



HAL
open science

Assimilation rétrospective de données par lissage de rang réduit : application et évaluation dans l'Atlantique Tropical

Nicolas Freychet

► **To cite this version:**

Nicolas Freychet. Assimilation rétrospective de données par lissage de rang réduit : application et évaluation dans l'Atlantique Tropical. Sciences de la Terre. Université de Grenoble, 2012. Français. NNT : 2012GRENU003 . tel-00683971

HAL Id: tel-00683971

<https://theses.hal.science/tel-00683971>

Submitted on 30 Mar 2012

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Terre, Univers, Environnement**

Arrêté ministériel : 7 aout 2006

Présentée par

Nicolas Freychet

Thèse dirigée par **Emmanuel Cosme**
et codirigée par **Pierre Brasseur**

préparée au sein du **Laboratoire des Ecoulements Géophysiques et Industriels**
et de l'**Observatoire des Sciences de l'Univers de Grenoble**

Assimilation rétrospective de données par lissage de rang réduit : application et évaluation dans l'Atlantique tropical.

Thèse soutenue publiquement le **11 janvier 2012**,
devant le jury composé de :

M. Eric BLAYO

Professeur UJF, Président

M. Gérald DESROZIERS

Ingénieur Divisionnaire des Travaux de la Météorologie (HDR), Météo-France,
Rapporteur

M. Lionel GOURDEAU

DR, IRD, LEGOS/CNES, Rapporteur

M. Joaquim BALLABRERA-POY

Chargé de recherche, CSIC, Examineur

M. Emmanuel COSME

Maitre de conférences UJF, Directeur de thèse

M. Pierre BRASSEUR

DR CNRS, LEGI, Co-Directeur de thèse



Après 3 années de travail acharné (si si...), il est enfin venu le temps de rendre justice à tous ceux ayant participé (volontairement ou pas) à cette aventure.

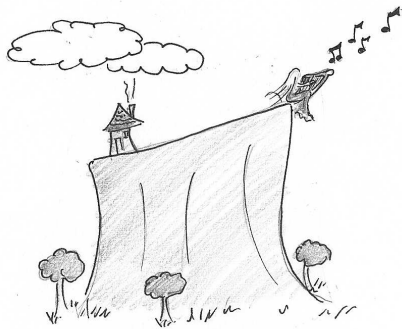
Dans un premier temps, je tiens à remercier ceux qui ont été derrière moi pour surveiller mon efficacité jours et nuits (ou presque), c'est-à-dire mes deux directeurs de thèse : Pierre et Manu. Tant scientifiquement que humainement, ces 3 années passées sous votre direction ont été une expérience unique. Manu, pour ta première thèse tu t'en ai très bien sorti ! Merci de m'avoir donné la chance d'en apprendre autant sur un sujet dont j'ignorais l'existence même avant de commencer... Et puis j'ai été très heureux d'avoir droit à mon petit bout de casier dans ton bureau, Pierre, au milieu de tes millions de projets. Je n'oublie pas non-plus celui qui pourrait être mon troisième directeur de thèse, ou ma bonne conscience, ou mon déblocueur de situations désespérées : Jean-Michel. Sans ton aide, je serais sans-doute toujours en train d'essayer de compiler SESAM... Je tiens également à remercier les rapporteurs de ces travaux, Gérald et Lionel, qui ont su apporter un regard critique, poser de nouvelles questions et donner de multiples directions pour des études futures. Merci également aux examinateurs, Eric et Joachim, pour avoir accepté "d'endurer" ma soutenance et apporté des remarques complémentaires aux rapports.

Enfin, je tiens à souligner (mais je sais pas trop comment faire avec Latex) que j'ai été très heureux de passer tout ce temps parmi l'équipe MEOM. Cette équipe possède quelque chose d'unique de par la qualité de ses gens, de ses travaux, et de son esprit de groupe. J'envie tous les futurs thésards qui auront la chance de vivre cette aventure. Je ne citerai pas les noms de chacun, mais je tiens quand même à remercier profondément Josiane qui arrive à gérer tous nos problèmes administratifs en gardant le sourire... Egalement mes soeurs de thèse, Mélanie, Marion et Carolina, qui ont toutes survécues à cette épreuve (sans doute grâce à mes réserves de gateaux que j'ai pourtant tenté de cacher au fond de mes tiroirs). Et puis y'a évidemment la Mère Dudu sur qui j'ai pu défouler mes nerfs et vider mon pistolet à fléchettes pendant qu'elle matait sa page facebook... (pas la peine de t'énerver, tout le monde sait que tu passais tes journées à ça...). Sans oublier les autres, Clément, Raph, Gildas, Albane, Angélique, et puis les nouveaux arrivants à qui je souhaite bon courage (c'est pas facile de me succéder après tout...)!

Bon, maintenant il faut aussi que je m'occupe un peu de ceux qui ont endurés cette période autant que moi... D'une part ma famille qui m'a toujours soutenu quelles que soient mes décisions (souvent changeantes d'ailleurs), et notamment ma grand-mère qui croit toujours que je suis ingénieur des Ponts et Chaussées... Et puis il y a ceux qui me suivent (ou que je harcèle) depuis pas mal d'années déjà, et qui sont restés là malgré mes sauts d'humeurs lors de cette fascinante période qu'est la thèse. Pour ça je vous remercie, et j'apprécie encore grandement la vision de mon avenir que vous ne cessez de m'exposer chaque jour, malgré mes vaines tentatives de vous expliquer ce que je fais... (voir figure 1). Grâce à vous, ma thèse a pris de nombreuses autres dimensions en dehors du boulot (j'entends déjà les remarques du genre "Ah mais parce que tu bossais vraiment?" ...).

D'abord, ça a été une envolée au pays des sushis partagée dans la joie (et la douleur (bobo genoux...)) avec Elo. Si j'avais pu imaginer qu'une simple pause café quelques années

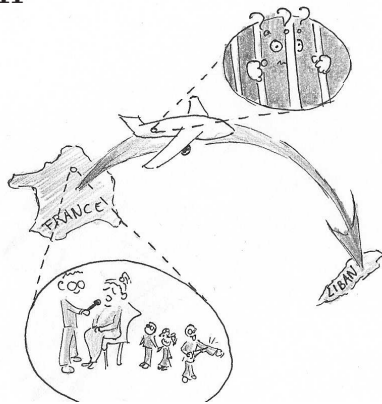
cas I



cas II



cas III



cas IV

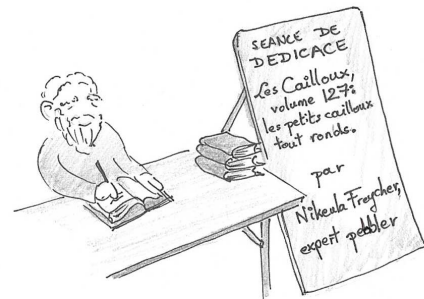


FIGURE 1 – L’avenir de Nicolas Freychet tel qu’il est prédit par différents experts : cas I : “Le Trièves : la harpe et la perruque blonde” par Quiquet, 2010 ; cas II : “Le Trièves bis : la poterie au bord de la route” par Quiquet, 2011 ; cas III : “Témoignage chez Delarue : retour aux origines” par Ronie-be, 2010 ; cas IV : “Nikeula Freychet : un expert peebler comme on n’en fait plus” par Varin, 2012, basé sur une étude de Buchonnet, 2009 et antérieur.

plus tôt allait conduire à tout ça...! J’attends la suite avec impatience maintenant!! Et puis il y a eu les débats, les déboires et les moments inoubliables (en bien ou en mal) avec Aurel. Je me demande comment on a fait pour tenir en se voyant tous les midi à EVE pendant 3 ans (enfin surtout comment TU as réussi à me supporter tout ce temps...). Ca vallait bien le coup de se taper Chamechaude de nuit pour ça tiens! Ah tiens d’ailleurs tu ne me devais pas un repas (très cher si possible...)? Ensuite, ma culture cinématographiquement improbable s’est grandement améliorée grâce à Rachel, et Jo et toutes nos soirées passées à nous demander “Mais heu... c’est quoi le but de regarder ça déjà?!”. La bandes des poulettes, Anne, Loren et Vilaine, compagnes de beuveries à l’Ouest et de débats (haute-ment philosophiques évidemment) en faisant des bains de pieds devant “La France a un incroyable talent” (je vais regretter d’écrire ça je crois...). Et puis y’a Roro (c’est où le Nord déjà hein?!). J’ai réservé nos places chez Delarue pour dans une dizaine d’années, donc faut qu’on s’active hein! Sans compter que je risque de me retrouver au Liban bientôt à force de passer les contrôles de douane!

Voilà, il y en a beaucoup d’autres qui sont intervenus, c’est long à écrire tout ça... Alex, malgré tes airs de Montpellierains j’ai appris à t’apprécier :-)

Et ta fidèle compagne Marie

qui nous a trahis pour aller rejoindre l'ennemi (en Suisse quoi..). Lucile, Dom et toute la bande de la MC2 avec qui je continue de passer des moments irréalistes (cf les Canaries par exemple...). Leo également, quand tu auras fini de squatter chez les chinois on pourra sûrement continuer de travailler ton français... A mes amis d'enfance, Mathieu et Emilie(x2), je vous promets que cette fois c'est bon j'ai fini les études! Et aussi la bande d'Auspex, PyB, Alex, Lio, Arthur et Fred (et re-Elo) qui ont eu la gentillesse de me surcharger un peu plus de boulot (mais ce fut et c'est encore une bonne collaboration!). Enfin, une ptite pensée aussi à toi qui a rendu ma fin de thèse un peu plus tumultueuse qu'elle ne l'était déjà!

Et puis qu'on se le dise une bonne fois pour toutes : les cailloux, y'en a partout!

Table des matières

Introduction	1
I L’Atlantique Tropical : dynamique, observations et modélisation	7
1 La dynamique de l’océan en Atlantique Tropical	9
1.1 Introduction	10
1.2 Circulation moyenne	10
1.2.1 Caractéristiques liées à la position tropicale	10
1.2.2 Circulation moyenne en Atlantique tropical	11
1.3 La variabilité saisonnière	12
1.3.1 Ondes tropicales d’instabilité (TIWs)	13
1.3.2 Anneaux du Brésil	15
1.4 La variabilité interannuelle	17
1.5 Conclusion	18
2 L’observation de l’océan	19
2.1 Introduction	20
2.2 Bref historique	20
2.3 Les relevés <i>in situ</i>	22
2.3.1 Les mouillages	22
2.3.2 Les bouées dérivantes	23
2.3.3 Autres instruments de mesures	24
2.3.4 Les campagnes de mesures océanographiques	25
2.4 Les données satellites	25
2.4.1 La mesure altimétrique	26
2.4.2 Les autres types d’observations	27
2.5 Les radars	28
2.6 Conclusion	28
3 Modélisation de l’océan	31
3.1 Introduction	32
3.2 L’outil numérique : le code OPA et le modèle NEMO	32
3.3 La physique résolue par le modèle	33
3.3.1 Les équations primitives	33
3.3.2 Les conditions limites	35
3.3.3 La hauteur de la mer : SSH	37
3.4 Le traitement des équations dans le milieu discret	38

3.4.1	La discrétisation spatiale et temporelle	38
3.4.2	Paramétrisation des échelles sous-mailles	38
3.5	La configuration utilisée : TATL4	40
3.5.1	Le domaine d'étude et les conditions aux frontières	41
3.5.2	Caractéristiques de la discrétisation	43
3.5.3	Les forçages utilisés	43
3.5.4	Représentation de la dynamique dans TATL4	45
3.5.5	Récapitulatif	47
3.6	Oscillation haute fréquence de la SSH et des forçages	47
3.7	Conclusion	49
 II Assimilation de données		51
4	Filtre de Kalman et lisseurs linéaires optimaux	53
4.1	Introduction	54
4.2	Théorie de l'estimation et le filtre de Kalman	55
4.2.1	Formulation du filtrage	56
4.2.2	Le filtre de Kalman	57
4.2.3	Les limites du filtrage	57
4.3	Problème du lissage : les différentes approches	58
4.3.1	Le lissage conjoint	58
4.3.2	Le lissage marginal	64
4.3.3	Les thématiques d'application du lissage	64
4.4	Le lisseur fixed-lag	65
4.4.1	Formulation du lisseur de Kalman	66
4.4.2	L'erreur modèle indispensable au lissage	66
4.5	Conclusion	67
5	Lisseur et filtre SEEK	69
5.1	Introduction	70
5.2	Mise en place du filtre et du lisseur de Kalman dans un environnement réaliste : développement du SEEK	70
5.2.1	La réduction d'ordre	70
5.2.2	Formulation SEEK du filtre	71
5.2.3	Formulation SEEK du lisseur	72
5.2.4	Résumé des équations du filtre et du lisseur SEEK	74
5.3	Paramétrisations du filtre et du lisseur SEEK	75
5.3.1	Initialisation : l'état \mathbf{x}_0^a	75
5.3.2	Initialisation : \mathbf{S}_0^a et méthode de réduction de la matrice des covariances d'erreur	75
5.3.3	Evolutivité des covariances d'erreur : les différents types de filtres	76
5.3.4	La prise en compte de l'erreur modèle \mathbf{Q}	78
5.3.5	Localisation de l'analyse	79
5.4	Implémentation du SEEK dans NEMO : le logiciel SESAM	81
5.5	Quelques précisions supplémentaires	82
5.5.1	Vecteur d'état et vecteur d'observation	82

5.5.2	Un outil de diagnostic : le calcul d'erreur RMS	82
5.6	Conclusion	83
III Préliminaires au lissage : mise en place du filtre évolutif		85
6	Méthode d'ensemble et perturbation de paramètres	87
6.1	Introduction	88
6.2	Outil d'analyse de la vitesse des ondes : la transformée de Radon	88
6.3	Effet d'une perturbation sur l'état de l'océan	89
6.3.1	Effets sur la vitesse des TIWs	90
6.3.2	Effets sur la température de surface	92
6.3.3	Effets sur les vitesses	94
6.3.4	Conclusion sur les expériences de sensibilité	95
6.4	Création de l'ensemble	95
6.4.1	Méthode de tirages aléatoires	96
6.4.2	Création de l'ensemble	97
6.4.3	Variances et corrélations	97
6.5	Conclusion	99
7	Initialisation et paramétrisation du filtre évolutif	101
7.1	Introduction	102
7.2	La variabilité interannuelle, une source d'incertitude	102
7.2.1	Etat de référence	102
7.2.2	Création d'un état faux	104
7.2.3	Construction des statistiques d'erreur initiales	105
7.3	Création d'observations (parfaites)	106
7.3.1	Traces satellites	107
7.3.2	Réseau <i>in situ</i>	107
7.4	Paramétrisation du filtre SEEK évolutif : efficacité de l'analyse	109
7.4.1	Taille de l'analyse locale	111
7.4.2	Détermination de la taille de la matrice d'erreur S	112
7.4.3	Influence des observations	114
7.5	Paramétrisation du filtre SEEK évolutif : propagation de l'erreur	115
7.5.1	Taille de l'analyse locale	116
7.5.2	Détermination de la taille de la matrice d'erreur S	116
7.5.3	Détermination de la matrice des erreurs d'observation	118
7.6	Conclusion	119
IV Expériences de lissage dans un modèle réaliste de circulation océanique		123
8	Mise en place et évaluation du lisseur - modes évolutifs	125
8.1	Introduction	126
8.2	Contexte de l'expérience	126
8.2.1	La génération de l'erreur	126
8.2.2	Les réseaux d'observation simulés	129

8.3	Assimilation de données par filtrage et lissage	130
8.3.1	Réanalyses	130
8.3.2	“Cohérence dynamique” des solutions analysées	144
8.4	Conclusion	149
9	Mise en place et évaluation du lisseur - interpolation optimale	151
9.1	Introduction	152
9.2	Mise en place des expériences	152
9.3	Représentation 4D des covariances d’erreur	153
9.4	Lissage avec une base de covariances d’erreur “semi-fixe”	159
9.4.1	Réseau “temps-réel”	159
9.4.2	Réseau “temps-groupé”	161
9.4.3	Bilan et compléments sur les réanalyses avec une base fixe	162
9.5	“Cohérence dynamique” des analyses avec une base de covariances d’erreur fixe	162
9.5.1	Réseau “temps-réel”	162
9.5.2	Réseau “temps-groupé”	163
9.5.3	Bilan et compléments sur la “cohérence dynamique” avec un système à base fixe	164
9.6	Conclusion	165
	Conclusions et perspectives	166
	Bibliographie	174
	Listes des figures	184
	Listes des tables	194

Introduction

Les origines de l'assimilation de données remontent aux années 50 (*Bergthorsson et Döös*, 1955 ; *Eliassen*, 1954 ; *Cressman*, 1959 et *Gandin*, 1963) avec l'analyse optimale en météorologie, et un peu plus tard (dans les années 80) en océanographie, lorsque les réseaux d'observations commençaient à être suffisamment denses et précis. Le terme *assimilation* est né à la fin des années 60 avec l'introduction systématique des observations dans les premiers modèles numériques de prévision météo. La première motivation de l'assimilation de données était de trouver le meilleur état possible pour initialiser une prévision. Les analyses combinaient alors la dynamique issue des modèles avec les informations tirées des observations pour en tirer le meilleur parti, et ce, tout en tenant compte des éventuelles erreurs imputées à l'une ou l'autre source d'information. Il s'agissait alors d'un problème de filtrage (donc tri-dimensionnel), où chaque observation était assimilée à la date où elle était observée.

Si le principe théorique de base demeure le même depuis, les méthodes n'ont cessé d'évoluer avec le temps. L'amélioration des outils informatiques et la naissance de nouveaux enjeux en sont les principales raisons. En parallèle de ces aspects, c'est également le degré de confiance donné aux différentes sources d'information lors de l'analyse qui a guidé la formulation des principales théories d'assimilation actuelles. En effet, chaque source d'information (représentation de la dynamique par les modèles numériques, observations directes ou indirectes) peut être associée à un certain degré de confiance, ou autrement dit, à un certain niveau d'erreur. Deux approches distinctes ont été formulées pour considérer le problème d'initialisation de prévision et de prise en compte des erreurs. L'approche variationnelle (*Sasaki*, 1955 ; *Sasaki*, 1958) repose sur la théorie du contrôle optimal (*Lions*, 1968). L'autre approche, basée sur l'analyse optimale (et donc statistique), utilise des considérations statistiques pour trouver la meilleure combinaison entre des observations et une représentation modèle au moment de l'analyse. La formulation la plus répandue des méthodes statistiques est le filtre de Kalman (*Kalman*, 1960).

Qu'il s'agisse des méthodes statistiques ou variationnelles, une question essentielle demeure au coeur de leur théorie : quelle place donner aux observations pour analyser un état numérique ? Au-delà des erreurs associées à ces observations, c'est également leur représentativité dans l'espace et le temps qui doit être définie. Naturellement, l'information des observations a toujours été considérée spatialement (analyse objective). La formulation du filtre de Kalman, forme la plus répandue de la branche statistique, a permis d'étendre l'analyse objective en propageant dynamiquement les résultats des analyses. Ce type de système prend ainsi en compte la dimension temporelle, mais seulement dans un sens (propagation vers le futur). Les systèmes opérationnels d'océanographie (par exemple Mercator¹) utilisent d'ailleurs ce type de configuration. Dans le cadre de réanalyses, cette formulation peut être discutée. En effet, si l'on considère qu'une série de mesures est déjà disponible, il est tout à fait légitime de vouloir utiliser cette série et propager ces informations de façon rétrograde pour corriger des états précédant ces observations.

Dans les problèmes géophysiques (météorologie ou océanographie), la dynamique peut être fortement liée dans le temps. Par exemple la situation météo en un jour est liée aux conditions atmosphériques sur les jours précédents (et réciproquement) par les lois géophysiques gouvernant la dynamique de l'atmosphère. Pour répondre à cet enjeu et définir une formulation quadri-dimensionnelle du problème d'assimilation, les méthodes séquentielles ont vu l'introduction d'algorithmes de **lissage** (voir par exemple *Anderson*

1. <http://www.mercator-ocean.fr>

et Moore, 1979 et Simon, 2006) qui permettent d'étendre l'information d'une observation aux états précédant la date de la mesure. En plus d'optimiser l'information tirée des observations, ces algorithmes permettent de fournir des solutions plus lisses dans le temps car l'introduction rétrograde des mesures casse la forme souvent séquentielle de ces méthodes. C'est sur ces algorithmes que les travaux de cette thèse vont s'orienter, et notamment sur le lisseur de Kalman de rang réduit.

L'assimilation de données telle que nous la concevons dans cette thèse est destinée à répondre à l'une des attentes actuelles : elle doit pouvoir *réanalyser* efficacement la trajectoire d'un modèle compte-tenu des séries d'observations disponibles et fournir des états analysés optimaux, cohérent avec la dynamique et de qualité homogène. Ces aspects sont au coeur des préoccupations actuelles de la modélisation océanique, puisque l'un des enjeux actuels est de fournir des simulations (réanalysées) les plus réalistes possibles pour la construction de bases de données climatiques. Cet aspect sera le premier point important de la thèse. Nous nous attarderons à définir dans quelle mesure le lissage permet d'obtenir une solution plus continue et plus efficace, en gardant à l'esprit le cadre de réanalyses. L'une des questions principales est de savoir si le lisseur permet d'améliorer la qualité des états analysés.

Si une représentation 4D des observations est justifiée, elle n'en demeure pas moins complexe à mettre en place. Elle nécessite notamment l'évolution des covariances d'erreur (permettant d'introduire les corrélations nécessaires dans le temps), aspect relativement coûteux numériquement selon la taille du système. Se pose alors la question de l'utilisation d'algorithmes de lissage dans un cadre simplifié (et donc plus opérationnel) où l'évolution des erreurs est par exemple négligée. Le lissage dans un cadre opérationnel sera donc un deuxième point que nous aborderons dans nos travaux.

Il existe différentes formulations du lisseur que l'on retrouvera par exemple dans Anderson et Moore (1979) et Simon (2006). Chacune de ces formulations est associée à une formulation particulière du filtre, mais toutes ont le même objectif : utiliser de façon optimale les observations dans l'espace et le temps. Pour nos travaux, nous avons choisi une approche du lissage présentée par Cohn et al. (1994) et Evensen et van Leeuwen (2000) et utilisée en rang réduit par Cosme et al. (2010) dans une configuration boîte carrée de circulation océanique haute résolution. Cette approche s'est révélée efficace dans une configuration idéalisée, mais nous allons nous attacher à l'implémenter dans un contexte réaliste. Pour ce faire, nous utiliserons un modèle haute résolution (NEMO du code OPA, Madec et al., 1998) dans une configuration régionale de l'Atlantique tropical. Le choix de cette configuration est essentiellement motivé par la dynamique particulière de la région qui peut se révéler intéressante dans la mise en place d'un lisseur. En effet, les phénomènes associés à cette région (ondes d'instabilités et tourbillons) présentent de forts signaux temporels. On les considère donc comme de bons cas d'étude pour l'assimilation 4D. De plus, chacun de ces signaux possède une sensibilité particulière : aux forçages en vent (pour les ondes) ou à la condition initiale (pour les tourbillons). Cela permet d'envisager plusieurs piste pour la mise en place d'expériences. Bien que travaillant en expériences jumelles, nous simulerons des jeux d'observations en accord avec les réseaux d'observations actuels.

Nous nous placerons donc dans une simulation de contexte réaliste. Bien qu'apportant des difficultés supplémentaires, cela permettra de valider nos résultats (dans une certaine mesure) directement pour des problèmes réels.

Nous allons donc dans une **première partie** détailler les éléments constituant le contexte général sur lequel nous effectuerons les expériences d'assimilation. Ces éléments regroupent à la fois la connaissance du milieu océanique concerné (afin de déterminer les enjeux de l'assimilation dans ce milieu), les réseaux réels d'observations disponibles (que nous devrons simuler) et la connaissance de l'outil de modélisation utilisé (afin, entre autres, d'en connaître ses limites).

Nous détaillerons dans une **deuxième partie** la théorie de l'assimilation de données statistique et sa mise en place dans un contexte réaliste. Cela nous permettra de revenir sur les limites de l'assimilation 3D et d'introduire plus précisément les enjeux et les attentes de l'assimilation 4D (c'est-à-dire du lisseur). Divers types de lissage seront présentés succinctement mais nous nous attarderons plus en détail sur la formulation du lisseur que nous avons choisie. Nous détaillerons également les approximations et les caractéristiques liées à notre configuration de l'assimilation. Ces caractéristiques peuvent varier fortement en fonction des configurations, il est donc nécessaire de préciser notre système et ses limites. Par la suite, la **troisième partie** du manuscrit exposera la paramétrisation des covariances d'erreur évolutives nécessaires au lissage, selon deux axes : une simulation de l'erreur modèle par une méthode d'ensemble, et une base d'erreur évolutive sans erreur modèle. Cette paramétrisation est l'un des points sensibles de l'assimilation 4D telle que nous la définissons. Elle occupera donc une place importante dans nos diagnostics préliminaires à la mise en place du lisseur.

Une **quatrième partie** présentera les expériences de lissage mises en place et les résultats que nous avons pu en tirer. Nous nous attarderons notamment sur l'influence du lissage sur les réanalyses (par rapport à un cas de filtrage) et sur la représentation 4D des corrélations. Ces résultats seront étendus ensuite à un cas plus proche des enjeux opérationnels (mais moins optimal sur le plan théorique), où l'évolution des covariances est simplifiée afin de limiter les coûts numériques de calcul.

L'ensemble de ces résultats sera lié aux questionnements principaux de cette thèse : en quoi l'extension 4D des méthodes d'estimation optimale dans un contexte réaliste de circulation océanique peut-elle s'avérer pertinente et efficace, notamment pour des réanalyses ? Et quels sont les principaux atouts ou difficultés de la mise en place de l'algorithme de lissage choisi ?

Une dernière partie permettra de synthétiser les conclusions tirées de nos expériences et d'exposer de nouvelles perspectives suite à ce travail.

Première partie

**L'Atlantique Tropical :
dynamique, observations et
modélisation**

Chapitre 1

La dynamique de l’océan en Atlantique Tropical

Sommaire

1.1	Introduction	10
1.2	Circulation moyenne	10
1.2.1	Caractéristiques liées à la position tropicale	10
1.2.2	Circulation moyenne en Atlantique tropical	11
1.3	La variabilité saisonnière	12
1.3.1	Ondes tropicales d’instabilité (TIWs)	13
1.3.2	Anneaux du Brésil	15
1.4	La variabilité interannuelle	17
1.5	Conclusion	18

1.1 Introduction

Les expériences d'assimilation menées dans cette thèse se font dans le cadre d'une modélisation réaliste de l'océan. Nous devons donc, avant de parler des expériences, avoir une certaine connaissance préalable de la région de l'océan qui nous concerne afin d'en retenir ses principales caractéristiques et d'identifier les points sensibles qui pourront s'avérer pertinents pour l'assimilation (par exemple des phénomènes mal représentés par la modélisation). Dans ce chapitre, nous allons donc exposer les grandes lignes de la dynamique océanique en Atlantique tropical (région définie par le cadre noir sur la figure 1.1). Nous regarderons d'abord la circulation moyenne du bassin, puis nous en préciserons les variabilités saisonnière et interannuelle ainsi que les phénomènes associés. Cela nous permettra également de définir des sous-régions où ces phénomènes sont observés (zone des TIWs, zone des anneaux du Brésil). Ces deux régions, présentées et étudiées par *Ubelmann* (2009), présentent des phénomènes à fortes variabilités temporelles. Ils sont donc intéressants dans le cadre de la mise en place d'une assimilation 4D. Nous les regarderons donc souvent indépendamment dans les diagnostics de nos expériences d'assimilation.

A noter que si la dynamique de l'Atlantique tropical est globalement comparable à celle du Pacifique Tropical, la disposition des continents apporte également quelques éléments spécifiques à l'Atlantique.

1.2 Circulation moyenne

1.2.1 Caractéristiques liées à la position tropicale

Donnons d'abord quelques caractéristiques générales de la circulation océanique tropicale (pour les détails sur ces mécanismes, on pourra se référer à *Bearman* (1989)).

D'une part, la force de Coriolis est nulle à l'équateur et de signe opposé de part et d'autre. L'influence de cette force sur la déviation des masses d'eau va donc rapidement varier selon la latitude.

Il en résulte que la dynamique de surface est fortement contrôlée par les vents. Cela implique également que la variabilité océanique en surface va suivre la variabilité atmosphérique, ce que l'on peut notamment observer dans le signal saisonnier (section 1.3).

D'autre part, il faut noter une certaine dissymétrie nord-sud par rapport à l'équateur. Cela est lié à la position de l'ITCZ (zone de convergence intertropicale), qui représente en quelque sorte l'équateur atmosphérique. A cause de l'influence du continent africain, l'ITCZ est légèrement décalée vers le Nord, notamment à l'Est du bassin (Golfe de Guinée). La convergence des vents s'effectue donc à quelques degrés au nord de l'équateur géographique et la circulation océanique en est tout autant décalée.

En gardant à l'esprit ces trois points, nous pouvons maintenant regarder la circulation plus en détail.

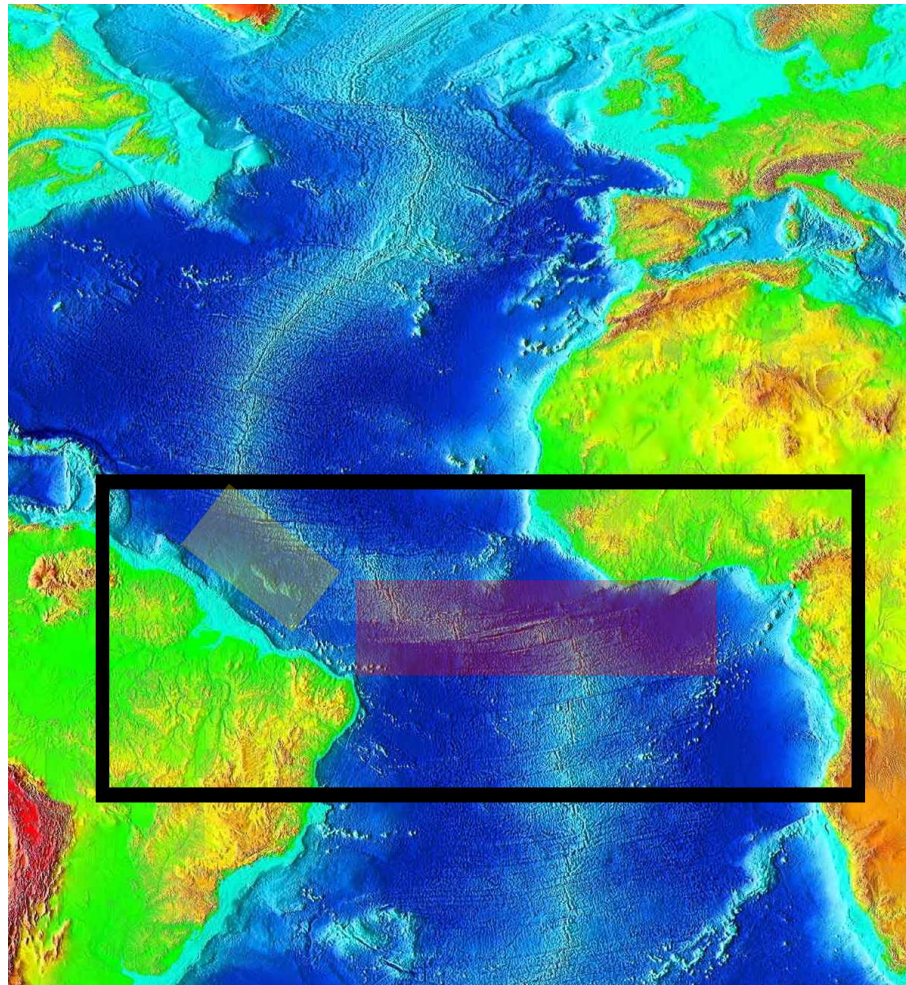


FIGURE 1.1 – Représentation de la bathymétrie dans l’océan Atlantique. Le rectangle noir indique la zone définie comme étant l’Atlantique tropical. Les rectangles rouge et jaune indiquent respectivement les zones définies comme la région des TIWs et la région des anneaux du Brésil.

1.2.2 Circulation moyenne en Atlantique tropical

En regardant la figure 1.2 tirée de *Stramma et Schott (1999)* on constate que la dynamique de surface est essentiellement zonale. Les Alizés (vents d’est) génèrent deux principaux courants de part et d’autre de l’ITCZ : le NEC (North Equatorial Current) et le SEC (South Equatorial Current). Ce dernier est décomposé en plusieurs branches de part et d’autre de l’équateur géographique : le nSEC (northern), le eSEC (equatorial), le cSEC (central) et le sSEC (southern). Le SEC est le courant dominant de la zone équatoriale (le NEC étant situé vers 10°N). Il transporte les eaux de surface d’est en ouest, accumulant les eaux à l’ouest du bassin, et maintenant ainsi un gradient permanent de SSH est-ouest sur le bassin. Cette accumulation est en partie compensée par un retour vers l’est des eaux avec l’EUC (Equatorial Under Current), situé entre 50 et 100m de profondeur au niveau de l’équateur géographique. Il existe également d’autres courants de retour, le NECC (North Equatorial Counter-Current) et le SECC (South Equatorial Counter-Current) situés à des latitudes un peu plus élevées. Le sSEC et l’eSEC alimentent également un courant

le long de la côte nord brésilienne, le NBC (North Brazil Current). Ce courant peut se retroflecter pour alimenter l'EUC et, selon la saison, pour alimenter le NECC entre 6 et 8°N (section 1.3). Enfin, on notera la présence de deux cellules cycloniques (ou domes) le long de la côte africaine : le GD (Guinea Dome) et l'AD (Angola Dome), lui-même situé à l'intérieur du gyre d'Angola (AG).

Bien que faible à l'équateur, la force de Coriolis n'en est pas moins suffisante pour induire des transports d'Ekman couplés aux différents courants (transport à droite de la direction principale du courant dans l'hémisphère nord et à gauche dans l'hémisphère sud). Ainsi, entre le nSEC et le sSEC, les transports d'Ekman de directions opposées écartent les masses d'eau de part et d'autre de l'équateur, créant la divergence équatoriale. Cette divergence entraîne à son tour une remontée des eaux de sub-surface (upwelling équatorial). Cet upwelling est surtout visible à l'est du bassin, car couplé avec des upwelling côtiers dus aux divergences de masses d'eau le long de la côte africaine. D'autres upwellings (parfois saisonniers) sont également observés à l'intérieur des domes, au large de Dakar où une branche du NEC longe la côte au printemps, et entre la Côte d'Ivoire et le Bénin où le NECC longe la côte d'ouest en est.

Dans les eaux intermédiaires (de 100 à 1200m), on retrouve le même schéma de circulation que nous ne redétaillerons pas ici. On pourra se référer aux figures 5 et 6 de *Stramma et Schott* (1999) pour une vision détaillée de cette circulation. On notera juste que certaines branches de courant de retour font leur apparition, comme par exemple le SEUC (South Equatorial Under Current) situé au sud de l'eSEC, ou le NEUC (North Equatorial Under Current) situé au nord du nSEC.

Enfin, il ne faut pas oublier la circulation thermohaline profonde et notamment le sous-courant de bord ouest (DWBC, Deep Water Boundary Current) situé vers 1500m de profondeur. Ce courant, qui longe la côte brésilienne du nord vers le sud, est lié à la plongée des eaux froides au niveau du gyre sub-polaire dans les hautes latitudes de l'Atlantique Nord.

Nous ne détaillerons pas plus cette circulation, mais le lecteur pourra se référer par exemple à *Gouriou et al.* (2001).

1.3 La variabilité saisonnière

Comme nous l'avons précisé auparavant, la variabilité de l'Atlantique tropical est fortement contrôlée par la variabilité atmosphérique. Plusieurs études ont montré que le cycle annuel pouvant être observé sur la thermocline est contrôlé par les vents à l'échelle du bassin (*Busalacchi et Picaut*, 1983 ; *Du Penhoat et Tréguier*, 1985 ; *Philander et Pacanowski*, 1986). En effet, l'intensification des Alizés à la fin du printemps va accentuer la divergence équatoriale et donc intensifier l'upwelling. La profondeur de la thermocline s'en trouvera alors tout autant amincie, notamment dans la partie est (*Houghton*, 1989), et une langue d'eau froide assez marquée se forme de l'est jusqu'au centre du bassin. De façon plus générale, les upwellings ont tendance à être amplifiés durant l'été et l'automne. On notera également que le NECC (figure 1.2) possède une étendue zonale beaucoup plus faible en hiver et au printemps.

L'intensification des courants tropicaux va également donner naissance à un signal typique de la variabilité en Atlantique tropical : les ondes d'instabilité tropicales (TIWs,

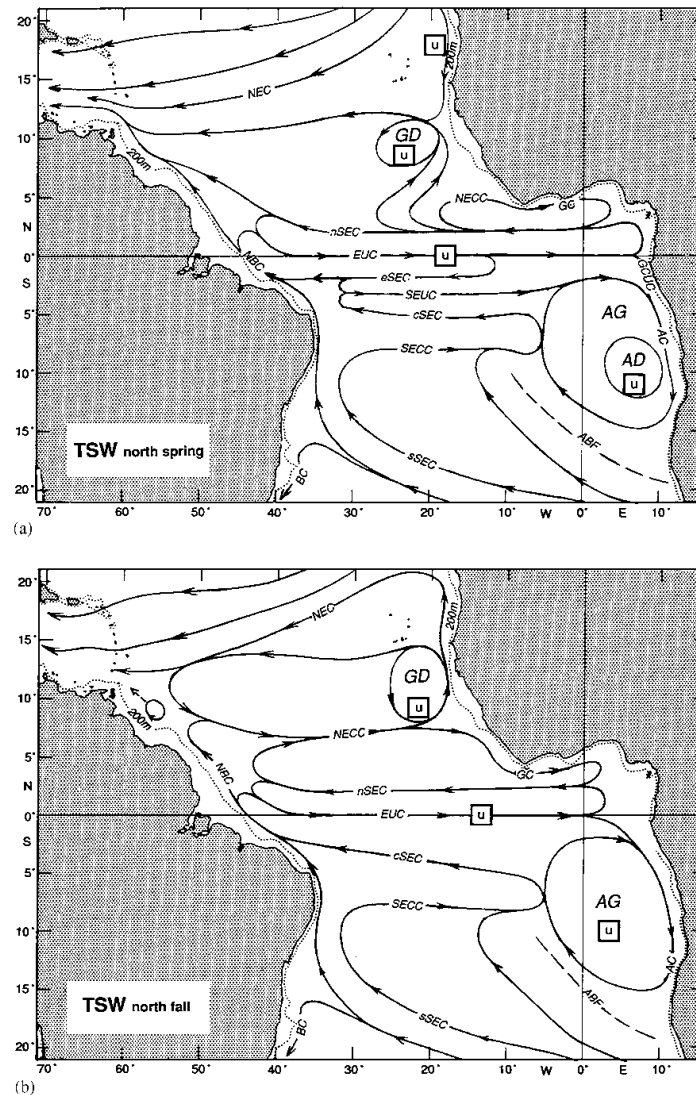


FIGURE 1.2 – Figure d'après *Stramma et Schott (1999)*. Circulation moyenne dans les 100 premiers mètres, au printemps (a) et à l'automne (b).

Tropical Instability Waves) qui se développent au centre du bassin. On observe également à la même période sur le bord ouest du bassin la formation des anneaux du Brésil. Nous allons donner un peu plus de détails sur ces deux phénomènes car ils seront souvent mis en avant dans les expériences d'assimilation.

1.3.1 Ondes tropicales d'instabilité (TIWs)

L'amplification des courants zonaux dans la bande tropicale augmente le cisaillement entre les différents courants du bassin. Ce cisaillement peut être horizontal (entre le SEC et le NECC ou entre l'EUC et le SEC en surface) et créer des instabilités barotropes, ou vertical (entre le SEC et l'EUC en profondeur) et créer des instabilités baroclines. Cela va conduire à la formation des ondes tropicales d'instabilité (TIWs). Ces ondes ont

d'abord été étudiées par *Legeckis* (1977) d'après les données radiométriques d'un satellite géostationnaire. Elles ont ensuite fait l'objet de nombreuses études d'après des données infrarouges satellitaires (*Allen et al.*, 1995), des données in-situ (*Halpern et al.*, 1988; *Hayes et al.*, 1989) ou des données issues de modèles numériques (*Philander et al.*, 1986; *Masina et Philander*, 1999). Les mécanismes exacts de la formation des ondes sont encore largement discutés. Si certains suggèrent un effet prédominant du mode barotrope (*Jochum et al.*, 2004), d'autres au contraire mettent en avant le mode barocline (*Hansen et Paul*, 1984; *Luther et Johnson*, 1990; *Baturin et Niiler*, 1997), et d'autres encore pointent une combinaison des deux modes (*Masina et al.*, 1999; *Masina*, 2002). Nous ne nous attarderons pas ici sur les différents arguments de formation des ondes. Nous allons simplement décrire leurs caractéristiques observables principales.

Les TIWs naissent au début de l'été, lorsque les courants ont été suffisamment amplifiés par les Alizés. Elles se développent à l'est du bassin et se propagent ensuite vers l'ouest de façon quasi zonale. Elles ont une longueur d'onde de l'ordre de 800 km, et une vitesse moyenne de 0.4 à 0.5 m.s^{-1} . Leur signature instantanée est nettement visible sur la température de surface (SST) ou la couleur de l'eau entre 2 et 5°N où les eaux froides de la bande équatoriale sont advectées vers les eaux chaudes plus au nord, et vice-versa. La figure 1.3, extraite de *Jochum et al.* (2004), est une image satellite de la couleur de l'eau où l'on voit la trace des TIWs. De même, leur propagation temporelle apparaît clairement sur des diagrammes de Hovmöller tracés autour de 3°N d'après des observations satellites (figure 1.4). Les TIWs ont également une signature topographique pouvant faire varier la SSH de 5 à 10 cm selon les ondes. Cette signature est suffisante pour être détectée par les satellites ou sur des diagrammes de Hovmöller.

Ces ondes sont observées de la même façon en Atlantique et dans le Pacifique. Néanmoins, l'Atlantique étant moins étendu zonalement, elles ont moins le temps de se propager et leur signal se perd rapidement à l'approche de la côte brésilienne. Enfin, on peut noter que les ondes d'instabilité existent aussi au sud de l'équateur, mais sont souvent moins marquées.

Le rôle des TIWs dans le bilan énergétique de la zone tropicale a également fait l'objet de nombreuses études. Initialement, l'advection horizontale de chaleur par les instabilités était estimée à 100 W.m^{-2} (par exemple *Hansen et Paul*, 1984; *Bryden et Brady*, 1989; *Weingartner et Weisberg*, 1991; *Swenson et Hansen*, 1999; *Wang et McPhaden*, 1999), ce qui est comparable aux variations saisonnières des flux de chaleur en surface. D'autres (*Vialard et al.*, 2001; *Jochum et al.*, 2004) ont montré que le réchauffement induit par l'advection horizontale était en partie compensé par un refroidissement de la couche de mélange lui-même dû aux TIWs. Dans leur étude, *Grodsky et al.* (2005) estiment à 10 W.m^{-2} le réchauffement de la couche de mélange dû aux TIWs, ce qui est relativement faible par rapport aux 100 W.m^{-2} initialement estimés. Enfin, d'autres études montrent également l'impact que peut avoir ces ondes sur l'atmosphère, et notamment sur la position de l'ITCZ (*Caltabiano et al.*, 2005).

Tous ces éléments de discussion indiquent que les TIWs sont un élément important de la dynamique tropicale. Dans le cadre de cette thèse nous utiliserons les TIWs comme un objet d'étude pertinent à caractériser d'un point de vue spatio-temporel pour l'assimilation de données. En effet, le fait qu'elles constituent un phénomène propagatif en plus du fait qu'elles possèdent une signature claire sur la température peut s'avérer pertinent dans le cadre d'assimilation 4D où la dimension temporelle est importante. D'une façon plus générale, la méconnaissance qui demeure encore sur la formation de ces ondes (et sur

les forçages en vent contrôlant ces ondes) justifie la mise en place d'une assimilation de données pour accorder le mieux possible la représentation de ces ondes dans les modèles avec les observations disponibles.

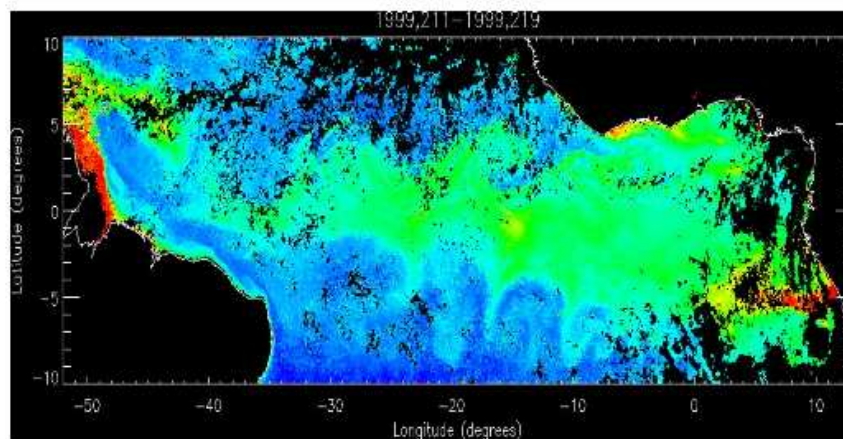


FIGURE 1.3 – Figures d'après *Jochum et al.* (2004). Image de la couleur de l'eau prise par le capteur SeaWiFS où l'on peut distinguer le passage des TIWs (par le contraste entre les zones vertes et bleues).

