_1$$

Le gradient se retrouve alors aisément par :

$$\nabla_{\eta_1} J = \sum_{i=1}^n \mathcal{A}_{(t_i, t_1)}^T \mathbf{x}_i^* + n \mathbf{Q}^{-1} \eta_1$$

Nous avons testé ici le cas de \mathcal{A} étant l'identité. D'autres formes d'évolution de l'erreur ont été envisagées (affine, sinusoïdale...) et des expériences très préliminaires ont été effectuées mais les résultats sont du même ordre que ceux obtenus pour un terme constant malgré une complexité accrue. Cette piste nécessite d'être explorée plus profondément.

La plupart des chaînes d'assimilation 4D-Var opérationnelles le sont sous la forme de la variante incrémentale (cf partie 1.3.2), c'est également le cas de OPALM que nous allons utiliser par la suite pour expérimenter le contrôle de la part systématique de l'erreur sur un cas réaliste. L'application de la méthode du 4D-SEC introduite plus haut à une formulation incrémentale ne pose pas trop de problème. Grâce à l'hypothèse linéaire tangente [1.13] on peut écrire l'état de l'océan au pas de temps n :

$$\mathbf{x}_n = M_{(t_0, t_n)}(\mathbf{x}_0) + \sum_{i=1}^n \mathbf{M}_{(t_i, t_n)} \eta_1$$

si on applique une petite perturbation $(\delta \mathbf{x}_0, \delta \mathbf{b}_0)$ au vecteur de contrôle, on peut réécrire

l'équation précédente :

$$\tilde{\mathbf{x}}_n = M_{(t_0, t_n)}(\mathbf{x}_0 + \delta \mathbf{x}_0) + \sum_{i=1}^n \mathbf{M}_{(t_i, t_n)}(\eta_1 + \delta \eta_1) \quad (7.12)$$

$$= M_{(t_0, t_n)}(\mathbf{x}_0) + \sum_{i=1}^n \mathbf{M}_{(t_i, t_n)}(\eta_1) + \mathbf{M}_{(t_0, t_n)}\delta \mathbf{x}_0 + \sum_{i=1}^n \mathbf{M}_{(t_i, t_n)}\delta \eta_1 \quad (7.13)$$

$$= \mathbf{x}_n + \delta \mathbf{x}_n \quad (7.14)$$

en posant

$$\begin{cases} \mathbf{x}_0 \\ \mathbf{x}_{i+1} = \mathbf{M}_{(t_i, t_{i+1})}\delta \mathbf{x}_i + \delta \eta_1 \end{cases} \quad (7.15)$$

C'est cette formulation qui sera appliquée à la configuration réaliste (OPA sur la fenêtre pacifique tropicale) utilisée dans les chapitres suivants.

Troisième partie

Application au modèle de circulation générale océanique OPA

Chapitre 8

Comment OPA réagit-il?

Ce chapitre est consacré dans sa première partie à l'étude de l'influence sur la dynamique d'OPA de diverses sources d'erreur. La seconde partie traitera des résultats du 4D-Var classique dans le cadre d'une expérience jumelle.

8.1 Influence des différentes sources d'erreur

Dans cette partie on ne fera pas d'assimilation de données mais on étudiera l'évolution temporelle du modèle en fonction de trois différents types d'erreur (erreur sur la condition initiale, erreur dans la dynamique du modèle et erreur sur les forçages).

Pour cela on se place dans le cadre d'expériences jumelles : la réalité est simulée par l'intégration du modèle sur l'année 1993, les vents utilisés pour forcer le modèle sont ceux fournis par le satellite ERS sur la même période de temps et la condition initiale est celle du 1^{er} janvier 1993 issue d'un long spin-up du modèle. Dans la suite on appellera cet état "bonne condition initiale" ou "état vrai" quand on sera dans le cadre d'expériences jumelles. Cette condition initiale et l'évolution du système qui y est associée seront référencées par l'appellation expérience GO4.

8.1.1 L'erreur sur la condition initiale (expérience CT1)

Pour étudier l'influence d'erreurs sur la condition initiale, on effectue une expérience jumelle en utilisant exactement le même modèle et les mêmes forçages mais on prend comme condition initiale l'état au 1^{er} janvier 1996 (à la place du 1^{er} janvier 1993).

La figure 8.1 représente l'évolution de la norme de l'erreur sur chacune des composantes du vecteur d'état en surface (en haut) et à 200 mètres (en bas) sur une période de 1 an. On s'aperçoit que les erreurs ont tendance à disparaître. Principalement les erreurs sur les vitesses de courants sont éliminées par la dynamique du modèle de façon impressionnante (jusqu'à 90% de l'erreur éliminée). La température rejoint de façon régulière l'état vrai sans qu'il soit nécessaire d'effectuer une assimilation de données. Les erreurs sur la salinité persistent un peu plus mais celles ci sont déjà très faibles au départ.

Ce résultat n'étonnera pas les océanographes; en effet, soumis aux bons forçages de vent, de part le comportement linéaire de sa dynamique, les courants de l'océan équatorial s'ajustent et s'équilibrent (PHILANDER, 1990 [96]) en à peu près six mois à la surface et un an en profondeur, comme on peut d'ailleurs le constater sur la figure. Il faut un peu plus de temps pour que cet équilibre se fasse aussi sur la température et la salinité.

Cette première expérience indique que si les erreurs étaient uniquement contenues dans la condition initiale, il ne serait, peut-être (restons prudent), pas nécessaire de faire de l'assimilation de données sur l'océan pacifique tropical. En pratique la question ne se pose pas : le modèle n'est pas une représentation idéale des phénomènes physiques de la réalité et surtout les forçages sur le pacifique tropical sont loin d'être exacts.

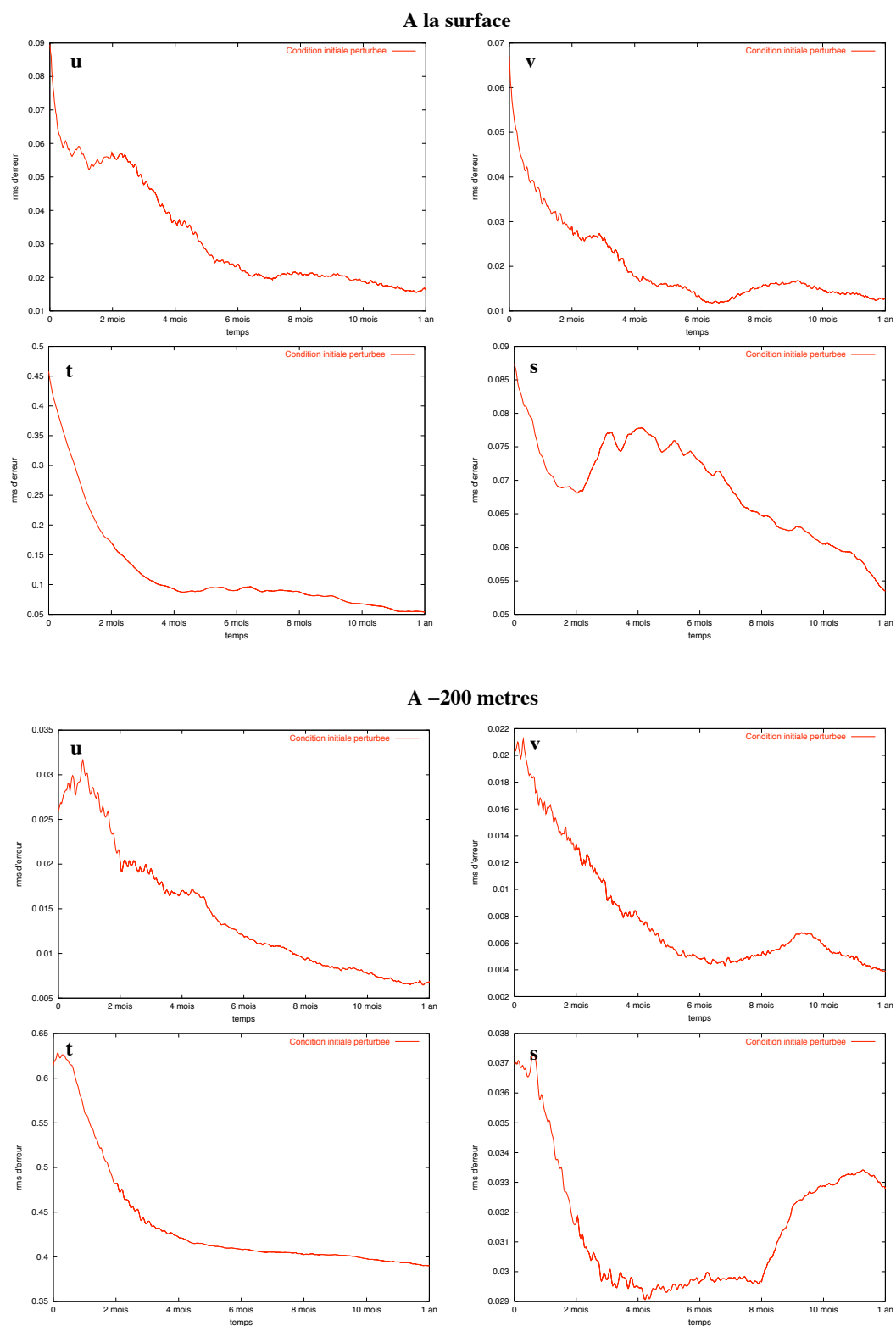


FIG. 8.1: Évolution de la norme de l'erreur sur chacune des composante du vecteur d'état en surface (en haut) et à 200 mètres (en bas) pour une erreur sur la condition initiale

8.1.2 L'erreur dans la physique du modèle (expérience CT2)

Cette fois ci la condition initiale du modèle est bien celle de janvier 1993 mais des différences dans les calculs des coefficients de diffusivité et de viscosité verticale tourbillonnaire ont été introduites. Dans la simulation de la réalité, pour ces calculs un schéma TKE (Turbulent Kinetic Energy) a été utilisé (BLANKE et DELECLUSE, 1993 [13]). Pour simuler un modèle inexact le calcul de ces coefficient à été obtenu à partir d'une formulation dépendant du nombre de Richardson (suivant PACANOWSKI et PHILANDER, 1981 [89])

Cette différence dans la physique du modèle se caractérise principalement par des gradients verticaux plus resserrés lors de l'usage du schéma TKE.

La figure 8.2 représente elle aussi l'évolution de la norme de l'erreur sur chacune des composantes du vecteur d'état en surface (en haut) et à 200 mètres (en bas) sur une période de 1 an. Contrairement à ce qu'on avait dans le cas de la seule condition initiale perturbée l'erreur ne cesse de croître. Les modèles étant (légèrement) différents, les deux solutions dérivent régulièrement. Comme on pouvait s'en douter les erreurs modèle jouent donc un rôle important dans l'évolution de l'erreur de prévision. À plus long terme l'erreur va saturer et donc se stabiliser comme cela à été décrit par BLANKE et DELECLUSE [13]. La période de temps considéré est trop courte pour que l'on parvienne à cette saturation

Sur certaines fenêtres de temps une assimilation 4D-Var classique ne pourra être efficace car la croissance de l'erreur est trop importante tandis que sur d'autres (le "replat" entre 4 et 6 mois) il ne sera peut-être pas nécessaire de prendre en compte l'erreur modèle. Cette remarque est faite à titre purement indicatif car rien ne permet de penser que dans un cas réel ce type d'erreur modèle est présent. Cela illustre simplement le fait que l'impact d'une même erreur modèle peut être différent suivant la période de temps étudiée. La prudence est de rigueur dès qu'il s'agit de prévoir l'impact de tel ou tel type d'erreur dans la physique du modèle. Suivant la nature de cette erreur elle peut être amplifiée par la dynamique du modèle ou au contraire dissipée

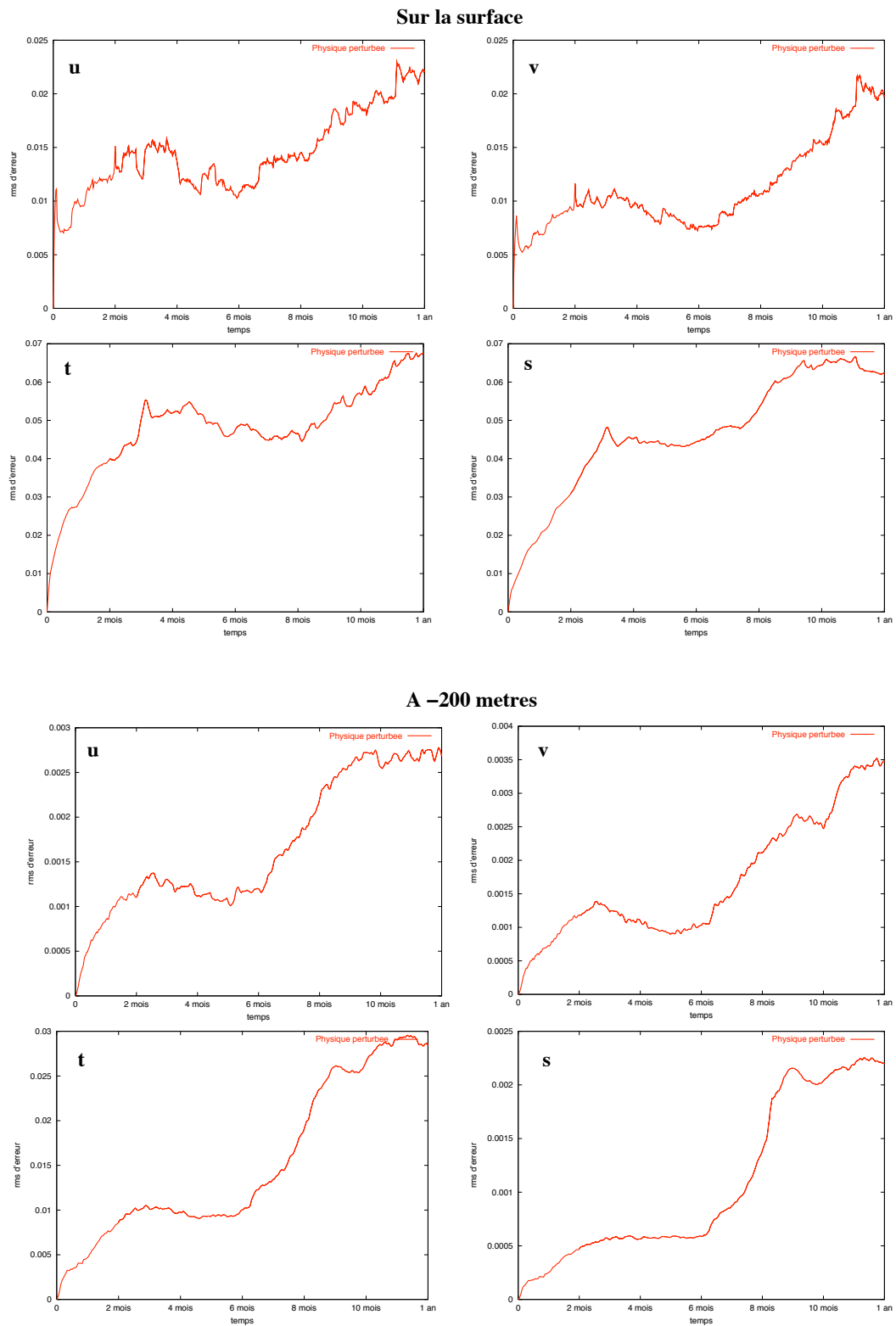


FIG. 8.2: Évolution de la norme de l'erreur sur chacune des composantes du vecteur d'état en surface (en haut) et à 200 mètres (en bas) pour des erreurs dans la physique du modèle

8.1.3 L'erreur sur les forçages (expérience CT3)

Comme dans l'expérience précédente la condition initiale est la bonne, cette fois ci la physique du modèle est aussi exacte (on reprend le schéma TKE pour estimer les paramètres du mélange vertical) mais des erreurs sont introduites dans le forçage de vent. Ces erreurs sont à la fois systématiques (on multiplie le vent par 1,5) et aléatoires (on ajoute un bruit blanc d'amplitude d'environ 10%).

On voit sur la figure 8.3 que la différence entre l'état du modèle de référence et celui du modèle au forçage défaillant a tendance à s'accroître rapidement durant les premiers mois d'intégration.

Une fois cette phase de croissance passée le modèle installe ses vitesses de courant dans un autre régime et l'erreur sur celles ci ne croit plus ou très peu. On a ici simplement augmenté la vitesse du vent, le modèle ajuste donc ses vitesses à ces nouveaux forçages. Comme on l'a précisé lors de l'expérience CT1 le modèle s'est ajusté en 6 mois / 1 an à la dynamique imposée par les forçages (malheureusement pour lui, ceux ci sont faux).

La température et la salinité en profondeur continuent à dériver lentement, on peut penser que l'ajustement au nouveau régime se fait moins rapidement (le lien entre le vent et ces variables n'est pas direct) mais que si on continue à intégrer le modèle pendant une certaine période de temps le niveau d'erreur se stabilisera.

La présence de ce type d'erreur rendra le 4D-var sans doute peu efficace. En effet dans ce cas la croissance de l'erreur sur l'état est la plus forte quand la condition initiale est exacte. Le 4D-Var sera donc contraint d'introduire de l'erreur dans la condition initiale pour avoir un état plus proche des observations sur la globalité de la fenêtre (phénomène observé au dernier chapitre de la partie précédente).

Contrôler l'erreur des tendances du modèle peut être vu comme contrôler les erreurs de manière générale, les incertitudes sur les forçages étant transmises grâce à la physique du modèle. On peut donc espérer que le 4D-SEC pourra être efficace en présence de ce type d'erreur.

8.1.4 Conclusion

Cette étude rapide sur l'impact de ces différents type d'erreurs montre que les erreurs sur la condition initiale sont amorties au cours du temps alors que les autres types d'erreur (dans la physique du modèle ou sur les forçages) persistent et/ou s'amplifient. Dans des applications pratiques sur le pacifique tropical, si seule la condition initiale était inexacte le 4D-Var serait très efficace : il corrigerait une grande partie de cette erreur et la dynamique du modèle se chargerait d'éliminer le reste. La présence d'erreurs dans la physique du modèle et sur les forçages ne permettra sans doute pas au 4D-Var de corriger efficacement la condition initiale. Dans des configurations réalistes ces trois types d'erreurs sont présents. Il est donc très vraisemblable que la prise en compte des erreurs modèle sera une avancée importante.

La part systématique de l'erreur est due principalement au erreurs dues aux erreur dans la physique du modèle (VIALARD et al, 2001 [122]) bien que l'incertitude sur les forçages soit très importante dans la fenêtre pacifique tropicale.

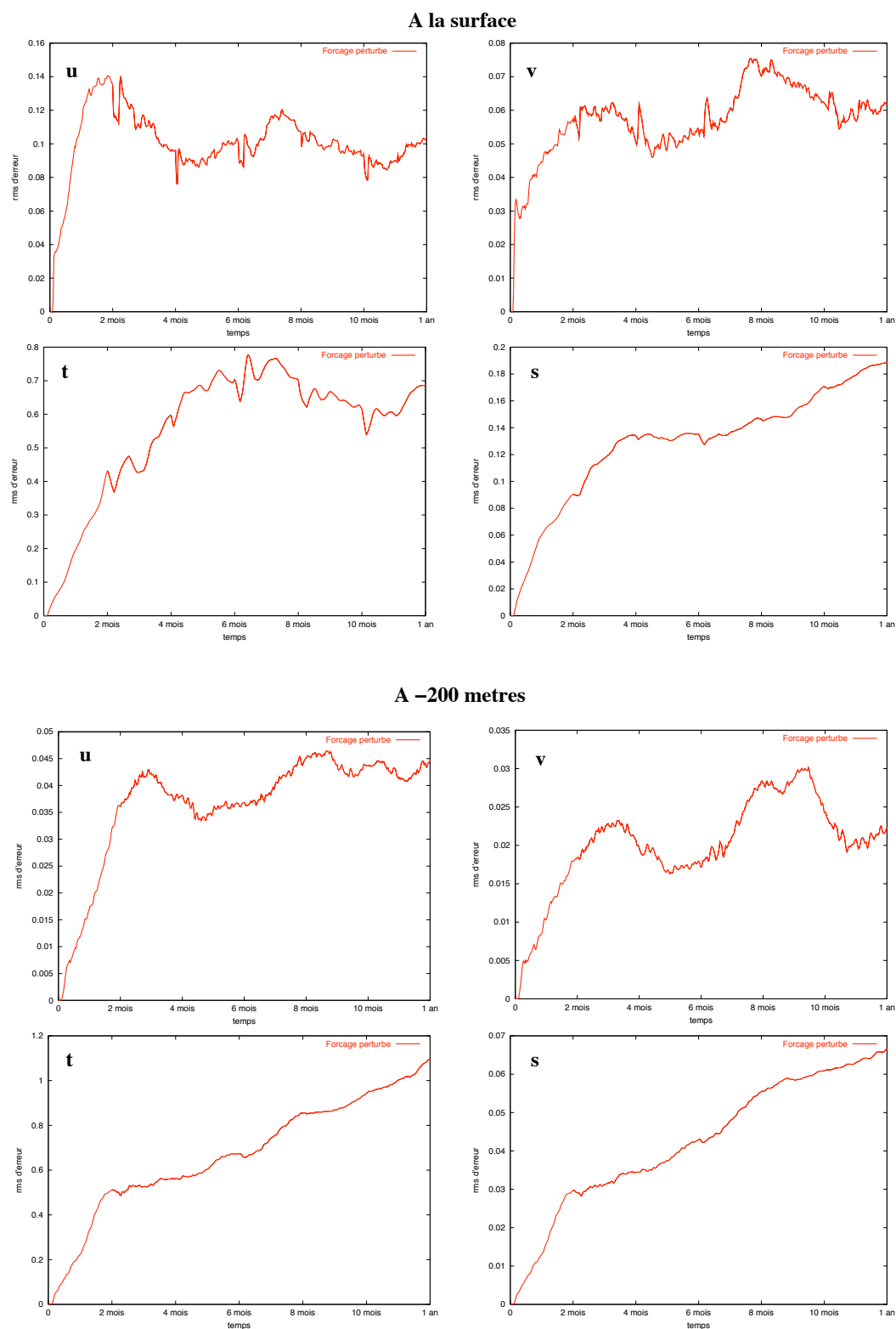


FIG. 8.3: Évolution de la norme de l'erreur sur chacune des composantes du vecteur d'état en surface (en haut) et à 200 mètres (en bas) pour des erreurs dans les forçages du modèle

8.2 Les résultats du 4D-Var incrémental sur OPA

Dans cette partie, on anticipe un peu sur la fin du chapitre prochain en présentant des résultats d'une assimilation 4D-Var incrémentale sur le modèle OPA dans le cadre d'expériences jumelles. On rappelle que le 4D-Var incrémental se caractérise par la minimisation de la fonctionnelle

$$J(\delta\mathbf{x}_0) = \frac{1}{2} \sum_{i=1}^n [\mathbf{H}_{t_i} \mathbf{M}_{t_i, t_0} \delta\mathbf{x}_0 - \mathbf{d}_i]^T \mathbf{R}_i^{-1} [\mathbf{H}_{t_i} \mathbf{M}_{t_i, t_0} \delta\mathbf{x}_0 - \mathbf{d}_i] + \frac{1}{2} [\delta\mathbf{x}_0]^T \mathbf{B}^{-1} \delta\mathbf{x}_0 \quad (8.1)$$

où $\mathbf{d}_i = \mathbf{y}_i^o - H_{t_i}(M_{t_i, t_0}(\mathbf{x}^b))$ (cf premier chapitre).