1.3.2 Anneaux du Brésil

Comme nous l'avons déjà évoqué dans la section 1.2, le NBC, qui longe la côte brésilienne entre 0 et 500m de profondeur, se réfléchit à deux niveaux : près de l'équateur pour alimenter l'EUC entre 100 et 200 m de profondeur, et entre 6 et 8°N pour alimenter le NECC en surface (figure 1.2). Cette deuxième réflexion n'a lieu qu'entre l'été et le début de l'hiver, lorsque le NBC est intensifié (*Richardson et Walsh*, 1986 ; *Garzoli et Katz*, 1993). La formation d'anneaux aux différents niveaux de réflexion a été détectée à la fin des années 80, à partir d'imagerie de la couleur de l'eau et de données de courantomètres au large des côtes brésiennes (*Johns et al.*, 1990 ; *Richardson et al.*, 1994). Leur signature altimétrique a ensuite pu être observée dans les années 90 (*Diden et Schott*, 1993 ; *Goni et Johns*, 2001). Ces anneaux, ayant un diamètre de l'ordre de 450 km, se propagent vers le nord-ouest en longeant la côte à une vitesse variant de 0.08 à 0.16 m.s⁻¹. Ils finissent par atteindre les petites Antilles et peuvent alors se désintégrer en se heurtant à la topographie complexe de la région. Les vitesses associées aux anneaux peuvent parfois atteindre 1 m.s⁻¹ en périphérie. Ils sont associés à une anomalie positive de SSH en leur centre, pouvant varier de quelques centimètres jusqu'à 15 ou 20 cm pour les plus gros. On peut également retrouver une signature sur la température ou la couleur de l'eau comme le montre l'image 1.5-b tirée de *Fratantoni et Glickson* (2002).

Différents types d'anneaux ont été mis en évidence par *Fratantoni et al.* (1995) puis plus précisément par *Garraffo et al.* (2003). Ces derniers distinguent les anneaux peu profonds (ou de surface), intermédiaires, profonds et de subsurface. Les trois premiers s'étendent de la surface jusqu'à des profondeurs respectives de 200, 500 et 900-1000 mètres. Les anneaux de subsurface n'ont quant à eux pas de signature en surface. Si les anneaux peu profonds

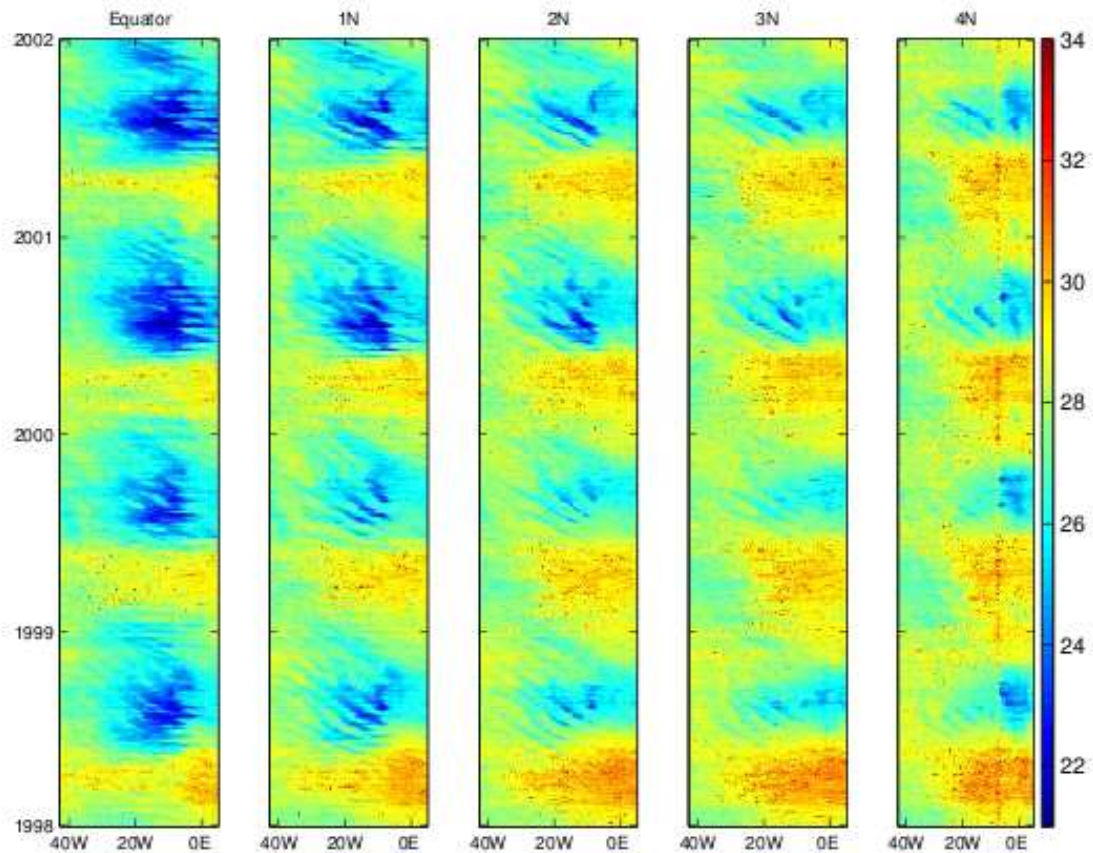


FIGURE 1.4 – Figures d'après *Caltabiano et al.* (2005). Diagrammes de Hovmöller de la température de surface issue des données du satellite TMI (Tropical Rainfall Measuring Mission Microwaves Imager), entre janvier 1998 et décembre 2001, tracés (de gauche à droite) : à l'équateur, à 1°N, à 2°N, à 3°N, à 4°N et à 5°N.

sont créés au niveau de la réflexion du NBC en NECC, les autres types d'anneaux ont plutôt tendance à être formés au niveau de la réflexion du NBC en EUC. On compte en moyenne de 6 à 8 anneaux formés chaque année. La figure 1.5-a extraite de *Fratantoni et Glickson* (2002) schématise la formation des anneaux.

Tout comme les TIWs, les anneaux du Brésil s'avèreront un objet d'étude intéressant pour l'assimilation de données et notamment la mise en place d'une méthode de type lisseur puisqu'il s'agit là aussi d'un signal propagatif temporellement. Leur formation est fortement corrélée au courant amont à la zone (NBC), une faible perturbation du modèle pourrait avoir un impact significatif sur les anneaux. Mais leur signal est aussi moins clair sur la température que le signal des TIWs, il s'agit donc d'un défi à relever pour l'assimilation de données.

La représentation de la dynamique par le modèle sera précisée dans le chapitre 3. Les deux zones à caractère particulier sont rappelées sur la figure 1.1. Ces zones feront souvent l'objet d'un diagnostic particulier dans le cadre des expériences d'assimilation, en plus des diagnostics globaux.

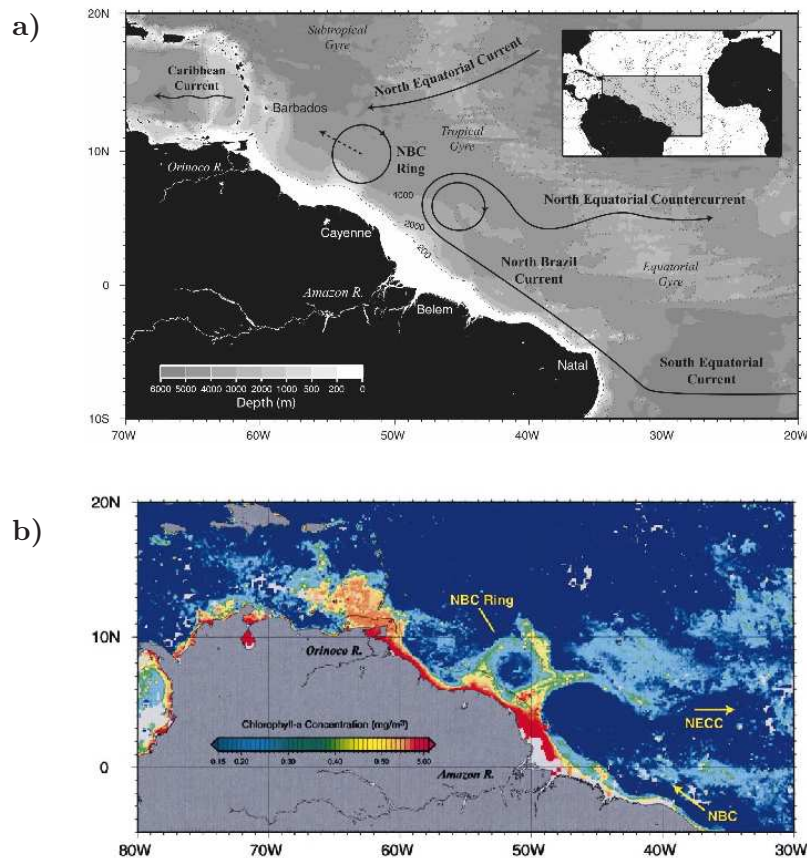


FIGURE 1.5 – Figures d’après *Fratantoni et Glickson* (2002). a) Schématisation des principaux courants et de la formation des anneaux du Brésil. b) Image de la chlorophylle prise par le satellite SeaWiFS où l’on peut distinguer un anneau.

1.4 La variabilité interannuelle

Si la variabilité de l’Atlantique tropical est essentiellement dominée par son cycle saisonnier, il existe également une variabilité interannuelle liée à des phénomènes d’échelle planétaire comme ENSO (El Niño and Southern Oscillation). Nous allons donner brièvement quelques éléments de cette variabilité (voir par exemple *Xie et Carton* (2004) pour plus de détails sur cette variabilité).

On peut observer parfois une anomalie positive de température lors de l’été boréal. Cette anomalie peut atteindre des pointes de 1°C, et est généralement située au niveau de la langue d’eau froide créée par l’upwelling équatorial, entre 6°S et 2°N et entre 20°W et 5°E. La fréquence de ces événements est d’environ 30 mois. La première étude détaillée de ces événements remonte à 1963 car l’évènement coïncida avec le programme d’observation EQUALANT (*Merle*, 1980). La diminution des vents et de la convection observée en conjoncture à cet événement a amené *Merle* (1980) et *Hisard* (1980) à utiliser le terme “Atlantic Niño”. Des études postérieures ont ensuite indiqué que ce réchauffement coïncidait également avec un approfondissement de la thermocline à l’est du bassin (*Philander*, 1986).

En plus de ce mode propre à l’Atlantique tropical, il existe également une variabilité

forcée par des phénomènes externes. Il a été observé que El Niño par exemple, qui se produit dans le Pacifique, provoque (avec quelques mois de décalage) un réchauffement anormal de la SST à l'échelle du bassin dans l'Atlantique tropical, de même que les précipitations (et donc l'apport d'eaux douces) ont tendance à diminuer lors de cet événement (*Covey et Hastenrath, 1978; Aceituno, 1988; Giannini et al., 2000; Mestas-Nunez et Enfield, 2001; Alexander et Scott, 2002*). A cela il faut aussi ajouter de fortes interactions et rétroactions avec l'atmosphère et l'ITCZ qui ont fait et font encore l'objet de nombreuses études (*Chang et al., 2000; Sutton et al., 2000; Okumura et al., 2001; Terray et Cassou, 2002*).

La variabilité interannuelle ne sera pas plus détaillée ici. On retiendra que même si les modes saisonniers dominent les variations en Atlantique tropical, des anomalies dues à des variations interannuelles peuvent aussi s'y superposer.

1.5 Conclusion

Nous avons pu voir dans ce chapitre les grandes caractéristiques de la circulation en Atlantique tropical. Cette circulation est marquée d'une part par une orientation très zonale des courants, et d'autre part par un cycle saisonnier important. Des traces d'une variabilité interannuelle de grande échelle peuvent également se superposer au cycle saisonnier sans pour autant présenter un signal dominant.

La variabilité saisonnière est associée à deux phénomènes particuliers (durant l'été et l'automne) : les ondes d'instabilité tropicales (TIWs) et les anneaux du Brésil. Ces deux phénomènes sont propagatifs et présentent un signal fort sur la température (pour les TIWs notamment), la SSH et les vitesses (pour les anneaux du Brésil). Cela en fait des objets d'étude pertinents dans le cadre de la mise en place de l'assimilation 4D car les corrélations temporelles sont *a priori* autant importantes que les corrélations spatiales.

Chapitre 2

L'observation de l'océan

Sommaire

2.1	Introduction	20
2.2	Bref historique	20
2.3	Les relevés <i>in situ</i>	22
2.3.1	Les mouillages	22
2.3.2	Les bouées dérivantes	23
2.3.3	Autres instruments de mesures	24
2.3.4	Les campagnes de mesures océanographiques	25
2.4	Les données satellites	25
2.4.1	La mesure altimétrique	26
2.4.2	Les autres types d'observations	27
2.5	Les radars	28
2.6	Conclusion	28

2.1 Introduction

Dans le chapitre précédent nous avons résumé de façon générale les caractéristiques connues de la dynamique en Atlantique tropical. Ces connaissances sont issues à la fois des recherches théoriques en dynamique des fluides, des modèles numériques, mais aussi et surtout des diverses observations qui se sont accumulées au fil des décennies. Nous allons donc préciser dans ce chapitre avec quels moyens l'océan peut-être observé. Par la suite, nous n'utiliserons pas de vraies observations dans le cadre de cette thèse (nous simulerons simplement des jeux d'observations synthétiques). Il paraît néanmoins nécessaire de prendre connaissance des moyens actuels (et passés) qui permettent d'avoir une description fiable de l'océan, puisque l'assimilation de données a justement pour rôle d'utiliser ces observations pour compléter ou améliorer la représentation des modèles numériques. Nous tenterons d'ailleurs dans nos expériences de simuler des réseaux d'observations proches de ceux existant réellement afin que cette étude puisse se rapprocher au mieux des problèmes réels.

2.2 Bref historique

Contrairement à l'atmosphère, les observations et surtout les mesures systématiques de l'océan ne sont pas très anciennes. L'observation a débuté naturellement avec les expéditions maritimes destinées à explorer les mers et à cartographier de nouveaux continents. Une première carte du Gulf Stream a par exemple été tracée en 1770 par B. Franklin. En 1851, la pose du premier câble télégraphique sous-marin entre la France et la Grande Bretagne a fait prendre conscience de la nécessité de mieux connaître les fonds océaniques, tant sur la topographie que sur la dynamique des courants. Mais c'est entre 1872 et 1876 que débutent vraiment les campagnes de mesures avec l'expédition britannique du navire Challenger. Néanmoins, ces expéditions restent peu fréquentes et surtout très coûteuses. Il faudra attendre les années 60 et le développement des mouillages hauturiers autonomes pour avoir accès à une réelle mesure systématique de l'océan.

A partir de ce moment, des réseaux d'observations vont être mis en place, avec notamment le déploiement dans les années 1980-1990 dans le Pacifique équatorial des mouillages autonomes ATLAS composant le réseau TAO/TRITON¹ du programme international TOGA : Tropical Ocean Global Atmosphere (*McPhaden, 1993; McPhaden et al., 1998*). La mise en place de ce réseau a été motivée en grande partie par un évènement El Niño d'une intensité extrême en 1982-1983. Le désastre humain et économique qu'engendra cet évènement et le fait que personne n'ait pu le prédire souligna l'importance d'avoir accès à des données en temps réel, au même titre que la météorologie. Il aura fallu attendre près de 10 ans pour que le réseau TAO soit opérationnel. Le réseau PIRATA² a ensuite été développé en Atlantique tropical à partir de 1997 pour étudier les interactions océan-atmosphère (*Servain et al., 1998*). Il est issu d'une collaboration entre la France, le Brésil et les USA, dans le cadre du programme CLIVAR³ (CLImate VARIability and predictability). Plus récemment, dans les années 2000, le réseau RAMA⁴ a été implanté dans l'Océan Indien dans le cadre de l'étude de la mousson indienne.

1. <http://www.pmel.noaa.gov/tao/>
2. <http://www.ifremer.fr/ird/pirata/>
3. <http://www.clivar.org>
4. <http://www.pmel.noaa.gov/tao/rama/>

Ces réseaux fixes ont également été complétés par la mise en place du réseau de flotteurs profileurs ARGO⁵ (*Argo Steering Team*, 1998), dont le déploiement a débuté en 2000 dans le cadre des programmes internationaux GCOS⁶ (Global Climate Observing System) et GOOS⁷ (Global Ocean Observing System). Ce réseau a permis d'augmenter considérablement le nombre de données *in-situ* en une décennie, comme le montre la figure 2.1 tirée de *Juza* (2011), avec la capacité de faire des mesures jusqu'à 2000 mètres.

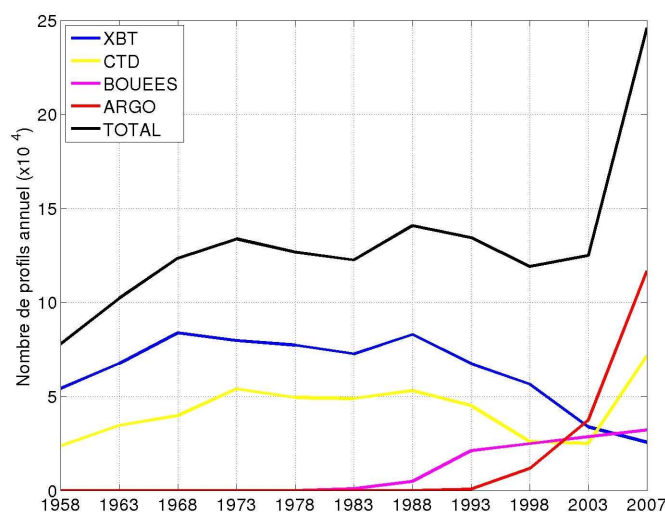


FIGURE 2.1 – Figures d'après *Juza* (2011). Evolution du nombre de profils *in situ* annuels disponibles pour différents types d'observations entre 1958 et 2007.

Si les relevés *in situ* ont apporté (et apportent toujours) une connaissance indispensable à la compréhension du système océanique, ils n'en demeurent pas moins très ponctuels et insuffisants pour décrire toute la dynamique. C'est l'arrivée des observations satellites qui va permettre de compenser cet aspect et, par la même occasion, de vraiment révolutionner les sciences océanographiques, avec notamment la mesure altimétrique. Les premières mesures de ce type furent réalisées au début des années 1970 depuis la station Skylab, puis avec la mission GEOS-3 en 1974 et Seasat en 1978. Néanmoins, le bruit sur les mesures et l'imprécision de la trajectoire des satellites rendaient ces données inexploitable. C'est l'arrivée du satellite Geosat en 1985 qui va fournir pour une durée de 3 ans la première base de données utilisables (*Douglas et Cheney*, 1990). Par la suite, divers missions vont voir le jour, certaines destinées à l'observation de la Terre et des océans (comme ERS⁸ ou Envisat⁹), et d'autres plus spécifiquement dédiées aux océans : Topex/Poséidon¹⁰ et plus récemment Jason¹² (2001) qui a pris la succession de Topex/Poséidon. Le lancement de ces

5. <http://www.argo.net>

6. <http://www.wmo.ch/web/gcos/gcoshome.html>

7. <http://www.ioc-goos.org>

8. <http://earth.esa.int/ers/>

9. <http://envisat.esa.int/>

10. <http://www.aviso.oceanobs.com/fr/missions/missions-passees/topexposeidon/index.html>

12. <http://smc.cnes.fr/JASON/Fr/>

missions a fait exploser le nombre de données disponibles. En plus de la hauteur du niveau de l'eau initialement mesurée, les satellites actuels fournissent également des données sur la température de surface, la couleur de l'eau, la salinité de surface et le géoïde. Si la révolution spatiale a existé, elle a aussi entraîné la révolution océanographique car les satellites ont permis d'avoir accès rapidement à des champs de surface (hauteur de l'eau ou température par exemple) quasi-complets. Aujourd'hui encore, de nombreuses missions sont en préparation, avec comme principal enjeu d'améliorer la précision des mesures.

Désormais, les observations de l'océan sont donc devenues quotidiennes. Elles donnent accès à des données en profondeur (profileurs) et à des données de surface (satellites). Bien qu'insuffisantes à elles-seules pour décrire l'état 3D complet de l'océan, ces observations sont un point essentiel à la compréhension de la dynamique et servent aussi à la validation des modèles numériques. Nous allons donner ci-dessous quelques caractéristiques principales de chaque type d'observation afin de pouvoir simuler des jeux de données pour la suite de nos expériences.

2.3 Les relevés *in situ*

Toute mesure directement effectuée dans l'océan peut être définie comme une donnée *in situ*. Ces mesures peuvent être de natures très diverses (température, salinité, conductivité, pression, courants...) et ont des domaines d'application tout autant variés (météorologie, océanographie, climatologie...). Il existe différents types de réseaux de mesures que l'on peut classer en trois catégories : les réseaux fixes (mouillages), les réseaux dérivants et les campagnes de mesure ponctuelles. Nous allons détailler chacun de ces réseaux dans les sections suivantes.

2.3.1 Les mouillages

Les réseaux de bouées fixes (ou mouillages) sont généralement mis en place dans le cadre de programmes visant à étudier la dynamique d'une région particulière. Ils permettent d'obtenir à des points fixes des séries temporelles de mesures, qui se traduisent généralement sous forme de profils verticaux de température, de salinité et parfois de vitesse. Parmi ces réseaux, on trouve notamment les mouillages ATLAS et TRITON composant les réseaux TAO/TRITON, PIRATA et RAMA (figure 2.2) destinés à étudier la bande équatoriale. Ces mouillages permettent d'obtenir des profils de température et de salinité entre la surface et 500m de profondeur (tous les 25m entre la surface et 150m, puis tous les 50m jusqu'à 300m, puis un dernier relevé à 500m), complétés par des relevés météorologiques de surface (vent, humidité relative, température de l'air, pluviométrie). Les relevés se font selon différentes fréquences en fonction des variables (par exemple toutes les 10 minutes pour la température) et sont ensuite moyennés sur 24 heures. Ces données sont transmises par satellite et accessibles en temps quasi réel sur internet.

En Atlantique Tropical, le réseau PIRATA permet d'accéder à des données continues de subsurface et donc à des séries temporelles riches en information. Le réseau est en revanche peu dense (une vingtaine de bouées réparties sur toute la bande tropicale de l'Atlantique). Il permet donc de capter des signaux localisés mais de haute fréquence (comme le passage d'une onde au niveau d'une bouée par exemple). Ce réseau étant placé à des points fixes statégiques, il permet d'étudier précisément la propagation des TIWs

et la variabilité de surface/subsurface autour des bouées. Ce réseau étant fixe et de haute fréquence temporelle, il ne convient pas pour l'étude de l'impact du lissage en assimilation de données. En effet, le lisseur ne permettrait pas dans ce cas d'améliorer la couverture spatiale (car les bouées sont fixes) ou temporelle (car séries de mesures déjà de haute fréquence). Nous ne tiendrons donc pas compte de ce réseau dans la simulation de nos observations.

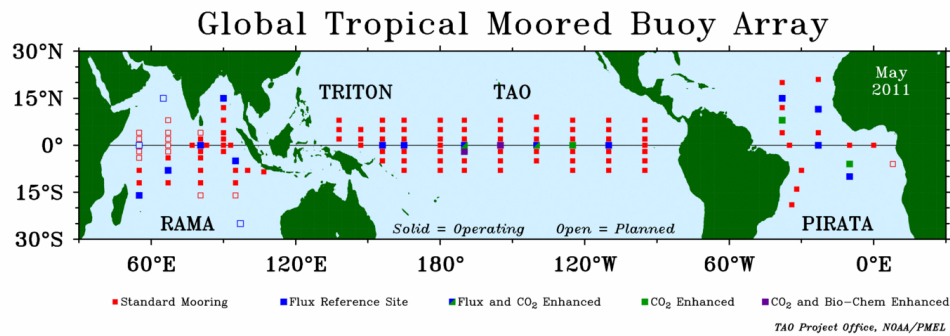


FIGURE 2.2 – Figure tirée du site <http://www.pmel.noaa.gov/tao/global/global.html>. Représentation des différents réseaux de mouillages dans la bande tropicale.

2.3.2 Les bouées dérivantes

Contrairement aux mouillages, les bouées et flotteurs dérivants ne sont pas localisés en un point fixe. Ils dérivent au gré des courants et permettent ainsi de suivre les déplacements des masses d'eau. Il existe plusieurs types de flotteurs dérivants : les flotteurs de surface, les flotteurs de subsurface (qui se stabilisent à une profondeur donnée par flottaison) et les flotteurs profileurs. Le réseau ARGO par exemple est constitué de profileurs. Ils sont actuellement de 3 types : les PROVOR (construits par KANNAD¹³ en France en collaboration avec l'IFREMER¹⁴), les APEX (fournis par la Webb Research Corporation¹⁵ aux Etats-Unis) et les SOLO (fournis par le Scripps Institution of Oceanography¹⁶ aux Etats-Unis). Ces profileurs sont situés à une profondeur dite "profondeur de parking" et dérivent avec les courants pendant une période préprogrammée. Ils remontent ensuite vers la surface en effectuant des relevés de température et de salinité en fonction de la pression (et donc de la profondeur) et transmettent ces données ainsi que leur position par le système ARGOS. Ils replongent ensuite à leur profondeur de parking et se remettent à dériver. Ces profileurs permettent d'avoir des mesures jusqu'à 2000m avec une fréquence de 10 jours. Le réseau est actuellement composé de plus de 3000 appareils, dont la durée de vie est de l'ordre de 4-5 ans, répartis sur l'ensemble des océans. La densité spatiale est de l'ordre de 3° en moyenne (un exemple de la répartition des flotteurs est donné sur la figure 2.3). Les données sont accessibles en temps quasi réel grâce au GDACs (Global Data Assembly Centers) après un contrôle de qualité automatique. Des contrôles de qualité d'une forme plus scientifique ont également lieu mais le délai d'accès aux données est alors de 6 mois en moyenne.

13. <http://www.kannad.com>

14. <http://www.ifremer.fr>

15. <http://www.webbresearch.com>

16. scripps.ucsd.edu

Les bouées dérivantes complètent bien les réseaux fixes de mouillages, mais la trajectoire des flotteurs est difficile à prévoir et on peut parfois noter des différences de densité dans leur répartition (avec notamment des densités plus élevées dans les zones où les courants convergent). Enfin, notons que ARGO n'est pas le seul réseau de flotteurs dérivants, avec par exemple le réseau MEDARGO plus spécifique à la mer Méditerranée.

Dans le cas de l'Atlantique Tropical, le réseau ARGO permet d'obtenir une couverture plus large des relevés *in situ* mais avec une fréquence plus faible (10 jours). Ce réseau sera donc plus apte à capter des phénomènes sur tout le bassin mais ne permet *a priori* pas d'étudier directement la variabilité haute-fréquence (inférieure à 10 jours). En combinant les informations de différentes bouées proches géographiquement (qui ne fournissent pas toutes leurs données en même temps), on peut espérer accéder à une variabilité de fréquence plus élevée. Ce réseau est donc utile pour accéder à une vision *in situ* plus étendue géographiquement que le réseau PIRATA. C'est essentiellement sur ce type de réseau que nous construirons les jeux d'observations simulées dans le cadre de cette thèse car elles permettent de couvrir tout le domaine. De plus, dans le cadre d'un réseau de faible fréquence temporelle, l'ajout d'un lisseur pour l'assimilation de données peut avoir un impact significatif pour augmenter la fréquence des observations et leur couverture spatiale (en tirant des informations de différentes bouées à des dates différentes). On espère ainsi pouvoir améliorer l'information issue de ce type de réseau d'observation.

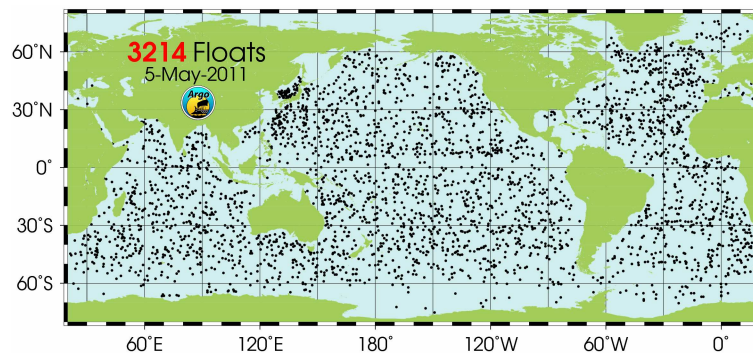


FIGURE 2.3 – Figure tirée du site <http://www.argo.net>. Représentation de la répartition des flotteurs du réseau ARGO qui ont émis un signal dans les 30 derniers jours au 5 mai 2011.

2.3.3 Autres instruments de mesures

En plus des mouillages et des profileurs, de nouveaux types d'instruments sont de plus en plus utilisés pour les mesures *in situ* : les AUVs (Autonomous Underwater Vehicle). Il en existe de différents types (Gliders, Véhicules de surface autonomes, AUVs motorisés). Ils permettent d'effectuer des mesures de la même manière que les profileurs (par une succession de plongées et de retour en surface pour transmettre les données) avec néanmoins la possibilité de contrôler leurs trajectoires (ce qui permet d'éviter les dérives dans les zones où le courant est trop fort). Mais ces systèmes restent encore marginaux du fait de la rareté des appareils, ils ne seront pas pris en compte dans les observations que nous allons simuler par la suite.

2.3.4 Les campagnes de mesures océanographiques

Indépendamment des réseaux permanents, les campagnes de mesures permettent ponctuellement de faire des relevés selon une route définie. Ces campagnes se font à partir de navires spécialement dédiés (navires scientifiques) ou plus simplement à partir de navires marchands dans le cadre du programme VOS (Voluntary Observing Ships). Ce deuxième cas est généralement moins coûteux mais le tracé de la route est alors imposé par les besoins commerciaux.

Les relevés se font à travers le largage de sondes ou capteurs (reliées au bateau par un câble électrique et mécanique) qui vont mesurer les paramètres hydrographiques de l'eau de la surface jusqu'à une profondeur donnée (en fonction de la longueur maximale des câbles...). On distingue par exemple les CTD (Conductivité, Température, Densité) ou encore les XBT (eXpandable BathyThermograph). Les mesures concernent généralement les grandeurs thermodynamiques (température et salinité) ou les champs de vitesse, mais permettent également d'obtenir des grandeurs physico-chimiques des masses d'eau grâce au prélèvement d'échantillons qui est effectué tout au long de la remontée des appareils (XBT).

Ces campagnes restent relativement coûteuses et sont donc très ponctuelles. Par exemple, les campagnes PIRATA-FR qui se déroulent dans le Golfe de Guinée n'ont lieu qu'une fois par an. Elles ne permettent donc pas d'acquérir des séries temporelles sur de longues périodes. En revanche, elles fournissent des relevés quasi-continus le long de leur route ce qui permet de reconstituer des sections de l'océan assez précises.

Nous ne simulerons pas ce type d'observation dans le cadre de cette thèse du fait de leur ponctualité temporelle et spatiale. Nous privilégierons des observations plus étendues et constantes.

2.4 Les données satellites

Les mesures *in situ* présentent le gros désavantage d'être très localisées. Ainsi, même si elles fournissent des données haute fréquence pour certaines, leur représentativité reste très limitée spatialement. Les données satellitaires permettent de combler ce défaut (au moins en surface) en fournissant des données couvrant de larges zones quotidiennement. Ces mesures, disponibles depuis les années 70, ont permis aux océanographes de cartographier de manière précise la topographie des océans et d'en observer les variations au cours du temps. Les données satellitaires sont utilisées à la fois dans le cadre de projets d'études de l'océan et du climat à l'échelle mondiale comme WOCE (World Ocean Circulation Experiment), WCRP (World Climate Research Programm), CLIVAR (CLimate VARiability and predictability) ou GOOS (Global Ocean Observing System), mais aussi pour des projets d'océanographie opérationnelle comme GODAE (Global Ocean Data Assimilation Experiment) ou Mercator. Des études plus spécifiques à l'Atlantique Tropical ont également utilisé ce type de données, comme par exemple *Gourdeau* (1991) ou *Ubelmann* (2009). Si l'altimétrie demeure encore aujourd'hui la mesure dominante pour la dynamique océanique, d'autres types d'observations sont également disponibles comme nous allons le voir ci-dessous.

2.4.1 La mesure altimétrique

Cette mesure consiste à déterminer la hauteur du niveau de l'océan par rapport à une référence donnée. Cette référence est appelée ellipsoïde de référence. Elle suit la surface équipotentielle de gravité de la Terre. Les satellites mesurent donc l'écart à l'ellipsoïde, appelé SSH (Sea Surface Height). Néanmoins, cette grandeur n'est pas utilisable directement pour l'océanographie car elle combine plusieurs effets (figure 2.4) :

- Il faut tout d'abord prendre en compte la forme du géoïde. En effet, bien que la Terre puisse paraître ronde et de gravité constante à sa surface, il existe tout de même de petites variations dues à l'inhomogénéité de la composition du manteau terrestre (la gravité étant liée à la masse des roches situées en profondeur). Le géoïde représente en quelques sortes la forme qu'aurait la surface de l'océan s'il était parfaitement homogène (en température, salinité...) et soumis à aucune force extérieure autre que la gravité terrestre. Il ne serait alors pas parfaitement rond mais suivrait la forme des variations de gravité de la Terre. Le géoïde ne varie heureusement pas beaucoup dans le temps (les échelles de temps géologiques étant bien supérieures aux échelles de variations océaniques ou même climatiques).
- L'autre contribution aux variations de hauteur de l'océan est celle de la dynamique. Elle comprend à la fois une part permanente due à la circulation globale, et aussi une part plus variable liée aux variations des vents, des courants... La part des variations de hauteur de l'océan liée à la dynamique (dite Hauteur Dynamique, DT) peut se calculer simplement en retranchant le géoïde à la SSH : $DT = SSH - geoid$.

La hauteur dynamique DT est donc la grandeur utile pour étudier les mouvements de l'océan. Néanmoins, sa précision est liée à celle des autres variables (SSH et géoïde). Si la SSH peut être mesurée de façon assez précise, le géoïde est généralement moins bien connu surtout sur les petites échelles spatiales. Ainsi, il en résulte une certaine imprécision sur la hauteur dynamique. A noter que dans les modèles numériques où le géoïde est considéré comme sphérique (et donc également confondu avec l'ellipsoïde de référence), la hauteur dynamique est identique à la SSH.

On utilise également la SLA (Sea Level Anomaly) pour les études océanographiques. Cette grandeur correspond aux anomalies de SSH par rapport à la SSH moyenne (MSSH), elle-même calculée sur la plus longue période de temps possible ($SAL = SSH - MSSH$). La SLA traduit donc la variabilité de courte période et filtre le signal dû à la circulation permanente ou aux variations sur de grandes échelles de temps. Elle présente l'avantage d'être mesurée précisément par les satellites.

Les mesures altimétriques peuvent se faire selon deux méthodes : la mesure au nadir (actuellement utilisée) ou à large fauchée (encore en phase de développement).

Les altimètres au nadir sont ceux le plus utilisés actuellement. Ils mesurent la SSH grâce au temps d'aller-retour d'une impulsion envoyée depuis le satellite et se réfléchissant sur la surface de l'océan. Néanmoins, ce signal étant souvent très bruité par les conditions atmosphériques, il faut moyenniser un grand nombre d'impulsions pour obtenir une mesure fiable. De plus, la forme du signal au retour permet d'étudier la rugosité de l'océan et donc la hauteur des vagues. On distingue actuellement deux types de bande pour l'émission des

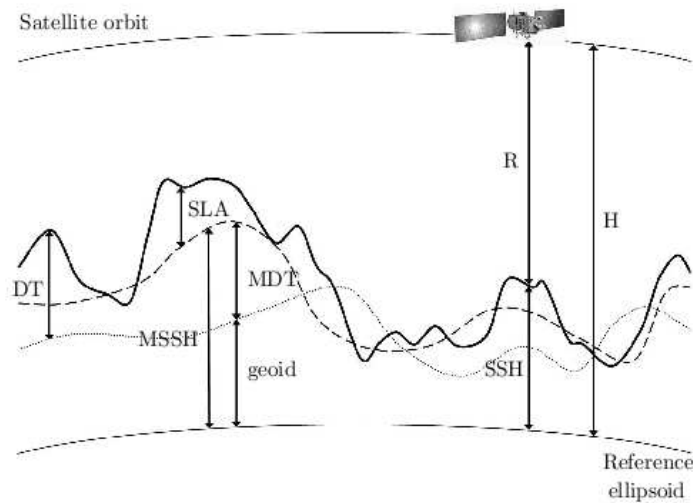


FIGURE 2.4 – Figure d'après *Castruccio* (2006). Principe de la mesure altimétrique et décomposition des différentes grandeurs liées à cette mesure. Notations : SSH (Sea Surface Height), MSSH (Mean Sea Surface Height), DT (Dynamic Topography), MDT (Mean Dynamic Topography), SLA (Sea Level Anomaly), R (distance entre le satellite et la surface de l'océan), H (distance entre le satellite et l'ellipsoïde de référence).

signaux : la bande Ku (entre 12 et 18 GHz) et la bande Ka (entre 27 et 40 GHz) qui présente de meilleurs résultats sur les zones côtières ou englacées.

Les altimètres à large fauchée, encore en développement (SWOT¹), utiliseront 2 antennes distinctes pour recevoir le signal de retour. Ainsi, par interférométrie, il sera possible d'avoir une mesure plus efficace.

Enfin, on peut noter qu'à l'imprécision de la mesure s'ajoute également l'imprécision de la position des satellites qui est de l'ordre du centimètre.

Si les satellites permettent une bonne couverture spatiale de la surface de l'océan, il a été montré par *Ubelmann et al.* (2009) qu'un satellite seul pouvait parfois être insuffisant pour capter le signal d'une TIW ou d'un anneau du Brésil, et ce en raison d'un phasage entre la trace satellite et la propagation des ondes. L'ajout de traces supplémentaires (provenant d'autres satellites par exemple) peut permettre d'éviter ces problèmes et de capter de manière plus précise les systèmes propagatifs. L'introduction d'un lisseur peut également limiter ces problèmes en augmentant le nombre de traces d'un même satellite disponibles à une date donnée. Dans la simulation d'observations satellites nous introduirons donc un seul type de traces afin de pouvoir caractériser l'impact du lissage.

2.4.2 Les autres types d'observations

En complément de la mesure altimétrique, il existe d'autres types de mesures de plus en plus exploités :

La température de surface : la température de surface (SST) est observée depuis 1972 par satellite avec l'utilisation de radiomètres mesurant l'émission infrarouge de la sur-

1. <http://swot.jpl.nasa.gov/>

face océanique. Depuis 1980, les mesures sont devenues plus précises grâce à une prise en compte d'une correction atmosphérique (présence de vapeur d'eau notamment).

La couleur de l'océan : lancé à la fin des années 70, le satellite CZCS² fut le premier dédié à observer la couleur de l'eau, quantifiant ainsi la concentration en chlorophylle A, et donc l'abondance du phytoplancton. En 1997 le capteur SeaWiFS du satellite SeaStar a permis de compléter ces observations. La couleur de l'océan est mesurée par photographie du rayonnement visible émis par la surface. En effectuant 14 fois le tour du globe par jour, le satellite SeaWiFS permet une couverture totale quotidienne à nos latitudes. Néanmoins, la présence de nuages limite fortement la quantité d'observations exploitables. La figure 1.3 du chapitre 1 est par exemple issue de ces mesures.

La salinité : l'observation de la salinité par satellite est très récente. La mission SMOS-ESA (Soil Moisture and Ocean Salinity) a par exemple débutée en novembre 2009. Elle fournit des observations de l'humidité des sols (variable importante pour comprendre les cycles hydrologiques et climatiques) et de la salinité de surface des océans. L'un des objectifs de SMOS est de fournir des cartes globales de salinité avec une résolution de 200 km tous les 10 jours et une précision de 0,1.

2.5 Les radars

Ce dernier type de données est le plus récent dans l'observation de l'océan. Il est à mi-chemin entre les données *in situ* et les données satellites. Les radars, généralement implantés sur le littoral, permettent de mesurer les vitesses de surface de façon continue sur une large zone (plusieurs dizaines ou centaines de kilomètres). Ce type de données est très utile pour étudier les courants de bord par exemple, et peut rapidement être adapté pour suivre des phénomènes en temps réels (en complément des gliders par exemple, qui eux permettent d'avoir des mesures en profondeur). Ces données sont néanmoins limitées aux voisinages des rivages et ne sont donc pas adaptées pour une étude complète d'un bassin océanique de grande dimension comme ce sera le cas dans cette thèse.

2.6 Conclusion

Comme nous avons pu le voir dans ce chapitre à travers la description non-exhaustive des réseaux de mesure, l'observation de l'océan s'est largement développée durant les dernières décennies. Les mesures *in situ* (mouillages, réseaux dérivants, campagnes océanographiques) permettent d'accéder à des données en profondeur (jusqu'à 2000m). Peu de données sont accessibles pour les grands fonds. Ces mesures ont le désavantage d'être ponctuelles, soit en temps (campagnes océanographiques), soit spatialement (réseaux dérivants ou mouillages). En complément, les données satellitaires permettent d'effectuer des séries quasi-continues en temps et en espace, mais elles restent limitées à la surface de l'océan.

L'ensemble de ces mesures permet de construire de solides jeux de données qui, couplés à des modèles numériques, aident à la compréhension de la dynamique océanographique. Ces données permettent aussi, grâce à l'assimilation de données, de corriger les sorties des modèles numériques dans le cadre de réanalyses ou d'initialisation pour la prévision.

2. <http://oceancolor.gsfc.nasa.gov/CZCS/>

Leur nombre est en constante augmentation et la nature de ces données est de plus en plus diverse, aussi l'intégration systématique des données pour corriger les modèles semble plus que jamais justifiée. Néanmoins, leur fréquence de mesure est souvent assez faible (quelques jours au moins pour les réseaux *in situ*) et malgré les satellites leur répartition spatiale est encore trop limitée. *Ubelmann et al.* (2009) ont montré que les réseaux *in situ* seuls ne permettent pas de capter efficacement les signaux liés aux TIWs, de même qu'un satellite seul peut être inefficace pour capter la propagation de tels signaux. Pour utiliser les informations issues des observations de façon optimale il semble donc nécessaire d'introduire non seulement des corrélations dans l'espace (entre les points de mesure et les zones non-observées) mais aussi et surtout des corrélations temporelles (afin de combler les vides temporels entre les dates d'observations, notamment pour des réseaux de type ARGO où la fréquence temporelle est faible). Nous avons donc choisi ces deux types de réseaux (ARGO et traces altimétriques), pas toujours optimaux pour un filtrage 3D, afin pouvoir mettre en avant l'intérêt d'une représentation 4D des observations.

L'assimilation de données pour utiliser correctement l'information des observations sera détaillée dans les chapitres 4 et 5. Il nous reste néanmoins à préciser avant, dans le chapitre 3, l'outil de modélisation que nous utiliserons pour nos travaux.

Chapitre 3

Modélisation de l’océan

Sommaire

3.1	Introduction	32
3.2	L’outil numérique : le code OPA et le modèle NEMO	32
3.3	La physique résolue par le modèle	33
3.3.1	Les équations primitives	33
3.3.2	Les conditions limites	35
	Les échanges de quantité de mouvement	35
	Les échanges de chaleur	36
	Les échanges de masse	37
3.3.3	La hauteur de la mer : SSH	37
3.4	Le traitement des équations dans le milieu discret	38
3.4.1	La discrétisation spatiale et temporelle	38
3.4.2	Paramétrisation des échelles sous-mailles	38
	Paramétrisation sous-maille horizontale	39
	Le schéma de diffusion verticale TKE	40
3.5	La configuration utilisée : TATL4	40
3.5.1	Le domaine d’étude et les conditions aux frontières	41
	Dérive de la SSH	41
	Situation géographique des frontières	42
3.5.2	Caractéristiques de la discrétisation	43
3.5.3	Les forçages utilisés	43
3.5.4	Représentation de la dynamique dans TATL4	45
3.5.5	Récapitulatif	47
3.6	Oscillation haute fréquence de la SSH et des forçages	47
3.7	Conclusion	49

3.1 Introduction

L'océanographie physique a connu un développement très rapide grâce à l'arrivée des observations satellites mais aussi grâce aux modèles numériques de la circulation générale. Ces derniers ont notamment permis (et permettent toujours) de vérifier l'accord entre la théorie et les observations, et de découvrir des processus océaniques. Ils sont également importants pour la mise en place de scénarios climatiques du fait du rôle de l'océan sur le climat (l'océan peut stocker 65 fois plus de carbone que l'atmosphère et 1200 fois plus de chaleur) et pour la prévision sur de courtes périodes au même titre que la météorologie. Ces modèles ne sont néanmoins pas parfaits. La représentation de la dynamique (océanique ou atmosphérique) qu'ils fournissent ne correspond pas forcément en tout point à la dynamique réelle. Cela vient essentiellement du fait que la résolution des modèles ne permet pas de résoudre tous les phénomènes physiques de petite échelle. L'introduction de l'assimilation de données dans les modèles est une solution pour parer à ces erreurs. Nous allons développer un peu dans ce chapitre les grands traits du modèle numérique que nous utilisons dans cette thèse et en préciser les approximations.

Tout comme les modèles de dynamique atmosphérique, les modèles océaniques se basent sur les équations de conservation de la quantité de mouvement dans un repère tournant (Navier-Stokes) :

$$\frac{D\vec{u}}{Dt} = -\frac{\vec{\nabla}P}{\rho} + \vec{g} - 2\vec{\Omega} \times \vec{u} + \vec{D} + \vec{F} \quad (3.1)$$

où t est le temps, \vec{u} le vecteur vitesse d'une particule de fluide considérée, P la pression, ρ la densité du fluide, \vec{g} l'accélération de la pesanteur et $\vec{\Omega}$ le vecteur rotation de la Terre. Le terme $\frac{D}{Dt}$ représente la dérivée Lagrangienne de la particule (c'est-à-dire les variations de cette dernière lorsque l'on suit son déplacement). Elle peut être reliée à la dérivée Eulérienne (c'est-à-dire lorsque l'on regarde les variations de la particule depuis un point fixe, ce qui inclut les variations de sa position) par : $\frac{D}{Dt} = \frac{\partial}{\partial t} + \vec{u} \cdot \vec{\nabla}$.

La différence entre les modèles atmosphériques et océaniques repose essentiellement sur les variables du vecteur d'état choisies et sur les approximations utilisées. En océanographie, on peut décrire l'évolution spatio-temporelle du système à travers sept variables : le champ de vitesse 3D, la pression, la température, la salinité et la densité. Certains modèles utilisent maintenant un système couplé océan-atmosphère, et d'autre font même intervenir les interactions glace-océan. Mais ces modèles dépassent le cadre de cette thèse. En effet, nous allons utiliser un code purement océanique. Tous les facteurs extérieurs pouvant interagir avec l'océan seront considérés sous forme de forçages (les vents, les précipitations, l'apport d'eau douce par les fleuves...) que nous détaillerons par la suite. Nous allons d'abord commencer par une description du modèle utilisé (NEMO) puis nous détaillerons la configuration régionale de l'Atlantique tropical (TATL4) utilisée.

3.2 L'outil numérique : le code OPA et le modèle NEMO

Le code OPA¹ (Océan PArallélisé, *Madec et al.*, 1998) a été développé au LOCEAN (Laboratoire d'Océanographie et du Climat - Expérimentation et Analyse Numérique) et

1. <http://www.lodyc.jussieu.fr/opa/>

dans la communauté océanographique européenne afin de simuler la circulation régionale et globale. Ce code est à la base de la plateforme de modélisation NEMO² (Nucleus for European Modelling of the Ocean), développé lui-même dans le cadre du projet européen MERSEA³ (Marine Environmental and Security for the European Area). Il a été développé et utilisé dans le cadre de nombreux projets français ou européens (CLIPPER⁴, DRAKKAR⁵, MERCATOR⁶, MERSEA³, MyOcean⁷) et son code est en constante amélioration (avec notamment le développement de maillages de plus en plus fins). Dans cette étude, nous utilisons la version 3.1 de NEMO avec une résolution au $1/4$ de degré.

3.3 La physique résolue par le modèle

3.3.1 Les équations primitives

Le code OPA résout les équations de la dynamique en milieu tournant. Ces équations sont :

- Les équations du mouvement (ou de Navier-Stokes) issues du principe de conservation de la quantité de mouvement
- L'équation de continuité, issue du principe de conservation de la masse
- L'équation d'évolution de la température, issue du principe de conservation de l'énergie
- L'équation d'évolution de la salinité, issue de la loi de conservation des substances dissoutes
- L'équation d'état de l'eau qui lie la température, la salinité et la densité

Compte tenu des échelles géophysiques et du milieu étudié, un certain nombre d'hypothèses et d'approximations peuvent être réalisées (pour plus de détails, on pourra se référer à *Chassignet et Verron, 1998*) :

- L'approximation sphérique : les surfaces géopotentielles sont supposées être des sphères, le vecteur d'accélération gravitationnelle \vec{g} est donc parallèle aux rayons de la Terre.
- L'approximation d'eau peu profonde : la profondeur de l'océan est considérée comme faible devant le rayon de la Terre (3.8 km en moyenne contre 6400 km).
- L'approximation de Boussinesq : les variations de densité de l'eau de mer sont faibles devant la valeur de référence $\rho_0 = 1020 \text{ kg.m}^{-3}$. On considère donc les variations de densité comme négligeables, sauf dans les termes de flottabilité. Le reste du temps elle est supposée constante et égale à ρ_0 .
- L'hypothèse d'incompressibilité : la densité étant considérée comme quasi constante, l'équation de conservation de la masse $\frac{DM}{Dt} = 0$ implique que la vitesse est non divergente $\nabla \cdot \vec{u} = 0$.

2. <http://www.lodyc.jussieu.fr/NEMO/>

3. <http://www.mersea.eu.org>

4. <http://www.ifremer.fr/lpo/clipper/>

5. <http://www.ifremer.fr/lpo/drakkar/>

6. <http://www.mercator-ocean.fr>

7. <http://www.myocean.eu.org/>

- L'approximation hydrostatique : dans la composante verticale de l'équation de la quantité de mouvement, l'accélération verticale et le terme de Coriolis sont négligés (car faibles devant les autres termes). Il en résulte alors un équilibre entre le gradient vertical de pression et les forces de flottabilité.
- L'hypothèse de fermeture turbulent : les flux turbulents correspondant aux processus d'échelle inférieure à celle de la discrétisation des variables (effets dits "sous-mailles ") peuvent être paramétrés en fonction des variables de grande échelle.

Considérant toutes ces hypothèses, les équations que doit résoudre le modèle sont les suivantes :

$$\frac{\partial \vec{u}_h}{\partial t} = - \left\{ (\nabla \times \vec{u}) \times \vec{u} + \frac{1}{2} \nabla (\vec{u}^2) \right\}_h - f \vec{k} \times \vec{u}_h - \frac{1}{\rho_0} \nabla_h p + \vec{D}^u \quad (3.2)$$

$$\frac{\partial p}{\partial z} = -\rho g \quad (3.3)$$

$$\nabla \cdot \vec{u} = 0 \quad (3.4)$$

$$\frac{\partial T}{\partial t} = -\nabla \cdot (\vec{u} T) + D^T \quad (3.5)$$

$$\frac{\partial S}{\partial t} = -\nabla \cdot (\vec{u} S) + D^S \quad (3.6)$$

$$\rho = \rho(T, S, p) \quad (3.7)$$

$\vec{u} = \vec{u}_h + w\vec{k} = (u, v, w)$ est le vecteur vitesse (en $m.s^{-1}$)

\vec{u}_h est la vitesse horizontale (en $m.s^{-1}$)

$w\vec{k}$ est la vitesse verticale (en $m.s^{-1}$)

T est la température potentielle (en $^{\circ}C$)

S est la salinité

ρ est la densité *in situ* (en $kg.m^{-3}$)

p est la pression (en *bar*)

z est la coordonnée verticale ascendante

\vec{k} est le vecteur vertical unitaire ascendant

f est le paramètre de Coriolis en fonction de la latitude ϕ ($f = 2\Omega \sin \phi$)

La discrétisation spatiale, que nous allons aborder dans la section 3.4.1, limite la résolution dans les petites échelles. Les effets non résolus sont donc paramétrés pour être pris en compte à travers des termes de diffusions D^u , D^T et D^S (correspondant respectivement au terme de diffusion de vitesse, de température et de salinité). L'hypothèse de fermeture turbulente suppose que ces termes peuvent être exprimés en fonction des variables de grande échelle.

3.3.2 Les conditions limites

L'océan n'est évidemment pas un milieu isolé. Des échanges d'énergie ont lieu avec les milieux environnants. Ses interactions avec l'atmosphère notamment doivent être prises en compte pour résoudre de façon réaliste la dynamique océanique. Les échanges peuvent se classer en trois groupes : les échanges de quantité de mouvement, de chaleur et de masse. Il existe deux représentations possible de ces forçages : les forçages "à flux spécifié" et les forçages de type "bulk". Dans le premier cas, les valeurs des différents flux sont préalablement calculées (sur une base d'observations et de modèles atmosphériques) puis imposées au modèle à l'interface océan-atmosphère. Dans le deuxième cas, une formulation introduite par *Large et al.* (1997) permet de déterminer la valeur des flux en fonction des variables pronostiques du modèle d'océan. Cette formulation ne correspond pas à un véritable couplage océan-atmosphère puisque les grandeurs atmosphériques sont déterminées préalablement de la même façon dans le cas des forçages à flux spécifié. Dans la configuration du modèle utilisée, les forçages sont de type *bulk*.

Les échanges de quantité de mouvement

- *Avec l'atmosphère :*

Les vents en surface exercent une tension sur les masses d'eau et leur transmettent donc une partie de leur énergie. Ce transfert de quantité de mouvement est défini par un vecteur $\vec{\tau}$ (appelé aussi vecteur de tension de vent). Il dépend essentiellement de la vitesse relative du vent par rapport à celle de la surface océanique. Il est généralement défini par une relation empirique du type :

$$\vec{\tau} = \rho_a C_D |U_{10} - U_{sfc}| (\vec{U}_{10} - \vec{U}_{sfc})$$

\vec{U}_{10} est le vecteur de vitesse du vent à une altitude de 10 m

\vec{U}_{sfc} est le vecteur de vitesse du courant océanique en surface

C_D est un coefficient d'échange turbulent déterminé en partie de manière empirique en fonction de la stabilité de l'atmosphère au niveau de la surface océanique, elle-même calculée en fonction de la température à la surface de l'océan et des variables atmosphériques

ρ_a est la densité de l'air

En surface ($z = 0$), la tension de vent est utilisée comme condition limite pour l'équation 3.2 :

$$\vec{D}_{z=0}^u = \left(A^{vm} \frac{\partial \vec{u}_h}{\partial z} \right)_{z=0} = \frac{\vec{\tau}}{\rho_0} \quad (3.8)$$

- *Avec les continents et le fond :*

Le frottement de l'eau sur le fond dissipe une certaine quantité de mouvement et ralentit les masses d'eau. Dans le modèle, un flux négatif est défini proportionnellement à la vitesse horizontale de la couche de fond et est paramétré par l'équation suivante :

$$\vec{F}_h = r \vec{u}_h^b \quad (3.9)$$

\vec{u}_h^b est le vecteur de vitesse horizontale dans la couche de fond

r est un coefficient de frottement variable selon la nature du sol

- La quantité de mouvement du flux d'eau douce apportée par les fleuves est généralement négligée dans OPA, sauf éventuellement dans des configuration côtières haute résolution qui ne nous concernent pas.

Les échanges de chaleur

- *Avec l'atmosphère :*

Les échanges de chaleur avec l'atmosphère se font à travers trois procédés physiques : transferts latents, radiatifs et conductifs.

Les transferts latents, notés \mathbf{Q}_{LAT} , sont liés à l'évaporation et à la condensation de l'eau à la surface de l'océan. L'évaporation, réaction majoritaire à la surface des eaux, est endothermique, elle consomme donc de la chaleur. La condensation en revanche est exothermique, mais cette réaction ne peut avoir lieu que dans le cas où l'eau de surface est plus froide que l'air chargé en humidité. Le flux \mathbf{Q}_{LAT} est donc généralement négatif (c'est-à-dire que le bilan est négatif pour les masses d'eau de surface). On calcule ses flux avec les formulations *bulk* :

$$\mathbf{Q}_{\text{LAT}} = \rho_a L_e C_E \vec{U}_{10} (q_a - q_w) \quad (3.10)$$

avec ρ_a la densité de l'air, L_e la chaleur latente de vaporisation de l'air, \vec{U}_{10} le vecteur vitesse du vent à 10m ($u_{10} + v_{10}$), q_a l'humidité spécifique de l'air et q_w l'humidité spécifique de l'air à saturation à la surface de l'océan.

Les transferts de chaleur radiatifs sont notés \mathbf{Q}_{LW} et \mathbf{Q}_{SW} . \mathbf{Q}_{LW} est lié à l'émission infrarouge de la surface océanique (qui perd donc de la chaleur par ce procédé) et à l'émission de même type de la couche nuageuse (qui fournit alors de la chaleur à l'océan). Le flux \mathbf{Q}_{LW} est d'autant plus négatif que le ciel est dégagé. Il se calcule par la formulation suivante :

$$\mathbf{Q}_{\text{LW}} = (1 - \alpha_{lw}) rad_{lw} - \epsilon \sigma T_w^4 \quad (3.11)$$

avec α_{lw} l'albédo de l'océan pour l'infrarouge (indiquant donc la quantité d'infrarouge non absorbée par la surface océanique), rad_{lw} une donnée satellite journalière du flux infrarouge reçu par l'océan, ϵ l'émissivité de la surface océanique et σ la constante de Stephan-Boltzmann.

\mathbf{Q}_{SW} est quant à lui lié au rayonnement du soleil atteignant la surface de l'eau. Ce flux, positif, dépend essentiellement de l'incidence du soleil et de la transparence de l'eau, et provient de jeux de données spécifiés.

Enfin, les transferts conductifs, notés \mathbf{Q}_{SENS} , sont liés directement à la différence de température entre l'air et la surface de l'océan. Ce flux de chaleur sensible peut être aussi bien positif que négatif, selon les températures respectives de l'atmosphère ou de l'océan. Dans les régions tropicales, \mathbf{Q}_{SENS} est faible devant les autres termes sans pour autant être négligeable. Ce flux est calculé par la formulation suivant :

$$\mathbf{Q}_{\text{SENS}} = \rho_a c_p^a C_H \vec{U}_{10} (T_a - T_w) \quad (3.12)$$

$$(3.13)$$

avec ρ_a la densité de l'air, c_p^a la capacité calorifique de l'air, \vec{U}_{10} le vecteur vitesse du vent à 10m, T_a et T_w les températures de l'air et de l'eau. A noter que tous ces flux, excepté \mathbf{Q}_{SW} , s'appliquent uniquement à la surface de l'eau.

- *Avec les continents et le fond :*

Les échanges de chaleur avec le fond des océans sont minimales. Ils sont essentiellement liés au flux géothermique qui réchauffe le fond des océans. Très localement, des épanchements magmatiques sous-marins ou côtiers peuvent aussi transférer de la chaleur à l'eau. Mais ces flux ne seront pas pris en compte dans le modèle tel que nous l'avons configuré.

Les échanges de masse

- *Avec l'atmosphère :*

Les échanges de masse avec l'atmosphère se font à travers l'évaporation ou la condensation (\mathbf{E}) et les précipitations (\mathbf{P}). Les précipitations doivent être spécifiées, mais l'évaporation est déduite de la formule suivante :

$$E = \frac{\mathbf{Q}_{\text{LAT}}}{L_e} \quad (3.14)$$

avec \mathbf{Q}_{LAT} le flux de chaleur latente (calculé par l'équation 3.10) et L_e la chaleur latente de vaporisation de l'air.

- *Avec les continents et le fond :*

Les continents apportent des flux d'eau douce (\mathbf{F}) à l'embouchure des fleuves. On notera par exemple l'embouchure de l'Amazone qui a une trace nettement visible sur la salinité de surface.

D'autre part, l'absence d'échange d'eau entre le fond et l'océan (ou sur les bords des continents) impose forcément que la vitesse normale à la paroi, doit être nulle ($w = 0$ au fond, $u = 0, v = 0$ le long des côtes).

3.3.3 La hauteur de la mer : SSH

Le modèle NEMO utilise une condition limite de surface libre (*Roulet et Madec, 2000*). Contrairement aux modèles dits à "toit rigide", la dénivellation de la surface libre η est prognostiquée d'après la solution modèle et les conditions limites (échanges de masse) :

$$\frac{\partial \eta}{\partial t} = \vec{w} + \mathbf{P} + \mathbf{F} - \mathbf{E} \quad (3.15)$$

avec \vec{w} la vitesse verticale en surface, \mathbf{P} , \mathbf{F} et \mathbf{E} les composantes du flux d'eau douce.

Bien que l'assimilation de données altimétriques puisse se faire avec un modèle de toit rigide (*Testut et al., 2003 ; Pinardi et al., 1995*), la configuration en surface libre présente plusieurs avantages. D'une part, elle est actuellement la plus utilisée par la communauté océanographique. De plus, l'approximation de toit rigide impose que le champ de vitesse

barotrope du modèle soit non divergent. Or il n'y a aucune raison intrinsèque pour qu'un champ analysé par l'assimilation de données satisfasse cette contrainte. Il faudrait donc, dans cette configuration, remplacer la partie divergente du champ barotrope par une partie non divergente (*Testut, 2000*) avant de redémarrer le modèle. La condition de surface libre permet quant à elle de redémarrer le modèle avec un champ de courant qui soit divergent. L'assimilation est donc dans ce cas beaucoup plus naturelle, au moins en apparence. Dans l'équation 3.2 de la quantité de mouvement, le terme $gT_c \nabla_h \frac{\partial \eta}{\partial t}$ ajouté artificiellement (qui correspond en quelque sorte à une tension superficielle), va permettre de dissiper en quelques pas de temps les ondes de gravité externes (barotropes) générées par la partie divergente du champ de vitesse lors du redémarrage du modèle.

3.4 Le traitement des équations dans le milieu discret

Les équations données ci-dessus ne peuvent pas être résolues directement dans tout l'espace spatio-temporel. L'espace doit être découpé assez finement de façon à pouvoir résoudre chaque équation par une approche discrète tout en restant en accord avec les équations primitives (non-linéaires).

3.4.1 La discrétisation spatiale et temporelle

Les équations sont en effet discrétisées par un système de différences finies centrées du second ordre. La discrétisation verticale se fait en niveau z , c'est-à-dire avec des niveaux d'épaisseur constante définie en mètres. Il existe d'autres types de discrétisations verticales (en fonction de la densité par exemple) que nous ne détaillerons pas ici. Les variables sont discrétisées selon une grille Arakawa C (*Arakawa et Lamb, 1977*). Les mailles sont centrées sur les grandeurs scalaires (comme la température ou la pression) et les grandeurs vectorielles (vitesses u , v et w) sont calculées au centre des faces de chaque maille (figure 3.1).

La dimension temporelle doit être elle aussi discrétisée. Pour cela, un schéma "Leapfrog" est utilisé, stabilisé par l'utilisation d'un filtre d'Asselin (*Asselin, 1972*) permettant d'éviter la dissociation des modes pairs et impairs (équation 3.17).

$$\text{Leapfrog : } u^{t+\Delta t} = u^{t-\Delta t} + 2\Delta t \text{ RHS}^t \quad (3.16)$$

$$\text{Asselin : } u_f^t = u^t + \gamma \left(u_f^{t-\Delta t} - 2u^t + u^{t+\Delta t} \right) \quad (3.17)$$

(l'indice f dénote les valeurs filtrées)

3.4.2 Paramétrisation des échelles sous-mailles

La circulation océanique est liée à des processus de tailles très différentes. En effet, on retrouve des phénomènes allant de la turbulence petite échelle jusqu'à la circulation grande échelle de plusieurs milliers de kilomètres. Les processus de petite échelle qui ont une longueur inférieure à la taille de la maille du modèle ne peuvent pas être résolus explicitement. Pourtant leur rôle est primordial, il faut donc les prendre en compte dans les équations du modèle. Cet effet est pris en compte dans les termes de diffusion D^x dans les équations 3.2, 3.5 et 3.6. Compte tenu de l'hypothèse de fermeture turbulente, on suppose que ces termes dépendent linéairement des gradients de grande échelle. De plus,

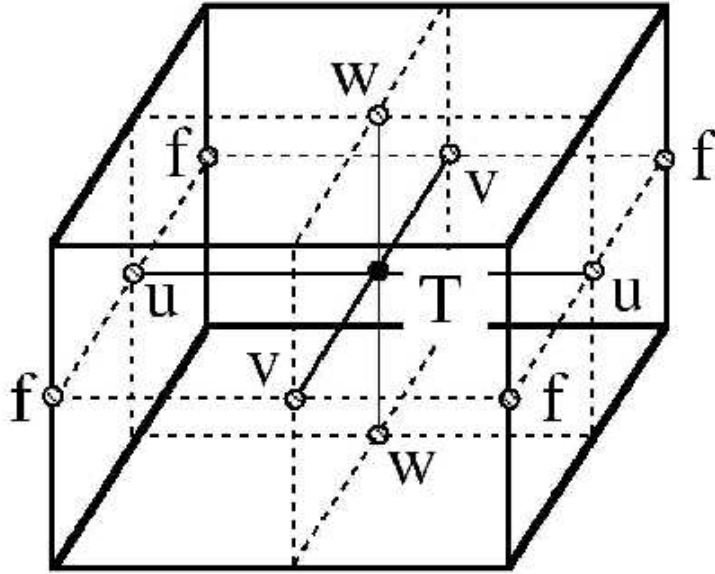


FIGURE 3.1 – Figure d’après *Castruccio* (2006). Disposition des variables du modèle NEMO sur la grille C. T indique le point où les grandeurs scalaires sont calculées. u , v et w indiquent les points de calcul des grandeurs vectorielles (vitesses).

compte tenu de l’anisotropie du milieu, les termes horizontaux $D^h x$ et verticaux $D^v x$ sont traités séparément :

$$D^x = D^{hx} + D^{vx}$$

Paramétrisation sous-maille horizontale

Dans la configuration de NEMO utilisée, les termes de diffusion des traceurs et des vitesses sont traités séparément.

Pour les traceurs, la paramétrisation choisie est celle proposée par *Gent et McWilliams* (1990). La diffusion est exprimée par la somme d’un terme diffusif (sous forme d’un laplacien à travers les isopycnes) et d’un terme d’advection :

$$D^{hT} = \nabla \cdot (A^{hT} \mathfrak{R} \nabla T) + \nabla (U^* T) \quad (3.18)$$

où A^{hT} (positif) est un coefficient de diffusion, U^* est la partie non divergente du transport tourbillonnaire et \mathfrak{R} est un opérateur (3×3) prenant en compte les pentes entre les surfaces isopycnales et les surfaces géopotentiels. Cette paramétrisation a pour effet de réduire l’énergie moyenne de l’océan.

Pour les vitesses en revanche, un opérateur bilaplacien (ou biharmonique) est choisi dans notre configuration afin d’être plus sélectif au niveau des petites échelles et de ne pas éliminer la turbulence résolue. En effet, le bilaplacien amortit moins les mouvements de grande échelle est permet donc la résolution de la turbulence océanique.

Le schéma de diffusion verticale TKE

Les processus de mélange vertical turbulent ont également des échelles bien inférieures à la résolution du modèle. Dans les modèles hydrostatiques, les mouvements turbulents verticaux sont toujours traités avec une paramétrisation. La paramétrisation verticale est notamment importante dans les couches de surface puisqu'elle permet la ventilation de l'océan et la pénétration des échanges avec l'atmosphère vers les eaux plus profondes. Les flux paramétrés dépendent linéairement des gradients verticaux locaux des variables grande échelle, par analogie avec la diffusion moléculaire :

$$\vec{D}^{vu} = \frac{\partial}{\partial z} \left(A^{vm} \frac{\partial \vec{u}_h}{\partial z} \right) \quad (3.19)$$

$$D^{vT} = \frac{\partial}{\partial z} \left(A^{vT} \frac{\partial T}{\partial z} \right) \quad (3.20)$$

$$D^{vS} = \frac{\partial}{\partial z} \left(A^{vT} \frac{\partial S}{\partial z} \right) \quad (3.21)$$

où D^{vu} , D^{vT} et D^{vS} sont respectivement les paramètres de diffusions pour la vitesse, la température et la salinité, chacun de ces paramètres étant associé à un coefficient de diffusion turbulent (A^{vm} pour la vitesse et A^{vT} pour les traceurs). Dans la configuration utilisée, ces coefficients sont calculés à partir du schéma TKE (Turbulent Kinetic Energy), basé sur une équation pronostique de l'énergie cinétique turbulente (qui détermine l'évolution de l'énergie en comparant sa production par cisaillement vertical et sa destruction par la stratification, sa diffusion et sa dissipation). Ce schéma de fermeture turbulente, d'ordre 1.5, a été implanté dans OPA par *Blanke et Delecluse* (1993). Pour une description détaillée, on pourra se référer à *Langlais* (2007) ou *Madec et al.* (1998). On notera simplement ici qu'une valeur seuil minimale est imposée pour les coefficients A^{vm} et A^{vT} afin de prévenir d'éventuelles instabilités numériques liées à une diffusion verticale trop faible.

Dans les régions où peuvent se développer des instabilités hydrostatiques (avec une fréquence de Brünt Väisälä négative), le déclenchement de la convection verticale permet d'éliminer ces instabilités. Or dans un modèle hydrostatique cette convection n'est pas permise. En considérant que cette convection est un agent de mélange plus que de transport, elle va être prise en compte par le schéma TKE dans les régions instables en utilisant un schéma de diffusion verticale augmentée. Cela schéma homogénéise la colonne d'eau en imposant une forte diffusion verticale des traceurs et de la quantité de mouvement (de l'ordre de $1 \text{ m}^2 \cdot \text{s}^{-1}$).

Les coefficients de diffusion liés à la paramétrisation sous-maille sont généralement déterminés arbitrairement de façon à ce que la physique modélisée corresponde à la physique observée (épaisseur de la couche de mélange par exemple). La valeur exacte de ces coefficients est donc une source d'incertitude intrinsèque au modèle, et elle sera utilisée par la suite pour tenter de mettre en place une première expérience d'assimilation (chapitre 6).

3.5 La configuration utilisée : TATL4

Nous allons maintenant détailler les spécificités de la configuration de NEMO que nous utilisons, avec notamment le découpage du domaine et les conditions aux frontières, ainsi

que les jeux de forçages atmosphériques choisis. Un tableau récapitule les caractéristiques principales de la configuration dans la section 3.5.5.

3.5.1 Le domaine d'étude et les conditions aux frontières

Puisque le cadre de nos expériences va être appliqué uniquement à l'Atlantique tropical, il n'est pas nécessaire d'utiliser directement un modèle global de circulation. Nous avons donc choisi une configuration régionale du modèle NEMO centrée sur l'Atlantique tropical, TATL4 (figure 3.2), similaire à celle développée dans le cadre de la thèse de *Ubelmann* (2009). Le domaine s'étend de 15°S à 17.75°N et de 61.5°W à 15°E. Bien entendu, le fait d'extraire un domaine régional de ce type implique également de spécifier les conditions à chacune des frontières (OBC, Open Boundary Conditions) que nous allons détailler ci-dessous. Il existe plusieurs types de frontières pour une configuration régionale :

Les frontières fermées ne laissent passer aucun flux de matière, de salinité ou de chaleur, elles ne sont donc pas bien adaptées lorsque la dynamique externe influence la dynamique du domaine, ou plus simplement lorsque des courants traversent les frontières du domaine. *Les frontières spécifiées* consistent à imposer autour du domaine les grandeurs océaniques issues généralement d'une simulation globale. En général, une "zone tampon" est utilisée le long des bordures du domaine dans laquelle la viscosité est fortement augmentée, permettant ainsi un écoulement plus laminaire proche des frontières. Cela évite les discontinuités trop fortes entre la dynamique générée à l'intérieur du domaine et les valeurs spécifiées par les frontières.

Les frontières radiatives quant à elles permettent de prendre en compte à la fois la dynamique issue du domaine et les valeurs spécifiées. C'est ce type de frontière que nous utiliserons dans cette thèse. On pourra se référer à la thèse de *Cailleau* (2004) pour une description plus détaillée du fonctionnement des frontières. Dans la configuration TATL4, les valeurs spécifiées par les OBC sont issues des sorties du run global interannuel DRAKKAR B83 (*Dussin et al.*, 2009), sorties correspondant à des moyennes sur 5 jours. Les OBC sont donc spécifiées tous les 5 jours, et interpolées linéairement pour les pas de temps intermédiaires.

Dérive de la SSH

Les OBC spécifient les valeurs des vitesses et des traceurs (température et salinité), mais pas la hauteur de l'eau. Ainsi, les frontières étant radiatives et non pas purement spécifiées, le bilan volumique peut s'avérer différent de celui du run global à l'intérieur du domaine. Si aucune correction n'est appliquée, le volume d'eau dans le domaine peut varier significativement au cours du temps et entraîner une dérive de la SSH. Pour pallier à ce problème, nous utilisons une condition de moyenne nulle sur les flux aux frontières. Ainsi, le volume est maintenu constant à l'intérieur du domaine. On aurait également pu utiliser une condition directement sur la SSH en imposant une moyenne constante au cours du temps, mais cela ne s'avérerait pas plus pertinent. En effet, le réchauffement ou le refroidissement des masses d'eau à l'échelle du bassin peut très bien entraîner une variation du niveau moyen de SSH, variation qui s'avérerait alors pertinente d'un point de vue physique. Il est donc préférable de maintenir l'équilibre à travers les frontières et non directement dans le domaine.

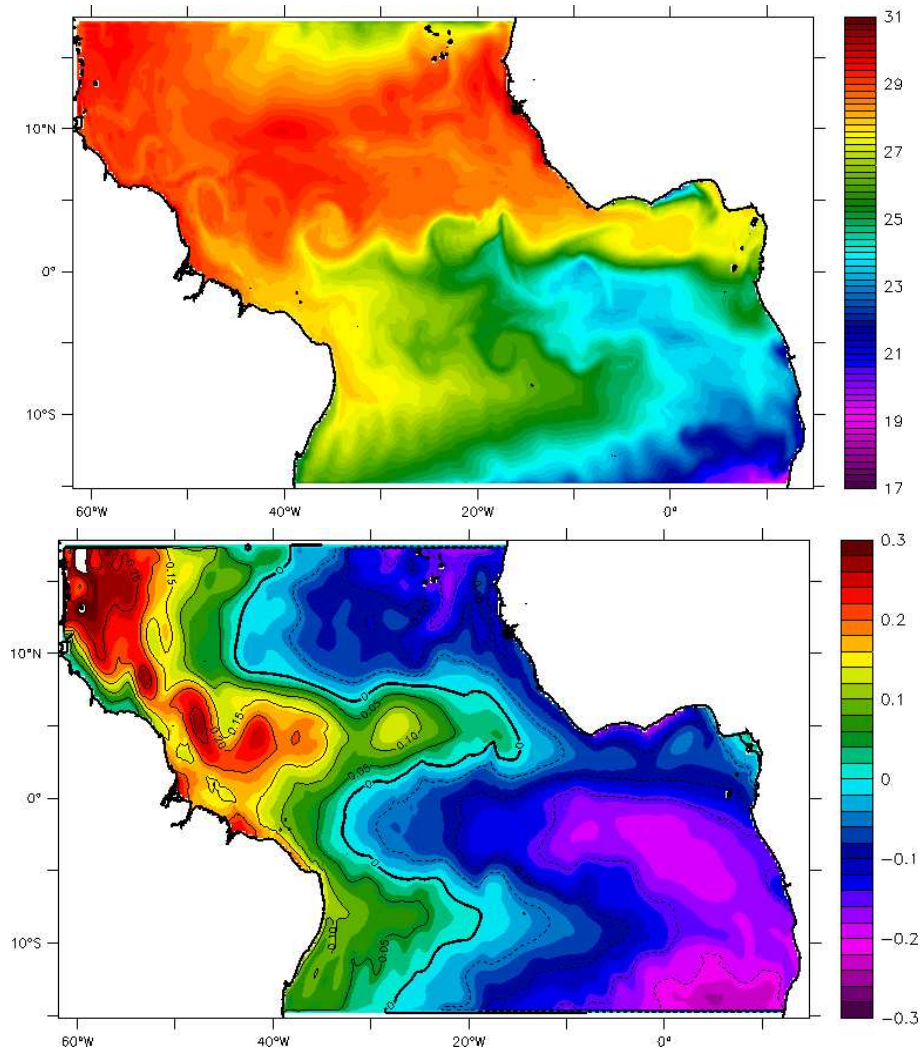


FIGURE 3.2 – Cartes de la température de surface instantanée en $^{\circ}\text{C}$ (haut) et de la SSH instantanée en m (bas) au 25 juillet 2005 pour le run libre de référence dans la configuration TATL4.

Situation géographique des frontières

Nous avons déjà précisé les limites du domaine (et donc la localisation des frontières) que l'on peut visualiser par exemple sur la carte du bas de la figure 3.2. Nous précisons juste un peu plus ici le choix de ces limites. D'une part, le domaine est choisi suffisamment grand pour éviter une influence trop grande entre les OBC et la dynamique centrale du domaine qui nous intéresse (avec notamment les TIWs et les anneaux du Brésil). En effet, dans le cadre d'une expérience d'assimilation de données, la dynamique de l'écoulement sera modifiée par l'analyse, il ne faudrait donc pas que cette correction soit effacée par des frontières trop proches. De plus, les frontières Nord et Ouest se croisent au niveau d'un point de terre (l'île de Saint Bartélémy), ce qui permet d'éviter des interactions trop fortes au niveau de la maille commune aux deux frontières. Enfin, le courant le plus marqué du domaine, le NBC, est un courant sortant (par la frontière Ouest). Donc même s'il existe

niveau	profondeur	niveau	profondeur	niveau	profondeur
1	3.046733	17	322.0169	33	2710.133
2	9.454049	18	382.1444	34	2939.812
3	16.36397	19	452.4429	35	3173.588
4	23.89871	20	534.0197	36	3410.756
5	32.20929	21	627.8525	37	3650.712
6	41.48185	22	734.715	38	3892.95
7	51.94513	23	855.1112	39	4137.047
8	63.87905	24	989.2289	40	4382.654
9	77.62451	25	1136.922	41	4629.485
10	93.59412	26	1297.724	42	4877.303
11	112.2835	27	1470.893	43	5125.919
12	134.2823	28	1655.472	44	5375.177
13	160.284	29	1850.365	45	5624.952
14	191.0925	30	2054.414	46	5875.141
15	227.6233	31	2266.454		
16	270.8962	32	2485.371		

TABLE 3.1 – Profondeurs (en mètres) des niveaux verticaux dans la configuration TATL4.

des différences entre les valeurs spécifiées par la frontière ouest et celles déterminées par le modèle, les mauvaises interactions se feront uniquement dans du côté nord-ouest et seront rapidement entraînées vers l’extérieur du domaine.

On note également que la frontière Est se situe sur le continent africain, il n’y aura donc aucune interaction à ce niveau. La limite Sud présente quant à elle une dynamique presque parallèle à la frontière, elle aura donc peu d’influence sur le domaine. Enfin, parmi les courants principaux, on notera un courant entrant de bord ouest (DWC, Deep Western Current), longeant la côte du Brésil en profondeur. La formation de ce courant provenant de la circulation à l’échelle globale, il ne devrait pas être beaucoup influencé par la dynamique du domaine.

3.5.2 Caractéristiques de la discrétisation

La configuration TATL4 est réalisée sur la grille ORCA025 (*Barnier et al.*, 2008) de résolution 0.25° en longitude et en latitude, avec 46 niveaux verticaux (voir table 3.1). Ces niveaux sont de l’ordre de quelques mètres en surface (3m pour le premier niveau) et augmentent de plus en plus en profondeur (le dernier niveau étant de 250m).

La discrétisation temporelle se fait avec un pas de temps de 4320 secondes (72 minutes), qui est un diviseur de 86400 secondes (1 jour). Cela permet d’éviter un décalage horaire de jour en jour.

3.5.3 Les forçages utilisés

Les flux à l’interface océan-atmosphère sont déterminés à partir des équations 3.10, 3.11, 3.13 et 3.14. Pour calculer tous ces flux, seules les grandeurs Ta (température de l’air), qa (humidité relative de l’air), \vec{U}_{10} (vitesse du vent à 10m) et rad_{lw} (rayonnement

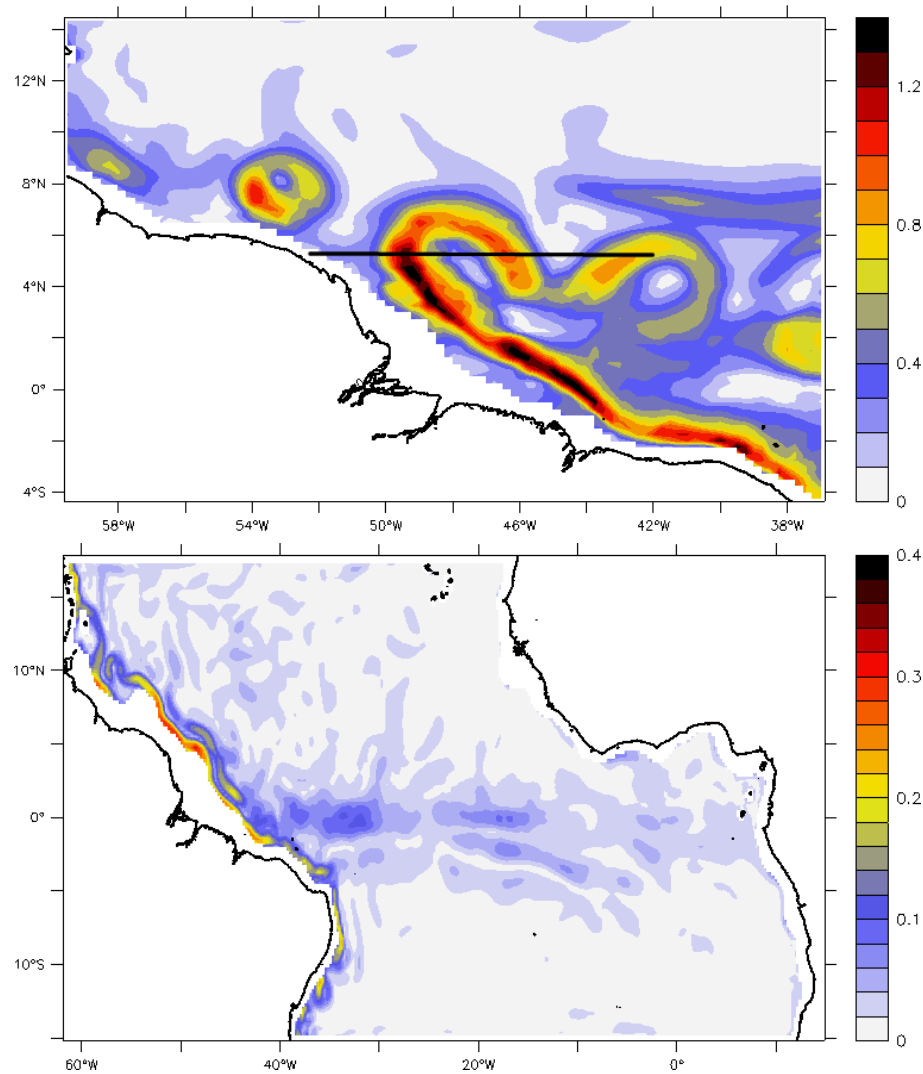


FIGURE 3.3 – Cartes de la vitesse absolue instantanée en $m.s^{-1}$: à 30 m de profondeur (haut) et moyennée entre 1500 et 2500 m (bas) au 25 juillet 2005 pour le run libre de référence dans la configuration TATL4. Le trait noir sur la figure du haut indique la position des coupes tracées sur la figure 3.5.

infrarouge) doivent être spécifiées à travers des forçages atmosphériques. Il faut également préciser la valeur des précipitations \mathbf{P} et du rayonnement solaire \mathbf{Q}_{sw} pour compléter l'ensemble des flux nécessaires. Afin d'être cohérent avec les OBC utilisées, les forçages atmosphériques ont été choisis identiques à ceux utilisés dans le run B83 (*Dussin et al.*, 2009). Il s'agit du jeu de forçages DFS4.1 (DRAKKAR Forcing Set 4.1) documenté dans *Brodeau et al.* (2010), construit à partir de différentes sources (CORE, ERA40...). Les valeurs de Ta , qa et \bar{U}_{10} (u_{10} et v_{10}) sont spécifiées toutes les 6 heures. rad_{lw} et \mathbf{Q}_{sw} sont donnés journalièrement. Enfin, \mathbf{P} est spécifié tous les mois. Le pas de temps du modèle étant inférieur à la fréquence des données, une interpolation linéaire est effectuée pour chaque pas de temps intermédiaire entre deux données disponibles.

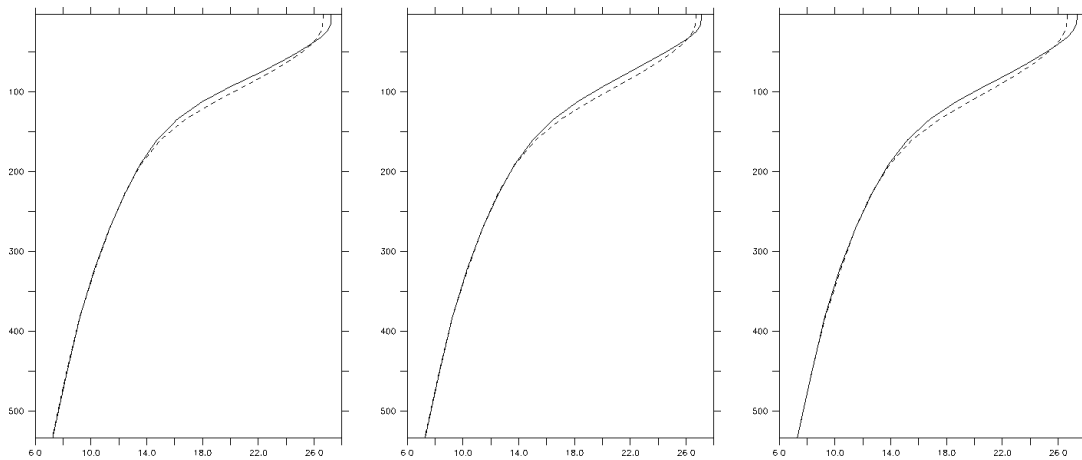


FIGURE 3.4 – Profils de température moyennée horizontalement sur tout le bassin, pour la configuration TATL4 (trait plein) et le domaine équivalent dans B83 (en pointillés). Les profils sont moyennés sur les années 2001 (à gauche), 2003 (au milieu) et 2005 (à droite).

La configuration TATL4 telle que nous l’avons définie diffère légèrement de celle de *Ubelmann* (2009). La situation géographique reste la même, en revanche les forçages atmosphériques que nous avons utilisés sont similaires à ceux du run B83. De même, les conditions aux frontières sont directement issues de ce même run. On peut donc considérer notre configuration comme une sous-configuration du run B83 en mode TATL4, aux différences près de la version du code utilisée (NEMO 2.3 pour B83, NEMO 3.1 pour notre run). Le run B83, renseigné dans le rapport technique de *Dussin et al.* (2009), couvre la période de 1958 à 2007. Notre état initial, issue du 1er janvier 1995, se situe donc plutôt vers la fin de B83. Nous évitons ainsi d’introduire d’éventuels problèmes liés au spin-up de B83 et on peut considérer que notre modèle est déjà stable dès le départ. Néanmoins, nous effectuerons les expériences d’assimilation sur l’année 2005, soit après 10 ans de runs TATL4, afin de permettre à la dynamique de se stabiliser dans le bassin en accord avec les OBC et les forçages atmosphériques. La figure 3.4 permet par exemple de comparer le contenu thermique du bassin entre TATL4 et B83. Globalement, les deux runs sont assez proches. De légers écarts sont visibles et peuvent être dus aux différences entre les versions du code NEMO utilisées. La mise en place d’OBC et leur paramétrisation peut également limiter les échanges entre le bassin et l’extérieur, et donc modifier en partie la dynamique intrinsèque.

3.5.4 Représentation de la dynamique dans TATL4

Pour terminer la mise en place de la configuration, nous effectuons une vérification très succincte de la représentation de la dynamique sur un run libre (c’est-à-dire sans assimilation de données). L’état initial du run est directement issu du run B83, au premier janvier 1995. La propagation se fait ensuite en étant forcée uniquement par les OBC et les forçages définis ci-dessus. Nous reviendrons sur les détails de ce run de référence dans le cadre des expériences d’assimilation du chapitre 7.

Une lecture qualitative des résultats permet de vérifier que les ondes d’instabilité sont correctement représentées, comme on peut le voir sur la figure 3.2 (haut) montrant la température de surface instantanée au 25 juillet 2005. Elles ont une longueur d’onde de 800

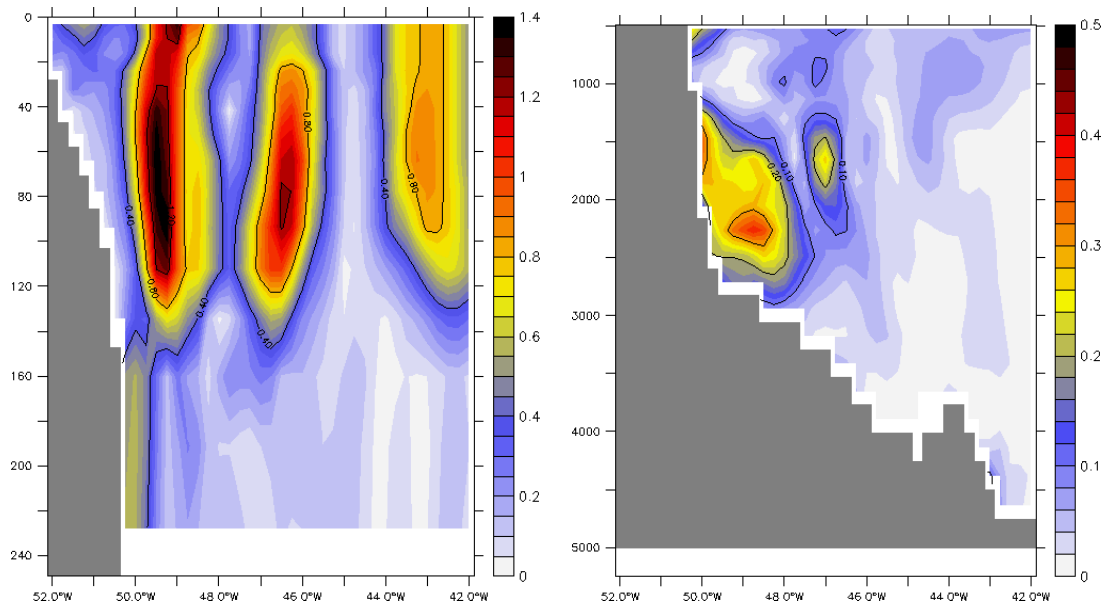


FIGURE 3.5 – Coupes de la vitesse absolue instantannée en $m.s^{-1}$ à $5^{\circ}N$: entre la surface et 250 m de profondeur (gauche) et entre 500 et 5000 m de profondeur (droite) au 25 juillet 2005 pour le run libre de référence dans la configuration TATL4. La position des coupes est rappelée sur la figure 3.3 (bas).