La présentation de ces résultats sera très brève le 4D-Var n'étant pas un des objectifs de cette thèse, ils seront un peu plus développés dans les chapitres suivants lors de la comparaisons avec les méthodes introduites dans cette thèse. Le but de cette partie est de mettre en évidence certaines limitations du 4D-Var qui pourront, ou ne pourront pas, être minorées par le contrôle de l'erreur systématique.

Pour toutes les expériences qui vont suivre on utilisera les mêmes matrices de covariances \mathbf{B} et \mathbf{R} . Les variances d'erreur d'ébauche qui servent à construire l'opérateur \mathbf{B} (voir chapitre 3) peuvent être différentes à chaque point de grille et ont été calculées grâce à la moyenne climatologique du modèle obtenue au moyen d'une longue intégration du modèle sans assimilation. Ces spécifications sont basées sur l'hypothèse que les erreurs d'ébauche sont plus importantes dans les régions de forte variabilité.

Les covariances d'erreur d'observation sont supposées ne pas être corrélées en espace et en temps (la matrice \mathbf{R} est diagonale). Les variances sont fixées à $(0.5\text{K})^2$ pour les données TAO et à $(1.0\text{K})^2$ pour les autres données. Ceci assurera plus de poids dans l'analyse aux données TAO qui sont généralement de meilleure qualité. L'opérateur \mathbf{H} d'observation consiste en une interpolation linéaire horizontale et verticale. Pour les données TAO, \mathbf{H} comprend également une moyenne temporelle car ces données sont disponibles en moyennes journalières.

Expériences jumelles

On a effectué dans le cadre d'expériences jumelles, un mois d'assimilation (janvier 1993). La réalité est simulée à l'aide de l'expérience GO4 de la section précédente et les trois différents types d'erreurs (sur la condition initiale, dans la physique du modèle et sur les forçages) sont introduits dans le modèle de prévision. Les observations sont distribuées comme si il s'agissait d'observations issues des bouées TAO (les profils de température uniquement, aux emplacements des bouées). L'idée ici est de se placer dans une configuration aussi proche que possible de celles qu'on pourra avoir dans des application opérationnelles (un modèle, des forçages et une condition initiale entachés d'erreur et des observations peu fournies).

La méthode d'assimilation utilisée étant un 4D-Var incrémental, il est nécessaire de réactualiser la trajectoire de référence (cf chapitre 1). Étant donnée la complexité du problème et le coût en temps de calcul d'une itération on ne peut pas effectuer la minimisation jusqu'à obtenir la convergence, on doit donc fixer un nombre maximal d'itérations. Dans l'expérience présentée ici on a fixé le nombre d'itérations à 60 avec une réactualisation de la trajectoire toutes les 10 itérations (1 boucle externe toutes les 10 boucles internes). Dans la suite cette expérience sera dénommée VI2

La figure 8.4 représente l'évolution temporelle de la norme de l'erreur sur les différentes composantes du vecteur d'état à différentes étapes de la minimisation. Chaque bande verticale représente cette évolution après chaque boucle externe de la minimisation, la première bande étant l'évolution de l'erreur si on ne fait pas d'assimilation.

Globalement le champ de température est amélioré surtout au début de la minimisation, on obtient de bons résultats sur les vitesses de courants en surface mais pas en profondeur et la salinité dérive complètement.

Sur la variable observée (la température), les effets de l'assimilation sont importants et bénéfiques que ce soit en surface ou à 200 mètres de profondeur et on retrouve la forme en "u" caractéristique du 4D-Var qui révèle la présence d'erreur modèle. La première boucle externe effectue l'essentiel du travail, les deux conditions initiales ("vraie" et *ébauche*) étant très différentes. Les boucles externes suivantes amélioreront un peu les résultats. Les vitesses de courant en surface sont plus proches de la réalité simulée au fur et à mesure que la minimisation avance jusqu'à la 3^{ème} ou la 4^{ème} boucle externe; par la suite les résultats se dégradent quelque peu sur \mathbf{v} et sont plus discutables sur \mathbf{u} (moins bon au début de la période mais meilleur à la fin en surface). Cette dégradation est vraisemblablement due au fait que la méthode qui dans un premier temps modifie les grandes échelles de la condition initiale pour se rapprocher des observations, modifie des échelles de plus en plus petites au fil de la minimisation pour arriver à des tailles caractéristiques trop petites pour la couverture spatiale des observations et/ou pas ou mal prises en compte dans la matrice de covariance d'erreur d'ébauche \mathbf{B} . Il apparaît donc à la lecture de ces résultats que de réactualiser trop fréquemment la trajectoire de référence peut nuire à la qualité du résultat si la résolution du modèle n'est pas ajustée aux réseaux d'observation (ou réciproquement). De plus, dans notre configuration, l'hypothèse linéaire tangente est bien respectée aux grandes échelles ce qui est moins vrai aux petites (WEAVER et VIALARD, discussions personnelles).

L'impact de l'assimilation sur la salinité est pour le moins catastrophique notamment en surface. Ce phénomène, d'une ampleur plus ou moins importante, est observé sur l'ensemble des résultats d'assimilation 4D-Var qui sont présentés dans cette thèse (avec ou sans contrôle de l'erreur systématique). Une des explications possible peut provenir du fait que, modifier la salinité de façon importante (et peu cohérente avec la physique) permet d'améliorer les résultats sur la température. Celle ci étant la seule composante observée, c'est ce qui prime lors de la minimisation. La salinité n'apparaissant que dans le terme J_b de la fonction coût on pourrait tenter de résoudre ce problème en accordant plus d'importance à la condition initiale en salinité, c'est à dire en diminuant les variances

d'erreur spécifiés en salinité dans la matrice \mathbf{B} . Cependant le compromis est difficile à trouver.

Ces résultats médiocres obtenus sur les vitesses de courant et surtout sur la salinité sont principalement imputables au fait que la matrice \mathbf{B} utilisée ici est sub optimale. En effet celle ci est monovariée, ainsi la transmission de l'information aux variables non observées ne se fait que grâce à l'intégration rétrograde du modèle adjoint (cf chapitre 1) ce qui est visiblement insuffisant; l'importance de ce phénomène sera illustrée au chapitre suivant. De plus les échelles de corrélation verticales de cette matrices sont difficiles à bien spécifier : soit on induit des mélanges verticaux non désirés, soit on lisse l'influence des observations sur la verticale. Une fois encore se pose le problème de l'adéquation de la résolution du modèle avec le champ d'observation utilisée. L'introduction du contrôle de l'erreur systématique ne devrait pas permettre de corriger pleinement ce problème, seule une matrice \mathbf{B} plus élaborée pourrait y parvenir. Cependant on verra dans la suite que cela réduira l'ampleur de ce phénomène, améliorera les résultats globaux et atténuera la forme en "u" de l'erreur sur la composante observée.

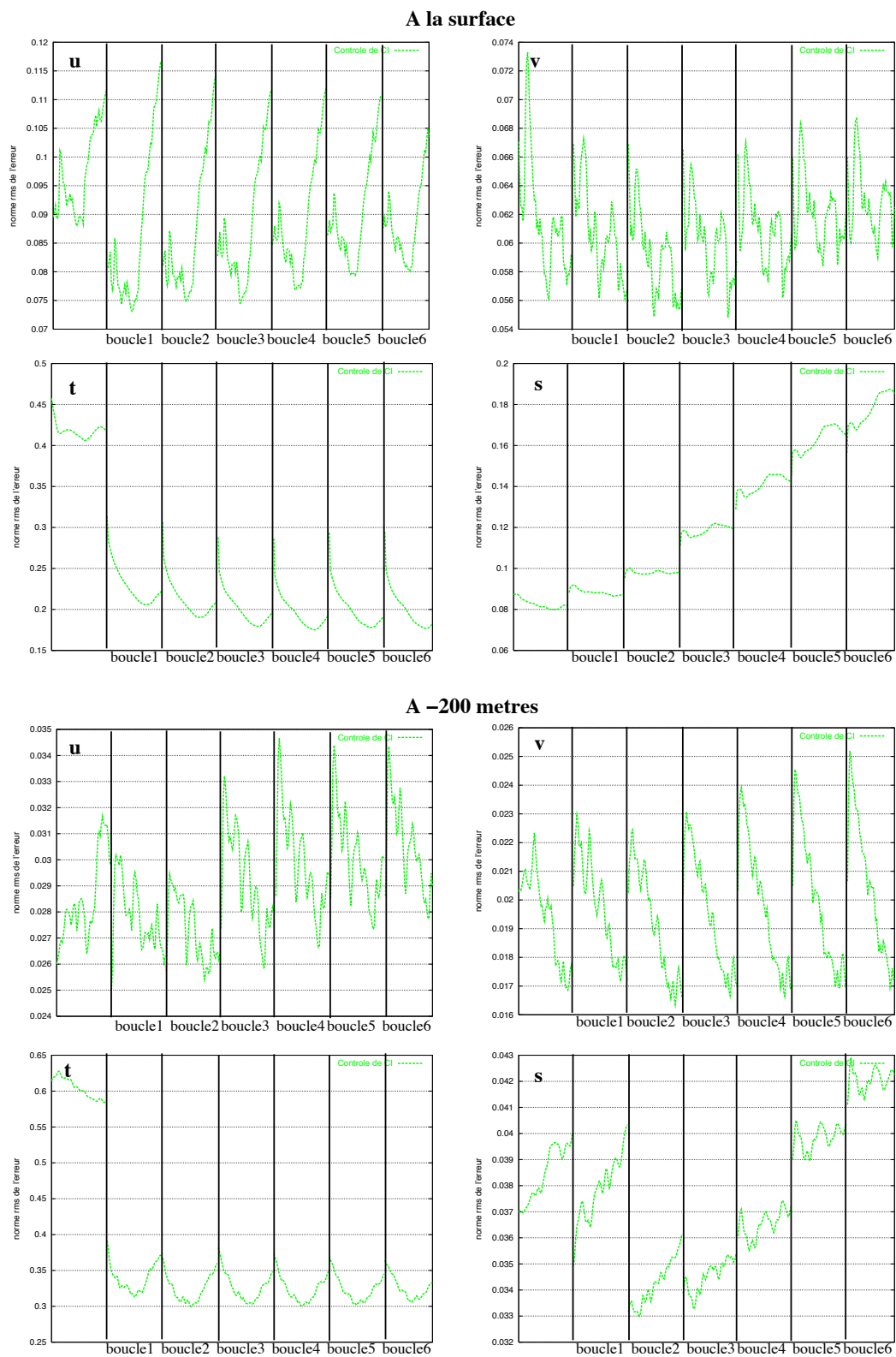


FIG. 8.4: Évolution de la norme de l'erreur sur chacune des composantes du vecteur d'état en surface (en haut) et à 200 mètres (en bas) suivant plusieurs boucles externes

Chapitre 9

Contrôle de l'erreur systématique : Expériences jumelles

Dans ce chapitre, on se place une nouvelle fois dans le cadre d'expériences jumelles. La réalité est, comme à la fin du précédent chapitre, simulée à l'aide de l'expérience GO4, c'est à dire en utilisant un schéma TKE pour résoudre le mélange vertical, la bonne condition initiale et les bons forçages.

Les trois différents types d'erreur introduits au chapitre précédent (sur la condition initiale, dans la physique du modèle, sur les forçages) sont, là encore, présents dans le modèle de prévision.

Grâce à l'utilisation de ces deux modèles (modèle "vrai" et modèle de prévision) on a simulé la présence d'erreur dans le modèle de prévision. Celle ci a été calculée en intégrant le modèle de prévision à partir de l'état initial "vrai" et en faisant un rappel à la trajectoire "vraie" à chaque pas de temps. L'erreur systématique étant la moyenne de l'erreur η_i à chaque pas de temps (cf figure 9.1).

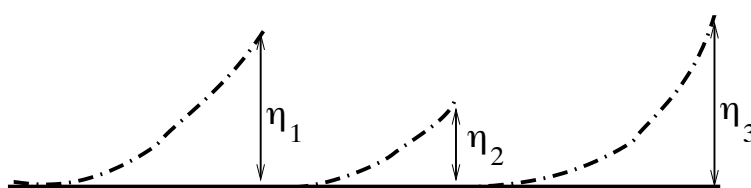


FIG. 9.1: calcul de l'erreur η_i par intégration du modèle de prévision (trait pointillé) à partir de la trajectoire "vraie" (trait continu)

La composante en température de la part systématique de l'erreur introduite dans le modèle est représentée sur les figures 9.2.

Comme on l'a vu à la fin du chapitre précédent, les informations sur les autres composantes sont insuffisantes (observations ou matrice de covariances). Afin de limiter un peu le nombre des figures on ne présentera donc généralement que les résultats sur la température, plus représentatifs.

La combinaison des erreurs modèle (dans la physique et dans les forçages) et l'utilisation d'une condition initiale fautive impliquent d'importantes erreurs dans l'estimation des champs

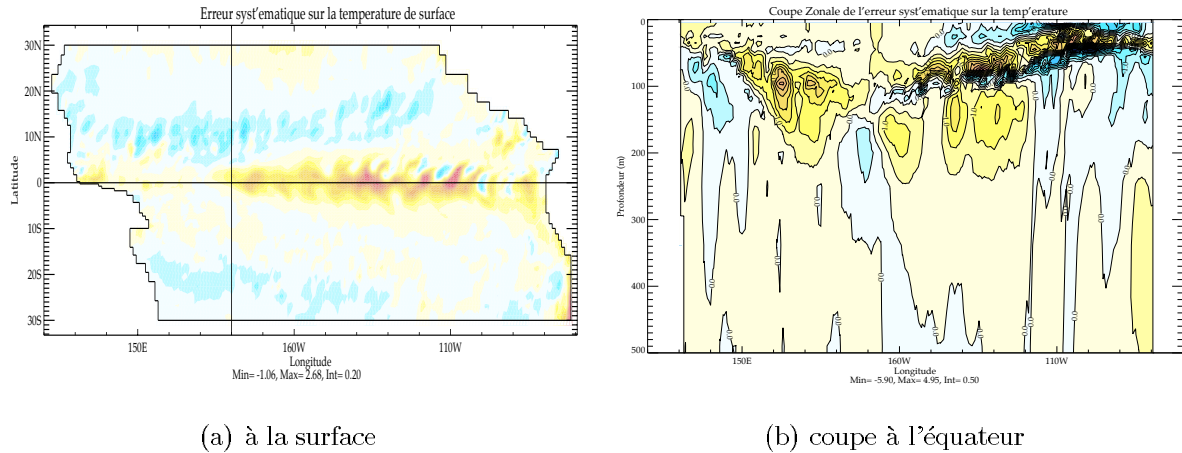


FIG. 9.2: Champ de surface et coupe à l'équateur de l'erreur systématique présente dans le modèle sur la composante de température du vecteur d'état

des quatre composantes du vecteur d'état. La figure 9.3 (a) représente cette erreur sur la température à l'instant initial de la période d'intégration du modèle, en coupe à l'équateur (on ne fait pas encore d'assimilation). Celles-ci représentent uniquement l'écart entre la condition initiale utilisée pour simuler la réalité et celle utilisée pour notre expérience d'assimilation. La figure 9.3 (b) décrit ces mêmes erreurs après 1 mois d'intégration du modèle. On s'aperçoit que la zone de température trop chaude à l'ouest du bassin (entre 100 et 200 mètres) a diminué en taille et en amplitude alors que la zone d'eau trop froide à l'est a augmenté en intensité. Ce phénomène s'explique facilement : comme on l'a montré au chapitre précédent l'erreur sur la condition initiale s'atténue avec le temps, mais dans la zone à l'est dans les 200 premiers mètres, les erreurs modèle induisent un biais (voir figure 9.2) qui amplifie l'erreur initiale.

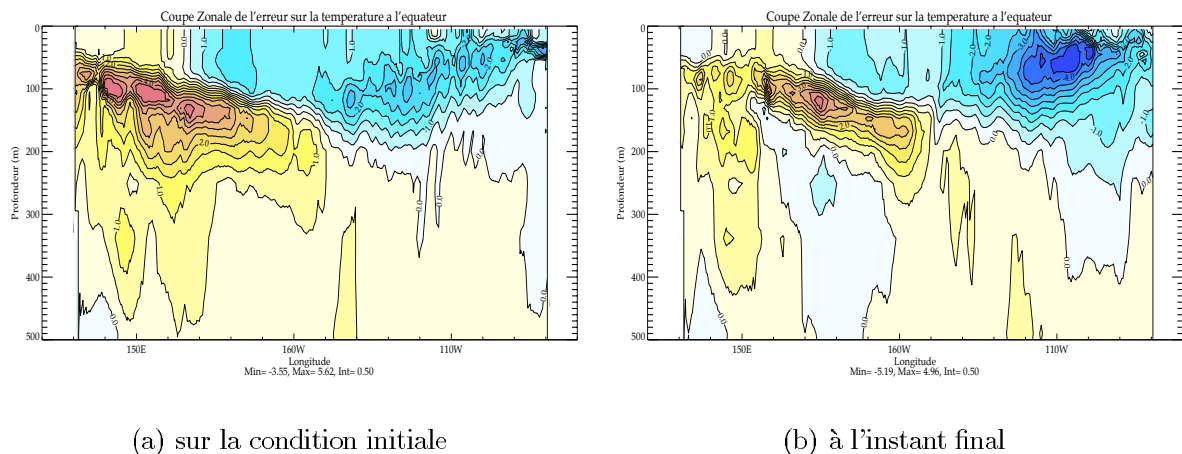


FIG. 9.3: coupe à l'équateur de l'erreur sur la condition initiale du modèle libre (a) et après un mois d'intégration du modèle libre (b) sur la composante de température du vecteur d'état

Afin de tenter de corriger ces erreurs on va appliquer notre méthode de contrôle de

la part systématique de l’erreur modèle et on la comparera au contrôle de la condition initiale dans deux configurations différentes. La première sera une configuration favorable aux deux méthodes d’assimilation de données avec un nombre conséquent d’observations (le problème sera surdéterminé) alors que la deuxième sera un cas quasi-réaliste simulant un réseau d’observation identique à celui que nous utiliserons au chapitre suivant pour assimiler des données réelles. Dans les deux cas on a fixé le nombre d’itérations à 60 (6 boucles externes de 10 itérations).

9.1 Configuration favorable

9.1.1 Description de l’expérience

À partir du modèle de prévision décrit dans la section précédente, on effectue un mois d’assimilation (janvier 1993) en introduisant le contrôle de l’erreur systématique en plus de la condition initiale (expérience VT2). Dans un premier temps on utilisera comme observation le champ total de température issu de l’intégration du modèle “vrai” une fois tous les trois jours. On comparera les résultats de VT2 à ceux d’une assimilation 4D-Var classique (expérience VI2). L’opérateur \mathbf{B} est le même qu’à la fin du chapitre précédent et l’opérateur \mathbf{Q} en conserve les structures de corrélation mais les variances ont été recalculées.

9.1.2 Résultats

Dans cette configuration favorable pour toutes les méthodes d’assimilation de données (le problème est surdéterminé), l’introduction du contrôle de l’erreur systématique améliore sensiblement les résultats du 4D-var. Si on compare les champs de l’erreur systématique estimée à celle réellement présente dans le modèle on s’aperçoit qu’en surface la structure de l’erreur en température est bien respectée mais l’erreur est sous-évaluée (figure 9.4).

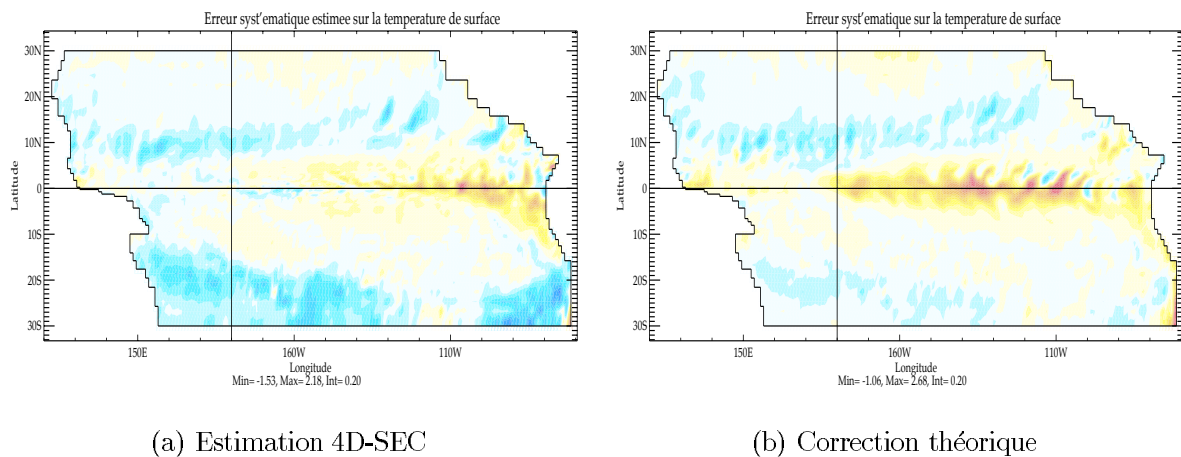


FIG. 9.4: Erreur systématique sur la température à la surface estimée par l’expérience VT2 (a) et présente dans le modèle (b)

De plus certaines petites structures sont mal ou ne sont pas représentées. Ceci est peut-être dû au fait que l'hypothèse linéaire tangente n'est pas respectée pour les plus petites échelles.

La vue en coupe à l'équateur laisse apparaître un tout autre problème : une bande de refroidissement (en bleu) présente dans le terme de correction idéal théorique n'est pas du tout représentée par le 4D-SEC (figure 9.5). Ce manque est certainement dû à des échelles de corrélation verticale trop importantes dans l'opérateur \mathbf{Q} . Cependant des échelles plus fines ne permettraient pas de transmettre suffisamment l'information dans d'autres zones. L'approche employée ici pour construire les opérateurs de covariance peut difficilement représenter de telles structures (fines strates), il conviendrait peut-être d'adopter une approche plus en relation avec la physique du modèle ou tout au moins à affiner encore plus les structures de corrélation verticales suivant les localisations géographiques.

On peut de plus constater qu'une bulle d'eau trop chaude apparaît en sub-surface à l'ouest du bassin. Il est possible que cette mauvaise estimation soit liée avec la très mauvaise estimation de l'erreur en salinité notamment sur cette zone. L'utilisation d'une matrice de covariance d'erreur \mathbf{Q} multivariée permettrait sans doute de corriger ce défaut.

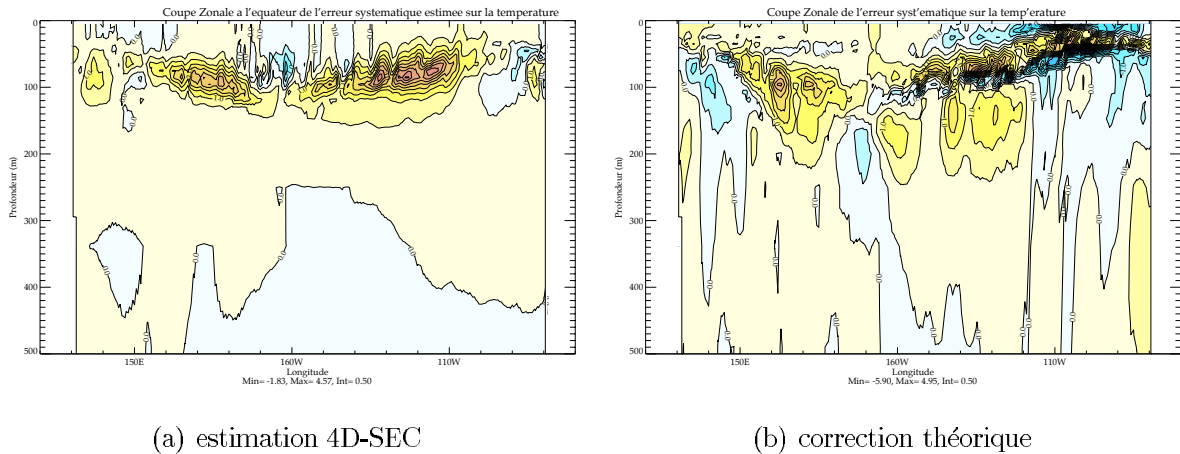


FIG. 9.5: coupe à l'équateur de l'erreur systématique sur la température estimée par l'expérience VT2 (a) et présente dans le modèle (b)

Les figures 9.6 représentent les désormais classiques évolutions de la norme de l'erreur en fonction du temps pour la période d'assimilation et la période de prévision associée, à différentes profondeurs de l'océan et pour chaque composante du vecteur d'état.