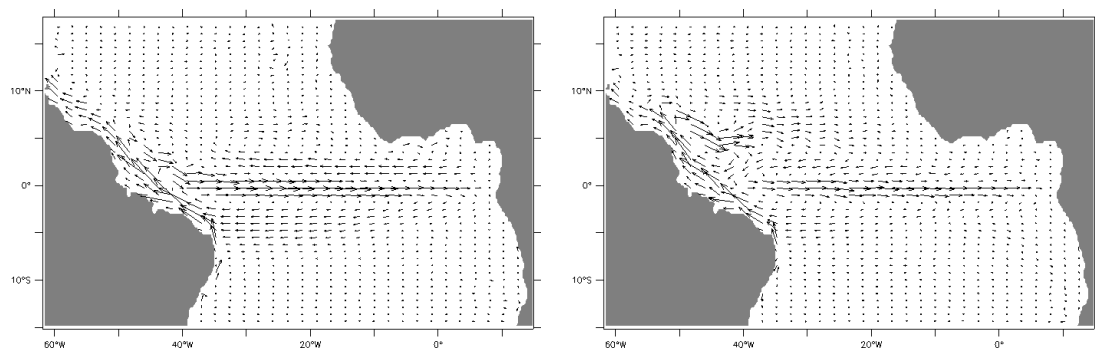


FIGURE 3.6 – Représentation des vecteurs vitesses de sub-surface (entre 10m et 110m) moyennés sur le printemps 2005 (à gauche) et l'automne 2005 (à droite).

à 900 km et une signature sur la SSH (figure 3.2, bas) de l'ordre de 5 à 10 cm, ce qui correspond aux descriptions issues des observations (voir section 1.3.1).

Les anneaux du Brésil, dont on peut voir la signature claire sur la vitesse absolue à 30 m de profondeur (figure 3.3, haut), ont également une signature en profondeur conforme aux observations (figure 3.5, gauche). Ils présentent des vitesses correctes, jusqu'à 1 m.s^{-1} (les vitesses les plus élevées correspondent au courant nord-brésilien NBC). Il en est de même sur la signature en SSH (figure 3.2, bas) qui peut atteindre jusqu'à 22 cm au centre des anneaux.

Enfin, on retrouve bien la signature du courant de bord ouest, comme l'illustrent les figures de vitesse absolue 3.5 (droite) et 3.3 (bas). La représentation des vitesses (figure 3.6) au printemps et à l'automne rappelle correctement le schéma théorique présenté dans la première section du manuscrit (figure 1.2). Les courants principaux (EUC par exemple) ont une trace bien visible, de même que l'activité tourbillonnaire dans la région au nord du Brésil qui se développe clairement à l'automne.

On note néanmoins qu'il est difficile de voir clairement certaines branches des courants, comme le NEC ou le sSEC. La modélisation est donc efficace pour représenter la dynamique globale mais reste parfois limitée pour certains courants faibles ou étroits. Le rôle de l'assimilation de données et de palier en partie ces problèmes.

3.5.5 Récapitulatif

Le tableau 3.2 permet de récapituler les paramètres principaux de la configuration TATL4.

Le domaine et sa discrétisation :	
Limites du domaine :	61.5°W à 15°E, 15°S à 17.75°N
Dimensions de la grille :	307 × 134 × 46
Résolution spatiale :	0.25° (soit environ 28 km à l'équateur)
Discrétisation temporelle :	$\Delta T = 4320$ secondes
Paramétrisation :	
Turbulence horizontale :	Laplacien pour les traceurs et bi-laplacien pour les vitesses
Turbulence verticale :	Paramétrisation TKE
Forçages (DFS4.1) :	
Température de l'air, humidité et vents (T_a , q_a et \vec{U}_{10}) :	fréquence de 6h
Flux solaire et rayonnement infrarouge (\mathbf{Q}_{sw} et \mathbf{Q}_{lw}) :	fréquence de 1 jour
Précipitations (\mathbf{P}) :	fréquence de 1 mois

TABLE 3.2 – *Caractéristiques principales de la configuration TATL4.*

3.6 Oscillation haute fréquence de la SSH et des forçages

Dans une expérience d'assimilation jumelle, les jeux d'observations sont généralement simulés d'après des sorties instantanées du modèle. Or nous avons pu constater lors de

nos premières expériences un problème lié à la SSH. En effet, les forçages en vent étant spécifiés toutes les 6 heures, la SSH répond à la même fréquence à ces forçages. Bien qu'ils soient normalement lissés linéairement entre deux valeurs spécifiées consécutives, lorsque des valeurs instantanées de SSH sont extraites on peut constater une oscillation haute fréquence dans le signal. Ces oscillations introduisent un bruit suffisamment fort pour brouiller le véritable signal de la SSH et le rendre inutilisable par l'assimilation de données. Cette oscillation n'est pas visible dans les sorties DRAKKAR du fait que ces sorties sont moyennées sur 5 jours.

Etant contraints d'utiliser des sorties instantanées dans les expériences jumelles d'assimilation, nous avons choisi de lisser les forçages (et non pas directement les sorties) avec une moyenne glissante de 5 jours. Ce lissage a permis d'annihiler les oscillations hautes fréquences sans changer la dynamique du système, comme le montre la figure 3.7 : les trajectoires des variables 3D sont identiques dans les deux cas (avec forçages lissés ou non). La trajectoire moyenne de la SSH sur des sorties dont les forçages en vent ont été lissés correspond également à la trajectoire de la SSH de base, mais sans les oscillations haute fréquence.

Nous utiliserons par la suite dans toutes nos expériences ces forçages lissés.

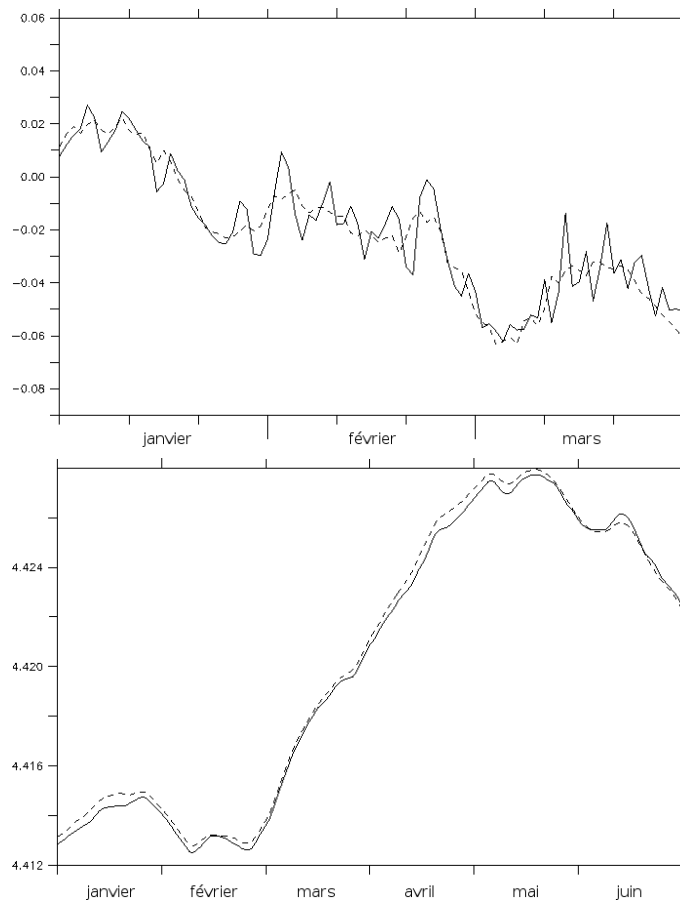


FIGURE 3.7 – Evolution de la SSH en m (en haut) et de la température en °C (en bas) pour le run effectué avec des forçages en vent haute fréquent (en trait plein) et celui effectué avec des forçages lissés sur 5 jours (en pointillés). La SSH est donnée en un point et la température est moyennée sur tout le domaine.

3.7 Conclusion

La modélisation de la dynamique océanique a connu et connaît encore de nombreux progrès. Si l'amélioration de la résolution des modèles et la précision des forçages extérieurs sont de plus en plus précises, la modélisation ne permet pas encore à l'heure actuelle de décrire de façon parfaite l'état de l'océan. Les nombreuses approximations encore liées aux modèles les rendent forcément différents de la réalité. Nous avons pu le constater dans ce chapitre lors de la mise en place de la configuration TATL4. La dynamique du bassin est dans l'ensemble représentée correctement, mais certains courants sont peu visibles.

L'assimilation de données est un rouage important permettant de faire le lien entre la réalité des observations et l'imprécision des modèles pour en tirer une description de l'état de l'océan la plus juste et la plus complète possible. Nous allons donc maintenant aborder la théorie de ces méthodes dans les chapitres 4 et 5.

Deuxième partie

Assimilation de données

Chapitre 4

Filtre de Kalman et lisseurs linéaires optimaux

Sommaire

4.1	Introduction	54
4.2	Théorie de l'estimation et le filtre de Kalman	55
4.2.1	Formulation du filtrage	56
4.2.2	Le filtre de Kalman	57
4.2.3	Les limites du filtrage	57
4.3	Problème du lissage : les différentes approches	58
4.3.1	Le lissage conjoint	58
	Fixed-interval smoother	60
	Fixed-lag smoother	60
	Ensemble smoother	60
4.3.2	Le lissage marginal	64
	Forward-backward smoother	64
	Two-filter smoother	64
4.3.3	Les thématiques d'application du lissage	64
4.4	Le lisseur fixed-lag	65
4.4.1	Formulation du lisseur de Kalman	66
4.4.2	L'erreur modèle indispensable au lissage	66
4.5	Conclusion	67

4.1 Introduction

Comme nous l'avons vu dans le chapitre précédent, la modélisation seule est insuffisante pour résoudre de façon réaliste tous les processus de la dynamique océanique. Les paramétrisations sous-maille, l'approximation linéaire des équations et les forçages externes au modèle introduisent des erreurs ou imprécisions, rendant la solution modèle inévitablement différente de la réalité. Quant aux observations, malgré leur nombre en constante augmentation, elles restent insuffisantes pour reconstituer à elles-seules le champ 3D de l'océan. Les deux sources d'information disponibles sont donc insuffisantes lorsqu'on les considère indépendamment pour avoir un bon estimé de la réalité. L'assimilation de données intervient donc pour combiner la connaissance de l'état réel de l'océan (à travers les observations) et la connaissance de la théorie gouvernant la dynamique océanique (à travers la solution déterminée par le modèle). L'état produit, dit analysé, devrait alors en théorie être le plus proche possible de la réalité tout en restant cohérent avec les équations du modèle. Cette analyse permet d'une part de corriger les différentes erreurs dues au modèle, et permet d'autre part de reconstituer un état réaliste continu à partir d'observations ponctuelles.

La recherche d'une solution analysée optimale est complexe et dépend fortement du positionnement du problème. Les enjeux actuels ne sont plus simplement d'estimer le meilleur état initial pour une prévision. Désormais, l'assimilation est aussi utilisée pour corriger certains paramètres des modèles encore mal connus (comme les flux océan-atmosphère, par exemple *Bonekamp et al.*, 2001 ; *Stammer et al.*, 2004 ou *Skachko et al.*, 2009, les paramètres des processus sous-maille...) ou encore pour optimiser les réseaux d'observation (*Verron et Holland*, 1989 ou *Ubelmann et al.*, 2009). Outre ces enjeux, l'assimilation a également évolué dans ses méthodes. Il existe de nombreux angles pour aborder un problème d'estimation optimale, notamment de part la place donnée à l'observation. L'information apportée par les mesures peut en effet être considérée comme ponctuelle ou être plus ou moins étendue dans l'espace et dans le temps grâce à des corrélations. Si les corrélations spatiales sont traditionnellement bien prises en compte, les corrélations temporelles (et donc une représentation 4D des observations) restent encore marginales, notamment dans les méthodes séquentielles.

La figure 4.1 illustre schématiquement les représentations possibles d'une observation. Dans le cas a), l'influence d'une observation se fait uniquement au point et à l'instant de mesure (ou dans son voisinage), il s'agit donc d'un cas unidimensionnel. Dans le cas b), l'information de l'observation est étendue spatialement grâce à des corrélations tridimensionnelles (il s'agit donc d'un cas 3D) mais uniquement à l'instant de la mesure. Ce cas peut être simplement propagé ensuite (par la dynamique du modèle par exemple), étendant donc l'influence de l'observation vers les instants futurs. Il s'agit alors du cas c), que l'on qualifie de dimension 3,5. Enfin, dans le cas d), l'influence de l'observation est également étendue vers le passé, on est alors dans un cas quadridimensionnel. Les différentes perceptions de l'observation dans l'espace et le temps conduisent à autant de variantes dans les méthodes d'assimilation. Dans le cadre de cette thèse, nous utilisons une méthode d'assimilation basée sur la théorie de l'estimation, qui est donc une méthode statistique. Nous donnons les bases de cette théorie dans la section 4.2 et de sa formulation 3D (filtrage) avant de discuter des diverses possibilités d'une formulation 4D (lissage) dans la section 4.3. Cette section exposera les approches de lissages possibles telles qu'elles sont présentées dans *Cosme et al.* (in press), ainsi que leurs applications privilégiées. Nous

détaillerons plus en détail dans la section suivante la méthode de lissage utilisée pour nos travaux.

Les autres méthodes d'assimilation ne seront pas traitées ici. Notons simplement qu'il existe, en dehors de l'approche d'estimation optimale, la méthode du nudging (voir par exemple *Hoke et Anthes, 1976*; *Verron et Holland, 1989*; *Verron et al., 1992* ou *Blayo et al., 1994*) et les méthodes de contrôle optimal ou variationnelles (*Sasaki, 1955*; *Sasaki, 1958*; *Lions, 1968*), elles-mêmes déclinées en méthodes 3D (*Courtier et al., 1998*; *Rabier et al., 1998*; *Andersson, 1998*) ou 4D (*Bennett, 1992*; *Amodei, 1995*; *Bennett, 2002*; *Bennett, 2002*).

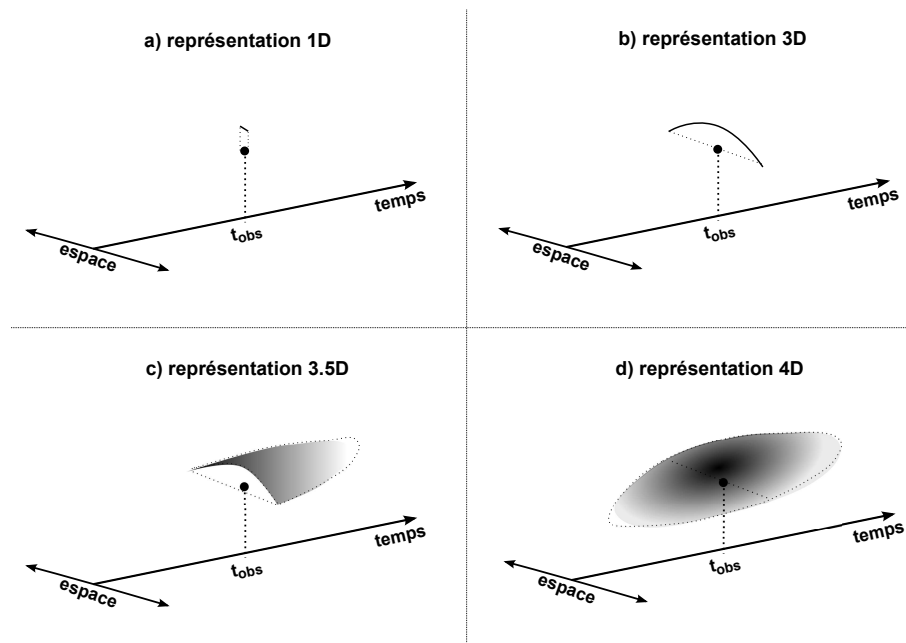


FIGURE 4.1 – Schématisation des différentes représentations d'une observation dans l'espace et le temps. Pour plus de lisibilité la dimension spatiale est représentée sur un seul axe mais il s'agit en réalité d'une dimension 3D.

4.2 Théorie de l'estimation et le filtre de Kalman

La théorie de l'estimation consiste à trouver la meilleure description statistique possible d'un état inconnu compte tenu d'informations disponibles (par exemple déterminer l'état de l'océan d'après des observations et une solution modèle). Historiquement, l'assimilation de données consiste à déterminer le meilleur état initial pour une prévision. Il s'agit donc d'un problème de filtrage d'un point de vue de la théorie de l'estimation. Le filtrage peut se traduire par une formulation Bayésienne qui consiste alors à déterminer la fonction de probabilité (pdf) d'un état \mathbf{x} à un temps k (\mathbf{x}_k) compte tenu d'observations $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k\} = \mathbf{y}_{1:k}$ passées et présentes (à partir du temps d'indice 1 par exemple) :

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \quad (4.1)$$

4.2.1 Formulation du filtrage

Dans le cadre de problèmes géophysiques, l'assimilation de données consiste généralement à analyser une succession d'états propagés dans le temps. Il s'agit alors d'une chaîne de Markov (*Wikle et Berliner, 2007; Ihler et al., 2007*) du type $\mathbf{x}_{0:K} = \{\mathbf{x}_0, \dots, \mathbf{x}_K\}$, où le premier état d'indice 0 n'est pas observé et le dernier état d'indice K correspond à la dernière analyse. Chaque état est déterminé en fonction des états précédents, ce qui suppose la connaissance de la pdf de transition d'un état $(k-1)$ au suivant (k) , $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. En pratique, un modèle physique est souvent disponible, ce qui permet de réduire l'incertitude associée à cette transition. La transition physique peut être exprimée par :

$$\mathbf{x}_k = \mathcal{M}_{k-1,k}\mathbf{x}_{k-1} + \eta_{k-1,k} \quad (4.2)$$

où $\mathcal{M}_{k-1,k}$ est le modèle et $\eta_{k-1,k}$ l'erreur modèle, dont la pdf détermine la pdf de transition. Dans le cadre du filtre de Kalman, cette erreur est considérée comme gaussienne, de moyenne nulle (pas de biais) et de matrice de covariance notée $\mathbf{Q}_{k-1,k}$. L'opérateur modèle $\mathcal{M}_{k-1,k}$ est supposé linéaire (et s'écrit alors $\mathbf{M}_{k-1,k}$).

Les observations sont également distribuées dans le temps et s'apparentent à une chaîne de Markov. Il faut donc connaître la pdf de transition $p(\mathbf{y}_k|\mathbf{x}_k)$ de ces observations. On peut formuler la description d'une observation à un instant k en fonction des observations précédentes par :

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \epsilon_k^o \quad (4.3)$$

où $\mathcal{H}_k(\mathbf{x}_k)$ est la projection de la solution modèle dans l'espace des observations, et ϵ_k^o l'erreur d'observation dont la pdf détermine la pdf de transition. Dans le cadre du filtre de Kalman, cette erreur est considérée comme gaussienne, de moyenne nulle et de matrice de covariance notée \mathbf{R}_k . L'opérateur d'observation \mathcal{H}_k est défini comme linéaire (et s'écrit alors \mathbf{H}_k).

Le filtrage peut intuitivement être exprimé par une succession d'étapes de prévision et d'analyse. En utilisant les règles de Bayes :

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \quad (4.4)$$

et de marginalisation :

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y})d\mathbf{y} = \int p(\mathbf{x}|\mathbf{y})p(\mathbf{y})d\mathbf{y} \quad (4.5)$$

ces étapes peuvent s'écrire :

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1} \quad \text{prévision} \quad (4.6)$$

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{x}_k|\mathbf{y}_{1:k-1})p(\mathbf{y}_k|\mathbf{x}_k) \quad \text{analyse} \quad (4.7)$$

avec $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ la pdf de transition entre le temps t_{k-1} et t_k , $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ la pdf d'analyse au temps t_{k-1} , $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ la pdf de prévision au temps t_k et enfin $p(\mathbf{y}_k|\mathbf{x}_k)$ la pdf l'observation au temps t_k .

4.2.2 Le filtre de Kalman

Dans un cas où les équations 4.2 et 4.3 sont linéaires et de bruit gaussien de moyenne nulle, les équations du filtrage peuvent s'écrire sous la forme définie par *Kalman* (1960) :

Initialisation	
	\mathbf{x}_0^a (4.8)
	\mathbf{P}_0^a (4.9)
Etape de prévision	
	$\mathbf{x}_{k k-1}^f = \mathbf{M}_{k-1,k} \mathbf{x}_{k-1 k-1}^a$ (4.10)
	$\mathbf{P}_{k k-1}^f = \mathbf{M}_{k-1,k} \mathbf{P}_{k-1 k-1}^a \mathbf{M}_{k-1,k}^T + \mathbf{Q}_{k-1,k}$ (4.11)
Etape d'analyse	
	$\mathbf{K}_{k k} = \mathbf{P}_{k k-1}^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k k-1}^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$ (4.12)
	$\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_{k k-1}^f$ (4.13)
	$\mathbf{x}_{k k}^a = \mathbf{x}_{k k-1}^f + \mathbf{K}_{k k} \mathbf{d}_k$ (4.14)
	$\mathbf{P}_{k k}^a = (\mathbf{I} - \mathbf{K}_{k k} \mathbf{H}_k) \mathbf{P}_{k k-1}^f$ (4.15)

TABLE 4.1 – Les différentes étapes du filtre de Kalman.

Les exposants f et a désignent respectivement les états liés à la prévision (forecast) ou à l'analyse. Bien qu'ils ne soient pas indispensables pour le filtre, les indices temporels plus spécifiques à la problématique 4D (définis dans *Cosme et al.*, 2010) sont introduits ici. Ils sont écrits sous la forme $X_{i|j}$, signifiant que l'état X est donné au temps i et tient compte de toutes les observations disponibles jusqu'au temps j (ces notations concernent uniquement les états dépendant des observations). Ces indices seront nécessaires dans la description du lisseur, nous préférons donc garder une cohérence des notations dans toute la description.

Bien que le filtre s'écrive en une succession de cycles où les conditions pour une analyse ou une prévision sont fournies par les étapes précédentes, il faut évidemment lui préciser des conditions initiales au temps t_0 . La connaissance de ces conditions, notamment la détermination des statistiques d'erreur initiales, n'est pas toujours aisée. Nous reviendrons sur cet aspect dans la section 5.2.4.

4.2.3 Les limites du filtrage

Bien que le filtre détermine *a priori* une solution optimale de l'état analysé, sa forme séquentielle propagative uniquement vers le futur limite quelque peu son potentiel. En effet, les observations utilisées lors d'une analyse ont un impact direct sur l'état analysé et

un impact indirect (par la propagation du modèle) sur les états ultérieurs. Leur influence spatiale est bien tri-dimensionnelle (grâce aux corrélations), en revanche la dimension temporelle n'est qu'à moitié représentée. L'influence du passé sur une observation (ou inversement) n'est pas prise en compte dans la formulation du filtre. Ainsi, dans le cas d'une dynamique présentant de fortes corrélations temporelles, les observations ne sont pas utilisées de façon optimale. En effet, le filtre ne tiendra pas compte à un temps $k - 1$ de l'information que pourrait lui apporter une observation au temps k . C'est pour palier à ce problème qu'une extension 4D du filtre a été développée, le *lissage*, dont nous allons présenter les différentes approches dans la section suivante. La forme de lissage choisie pour cette thèse sera détaillée dans la partie 4.4 et étudiée plus en détail à travers nos résultats.

4.3 Problème du lissage : les différentes approches

Nous nous plaçons maintenant dans le cas où des observations sont disponibles jusqu'à un temps K . Le lissage peut se traduire par une formulation Bayésienne (à comparer avec la formulation 4.1) qui consiste alors à déterminer la pdf d'un état \mathbf{x}_k compte tenu d'observation \mathbf{y} passées, présentes au temps k et futures jusqu'au temps K (avec $K > k$) :

$$p(\mathbf{x}_k | \mathbf{y}_{1:K}) \quad (4.16)$$

Dans cette formulation, toutes les observations disponibles sont destinées à analyser un unique état \mathbf{x}_k . Il s'agit alors d'un problème de *lissage marginal*. Dans le cas où l'on voudrait maintenant analyser toute une série d'états compte tenu des observations, cela reviendrait à déterminer la pdf suivante :

$$p(\mathbf{x}_{0:K} | \mathbf{y}_{1:K}) \quad (4.17)$$

Il s'agit dans ce cas d'un problème de *lissage conjoint*.

Les lisseurs conjoint et marginal ne sont *a priori* pas destinés aux mêmes applications. Les premiers sont de préférence appliqués à des problèmes d'estimations (de l'état de l'océan, des paramètres de flux...) sur toute une fourchette de temps, alors que les seconds permettent essentiellement de caler une condition initiale. De plus, les deux familles de lisseurs sont elles-mêmes déclinées en plusieurs formulations que nous allons maintenant décrire. Ces descriptions resteront plutôt qualitatives et graphiques (à l'aide de figures) afin de donner simplement un aperçu des différentes approches possibles. L'implémentation des différents lisseurs sera toujours décrite dans le cadre des hypothèses de Kalman. Par la suite, nous explorerons un seul type de lisseur dans ce manuscrit, que nous détaillerons dans la section 4.4. A noter également que nous garderons de préférence dans ces sections les noms anglais des lisseurs, plus généralement utilisés dans la communauté d'assimilateurs.

4.3.1 Le lissage conjoint

Cette famille de lisseurs se décline en 3 types : le *fixed-interval smoother* (lisseur à interval fixe), le *fixed-lag smoother* (lisseur à fenêtre fixe) et le *ensemble smoother* (lisseur d'ensemble).

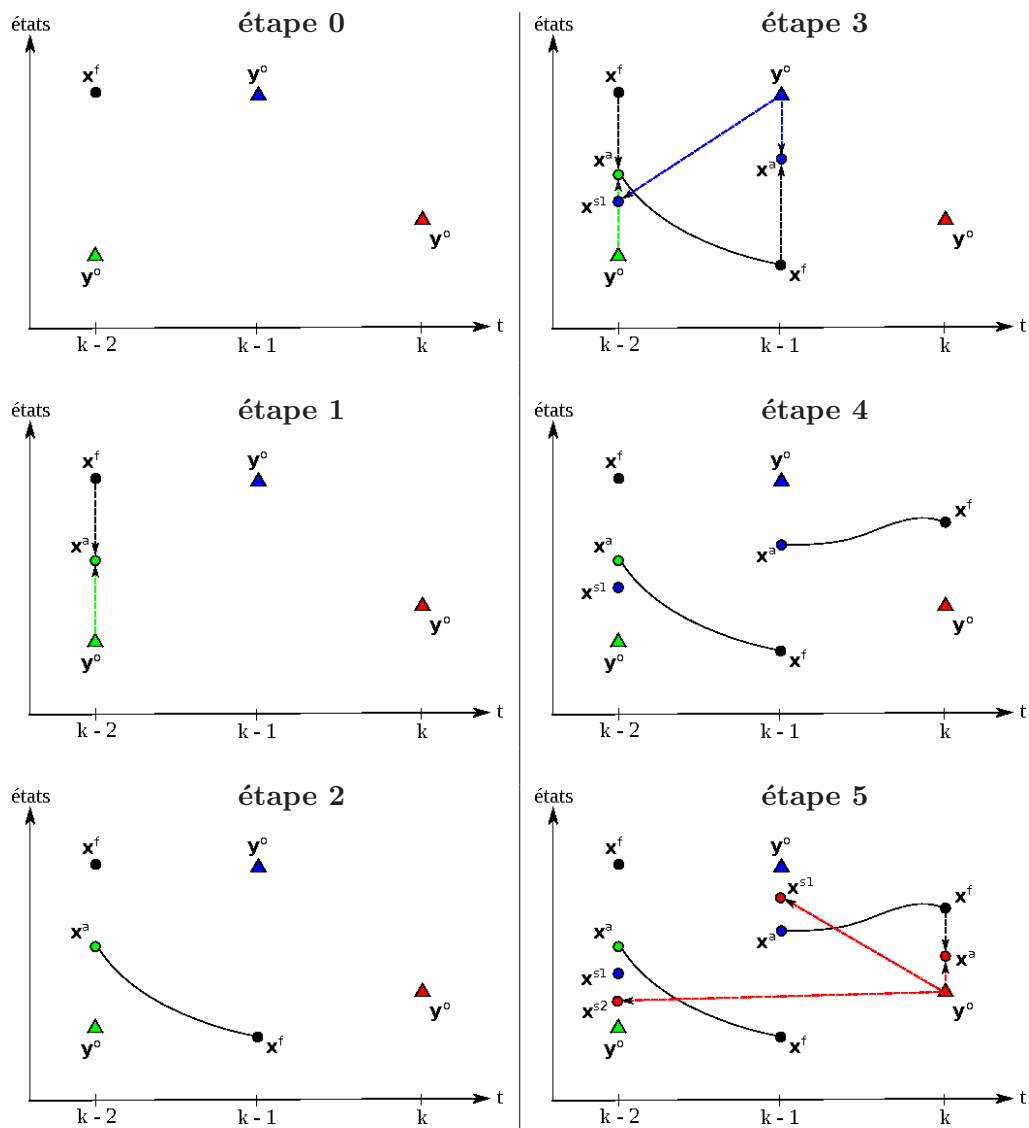


FIGURE 4.2 – Représentation schématique du lissage séquentiel de type *fixed-interval* ou *fixed-lag*. Les notations sont : y^o (triangles) pour les observations, x^a (ronds) pour les états analysés par le filtre, x^s (ronds) pour les états analysés par le lisseur et x^f (ronds noirs) pour les états prédits par le modèle. L'axe des abscisses représente différentes dates, avec k la date de la dernière observation disponible. **Étape 0** : au temps $k - 2$ un état modèle de background et une observation y^o (en vert) sont disponibles ; **Étape 1** : le filtre de Kalman détermine un état analysé x^a au temps $k - 2$; **Étape 2** : l'état analysé précédemment, ainsi que les statistiques issues de l'analyse, sont propagés par le modèle jusqu'au temps $k - 1$ où une nouvelle observation y^o (en bleu) est disponible ; **Étape 3** : le filtre effectue une analyse au temps $k - 1$, et le lisseur peut utiliser l'observation y^o (en bleu) pour réanalyser l'état x^a au temps $k - 2$; **Étape 4** : l'état analysé par le filtre au temps $k - 1$, ainsi que les statistiques issues de l'analyse, sont propagés par le modèle jusqu'au temps k où une nouvelle observation y^o (en rouge) est disponible ; **Étape 5** : le filtre effectue une analyse au temps k , et le lisseur peut utiliser l'observation y^o (en rouge) pour réanalyser les états x^a et x^{s1} respectivement aux temps $k - 1$ et $k - 2$.

Fixed-interval smoother

Ce lisseur fonctionne séquentiellement et se greffe directement sur le filtre (se référer à la figure 4.2). Lorsqu'une observation est introduite au temps t_k , le filtre effectue une analyse seulement à ce temps là. Le lisseur va lui effectuer des analyses rétrospectives (du temps t_{k-1} jusqu'au temps t_1). Pour ce faire, il doit avoir accès à deux types d'information. Le premier type d'informations concerne les statistiques d'erreur issues des analyses antérieures de chaque date (analyses effectuées par un filtrage ou un lissage précédent). Il s'agit par exemple des matrices \mathbf{P}^a dans la formulation du filtre de Kalman. Pour la seconde source d'information, il s'agit de tenir compte de la distance temporelle séparant l'observation utilisée (au temps t_k) et la date de l'analyse. Cette information est fournie par la propagation dynamique des statistiques d'erreur du filtrage (matrices \mathbf{P}^f dans le filtre de Kalman). Le lisseur procède de la même façon pour chaque nouvelle observation introduite aux temps t_{k+1}, \dots, t_K . Les analyses rétrospectives sont effectuées sur toute la durée de l'expérience (de t_1 à t_K).

Bien que le lisseur, dans une formulation d'ensemble, ne présente pas de coût de calcul supplémentaire de propagation dynamique par rapport au filtre (puisque toutes les informations dont il a besoin pour effectuer les analyses sont déjà disponibles suite au filtrage), le coût en stockage peut vite devenir élevé si l'expérience est longue (car il faut stocker toutes les statistiques issues du lissage). De plus, le coût des analyses du lisseur peut également devenir élevé puisque pour chaque nouvelle observation, il y a autant d'analyses rétrospectives à effectuer qu'il y a d'états antérieurs. Enfin, notons que cette formulation peut bénéficier des mêmes simplifications que le filtre pour limiter le coût de calcul (réduction d'ordre des covariances d'erreur...), aspect qui sera traité dans le chapitre suivant.

Fixed-lag smoother

Ce lisseur est proche du fixed-interval dans son approche, mais cette fois la fenêtre temporelle des analyses rétrospectives est de longueur constante, quelque soit la durée de l'expérience (la figure le représentant est la même que celle du fixed-interval). Cela signifie que chaque nouvelle observation introduite aura une influence limitée dans le passé. Cette limite peut évidemment être définie selon les souhaits de l'utilisateur. La pdf correspondant à ce lisseur peut s'écrire (pour des observations allant de 1 à k) :

$$p(\mathbf{x}_{k-L:k} | \mathbf{y}_{1:k}) \quad (4.18)$$

avec L définissant la taille de la fenêtre temporelle des analyses du lisseur.

Le fixed-lag smoother présente les mêmes avantages que le fixed-interval (il se greffe directement sur le filtre, il est peu coûteux numériquement par rapport au filtre et peut bénéficier de la réduction d'ordre). Il possède en plus l'avantage de limiter le coût du stockage puisque le nombre d'analyses rétrospectives est maintenant limité. Cet avantage le rend particulièrement intéressant dans le cadre de réanalyses (généralement effectuées sur de longues périodes).

Ensemble smoother

Ce lisseur a été introduit par *van Leeuwen et Evensen* (1996). Il consiste (figure 4.3) à utiliser un run d'ensemble pour constituer un ensemble de vecteurs d'état étendu (4D). Ce

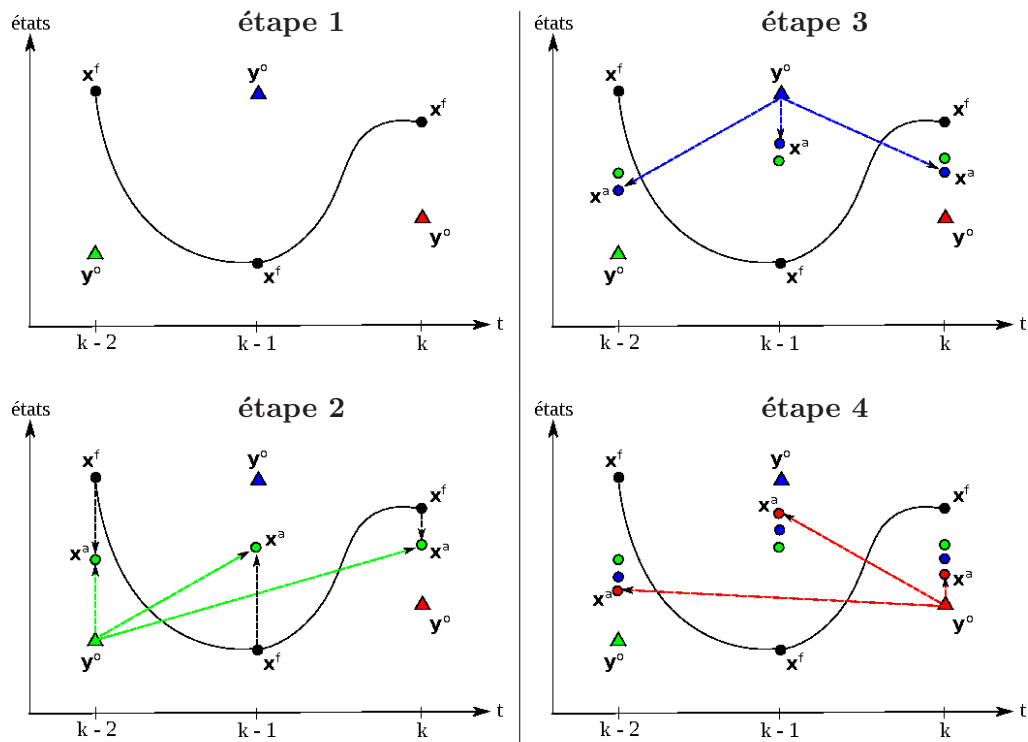


FIGURE 4.3 – Représentation schématique de l'*ensemble smoother*, dans une version séquentielle. Les notations sont les mêmes que pour la figure 4.2 excepté que tous les états analysés sont notés x^a . **Étape 1** : une expérience d'ensemble (sans assimilation) est menée sur toute la fenêtre temporelle de $k-2$ à k ; **Étape 2** : la première observation disponible au temps $k-2$ est utilisée pour analyser les états sur toute la durée de l'expérience (les états x^a sont donc des états issus du lissage et non pas du filtrage); **Étape 3** : la deuxième observation disponible au temps $k-1$ est à son tour utilisée pour analyser les états sur toute la durée de l'expérience; **Étape 4** : les étapes d'analyses sont répétées jusqu'à la dernière observation disponible (au temps k).

vecteur est ensuite analysé soit de façon globale soit séquentiellement. Dans le premier cas, le vecteur d'observation doit être étendu de la même façon selon l'axe temporel (constituant ainsi un vecteur d'observations 4D), puis une analyse BLUE (Best Linear Unbiased Estimator) est effectuée sur le vecteur d'état étendu. Dans le second cas (qui suppose des erreurs d'observations décorréliées dans le temps), les observations sont introduites une à une et servent à analyser successivement le vecteur d'état étendu.

Ce lisseur présente l'avantage d'avoir un coût comparable, voire légèrement moindre par rapport à un filtre d'ensemble. De plus, une fois que l'ensemble de base est défini, il est possible d'effectuer autant d'expériences de lissage que voulu sans nouvelle itération du modèle, rendant cette formulation particulièrement intéressante pour des études de réseaux d'observations par exemple (OSSE). Notons également que ce lisseur peut bénéficier d'une simplification du type décomposition en racine carrée. De plus, si initialement il fonctionne sur le même principe que le fixed-interval smoother (une observation ayant une influence sur toute la durée de l'expérience), il est possible de localiser temporelle l'influence des observations au même titre que le fixed-lag smoother, mais cela n'a encore jamais été étudié.

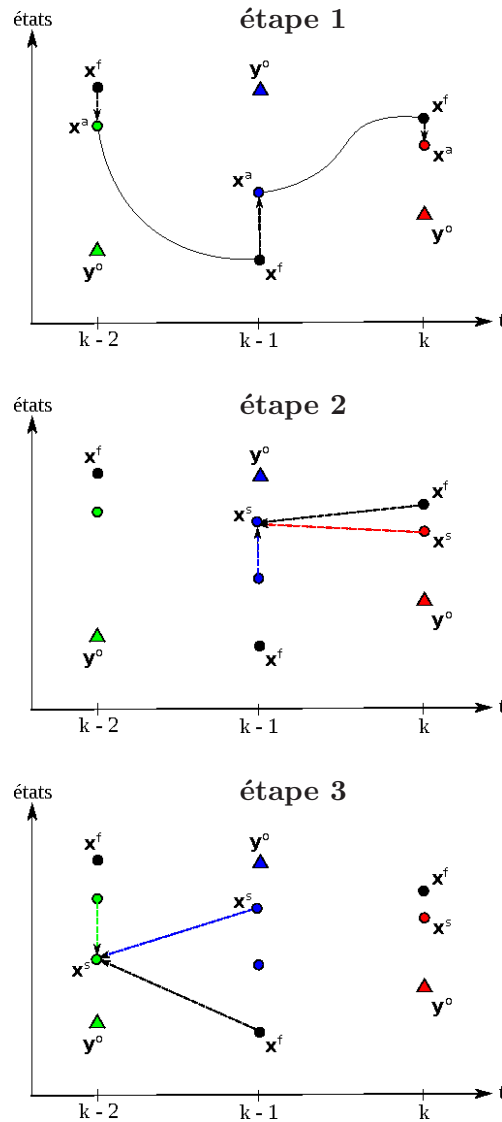


FIGURE 4.4 – Représentation schématique du *forward-backward smoother*. Les notations sont les mêmes que pour la figure 4.2. **Étape 1** : un filtrage est effectué sur toute la fenêtre temporelle, du temps $k-2$ jusqu'à la dernière observation disponible au temps k ; **Étape 2** : le dernier état analysé par le filtre peut être considéré comme le premier état lissé ($x^a = x^s$ au temps k), le lisseur rétropropage alors l'information extraite de la dernière analyse (d'après les états prédit x^f et analysé x^s) et l'utilise pour réanalyser l'état filtré précédent (en bleu), donnant ainsi x^s au temps $k-1$; **Étape 3** : le lissage rétropropage de nouveau l'information en utilisant les états prédit x^f et lissé x^s au temps $k-1$ pour réanalyser l'état filtré précédent (en vert), donnant ainsi x^s au temps $k-2$.

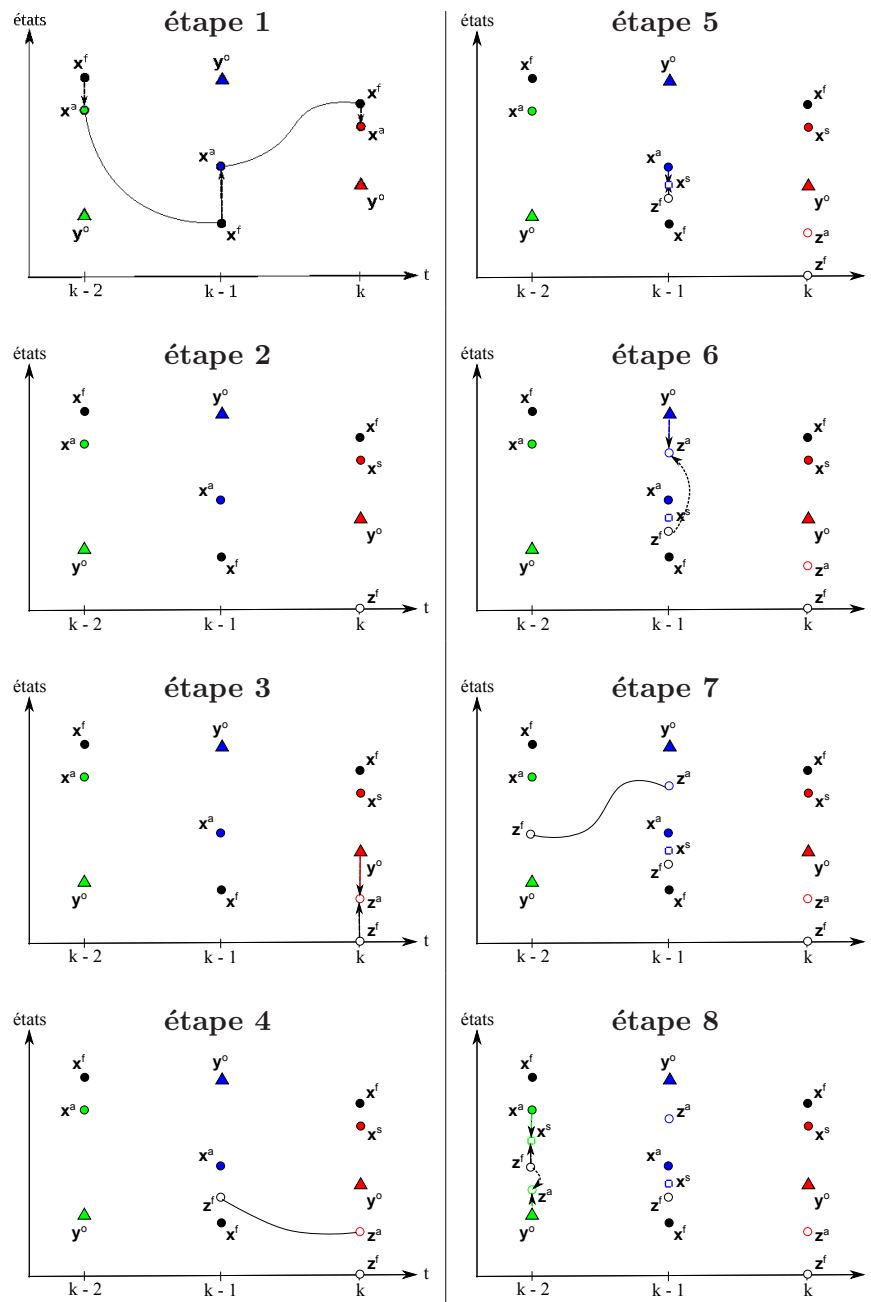


FIGURE 4.5 – Représentation schématique du *two-filter smoother*. Les notations sont les mêmes que pour la figure 4.2, avec en plus les états \mathbf{z} correspondant au filtrage rétrograde. **Étape 1** : un filtrage est effectué sur toute la fenêtre temporelle, du temps $k - 2$ jusqu'à la dernière observation disponible au temps k ; **Étape 2** : au temps k , on utilise $\mathbf{P} = +\infty$ (non informatif), et l'on définit $\mathbf{N} = \mathbf{P}^{-1}$ et $\mathbf{z} = \mathbf{P}^{-1}\mathbf{x}$, qui permettent d'initialiser le filtre rétrograde avec $\mathbf{N} = 0$ et $\mathbf{z} = 0$; d'après un état z^f contenant une information non-descriptive ou nulle; **Étape 3** : la première analyse du filtre rétrograde (z^a) est effectuée d'après l'état z^f et l'observation y^o au temps k ; **Étape 4** : l'état z^a est propagé par le modèle adjoint jusqu'au temps précédent $k - 1$, donnant ainsi z^f ; **Étape 5** : une analyse est effectuée par le filtre rétrograde d'après l'état analysé par le filtre direct au temps $k - 1$ et l'état z^f au même temps, donnant ainsi un état lissé a^s ; **Étape 6** : le même état z^f est également analysé directement avec les observations du temps $k - 1$, donnant un nouvel état analysé du filtre rétrograde z^a ; **Étape 7** : cet état z^a est de nouveau propagé par le modèle adjoint jusqu'au temps $k - 2$; **Étape 8** : de nouveau, des analyses permettent d'obtenir un état lissé x^s et un état filtré z^a au temps $k - 2$.

4.3.2 Le lissage marginal

Cette famille de lisseurs se décline en 2 types : le *forward-backward smoother* (lisseur en “aller-retour”) et le *two-filter smoother* (lisseur à double filtre).

Forward-backward smoother

Ce lisseur, appelé aussi RTS (Rauch-Tung-Striebel) smoother, a été introduit dans un cadre linéaire gaussien par *Rauch et al.* (1965). Comme son nom l’indique, ce lisseur fonctionne selon un principe d’aller-retour (figure 4.4). L’aller correspond simplement à un filtrage (du temps 1 au temps k). Le retour commence par la réanalyse de l’état issu du filtrage au temps $k - 1$ en tenant compte de l’analyse effectuée au temps k . Puis une seconde analyse du lisseur est effectuée au temps $k - 2$ en tenant compte de l’analyse du lissage au temps $k - 1$, et ainsi de suite jusqu’à rétropropager l’information jusqu’au premier état. La seule difficulté de ce lisseur est qu’il nécessite l’inversion des matrices de covariances d’erreur de prévision, mais cette difficulté peut être contournée, par exemple par une décomposition en racine carrée (*Cosme et al.*, in press).

Ce lisseur a été le plus utilisé en géophysique. Il présente l’avantage de ne pas nécessiter d’autres itérations modèle : toutes ses analyses rétrospectives sont purement statistiques et se basent sur celle du filtre. Il peut également bénéficier d’une formulation d’ensemble (décomposition en racine-carrée...) mais il reste tout de même plus coûteux que le filtre.

Two-filter smoother

Ce lisseur a été introduit dans un cadre linéaire gaussien par *Fraser et Potter* (1969). Il consiste, de façon simplifiée (figure 4.5), à utiliser dans un premier temps un filtre propagatif jusqu’à la dernière date d’observation, puis, dans un second temps, à utiliser un filtre inverse pour rétropropager les informations jusqu’au temps 1. Cette deuxième étape implique l’utilisation d’un modèle adjoint.

A cause de sa complexité, ce lisseur n’a pas été utilisé dans le cadre de problèmes géophysiques. Son coût (au moins le double d’un filtre classique) et la nécessité de développer un modèle adjoint sont prohibitifs. De plus, ce lisseur ne peut *a priori* pas bénéficier d’une formulation d’ensemble, bien que cette option n’ait jamais été étudiée en profondeur.

4.3.3 Les thématiques d’application du lissage

A travers la description simplifiée des différents lisseurs, nous avons pu voir quels étaient les atouts et les intérêts de chacun. Tous permettent d’étendre l’influence d’une observation à une dimension 4D, mais avec des approches différentes et plus ou moins complexes. La figure 4.6 récapitule de façon simplifiée quelles sont les applications privilégiées pour chaque type de lisseur. En effet, si l’extension 4D de l’influence des observations semble particulièrement intéressante pour les réanalyses, cette extension peut également être avantageuse pour d’autres thématiques. Par exemple, l’estimation de paramètres liés aux forçages ou à des processus chimiques, l’optimisation de réseaux d’observations ou encore l’estimation d’états initiaux (notamment dans le cas du lissage marginal).

Dans l’approche que nous avons adoptée dans cette thèse, nous avons choisi le lisseur de type *fixed-lag*. D’une part, comme nous allons le voir plus en détail par la suite, son

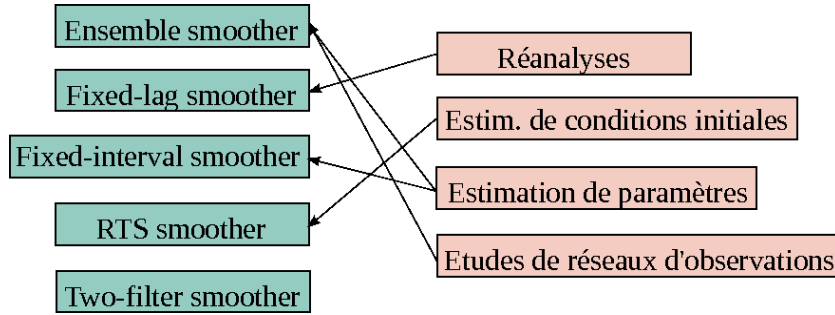


FIGURE 4.6 – Récapitulatif des applications privilégiées propres à chaque lisseur.

implémentation est très facile lorsqu’un filtre est déjà en place. De plus, son application privilégiée pour les réanalyses et son coût relativement faible sont tout à fait appropriés pour nos expériences. Nous allons donc à présent définir plus précisément les caractéristiques de ce lisseur et donner notamment ses équations.

A noter que, par la suite, lorsque nous parlerons du “lisseur” cela concernera uniquement le fixed-lag smoother.

4.4 Le lisseur fixed-lag

Les indices et les notations utilisées sont celles définies dans *Cosme et al. (2010)* et rappelées ici. On se place dans le cas où un filtre de Kalman a analysé les états du modèle jusqu’au temps t_{k-1} . Une nouvelle observation est alors disponible au temps t_k .

Le lisseur peut corriger un état à un temps t_i antérieur à t_k ($t_i \leq t_k$) avec les observations disponibles jusqu’à t_k ($1 : k$). L’état ainsi analysé au temps t_i est noté $x_{i|k}^a$. Bien entendu, avec une observation au temps t_k il est possible de faire plusieurs analyses antérieures pour différentes dates t_i . On note donc Σ_k l’ensemble des indices i des différentes analyses rétrospectives produites avec l’observation de la date t_k . Dans le cas où on aurait $\Sigma_k = \{k\}$, on retomberait sur la formulation du filtrage (c’est-à-dire l’observation est utilisée pour analyser l’état au temps t_k uniquement). Dans le cadre du lisseur fixed-lag, Σ_k est une série d’indices dont le nombre d’éléments L est fixe mais dont les éléments eux-mêmes dépendent de l’indice k . Ces éléments précèdent l’indice de la date d’observation k ($\Sigma_k = \{k - L, \dots, k - 1\}$). Il s’agit donc en quelque sorte d’un “lisseur glissant”.

Pour effectuer une analyse, le lisseur doit avoir accès à différentes statistiques d’erreur. D’une part, il faut connaître les erreurs de prévision au temps t_k (notées $\mathbf{e}_{k|k-1}^f$), c’est-à-dire au temps où une nouvelle observation est disponible. D’autre part, pour analyser un état au temps t_i (avec $t_i < t_k$), il faut connaître les erreurs résiduelles estimées à cette date, que nous notons $\mathbf{e}_{i|k-1}^a$. Ces dernières tiennent compte de toutes les observations disponibles jusqu’au temps t_{k-1} , mais pas encore de l’observation au temps t_k . Le lisseur utilise alors des matrices de covariances d’erreur croisées définies par :

$$\mathbf{P}_{k,i|k-1}^{fa} = E[\mathbf{e}_{k|k-1}^f \mathbf{e}_{i|k-1}^a{}^T] \quad (4.19)$$

$$\mathbf{P}_{k,i|k}^{aa} = E[\mathbf{e}_{k|k}^a \mathbf{e}_{i|k}^a{}^T] \quad (4.20)$$

avec E l’opérateur d’espérance, $\mathbf{P}_{k,i|k-1}^{fa}$ la matrice des covariances croisées d’erreur de prévision (au temps t_k) et d’analyse (au temps t_i), et $\mathbf{P}_{k,i|k}^{aa}$ la matrice des covariances

croisées d'erreur d'analyse. Ces matrices permettent de mettre en lien les corrections passées (t_i) et les informations futures (t_k). Elles interviennent dans le calcul du gain du lisseur (équation 4.24).

Les éléments définis ci-dessus permettent maintenant d'exposer la formulation générale du lisseur.

4.4.1 Formulation du lisseur de Kalman

Le lissage se superpose au filtrage et s'effectue donc en deux étapes : celle de la prévision et celle de l'analyse. La propagation de l'état $\mathbf{x}_{k-1|k-1}^a$ par le modèle entre le temps t_{k-1} et t_k s'effectue comme dans le cas du filtre (équation 4.21). En revanche, la propagation des covariances d'erreur est divisée en deux équations (équations 4.22 et 4.23) car les matrices des covariances croisées (équation 4.22) sont nécessaires pour l'analyse du lisseur par la suite. L'étape d'analyse du filtre n'est quant à elle pas modifiée. Seules les équations liées au lissage se rajoutent à cette étape. Les équations du lisseur sont résumées ci-dessous :

Etape de prévision	
$\mathbf{x}_{k k-1}^f$	$= \mathbf{M}_{k-1,k} \mathbf{x}_{k-1 k-1}^a \quad (4.21)$
$\mathbf{P}_{k,i k-1}^{fa}$	$= \mathbf{M}_{k-1,k} \mathbf{P}_{k-1,i k-1}^{aa}, \quad i \in \Sigma_k \quad (4.22)$
$\mathbf{P}_{k k-1}^f$	$= \mathbf{M}_{k-1,k} (\mathbf{P}_{k,k-1 k-1}^{fa})^T + \mathbf{Q}_{k-1 k} \quad (4.23)$
Etape d'analyse du lisseur	
$\mathbf{K}_{i k}$	$= \mathbf{P}_{k,i k-1}^{fa} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k k-1}^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (4.24)$
$\mathbf{x}_{i k}^a$	$= \mathbf{x}_{i k-1}^a + \mathbf{K}_{i k} \mathbf{d}_k, \quad i \in \Sigma_k \quad (4.25)$
$\mathbf{P}_{k,i k}^{aa}$	$= (\mathbf{I} - \mathbf{K}_{i k} \mathbf{H}_k) \mathbf{P}_{k,i k-1}^{fa}, \quad i \in \Sigma_k \quad (4.26)$
$\mathbf{P}_{i k}^a$	$= \mathbf{P}_{i k-1}^a - \mathbf{K}_{i k} \mathbf{H}_k \mathbf{P}_{k,i k-1}^{fa}, \quad i \in \Sigma_k \quad (4.27)$

TABLE 4.2 – Equations du lisseur de Kalman.

L'équation 4.27 est purement diagnostique, elle n'est pas réutilisée dans les analyses du lisseur.

4.4.2 L'erreur modèle indispensable au lissage

Les équations ci-dessus décrivent donc la théorie du lissage séquentiel qui permet une représentation 4D des observations (en rajoutant la dimension *passée* par rapport au temps).

Les corrélations des erreurs dans la dimension temporelle sont intrinsèquement liées à la dynamique des systèmes et à ses incertitudes. Un modèle ayant des erreurs importantes et décorréelées dans le temps doit conduire à une décorrélation des erreurs dans le temps.

Pour que ce processus soit correctement représenté, il est théoriquement indispensable d’inclure la dynamique et l’erreur du modèle, donc d’utiliser la partie “évolutive” du filtre dans sa forme originale (équation 4.11), et non sous une forme simplifiée du type interpolation optimale. Nous discuterons d’ailleurs en détail de ce point plus tard dans cette thèse (chapitre 9). De plus, le terme d’erreur modèle \mathbf{Q} est généralement nécessaire pour décorréler de façon réaliste les erreurs dans le temps, à condition que \mathbf{Q} soit elle-même décorrélée dans le temps. Ce terme permet d’introduire explicitement des directions de décorrélation et de contrebalancer une propagation dynamique des modes éventuellement trop linéaire (et donc corrélée).

Dans un cas où l’on négligerait les deux termes de la définition de \mathbf{S}^f (terme évolutif et erreur modèle), cela reviendrait à ne plus prendre en compte la distance temporelle séparant une observation de la date d’analyse du lisseur (autrement dit, cela reviendrait à considérer que les observations, normalement futures, sont mesurées à la date de l’analyse). Dans de telles conditions, il est facile d’imaginer que le lissage pourrait dégrader la solution analysée.

En pratique, ces deux points ne sont pas forcément aisés à implémenter. Nous reviendrons sur ces aspects dans les parties 5.3.3 et 5.3.4 du chapitre suivant.

4.5 Conclusion

Ce chapitre a permis de définir les grandes lignes d’une méthode d’assimilation de données statistique (filtre de Kalman) et ses limites quant à une utilisation optimale des observations. Nous avons exposé diverses méthodes de lissage pour répondre à ces limites et notre choix s’est porté sur un lisseur, le fixed-lag smoother, dont nous avons exposé plus précisément les caractéristiques et la théorie. Le filtre est largement utilisé dans la communauté océanographique, mais le lissage reste peu courant, en partie à cause de la paramétrisation des covariances d’erreurs qui sont rarement optimales pour le lisseur (aspect que nous évoquerons dans la section 5.3.3 du chapitre suivant). L’implémentation du lisseur dans une configuration réaliste représente donc un enjeu important.

Ces méthodes, filtre ou lisseur, ne peuvent néanmoins pas être utilisées directement avec leur formulation décrite précédemment dans le cadre d’un problème géophysique telle que l’océanographie. Cela est dû aux incertitudes sur la connaissance des matrices de covariances d’erreur, mais aussi et surtout à la taille de ces matrices. La mise en place pratique des méthodes statistiques a donc nécessité le développement d’une autre formulation que nous allons définir dans le chapitre suivant.

Chapitre 5

Formulations SEEK du filtre et du lisseur de Kalman

Sommaire

5.1	Introduction	70
5.2	Mise en place du filtre et du lisseur de Kalman dans un environnement réaliste : développement du SEEK	70
5.2.1	La réduction d'ordre	70
5.2.2	Formulation SEEK du filtre	71
5.2.3	Formulation SEEK du lisseur	72
5.2.4	Résumé des équations du filtre et du lisseur SEEK	74
5.3	Paramétrisations du filtre et du lisseur SEEK	75
5.3.1	Initialisation : l'état \mathbf{x}_0^a	75
5.3.2	Initialisation : \mathbf{S}_0^a et méthode de réduction de la matrice des covariances d'erreur	75
5.3.3	Evolutivité des covariances d'erreur : les différents types de filtres	76
5.3.4	La prise en compte de l'erreur modèle \mathbf{Q}	78
5.3.5	Localisation de l'analyse	79
5.4	Implémentation du SEEK dans NEMO : le logiciel SESAM	81
5.5	Quelques précisions supplémentaires	82
5.5.1	Vecteur d'état et vecteur d'observation	82
5.5.2	Un outil de diagnostic : le calcul d'erreur RMS	82
5.6	Conclusion	83

5.1 Introduction

Nous avons précédemment exposé la théorie du filtre de Kalman et du lisseur associé. Néanmoins, les formulations théoriques telles qu'elles ont été énoncées ne peuvent pas être appliquées directement à un problème réaliste de géophysique et ce à cause de la taille des systèmes. En effet, dans un problème géophysique (océanographie ou météorologie par exemple), la dimension des vecteurs d'état n peut être de l'ordre de 10^6 ou bien plus. Cette dimension dépend essentiellement de la résolution du modèle considéré et de la zone étudiée. Les matrices associées à la taille du système (\mathbf{P}^f et \mathbf{P}^a par exemple) sont alors de dimension $n \times n$. Les calculs associés à ces matrices et l'espace nécessaire pour les stocker s'avèrent beaucoup trop importants compte tenu de la puissance des outils de calcul. Le stockage est encore plus problématique dans le cadre du lissage puisqu'il est nécessaire de stocker des matrices de covariances croisées supplémentaires. La version SEEK (Singular Evolutive Extended Kalman) du filtre a été développée par *Pham et al.* (1998) dans le but de parer à ce problème. Son implémentation dans les problèmes d'océanographie a déjà été largement renseignée (*Brasseur et Verron*, 2006 et *Rozier et al.*, 2007 par exemple). Il existe d'autres méthodes de réduction d'ordre, comme par exemple le filtre de Kalman d'Ensemble (EnKF : Ensemble Kalman Filter) introduit par *Evensen* (1994), le filtre racine carrée de rang réduit (RRSQRT : Reduced-Rank Square-Root filter) développé par *Verlaan et Heemink* (1997) ou encore l'algorithme d'estimation statistique du sous-espace d'erreur (ESSE : Error Subspace Statistical Estimation) présenté par *Lermusiaux et Robinson* (1999). Chacune de ces méthodes a également été associée à la mise en place du lissage selon différentes formulations. Citons par exemple le cas du lisseur d'ensemble (EnKF) testé par *van Leeuwen* (1999) et *van Leeuwen* (2001) avec un modèle 2-couches quasigéostrophique. *Lermusiaux et Robinson* (1999) ont quant à eux mis en place en lisseur s'apparentant à un RTS d'après le filtre ESSE, testé par la suite dans un contexte réaliste d'assimilation haute-résolution (*Lermusiaux*, 1999a; *Lermusiaux*, 1999b et *Lermusiaux et al.*, 2002). *Cosme et al.* (2010) ont implémenté un lisseur SEEK dans une configuration idéalisée boîte carrée de la circulation océanique.

Nous allons nous concentrer dans ce chapitre sur le schéma SEEK (section 5.2) et redéfinir les formulations associées du filtre et du lisseur (section 5.2.4). Nous évoquerons également diverses paramétrisations nécessaires à l'implémentation de la version SEEK (section 5.3). Nous terminerons en précisant la mise en place technique de ces algorithmes grâce au logiciel SESAM (section 5.4).

5.2 Mise en place du filtre et du lisseur de Kalman dans un environnement réaliste : développement du SEEK

Commençons par aborder l'hypothèse de la réduction d'ordre.

5.2.1 La réduction d'ordre

L'objectif principal de la réduction d'ordre est de permettre la mise en œuvre numérique d'un filtre (et par extension d'un lisseur) dans le cadre de problèmes de grande dimension, tout en conservant les caractéristiques principales de ces schémas d'assimilation. Au-delà de l'aspect de coût de calcul, d'autres arguments justifient également la réduction d'ordre. D'une part, il a été démontré que le système océan peut être considéré comme un système

dynamique possédant un attracteur de dimension finie (*Pham et al.*, 1998). Pour les équations de Navier-Stokes, l'existence de cet attracteur a été démontrée dans *Lions et al.* (1997). La géostrophie en est l'une des propriétés dominantes, ce qui permet d'imposer des contraintes fortes entre certaines variables. Il s'avère ainsi inutile de vouloir corriger le modèle selon des directions qui vont à l'encontre de la nature attractive du système. D'autre part, le nombre d'observations disponibles est généralement très insuffisant pour décrire complètement les statistiques d'erreur du système. L'utilisation d'un filtre de Kalman de rang complet est donc superflue (*Cane et al.*, 1996).

5.2.2 Formulation SEEK du filtre

La réduction d'ordre passe d'abord par la ré-écriture des équations grâce à une décomposition racine carrée de la matrice des covariances d'erreur : $\mathbf{P}^f = \mathbf{S}^f \mathbf{S}^{fT}$. La matrice \mathbf{P} reste de dimension $n \times n$, mais la matrice \mathbf{S}^f est de même dimension initialement mais peut être tronquée pour devenir de dimension $n \times r$, avec dans l'idéal $r \ll n$ comme nous le verrons un peu plus loin. Le système reste cohérent avec la dimension du vecteur d'état. Les colonnes de \mathbf{S}^f contiennent les r premiers modes de \mathbf{P}^f , c'est-à-dire les informations considérées comme les plus représentatives de la matrice complète. Il existe différentes approches pour déterminer les composantes principales de \mathbf{P} . Nous décrirons celles que nous allons utiliser au cours de nos travaux dans la section 5.3.2. Avec la décomposition en racine carrée, les équations du filtre peuvent être reformulées comme suit (les indices utilisés sont les mêmes que dans le cadre du lisseur).

Etape de prévision

$$\mathbf{x}_{k|k-1}^f = \mathbf{M}_{k-1,k} \mathbf{x}_{k-1|k-1}^a \quad (5.1)$$

$$\mathbf{P}_{k|k-1}^f = (\mathbf{M}_{k-1|k} \mathbf{S}_{k-1|k-1}^a) (\mathbf{M}_{k-1|k} \mathbf{S}_{k-1|k-1}^a)^T + \mathbf{Q}_{k-1|k} \quad (5.2)$$

Dans cette formulation, le nombre d'intégrations du modèle à effectuer entre les étapes $k-1$ et k pour propager les covariances d'erreur devient simplement égal au nombre de colonnes de la matrice \mathbf{S}^a , c'est-à-dire r , +1 pour la propagation de l'état. En revanche, la matrice $\mathbf{Q}_{k-1|k}$ reste problématique, non pas à cause de sa dimension mais plutôt par le manque de connaissance la concernant. Nous reviendrons sur cet aspect dans la section décrivant les différentes paramétrisations nécessaires pour le filtre (et le lisseur) SEEK (section 5.3.4).

Etape d'analyse

$$\mathbf{\Gamma}_k = (\mathbf{H}_k \mathbf{S}_{k|k-1}^f)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{S}_{k|k-1}^f) \quad (5.3)$$

$$\mathbf{K}_k = \mathbf{S}_{k|k-1}^f [\mathbf{I} + \mathbf{\Gamma}_k]^{-1} (\mathbf{H}_k \mathbf{S}_{k|k-1}^f)^T \mathbf{R}_k^{-1} \quad (5.4)$$

$$\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_{k+1} \mathbf{x}_{k|k-1}^f \quad (5.5)$$

$$\mathbf{x}_{k|k}^a = \mathbf{x}_{k|k-1}^f + \mathbf{K}_{k|k} \mathbf{d}_k \quad (5.6)$$

$$\mathbf{S}_{k|k}^a = \mathbf{S}_{k|k-1}^f (\mathbf{I} + \mathbf{\Gamma}_k)^{-1/2} \quad (5.7)$$

Dans cette formulation, deux conditions doivent être vérifiées pour que l'algorithme soit relativement peu coûteux.

D'une part, \mathbf{R}_k doit être inversible et à moindre coût. Une hypothèse simple est de considérer cette matrice diagonale (c'est-à-dire de considérer que les erreurs sur les observations ne sont pas corrélées entre elles). Cette hypothèse n'est pas toujours justifiée dans le cas où un même appareil effectue plusieurs mesures (une erreur sur l'appareillage se répercute donc sur chacune des mesures, il existe donc bel et bien une corrélation entre les erreurs de mesure). Néanmoins l'approximation diagonale demeure largement utilisée. D'autres méthodes ont été mises en place pour garder l'avantage d'une matrice diagonale tout en paramétrant des corrélations entre les erreurs (notamment *Testut et al.*, 2003; *Brankart et al.*, 2003 et *Brankart et al.*, 2009).

Le deuxième terme dont l'inversion peut s'avérer délicate est $\mathbf{I} + \mathbf{\Gamma}_k$. En rang complet, cette matrice est de dimension $n \times n$ ce qui est beaucoup trop grand pour un calcul efficace. Dans la formulation SEEK, grâce à la réduction d'ordre, elle est de dimension $r \times r$.

5.2.3 Formulation SEEK du lisseur

Pour ré-écrire les équations du lisseur dans une forme SEEK, nous allons considérer les sorties d'un filtre au temps t_{k-1} (un état analysé $x_{k-1|k-1}^a$ et les covariances d'erreur associées $\mathbf{S}_{k-1|k-1}^a$ sont donc disponibles). On considère alors une nouvelle observation disponible au temps t_k . On se place dans une situation où l'erreur modèle est négligée. Cela permet de simplifier les équations mais surtout permet de rester cohérent avec le fait que nous n'utiliserons pas directement de terme d'erreur modèle dans nos expériences.

Etape de prévision

La décomposition en racine carrée permet de ré-écrire l'équation de prévision du lisseur comme suit :

$$\mathbf{x}_{k|k-1}^f = \mathbf{M}_{k-1,k} \mathbf{x}_{k-1|k-1}^a \quad (5.8)$$

$$\mathbf{P}_{k,k-1|k-1}^{fa} = \mathbf{M}_{k-1,k} \mathbf{S}_{k-1|k-1}^a \mathbf{S}_{k-1|k-1}^{aT} = \mathbf{S}_{k|k-1}^f \mathbf{S}_{k-1|k-1}^{aT} \quad (5.9)$$

$$\mathbf{P}_{k|k-1}^f = \mathbf{M}_{k-1,k} \mathbf{S}_{k-1|k-1}^a \mathbf{S}_{k|k-1}^{fT} = \mathbf{S}_{k|k-1}^f \mathbf{S}_{k|k-1}^{fT} \quad (5.10)$$

où $\mathbf{S}_{k|k-1}^f = \mathbf{M}_{k-1,k} \mathbf{S}_{k-1|k-1}^a$ est la matrice de covariance d'erreur de prévision du filtre au temps t_k . On peut noter que la matrice des covariances croisées ($\mathbf{P}_{k,k-1|k-1}^{fa}$) est entièrement déterminée par les sorties du filtre.

Etape d'analyse

Les différents termes de l'analyse du lisseur (équations 4.24 à 4.27) peuvent être reformulés comme suit :

$$\begin{aligned} \mathbf{K}_{k-1|k} &= (\mathbf{H}_k \mathbf{S}_{k|k-1}^f \mathbf{S}_{k-1|k-1}^{aT})^T \mathbf{G}_k^{-1} \\ &= \mathbf{S}_{k-1|k-1}^a (\mathbf{H}_k \mathbf{S}_{k|k-1}^f)^T \mathbf{G}_k^{-1} \end{aligned}$$

$$= \mathbf{S}_{k-1|k-1}^a [\mathbf{I} + \mathbf{\Gamma}_k]^{-1} (\mathbf{H}_k \mathbf{S}_{k|k-1}^f)^T \mathbf{R}_k^{-1} \quad (5.11)$$

$$\mathbf{x}_{k-1|k}^a = \mathbf{x}_{k-1|k-1}^a + \mathbf{K}_{k-1|k} \mathbf{d}_k \quad (5.12)$$

$$\begin{aligned} \mathbf{P}_{k,i|k}^{aa} &= (\mathbf{I} - \mathbf{K}_{k|k} \mathbf{H}_k) \mathbf{S}_{k|k-1}^f \mathbf{S}_{k-1|k-1}^a{}^T \\ &= \mathbf{S}_{k|k-1}^f [\mathbf{I} + \mathbf{\Gamma}_k]^{-1} \mathbf{S}_{k-1|k-1}^a{}^T \end{aligned} \quad (5.13)$$

$$\begin{aligned} \mathbf{P}_{k-1|k}^a &= \mathbf{S}_{k-1|k-1}^a \mathbf{S}_{k-1|k-1}^a{}^T - \mathbf{K}_{k-1|k} \mathbf{H}_k \mathbf{S}_{k|k-1}^f \mathbf{S}_{k-1|k-1}^a{}^T \\ &= \mathbf{S}_{k-1|k-1}^a [\mathbf{I} + \mathbf{\Gamma}_k]^{-1} \mathbf{S}_{k-1|k-1}^a{}^T \end{aligned} \quad (5.14)$$

avec $\mathbf{\Gamma}_k$ défini par l'équation 5.3 du filtre SEEK. L'équation 5.14 peut être ré-écrite en racine carrée :

$$\mathbf{S}_{k-1|k}^a = \mathbf{S}_{k-1|k-1}^a [\mathbf{I} + \mathbf{\Gamma}_k]^{-1/2} \quad (5.15)$$

De même, l'équation 5.13 des covariances d'analyse croisées peut être redéfinie grâce aux équations 5.7 et 5.15 :

$$\mathbf{P}_{k,k-1|k}^{aa} = \mathbf{S}_{k|k}^a \mathbf{S}_{k-1|k}^a{}^T \quad (5.16)$$

On remarque donc que les matrices de covariances et de covariances croisées d'analyse sont entièrement définies par les matrices racine carrée $\mathbf{S}_{k|k}^a$ et $\mathbf{S}_{k-1|k}^a$.

Étape d'analyse rétrospective généralisée

L'étape d'analyse du lisseur décrite ci-dessus permet de corriger l'état à la date t_{k-1} grâce à l'observation introduite au temps t_k . Cette étape fournit à la fois les matrices de covariances d'analyse mais aussi les matrices de covariances croisées. Les analyses du lisseur peuvent alors être étendues à n'importe quel état antérieur au temps i (avec $i < k-1$). Il s'agit alors de déterminer l'état lissé $\mathbf{x}_{i|k}^a$ et les covariances d'erreur associées $\mathbf{S}_{i|k}^a$ à partir des sorties de l'analyse du filtre ($\mathbf{x}_{k-1|k-1}^a$ et $\mathbf{S}_{k-1|k-1}^a$) et du lisseur ($\mathbf{x}_{i|k-1}^a$ et $\mathbf{S}_{i|k-1}^a$). Le lisseur suit alors les mêmes étapes que celles décrites précédemment. Avec les indices i , la matrice croisée de prévision et d'analyse peut alors se ré-écrire :

$$\mathbf{P}_{k,i|k-1}^{fa} = \mathbf{M}_{k-1,k} \mathbf{S}_{k-1|k-1}^a \mathbf{S}_{i|k-1}^a{}^T = \mathbf{S}_{k|k-1}^f \mathbf{S}_{i|k-1}^a{}^T \quad (5.17)$$

De même que le gain du lisseur et la matrice racine carrée des covariances d'analyse s'écrivent :

$$\mathbf{K}_{i|k} = \mathbf{S}_{i|k-1}^a [\mathbf{I} + \mathbf{\Gamma}_k]^{-1} (\mathbf{H}_k \mathbf{S}_{k|k-1}^f)^T \mathbf{R}_k^{-1} \quad (5.18)$$

$$\mathbf{S}_{i|k}^a = \mathbf{S}_{i|k-1}^a [\mathbf{I} + \mathbf{\Gamma}_k]^{-1/2} \quad (5.19)$$

Finalement, les matrices des covariances d'analyse et des covariances croisées d'analyse s'écrivent :

$$\mathbf{P}_{i|k}^a = \mathbf{S}_{i|k}^a \mathbf{S}_{i|k}^a{}^T \quad (5.20)$$

$$\mathbf{P}_{k,i|k}^{aa} = \mathbf{S}_{k|k}^a \mathbf{S}_{i|k}^a{}^T \quad (5.21)$$

Et l'état analysé par le lisseur au temps i avec les observations disponibles jusqu'au temps k sera défini par :

$$\mathbf{x}_{i|k}^a = \mathbf{x}_{i|k-1}^a + \mathbf{K}_{i|k} \mathbf{d}_k \quad (5.22)$$

ce qui complète la formulation SEEK du lisseur.

Les équations définies ci-dessus permettent donc de mettre en place le filtre SEEK et le lisseur SEEK (de rangs réduits) dans un contexte géophysique de grande dimension. La section suivante résume le schéma d'assimilation SEEK et précise l'étape d'initialisation du schéma d'assimilation.

5.2.4 Résumé des équations du filtre et du lisseur SEEK

Le développement des équations du filtre et du lisseur SEEK exposées précédemment est résumé dans le tableau 5.1. La figure 5.1 rappelle de façon schématique le principe du lissage tel que nous l'utilisons.

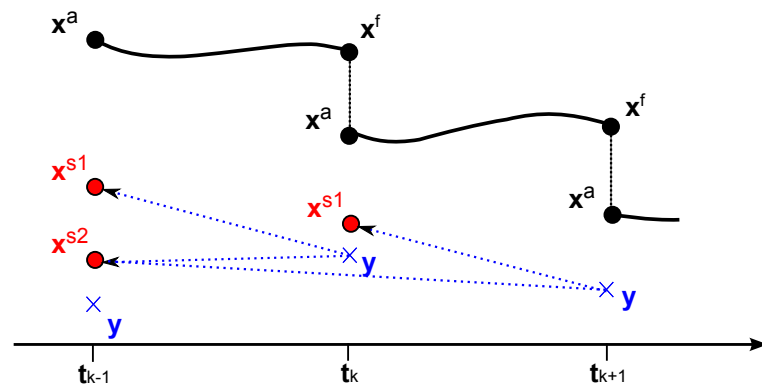


FIGURE 5.1 – Schématisation du principe de filtrage séquentiel (en noir) et du lissage (en rouge), compte-tenu de divers observations (croix bleues) disponibles aux dates t_i . Les points x^a désignent les analyses du filtre, les points x^f désignent les prévisions, les points x^{si} désignent les i^{eme} analyses rétrospectives du lisseur.

Notons que le lisseur représente uniquement 3 étapes supplémentaires en plus du filtre, son implémentation pratique est donc relativement peu coûteuse sous cette formulation. Il ne nécessite pas d'intégration modèle supplémentaire. Il présente également l'avantage d'être indépendant du filtrage, cela signifie que le lisseur peut être appliqué en même temps ou après le filtre sans besoin de nouvelle itération modèle (à condition d'avoir accès à toutes les sorties et statistiques issues du filtrage). Cet aspect est par exemple un atout majeur dans le cadre de réanalyses ayant déjà été effectuées, le lisseur pourrait permettre de réanalyser les sorties du filtrage pour éventuellement les améliorer.

L'algorithme défini ci-dessus nécessite une phase d'initialisation (conditions initiales du problème) et d'autres paramétrisations que nous allons exposer dans la section suivante.

Initialisation	
\mathbf{x}_0^a et \mathbf{S}_0^a	
Etape de prévision	
$\mathbf{x}_{k k-1}^f = \mathbf{M}_{k-1,k} \mathbf{x}_{k-1 k-1}^a$	propagation de l'état (5.23)
$\mathbf{S}_{k k-1}^f = \mathbf{M}_{k-1,k} \mathbf{S}_{k-1 k-1}^a + \mathbf{Q}$	propagation des erreurs (5.24)
Etape d'analyse du filtre	
$\mathbf{\Gamma}_k = (\mathbf{H}_k \mathbf{S}_{k k-1}^f)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{S}_{k k-1}^f)$	(5.25)
$\mathbf{K}_k = \mathbf{S}_{k k-1}^f [\mathbf{I} + \mathbf{\Gamma}_k]^{-1} (\mathbf{H}_k \mathbf{S}_{k k-1}^f)^T \mathbf{R}_k^{-1}$	gain de Kalman (5.26)
$\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_{k+1} \mathbf{x}_{k k-1}^f$	calcul de l'innovation (5.27)
$\mathbf{x}_{k k}^a = \mathbf{x}_{k k-1}^f + \mathbf{K}_{k k} \mathbf{d}_k$	analyse du filtre (5.28)
$\mathbf{S}_{k k}^a = \mathbf{S}_{k k-1}^f (\mathbf{I} + \mathbf{\Gamma}_k)^{-1/2}$	covariances d'analyse (filtre) (5.29)
Etape d'analyse du lisseur	
$\mathbf{K}_{i k} = \mathbf{S}_{i k-1}^a [\mathbf{I} + \mathbf{\Gamma}_k]^{-1} (\mathbf{H}_k \mathbf{S}_{k k-1}^f)^T \mathbf{R}_k^{-1}$	gain de Kalman (lisseur) (5.30)
$\mathbf{x}_{i k}^a = \mathbf{x}_{i k-1}^a + \mathbf{K}_{i k} \mathbf{d}_k, i \in \Sigma_k$	analyse du lisseur (5.31)
$\mathbf{S}_{i k}^a = \mathbf{S}_{i k-1}^a [\mathbf{I} + \mathbf{\Gamma}_k]^{-1/2}, i \in \Sigma_k$	covariances d'analyse (lisseur) (5.32)

TABLE 5.1 – Les différentes étapes du filtre de Kalman.

5.3 Paramétrisations du filtre et du lisseur SEEK

5.3.1 Initialisation : l'état \mathbf{x}_0^a

L'état initial du système est généralement issu d'une expérience de simulation libre passée (dans le cadre d'un problème de modélisation) ou climatique. Il est également parfois issu d'une expérience d'assimilation passée. Il peut être parfaitement en accord avec la réalité ou présenter une certaine erreur ou incertitude quant à sa véracité (ce qui est généralement le cas). Ainsi, cet état représente une *estimation* de l'océan au temps t_0 .

5.3.2 Initialisation : \mathbf{S}_0^a et méthode de réduction de la matrice des covariances d'erreur

L'initialisation de la matrice des covariances d'erreur \mathbf{P}^a sous sa forme racine carrée \mathbf{S}_0^a est une phase délicate et importante (Ballabrera-Poy et al., 2001) puisque toutes les analyses vont être liées à elle. \mathbf{S}_0^a , qui est de rang r , doit représenter les directions principales des erreurs. Pour choisir ces directions, il existe diverses méthodes dont une couramment employée : la décomposition en EOFs (Empirical Orthogonal Functions). Cette décomposition permet de trier et classer les modes de variabilité d'après une série d'états. Généralement,

cette série correspond à une succession d'états issus d'une simulation libre sur plusieurs années afin de pouvoir capter toute la variabilité du système. Chaque EOF représente une colonne de \mathbf{S}_0^a (ou une direction d'erreur) et est associée à une valeur propre permettant de les classer par ordre d'importance. La réduction d'ordre va pouvoir être effectuée en gardant uniquement les r premières EOFs représentant les modes principaux de l'erreur. En pratique, plusieurs critères peuvent permettre de définir le nombre d'EOFs à sélectionner. Dans le cas où la contrainte de coût de calcul est forte (notamment avec un filtre évolutif), r est choisi le plus petit possible. Ainsi, seules les directions d'erreur les plus globales seront conservées (et les erreurs liées à des phénomènes de plus petites amplitudes ou échelles spatiales seront filtrés).

Si la contrainte numérique est moins forte, un critère plus objectif peut être choisi en admettant par exemple qu'il faut conserver assez d'EOFs pour représenter au moins 80% de la variance du signal total. r est alors très variable et dépend de la forme du signal. Si ce dernier est dominé par quelques modes principaux, alors ils suffiront à représenter à eux-seuls suffisamment de la variance totale. En revanche si le signal est contrôlé par de nombreux phénomènes indépendants et d'échelles comparables, alors la décomposition en EOFs ne fera pas forcément ressortir de grands modes principaux. Il faudra dans ce cas en garder un grand nombre pour représenter suffisamment bien le signal de base.

Un autre critère, comparable à celui énoncé ci-dessus, consiste à sélectionner les EOFs jusqu'à ce que leur valeur propre atteigne un seuil minimum. En-dessous de ce seuil, elles sont considérées comme peu représentatives et ne sont pas sélectionnées.

L'utilisation de la décomposition en EOFs pour initialiser la base des covariances d'erreur repose sur diverses hypothèses implicites (*Brankart et al., 2003*) :

- la covariance associée à la variabilité du système est représentative de l'erreur initiale
- la variabilité du modèle est assimilable à la variabilité du système
- la série d'états échantillonnée est suffisante pour représenter correctement la variabilité du système
- l'information contenue dans les EOFs non sélectionnées est négligeable.

Bien entendu ces hypothèses ne sont parfois pas totalement respectées. Nous aurons l'occasion de mettre en place une méthode de ce type dans les expériences du chapitre 7 et d'en discuter les limites.

5.3.3 Evolutivité des covariances d'erreur : les différents types de filtres

Dans la formulation SEEK du filtrage, la matrice racine carrée des covariances d'erreur \mathbf{S} à un instant t_{k+1} est déterminée par deux termes (équation 5.24) : d'une part la propagation par le modèle des covariances d'erreur issues des analyses antérieures ($\mathbf{M}_{k-1,k} \mathbf{S}_{k-1|k-1}^a \mathbf{M}_{k-1,k}^T$), et d'autre part un terme d'erreur lié aux approximations du modèle ($\mathbf{Q}_{k-1,k}$). L'aspect propagatif des erreurs peut amener plusieurs problèmes. Selon la taille de la matrice, malgré la réduction d'ordre, cette propagation par le modèle peut se révéler très coûteuse numériquement (il faut autant d'itérations du modèle supplémentaires à chaque étape de prévision qu'il y a de membres dans la matrice \mathbf{S}). De plus, dans le cas où une analyse serait mal paramétrée et entraînerait de mauvaises corrections (mauvaises corrélations des observations suite à la troncature de \mathbf{P}), les covariances d'erreur \mathbf{S}^a is-

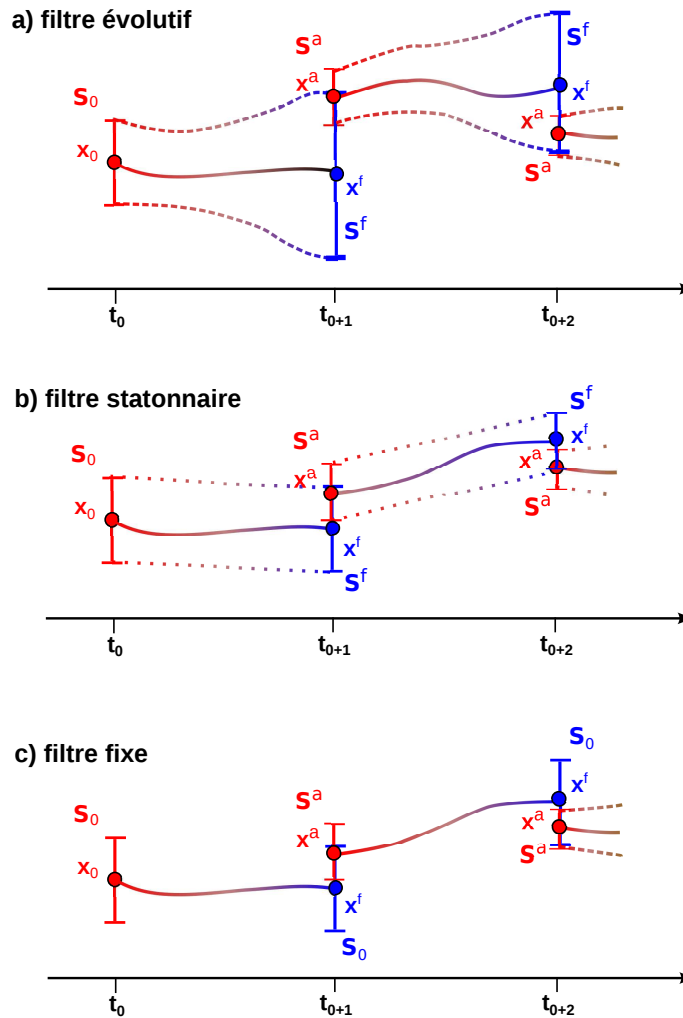


FIGURE 5.2 – Schématisation des différents types de filtrage en fonction de la propagation ou non des covariances d'erreur : a) pleinement propagée ($S^f = MS^aM^T + Q$); b) non propagée explicitement par le modèle mais l'information est tout de même prise en compte d'une étape à l'autre ($S^f = S^a + Q$); c) non propagée, les covariances d'erreur de prévision sont toujours identiques et égale à la matrice de covariances initiale ($S^f = S_0$).

sues de cette analyse en seraient tout autant détériorées. Leur prise en compte pour des analyses postérieures seraient donc dans ce cas problématique.

Ces différentes contraintes ont justifié la mise à place de diverses formulations de filtrage dont nous donnons ci-dessous les caractéristiques de trois principaux (illustrés par la figure 5.2) :

Le filtre stationnaire : il s'agit du cas le plus simple où les covariances d'erreur sont définies constantes quelle que soit l'étape d'analyse. Cela revient à décrire les covariances d'erreur (équation 4.11) par : $S_{k|k-1}^f = S_0$. Cette formulation est comparable à de l'interpolation optimale. On note tout de même que les matrices issues des analyses peuvent être utilisées à des fins diagnostiques.

Le filtre fixe : cette formulation élimine le coût de propagation des covariances. Les analyses précédentes sont prises en compte, mais pas la propagation dynamique des

covariances. Ainsi, l'équation 4.11 se reformule par : $\mathbf{S}_{k|k-1}^f = \mathbf{S}_{k-1|k-1}^a + \mathbf{Q}_{k-1,k}$.