Durant la période d'assimilation on a amélioré la qualité de la trajectoire estimée et réduit le niveau d'erreur modèle (la courbe d'erreur moins convexe) sur l'ensemble des composantes du vecteur d'état (y compris les variables non observées, à l'exception peut-être de la vitesse méridionale en surface). L'erreur systématique est, pour la température et la vitesse de courant zonale, bien estimée par la méthode (l'erreur de prévision croît moins vite que après une assimilation 4D-var classique), ce qui est moins évident pour la salinité et surtout la vitesse méridionale. Cependant, même pour ces deux composantes, le niveau d'erreur durant la période de prévision reste plus faible.

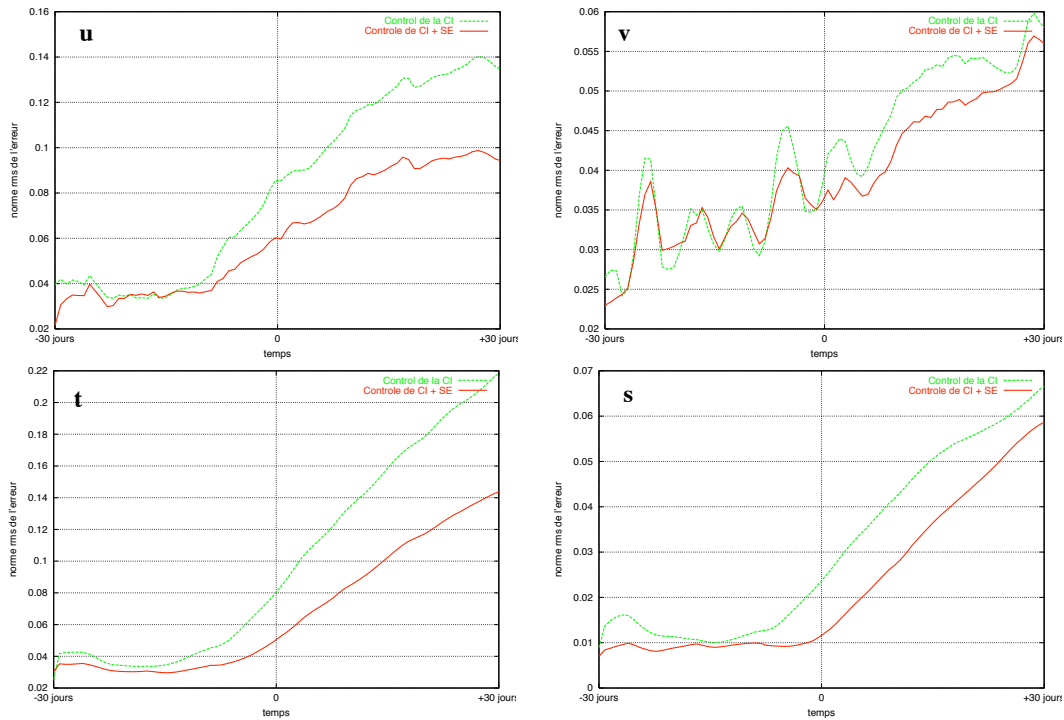


FIG. 9.6: Évolution de la norme de l'erreur sur chacune des composantes du vecteur d'état en surface après assimilation (expérience VT2 et VI2)

9.1.3 Premières conclusions

Les améliorations dues à l'introduction du contrôle de l'erreur systématique dans l'assimilation 4D variationnelle sont significatives même sur les composantes du vecteur d'état qui ne sont pas observées. Cependant, cet apport n'est pas aussi évident que dans le cas des expériences avec le modèle Shallow-Water. Les causes de cette moins bonne efficacité peuvent être multiples :

- Le principal problème provient sans doute, encore et toujours, du fait que les matrices de covariance d'erreur ne sont pas multivariées. Même si ce défaut handicape aussi le 4D-Var classique on peut considérer que améliorer, par l'utilisation d'une matrice \mathbf{B} multivariée, l'estimation de la bonne condition initiale permettra également d'améliorer l'estimation de l'erreur systématique et l'utilisation d'une matrice \mathbf{Q} multivariée sera un apport supplémentaire.
- Plus spécifique au contrôle de l'erreur, le fait d'utiliser une formulation incrémentale : il est très possible que si l'hypothèse linéaire tangente n'est pas suffisamment respectée, le 4D-SEC tente de corriger l'erreur $M(\mathbf{x} + \delta\mathbf{x}) - (M(\mathbf{x}) + \mathbf{M}\delta\mathbf{x})$ en plus de l'erreur modèle proprement dite. De plus une partie de l'erreur introduite (sur les forçages) n'est plus présente dans le modèle tangent et n'est donc pris en compte que par la réactualisation de la trajectoire de référence.

- Le coût en calcul de la minimisation nous impose de n'effectuer qu'une soixantaine d'itérations alors qu'on en faisait une centaine avec le Shallow-Water. On avait alors à peu près 40 000 valeurs à estimer (la taille du vecteur d'état multiplié par 2) et on est porté dans le cas présent à un peu plus de 2 000 000. On est naturellement amené à penser qu'augmenter le nombre d'itérations aurait un effet bénéfique sur l'efficacité de la minimisation. Il est vrai qu'on a vu dans le chapitre précédent sur l'expérience VI4 qu'augmenter le nombre d'itérations pouvait être néfaste; cependant ce problème sera sans doute partiellement éliminé par l'utilisation de matrices multivariées (décidément !). Cela ne suffira sans doute pas car, au fur et à mesure des boucles externes, on s'intéresse à des échelles de plus en plus petites pour arriver à des tailles caractéristiques qui ne sont plus résolues par un ou plusieurs des acteurs de l'assimilation (modèle, observations, \mathbf{B} , hypothèse linéaire tangente).

9.2 Comme en “vrai”

Jusqu'ici on a assimilé des observations simulées représentant la totalité de la température, ceci est évidemment peu réaliste. Dans cette section on va se placer dans une configuration très proche de ce que l'on pourrait avoir si on avait à assimiler des données réelles. Cette approche permettra de cerner les éventuelles difficultés que l'on sera amené à rencontrer lors des applications réalistes.

9.2.1 Description de l'expérience

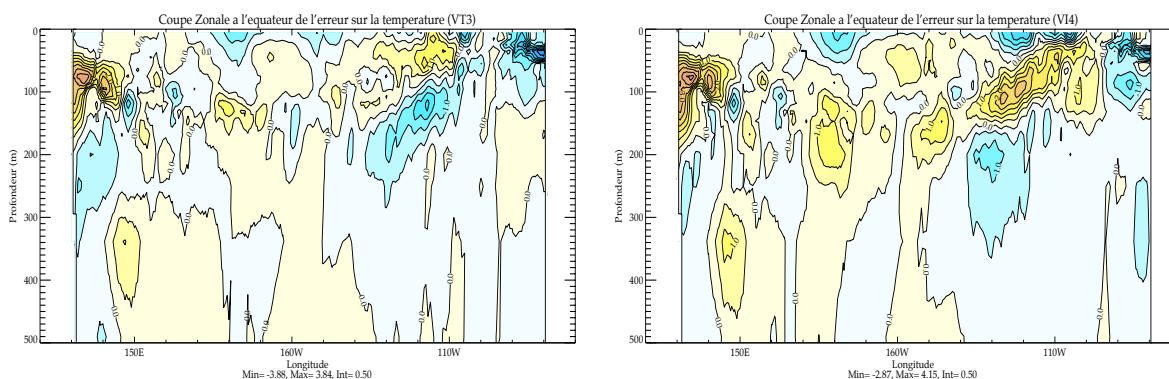
On se place dans le même cas que dans la section précédente : les observations sont générées par l'expérience GO4 et le modèle de prévision est le même que dans l'expérience VT2. La période étudiée est toujours la même (janvier 1993), cependant les observations simulées ne sont plus composées du champ total de température, une fois tous les trois jours, mais de ce qu'on aurait comme observations fournies par des bouées TAO sur la même période (uniquement les profils de température). Les matrices de covariances \mathbf{B} , \mathbf{Q} et \mathbf{R} sont les mêmes que dans l'expérience précédente.

9.2.2 Résultats

On présentera les résultats du 4D-SEC sur cette configuration (expérience VT3) et on les comparera avec ceux de l'expérience 4D-Var dans la même configuration (expérience VI4 déjà présentée au chapitre précédent).

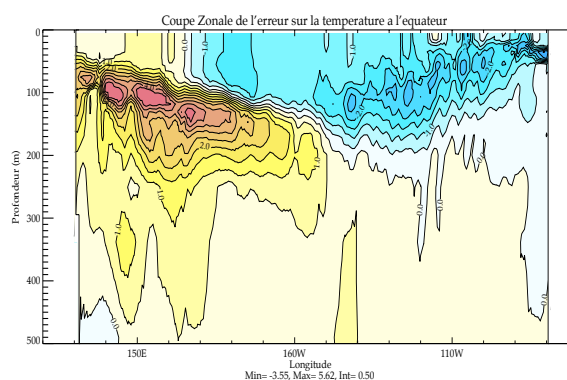
Les résultats sur la température sont similaires pour les deux expériences. Si on compare (a) et (b) de la figure 9.7, on s'aperçoit que la condition initiale est légèrement mieux estimée par VT3 que par VI4 (cette dernière étant pourtant déjà efficace en comparaison de l'ébauche (c)).

Si on compare les ordres de grandeur des deux incréments (d) et (e) calculés par ces méthodes et celui de l'erreur d'ébauche (on compare ici une erreur et un terme de correction, ils sont donc de signes opposés) on voit que le 4D-Var surévalue la correction dans certaines zones. Ceci est expliqué par le principe déjà exposé : le 4D-Var tente de corriger les erreurs

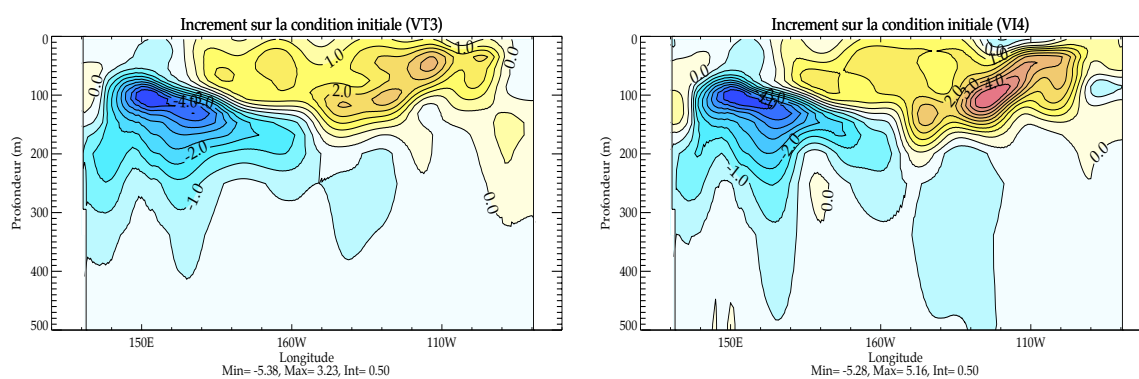


(a) erreur résiduelle 4D-SEC

(b) erreur résiduelle 4D-VAR



(c) erreur d'ébauche



(d) incrément 4D-SEC

(e) incrément 4D-Var

FIG. 9.7: coupe à l'équateur de l'erreur résiduelle sur la condition initiale sur la composante de température du vecteur d'état pour les expériences VT3 (a), VI4 (b) après application de l'incrément calculé par ces deux méthode (d) et (e) en correction de l'erreur d'ébauche(c)

modèle (systématiques et aléatoires) en agissant sur la condition initiale et introduit donc de l'erreur sur celle ci.

Ici on assiste également au phénomène inverse : Le 4D-SEC tente de corriger une partie

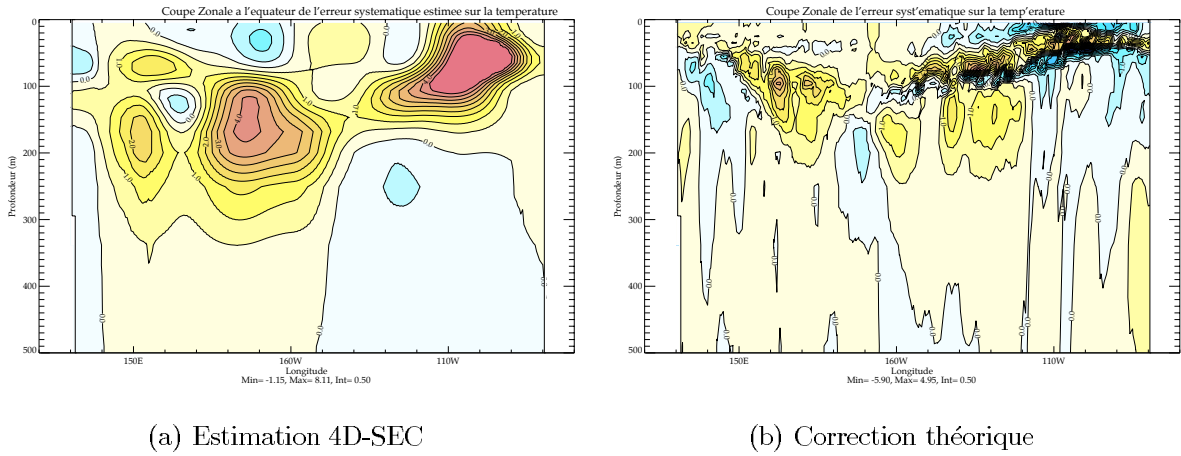


FIG. 9.8: coupe à l'équateur de l'erreur systématique sur la température estimée par l'expérience VT3 (a) et présente dans le modèle (b)

de l'erreur sur la condition initiale en agissant sur le terme de correction d'erreur modèle (surestimation du terme de correction de l'erreur systématique à l'ouest et bulle bleue à l'est du (a) de la figure 9.8). Si on effectue moins d'itérations de minimisation, l'erreur systématique est mieux estimée, on ne retrouve toujours pas la fine couche bleue, mais les ordres de grandeur et les emplacements sont meilleurs. Cependant on obtient alors un moins bon résultat global. Un bon compromis semble être ces 10×6 itérations.

Profitons du fait que l'on soit encore dans le cas d'expériences jumelles (c'est la dernière fois) pour étudier l'évolution de la norme de l'erreur sur les figures habituelles 9.9. On ne présentera cette fois que la rms à la surface car elle est représentative de l'ensemble du bassin.

À première vue les résultats sont surprenants : L'introduction du contrôle de l'erreur modèle améliore les résultats du 4D-var de façon nettement plus importante que dans le cas précédent pourtant plus favorable (sauf peut-être sur \mathbf{u}). À y regarder de plus près c'est surtout du au fait que le 4D-var fournit des résultats assez contrastés.

Comme on l'avait remarqué dans la brève introduction de l'expérience VI4 dans le chapitre précédent, les résultats du 4D-Var sont plutôt bons sur la température, mitigés sur les vitesses de courant et catastrophiques sur la salinité. La figure 9.10 représente l'évolution temporelle de la norme de l'erreur sur les différentes composantes du vecteur d'état à différentes étapes de la minimisation. Chaque bande verticale représente cette évolution après chaque boucle externe de la minimisation, la première bande représentant l'évolution de l'erreur si on ne fait pas d'assimilation. Dans le cas plus favorable du début de chapitre (expériences VI2 et VT2), le niveau d'erreur diminuait au fil de la minimisation sur l'ensemble des variables. Dans les expériences VT3 et VI4 cela reste vrai pour la variable observée (température). L'erreur sur les vitesses de courant continue à diminuer légèrement après la première boucle externe de VT3 alors que les résultats se dégradent pour l'expérience VI4. La différence entre les résultats des deux méthodes est encore plus évident sur la salinité : même si l'assimilation VT3 dégrade aussi légèrement la qualité de l'estimation de cette composante du vecteur d'état, c'est sans commune mesure avec

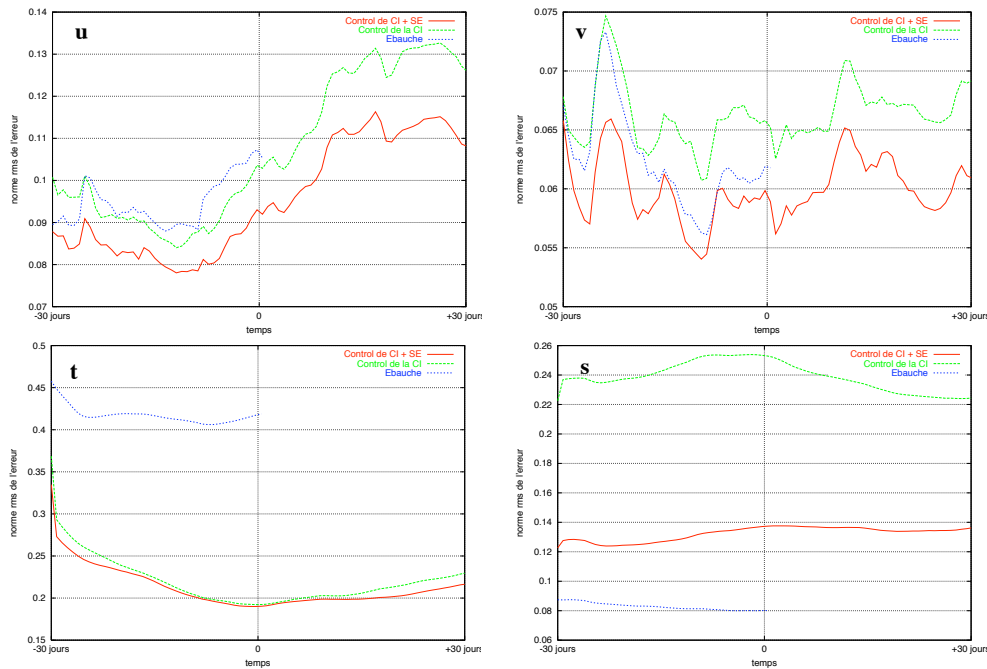


FIG. 9.9: Évolution de la norme de l'erreur sur chacune des composantes du vecteur d'état en surface après assimilation pour les expériences VI4 (trait vert pointillé) et VT3 (rouge continu)

l'action (néfaste) de VI4.

L'introduction du contrôle de l'erreur systématique a permis d'enrayer ces dégradations au fil de la minimisation. Comme seule la température est observée et que la matrice \mathbf{B} est univariée, les champs en salinité ne sont pas pris directement en compte pour la minimisation. Il est peut-être plus efficace pour le 4D-Var, pour réduire l'écart aux observations, de modifier fortement la salinité (ce qui induira des modifications sur les températures) que de modifier légèrement et directement la température. La présence d'erreur modèle implique aussi que tout ne peut pas être corrigé par des ajustements sur la condition initiale, le 4D-Var tente vraisemblablement de corriger une partie du biais induit par ces erreurs sur la température en “prenant des libertés” sur la salinité. Le fait que Le 4D-SEC agisse différemment sur le modèle (un incrément à chaque pas de temps) atténue ces phénomènes.

Si on focalise notre attention sur les résultats obtenus par nos deux méthodes sur la zone où se trouvent nos bouées TAO (10° Nord-10° Sud, 150° Est-90° Ouest, sur les 500 premiers mètres) on s'aperçoit que la qualité de l'estimation de l'erreur est améliorée grâce à la bonne densité d'information fournie sur cette zone (figure 9.11). Ceci se traduit par une prévision de plus grande qualité. Si on compare les résultats des deux expériences VT3 et VI4 sur les vitesses de courant avec celui issu de l'ébauche (courbe bleue ou pointillés

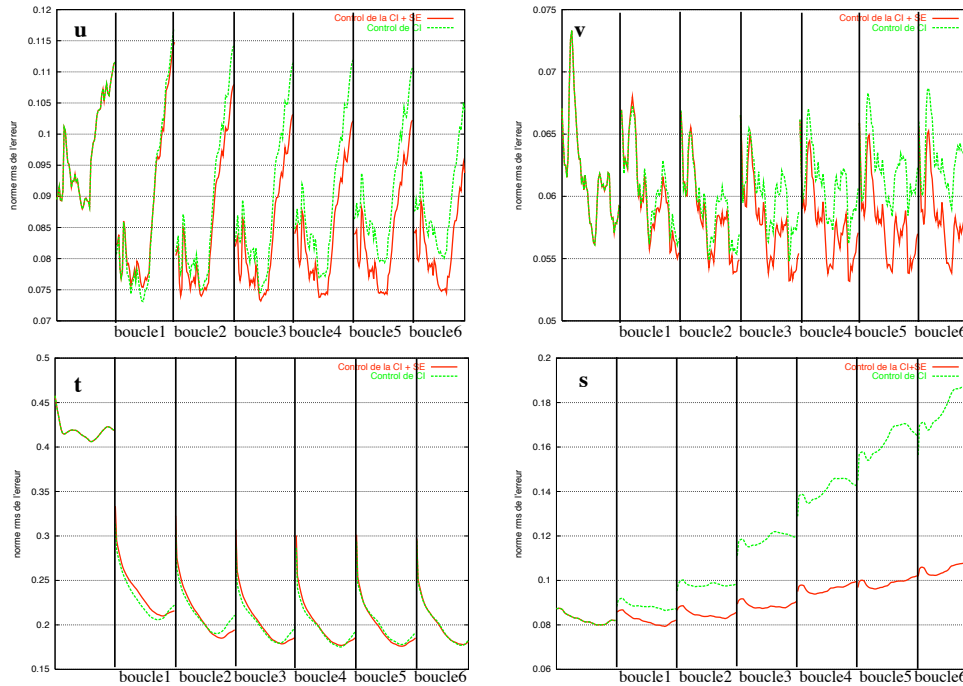


FIG. 9.10: *Évolution de la norme de l'erreur sur chacune des composantes du vecteur d'état en surface suivant plusieurs boucles de minimisation pour les expériences VI4 (trait vert pointillé) et VT3 (rouge continu)*

épais), on s'aperçoit qu'à la fin de la minimisation le fait d'appliquer une assimilation 4D-Var classique (VI4) a un effet plutôt néfaste sur ces composantes alors que le contrôle de l'erreur modèle permet d'améliorer les résultats. Cela s'explique par le fait que les erreurs modèle introduites (le vent et le mélange vertical) influent directement sur ces deux composantes. Il est donc difficile d'en corriger efficacement la trajectoire en ne contrôlant que la condition initiale.

9.3 Conclusion des expériences jumelles

Ces expériences ont une nouvelle fois montré l'intérêt de prendre en compte les erreurs modèle en assimilation variationnelle. Quelle que soit la configuration testée en expériences jumelles, le 4D-SEC fournit de meilleurs résultats pour un coût en temps de calcul équivalent. De plus l'introduction du contrôle de l'erreur systématique dans une chaîne d'assimilation 4D-Var déjà existante ne requiert quasiment pas de développement supplémentaire. Cependant, contrairement à ce qui se passait dans le cas du Shallow-Water, les marges de progressions sont encore importantes. En effet, de nombreux développements peuvent encore être faits principalement sur les modèles adjoint et linéaire tangent et sur les matrices de covariance d'erreur. Ce dernier point permettra sans doute d'améliorer grandement les résultats sur les deux méthodes. Des opérateurs multivariés résoudraient sans doute une grande partie des problèmes mis en évidence sur la salinité et dans une moindre mesure sur les vitesses de courant. De plus il serait peut-être utile d'explorer d'autres voies pour

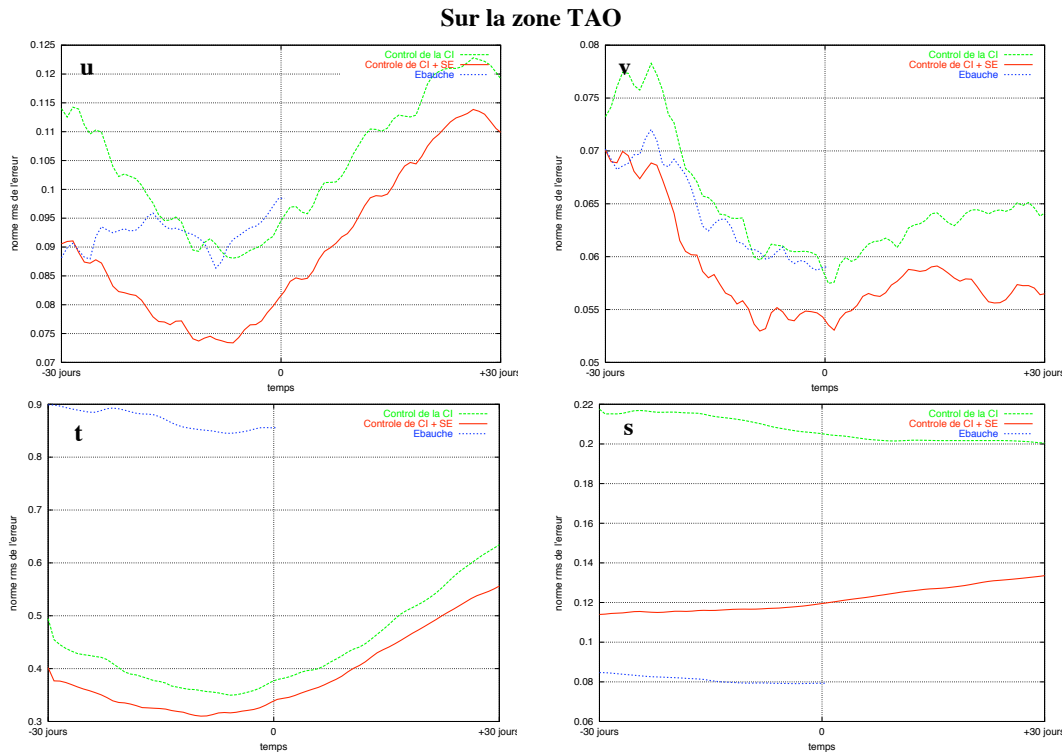


FIG. 9.11: Évolution de la norme de l'erreur sur chacune des composantes du vecteur d'état sur l'ensemble de la zone TAO pour les expériences VI₄ (trait vert pointillé) et VT₃ (rouge continu) et sans assimilation (pointillé bleu)

la construction des opérateurs de covariance d'erreur : sur l'ensemble des incréments issus de l'assimilation (sur la condition initiale ou sur l'erreur) représentés dans ce chapitre, on voit aisément la structure en "bulle" de ces incréments engendrés par la construction intrinsèque de ces opérateurs. Il se peut que le choix de ces corrélations spatiales ne soit pas judicieux au regard de celles des erreurs réellement présentes dans le modèle et dans l'ébauche.

En résumé on peut dire que d'autres facteurs pèsent plus que la présence d'erreur modèle sur les performances du 4D-Var utilisé ici comme référence et qu'il conviendra de les traiter si on veut que cela deviennent une chaîne d'assimilation opérationnelle compétitive. Cependant, en raison de son faible surcoût et de sa facilité de mise en place, il n'est pas nécessaire d'attendre que le 4D-Var ait atteint son maximum de performance pour y introduire le contrôle de l'erreur systématique et d'appliquer cette méthode à une assimilation de données réelles. C'est ce que l'on se propose de faire dans le chapitre suivant.

Chapitre 10

Contrôle de l'erreur systématique : Expériences réalistes

Cet ultime chapitre traite des expériences d'assimilation de données variationnelles avec contrôle de l'erreur systématique en assimilant des données réelles. Le but de ce chapitre n'est pas d'étudier de façon exhaustive la qualité des résultats physiques, ce qui sort du cadre de cette thèse, et devra être confié aux océanographes. Seuls quelques résultats précis sont révélés ici et comparés avec ceux du 4D-Var incrémental classique afin de mettre en évidence les améliorations apportées et les défauts à régler.

10.1 Configuration

La configuration utilisée pour cette expérience est presque identique à celle présentée au chapitre précédent, à savoir le modèle OPA restreint au domaine pacifique tropical. Les forçages sont fournis par les moyennes hebdomadaires des vents mesurés par le satellite ERS, ainsi que les flux ECMWF et le rappel à la température de surface se fait grâce à la climatologie hebdomadaire de Reynolds. L'assimilation de données est appliquée sur une période de 6 ans (début 1993- fin 1998). L'hypothèse linéaire tangente n'étant valide que sur approximativement un mois, on a effectué ces 6 ans d'assimilation par cycles de 1 mois avec, là encore, 6 boucles externes de 10 boucles internes chacune. Les données assimilées sont cette fois réelles et issues des mouillages TAO (profils de température seulement) auxquelles viennent s'ajouter les profils de sondes XBT (température également).

L'opérateur \mathbf{B} et la matrice \mathbf{R} sont construits de la même façon que pour les précédentes expériences d'assimilation. Comme c'est le cas pour toutes les expériences sur OPA présentées dans cette thèse on utilise des variantes incrémentales du 4D-Var et du 4D-SEC. L'opérateur de covariance d'erreur modèle \mathbf{Q} est ici pris égal à \mathbf{B} . Ce choix est purement arbitraire car rien ne dit que les covariances de l'erreur modèle soient les mêmes que celles d'ébauche. Cependant ce choix permet de s'assurer d'obtenir un bon conditionnement pour la minimisation et de réaliser des économies sur le calcul de \mathbf{Q} ; de plus, l'opérateur \mathbf{B} ne représentant pas exactement les covariances d'erreur d'ébauche, mais étant une approximation basée sur l'hypothèse que les erreurs d'ébauche sont plus importantes dans les

régions de forte variabilité, il ne semble pas absurde de penser qu'il en va de même pour les erreurs modèle. Une étude plus poussée sur les différences entre ces deux opérateurs serait à mener en considérant les connaissances physiques que l'on peut avoir sur ces deux types d'erreur.

Un autre aspect de ces opérateurs de covariance d'erreur qui n'a pas été traité ici est le fait que, dans un 4D-Var classique, si le modèle n'est pas exact, le plus efficace n'est pas de retrouver exactement la vraie condition initiale, mais une condition initiale légèrement erronée afin de corriger une partie des erreurs modèle. Ceci fait que l'opérateur \mathbf{B} du 4D-Var ne devra pas représenter les covariances d'erreur d'ébauche exactement mais prendre en compte ce fait. Si on contrôle également un terme de correction d'erreur, il ne sera plus nécessaire de prendre en compte les erreurs modèle dans l'opérateur \mathbf{B} . On devra donc distinguer les opérateurs \mathbf{B} du 4D-Var et du 4D-SEC.

10.2 Résultats

Les résultats du 4D-Var présentés dans cette section sont issus des simulations effectuées par *WEAVER et al* [128] elles serviront de base de comparaison avec la méthode de contrôle de la part systématique de l'erreur modèle.

Globalement, les résultats de ces simulations 4D-Var sont relativement satisfaisants, notamment en ce qui concerne les champs de température et de courants de surface. Ces deux caractéristiques justifient à elles-seules le fait de faire de la modélisation océanique de part leur importance dans la prévision climatique et notamment des phénomènes El niño. Cependant tous les aspects des résultats du 4D-Var ne sont pas satisfaisants. C'est sur ces faiblesses que nous allons insister dans cette section afin de mettre en évidence les améliorations qui pourront être apportées par le contrôle de l'erreur systématique.

10.2.1 Une qualité conservée

On compare donc ces résultats avec ceux issus de l'assimilation 4D-SEC dans la même configuration. Avant d'entrer plus dans les détails, on peut remarquer que, d'une façon générale, lorsque le 4D-Var est efficace, le 4D-SEC conserve la bonne qualité des résultats. Sans entrer trop dans les détails, on peut en citer deux exemples :

- L'altimétrie :

On dispose de nombreuses informations sur le niveau de l'océan grâce aux mesures altimétriques fournies par les satellites. Ces données ne sont pas directement comparables aux sorties du modèle car on a opté pour une configuration "à toit rigide" qui implique que le niveau de la mer ne varie pas. Cependant on peut calculer la hauteur dynamique de l'océan au moyen des valeurs de \mathbf{T} et \mathbf{s} .

On peut aussi considérer l'altimétrie dans une zone où le 4D-Var est performant. À cet endroit les résultats sont d'une qualité similaire pour le 4D-Var et le 4D-SEC et même légèrement meilleur pour ce dernier, mais la différence n'est pas significative.

Les deux méthodes améliorent grandement les résultats du modèle libre (l'intégration du modèle libre de toute assimilation) (figure 10.1).

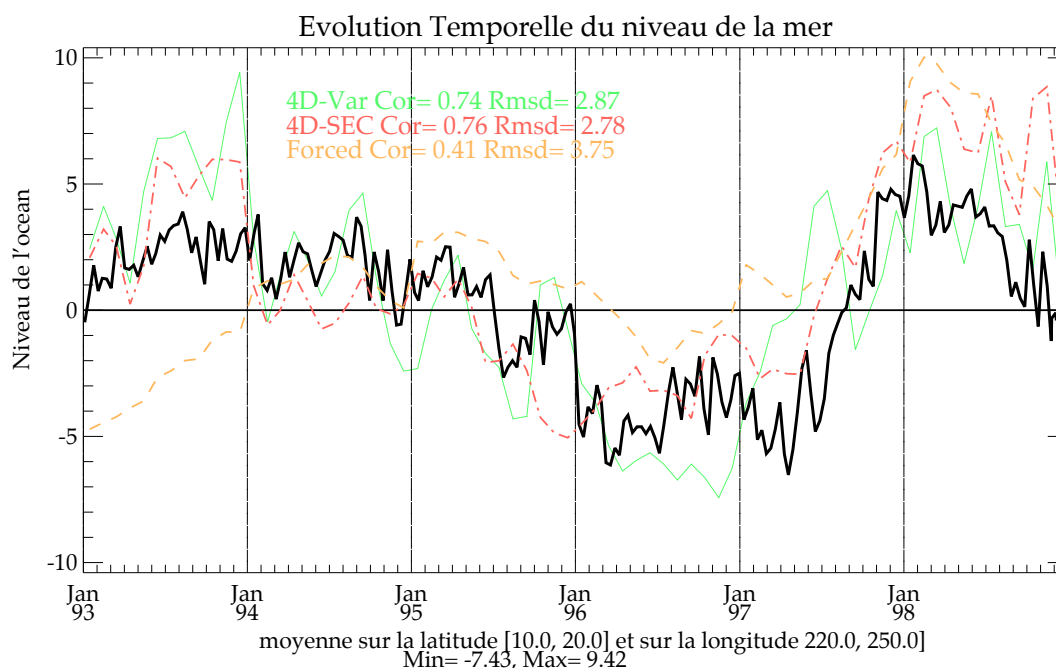


FIG. 10.1: Comparaison aux données altimétriques TOPEX/POSEIDON (trait noir épais) des hauteurs d'eau issues du 4D-Var (trait rouge continu), du 4D-SEC (vert mixte) et du modèle libre (orange pointillé)

- La thermocline

La thermocline principale est une couche d'eau caractérisée par une diminution rapide de la température en fonction de la profondeur, elle se situe en deçà de la couche de mélange, qui représente en général les 40 premiers mètres dans la zone qui nous intéresse, et se prolonge jusqu'à ce que l'eau atteigne à peu près 12°C (représentée sur les deux premiers dessins de la figure 10.2 par les bandes resserrées entre 24 et 12°C). Un phénomène bien connu des océanographes est le resserrement de cette thermocline dans l'océan pacifique tropical notamment au niveau de la latitude 10°Nord . Le modèle libre représente assez mal cette caractéristique ((c) de la figure 10.2 et/ou VIALARD et al, 2001 [122]); les assimilations 4D-Var (a) et 4D-SEC (b) permettent de corriger efficacement cette imperfection et de redresser la thermocline. Une bonne représentation de cette thermocline est importante car sa pente et son resserrement vont induire les gradients de pression responsables des courants équatoriaux.

On s'aperçoit notamment que, dans le modèle libre, le resserrement entre 5°Nord et 10°Nord est très mal représenté, on peut donc s'attendre à ce que le contre courant nord-équatorial soit mal représenté.

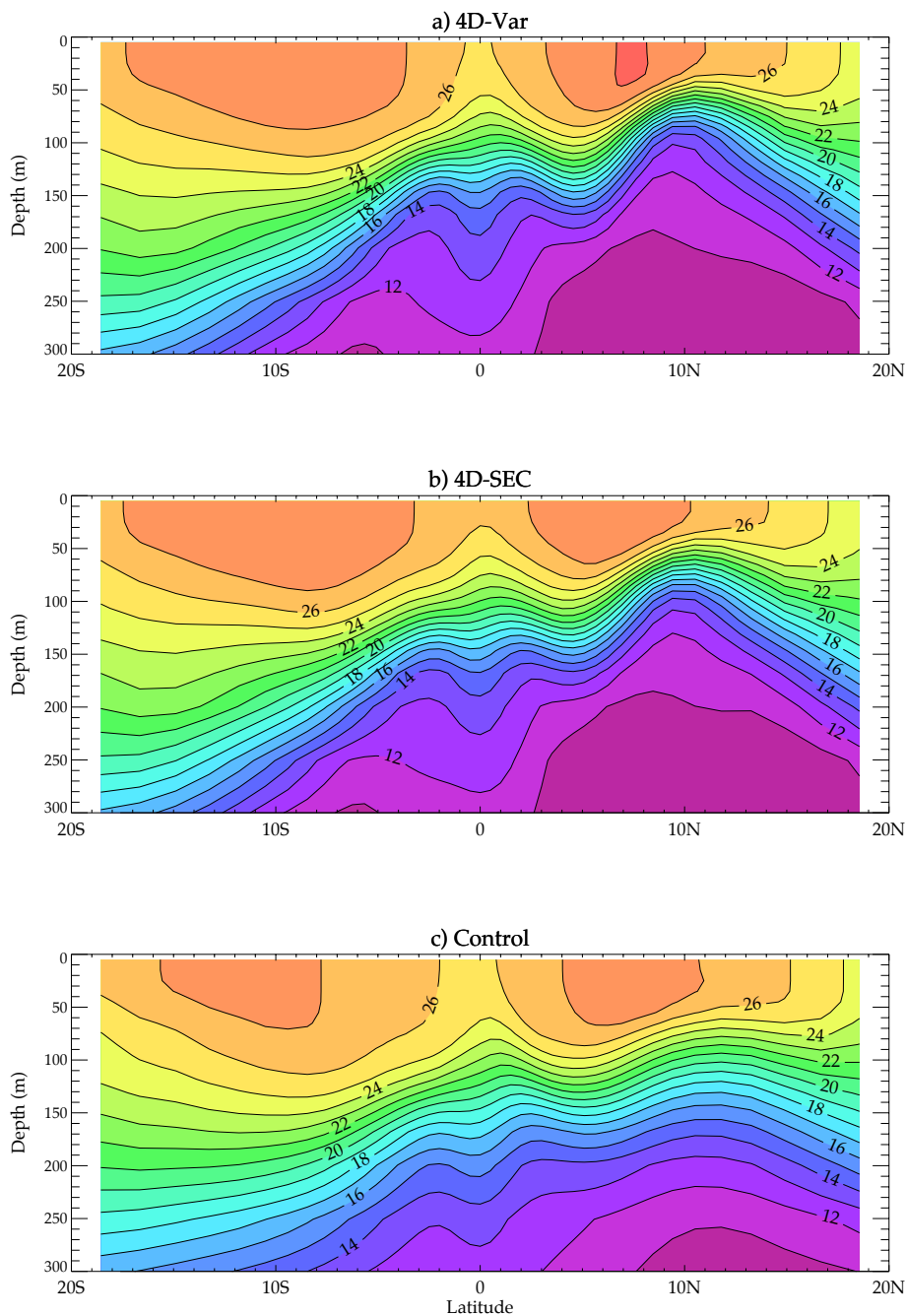


FIG. 10.2: Coupe à l'équateur de la moyenne du champ de température pour le 4D-Var (a) et le 4D-SEC (b)

10.2.2 Comparaison aux données indépendantes

- La hauteur du niveau de la mer

On a vu que sur certaines zones, l'assimilation 4D-Var permettait d'améliorer grandement les résultats sur la hauteur d'eau, ceci n'est malheureusement pas vrai sur tout le domaine. Sur les zones situées au loin de l'équateur le bilan est plus contrasté. Si on prend pour exemple une petite zone du nord du domaine (figure 10.3) l'assimilation 4D-Var dégrade les résultats du modèle libre pourtant déjà de piètre qualité.

L'introduction du contrôle de l'erreur systématique a permis d'améliorer ces résultats dans une proportion non négligeable (la rms est passée de 4 à 3) mais cela nécessite d'être encore amélioré.

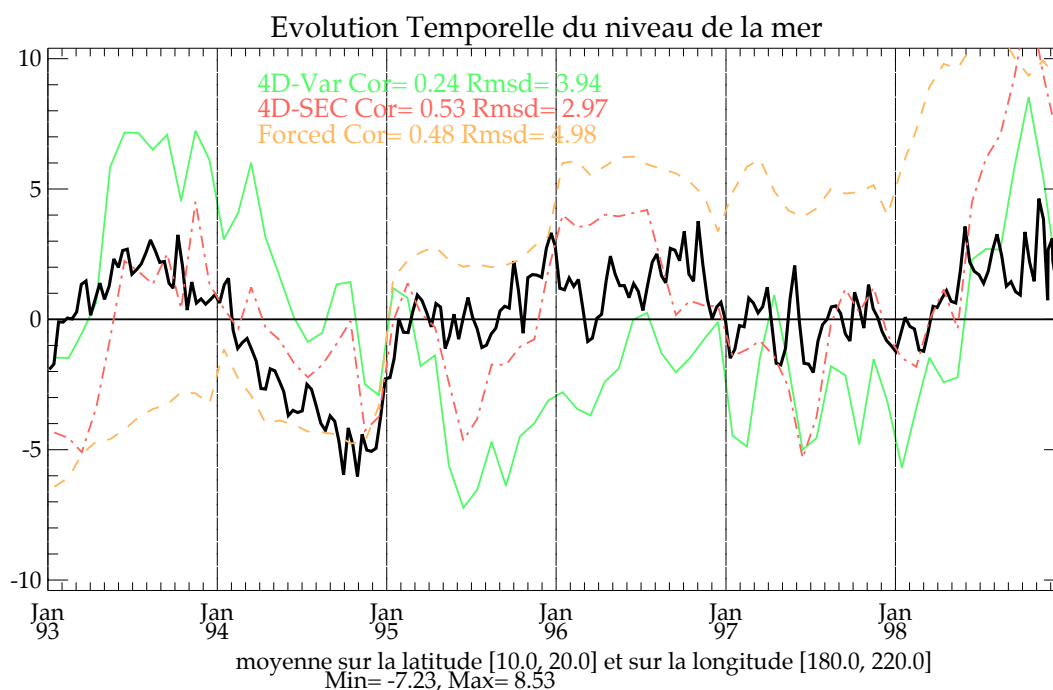


FIG. 10.3: Comparaison aux données altimétriques TOPEX/POSEIDON (trait noir épais) des hauteurs d'eau issues du 4D-Var (trait rouge continu), du 4D-SEC (vert mixte) et du modèle libre (orange pointillé)

- Les courants en profondeur

Une autre information disponible et non utilisée pour l'assimilation provient des courantomètres embarqués sur les bouées TAO, elle peut donc être utilisée pour valider les résultats des expériences d'assimilation (figure 10.4). Là encore les résultats issus de l'assimilation 4D-Var ne sont pas pleinement satisfaisants : sans être catastrophiques, ils sont de moins bonne qualité que le modèle libre. Ceci est vraisemblablement du aux échelles de corrélations verticales choisies dans l'opérateur \mathbf{B} qui induisent un trop fort mélange vertical (VIALARD communication personnelle).

Ne plus contrôler uniquement la condition initiale mais aussi un terme de correction d'erreur modèle permet de corriger ce défaut bien que celui ci ne soit pas imputable aux erreurs modèle.

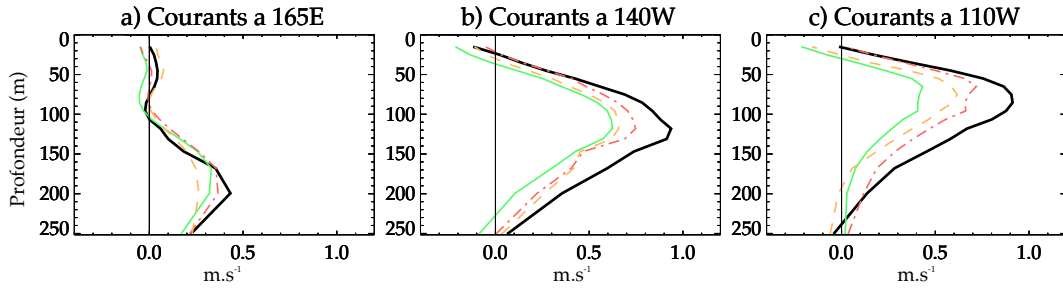


FIG. 10.4: Comparaison des courants du 4D-var (trait plein), du run libre (pointillés) et du 4D-SEC (trait mixte) avec les courants TAO (trait gras)

10.2.3 Un biais réduit

Afin de justifier l'application de la méthode de contrôle de l'erreur systématique à la configuration pacifique tropicale du modèle OPA, il convient de se poser la question de la présence de l'erreur systématique ou non dans le modèle sur cette configuration.

- Les profils de température

Outre les rumeurs persistantes sur cette présence et la forte incertitude sur les forçages qui devrait théoriquement induire une erreur systématique, on peut simplement regarder la moyenne des incréments. En théorie, si le modèle est exempt d'erreur systématique, cette moyenne doit être nulle (si on enlève le premier mois d'assimilation car l'ébauche de celui ci peut contenir un biais en fonction de la méthode utilisée pour la générer). Sur les zones niño 3 et niño 4 (figure 10.5) par exemple, on s'aperçoit que l'assimilation 4D-Var a tendance en moyenne à réchauffer les 100 premiers mètres et à refroidir les 100 suivants, ce qui est symptomatique d'un modèle qui est trop froid sur les 100 premiers mètres et trop chaud sur les 100 suivants.

L'introduction du contrôle de l'erreur systématique n'a pas permis d'éliminer complètement ce phénomène mais l'a considérablement réduit ((b) et (d) de la figure 10.5) cela a également permis, tout en conservant quasiment le même profil de température ((a) et (c)), de réduire sensiblement les écart types de ces incréments.

Grosso modo la moyenne des incréments représente la part systématique de l'erreur et l'écart type sa part aléatoire. On s'aperçoit que même si elle n'est pas négligeable, la part systématique de l'erreur est moins importante que la part aléatoire. En théorie, ceci limite l'efficacité du contrôle de l'erreur systématique mais en pratique, les périodes d'assimilation étant assez courtes, notre terme de correction de l'erreur systématique parvient à corriger aussi une partie de l'erreur aléatoire.