Le filtre évolutif : cette formulation est en fait la formulation générale du filtrage. Les covariances d'erreurs issues des analyses sont prises en compte et propagées explicitement par le modèle. L'équation 4.11 reste donc inchangée dans la formulation du filtrage.

Dans le cadre de la mise en place du lisseur, l'évolution des erreurs est théoriquement indispensable (comme précisé dans le chapitre précédent). Nos travaux utiliseront donc de préférence un filtre évolutif dans un premier temps, mais nous verrons dans le dernier chapitre de ce document que l'utilisation du lissage basé sur de l'interpolation optimale, qui est très répandue dans l'océanographie opérationnelle, peut également donner se révéler avantageux.

5.3.4 La prise en compte de l'erreur modèle \mathbf{Q}

Comme nous l'avons précisé dans le chapitre précédent, l'erreur modèle \mathbf{Q} doit généralement être prise en compte pour décorréler efficacement les modes d'erreur avec le temps. Néanmoins cette erreur est souvent mal connue, il faut donc la prendre en compte différemment, sous forme d'une paramétrisation par exemple. Il existe diverses méthodes permettant de prendre en compte indirectement l'erreur modèle, comme par exemple *Verlaan et Heemink* (1997) et le filtre RRSQRT ou encore *Lermusiaux* (1999b) et le filtre ESSE. Dans le cadre du filtre SEEK, une approche simple (et souvent utilisée) est celle de l'inflation des covariances (au lieu de prendre en compte la matrice \mathbf{Q}). Dans cette approche, l'équation 5.24 du filtre devient :

$$\mathbf{S}_{k|k-1}^f = \frac{1}{\sqrt{\rho}} \mathbf{M}_{k-1,k} \mathbf{S}_{k-1|k-1}^a \quad (5.33)$$

avec ρ un facteur "d'oubli" positif et inférieur à 1 (dans le cas où $\rho = 1$, le modèle est considéré comme parfait). En revanche pour le lisseur, l'ajout d'un facteur d'oubli se répercute sur deux équations, le gain de Kalman et les covariances d'analyse :

$$\mathbf{K}_{i|k} = \sqrt{\rho} \mathbf{S}_{i|k-1}^a [\mathbf{I} + \mathbf{\Gamma}_k]^{-1} (\mathbf{H}_k \mathbf{S}_{k|k-1}^f)^T \mathbf{R}_k^{-1} \quad (5.34)$$

$$\mathbf{S}_{i|k}^a = \sqrt{\rho} \mathbf{S}_{i|k-1}^a [\mathbf{I} + \mathbf{\Gamma}_k]^{-1/2}, \quad i \in \Sigma_k \quad (5.35)$$

Le problème est qu'avec cette définition de $\mathbf{S}_{i|k}^a$, il n'est plus possible de remonter directement à la matrice de covariances d'analyse du lisseur ($\mathbf{P}_{i|k}^a = \mathbf{S}_{i|k}^a \mathbf{S}_{i|k}^{aT}$). Pour retrouver cette matrice il faut la redéfinir sous la forme $\mathbf{P}_{i|k}^a = \mathbf{S}_{i|k}^a \mathbf{T}_{i|k}^a \mathbf{T}_{i|k}^{aT} \mathbf{S}_{i|k}^{aT}$ où $\mathbf{T}_{i|k}^a$ doit être définie de manière récursive (pour plus de détails, se reporter à *Cosme et al.*, 2010). Néanmoins, la matrice $\mathbf{P}_{i|k}^a$ n'est pas nécessaire au lisseur (elle est utilisée uniquement à des fins diagnostiques), l'ajout d'un facteur d'oubli n'est donc pas un frein au lissage. Notons également que le facteur d'oubli intervient surtout sur l'efficacité du filtre SEEK, en augmentant le poids donné aux analyses. Il ne permet en aucun cas de mieux décorréler les modes (la paramétrisation de \mathbf{Q} par un facteur d'oubli entraîne une erreur modèle

parfaitement corrélée dans le temps).

Dans notre étude, nous avons choisi deux approches différentes :

- Dans un premier temps, nous allons considérer l'erreur modèle grâce à la création d'un ensemble basé sur la perturbation de paramètres intrinsèques au modèle (chapitre 6).
- Par la suite, pour simplifier le problème et parce que nous utiliserons une configuration en expériences jumelles (où l'erreur simulée ne sera pas liée au modèle), le modèle sera considéré comme parfait et nous garderons simplement le terme propagatif des covariances d'erreur.

5.3.5 Localisation de l'analyse

Les méthodes de réduction d'ordre (par décomposition en EOFs, ou plus généralement par des méthodes d'ensemble) sont efficaces pour diminuer la taille du système mais elles imposent en contrepartie de nouvelles contraintes. En effet, la troncature des directions d'erreur est une source de difficulté pour représenter efficacement les covariances entre des variables éloignées. En théorie, on pourrait s'attendre à trouver des covariances quasi-nulles lorsque les distances entre variables sont importantes. Mais la réduction d'ordre introduit parfois des corrélations non justifiées entre ces variables. *Houtekamer et Mitchell* (1998) ont par exemple montré que l'utilisation d'un filtre d'ensemble avec une analyse globale nécessitait un très grand nombre de modes (ou de degrés de liberté) pour contraindre efficacement la mésoéchelle (c'est-à-dire pour représenter efficacement les corrélations longue distance).

Pour s'affranchir de ce problème, une méthode de localisation doit être appliquée afin d'éliminer les mauvaises corrélations spatiales. Deux techniques sont principalement utilisées : la *localisation des covariances* (*Houtekamer et Mitchell*, 2001 ; *Hamill et al.*, 2001) et la *localisation de l'analyse* (*Evensen*, 2003 ; *Hunt et al.*, 2007). La première méthode joue directement sur la matrice des covariances \mathbf{P}^f alors que la seconde s'applique plutôt sur la matrice d'erreur d'observation \mathbf{R} . Ces approches sont justifiées par le fait que la dynamique locale est représentable dans un sous-espace de bien plus petite taille que l'espace du modèle complet (*Patil et al.*, 2001).

La localisation des covariances est généralement effectuée par une multiplication de Schur (qui coupe les corrélations longues-distances) de la matrice des covariances de prévision \mathbf{P}^f pour le filtre avec une matrice de corrélation à support local. Dans le cadre du lissage de type conjoint (que nous utilisons), la localisation est effectuée de la même façon sur les matrices de covariances croisées $\mathbf{P}_{k,i|1:k-1}^{fa}$ et $\mathbf{P}_{k,i|1:k-1}^a$ (voir par exemple *Khare et al.*, 2008).

La localisation de l'analyse est quant à elle effectuée très simplement en réduisant la correction liée à une observation en fonction de la distance à cette observation. En pratique, on fait tendre la matrice \mathbf{R} vers l'infini avec la distance. Les corrélations de plus grandes distances seront donc automatiquement annulées (voir la figure 5.3 pour une illustration de ce principe). Cette méthode est valable de la même façon pour le filtre et les lisseurs conjoints.

Notons également que dans le cas d'un lisseur marginal de type forward-backward, la méthode de localisation des analyses n'est pas valable pour l'étape rétrograde (qui n'utilise pas directement les observations). Dans le cas de ce lisseur une méthode de localisation

des covariances est donc préférable.

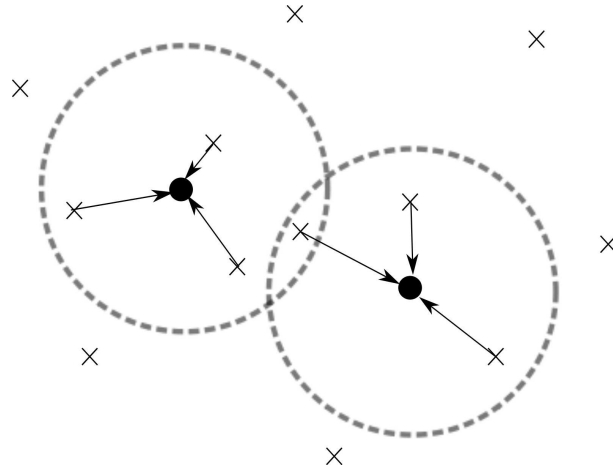


FIGURE 5.3 – Schématisation du principe de localisation de l'analyse. Les points noirs indiquent les lieux des analyses du filtre, les croix indiquent la position des différentes observations disponibles et les cercles pointillés gris indiquent la limite des zones dans lesquelles les observations sont prises en compte pour les analyses.

Avec le lisseur fixed-lag que nous utilisons, nous avons choisi une méthode de localisation des analyses. Nous aborderons plus en détail dans les expériences du chapitre 7 le choix de la zone d'influence des observations et les conséquences sur la qualité des analyses. On peut tout de même noter que le fait de réduire la longueur d'influence d'une observation limite en conséquent le nombre d'observations disponibles pour une analyse en un point (puisque l'analyse tiendra compte uniquement des observations proches et non plus des corrélations longues distances). Dans le cas d'un réseau éparse, la localisation pourrait entraîner des zones coupées de toute observation. L'inefficacité de réseaux d'observations peu denses pour corriger l'état 3D de l'océan a d'ailleurs été pointée par *Ubelmann (2009)*. Le lisseur peut présenter l'avantage de densifier le nombre d'observations en complétant le réseau disponible au moment de l'analyse par des observations issues de dates ultérieures.

Enfin, il faut noter que dans le cas d'un lisseur de rang réduit, la question de la localisation temporelle peut également être abordée (au même titre que la localisation spatiale). Cette localisation est normalement effectuée naturellement par la prise en compte de l'erreur modèle (qui limite l'influence d'une observation dans le temps). Mais dans le cas où cette erreur est simplifiée (paramétrisation) ou même ignorée, la question demeure. Avec le lisseur fixed-lag, l'influence des observations est artificiellement limitée par la fenêtre temporelle choisie pour les analyses rétrospective. On peut donc considérer que ce lisseur introduit automatiquement une localisation temporelle des observations.

5.4 Implémentation du SEEK dans NEMO : le logiciel SESAM

La mise en place de l'assimilation de données dans le modèle OPA a été effectuée grâce au logiciel SESAM¹ (*Brankart et al.*, 2002). Ce logiciel présente l'avantage de n'être pas intrusif : aucune modification du code du modèle n'est donc nécessaire. Le fait d'utiliser un module d'assimilation externe au modèle d'océan est avantageux sur plusieurs points. D'une part cela permet de mettre en oeuvre l'assimilation facilement sur n'importe quelle version du code (et plus généralement sur n'importe quelle configuration ou modèle utilisé). D'autre part, les modules d'assimilation peuvent facilement être modifiés sans avoir à toucher au code du modèle.

SESAM permet de mettre en place le filtre SEEK sous ses différentes formes (évolutive ou non), mais nous avons eu à rajouter les quelques étapes spécifiques au lisseur. Ci-dessous est résumé l'algorithme d'assimilation tel qu'il est codé dans SESAM (les indices ne sont pas précisés pour plus de clareté) :

I. Etapes de filtrage :

1. Calcul de Γ :

$$\Gamma = (\mathbf{HS}^f)^T \mathbf{R}^{-1} (\mathbf{HS}^f) \quad (5.36)$$

2. Décomposition U-D (choleski) de Γ :

$$\Gamma = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (5.37)$$

où $\mathbf{\Lambda}$ est une matrice diagonale contenant les valeurs propres.

3. Calcul du vecteur d'innovation dans l'espace réduit :

$$\delta = (\mathbf{HS})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{Hx}^f) \quad (5.38)$$

4. Calcul de la correction dans l'espace réduit :

$$\gamma = \mathbf{U} [\mathbf{I} + \mathbf{\Lambda}]^{-1} \mathbf{U}^T \delta \quad (5.39)$$

5. Calcul de la matrice de transformation de \mathbf{S}^f en \mathbf{S}^a :

$$\mathbf{L} = \mathbf{U} [\mathbf{I} + \mathbf{\Lambda}]^{-1/2} \mathbf{U}^T \quad (5.40)$$

6. Calcul de l'état analysé :

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{S}^f \gamma \quad (5.41)$$

7. Calcul des covariances d'erreur d'analyse :

$$\mathbf{S}^a = \mathbf{S}^f \mathbf{L} \quad (5.42)$$

II. Etapes de lissage :

1. Calcul des analyses rétrospectives :

$$\mathbf{x}_{i|k}^a = \mathbf{x}_{i|k-1}^a + \mathbf{S}_{i|k-1}^a \gamma \quad (5.43)$$

1. <http://www-meom.hmg.inpg.fr/Web/Outils/SESAM/>

2. Calcul des covariances d'analyse rétrospectives :

$$\mathbf{S}_{i|k}^a = \mathbf{S}_{i|k-1}^a \mathbf{L} \quad (5.44)$$

III. Etapes de prévision :

1. Propagation de l'état analysé :

$$\mathbf{x}^f = \mathbf{M}\mathbf{x}^a \quad (5.45)$$

2. Propagation des covariances d'erreur d'analyse du filtre :

$$\mathbf{S}^f = \mathbf{M}\mathbf{S}^a \quad (5.46)$$

Le lissage est donc calculé par seulement deux étapes supplémentaires au filtrage, étapes pouvant être effectuées en même temps qu'une expérience de filtrage ou après (en offline).

Enfin, précisons que le logiciel SESAM n'est pas uniquement un logiciel d'assimilation. Il permet également d'effectuer des opérations mathématiques, statistiques ou diagnostiques sur des fichiers de sortie modèle (comme par exemple la décomposition en EOFs, le calcul d'écart RMS...). Ces outils vont bien sûr de paire avec l'assimilation de données. SESAM permet donc à lui-seul d'avoir accès à une panoplie complète d'outils pour mettre pleinement en oeuvre l'assimilation de données avec n'importe quel modèle numérique.

5.5 Quelques précisions supplémentaires

Pour terminer la partie sur l'assimilation de données et avant de décrire plus en détail les expériences menées dans cette thèse, nous donnons quelques précisions générales sur les caractéristiques du système d'assimilation tel qu'il sera utilisé.

5.5.1 Vecteur d'état et vecteur d'observation

Le vecteur d'état doit contenir suffisamment de variables indépendantes pour pouvoir décrire toute la dynamique de l'océan. Les variables retenues dans notre cas sont la température, la salinité et les vitesses zonales et méridiennes comme variables 3D, ainsi que la SSH comme variable 2D. Ces 5 variables constitueront donc notre vecteur d'état.

Le vecteur d'état n'est pas considéré comme totalement connu, c'est-à-dire qu'il est sous-échantillonné dans les observations. Le réseau d'observation est défini dans la section 7.3. Il contient uniquement trois variables : température, salinité et SSH, elle-même sous-échantillonnées spatialement et temporellement. Les corrections sur les variables non-observées se feront par l'intermédiaire de corrélations multivariées.

5.5.2 Un outil de diagnostic : le calcul d'erreur RMS

Nous aurons l'occasion de développer plus en détail les différents diagnostics utilisés pour quantifier l'efficacité de l'assimilation de données. Nous précisons ici une méthode couramment utilisée pour définir rapidement l'écart entre deux états. Il s'agit de l'erreur

RMS (Root Mean Square), définie par (dans le cadre de nos expériences) :

$$\mathbf{RMS} = \sqrt{\frac{1}{N \times M \times P} \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^P (y_{i,j,k}^0 - Hx_{i,j,k}^a)^2} \quad (5.47)$$

à un instant t donné, où $i = 1 : N$ représente l'ensemble des longitudes, $j = 1 : N$ représente l'ensemble des latitudes et $k = 1 : P$ représente l'ensemble des niveaux verticaux. Cet écart (ou erreur) quadratique moyen peut être calculé sur l'ensemble du domaine (auquel cas on parlera d'erreur RMS globale) ou sur des sous-domaines choisis.

L'erreur RMS permet d'avoir une première mesure objective des différences existant entre deux états. Bien que très utile, l'erreur RMS possède certaines limites. Elle est en effet très sensible aux grands écarts (du fait de l'élévation au carré des écarts) et a donc tendance à moins tenir compte des petites différences. De plus, ce calcul automatique tient compte uniquement des différences point par point, ce qui n'est pas forcément le plus approprié dans un système dynamique tel que l'océan. Prenons par exemple le cas d'une température élevée sur une seule maille. Dans un cas (1) où cette température serait décalée d'une maille entre deux runs, l'erreur RMS serait considérée comme très importante. Dans un cas (2) où la température serait étalée sur plusieurs mailles dans l'un des runs tout en étant un peu plus faible, l'erreur RMS pourrait être identique à celle du cas (1). A niveau d'erreur RMS identique, on peut se demander lequel des cas (1) ou (2) est le plus réaliste. Mais cette différence ne sera pas visible sur l'erreur RMS. Il est donc indispensable de compléter ce diagnostic de premier ordre par d'autres plus précis.

5.6 Conclusion

Comme nous l'avons énoncé dans le chapitre précédent, la mise en place de l'assimilation de données dans un contexte réaliste est fortement contrainte par la taille des systèmes. L'algorithme SEEK du filtre et son extension au lisseur permet de répondre à ces contraintes. Nous avons pu aborder dans ce chapitre ses hypothèses et leur mise en place pratique, notamment la réduction d'ordre des matrices de covariances. Cette réduction permet d'appliquer la théorie du lissage (et du filtrage) à moindre coût sur des systèmes de grande dimension. En contrepartie, elle implique une phase délicate de paramétrage des matrices d'erreur, notamment l'erreur modèle \mathbf{Q} et les covariances d'erreur initiales \mathbf{S}_0^a . Ces aspects sont d'autant plus sensibles dans le cadre du lissage puisque l'efficacité d'un lisseur est fortement contrôlée par la représentation 4D des erreurs.

Les aspects pratiques et théoriques nécessaires à nos expériences ont maintenant tous été définis. Nous allons donc pouvoir aborder plus précisément les travaux d'assimilation ayant été effectués. Dans un premier temps, nous nous intéresserons à la mise en place d'une expérience d'ensemble pour représenter une erreur modèle (chapitre 6). Ces expériences, peu concluantes, nous amèneront ensuite à considérer une paramétrisation de l'erreur sur la condition initiale. Nous traiterons alors dans le chapitre 7 des aspects de mise en place d'un filtre évolutif (théoriquement indispensable au lissage), puis de l'efficacité des corrections 4D apportées par le lisseur dans un tel cas (chapitre 8). Afin d'évaluer plus précisément l'effet de l'évolution des covariances d'erreur sur le lissage, nous étudierons

également le comportement du lisseur dans le cas de covariances d'erreur fixes (qui est un cas très courant dans les systèmes opérationnels) dans le chapitre 9.

Troisième partie

Préliminaires au lissage : mise en place du filtre évolutif

Chapitre 6

Simulation d'erreur modèle par une méthode d'ensemble et perturbation de paramètres

Sommaire

6.1	Introduction	88
6.2	Outil d'analyse de la vitesse des ondes : la transformée de Radon	88
6.3	Effet d'une perturbation sur l'état de l'océan	89
6.3.1	Effets sur la vitesse des TIWs	90
6.3.2	Effets sur la température de surface	92
6.3.3	Effets sur les vitesses	94
6.3.4	Conclusion sur les expériences de sensibilité	95
6.4	Création de l'ensemble	95
6.4.1	Méthode de tirages aléatoires	96
6.4.2	Création de l'ensemble	97
6.4.3	Variances et corrélations	97
6.5	Conclusion	99

6.1 Introduction

Les expériences d'assimilation que nous avons mises en place dans cette thèse n'utilisent pas d'observations réelles. Il s'agit d'expériences jumelles, où un run modèle est défini comme la référence (ou l'état vrai de l'océan) de laquelle sont extraits des jeux d'observation, et un autre run modèle légèrement perturbé est défini comme un run faux à corriger par l'assimilation de données (le but étant de rapprocher au mieux le run faux du run de référence). Un aspect important de ce type d'expérience est donc de déterminer quelles perturbations vont permettre de différencier un run de référence d'un run faux. Dans le cadre d'expériences jumelles, le modèle utilisé doit rester globalement identique entre les deux runs (dans sa résolution, ses équations...). Utiliser deux modèles différents conduirait à des expériences cousines, qui peuvent présenter certaines complications pour l'assimilation (grilles de modèles différentes...).

Pour créer une expérience jumelle, la première piste que nous avons explorée et que nous allons détailler dans ce chapitre consiste à perturber certains paramètres intrinsèques au modèle (il s'agit donc de créer une erreur modèle). Trois paramètres ont été choisis en lien avec la modélisation de l'énergie turbulente dans le modèle. Il s'agit des coefficients de diffusion liés à la paramétrisation sous-maille, horizontale ou verticale (voir sections 3.4.2 et 3.4.2). Nous avons choisi ces paramètres pour leur impact *a priori* marqué sur la propagation des ondes d'instabilité tropicales (*Caltabiano et al.*, 2005) et parce que leur définition demeure encore incertaine. Ainsi, une perturbation entre deux runs (ayant des paramètres différents) pourrait se retranscrire clairement sur les TIWs et donc représenter un objet d'étude intéressant pour l'assimilation.

Nous allons étudier l'impact effectif sur les TIWs d'une perturbation sur chacun des paramètres choisis, puis nous étendrons ces résultats à une expérience d'assimilation par une méthode d'ensemble. Mais avant de présenter ces résultats, nous détaillons dans une première partie la méthode utilisée pour caractériser de façon automatique la vitesse de propagation des ondes à partir de mesure de la SSH : la transformée de Radon.

6.2 Outil d'analyse de la vitesse des ondes : la transformée de Radon

La perturbation des coefficients de diffusion peut avoir divers effets sur les TIWs, et notamment perturber leur vitesse de propagation. Ce point nous intéresse tout particulièrement dans le cadre d'une assimilation 4D où la composante temporelle joue un rôle important. Nous allons donc chercher à caractériser la vitesse des ondes dans le cas du run de référence et des runs perturbés.

Comme nous l'avons vu dans le chapitre 1, les TIWs ont une empreinte forte sur la température de surface et la SSH. Il est donc facile de voir leur propagation à travers le bassin en traçant un diagramme de Hovmöller sur l'une de ses variables autour de la latitude où leur signal est le plus fort (vers 3°N). Néanmoins, cette latitude n'étant pas forcément la meilleure, il faudrait tracer un diagramme de Hovmöller à chaque latitude voisine pour être sûr de capter l'empreinte la plus représentative. Une fois ce diagramme tracé (qui représente un temps en fonction d'une distance), la vitesse des ondes peut se déterminer facilement en cherchant la pente la plus proche de la propagation du signal. Mais cette recherche de pente peut s'avérer délicate si elle est faite à la main et si le

signal est imprécis. Une imprécision de quelques dixièmes de degrés sur un angle peut se traduire par quelques cm.s^{-1} sur les vitesses. Or, comme nous cherchons à comparer des signaux qui auront le même ordre de grandeur mais qui se caractériseront par de petites différences, il nous faut adopter une méthode plus systématique (et plus précise) pour des résultats plus cohérents. Nous avons donc opté pour la transformée de Radon (*Radon*, 1917; *Lecointre*, 2009).

Cette méthode consiste à effectuer les étapes décrites ci-dessus de façon automatique grâce à l'étude de la variance du signal projeté selon différents angles. Nous allons détailler le principe de façon schématique ci-dessous (et sur la figure 6.1) :

- a) Un diagramme de Hovmöller est tracé à une latitude donnée (sur une variable et une période choisies).
- b) Le diagramme est projeté perpendiculairement à une droite horizontale, c'est-à-dire dans la direction Est-Ouest (cela revient à faire la somme de toutes les valeurs selon une ligne perpendiculaire à la droite de projection).
- c) L'opération (b) est répétée pour des angles de projection différents (de 0° à 90°).
- d) Le signal projeté selon chaque angle X possède une variance (ou amplitude) différente. Cette variance dépend de la façon dont les structures de même nature se sont sommées ou non. L'angle correspondant à une projection de variance maximale correspondra à l'angle Xm de propagation du signal dominant (car toutes les structures de même nature s'additionneront parfaitement). L'angle Xm est ensuite converti en vitesse (grâce au rapport distance-temps du diagramme de Hovmöller). La vitesse du signal dominant pour une latitude donnée est ainsi obtenue.
- e) Les étapes (a) à (d) sont répétées pour différentes latitudes (dans notre cas, de l'équateur jusqu'à 5°N).
- f) Les vitesses du signal dominant de chaque latitude sont finalement combinées sous un seul diagramme "latitude-vitesse".

La transformée de Radon fournit donc un profil de vitesses en fonction des latitudes. La traces des TIWs, en admettant qu'elles représentent le signal dominant à leurs latitudes et sur les variables considérées (SST ou SSH), devrait apparaître distinctement autour de 3°N . Nous aurons donc une mesure systématique de leur vitesse pour chaque run.

6.3 Effet d'une perturbation sur l'état de l'océan

Nous allons étudier dans cette partie l'impact d'une perturbation du modèle sur le système océanique. Les paramètres que nous allons perturber à l'intérieur du modèle sont liés aux effets de diffusion dynamique sous-maille. Plus précisément, il s'agit du coefficient de diffusion horizontal sur la vitesse A^{hm} (section 3.4.2) et des coefficients de diffusion verticaux sur la vitesse en surface A_0^{vm} et en profondeur A^{vm} (voir section 3.4.2). Le coefficient de diffusion horizontal est imposé constant dans tout le domaine. En revanche les deux autres coefficients correspondent uniquement à des valeurs minimales limites pour le schéma TKE. Ils peuvent donc varier en fonction des régions ou des niveaux verticaux. Nous choisissons de perturber chacun de ces paramètres de façon autonome, en modifiant leur valeur d'un ordre de grandeur en-dessous et au-dessus de leur valeur de référence. Cela donne donc 6 runs perturbés différents, que nous noterons ES1 à ES6 (pour Expérience de

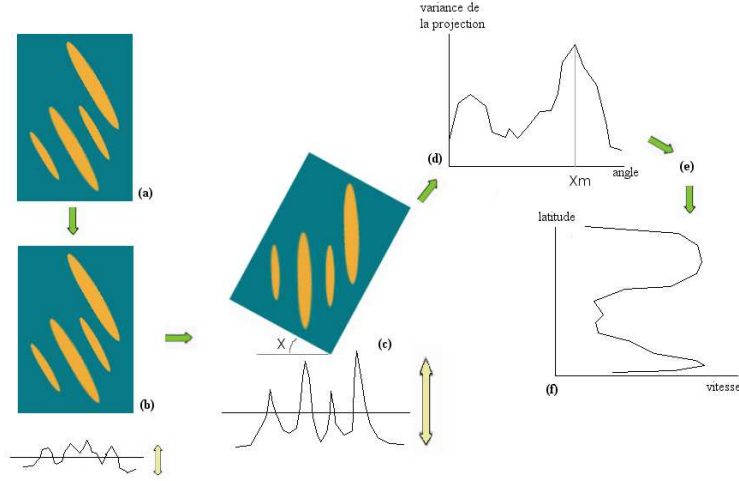


FIGURE 6.1 – Représentation schématique du principe de la transformée de Radon.

Sensibilité). Les valeurs des paramètres de référence et perturbés sont récapitulées dans le tableau 6.1.

Chacun des runs (perturbé ou référence) possède les mêmes forçages et la même condition initiale au Jour Julien 20170 (soit le 5 avril 2005). A partir de cette même condition initiale, chaque run est propagé librement avec les valeurs des paramètres qui lui sont propres. Les premières TIWs naissent plusieurs semaines après le début du run (mois de juin), mais l'énergie nécessaire à leur formation est introduite bien avant. Il apparaît donc nécessaire de perturber les runs à partir du mois d'avril afin que ces perturbations se répercutent à la fois sur la formation des ondes, puis sur leur propagation. La durée totale de chaque run est de 265 jours, soit jusqu'au jour 20435 (27 décembre 2005).

nom du run	A^{vm}	A_0^{vm}	A^{hm}
REF-ES	1×10^{-6}	1×10^{-4}	$-1,5 \times 10^{11}$
ES1	1×10^{-5}	1×10^{-4}	$-1,5 \times 10^{11}$
ES2	1×10^{-7}	1×10^{-4}	$-1,5 \times 10^{11}$
ES3	1×10^{-6}	1×10^{-5}	$-1,5 \times 10^{11}$
ES4	1×10^{-6}	1×10^{-3}	$-1,5 \times 10^{11}$
ES5	1×10^{-6}	1×10^{-4}	$-1,5 \times 10^{10}$
ES6	1×10^{-6}	1×10^{-4}	$-1,5 \times 10^{12}$

TABLE 6.1 – Tableau récapitulant les valeurs des différents paramètres de diffusion dynamique (diffusion horizontale A^{hm} , diffusion verticale en surface 1×10^{-4} et diffusion verticale en profondeur A^{vm}) pour le run de référence REF-ES et les run perturbés ES1 à ES6. Les valeurs perturbées par rapport aux valeurs de référence sont indiquées en gras.

6.3.1 Effets sur la vitesse des TIWs

La figure 6.2 récapitule l'ensemble des profils de vitesse calculés par transformée de Radon sur chacun des runs perturbés (en bleu) et le run de référence (en pointillés rouges). Entre 3°N et 4°N une vitesse marquée de l'ordre de 40 à $50 \text{ cm}\cdot\text{s}^{-1}$ est visible sur presque

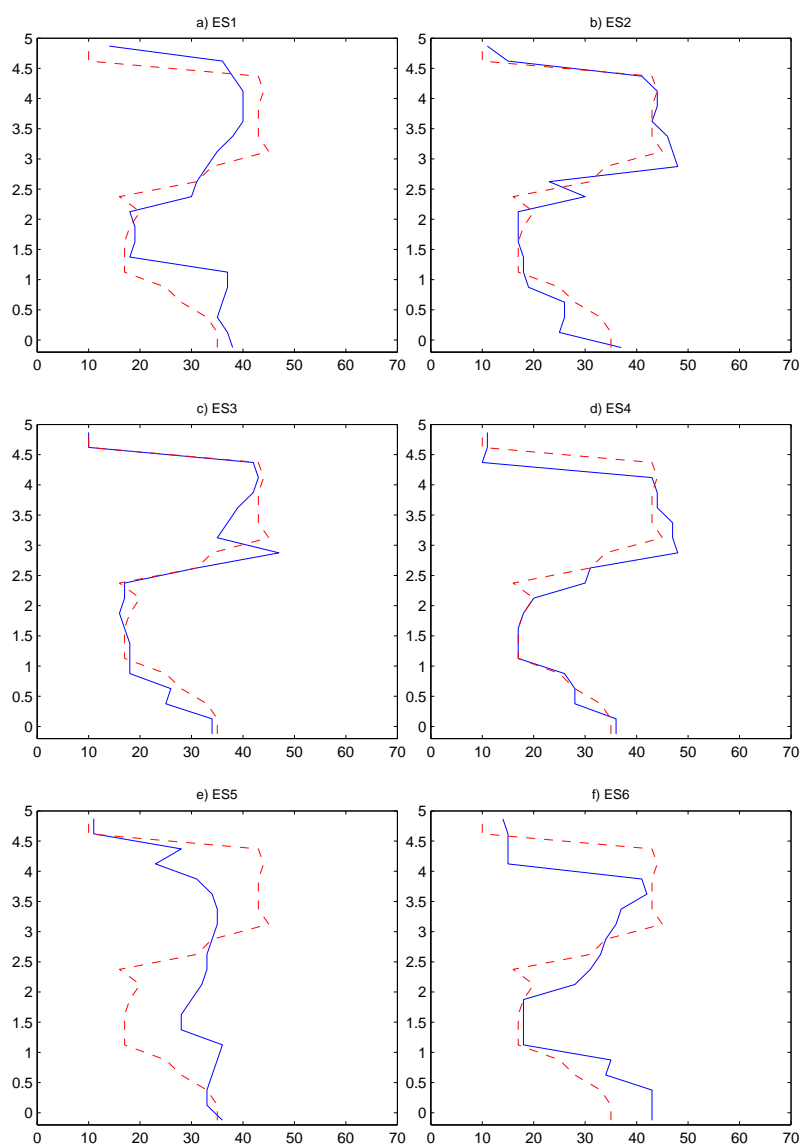


FIGURE 6.2 – Profils de vitesse (en abscisse, en cm.s^{-1}) calculés par transformée de Radon sur la SSH entre les jours 20170 et 20435 en fonction de la latitude (en ordonnée, en $^{\circ}\text{N}$). Les dénominations ES1 à ES6 font référence aux runs perturbés (voir tableau 6.1). Les courbes bleues indiquent les vitesses calculées pour chaque run perturbé, et les courbes pointillées rouges indiquent le profil des vitesses dans le run de référence.

chaque profil. L'ordre de grandeur de la vitesse et sa localisation correspondent bien aux caractéristiques associées aux TIWs.

En comparant les différents profils, on constate que certains sont très proches de la référence (ES2 et ES3). Il présentent uniquement de petites différences qui peuvent facilement être liées à la qualité du signal. Au contraire, ES5 présente un profil très perturbé et des différences de presque 10 cm.s^{-1} par rapport à la référence. Les autres runs ont également des différences, mais plutôt dans la forme de leur profil que dans la vitesse des ondes (ES6 par exemple). Ces résultats suggèrent que la diffusion horizontale (ES5 et ES6) est la plus à même d'influencer la vitesse des TIWs.

On notera tout de même que des profils de vitesse calculés sur des périodes légèrement différentes peuvent afficher des structures très diverses (non montré). De plus, en regardant les diagrammes de Hovmöller (figure 6.3) on se rend compte qu'il est difficile d'y voir de nettes différences entre les runs. Les résultats de la transformée de Radon, bien que systématiques, doivent donc être pris avec un certain recul. Par conséquent, nous allons regarder dans la section qui suit s'il est possible d'identifier des différences plus claires sur la structure même des TIWs (sur la SST).

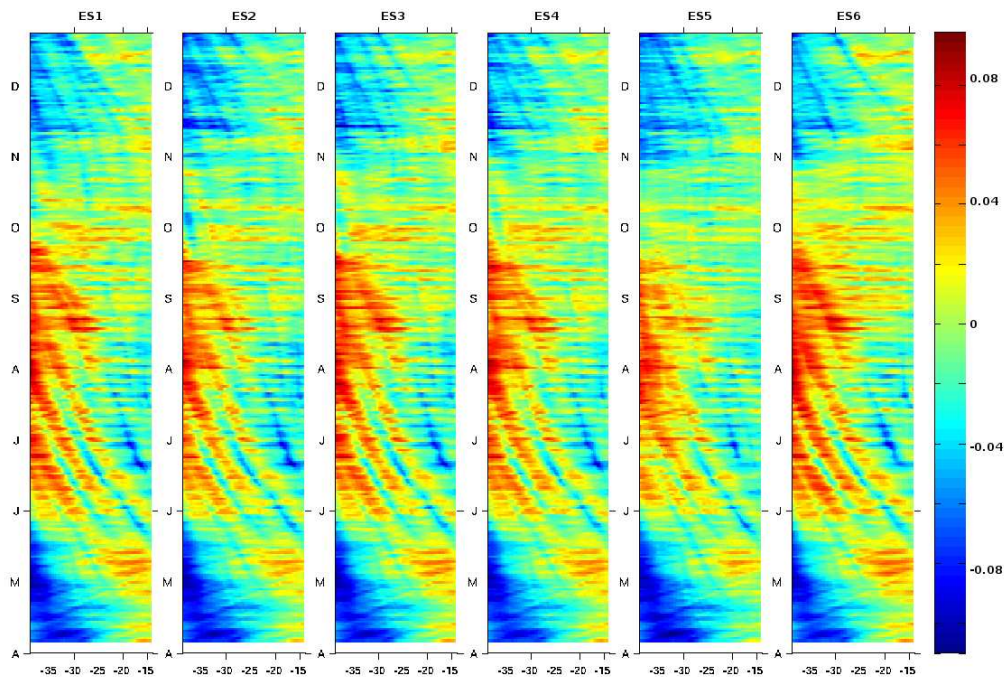


FIGURE 6.3 – Diagrammes de Hovmöller sur la SLA (en m) pour les différents runs perturbés (voir tableau 6.1).

6.3.2 Effets sur la température de surface

Les figures 6.4 et 6.5 montrent la température de surface instantanée pour les différents runs au milieu de la saison des TIWs (jour 20351, soit le 4 octobre 2005) et à la fin de la saison (jour 20435, soit le 27 décembre 2005). L'isotherme 28°C permet de bien visualiser la forme des ondes d'instabilités. Au début de l'automne (figure 6.4), la structure des TIWs commence déjà à être perturbée dans la plupart des runs. Ces perturbations se traduisent à la fois en terme de déphasages locaux (notamment sur ES5 et ES6) mais aussi en terme d'amplitude. Certaines ondes sont phasées d'un côté du bassin et déphasées de l'autre (par exemple ES1 ou ES2), cela suggère donc que leur propagation ne se fait pas de la même façon. Seuls ES3 et ES4 ne semblent pas vraiment perturbés (excepté à l'ouest du bassin, mais le décalage reste minime). Le fait de modifier la diffusion verticale en surface n'a *a priori* pas beaucoup d'effet. A la fin de l'automne (figure 6.5), les anomalies sur les TIWs sont plus marquées. Sur ES1 et ES6, le déphasage est clairement visible. Même ES3, qui n'était précédemment pas trop perturbé, présente maintenant de sérieuses anomalies. Sur ES5, on note de façon très claire une onde en phase total à l'est et des ondes déphasées à

l'ouest, ce qui souligne encore une fois les différences de propagation. Sur ES2, le déphasage n'est pas visible, mais les ondes n'ont pas la même amplitude méridionale. Enfin, ES4 reste assez peu perturbé comme précédemment, même si on note une des formes d'onde légèrement différentes à l'ouest.

Ces résultats suggèrent que, en dehors de ES4, toutes les perturbations introduites semblent avoir un effet visible sur les TIWs. Cet effet est plus marqué vers la fin de la période, il faut donc *a priori* quelques mois pour que la dynamique soit affectée par les perturbations. Les résultats indiquent également que même sur des runs où la transformée de Radon indiquait peu de dispersion, on observe quand même des différences notables (ES2 par exemple). La perturbation des paramètres de diffusion semble donc justifiée pour introduire une erreur modèle dans le cadre d'expériences jumelles. Avant d'étendre ces résultats à une méthode d'ensemble, nous allons vérifier rapidement l'influence des perturbations sur les vitesses du modèle. Cela permettra de nous assurer que c'est bien la vitesse des ondes qui est directement affectées et non la vitesse zonale des courants.

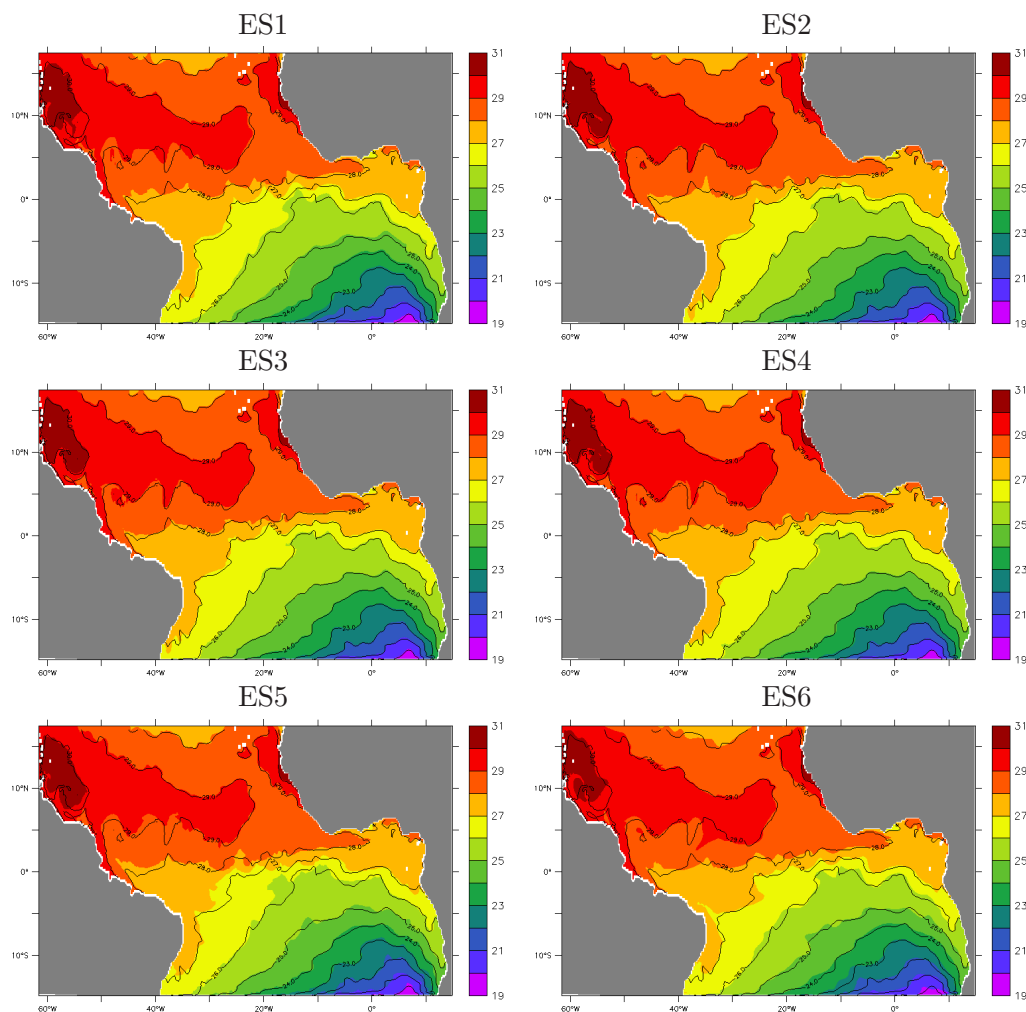


FIGURE 6.4 – Cartes de température de surface (en °C) instantannée au jour 20351 (4 octobre 2005) pour les différents runs perturbés (en couleur), et pour le run de référence (en contours).

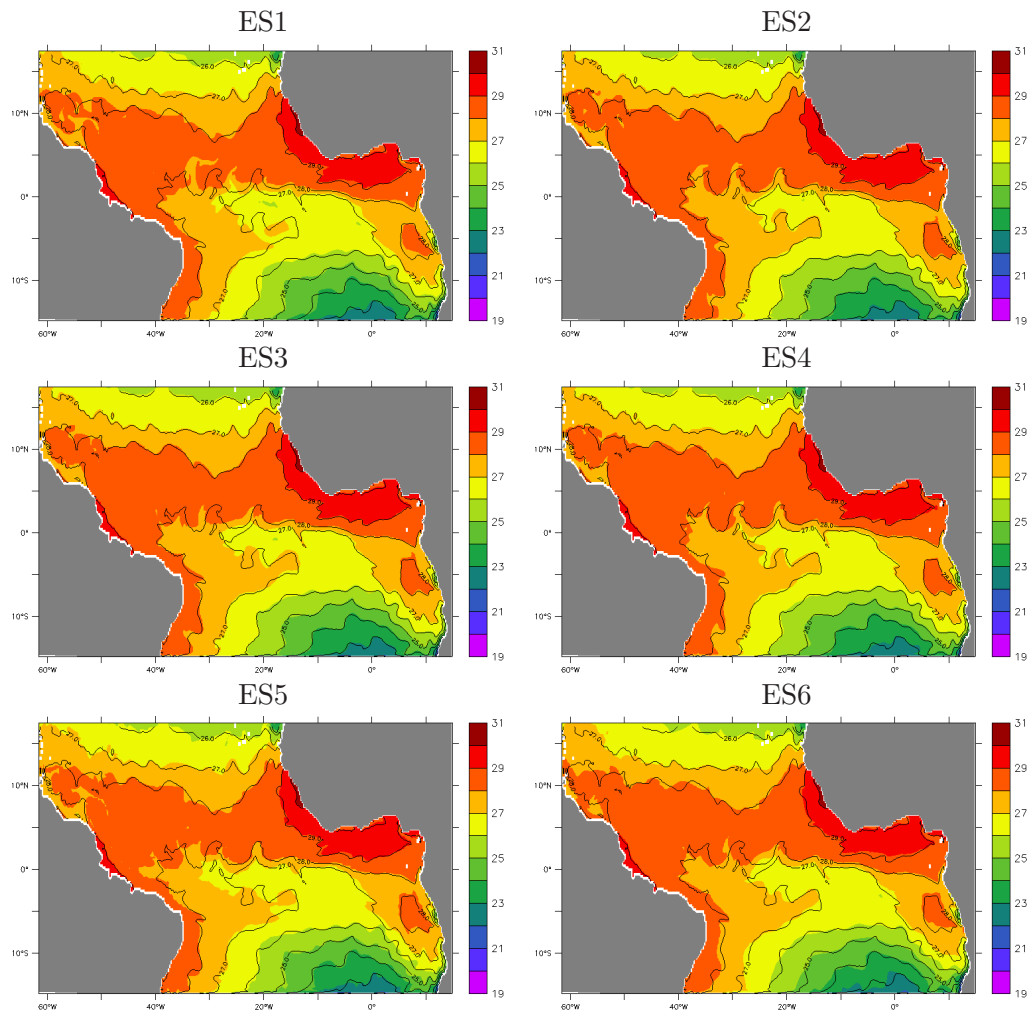


FIGURE 6.5 – Cartes de température de surface (en °C) instantannée au jour 20435 (27 décembre 2005) pour les différents runs perturbés (en couleur), et pour le run de référence (en contours).