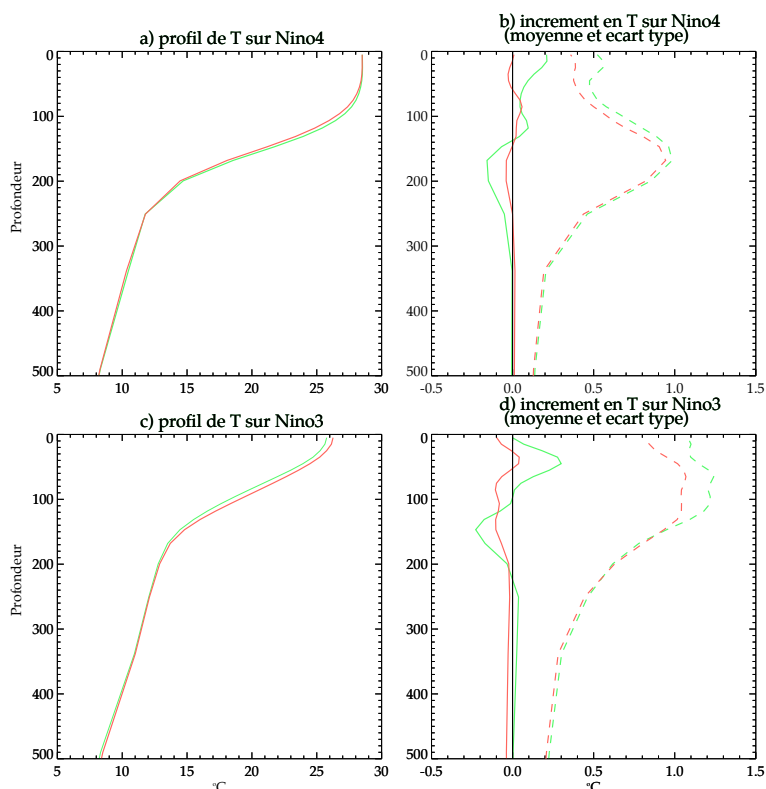


FIG. 10.5: Profils de température pour les expériences 4D-Var (trait vert) et 4D-SEC (trait rouge) et moyenne (trait continu) et écart type (pointillé) des incréments en température sur les zone niño 3 (en bas) et niño 4 (en haut) pour toute la période d'assimilation 1993–1998

- Les incréments en température

Si on s'intéresse directement aux champs de ces incréments (figures 10.6 et 10.7) on s'aperçoit qu'on a considérablement réduit l'amplitude des incréments de condition initiale, ce qui peut aussi indiquer que l'on a réduit les effets de l'erreur modèle systématique. Si la moyenne des incréments de condition initiale du 4D-SEC sont petits par rapport à ceux du 4D-Var classique, ils en conservent globalement la structure générale. Le reste de la correction est apporté par le terme de contrôle de l'erreur systématique chaque pas de temps.

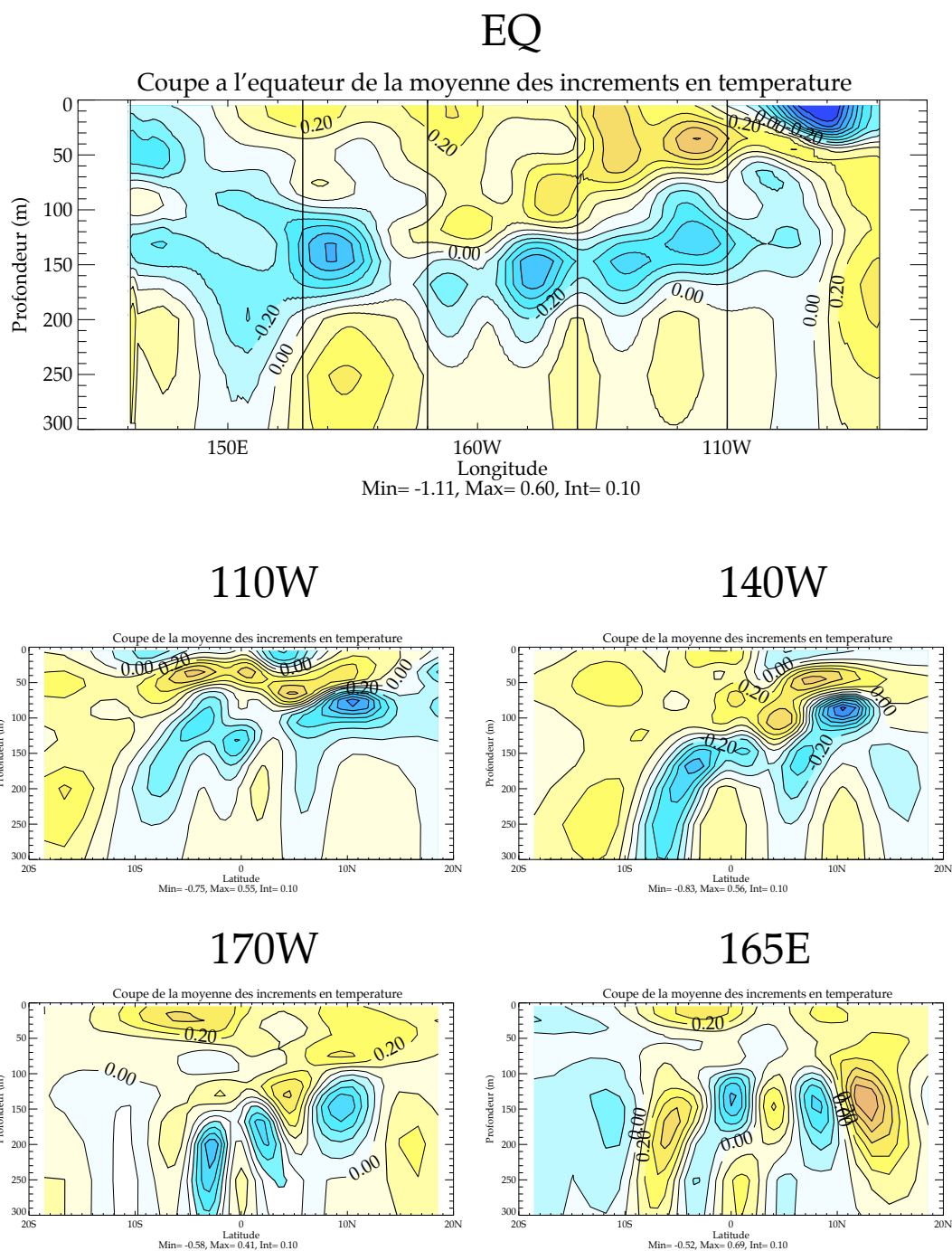


FIG. 10.6: Différentes coupes de la moyenne des increments sur T pour le 4D-Var

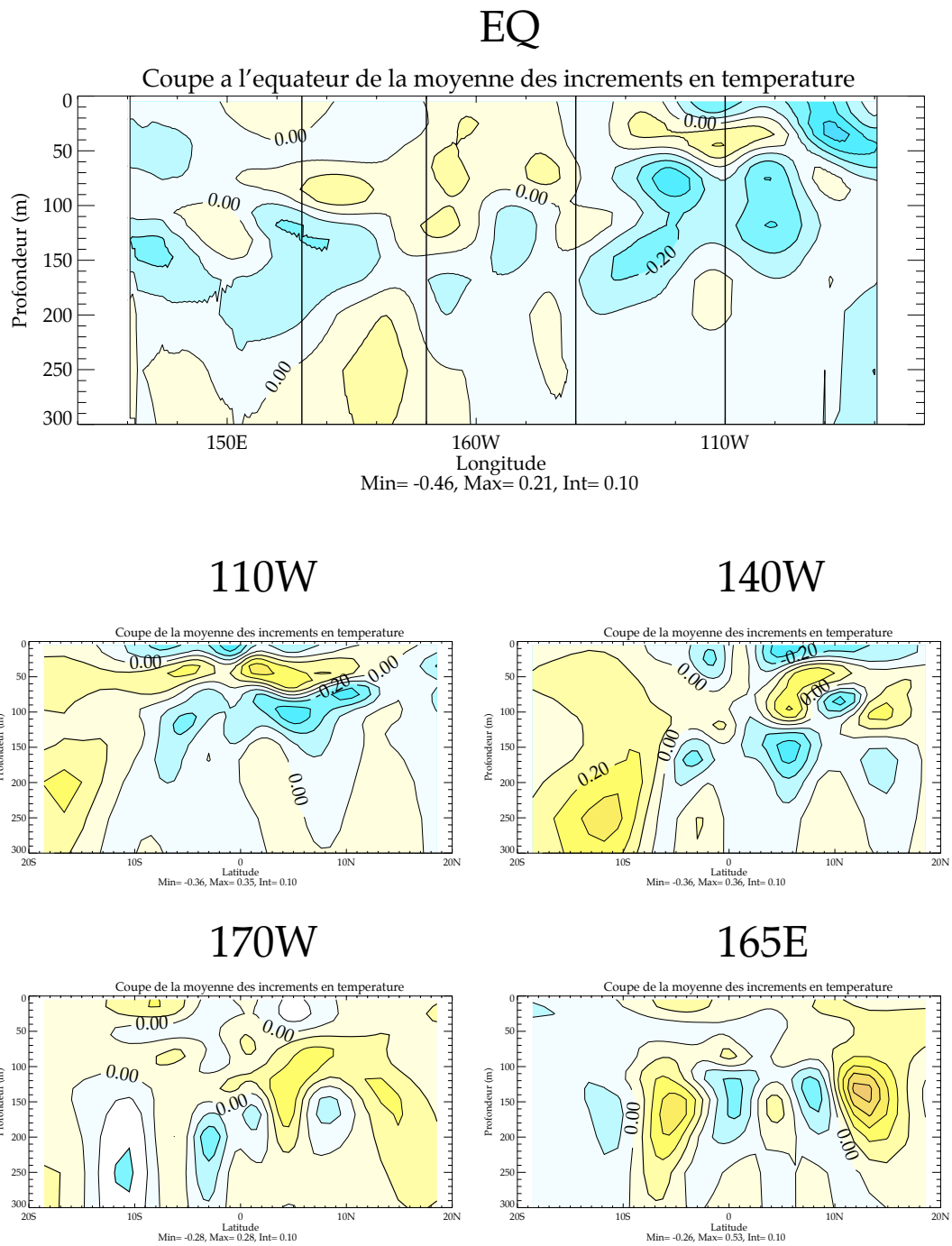


FIG. 10.7: Différentes coupes de la moyenne des increments sur T pour le 4D-SEC

10.2.4 Une assimilation “en douceur”

- Des chocs à chaque début de cycle

Le fait d'avoir des incréments de condition initiale plus importants se traduit par de brusques variations des champs corrigés. Ces sauts (figure 10.8) ne sont pas sans conséquences sur le comportement du système (voir paragraphe suivant). De plus ils ne correspondent en rien à d'éventuels phénomènes physiques.

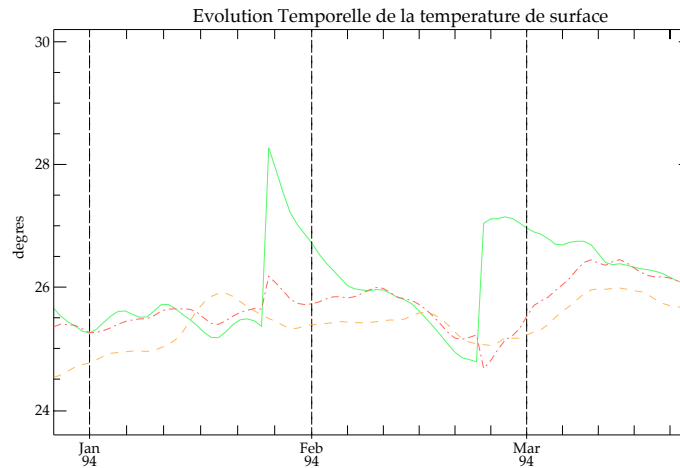


FIG. 10.8: Séries temporelles de la moyenne de la température de surface pour le modèle libre (trait pointillé orange), le 4D-Var (trait vert continu) et le 4D-SEC (rouge mixte) sur trois cycles d'assimilation de 30 jours

L'approche différente employée par le 4D-SEC réduit les incréments de condition initiale et répartit les corrections tout au long de la fenêtre d'assimilation. Sans les faire disparaître, cette approche permet de réduire de façon importante les chocs à chaque début de fenêtre d'assimilation.

- Les ondes de gravité

Comme on vient de le montrer, à chaque assimilation (une fois par mois, donc) le modèle subit un choc induit par l'ajout de l'incrément sur la condition initiale, ce qui crée des ondes de gravité mises en évidence sur la figure 10.9. On remarque aisément sur cette figure le choc en question un petit peu avant le début de chaque mois et les ondes de gravité qui en découlent et s'amenuisent au long de la fenêtre d'assimilation. Comme on l'a mis en évidence dans le chapitre 3 ceci provient du fait que l'incrément n'est pas équilibré d'un point de vue physique. Un opérateur \mathbf{B} multivarié devrait pouvoir là aussi résoudre en partie ce problème. Cela ne sera sans doute pas suffisant, l'utilisation d'un filtre digital pourra être également nécessaire.

Une fois encore, bien que ce problème ne soit pas lié directement aux erreurs modèle, l'introduction du contrôle de l'erreur systématique a permis de les réduire de façon importante (figure 10.9) par sa façon plus régulière d'apporter de l'information (une correction à chaque pas de temps). Elles restent cependant plus importantes que dans le cas du modèle libre.

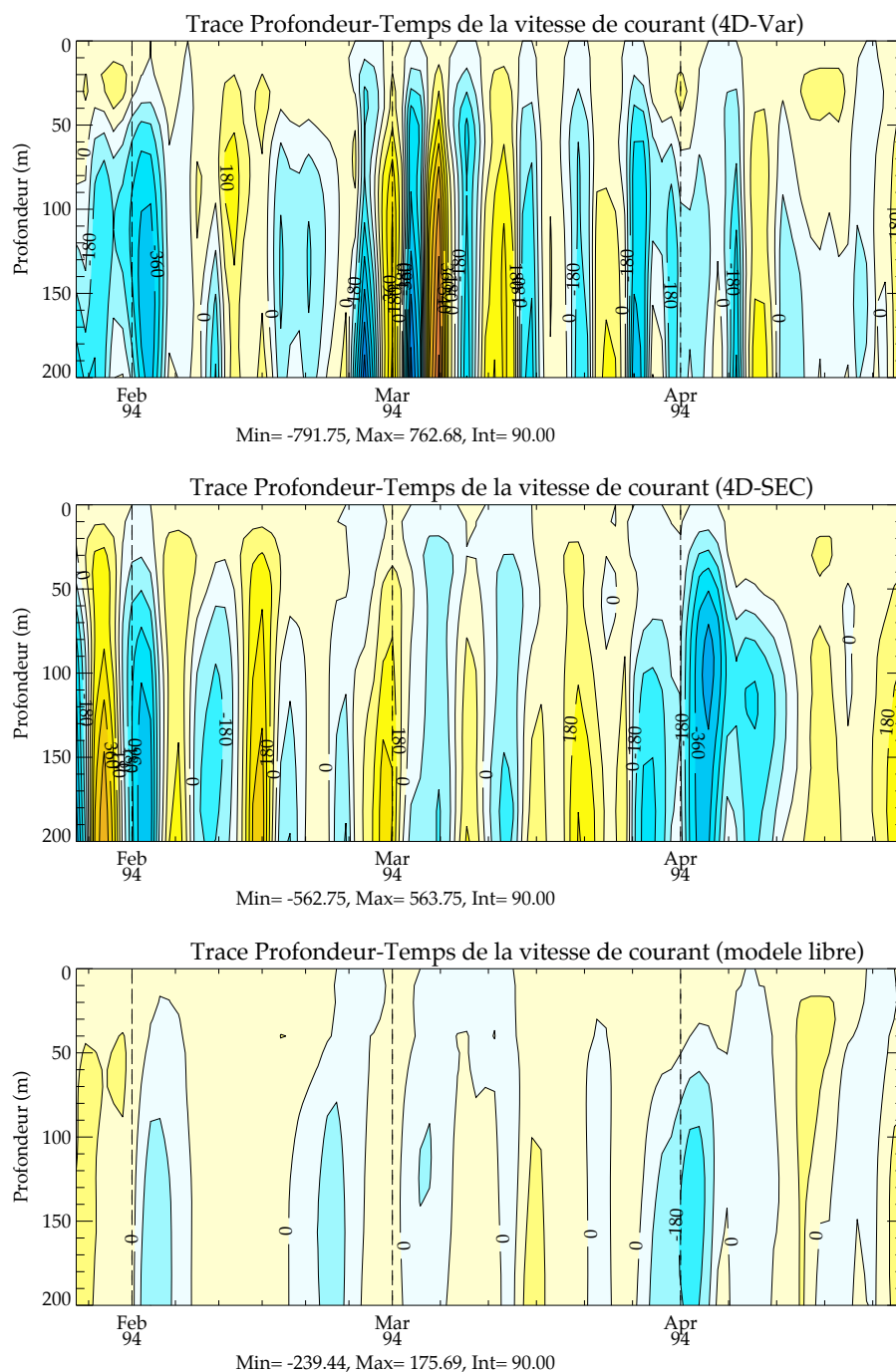


FIG. 10.9: Exemples de tracés profondeur/temps de vitesses de courant verticales pour le 4D-Var (en haut) et le 4D-SEC (en bas) sur trois cycles d'assimilation de 30 jours

10.2.5 La salinité

La salinité, c'est le sujet qui fâche le 4D-Var (ou plutôt ce 4D-Var). Comme c'était le cas pour les expériences jumelles de la section précédente, on assiste à une dérive de la salinité tout au long de l'assimilation. Si ce phénomène était facile à mettre en évidence dans le cas des expériences jumelles, dans les applications réalistes la manoeuvre est plus délicate en raison de l'extrême pauvreté en observation de la salinité. On en est donc réduit à croire les océanographes sur parole (ils se basent tout de même sur des constatations sur le terrain). Le modèle libre donne de bons résultats que le 4D-Var dégrade lors de la minimisation. Sur la figure 10.10, la moyenne des champs de salinité tout au long de l'expérience montre une surestimation globale de celle-ci de l'ordre de 0,3 p.s.u. Comme c'était le cas lors des expériences jumelles, on peut estimer que cela provient du fait qu'il est plus facile au 4D-Var de corriger l'évolution de la température, en agissant de façon importante sur la salinité, que par le seul incrément de condition initiale en température.

Le 4D-SEC ne reproduit pas ce type de dérive. Si on regarde les incréments de condition initiale et les termes de correction d'erreur en salinité de cette méthode (on ne les a pas représentés), on s'aperçoit qu'ils sont quasiment inexistantes. Le terme de correction d'erreur permet de corriger la température sans avoir à modifier la salinité de façon importante et peu cohérente avec la physique. Le manque d'information sur la salinité fait que celle-ci n'est quasiment pas modifiée.

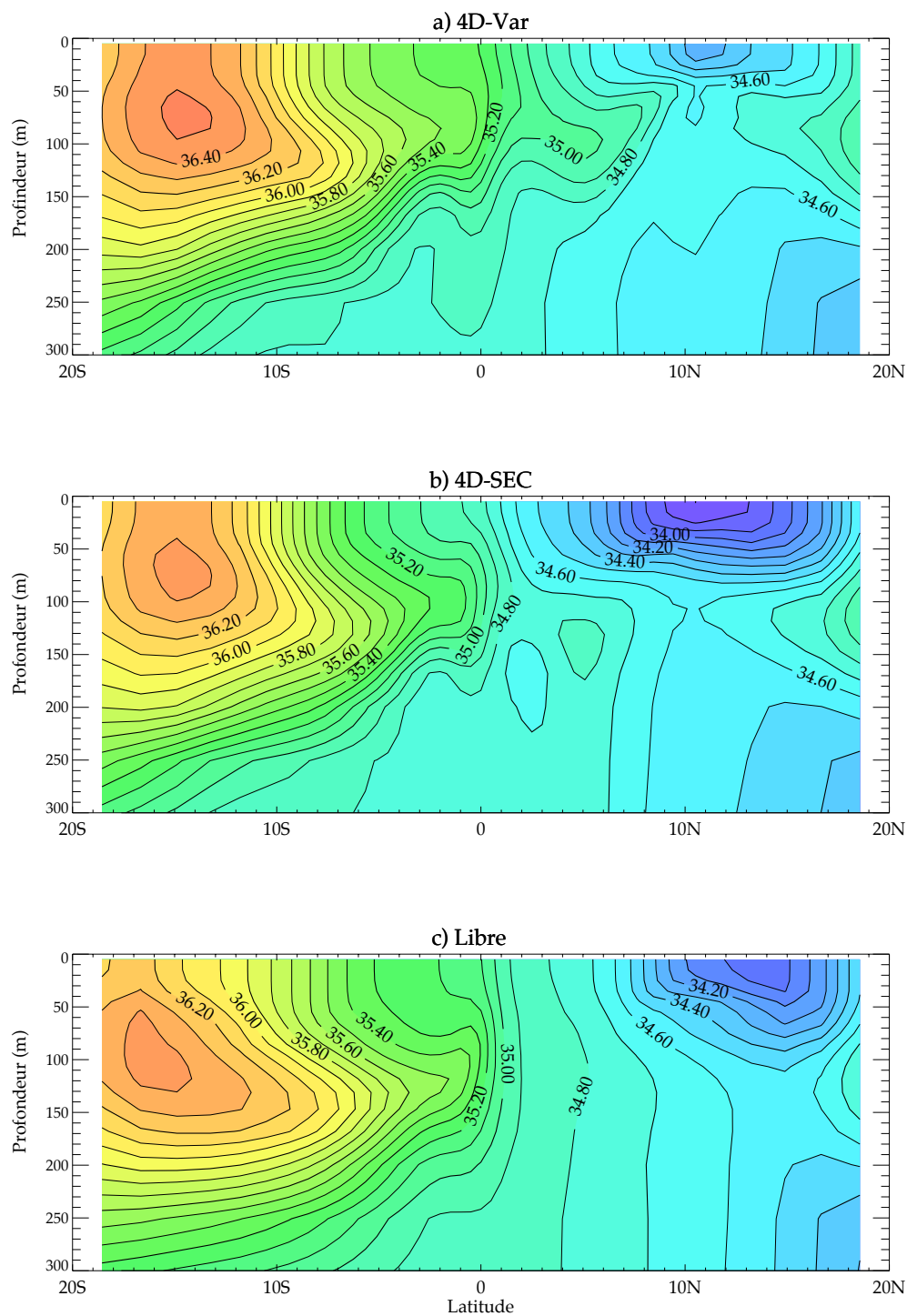


FIG. 10.10: Coupe à l'équateur de la moyenne du champ de salinité pour le 4D-Var (a), le 4D-SEC (b) et le modèle libre (c)

10.3 Pas toujours plus efficace

- Les courants de surface

On dispose de trop peu de mesures des vitesses de courants pour nous permettre de valider les résultats des diverses assimilations, cependant Reverdin et al [100] ont proposé une climatologie annuelle de ces courants, c'est à dire une moyenne annuelle des courants calculée grâce aux observations de nombreuses années ((a) de la figure 10.11). On y aperçoit le contre courant nord-équatorial (la langue jaune) et le courant sud-équatorial (les deux langues bleues). Si les résultats des assimilations 4D-Var (b) et 4D-SEC (c), et du modèle libre (d) n'ont pas de raison d'être rigoureusement identiques à cette climatologie, les moyennes sur les six années de l'expérience devraient s'en rapprocher fortement.

À la lecture de la figure 10.11, on constate que sans assimilation le modèle fournit déjà d'excellents résultats pour le courant sud-équatorial mais que le contre-courant nord-équatorial est fortement sous évalué (0,10 m/s au lieu de 0,20 m/s). Cette mauvaise estimation est à imputer à la thermocline qui est trop diffuse comme cela a été mis en évidence un peu plus haut.

Le 4D-Var a permis de retrouver un contre courant nord-équatorial convenable sans dégrader les bons résultats sur le courant sud-équatorial même si les deux phases de ce courant ne sont pas bien séparées.

Si jusqu'à présent on a principalement présenté des résultats montrant que le 4D-SEC était d'une efficacité égale ou supérieure au 4D-Var. Il réside néanmoins quelques aspects moins bien traités notamment en surface. Si on reprend la comparaison à la climatologie de Reverdin, on s'aperçoit que si l'introduction du contrôle de l'erreur systématique permet de mieux séparer le courant équatorial et le courant sud-équatorial, ceux ci sont fortement sous-évalués. Le contre courant nord-équatorial est, quant 'a lui, bien représenté.

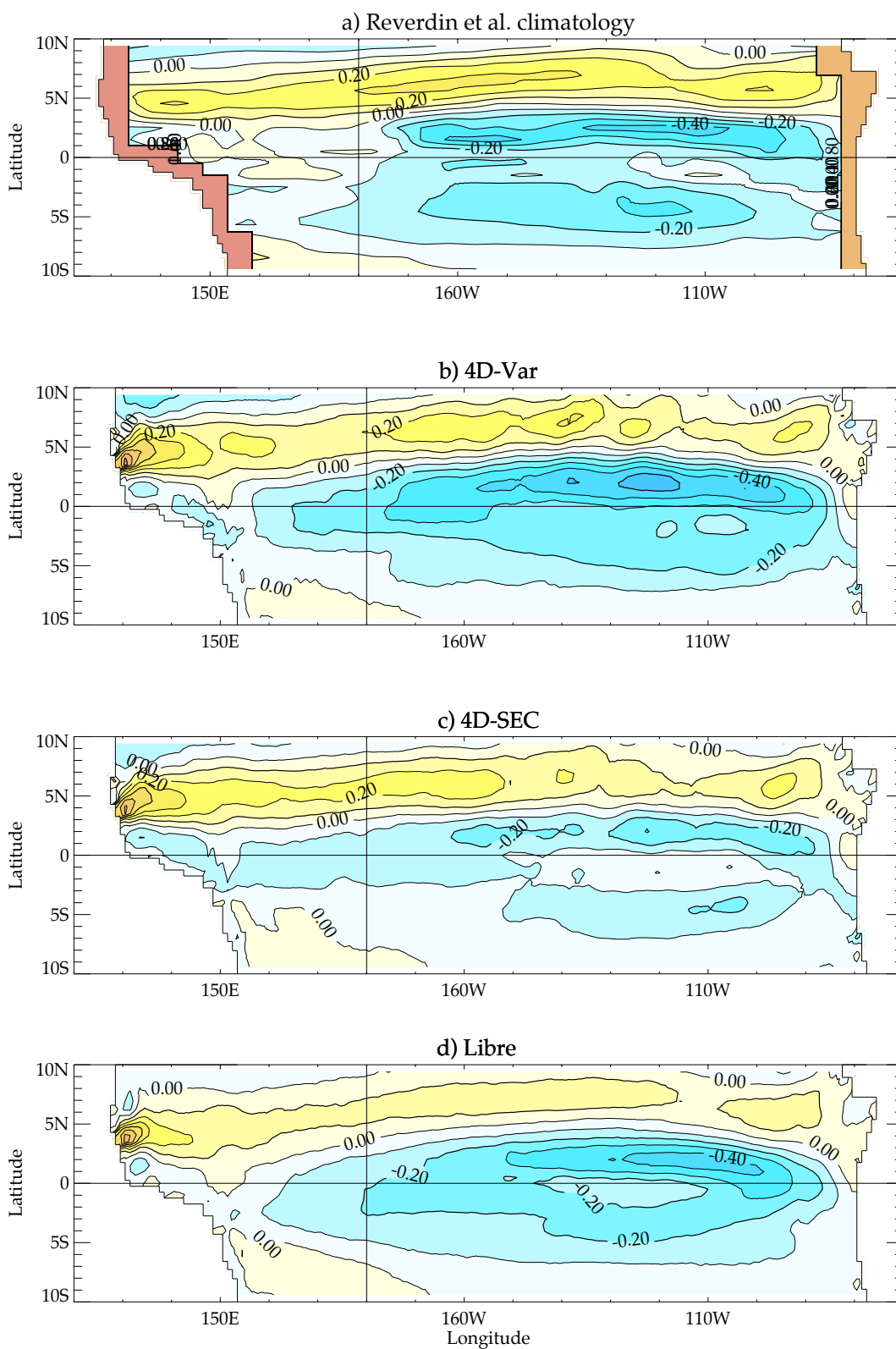


FIG. 10.11: Comparaisons aux courants de surface de Reverdin

10.4 Conclusion des expériences réalistes

Ces premières expériences avec des données réelles ne permettront pas de définitivement affirmer qu'il faut inclure le contrôle de l'erreur systématique dans les chaînes d'assimilation 4D-Var opérationnelles. En effet, les résultats exposés dans ce chapitre sont "pollués" par de nombreux autres problèmes que la présence d'erreur modèle. Avant de pouvoir vraiment mettre en évidence les apports du 4D-SEC par rapport au 4D-Var il faudra affiner ce dernier notamment en ce qui concerne la matrice \mathbf{B} de covariance d'erreur d'ébauche (que ce soit sur les variances, les échelles de corrélation et peut-être même sur l'approche utilisée pour la construire).

Cependant, outre les résultats des expériences jumelles qui ont démontré les intérêts d'une telle approche, on a pu voir dans ce chapitre quelques résultats intéressants. À part quelques problèmes en surface, les bons résultats du 4D-Var ont été, en règle générale, conservés et les points faibles du 4D-Var ont été améliorés notamment sur les courants en profondeur et la salinité.

Un des autres principaux intérêts du 4D-SEC est d'atténuer les chocs induits par l'ajout des incréments de condition initiale. Ceci a pour effet immédiat de réduire les ondes de gravité induites. Ce phénomène, qui est la conséquence directe de la répartition de la correction sur l'ensemble de la fenêtre, n'est pas vraiment en rapport avec les erreurs modèles et répond plutôt à "l'assimilation 4D-Var en continue" proposée par DERBER [40]. D'autres expériences ont été menées en appliquant exactement les propositions de DERBER (c'est à dire en ne contrôlant plus la condition initiale, mais seulement un terme de correction à chaque pas de temps). Celles-ci n'ont pas donné de très bons résultats notamment en raison d'une difficulté chronique à converger, et cela, même en partant d'une condition initiale analysée par un 4D-Var préalable. Si cette piste a été abandonnée dans cette thèse, elle mérite sans doute une étude plus approfondie.

Afin de rendre plus efficace le 4D-SEC, et en raison du relativement petit nombre d'observation, il sera sans doute pertinent, à l'instar de ce que l'on avait fait pour le modèle Shallow-Water dans la variante "grande échelles", d'introduire une forme *a priori* de l'erreur systématique recherchée. Le choix de cette structure nécessitera une étude plus poussée par des vrais océanographes.

10.5 Assimiler des données altimétriques

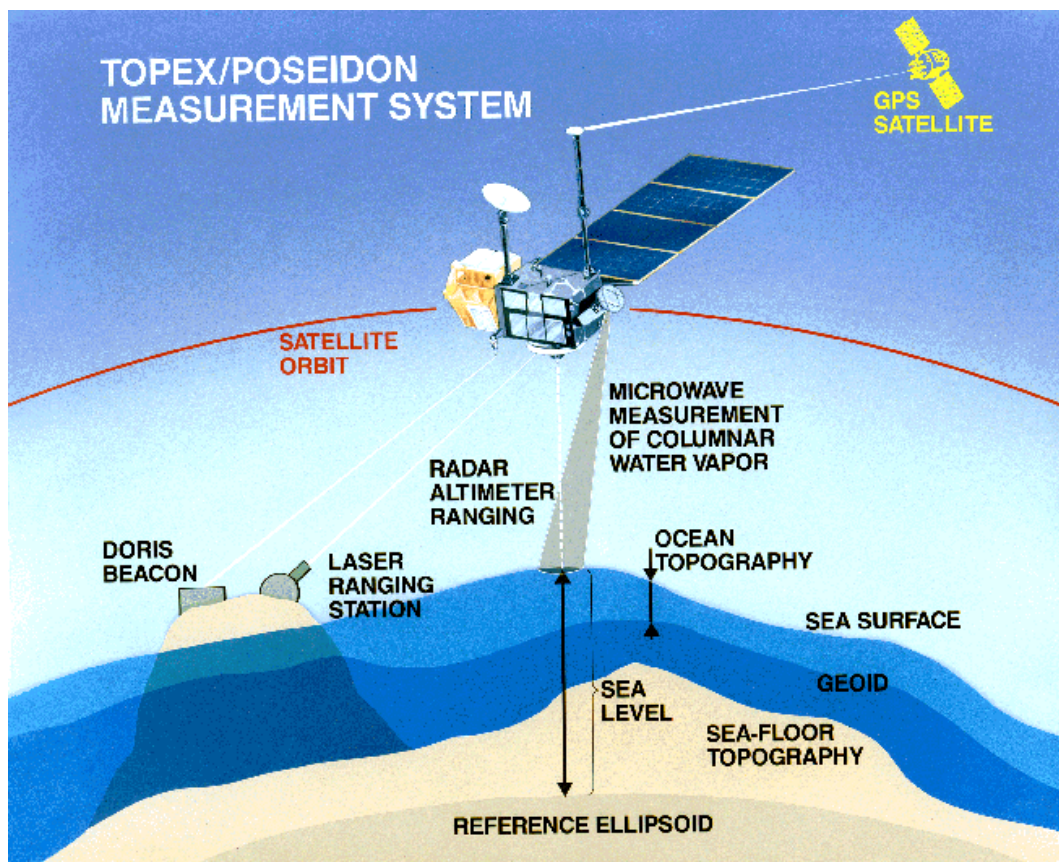
La présence de biais dans les modèles a de nombreuses fois été mise en évidence (overshooting du gulf stream, pincement de la thermocline dans le pacifique tropical, ...) et les sources en sont nombreuses (les forçages, une mauvaise paramétrisation, etc ...). En ce qui concerne les observations, et plus précisément celles qui nous intéressent ici, les mesures altimétriques, si l'on considère que l'orbite du satellite, sa dérive, la calibration de ses instruments, etc... sont bien connues (de nombreux travaux sont effectués dans ce sens, voir par exemple ceux cités dans [6]), la difficulté principale réside dans la connaissance

approximative qu'on a du géoïde¹, bien que de nombreux efforts sont faits dans cette direction notamment avec les missions CHAMP puis bientôt GRACE (2001) et GOCE (2005); difficulté à laquelle peuvent se rajouter les problèmes dus à l'interpolation sur les points de grille. En effet, la donnée océanographique intéressante est la variation par rapport à l'état idéalisé de l'océan au repos que représente le géoïde. En pratique, c'est la variabilité par rapport à la surface moyenne océanique observée qui est utilisée comme observation et elle est comparée à la variabilité du modèle par rapport à la surface moyenne "simulée" (issue du modèle). La différence entre ces deux surfaces moyennes océaniques représente de façon évidente un biais dans les observations. Une meilleure estimation de la surface moyenne océanique est donc un enjeu pour la bonne utilisation des données altimétriques [67] [87].

Une variante de la méthode 4D-SEC pourra être appliquée dans le but de corriger le biais contenu dans les observations. L'altimétrie satellitale est potentiellement capable de déterminer la dénivellation de la surface de la mer par rapport au géoïde avec une précision suffisante pour fournir une estimation des courants géostrophiques absolus à la surface de l'océan. Malgré un effort certain pour améliorer notre connaissance du géoïde, il reste imprécis, en particulier l'erreur du géoïde s'accroît fortement avec la résolution. Cette erreur peut raisonnablement être considérée comme étant un biais sur les observation et les échelles temporelles de l'évolution de cette erreur étant extrêmement importantes comparées à celles qui nous intéressent, ce biais pourra être considéré comme étant constant, ne nécessitant donc pas de "modèle de biais". La principale difficulté résidera sans doute dans l'équilibre adéquat à trouver dans la fonction coût entre les termes concernant le modèle (ébauche et éventuellement erreur systématique) et ceux concernant les observations.

Carton et al. [22] ont démontré que dans le pacifique tropical l'assimilation de données altimétriques a un plus grand impact sur le niveau de la mer et la variabilité des courants que l'assimilation des données issues de bouées dérivantes et des réseaux de bouées fixes (TAO). Les données utilisées pourront être les données altimétriques fournies par le satellite TOPEX/POSEIDON combinées avec le réseau d'observations in-situ TAO, ces dernières apportant une information non biaisée au système. Ceci permettra, si on assimile les données altimétriques référencées à la surface moyenne modèle en même temps que les données in-situ, de caractériser l'erreur systématique "pure" du modèle. D'autre part si on réalise la même expérience mais en utilisant les données altimétriques référencées par rapport au géoïde cette fois, on obtiendra de l'information sur la vraie surface moyenne en plus de l'erreur modèle.

¹surface d'équipotentiel de l'océan, c'est à dire surface de l'océan si celui ci était au repos (pas de circulation océanique)



Conclusions et perspectives

Conclusions...

Que retenir de ces trois années de travail...

L'utilisation de modèles simples comme l'équation de Burger et le Shallow-Water a permis de bien mettre en évidence les comportements du 4D-Var en présence d'erreur modèle. Leur facilité d'utilisation a également permis d'introduire les méthodes de contrôle de l'erreur modèle et de préparer le terrain pour les applications à des problèmes plus réalistes de la dernière partie.

Grâce à cette approche à donc pu être mis en place la méthode 4D-Var avec contrôle de la part systématique de l'erreur modèle (le 4D-SEC). Sur ces petites configurations le 4D-SEC a démontré sa supériorité sur le 4D-Var et ce, sur l'ensemble des aspects de l'assimilation. Le niveau d'erreur a été réduit sur l'ensemble des variables, y compris sur celles qui ne sont pas observées, et lorsque le nombre d'observations est suffisant le terme de correction d'erreur estimé est suffisamment pertinent pour permettre de corriger également le modèle durant la période de prévision. Si les observations ne sont pas suffisamment nombreuses, on a montré que l'on pouvait à nouveau obtenir ce type de résultat, en restreignant la recherche à certaines structures de l'erreur connues *a priori*. Le passage à une configuration réaliste a montré que si tout n'était plus aussi évident, les premiers résultats étaient encourageants. En règle générale le 4D-SEC a amélioré ou conservé les résultats du 4D-Var, seuls quelques points noirs restant à améliorer.

Une autre méthode a été testée dans cette thèse : le Nudging Optimal. Elle n'a été appliquée qu'aux modèles jouets (équation de Burger et modèle Shallow-Water) et a fait montre, sur ces configurations, d'une efficacité accrue comparée au 4D-Var. Il n'a pas été montré de résultat du Nudging Optimal sur les configurations réalistes, cela veut-il dire que cette méthode est à jeter aux oubliettes ? La réponse est évidemment non, seulement il a fallu faire des choix dans cette thèse, et pour des raisons de temps et surtout de ressources en calcul le nombre d'expérience de Nudging Optimal appliquée à la fenêtre pacifique tropicale du modèle OPA était insuffisant pour pouvoir en tirer quelques conclusions. Cette méthode étant plus délicate à calibrer que le 4D-SEC, on a préféré développer plus avant ce dernier.

De l'ensemble des expériences d'assimilation que l'on a effectuées il ressort l'extrême

importance du calibrage de la fonction coût, que ce soit pour un aspect conditionnement de la minimisation ou un aspect cohérence physique de la correction (de condition initiale ou d'erreur modèle). La définition des variances et des corrélations revêt donc une importance capitale pour l'efficacité des méthodes d'assimilation de données.

On ne peut déceimment clore ces conclusions sans dire un mot du coupleur d'assimilation PALM. La part consacrée à ce logiciel dans ce rapport est assez réduite alors qu'il a accompagné toutes les expérimentations qui ont été effectuées tout au long de la thèse. Sa modularité, l'aide qu'il apporte à la parallélisation et sa simplicité d'utilisation ont fourni, malgré des débuts difficiles, un gain de temps certain dans la construction et la manipulations des méthodes présentées ici. Avec l'aide de PALM et une fois la première méthode (le 4D-Var) palmisée, il a été facile de créer une tribu d'autres algorithmes appliqués à OPA (4D-SEC, 4D-ON) ou au modèle Shallow-Water (3D-Var, 3D-FGAT, 3D-PSAS, 4D-Var incrémentale, 4D-SEC, 4D-ON).

... et perspectives

Ce qu'on peut surtout retenir c'est qu'il reste encore beaucoup de travail, et ce dans beaucoup de directions possibles. Sans prétendre être exhaustif on peut discerner quelques directions principales.

Pour améliorer les performances du 4D-SEC et le rendre opérationnel sur des applications réalistes, il conviendra de définir plus précisément la structure *a priori* du terme de correction recherché à l'instar de ce qui a été présenté sur le modèle Shallow-Water. Cette étude devra être effectuée par de vrais océanographes, ayant une connaissance plus approfondie des phénomènes responsables de ces erreurs.

Une expérience qui n'a pas été exposée ici à montré que pour des fenêtres d'assimilation plus longues la minimisation du 4D-Var incrémental ne parvient pas à converger car l'hypothèse linéaire tangente n'est plus respectée. L'approche 4D-SEC, grâce à ces incréments plus petits, permet d'augmenter la taille des fenêtres d'assimilation. Sans chercher à établir un record, il a été possible d'effectuer des assimilation de données sur des cycles d'assimilation de 3 mois (contre 1 mois pour le 4D-Var) sans problème notable pour la minimisation. Afin de pouvoir faire les comparaisons avec le 4D-Var sans trop de différences algorithmiques ces expériences longues n'ont pas été exploitées. Il serait intéressant d'explorer plus avant cette piste notamment dans un cadre non-incrémental et d'en approfondir les justifications théoriques.

La combinaison de données *in situ* avec des données altimétriques pourra également être une direction de recherche décisive.

Parallèlement à ces développements du 4D-SEC, il sera certainement intéressant d'appliquer la méthode de Nudging Optimal a une configuration réaliste et avec des données réelles. Le réseau TAO ne sera peut-être pas suffisamment fourni pour cette méthode, il sera sans doute nécessaire d'assimiler également des données satellites.

D'une manière plus générale la chaîne d'assimilation 4D-Var sur OPA fenêtre pacifique tropical dont on a présenté les résultats dans la dernière partie est loin d'avoir atteint sa version définitive, sa marge de progression reste donc encore importante. Un important travail doit notamment être fait concernant la matrice de covariance d'erreur d'ébauche \mathbf{B} (opérateur d'équilibre, structures de corrélation plus "physiques", explorer d'autres méthodes de construction, ...), ainsi que l'utilisation de filtres digitaux.

Pour clore les aspects pratiques de ces perspectives, il faudra évidemment étendre ces expériences à d'autres régions de l'océan. L'océan pacifique tropical a un comportement assez linéaire, ce qui n'est pas forcément toujours le cas; qu'en sera-t-il alors du comportement de nos méthodes ?

D'un point de vue plus théorique de nombreuses directions peuvent encore être explorées. Comme on l'a évoqué à plusieurs reprises dans ce rapport, il serait nécessaire de réfléchir un peu à l'adéquation modèle/réseau d'observation. Ce sujet n'a, sauf erreur, pas, ou très peu, intéressé les spécialistes. Pourtant une course à la résolution pour les modèles et au nombre d'observation, de façon indépendante l'une de l'autre, n'a pas énormément de sens. On comprend bien que pour des raisons pratiques on ne va pas maintenir toute une série de modèles à des résolutions différentes pour s'accorder au réseau d'observation, mais cet aspect mérite tout de même d'être développé.

Tout au long de cette thèse on s'est appliqué à corriger les erreurs modèles "en général". En raison de la nature multiple de ces erreurs (erreur numériques, de discrétisation, physique manquante, forçages imprécis, etc.), pour être efficace, les méthodes devraient prendre en compte les différents types d'erreur de façon séparées et appropriées.

On a évoqué un peu plus haut, le travail qu'il restait à faire concernant l'opérateur de covariance d'erreur d'ébauche \mathbf{B} . Une des voies pour y parvenir pourrait être de mettre en places des méthodes hybrides SEEK/4D-Var comme le suggèrent VERSEE *et al* [119]. Le SEEK servirait alors à construire les matrices de covariance utilisées ensuite dans les fonctions objectives du 4D-Var ou de ses variantes "modèle imparfait".

Annexes

Annexe A

Résultats sur le 3D-Var et le 4D-Var incrémental.

Dans cette annexe sont regroupés des résultats de différentes méthodes d'assimilation de données : un 4D-Var, un 4D-Var incrémental et un 3D-Var. Ces différentes démarches sont appliquées à la même configuration du modèle Shallow-Water dans le cadre d'expériences jumelles avec des données simulées sous la forme de traces satellites. Les objectifs de cette annexe sont d'une part d'illustrer les différences de comportement entre une méthode séquentielle (le 3D-Var) et des méthodes globales comme le 4D-Var et ses variantes et d'autre part de mettre en évidence les caractéristiques de fonctionnement du 4D-Var incrémental tout au long de la minimisation évoquées dans les expériences sur OPA. Ces graphiques sont issus des expériences effectuées par Valérie AUFRAY que j'ai eu le plaisir de co-encadrer lors de son stage de 4^{ème} année à l'INSA Toulouse (AUFRAY, 2000 [3]).

Les courbes des figures A.1, A.3, A.2 représentent l'évolution temporelle de la norme de l'erreur globale (en haut à gauche), sur la hauteur d'eau (en haut à droite), sur la composante zonale de la vitesse (en bas à gauche) et sur sa composante méridionale (en bas à droite). Sur chacune de ces figures la courbe rouge représente la prévision du modèle et la courbe verte l'analyse résultante de l'assimilation de données. Pour le 4D-Var incrémental les deux courbes bleues représentent l'évolution de l'erreur après les deux premières itérations de la boucle externe (les réactualisation de la trajectoire de référence). L'échelle des ordonnées reste la même sur les trois figures alors que l'échelle des abscisses diffère quelque peu. Pour l'expérience 4D-Var trois cycles d'un mois d'assimilation/prévision sont tracés (1 mois = 1440 pas de temps); pour la variante incrémentale, seul 1 mois d'assimilation a été effectué ainsi que la prévision correspondante (1 mois d'assimilation + 1 mois de prévision), et pour l'expérience 3D-Var les résultats de 10 cycles d'assimilation de trois jours sont représentés.

Si on compare les résultats des 1440 premiers pas de temps des assimilations 4D-Var et 4D-Var incrémental à ceux du 3D-Var, on voit bien apparaître les problèmes de discontinuité qu'implique l'approche séquentielle. De plus le niveau global de l'erreur est sensiblement plus haut pour le 3D-Var que pour les deux variantes du 4D-Var. Ceci est principalement dû au fait que l'information prise en compte ici est seulement passée (in-

directement) et présente (directement) contrairement au 4D-Var pour lequel l'information est prise en compte sur la globalité de la période d'assimilation simultanément. Le principal avantage de l'approche 3D-Var est son coût en temps de calcul, en effet on économise ainsi les intégrations de M et \mathbf{M} durant la minimisation.