6.3.3 Effets sur les vitesses

La modification des coefficients de diffusion dynamique pourrait avoir un impact sur la vitesse des courants. Ainsi, les décalages observés précédemment sur les TIWs seraient alors plutôt dus à l'advection des ondes par les courants zonaux plutôt qu'à des différences sur la vitesse même des ondes. La figure 6.6 montre des coupes au niveau de l'équateur de la différence entre la vitesse zonale des runs perturbés et celle du run de référence. Il apparaît que la vitesse est clairement modifiée par la perturbation des paramètres, mais pas de façon uniforme et uniquement sur les 200 premiers mètres. Dans la plupart des runs, on observe une alternance d'anomalies positives et négatives, notamment dans les couches superficielles. Seul ES5 présente une anomalie négative plus générale en profondeur au niveau du courant sous-équatorial. On peut donc estimer que, de façon générale, en surface, la modification de la vitesse n'est pas suffisamment uniforme pour advecter les ondes (il n'y a pas de ralentissement ou d'accélération globale des courants). Elle peut néanmoins influencer localement les masses d'eau et la trace apparente des ondes sur la

SST. De plus, si les vitesses présentent des anomalies conséquentes au jour 20435, elles sont beaucoup moins affectées lors des premiers mois. Il faut attendre plusieurs semaines après le début des runs pour voir de sérieuses anomalies.

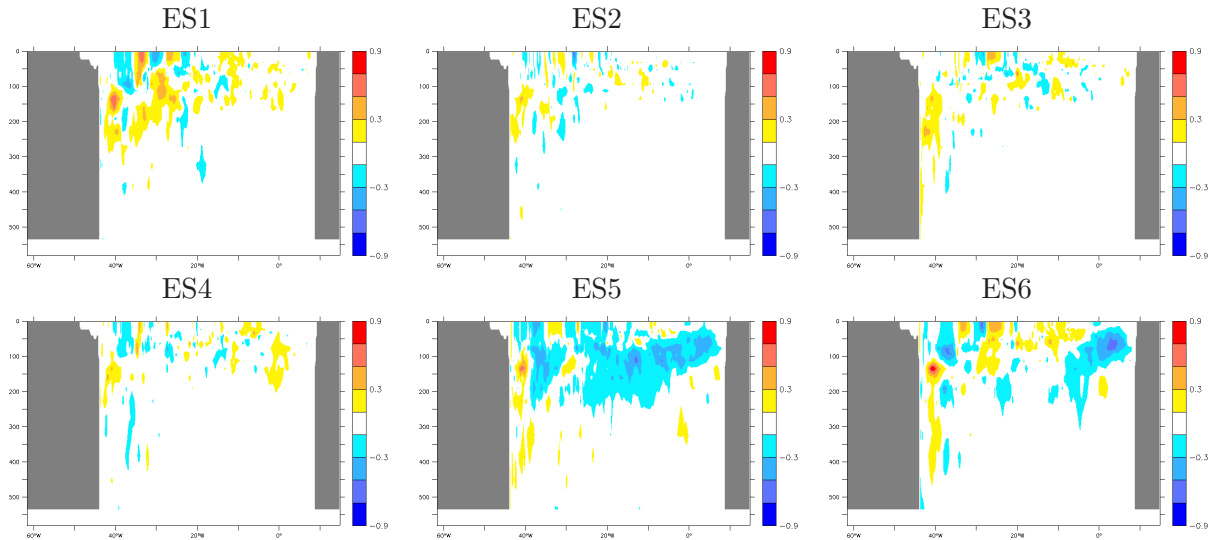


FIGURE 6.6 – Coupes à l'équateur des différences de vitesses zonales (en m.s^{-1}) au jour 20435 (27 décembre 2005) entre le run de référence et les différents runs perturbés.

6.3.4 Conclusion sur les expériences de sensibilité

A travers les résultats précédents nous avons pu montrer que la modification des paramètres de diffusion dynamique (horizontale ou verticale) impacte de façon significative la dynamique des ondes (et plus généralement la dynamique du bassin). La transformée de Radon ne suffit pas à elle-seule pour caractériser efficacement les modifications des ondes et doit être complétée par une analyse des champs de température. Grâce à ces deux analyses nous avons pu constater que les TIWs sont modifiées à la fois dans leur phasage (et donc leur vitesse) et dans leur amplitude (seul le run ES4 présente peu de modifications). Ces résultats sont suffisamment convaincants pour que nous choissions par la suite de construire une expérience d'ensemble à travers la perturbations des 3 paramètres.

6.4 Création de l'ensemble

Pour construire une expérience d'assimilation jumelle, nous avons besoin d'une vérité (le run de référence) et d'un run faux que nous tentons de corriger (run avec des paramètres perturbés). La connaissance de la vérité se fait à travers des observations simulées, il n'y a donc qu'une partie de l'océan vrai qui est connue. Par ailleurs, nous avons besoin d'une paramétrisation des statistiques d'erreur sur le run perturbé. En effet, dans une expérience réaliste d'assimilation, les erreurs associées au modèle ne sont pas connues. Ici cela revient à supposer ne pas connaître la valeur exacte des paramètres perturbés. Ainsi, pour obtenir ces statistiques d'erreur, nous allons utiliser une méthode de type ensemble qui permettra de simuler une "fourchette d'erreur" sur l'erreur modèle. Cette méthode est basée sur un tirage aléatoire de différentes valeurs de paramètres de diffusion dans

une gamme donnée, définissant une fourchette d'erreur liée au modèle. Pour paramétrer l'erreur sur les paramètres, nous choisissons une distribution lognormale (voir section 6.4.1) centrée autour d'une valeur de référence. Nous allons d'abord préciser la méthode de tirage aléatoire des paramètres puis nous détaillerons l'ensemble dans la section suivante.

6.4.1 Méthode de tirages aléatoires

Un tirage aléatoire ne se fait pas de façon aveugle. Il faut définir un cadre précis et pertinent pour notre problème. D'une part, on admet que la valeur des paramètres de diffusion du modèle de référence est une valeur réaliste, les autres valeurs tirées aléatoirement doivent donc rester dans le voisinage des valeurs de référence. D'autre part, il nous faut définir un écart maximal à la valeur de référence qui soit raisonnable. Comme les expériences de sensibilité menées précédemment utilisaient des valeurs à un ordre de grandeur près sans rendre la dynamique du modèle irréaliste, on peut supposer que cet écart est suffisant. Nous avons donc une valeur centrale (la valeur de référence) et une fourchette de plus ou moins un ordre de grandeur. Cela n'est néanmoins pas suffisant. On pourrait tirer aléatoirement des valeurs en tenant compte uniquement de ces deux paramètres, mais on aurait alors autant de chance d'avoir des valeurs extrêmes que des valeurs proches du centre. Si on suppose que la valeur de référence est quand même plutôt bien estimée (ou réaliste) et que les valeurs extrêmes sont des cas limites (et donc peu probables), il faut que les tirages aléatoires tiennent compte de ces probabilités. Nous avons donc choisi une distribution de probabilité log-normale (équation 6.1). Cette fonction permet d'obtenir un tirage aléatoire concentré symétriquement autour des valeurs de référence (la probabilité d'aller vers les valeurs extrêmes est d'autant plus faible). Cette loi est donnée en fonction des paramètres μ et σ qui permettent de contrôler le centrage et la dispersion des valeurs.

$$f(x) = \frac{1}{x\sigma\sqrt{2\Pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad (6.1)$$

Les 3 paramètres sont considérés comme indépendants, le tirage des valeurs pour chacun d'eux est donc effectué de façon autonome. Pour chaque paramètre, on effectue 100 tirages successifs. On regroupe ensuite aléatoirement les trois paramètres sous forme de triplets, donnant au final un jeu de 100 triplets aléatoires. Après avoir fait différents essais, nous avons fixé les valeurs de ces paramètres pour avoir une dispersion raisonnable. Pour $\mu = 0$ et $\sigma = 1$ on obtient des répartitions telles que montrées sur les figures 6.7 et 6.8. La figure 6.7 permet de vérifier que la répartition des tirages a bien la forme d'une loi log-normale. La densité des valeurs pour chaque paramètre est bien concentrée autour de la valeur de référence (indiquée par une croix rouge), avec un nombre de tirages dégressif vers les valeurs extrêmes. La figure 6.8 quant à elle permet de vérifier que chaque tirage est indépendant. Chaque tirage est ainsi représenté en fonction des 2 autres de son triplet sur les 3 premiers graphiques, et un quatrième graphique représente les triplets dans l'espace des 3 paramètres. Dans tous les cas, on ne peut observer aucune forme de corrélation entre les paramètres (les triplets se présentent sous forme de nuages de points).

A travers une méthode de tirages aléatoires suivant la probabilité de distribution d'une loi log-normale nous avons pu créer 100 triplets de paramètres de diffusion dynamique *a priori* indépendants entre eux. Ces 100 triplets vont nous permettre de simuler une erreur

modèle grâce à la création d'un ensemble que nous allons détailler dans la section qui suit.

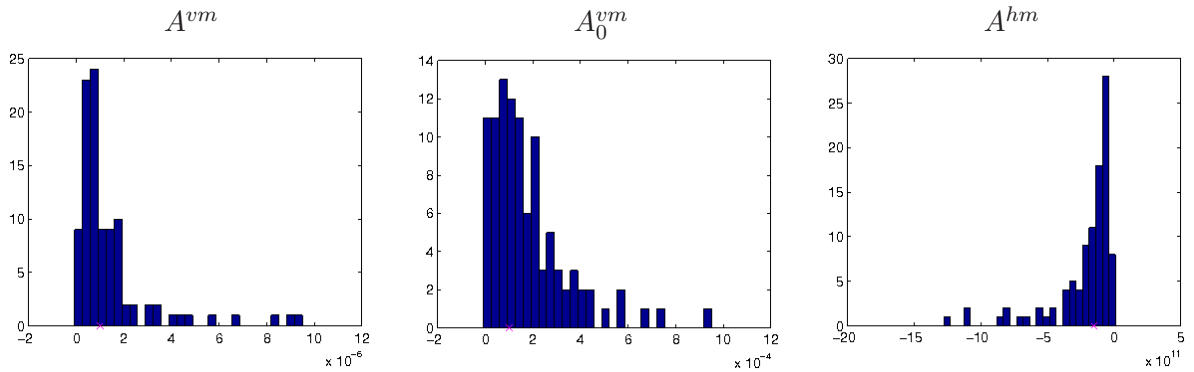


FIGURE 6.7 – Densités de répartition des valeurs tirées aléatoirement pour chaque paramètre de diffusion. Le total des tirages de chaque paramètre est de 100. L'axe des abscisse indique la valeur des tirages (et le facteur d'échelle associé), l'axe des ordonnées indique le nombre de tirages par tranches de $\frac{1}{3}$. Les croix rouges indiquent la position de la valeur de référence de chaque paramètre.

6.4.2 Création de l'ensemble

L'ensemble que nous allons créer consiste à décrire les statistiques d'erreur modèle considérée comme inconnue. Pour cela, nous allons effectuer une succession de runs libres à partir de la même date initiale que le run de référence et le run perturbé (jour 20170, soit le 5 avril 2005). Chacun de ces runs libres possède la même condition initiale et est propagé avec les mêmes forçages, mais il possède un triplet unique de paramètres de diffusion (issu du tirage aléatoire décrit précédemment). L'ensemble de ces runs crée donc un *ensemble* qui décrit les différentes trajectoires modèle possibles pour une gamme de paramètres donnés. Dans un premier temps, l'ensemble créé est de type fixe, c'est-à-dire qu'une fois défini il n'est pas influencé par l'assimilation de données.

La figure 6.9 montre l'écart RMS entre les runs de l'ensemble et la référence sur la température et sur la vitesse zonale. Le run perturbé est également indiqué par la courbe noire sur cette même figure. A première vue, l'ensemble est bien dispersé. L'écart RMS global évolue rapidement dans les premières semaines puis tend vers une croissance plus lente. De plus, le run perturbé est toujours situé au milieu des autres runs, l'ensemble permet donc *a priori* de donner le bon ordre de grandeur de l'erreur sur le run perturbé, en moyenne. Il donne, quelle que soit la date considérée, une représentation statistique de l'erreur introduite dans le modèle par la perturbation des paramètres. Avant de commencer les expériences d'assimilation, nous devons tout de même effectuer quelques vérifications sur la cohérence de l'ensemble.

6.4.3 Variances et corrélations

Les courbes d'évolution des écarts RMS calculées précédemment laissent supposer que l'ensemble se disperse correctement et que chaque membre est autonome (car issu d'un tirage aléatoire indépendant). Cette indépendance des membres est indispensable pour avoir un ensemble efficace. En effet, c'est cela qui permet d'avoir un espace d'erreur (ou un degré de liberté) maximal. Avec 100 membres, on aura donc au mieux 100 directions

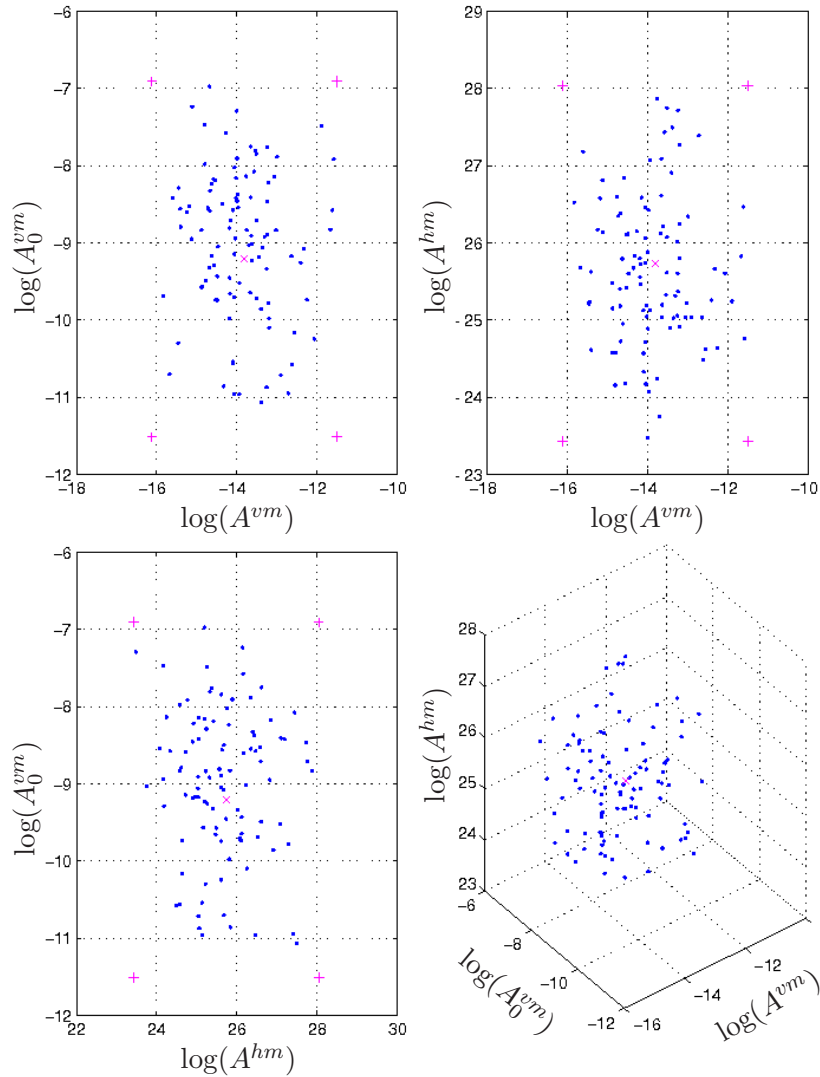


FIGURE 6.8 – Représentation des valeurs tirées aléatoirement de chaque paramètre de diffusion en fonction des autres : A_0^{vm} en fonction de A^{vm} (en haut à gauche), A^{hm} en fonction de A^{vm} (en haut à droite), A_0^{vm} en fonction de A^{hm} (en bas à gauche), A^{hm} en fonction de A^{vm} et de A_0^{vm} (en bas à droite). Les croix rouges (x) indiquent la position des valeurs de référence de chaque couple ou triplet de paramètres. Les + indiquent la position des couples extrêmes.

d'erreurs orthogonales entre elles. Pour nous assurer de ce point, nous avons effectué une décomposition en EOFs des 100 membres. Le résultat de cette décomposition est représenté sur les figures 6.10 et 6.11. Le pourcentage de variance (par rapport à la variance totale de l'ensemble) exprimé par chaque EOF (figure 6.10) indique clairement un problème d'indépendance. Si chaque membre était autonome, on devrait avoir quasiment une droite. Ici, on constate que presque 90% de la variance totale se situe sur la première EOF. Cela indique que l'on pourrait réduire l'espace d'erreur à 4 ou 5 directions. Dans le cadre de l'assimilation de données cet aspect peut être un avantage car le problème est *a priori* assez simple (les erreurs sont représentables dans un espace de 4 ou 5 dimensions seulement). De plus, la forme de la première EOF (illustrée par la figure 6.11) indique que la variabilité de

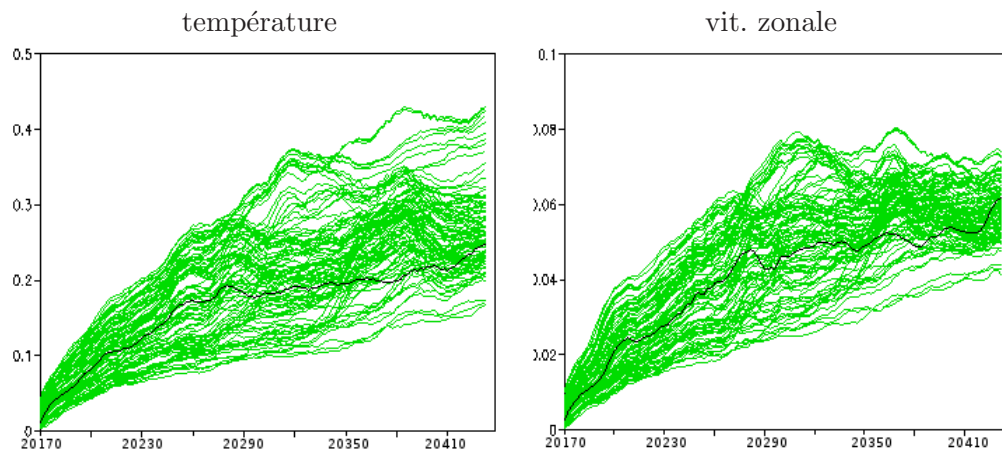


FIGURE 6.9 – Ecart RMS global entre les runs d'ensemble et le run de référence (en vert), sur la température en $^{\circ}\text{C}$ (à gauche) et la vitesse zonale en m.s^{-1} (à droite). La courbe noire indique l'écart RMS entre le run perturbé à corriger et le run de référence.

premier ordre s'effectue déjà selon des échelles spatiales très courtes. Les modes supérieurs présentent des échelles encore plus courtes. Cela est un point problématique : si les échelles spatiales sont trop réduites, alors les observations (ponctuelles) vont avoir un impact très localisé. En plus de ne pas corriger efficacement l'état global, cela peut aussi conduire à de fortes différences entre les zones observées et le reste du bassin, et donc conduire à des instabilités numériques. Pour contrôler efficacement un tel signal, il faudrait un réseau d'observations très dense, ce qui n'est pas réaliste ici.

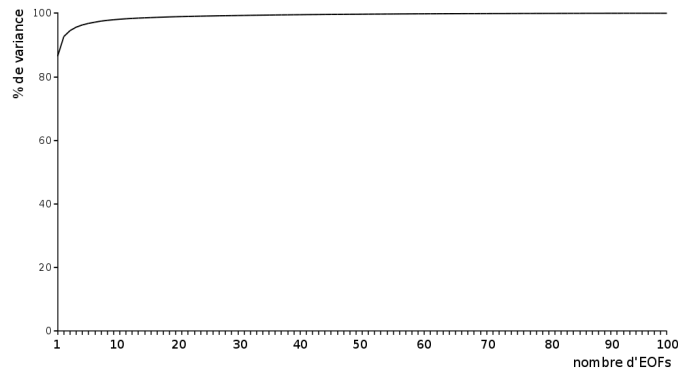


FIGURE 6.10 – Représentation de la répartition de la variance (en pourcentage de la variance totale) sur une décomposition en EOFs des runs d'ensemble.

6.5 Conclusion

L'ensemble créé par la perturbation des paramètres de diffusion était à première vue concluant. Les tests de sensibilité ont montré l'impact significatif de la perturbation des paramètres sur la dynamique du système et notamment sur la propagation des TIWs. La méthode de tirage aléatoire s'est avérée efficace pour créer un jeu de triplets de paramètres

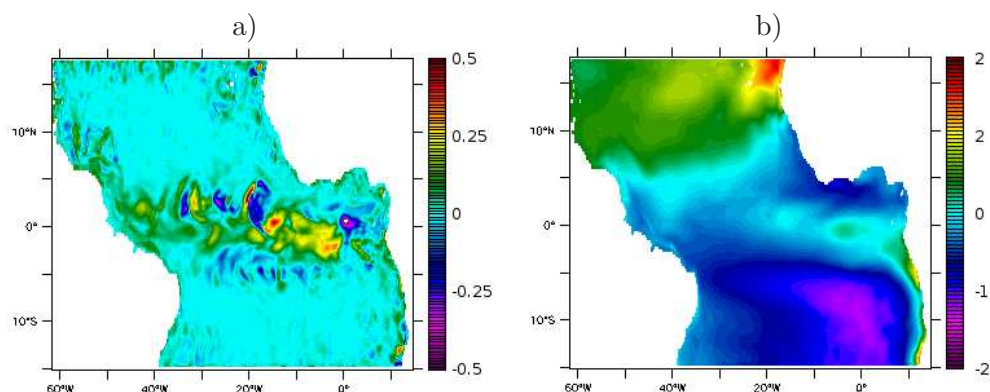


FIGURE 6.11 – Carte de la température de surface (en °C) de la 1ere EOF calculée à partir de l'ensemble perturbé (à gauche) et d'un run interannuel (à droite).

indépendants les uns des autres. L'étude plus précise de l'indépendance des membres associés à ces triplets a mis en évidence la faible dimension de l'espace de cet ensemble, ce qui aurait pû être un atout pour un problème d'assimilation de données. Néanmoins, les structures des erreurs générées (et les échelles des covariances d'erreur), comme montrées sur les figures 6.6 ou 6.11, se sont révélées beaucoup trop complexes pour pouvoir être contrôlées avec un système d'observations réduit. Pour que l'erreur modèle générée par la perturbation de paramètres soit efficace, il aurait fallu que ces perturbations se traduisent par des variations de grande échelle spatiale sur les grandeurs physiques du système, ou alors il faudrait disposer d'un réseau d'observations suffisamment dense pour contrôler le signal des perturbations, ce qui est exclu dans un cadre réaliste. Les quelques tests effectués (non détaillés ici) ont confirmé l'inefficacité du filtre SEEK et encore plus du lisseur. Dans ces expériences, l'assimilation de données a tendance à tuer les grandes échelles et créer une solution en déséquilibre numérique.

Nous avons donc choisi par la suite de nous tourner vers une expérience d'assimilation plus classique en nous appuyant directement sur la variabilité interannuelle du système pour créer une erreur sur la condition initiale. Le problème n'est donc plus posé de la même façon, et nous allons décrire dans le chapitre suivant la mise en place de cette nouvelle piste.

Chapitre 7

Initialisation et paramétrisation du filtre évolutif

Sommaire

7.1	Introduction	102
7.2	La variabilité interannuelle, une source d'incertitude	102
7.2.1	Etat de référence	102
7.2.2	Création d'un état faux	104
7.2.3	Construction des statistiques d'erreur initiales	105
7.3	Création d'observations (parfaites)	106
7.3.1	Traces satellites	107
7.3.2	Réseau <i>in situ</i>	107
7.4	Paramétrisation du filtre SEEK évolutif : efficacité de l'analyse	109
7.4.1	Taille de l'analyse locale	111
7.4.2	Détermination de la taille de la matrice d'erreur S	112
7.4.3	Influence des observations	114
7.5	Paramétrisation du filtre SEEK évolutif : propagation de l'erreur	115
7.5.1	Taille de l'analyse locale	116
7.5.2	Détermination de la taille de la matrice d'erreur S	116
7.5.3	Détermination de la matrice des erreurs d'observation	118
7.6	Conclusion	119

7.1 Introduction

Dans le chapitre précédent, nous avons vu que créer un ensemble en perturbant certains paramètres intrinsèques du modèle n'était pas forcément pertinent pour concevoir une expérience d'assimilation convenable. Toujours dans un souci d'étudier la mise en place d'un filtre évolutif dans les tropiques, nous allons donc dans ce chapitre utiliser une autre source d'incertitude pour créer notre expérience jumelle : la variabilité interannuelle de l'océan. Cette variabilité a l'avantage d'être une source d'incertitude réaliste et totalement naturelle. Elle peut intervenir lorsqu'il s'agit de prévoir l'état futur de l'océan sur plusieurs années : on considère alors que l'état prédit doit se trouver dans la gamme des oscillations interannuelles connues, avec tout de même une certaine marge de tolérance car certaines années peuvent varier plus fortement que d'autres. A la différence du chapitre précédent, nous utiliserons la source d'incertitude pour créer une erreur sur la condition initiale et non plus sur le modèle. L'erreur modèle ne sera plus prise en compte directement dans ces expériences. Seul le terme propagatif des covariances d'erreurs permettra de lier les erreurs temporellement. Ce terme sera particulièrement étudié dans ce chapitre puisqu'il s'agit de l'aspect le plus délicat dans la mise en place d'un filtre évolutif. Nous allons présenter dans un premier temps la mise en place des expériences jumelles (section 7.2) avant de nous concentrer sur la paramétrisation du filtre évolutif dans le cadre de ces expériences.

7.2 La variabilité interannuelle, une source d'incertitude

Un run libre interannuel (appelé REF) est initialisé au premier janvier 1995 avec une sortie du run global DRAKKAR B83 (*Dussin et al.*, 2009). Il est forcé aux frontières par des sorties moyennées sur cinq jours de B83. Les forçages atmosphériques sont identiques au run B83 (DFS4, voir chapitre 3), lissés sur cinq jours pour éviter les oscillations parasites de la SSH (voir section 3.6). REF est propagé librement jusqu'au 31 décembre 2005. La figure 7.1 montre l'évolution des différentes variables moyennées sur les cent premiers mètres (là où la variabilité est la plus forte car directement influencée par les forçages atmosphériques) et permet d'avoir une idée des amplitudes de variabilité. On peut déjà noter que la SSH a une moyenne constante (pas de biais) et de très petites oscillations : cela provient de la conservation des flux imposés aux frontières qui permet de prévenir une éventuelle dérive du volume d'eau dans le bassin et donc une dérive en SSH, voir chapitre 3. Les oscillations sur les autres variables sont plus significatives. On peut tout de même noter qu'il n'y a *a priori* pas de dérive de ces variables et que leur moyenne reste constante. Seule la salinité présente un saut autour de 1997, sans doute associé aux forçages imposés. Si ce saut ne perturbe pas la dynamique du domaine, il faut tout de même noter qu'il se trouve dans la période où seront calculées les EOFs par la suite (voir section 7.2.3). Cela peut donc limiter la bonne représentation des statistiques par la méthode de décomposition en EOFs sur cette période. Néanmoins l'ordre de grandeur de ce biais reste tout relatif par rapport à la valeur moyenne de la salinité.

7.2.1 Etat de référence

L'année 2005 de REF est considérée dans notre expérience comme étant l'état de référence, c'est-à-dire l'océan vrai (celui que l'on cherchera à atteindre). L'état 3D complet

7.2.2 Création d'un état faux

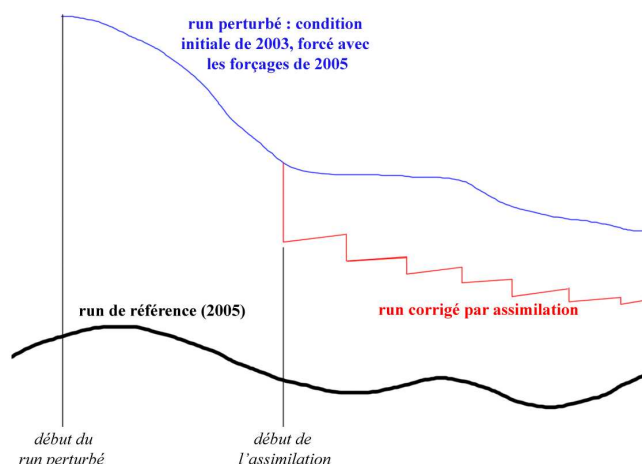


FIGURE 7.2 – Schématisation de l'expérience d'assimilation. L'état de référence (année 2005) est indiquée en trait noir épais; l'état perturbé par la condition initiale est indiqué par un trait bleu fin; enfin l'état corrigé par assimilation est indiqué en trait rouge fin.

Un moyen simple pour générer une erreur en utilisant la variabilité interannuelle est d'invertir l'état à une date donnée sur l'année de référence avec l'état de la même date mais d'une année différence. On crée alors l'état initial d'une simulation dite fausse. Il s'agit par exemple ici d'utiliser le 26 Mai 2003 pour initialiser un run sur l'année 2005 commençant au 26 Mai 2005 (figure 7.2). Cette date est choisie car elle se trouve juste avant la période des TIWs, période à laquelle nous nous intéressons particulièrement. L'écart entre le 26 Mai 2005 et le 26 Mai 2003 représente une erreur sur la condition initiale. Cette source d'erreur est pertinente dans le cadre de prévisions par exemple. Le run libre propagé à partir de la fausse condition initiale diffèrera alors de la trajectoire de référence uniquement à cause de l'erreur initiale (les forçages et les conditions aux frontières restant ceux de 2005). Dans les expériences réelles, on peut espérer que la condition initiale pour un run de prévision soit déjà en équilibre approximatif avec le modèle et les forçages extérieurs. De plus, l'erreur sur cette condition est minimisée par les analyses précédentes, elle est donc *a priori* moins importante que l'erreur que nous avons générée en expérience jumelle. Pour nous rapprocher un peu plus d'un problème d'initialisation réaliste et pour éviter les chocs au redémarrage du modèle dus à un déséquilibre entre la condition initiale et les forçages, nous choisissons de laisser le run faux s'équilibrer naturellement durant 1 mois. Les expériences d'assimilation à proprement parler commencent donc au 25 Juin 2005. L'erreur initiale de ces expériences (notée ϵ_0) représente alors la part de la dynamique qui ne s'est pas totalement équilibrée avec les forçages de 2005 (figure 7.2). On peut déjà noter que l'erreur sur la SSH diminue naturellement de 50% en environ 3 mois (figure 7.3). Elle est essentiellement située dans le centre et à l'Ouest du bassin, avec des traces très spécifiques au niveau de la côte nord-est du Brésil caractérisant le passage de tourbillons. L'erreur sur la salinité décroît rapidement dans un premier temps avant de se stabiliser. Elle est essentiellement liée aux apports d'eaux douces par l'amazone au Nord-Est du Brésil qui

peuvent varier fortement d'une année à l'autre. L'erreur en température a plutôt tendance à décroître régulièrement avec le temps. Ces deux erreurs sont essentiellement concentrées dans les couches de surface (figure 7.5 a et b). Quand aux erreurs sur les vitesses, la plus grande part est également concentrée sur les couches de surface et décroît légèrement naturellement (figure 7.5 c et d) mais il existe également une erreur en profondeur de l'ordre de 2cm.s^{-1} qui n'est pas corrigée naturellement. Cette erreur est essentiellement liée au courant de bord ouest le long de la côte du Brésil comme l'illustre la figure 7.6 et elle ne semble pas s'atténuer clairement dans le temps. Dans le reste du bassin, l'erreur sur les vitesses en profondeur est assez faible. La correction des champs de vitesse s'avère donc être l'un des défis à relever. De plus la différence de temps de réponse entre la SSH et les vitesses laissent supposer qu'un réseau altimétrique seul ne sera pas capable de corriger les champs dynamiques efficacement dans cette région, et les réseaux d'observation in-situ ne semblent pas assez profonds pour atteindre les zones d'erreur profonde en vitesse. Il faudra donc être conscient que la correction de cette erreur en profondeur est l'une des limites de notre expérience d'assimilation et nous serons amenés à nous concentrer essentiellement sur l'erreur dans les couches superficielles.

L'erreur ϵ_0 générée dans cette expérience est dûe à l'incertitude sur la variabilité du système et reste *a priori* incluse dans la gamme de variabilité interannuelle. Les statistiques d'erreur du système doivent donc représenter cette incertitude (section 7.2.3).

7.2.3 Construction des statistiques d'erreur initiales

Puisque l'erreur considérée dans cette expérience est uniquement liée à la variabilité interannuelle, nous choisissons de construire notre base de statistiques d'erreurs initiale (\mathbf{S}_0^f) par une décomposition en EOFs du signal contenu dans les premières années du run de référence REF. Les EOFs sont calculées sur une période de 6 ans, de 1995 à 2000 incluses, et prennent en compte un état tous les 4 jours durant les périodes été-automne (de juin à décembre). Ainsi, la haute fréquence (inférieur à 4 jours) est naturellement éliminée. De plus, seul le signal saisonnier du modèle sur lequel nous faisons nos expériences sera pris en compte. On peut espérer, en regardant la figure 7.1, que la variabilité de 1995 à 2000 soit représentative des variations de 2005.

La décomposition en EOFs permet de faire ressortir les modes principaux de variabilité. Ainsi, selon le nombre d'EOFs que l'on garde, on peut tronquer plus ou moins la variance de notre système et ainsi éliminer les modes de variabilité considérés comme non significatifs. La figure 7.7 montre la part de la variance totale exprimée en fonction du nombre d'EOFs sélectionnés. On peut noter que plus de 90% de la variance totale est concentrée sur les 25 premières EOFs. Le nombre d'EOFs à sélectionner sera étudié plus en détail dans les parties 7.5.2 et 7.4.2.

On note des modes de grande échelle sur la première EOF, comme l'illustre la figure 7.8, avec notamment une grande variabilité autour de la bande équatoriale. Cette variabilité tend ensuite vers de la plus petite échelle, comme l'illustre le 25ème EOF (figure 7.8), et fait ressortir des modes au nord-est du Brésil, au niveau de la réflexion du NBC (North Brazil Current).

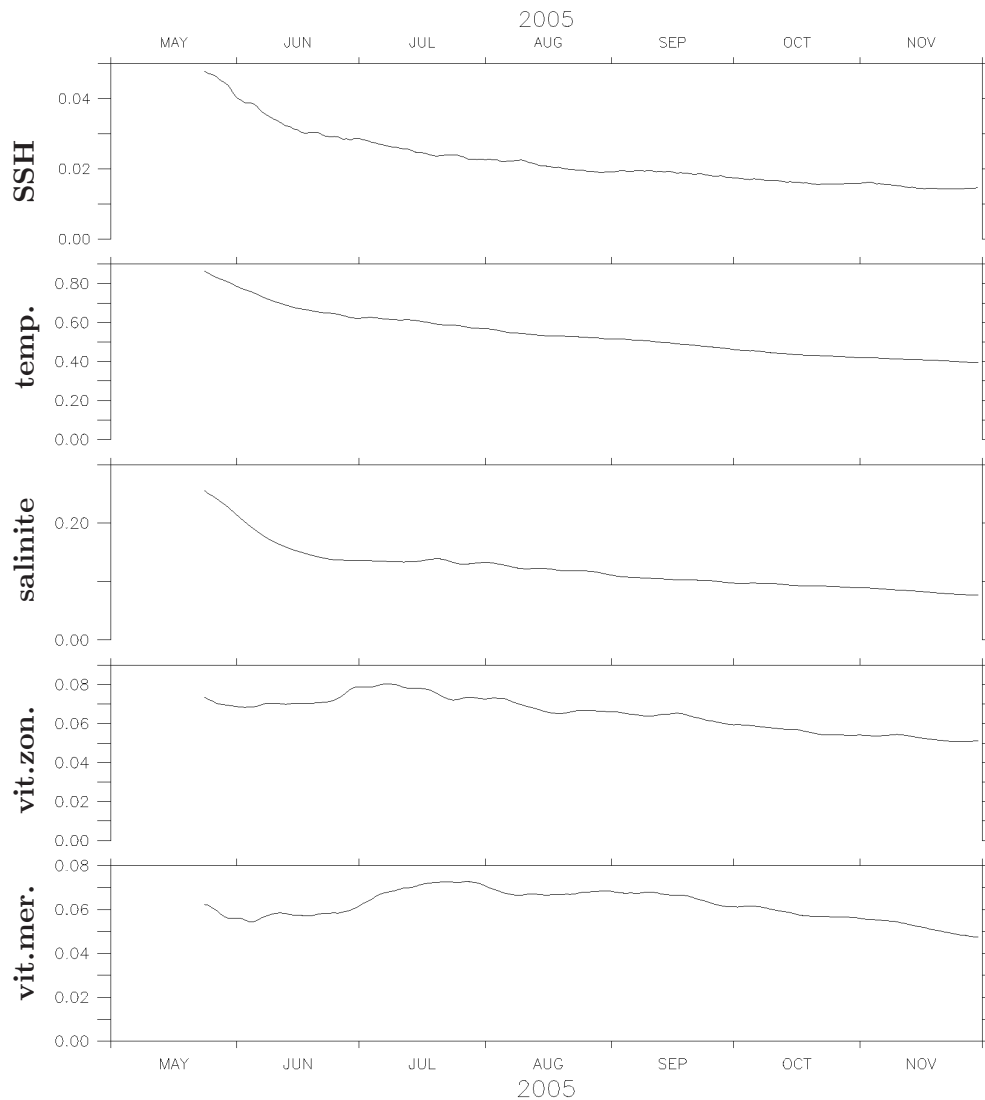


FIGURE 7.3 – Evolution de l'écart RMS sur le champ 3D (sur toute la profondeur) entre le run faux et le run de référence durant les premiers mois du run libre faux (de haut en bas : SSH (en m), température (en °C), salinité, vitesse zonale (en m.s^{-1}), vitesse méridienne (en m.s^{-1})).

7.3 Création d'observations (parfaites)

Comme nous l'avons vu dans le chapitre 2, les observations de l'océan sont de plus en plus nombreuses et diversifiées. L'avantage de travailler en expérience jumelle est de pouvoir choisir et simuler n'importe quel type d'observation. Dans notre cas, nous avons simulé des traces de satellite de type Jason donnant une information sur la hauteur de l'océan (section 7.3.1) et d'autre part des relevés in-situ de types flotteurs ARGO ou PIRATA donnant des informations sur la température et la salinité (section 7.3.2). Chacune de ces observations est extraite de l'année 2005 et ne présente pas d'erreur de mesure (il n'y a pas d'ajout de bruit artificiel). Ce choix de négliger l'erreur de mesure vient du fait

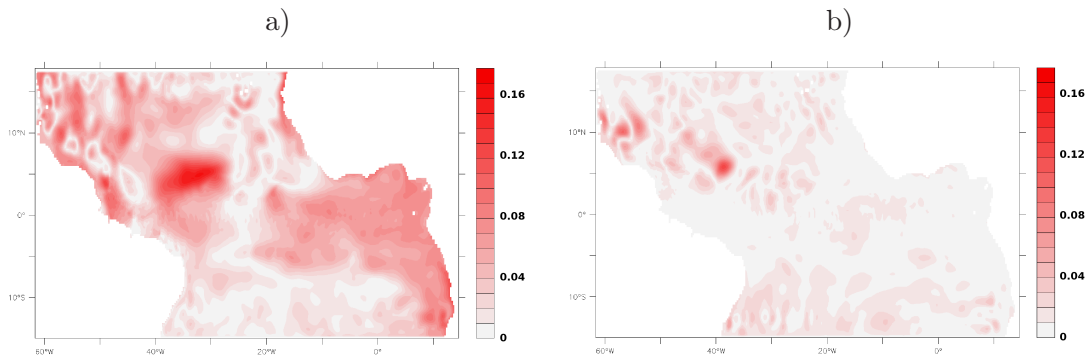


FIGURE 7.4 – Cartes de l'écart absolu en SSH (en m) entre le run faux et le run de référence au 26 Mai 2005 (a) et au 30 novembre 2005 (b).

que l'erreur associée aux observations sera essentiellement due à l'erreur de troncature dans notre système, comme nous le verrons dans la section 7.5.3.

7.3.1 Traces satellites

Les observations de SSH simulées sont extraites sous forme de traces satellite de type Jason. Un cycle complet est effectué en 10 jours (figure 7.9,a). A chaque étape d'assimilation (tous les deux jours), une partie seulement de cette couverture est disponible. Un exemple de ce sous-échantillonnage est donné par la figure 7.9,c). Bien que la forme des traces satellites soit réaliste, il existe deux simplifications fondamentales par rapport à des relevés satellites réels. D'une part ici chaque observation en un point correspond directement à un point de grille complet. Il n'y a donc pas d'erreur de représentativité spatiale. Les vraies traces satellite de SSH sont assez ponctuelles (de l'ordre de $7 \times 7 \text{ km}^2$), il faudrait donc prendre en compte une erreur liée au fait qu'un point précis représenterait une maille de grille complète (de l'ordre de $27 \times 27 \text{ km}^2$). D'autre part, à chaque cycle d'assimilation nous considérons que le sous-échantillonnage de traces disponibles est issu d'un même instant (il s'agit en quelque sorte d'une photo d'un instant donné). Il n'y a donc pas non plus d'erreur de représentativité temporelle. Il en serait autrement avec des traces réelles puisque le long d'une trace chaque point est observé à un instant différent. Pour avoir une couverture identique à celle dont nous disposons pour une analyse il faudrait regrouper des relevés sur deux jours consécutifs et estimer que malgré le décalage l'ensemble de ces relevés est représentatif du même état. Ces deux simplifications permettent de limiter les difficultés liées à l'assimilation d'observations très ponctuelles et continues dans le temps.

7.3.2 Réseau *in situ*

Les observations de température et de salinité simulées sont extraites sous forme de profils verticaux ponctuels (de type profileurs ARGO) allant jusqu'à une profondeur de 2000m. Cette profondeur devrait suffire pour corriger les erreurs en température et salinité qui se concentrent dans les couches superficielles mais elle pourrait se révéler insuffisante pour corriger les champs de vitesse dans les grandes profondeurs (7.5). On considère que tout le réseau n'émet pas un signal au même moment. Ainsi, à une date donnée, des profils sont disponibles tous les 18 points de grille (soit $4,5^\circ$). A la date suivante d'autres profils

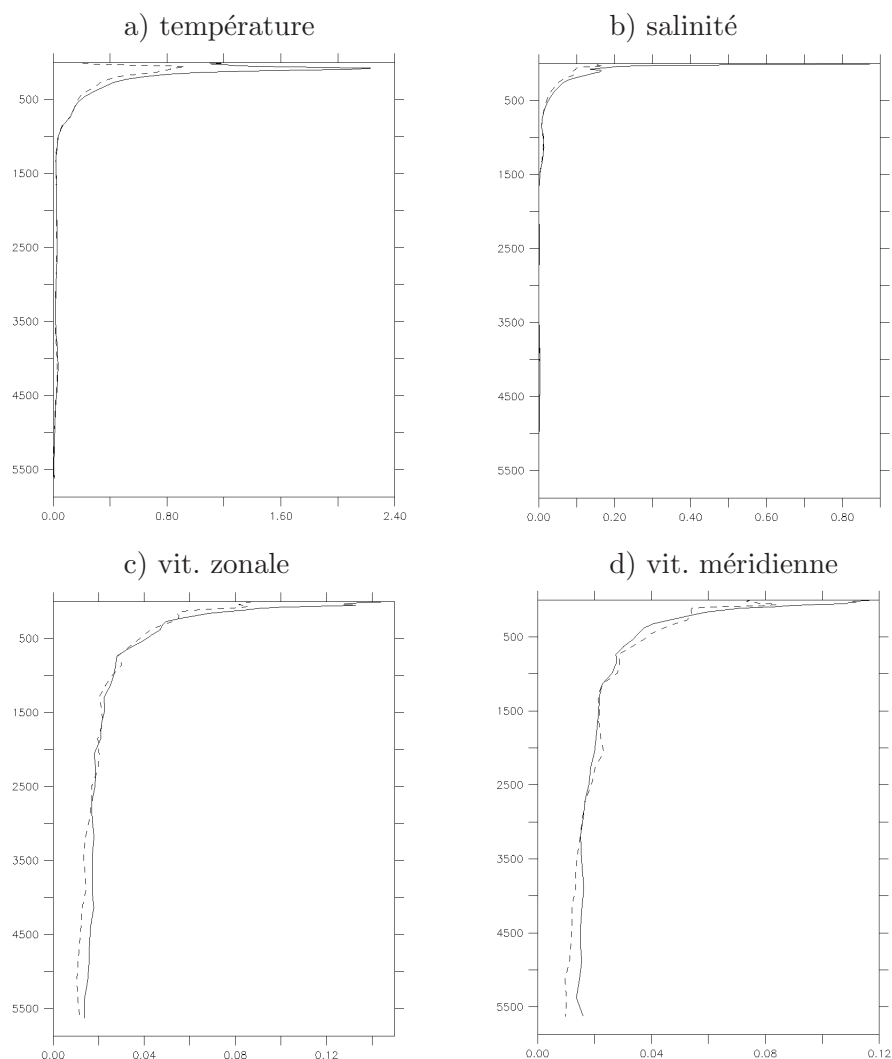


FIGURE 7.5 – Profils verticaux de l'écart RMS entre le run faux et le run de référence, moyenné horizontalement au 26 Mai 2005 (trait plein) et au 30 novembre 2005 (trait en pointillés) pour : a) la température (en $^{\circ}\text{C}$), b) la salinité, c) la vitesse zonale (en m.s^{-1}) et d) la vitesse méridienne (m.s^{-1}).

seront disponibles (toujours espacés de $4,5^{\circ}$) mais pas aux mêmes endroits (figure 7.9,b)). Les profils seront décalés ainsi de suite à chaque date jusqu'à revenir à leur position initiale. Un cycle complet est effectué en 20 jours (la couverture spatiale complète est donnée sur la figure 7.9,d)). La densité de profils et donc de 6 points de grilles ($1,5^{\circ}$) sur un intervalle de 20 jours, soit en moyenne 3° sur 10 jours. Cette densité se rapproche de la couverture théorique du réseau ARGO. Tout comme pour les observations de la SSH, les relevés *in situ* présentent des simplifications. Une observation simulée correspond directement à une maille du modèle, et les observations pour une date donnée sont représentatives du même instant.