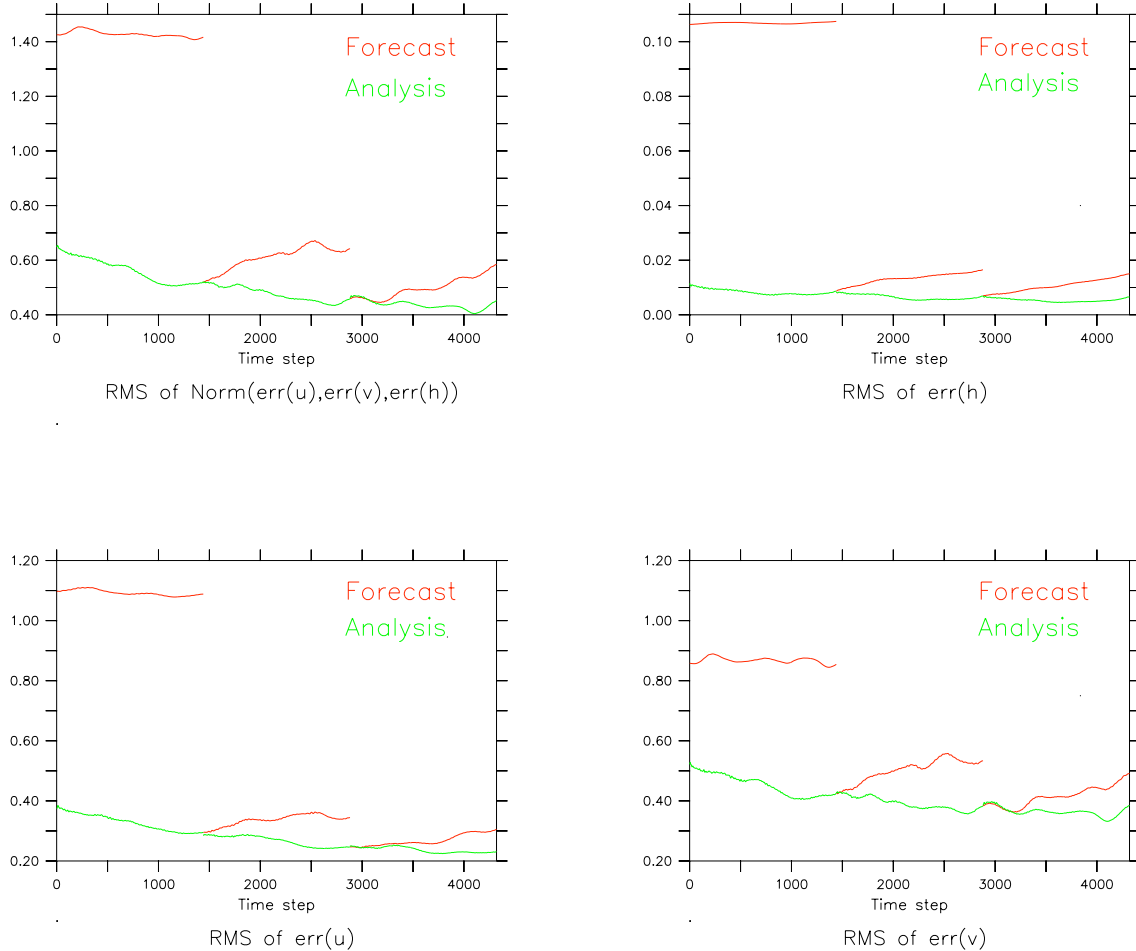


FIG. A.1: norme rms de l'erreur sur le vecteur d'état en fonction du temps après un 4D-Var

Les résultats de la dernière boucle externe du 4D-Var incrémental, bien que très proche de ceux du 4D-Var, restent globalement de moins bonne qualité. Ceci est dû au fait que l'hypothèse linéaire tangente n'est pas rigoureusement exacte. Cependant la fonction objective est alors quadratique ce qui permet d'avoir une minimisation plus efficace (et donc d'économiser du temps de calcul). De plus l'approche incrémentale permet facilement d'utiliser un modèle linéaire tangent simplifié (pas le cas ici) et ainsi de réaliser, au prix d'une légère dégradation de résultats, une nouvelle économie en temps de calcul.

On constate que la correction apportée sur la condition initiale lors de la première remise à jour est assez importante, la deuxième l'est un peu moins et la troisième encore moins. Les figures A.4, A.5 et A.6 représentent les incréments sur les vecteurs d'état pour chaque réactualisation de la trajectoire de référence (respectivement première, deuxième et troisième). Comme cela a été évoqué lors des expériences du 4D-Var incrémental appliqué

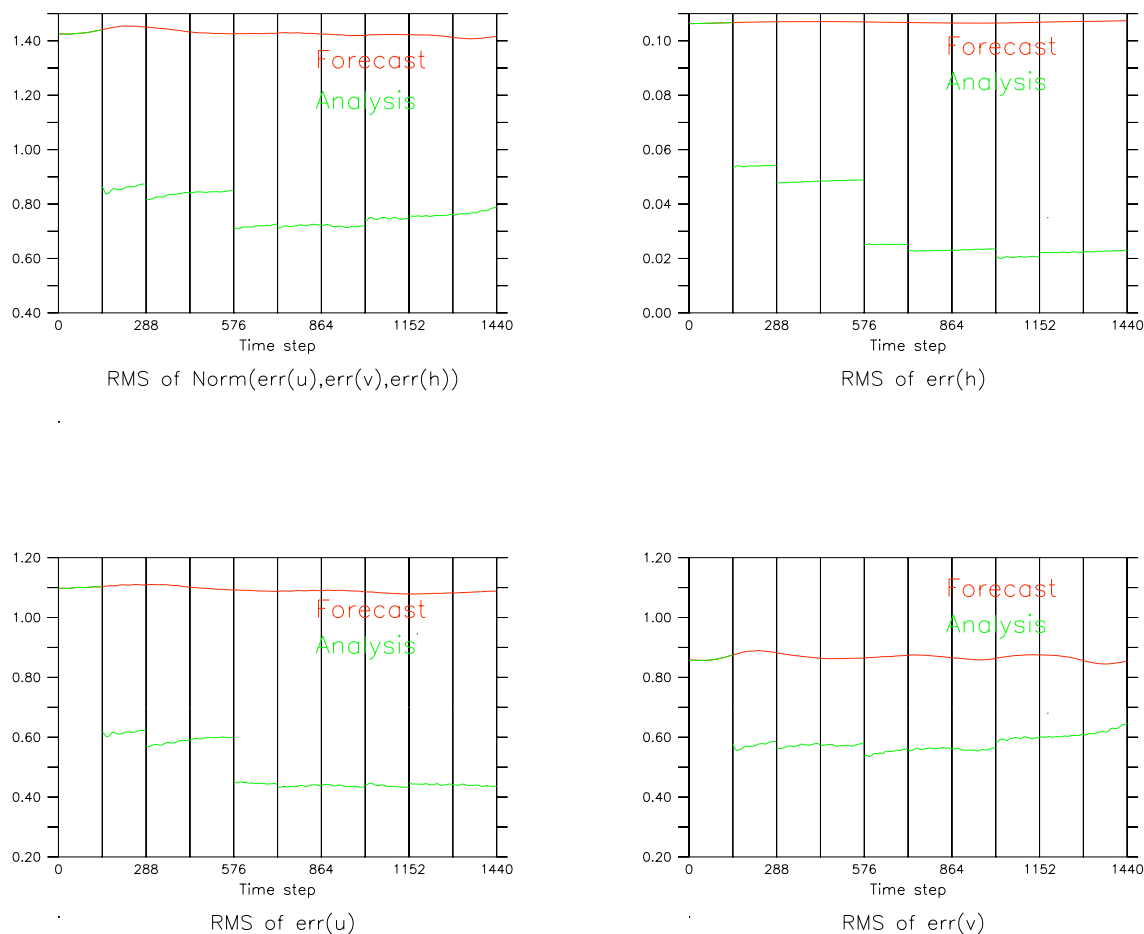


FIG. A.2: norme rms de l'erreur sur le vecteur d'état en fonction du temps après un 3D-Var

à OPA, on constate que le début de la minimisation est consacré à corriger l'ébauche suivant les grandes échelles et, plus on avance dans la minimisation, plus on corrige de petites structures.

Lors de la première boucle de minimisation, l'incrément propagé par le modèle linéarisé est une correction sur les grandes échelles. Ces dernières étant dominées par l'équilibre géostrophique qui est linéaire, l'hypothèse linéaire tangente reste valide. Cependant, ceci devient de moins en moins évident lors des réactualisations suivantes de la trajectoire car l'incrément concerne des échelles de plus en plus petites qui sont dominées par des termes non linéaires.

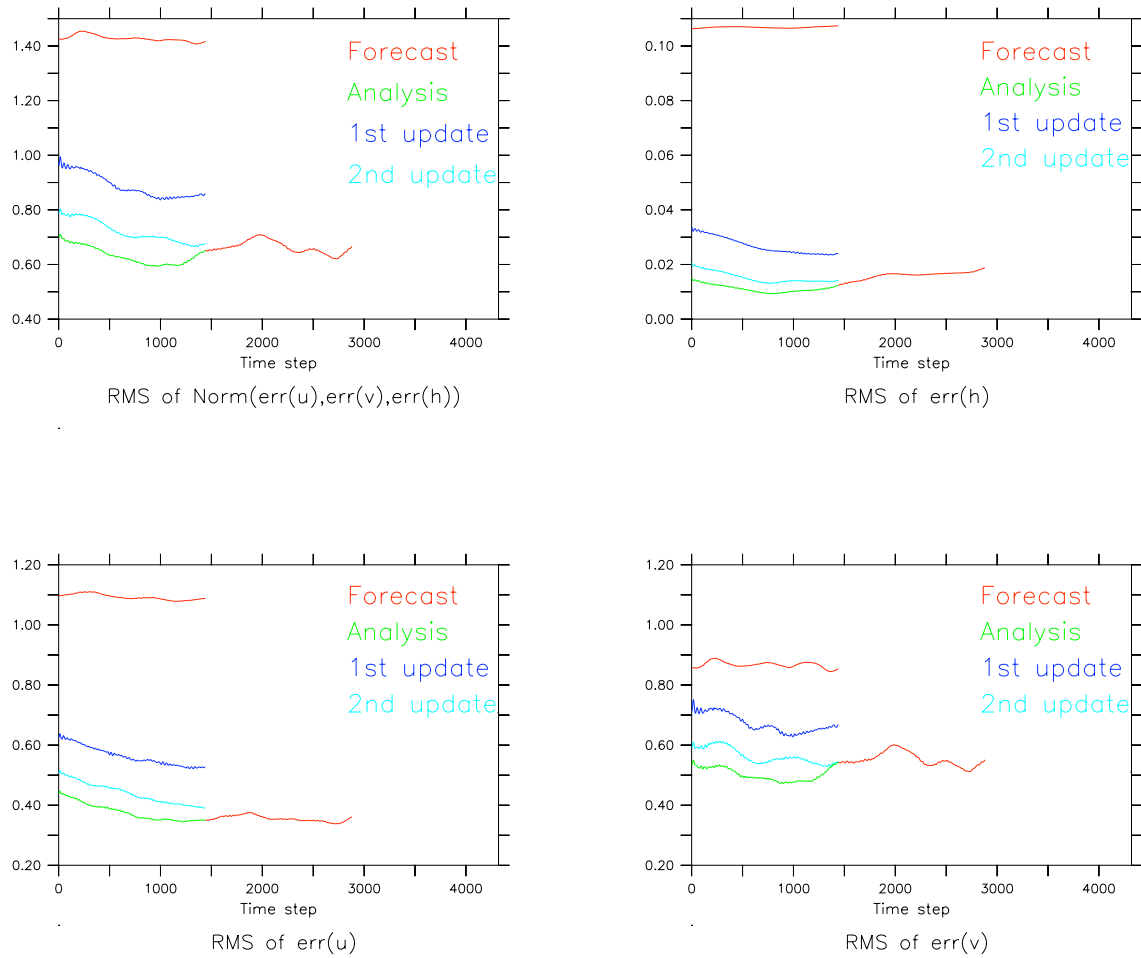


FIG. A.3: norme rms de l'erreur sur le vecteur d'état en fonction du temps après un 4D-Var incrémental

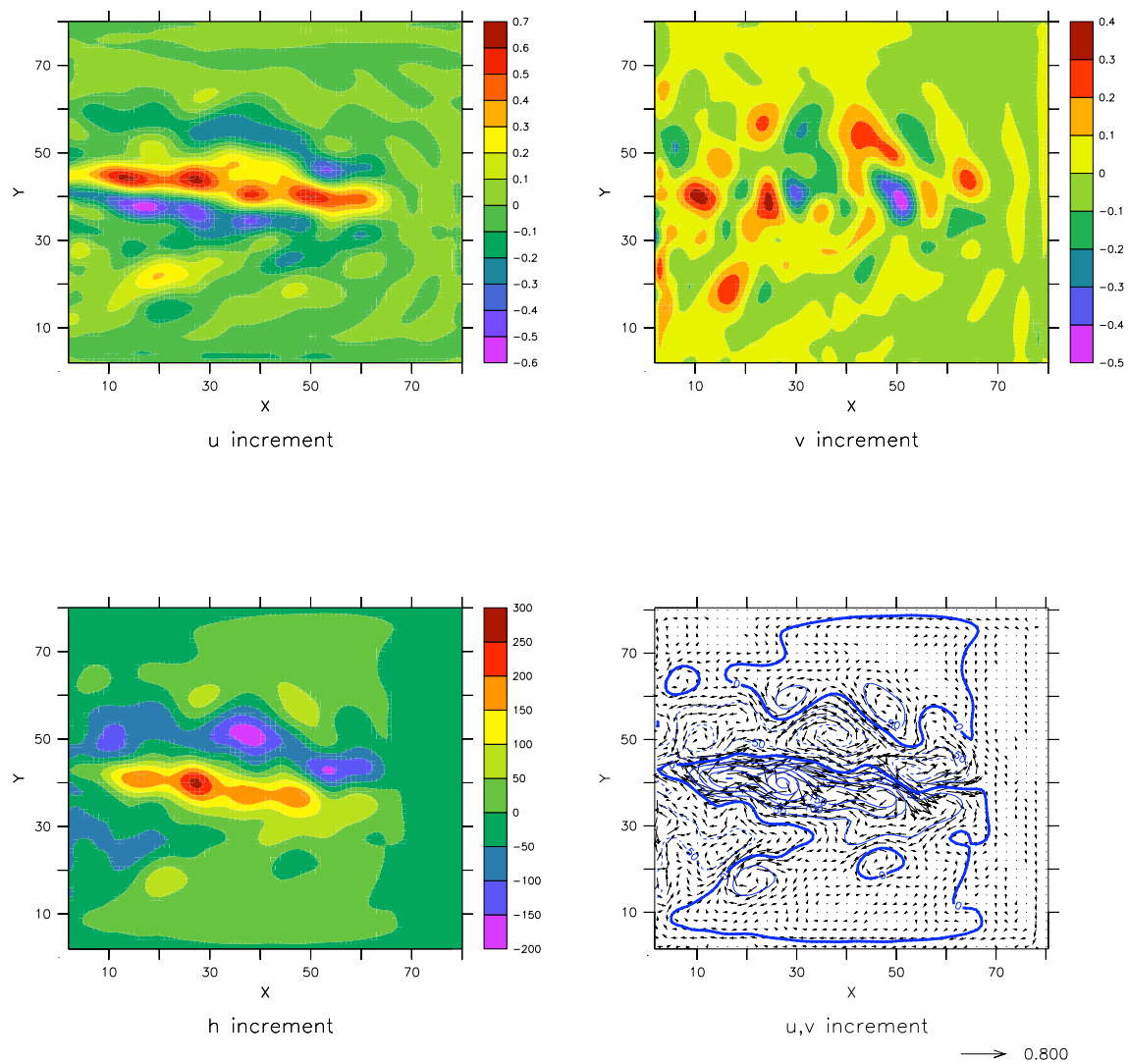


FIG. A.4: *Incrément sur les différentes composantes du vecteur d'état à l'issue de la première boucle externe*

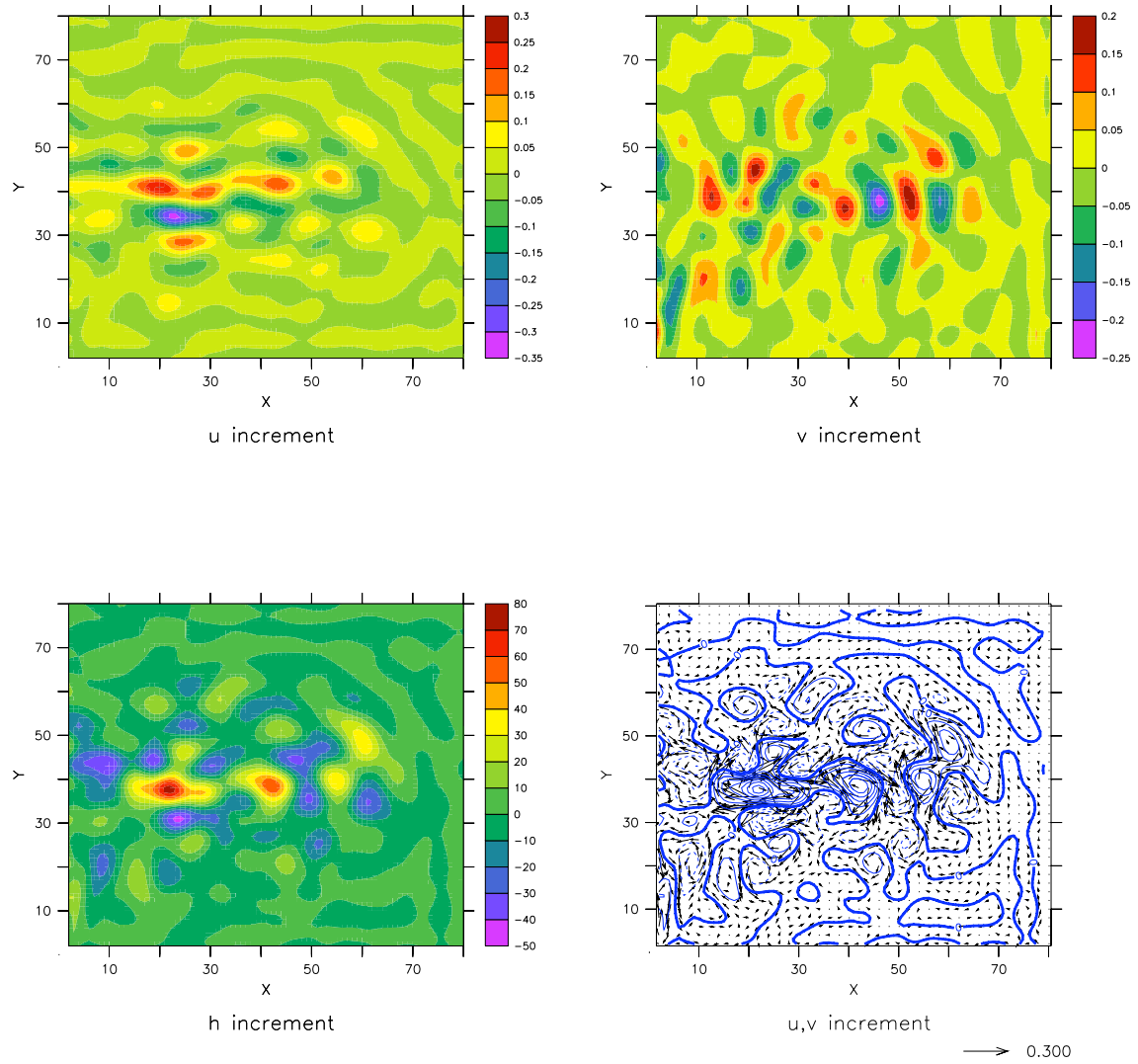


FIG. A.5: *Incrément sur les différentes composantes du vecteur d'état à l'issue de la deuxième boucle externe*

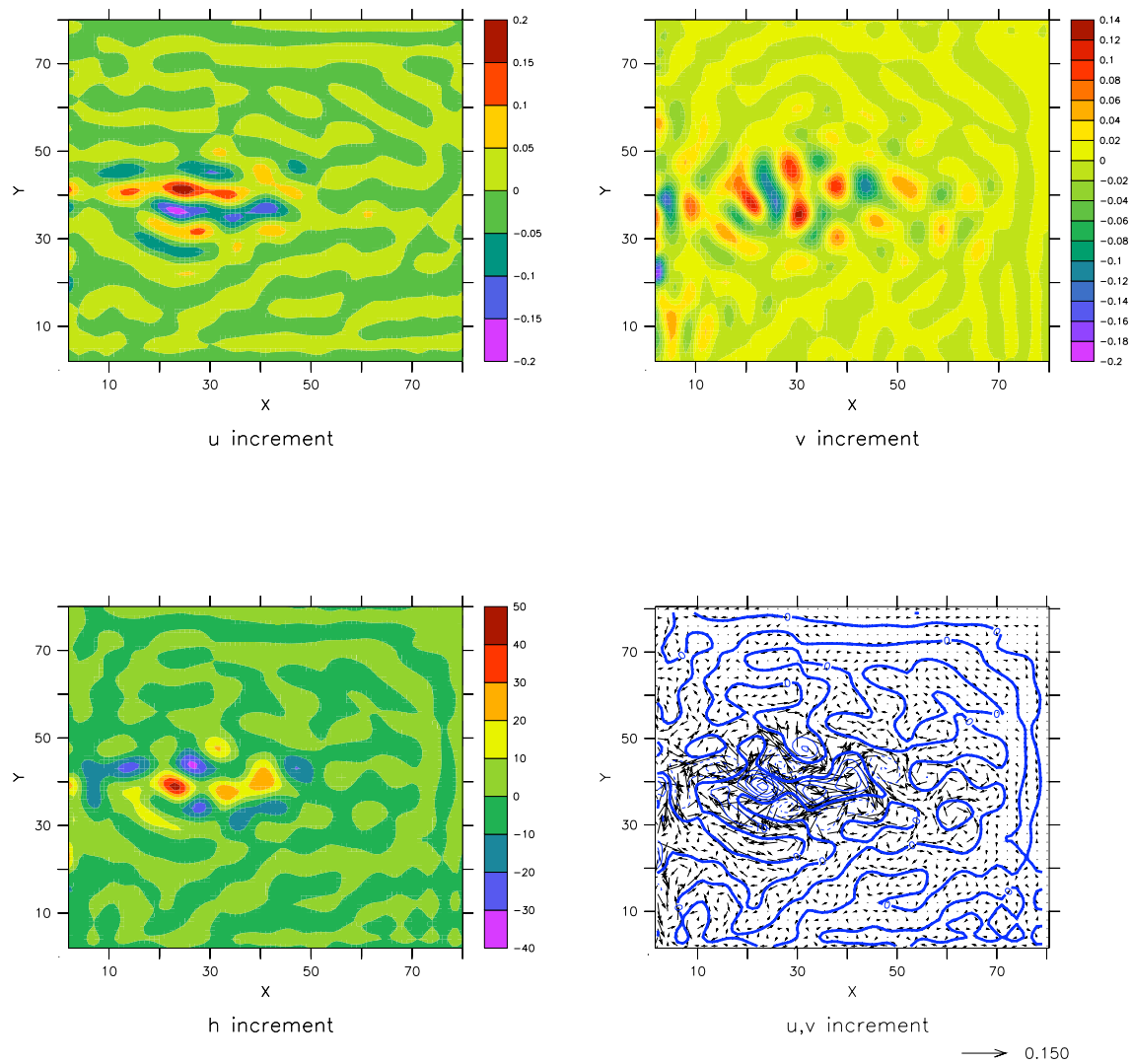


FIG. A.6: *Incrément sur les différentes composantes du vecteur d'état à l'issue de la troisième boucle externe*

Bibliographie

- [1] A. Adcroft and D. Marshall. How slippery are piecewise-constant coastlines in numerical ocean models. *Tellus*, 50A:95–108, 1998.
- [2] L. Amodei. Solution approchée pour un problème d’assimilation de données météorologiques avec prise en compte de l’erreur rmodèle. *C. R. Acad. Sci. Paris*, 321(12):1087–1094, 1995.
- [3] V. Auffray. Assimilation de données sur un modèle ”shallow water” à l’aide du coupleur palm. Rapport de stage insa, CERFACS, aug 2000. Available on the PALM web.
- [4] J. Ballabrera, P. Brasseur, and Verron J. Dynamical evolution of the error statistics with the seek filter to assimilate altimetric data in eddy-resolving ocean models. *Q.J.R.Met.Soc.*, submitted, 2000.
- [5] J. Barkmeijer, M. Van Gijzen, and F. Bouttier. Singular vectors and estimates of the analysis-error covariance metric. *Q.J.R. Meteorol. Soc.*, 124:1695–1713, 1997.
- [6] F. Barlier, P. Exertier, and P. Bonnefond. Regional validation in the mediterranean. *AVISO newsletter*, 1998.
- [7] A. Bennett. *Inverse methods in physical oceanography*. Cambridge monographs on mechanics and applied mathematics. Cambridge University Press, 1995.
- [8] A. Bennett, C. Boon, E. Harrison, and M. McPhaden. Generalized inversion of tropical atmosphere-ocean data and a coupled model of the tropical pacific. *Journal of Climate*, 11:1768–1792, July 1998.
- [9] A. F. Bennett, B. S. Chua, and L. M. Leslie. Generalized inversion of a global numerical weather prediction model. *Meteorol. Atmos. Phys.*, 60:165–178, 1996.
- [10] A. F. Bennett, B. S. Chua, and L. M. Leslie. Generalized inversion of a global numerical weather prediction model,ii: Analisis and implementation. *Meteorol. Atmos. Phys.*, 62:129–140, 1997.
- [11] P. Bergthorsson and B. R. Doos. Numerical weather map analysis. *Tellus*, 7:329–340, 1955.
- [12] I. Blanchet, C. Frankignoul, and M. A. Cane. A comparison of adaptive kalman filters for a tropical pacific ocean model. *Mon. Wea. Rev.*, 125:40–58, 1997.

-
- [13] B. Blancke and P. Delecluse. Variability of the tropical atlantic ocean simulated by a general circulation model with two different mixed layer physics. *J. Phys. Oceanogr.*, 23:1363–1388, 1993.
- [14] E. Blayo, J. Blum, and J. Verron. *Equation aux Dérivées Partielles et Applications*, chapter Assimilation variationnelle de données en océanographie et réduction de la dimension de l'espace de contrôle, page 199/219. Gauthier-Vilars, 1998.
- [15] E. Blayo, J. Verron, and J.-M. Molines. Assimilation of topex/poseidon altimeter data into a circulation model of the noerth atlantic. *J. Geophys. Res.*, 99:24691–24705, 1994.
- [16] R. Bleck and D.B. Boudra. Wind-driven spin-up in a eddy-resolving ocean models formulated in isopycnic and isobaric coordinates. *J. Geophys. Res.*, 91C:7611–7621, 1986.
- [17] F. Bouttier and P. Courtier. Data assimilation concepts and methods. Internal report, ECMWF, march 1999.
- [18] A. M. Bratseth. Statistical interpolation by means of successive corrections. *Tellus*, 38A:439–447, 1986.
- [19] N.P. Budgell. Nonlinear data assimilation for shallow water equations in branched channels. *J. Geophys. Res.*, 91:10633–10644, 1986.
- [20] M.A. Cane, A. Kaplan, R.N. Miller, B. Tang, E.C. Hackert, and A.J. Busalacchi. Mapping tropical pacific sea level: data assimilation via a reduced state space kalman filter. *J. Geophys. Res.*, 101:22599–22618, 1996.
- [21] S. Carme. *Méthodes d'assimilation de données oar Filtrage de Kalman dans un modèle réaliste de l'Atlantique Nord*. PhD thesis, Université Joseph Fourier, Grenoble I, 1999.
- [22] J. A. Carton, B. S. Giese, and X. Cao. Impact of altimeter thermistor, and expandable bathythermograph data on retrospective analyses of the tropical pacific ocean. *J. Geophys. Res.*, 101:14147–14159, 1996.
- [23] W. C. Chao and L-P. Chang. Development of a four-dimensional variational analysis system using the adjoint method at gla. part 1: Dynamics. *Monthly Weather Review*, 120:1661–1673, August 1992.
- [24] I. Charpentier and F.-X. Le Dimet. *Equation aux Dérivées partielles et Applications*, chapter Méthodes du second ordre en assimilation de données. Gauthier-Vilars, 1998.
- [25] P. G. Ciarlet. *Introduction à l'Analyse Numérique Matricielle et à l'Optimisation*. Masson, Paris, 1982.
- [26] S. E. Cohn. An introduction to estimation theory. *J. Meteor. Soc. Japan*, 75:257–288, 1997.

- [27] P. Courtier. Variational methods. *Geophys. Mag.*, 1, Spec. Issue:IV–i–1–19, 1995.
- [28] P. Courtier. Dual Formulation of Four-Dimensional Variational Assimilation. *Q. J. Royal Met. Soc.*, 123:2449–2461, 1997.
- [29] P. Courtier, J. C. Derber, R. Errico, J. F. Louis, and T. Vukićević. Important Literature on the Use of Adjoint, Variational Methods and the Kalman Filter in Meteorology. *Tellus*, 45A(1B):342–357, 1993.
- [30] P. Courtier, J.N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, 120:1367–1387, 1994.
- [31] P. Courtier et al. The ecmwf implementation of the three dimensional variational data assimilation (3d-var). i: Formulation. *Q.J.R.Met.Soc.*, 124:1783–1807, 1998.
- [32] G. Cressman. An optimal objective analysis system. *Mon. Weather Rev.*, 87:367–374, 1959.
- [33] R. Daley. *Atmospheric data analysis*. Cambridge University Press, 1986.
- [34] F. D’Andrea and R. Vautard. Reducing Systematic Errors by Empirically Correcting Model Errors. *Tellus*, 52A:21–41, 2000.
- [35] P. De Mey. Réglage du modèle quasi géostrophique en vue de l’assimilation dans la région açores-madère. Technical report, Rapport Quadran, 1993.
- [36] D. Dee and R. Todling. Data assimilation in the presence of forecast bias: The geos moisture analysis. *Monthly Weather Review*, 128:3268–3282, 2000.
- [37] D. P. Dee and A. M. Da Silva. Data assimilation in the presence of forecast bias. *Q. J. R. Meteorol. Soc.*, 124:269–295, 1998.
- [38] Dick P. Dee. On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*, 123:1128–1143, April 1995.
- [39] J Derber and F Bouttier. A reformulation of the background error covariance in the ecmwf global data assimilation system. *Tellus*, 51A:195–221, 1999.
- [40] J. C. Derber. A Variational Continuous Assimilation Technique. *Mon. Wea. Rev.*, 117:2437–2446, 1989.
- [41] John C. Derber. Variational four-dimensional analysis using quasi-geostrophic constraint. *Monthly Weather Review*, 115:998–1008, 1987.
- [42] S. Durbiano. *Vecteurs caractéristiques de modèles océaniques pour la réduction d’ordre en assimilation de données*. PhD thesis, Université Joseph Fourier, Grenoble I, 2001.

-
- [43] G. Egbert and A. Bennett. *Modern Approaches to Data Assimilation in Ocean Modeling*, chapter Data Assimilation Methods for Ocean Tides, pages 147–179. Malanotte-Rizzoli Editor, Elsevier Press, 1996.
- [44] M. Eherendorfer. Four-dimensional data assimilation: comparison of variational and sequential algorithms. *Quarterly Journal of the Royal Meteorological Society*, 118(506):673–714, 1992.
- [45] Anderson et al. The ecmwf implementation of the three dimensional variational data assimilation (3d-var). iii: Experimental results. *Q.J.R.Met.Soc.*, 124:1831–1860, 1998.
- [46] G. Evensen. Advanced data assimilation for strongly non linear dynamics. *Mon. Wea. Rev.*, 125:1342–1354, 1997.
- [47] C. FAURE and Y. Papegay. Odysée version 1.6, the language reference manual. Technical report, 211, INRIA, 1997.
- [48] M. Fisher. Efficient minimization of quadratic penalty functions. *Q. J. R. Meteorol. Soc.*, 999:1–14, 1997.
- [49] M. Fisher and P. Courtier. Estimating the covariance matrices of analysis and forecast error in variational data assimilation. Technical memorandum, ECMWF, August 1995.
- [50] L.S. Gandin. Objective analysis of meteorological fields. *Gidrometeorol. Isd. Leningrad*, 1963.
- [51] G. Gaspari and S. E. Cohn. Construction of correlation functions in two and three dimensions. *Q. J. R. Metoreol. Soc.*, 125:723–757, 1999.
- [52] J.-C. Gilbert and C. Lemaréchal. The modules M1QN3 and N1QN3. version 2.0c. Technical report, Inria, June 1995.
- [53] E. Greiner, S. Arnault, and A. Moliere. Twelve monthly experiments of 4d-variational assimilation in the tropical atlantic during 1987. part 1: Method and statistical results. *Progress in Oceanography*, 41:141–202, 1998.
- [54] E. Greiner and C. Perigaud. Assimilation of geosat altimetric data in a nonlinear reduced-gravity model of the indian ocean by adjoint approach, part 1 : method and statistical results. *J. Phys. Oceanogr.*, 24:1783–1804, 1994.
- [55] A. K. Griffith. *Data Assimilation for Numerical Weather Prediction Using Control Theory*. PhD thesis, University of Reading. Dept of Mathematics, 1997.
- [56] A. K. Griffith and N. K. Nichols. Accounting for model error in data assimilation using adjoint methods. *Computational Differentiation : Techniques, Applications and tools*, SIAM, pages 195–204, 1996.

-
- [57] A. K. Griffith and N. K. Nichols. Adjoint methods in data assimilation for estimating model error. *J. Flow, Turbulence and Combustion.*, submitted, 2000.
- [58] J. Hadamard. *La théorie des équations aux dérivées partielles*. Editions Scientifiques, Pekin, 1964.
- [59] K. Haines, P. Malanotte-Rizzoli, R. E. Young, and W. R. Holland. A comparison of two methods for the assimilation of altimetry data into a shallow water model. *Dyn. Atmos. Oceans*, 17:89–133, 1993.
- [60] W. R. Holland and P. Malanotte-Rizzoli. Along track assimilation of altimeter data into an ocean circulation model: Space versus time resolution studies. *J. Phys. Oceanogr.*, 19:1507–1534, 1989.
- [61] I. Hoteit. *Filtres de Kalman réduits et efficaces pour l'assimilation de donnée en océanographie*. PhD thesis, Université Joseph Fourier (Grenoble I), 2001.
- [62] I. Hoteit, D.-T. Pham, and J. Blum. Efficient reduced kalman filtering and application in tropical pacific. Rapport de recherche simulation et optimisation de systemes complexes, INRIA, 655, avenue de l'europe, 38330 Montbonnot St Martin (france), 2000.
- [63] K. Ide, P. Courtier, M. Ghil, and A.C. Lorenc. Unified notation for data assimilation: operational, sequential and variational. *J. Meteor. Soc. Japan*, 75 1B:181–189, 1997.
- [64] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
- [65] R. H. Jones. Optimal Estimation of Initial Conditions for Numerical Analysis. *J. Atmos. Sci.*, 22:658–663, 1965.
- [66] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transaction of the ASME—Journal of Basic Engineering*, pages 35–45, 1960.
- [67] P. Knudsen and O. B. Andersen. Sea level and sea surface temperature changes from satellites. In *Physics and Chemistry of the Earth*. Ed. H.P. Plag, Elsevier Science, 1997.
- [68] T. N. Krishnamurti, X. Jishan, H. S. Bedi, K. Ingles, and D. Oosterhof. Physical initialization for numerical weather prediction over the tropics. *Tellus*, 43A:53–81, 1991.
- [69] J.-F. Lacarra and O. Talagrand. Short range evolution of small perturbations in a barotropic model. *Tellus*, 40A:81–95, 1988.
- [70] T. Lagarde. *Nouvelle approche des methodes d'assimilation de donnees : les algorithmes de point selle*. PhD thesis, Universite Paul Sabatier (Toulouse III), 2000.
-

-
- [71] F.-X. Le Dimet. Une étude générale d'analyse objective variationnelle des champs météorologiques. Rapport scientifique du laboratoire associé de météorologie physique 28, Université de Clermont II, september 1980.
- [72] F.-X. Le Dimet and O. Talagrand. Variational Algorithms for Analysis and Assimilation of Meteorological Observations: Theoretical Aspects. *Tellus*, 38A:97–110, 1986.
- [73] J-M Lellouche, J-L Devenon, and I Dekeyser. Boundary control of burgers'equation - a numerical approach. *Computer Math Applic*, 28(5):33–44, 1994.
- [74] J. M. Lewis and J. C. Derber. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, 37A:309–322, 1985.
- [75] J.L. Lions. *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles*. Dunod, Paris, 1968.
- [76] D.-C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45:503–528, 1989.
- [77] A. C. Lorenc. Analysis Methods for Numerical Weather Prediction. *Q. J. Royal Met. Soc.*, 112:1177–1194, 1986.
- [78] A. C. Lorenc, R. S. Bell, and B MacPherson. The met. office analysis correction data assimilation scheme. *Quart. J. R. Met. Soc.*, 117:59–89, 1991.
- [79] S Louvel. *Étude d'un algorithme d'assimilation variationnelle de données à contrainte faible. Mise en oeuvre sur le modèle océanique MICOM*. Mathématiques appliquées à l'océanographie, Université Paul Sabatier (Toulouse III), LEGOS 14, avenue Edouard Belin - 31400 Toulouse, December 1999.
- [80] B. Luong, J Blum, and J Verron. A variational method for the resolution of data assimilation problem in oceanography. *Inverse problems*, 14:979–997, 1998.
- [81] W. H. Lyne, R. Swinbank, and N. T. Birch. A data assimilation experiment and the global circulation during the fgge special observing periods. *Q. J. Roy. Meteorol. Soc.*, 108:575–594, 1982.
- [82] G. Madec, P. Delecluse, M. Imbard, and C. Lévy. Opa 8.1 ocean general circulation model : Reference manuel. Note 11, IPSL, dec 1998.
- [83] J. M. Martin, N. K. Nichols, and M. J. Bell. Treatment of systematic errors in sequential data assimilation. Ocean applications Technical note 21, The Met. Office, july 1999.
- [84] A.M. Moore. Data assimilation in a quasigeostrophic open ocean model of the gulf stream region using the adjoint method. *J. Phys. Oceanogr.*, 21:382–427, 1991.
- [85] M. Mu. Nonlinear singular vectors and nonlinear singular values. ??, ??(??):??, 1999.

-
- [86] K.A. Myers and B.D. Tapley. Adaptive sequential estimation with unknown noise statistics. *IEEE Trans, Automat. Contr.*, AC-21:520–523, 1975.
- [87] R. S. Nerem, B. J. Haines, J. Hendricks, J. F. Minster, G. T. Mitchum, and W. B. White. Improved determination of global mean sea level variations using topex/poseidon altimeter data. *Geophys. Res. Ltr.*, 24:1331–1334, 1997.
- [88] H. E. Ngodock. *Assimilation de Données et Analyse de sensibilité : Une application a la circulation oceanique*. PhD thesis, Université Joseph Fourier (Grenoble I), 1996.
- [89] R. C. Pacanowski and S. G. H. Philander. Parametrization of vertical mixing in numerical models of tropical oceans. *J. Phys. Oceanogr.*, 11:1443–1451, 1981.
- [90] Yu. B. Panteleev, G. G. Filyushkin. A comparative analysis of the methods of optimal and variational interpolation of data on the velocity of oceanic currents. *Oceanology, english translation*, 35(1):16–19, 1995.
- [91] L. Parent. *Assimilation de données dans l’océan Pacifique Tropical sur la période 1994-1998*. PhD thesis, Université Joseph Fourier (Grenoble I), 2000.
- [92] D. F. Parish and S. E. Cohn. A kalman filter for a two dimensional shallow-water model: Formulation and preliminary experiments. office note 304, NMC, 1985.
- [93] D. F. Parish and J. C. Derber. The national meteorological center’s spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, 120:1747–1766, 1992.
- [94] D.T. Pham, J. Verron, and L. Gourdeau. Filtre de kalman sinulier évolutif pour l’assimilation de données en océanographie. *C. R. Acad. Sci. Paris*, in-press, 1997.
- [95] D.T. Pham, J Verron, and M. C. Roubaud. Singular evolutive kalman filter with eof initialization for data assimilation in oceanography. *J. Mar. Syst.*, 16:323–340, 1997.
- [96] S. G. H. Philander. *El Niño, La Niña, and the Southern Oscillation*, page 293. Academic Press, San Diego, 1990.
- [97] C. Pires, R. Vautard, and O. Talagrand. On extending the limit of variational assimilation in non-linear chaotic systems. *Tellus*, 47A, 1995.
- [98] F. Rabier and P. Courtier. Four-dimensional assimilation in presence of baroclinic instability. *Q. J. R. Meteorol. Soc.*, 118:649–672, 1992.
- [99] F. Rabier, Thépaut J.-N., and P. Courtier. Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Q. J. R. Meteorol. Soc.*, 124:1861–1887, 1998.
- [100] G. Reverdin, E. Frankignoul, E. Kestenare, and M. J. McPhaden. Seasonal variability in the surface currents of the equatorial pacific. *J. Geophys. Res.*, 99:20323–20344, 1984.
-

-
- [101] L. P. Riishøjgaard. A method for estimating the analysis error variance in a physical space data assimilation system. *Q.J.R. Meteorol. Soc.*, 126:1367–1385, 2000.
- [102] M. L. Salby. Sampling theory for asynoptic satellite observations. part i: Space-time spectram resolution, and aliasing. *Journal of the Atmospheric Sciences*, 39:2577–2600, November 1982.
- [103] Y. Sasaki. A fundamental study of the numerical prediction based on the variational principle. *J. Met. Soc. Japan*, 33:262–275, 1955.
- [104] Y. Sasaki. An objective analysis based on the variational method. *J. Met. Soc. Japan*, 36:77–78, 1958.
- [105] Y. Sasaki. Some basic formalisms in numerical variational analysis. *Mon. Wea. Rew.*, 98:875–883, 1970.
- [106] J. Shröter, U. Seiler, and M Wensel. Variational assimilation of geosat data into a eddy-resolving model of the gulf stream area. *J. Phys. Oceanogr.*, 23, 1993.
- [107] K. Swanson, T. Palmer, and R. Vautard. Error structure and the value of 4dvar. *Tellus*, submitted, 1999.
- [108] K. L. Swanson, R. Vautard, and C. Pires. Four-dimensional variational assimilation and predictability in a quasi geostrophic model. *Tellus*, 50A:369–390, 1997.
- [109] L. L. Takacs. A simple bias correction algorithm for use in data assimilation. Office note Series on Global Modeling and Data Assimilation 96-25, DAO, December 1996.
- [110] O. Talagrand. Adjoint Methods for the Atmospheric Sciences. Paper presented at the ECMWF annual seminar, September 1991.
- [111] O. Talagrand. The Use of Adjoint Equations in Numerical Modelling of the Atmospheric Circulation. In A. Grivank and G. G. Corliss, editors, *Proceedings of Workshop on Automatic Differentiation of Algorithms: Theory, Implementation and Application*, Breckenridge, Colorado, USA, January 1991. SIAM.
- [112] O. Talagrand. A Posteriori Evaluation and Verification of Analysis and Assimilation Algorithmsa. In *Proceedings of Workshop on Diagnosis of Data Assimilation Systems*. European Centre for Medium-range Weather Forecasts, December 1998.
- [113] O. Talagrand and P. Courtier. Variational Assimilation of Meteorological Observations with the Adjoint Vorticity Equation. Part I: Theory. *Q. J. Royal Met. Soc.*, 113:1311–1328, 1987.
- [114] C.-E. Testut. *Assimilation de données satellitales avec un filtre de Kalman de rang réduit dans un modèle aux équations primitives de l’océan Atlantique*. PhD thesis, Université Joseph Fourier (Grenoble I), 2000.

- [115] J.N. Thépaut and P. Courtier. Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Q. J. Royal Met. Soc.*, 117:1225–1254, 1991.
- [116] J.N. Thépaut and et al. The ecmwf implementation of the four-dimensional variational data assimilation (4d-var). i: Formulation. *Q. J. Royal Met. Soc.*, in press, 1999.
- [117] L Trondheim. *Lapinot et le carottes de Patagonie*. L’association, 1986.
- [118] F. Uboldi and M. Kamachi. Time–space weak–constraint data assimilation for non-linear models. *Tellus*, Submitted, 1999.
- [119] F. Veersé, D.-T. Pham, and J. Verron. 4d-var/seek : A consistent hybrid variational-smoothing data assimilation method. Rapport de recherche 3902, INRIA, March 2000.
- [120] F. Veersé and J. N. Thépaut. Multiple–truncation incremental approach of four-dimensional variational data assimilation. *Q. J. Royal Met. Soc.*, 124:1889–1908, 1998.
- [121] J. Verron. Nudging satellite altimeter data into quasi-geostrophic ocean model. *J. Geophys. Res.*, 97:7479–7491, 1992.
- [122] J. Vialard, C. Menkes, J.-P. Boulanger, P. Delecluse, E. Guilyardi, M. McPhaden, and G. Madec. A model study of oceanic mechanisms affecting equatorial pacific sea surface temperature during the 1997-98 el niño. *Journal of Physical Oceanography*, ?in press, 2001.
- [123] A. Vidard. Assimilation de données avec contrôle de l’erreur modèle. Rapport de dea, LMC-IMAG, june 1998.
- [124] A. Vidard, E. Blayo, F.-X. Le Dimet, and A. P. 4d-variational data analysis with imperfect model. reduction of the size of control. *Journal of Flow, Turbulence and Combustion*, ?in press, 2001.
- [125] A. Vidard, F.-X. Le Dimet, and A. Piacentini. Optimal determination of nudging coefficient. *Tellus*, ?submitted, 2001.
- [126] A. Vidard, F.-X. Le Dimet, A. Piacentini, and A. Weaver. Controle du biais systematique du modele en assimilation de donnee variationnelle. In *AMA 2000*, 2000.
- [127] A. Weaver and P. Courtier. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.*, 127:1815–1846, 2001.
- [128] A. Weaver, J. Vialard, D.L.T. Anderson, and P. Delecluse. Three- and four dimensional variational assimilation with a general circulation model of the tropical pacific ocean. *in preparation*, 2001.

- [129] W. Wergen. The effect of model errors in variational assimilation. *Tellus*, 44A:297–313, 1992.
- [130] C. Wunsch. *The Ocean Circulation Inverse Problem*. Cambridge University Press, 1996.
- [131] L. Zhijin and I. M. Navon. Optimality of 4d-var and its relationship with the kalman filter and kalman smoother. *Q. J. R. Meteorol. Soc.*, submitted, 1999.
- [132] X Zou, I. M. Navon, and F-X. Le Dimet. An optimal nudging data assimilation scheme using parameter estimation. *Q. J. R. Meteorol. Soc.*, 118:1163–1186, 1992.
- [133] M. Zupanski. Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rew.*, 121:2396–2408, 1993.