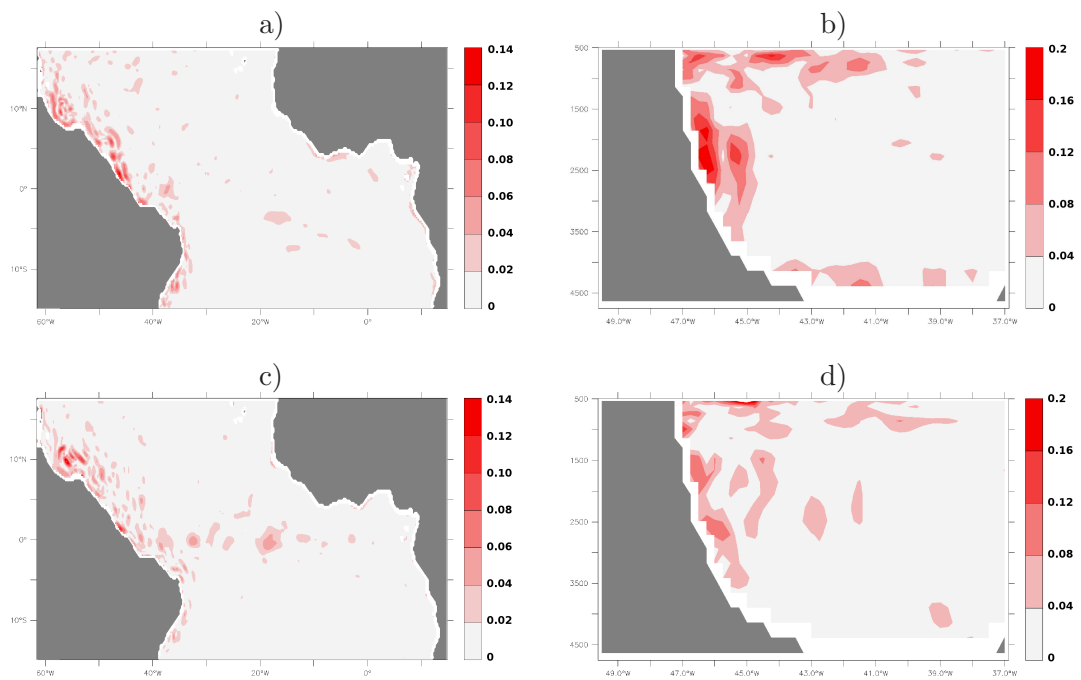


FIGURE 7.6 – Erreur RMS au 26 mai 2005 (en haut) et au 30 novembre 2005 (en bas) entre le run faux et le run de référence pour le champ de vitesse absolue, moyennée sur la vecticale entre 1000m et 4000m (à gauche) et en coupe E-W à 3°N au niveau de la côte du Brésil (à droite).

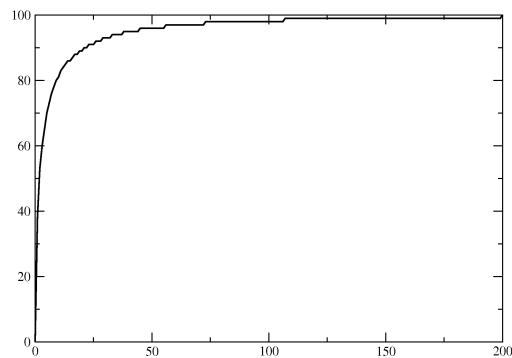


FIGURE 7.7 – Pourcentage de la variance totale en fonction du nombre d'EOFs.

7.4 Paramétrisation du filtre SEEK évolutif : efficacité de l'analyse

Avant de mettre en place le lisseur, la première étape importante est de déterminer quels sont les paramètres optimaux pour obtenir un filtre efficace. Le système d'observation étant dans un premier temps fixé, il reste 3 éléments à définir : la zone d'influence de l'analyse, la taille de la matrice d'erreur d'ébauche (c'est-à-dire le nombre d'EOFs

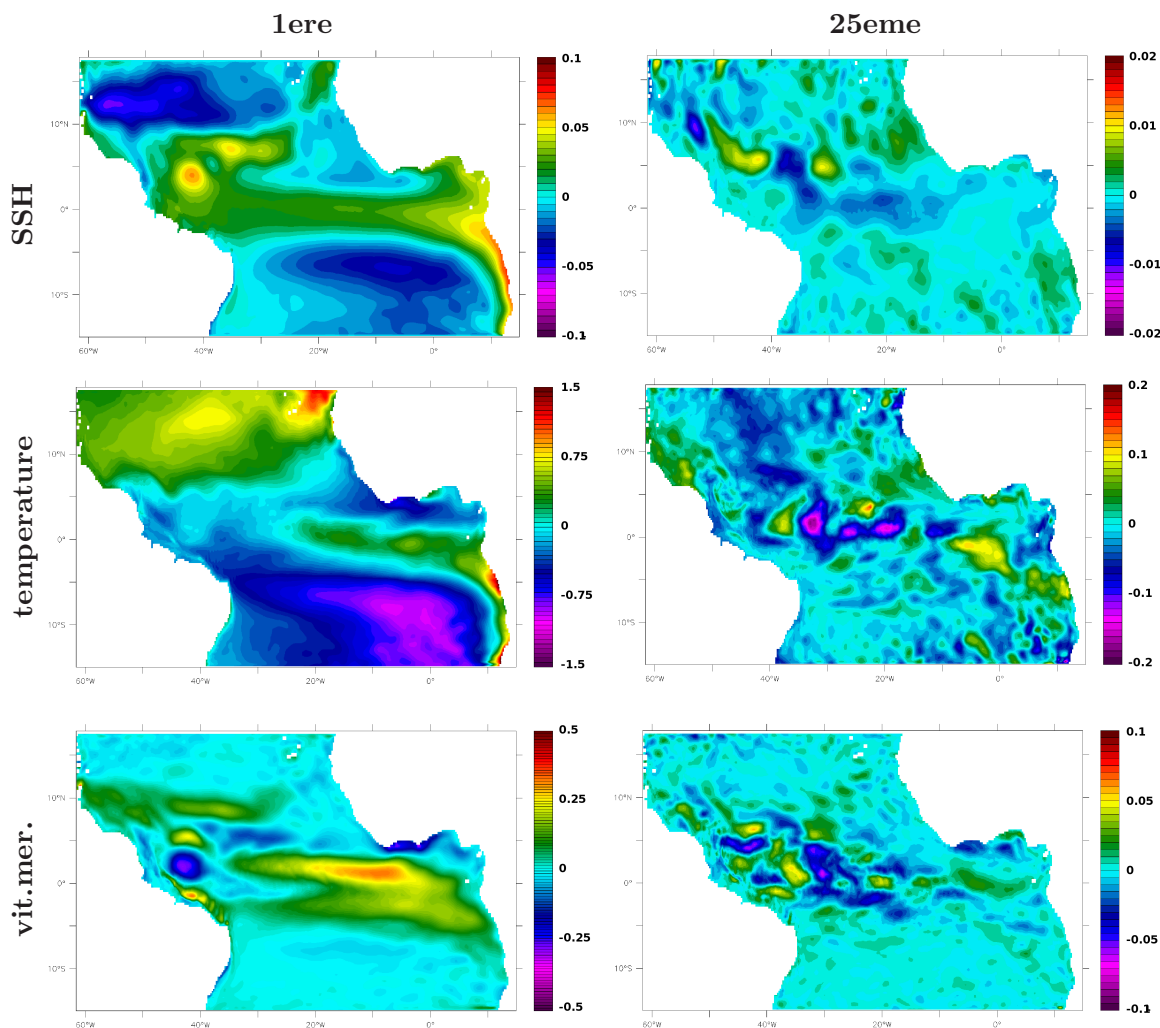


FIGURE 7.8 – Représentation des modes de variabilité par les EOFs (à gauche : 1ère EOF, à droite : 25ème EOF, en haut : la SSH (en m), au milieu : la température (en °C), en bas : la vitesse zonale (en $\text{m}\cdot\text{s}^{-1}$)).

utilisées pour représenter l'erreur d'ébauche) et la paramétrisation de la matrice des erreurs d'observation. Ces paramètres doivent être choisis selon deux critères (le tout étant bien entendu soumis à une contrainte de coût numérique qui doit rester raisonnable) : d'une part il faut que l'analyse soit la plus efficace possible à un instant donné, mais il faut également que les statistiques propagatives du filtre restent cohérentes dans le temps. Ce deuxième point est l'un des aspects les plus délicats à mettre en place et sera traité dans la section 7.5.

Dans les sous-sections qui suivent nous détaillons donc l'influence des 3 paramètres sur l'efficacité du filtre SEEK pour une analyse unique, effectuée avec la base d'EOFs initiale au premier jour de l'expérience (25 Juin 2005). A noter que ces résultats sont valables dans le cadre de notre expérience et ne peuvent pas être généralisés sans prudence à n'importe quel système.

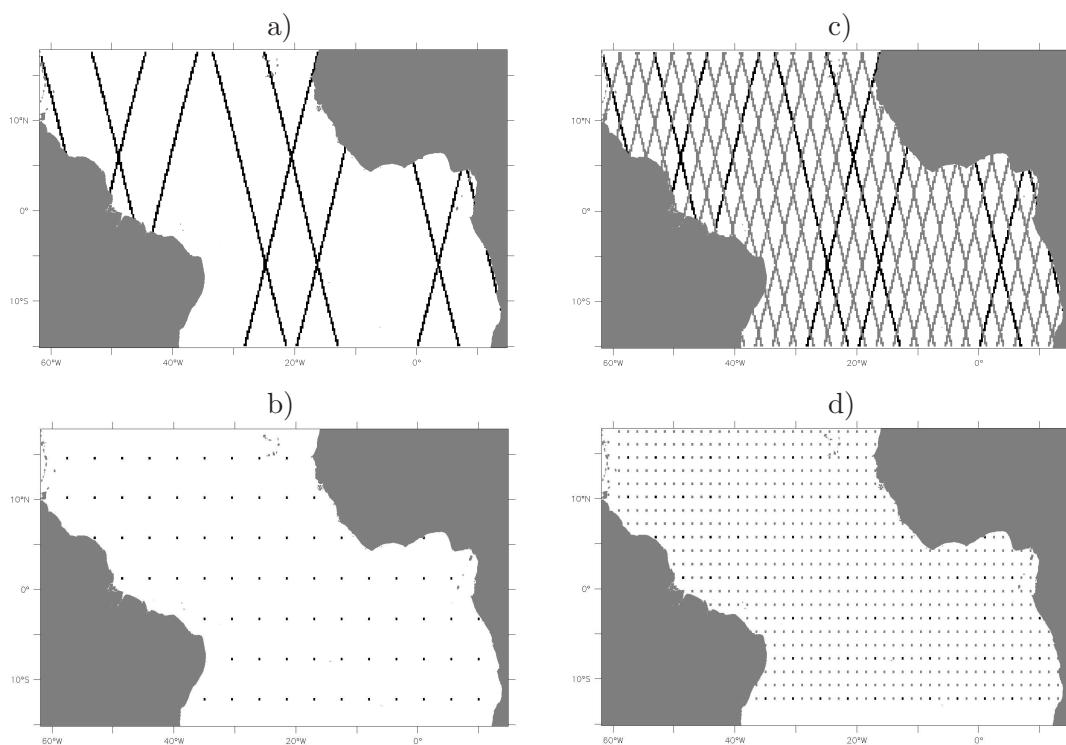


FIGURE 7.9 – Représentation du système d'observation de SSH (en haut) et de T/S (en bas), pour la couverture sur 10 jours (à droite) ou la couverture partiellement disponible pour une analyse, c'est-à-dire tous les deux jours (à gauche).

7.4.1 Taille de l'analyse locale

Pour étudier la taille de l'analyse locale, nous fixons un nombre d'EOFs à 24 et une erreur d'observation très faible (10^{-3}) afin de forcer au maximum l'analyse à tendre vers les observations. On se rapproche donc d'un cas où les observations sont parfaites et $R \rightarrow 0$. Pour différentes tailles de la zone d'influence de l'analyse locale, nous faisons une analyse unique avec la base d'EOFs initiale et le système d'observation décrit précédemment. Les zones d'influences utilisées sont résumées dans le tableau 7.1 et les résultats de chaque analyse sont indiqués sur la figure 7.10. L'erreur résiduelle pour une analyse globale (avec une zone d'influence couvrant tout le domaine) est également indiquée sur cette figure (elle correspond au point d'abscisse 150). Ces résultats montrent l'erreur moyennée sur l'état 2D pour la SSH et 3D pour la température et la vitesse zonale. Les résultats pour la salinité et la vitesse méridienne (non montrés) sont respectivement identiques à ceux de la température et de la vitesse zonale.

Il apparaît très clairement que les zones d'influences de grande échelle (les zones 30 à 100) sont les plus efficaces. Les zones de petites tailles obtiennent toujours un score plus faible quelles que soient les variables considérées. De plus, les zones de petite échelle présentent le désavantage de pouvoir introduire des perturbations dans le système. Si par exemple la correction se fait bien mais très localement autour des observations, on risque, lors du redémarrage, de voir naître des instabilités dues aux écarts trop forts et trop abrupts entre les zones bien corrigées et les zones où il n'y a pas du tout de correction. Les zones de très grandes tailles (60 ou 100) présentent quand à elle le désavantage d'être assez coûteuses en

temps de calcul, et elles n'améliorent pas de beaucoup les scores obtenus par les zones 30 ou 40. Pour la suite donc nous privilégierons des zones 30 ou 40 qui permettent d'obtenir une solution plus lisse entre les points d'observations pour un coût raisonnable. On peut également noter que le résultat avec une analyse globale (point d'abscisse 150) est moins bon que la plupart des analyses locales. Utiliser la version locale du filtre SEEK dans cette expérience est donc justifié. A noter tout de même que ces résultats sont sans doute très dépendant du type de réseau d'observation utilisé. Avec une couverture d'observation plus dense par exemple, les petites zones d'influences pourraient avoir un meilleur impact.

Dénomination	Longueur zonale	Longueur méridienne
4	4(3)	4(3)
8	8(2)	8(2)
12	12(5)	12(5)
20	20(7)	20(7)
30	30(20)	20(13)
40	40(30)	30(20)
60	60(45)	45(30)
100	100(80)	80(60)

TABLE 7.1 – Dénominations et tailles des différentes zones d'influence utilisées (en nombre de points de grille). Le premier chiffre indique la zone de coupure (c'est-à-dire la zone au-delà de laquelle les observations n'ont plus d'influence) et le chiffre entre parenthèses indique la zone d'influence maximale (c'est-à-dire la zone au-delà de laquelle le poids des observations diminue pour atteindre 0).

7.4.2 Détermination de la taille de la matrice d'erreur S

Comme nous l'avons déjà évoqué dans le chapitre 5, la troncature de la matrice d'erreur d'ébauche (autrement dit le choix du nombre d'EOFs initiales) peut se faire selon plusieurs critères.

Les critères quelque peu arbitraires (comme le pourcentage de variance totale représenté par un certain nombre d'EOFs) ne tiennent pas forcément compte de l'efficacité de la base d'EOFs. Il faut en effet considérer que la variance totale représentée par l'ensemble des EOFs n'est en réalité qu'une partie du signal océanique interannuel. En effet, le signal est déjà tronqué par la méthode de calcul des EOFs. De plus, la variabilité contenue dans les EOFs est issue d'années différentes de celle où les expériences d'assimilation sont menées. Il est donc tout à fait possible qu'une part de cette variabilité ne soit pas compatible avec l'erreur créée dans le cadre de notre expérience, ni avec le jeu d'observation.

Pour examiner l'efficacité de notre base d'EOFs, nous avons procédé à des tests semblables à ceux menés précédemment pour la zone d'influence de l'analyse. Ainsi, on fixe l'erreur d'observation à 10^{-3} et nous choisissons la zone d'influence 30 (tableau 7.1). Nous effectuons ensuite des analyses en sélectionnant différents nombres d'EOFs. Les résultats (figure 7.11) montrent l'erreur résiduelle moyennée sur l'état 3D après l'analyse en fonction du nombre d'EOFs choisi. Ils sont donnés pour la SSH, la température et la vitesse zonale, mais les résultats pour la salinité et la vitesse méridienne (non montrés) sont respectivement identiques à ceux de la température et de la vitesse zonale.

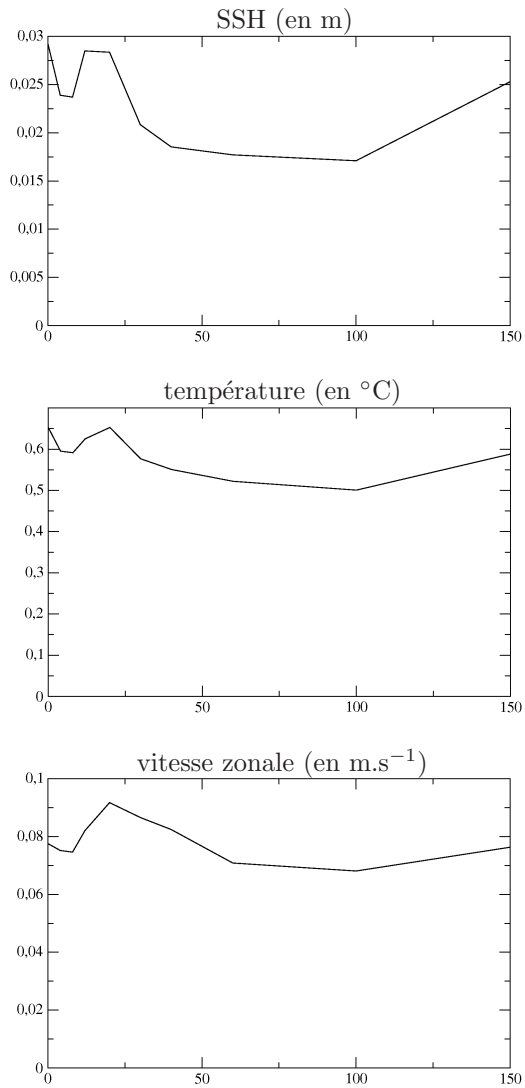


FIGURE 7.10 – Erreur RMS résiduelle sur le champ 3D après une analyse pour les différentes variables d'état, en fonction de la taille de la zone d'influence du filtre SEEK. Le point d'abscisse 0 correspond à l'erreur initiale avant l'analyse, et le point d'abscisse 150 correspond à une analyse globale.

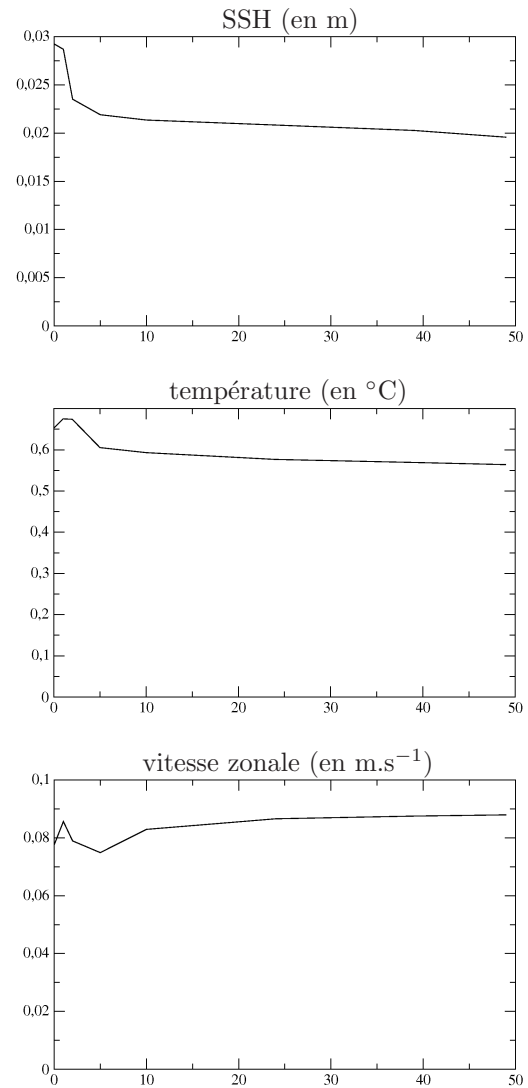


FIGURE 7.11 – Erreur RMS résiduelle sur le champ 3D après une analyse pour les différentes variables d'état, en fonction du nombre d'EOFs utilisées (la valeur à 0 EOF correspond à l'erreur initiale avant l'analyse).

On peut constater que l'erreur résiduelle sur les variables assimilées (SSH, température et salinité) décroît globalement quand le nombre d'EOFs augmente. Il faut tout de même noter que pour la température et la salinité les deux premières EOFs ont plutôt tendance à dégrader le système. Ainsi les principaux modes de variabilité issus des années 1995 à 2000 ne semblent pas forcément efficaces pour corriger l'erreur initiale sur ces champs durant l'année 2005. Une autre explication est également que la localisation de l'analyse ne tient pas compte des grandes échelles de ces premiers modes et ne permet donc pas

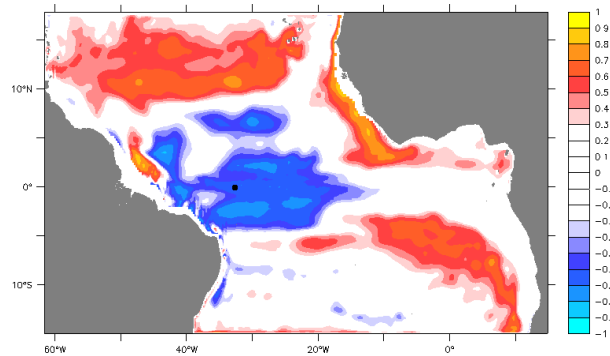


FIGURE 7.12 – Corrélation entre la vitesse méridienne en un point (symbolisé par le cercle noir) et la température sur le reste du bassin (en °C).

d’effectuer une analyse correcte. Sur les autres variables diagnostiquées (les composantes zonales et méridiennes du champ de vitesse) les résultats montrent plutôt une très légère tendance à l’augmentation de l’erreur avec le nombre d’EOFs au-delà des cinq premières EOFs. Cette tendance peut traduire le fait que les corrélations de courte distance entre les variables ne sont pas définies efficacement. Cela signifie que les statistiques d’erreur ne sont pas totalement conformes à l’erreur réelle. Le propagation dynamique des modes semble donc nécessaire pour approcher de façon plus réaliste les erreurs de petite échelle. De plus, si on regarde par exemple les corrélations entre un point de vitesse méridienne (où l’analyse dégrade l’état) et les autres variables (figure 7.12), on s’aperçoit que certaines corrélations semblent exagérées. Des points très éloignés en salinité peuvent avoir une corrélation de presque 100% avec la vitesse au point considéré. Il est donc primordial de limiter l’influence des observations, tant en amplitude (en paramétrant efficacement la matrice R) que spatialement.

7.4.3 Influence des observations

Dans cette dernière sous-partie nous allons regarder l’influence de l’assimilation de chaque variable sur la correction du champ 3D (la paramétrisation de la matrice d’erreur d’observation sera étudiée dans la partie 7.5. Ici nous conserverons des erreurs d’observations minimales (10^{-3}). Les expériences sont menées avec 24 EOFs et la zone d’influence 30 (tableau 7.1). Cette zone est choisie car avec un jeu d’observation complet (tout le champ 3D) une zone trop grande pourrait introduire des erreurs dues aux mauvaises corrélations longues distances. De plus, avec un champ complet, il est normalement peu utile d’aller chercher des informations trop éloignées du point à corriger. Chaque variable observée (SSH, température et salinité) est assimilée indépendamment, dans le cas de notre système d’observation partiel et dans le cas où l’on assimile le champ 3D complet. Les résultats (figure 7.13) montrent l’erreur RMS sur le champ 3D (pour la SSH, la température et la vitesse zonale) après l’analyse en fonction de la variable assimilée. Les résultats pour la salinité et la vitesse méridienne sont respectivement comparables aux résultats pour la température et la vitesse zonale.

Une première constatation est que l’efficacité de la correction sur le champ 3D avec notre système d’observation partiel est très relative lorsque l’on assimile les observations

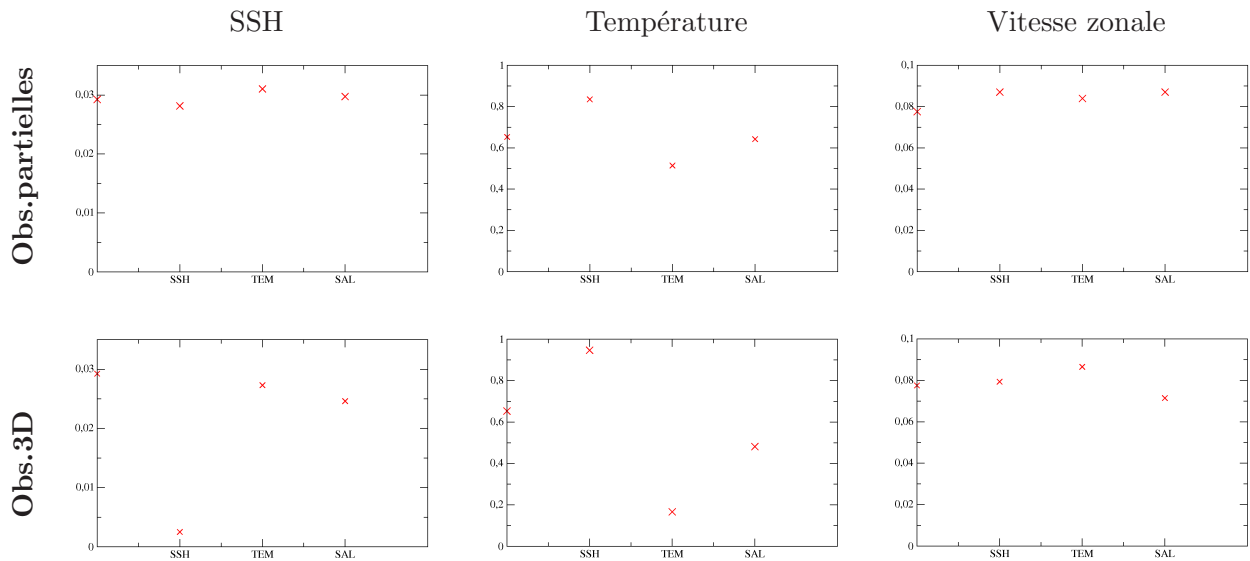


FIGURE 7.13 – Ecart RMS sur les variables de SSH (en m), température (en °C) et vitesse zonale (en $\text{m}\cdot\text{s}^{-1}$), en fonction des différentes variables assimilées de façon indépendantes (SSH, température(TEM) et salinité(SAL)), dans le cas où l'on assimile un réseau d'observations partiel (haut) ou un champ d'observations 3D complet (bas). La valeur de l'écart RMS au point d'abscisse 0 indique l'erreur initiale avant l'analyse.

indépendamment. Par exemple sur la SSH, un système de trace satellites tel que nous l'avons défini semble très insuffisant pour obtenir une bonne correction. De même que pour la correction du champ de température. On peut alors espérer que la mise en place d'un lisseur (qui permet de compléter la couverture d'observation) permettra d'obtenir des corrections plus proches de celles obtenues dans le cas de l'assimilation d'un champ complet 3D. Une deuxième constatation est qu'aucune variable ne semble suffisante à elle seule pour corriger le champ de vitesse. Néanmoins, nous pourrions observer que les vitesses peuvent être en partie corrigées par un ajustement dynamique du modèle aux autres variables assimilées (section 7.5).

7.5 Paramétrisation du filtre SEEK évolutif : propagation de l'erreur

Comme nous l'avons vu dans le chapitre 5, la propagation des statistiques d'erreur d'un filtre évolutif peut s'avérer délicate. Dans le cadre de nos expériences, nous devons trouver les bons paramètres qui permettent d'avoir une propagation de ces statistiques qui soit efficace et cohérente avec l'erreur réelle, c'est-à-dire l'écart entre le run perturbé et le run de référence (représentée par la courbe rouge). Nous effectuons donc des tests sur une période de 50 jours en faisant varier l'un des trois paramètres (nombre d'EOFs, matrice R , zone d'influence de l'analyse) à chaque fois. La durée de cette expérience a été choisie pour rester dans un coût de calcul raisonnable pour des expériences de sensibilité tout en permettant d'avoir un aperçu de la propagation du filtre. Néanmoins, rien ne garantit que le filtre ne puisse pas diverger sur du plus long terme. La date initiale des runs (le 25 Juin

2005, soit le Jour Julien 20250) est identique à celle utilisée pour les tests de sensibilité sur l'efficacité de l'analyse (7.4). Les figures 7.14, 7.15 et 7.16 présentent les évolutions de l'erreur moyenne réelle de l'océan (courbe rouge) et de l'erreur moyenne estimée par la base des covariances d'erreur du filtre (courbe noire pleine). La courbe noire pointillée sur les figures rappelle la trajectoire du run faux sans assimilation. Tous ces résultats sont donnés pour la température, la SSH et la vitesse zonale. Néanmoins les résultats sur les autres variables, salinité et vitesse méridienne, sont respectivement comparables aux résultats pour la température et la vitesse zonale.

A noter ici qu'il s'agit uniquement de comparer des erreurs moyennes pour vérifier que leur ordre de grandeur est cohérent. D'autres critères plus approfondis pourraient être utilisés pour vérifier l'efficacité du filtre, mais pour des raisons de mise en place nous nous contenterons ici d'une étude sur les erreurs moyennes.

7.5.1 Taille de l'analyse locale

Dans la section 7.4.1 nous avons vu que les zones de tailles 30 ou 40 (tableau 7.1) semblent les plus raisonnables pour allier efficacité d'analyse et coût de calcul. Nous effectuons ici un test pour comparer l'impact de ces deux zones sur la propagation des erreurs en utilisant deux cas différents : l'un où nous prenons uniquement 24 EOFs, et l'autre où nous prenons 49 EOFs (tableau 7.2). Les erreurs d'observations sont fixées à 0.15m pour la SSH, 2°C pour la température et 1.5 pour la salinité. Il s'agit d'erreurs d'observations intermédiaires dans la gamme étudiée dans la section 7.5.3. Les résultats sont donnés sur la figure 7.14.

Pour une base d'EOFs donnée, le fait d'augmenter la taille de la zone d'influence de l'analyse a pour conséquence de dégrader la cohérence du filtre (en comparant PeZL-A01 avec PeZL-A02 d'une part, et PeZL-B01 avec PeZL-B02 d'autre part). En effet, le résultat sur l'erreur réelle (courbe rouge) n'est pas meilleur et celle-ci est sous-estimée par le filtre (courbe noire). On peut néanmoins noter que le fait d'augmenter le nombre d'EOFs (en comparant PeZL-A02 avec PeZL-B02) limite cet effet. De toute évidence, lorsque la taille de la zone d'influence de l'analyse locale augmente, le nombre d'observations augmente en conséquence pour une analyse donnée. Il faut donc, pour garder une bonne cohérence, que la base d'erreur \mathbf{S} soit suffisamment importante et permette ainsi une projection efficace des erreurs issues de l'analyse. Autrement dit, pour assimiler plus d'observations, il faut augmenter les degrés de liberté de notre système et donc avoir une base d'EOFs plus riche.

Dénomination	Nombre d'EOFs	Zone d'influence
PeZL-A01	24	30
PeZL-A02	24	40
PeZL-B01	49	30
PeZL-B02	49	40

TABLE 7.2 – Zones d'influences et nombre d'EOFs associés à chaque expérience de sensibilité.

7.5.2 Détermination de la taille de la matrice d'erreur \mathbf{S}

Comme nous l'avons vu dans la section 7.4, la taille de la matrice d'erreur \mathbf{S} (et donc la troncature des statistiques d'erreur) peut avoir un impact significatif sur l'analyse du

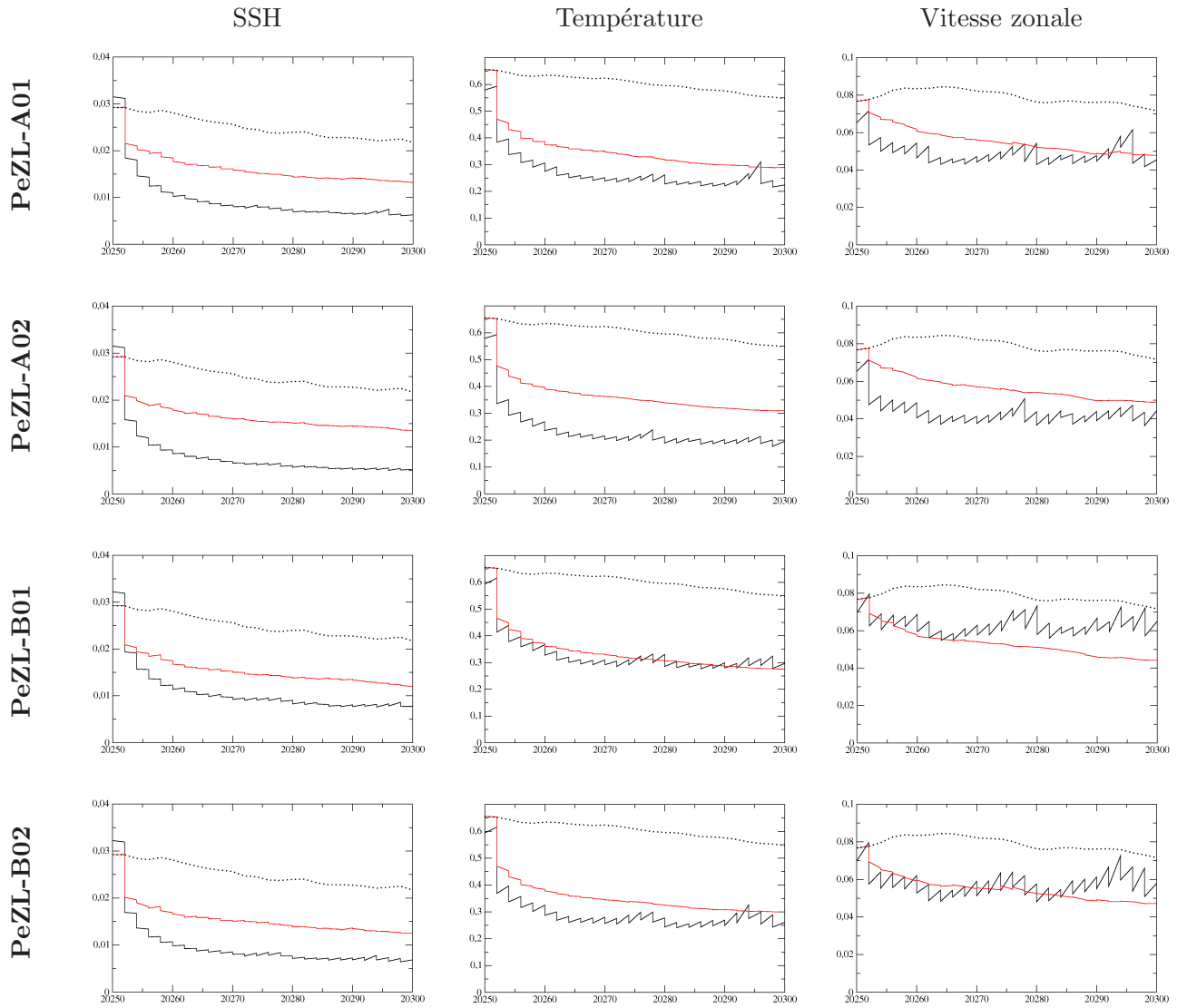


FIGURE 7.14 – Evolution des erreurs RMS réelles (courbe rouge) et estimées par le filtre (courbe noire pleine) pour les expériences PeZL-A01 à PeZL-B02, sur les variables : SSH (en m), température (en $^{\circ}\text{C}$) et vitesse zonale (en m.s^{-1}). La courbe pointillée noire rappelle la trajectoire du run faux sans assimilation.

filtre. Nous allons maintenant tester cet impact sur la propagation des erreurs. Nous fixons donc l'erreur d'observation (0.15m pour la SSH, 2°C pour la température et 1.5 pour la salinité) et nous choisissons la zone d'influence 30 (tableau 7.1). Nous faisons ensuite varier le nombre d'EOFs. Les expériences réalisées sont synthétisées dans le tableau 7.3 et les résultats sont donnés sur la figure 7.15

Les résultats montrent clairement qu'augmenter le nombre d'EOFs permet d'améliorer la cohérence entre l'erreur estimée par le filtre et l'erreur réelle et ce, quelle que soit la variable (assimilée ou non). On note néanmoins que l'erreur réelle (courbe rouge) ne diminue pas de beaucoup. Il s'agit donc surtout d'une amélioration sur la cohérence des

Dénomination	Nombre d'EOFs sélectionnées
PeEOF01	24
PeEOF02	39
PeEOF03	49

TABLE 7.3 – Nombre d'EOFs utilisé pour les expériences de sensibilité à la troncature de la base d'erreur.

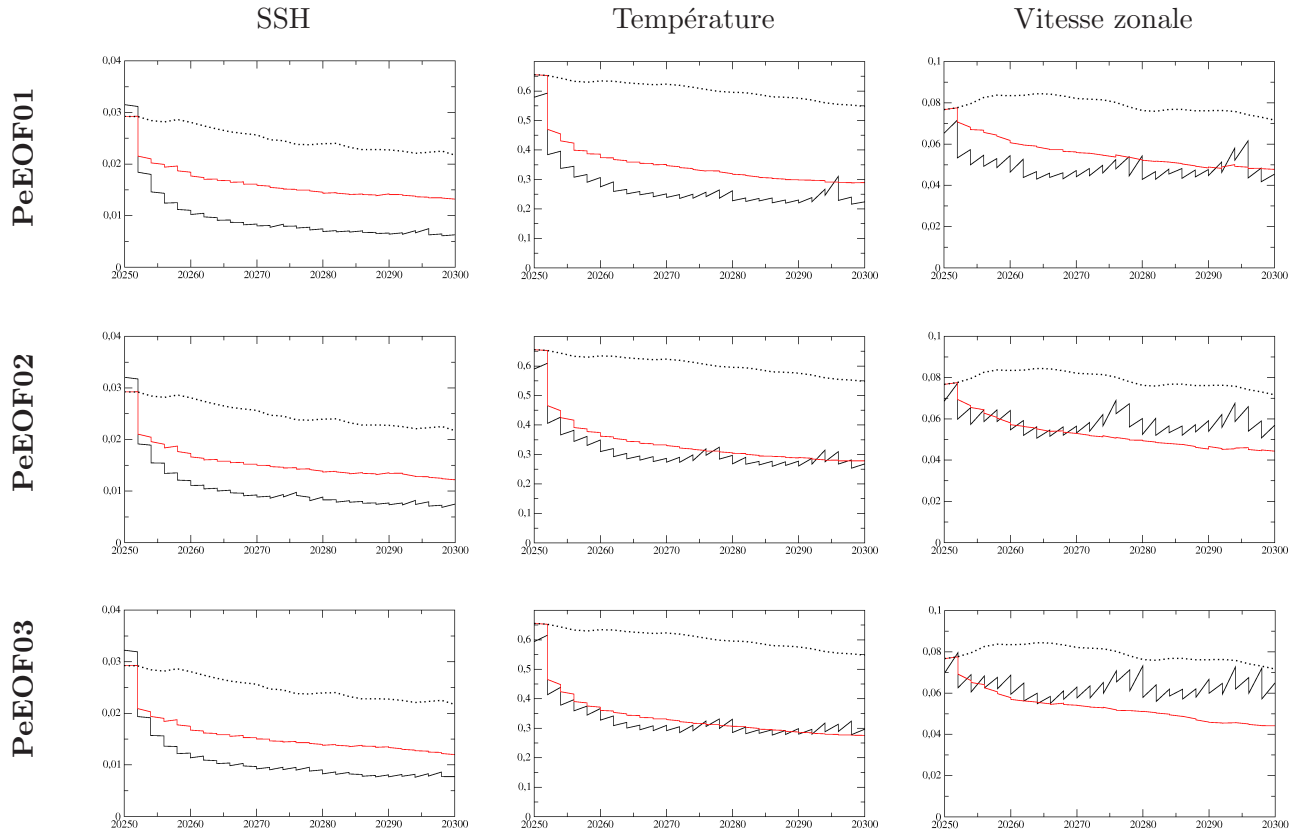


FIGURE 7.15 – Evolution des erreurs RMS réelles (courbe rouge) et estimées par le filtre (courbe noire pleine) pour les expériences PeEOF01 à PeEOF03, sur les variables : SSH (en m), température (en $^{\circ}\text{C}$) et vitesse zonale (en m.s^{-1}). La courbe pointillée noire rappelle la trajectoire du run faux sans assimilation.

statistiques d'estimation, l'efficacité de l'assimilation en elle-même n'est pas réellement influencée (de façon globale) dans cette expérience.

7.5.3 Détermination de la matrice des erreurs d'observation

Bien que nos observations soient parfaites, la matrice d'erreur d'observation ne doit pas être nulle. Cela s'explique par le fait que la matrice d'erreur d'observation R contient également les erreurs liées à la troncature de l'ensemble. Il faut donc prendre en compte dans R toute la part de l'erreur qui n'est pas prise en compte par les EOFs sélectionnées. Dans les expériences qui suivent, nous fixons la taille de l'erreur d'ébauche à 24 EOFs (pour des raisons de coût numérique), et choisissons la zone d'influence 30 (tableau 7.1).

Les différentes valeurs de l'erreur d'observation sont résumées dans le tableau 7.4. Les résultats sont donnés sur la figure(7.16).

Dénomination	Erreur en SSH	Erreur en température	Erreur en salinité
PeOb01	0.03	0.8	0.5
PeOb02	0.08	1.5	1
PeOb03	0.15	2	1.5
PeOb04	0.3	3	3
PeOb05	0.8	3	3

TABLE 7.4 – Paramètres d'erreur d'observation sur les variables assimilées : température (en °C.), salinité et SSH (en m), pour les différents tests de sensibilité de la propagation des statistiques d'erreur.

L'expérience PeOb01 est celle qui se rapproche *a priori* le plus des erreurs associées aux observations dans le cadre d'une expériences réaliste (tableau 7.4). Pourtant, les courbes d'évolution des erreurs montrent clairement que les statistiques sont mal évaluées dans ce cas (figure 7.16). En effet, l'erreur estimée par le filtre (courbe noire pleine) est bien inférieure à l'erreur réelle (courbe rouge). Cela a comme conséquence que l'assimilation n'a presque plus d'effet avec le temps (le trajectoire rouge se rapproche alors d'une trajectoire en run libre). La solution modèle est considérée comme quasi-parfaite, malgré la confiance que l'on accorde aux observations. Lorsque l'on augmente l'erreur d'observation (expériences PeOb02 à PeOb05), cela a deux conséquences : la correction est moins forte sur les premières analyses (et donc la solution corrigée est *a priori* moins bonne que pour l'expérience PeOb01) mais l'erreur estimée par le filtre reste plus cohérente avec l'erreur réelle, et l'analyse du filtre continue d'avoir de l'effet avec le temps. Néanmoins, on constate que le niveau d'erreur réelle final après deux mois d'expérience est sensiblement le même pour la température, la vitesse et la salinité (non montrée), et moins bon pour la SSH dans le cas des expériences PeOb02 à PeOb05. En revanche, les statistiques d'analyse sont beaucoup plus cohérentes, à la fois sur les variables observées mais aussi sur les champs de vitesse.

Pour avoir un filtre stable et cohérent, il faut donc une erreur d'observation plutôt grande. Cela traduit le fait qu'on ne peut pas négliger l'erreur de représentativité, comme cela a déjà été souligné par d'autres travaux, par exemple *Richman et al.* (2005) ou *Oke et Sakov* (2008).

7.6 Conclusion

Nous avons pu le voir dans ce chapitre, les paramètres liés à l'évolution des statistiques d'erreur du filtre SEEK évolutif ne sont pas simples à régler. Globalement, dans le cadre de cette expérience, choisir un nombre d'EOFs de 24 ou 39 semble un bon compromis entre efficacité et coût de calcul. De même, les zones d'influences 30 ou 40 (tableau 7.1) sont suffisamment étendues pour assimiler de façon correcte des observations même éparées. Notons aussi que la taille de la zone d'influence ici est calée en fonction des covariances initiales. Pour un réglage optimal de ce paramètre, il faudrait certainement adapter la zone d'influence à la portée des covariances (ce qui ne sera pas fait dans le cadre des ces travaux). Enfin, concernant l'erreur d'observation, il est nécessaire de la maintenir à

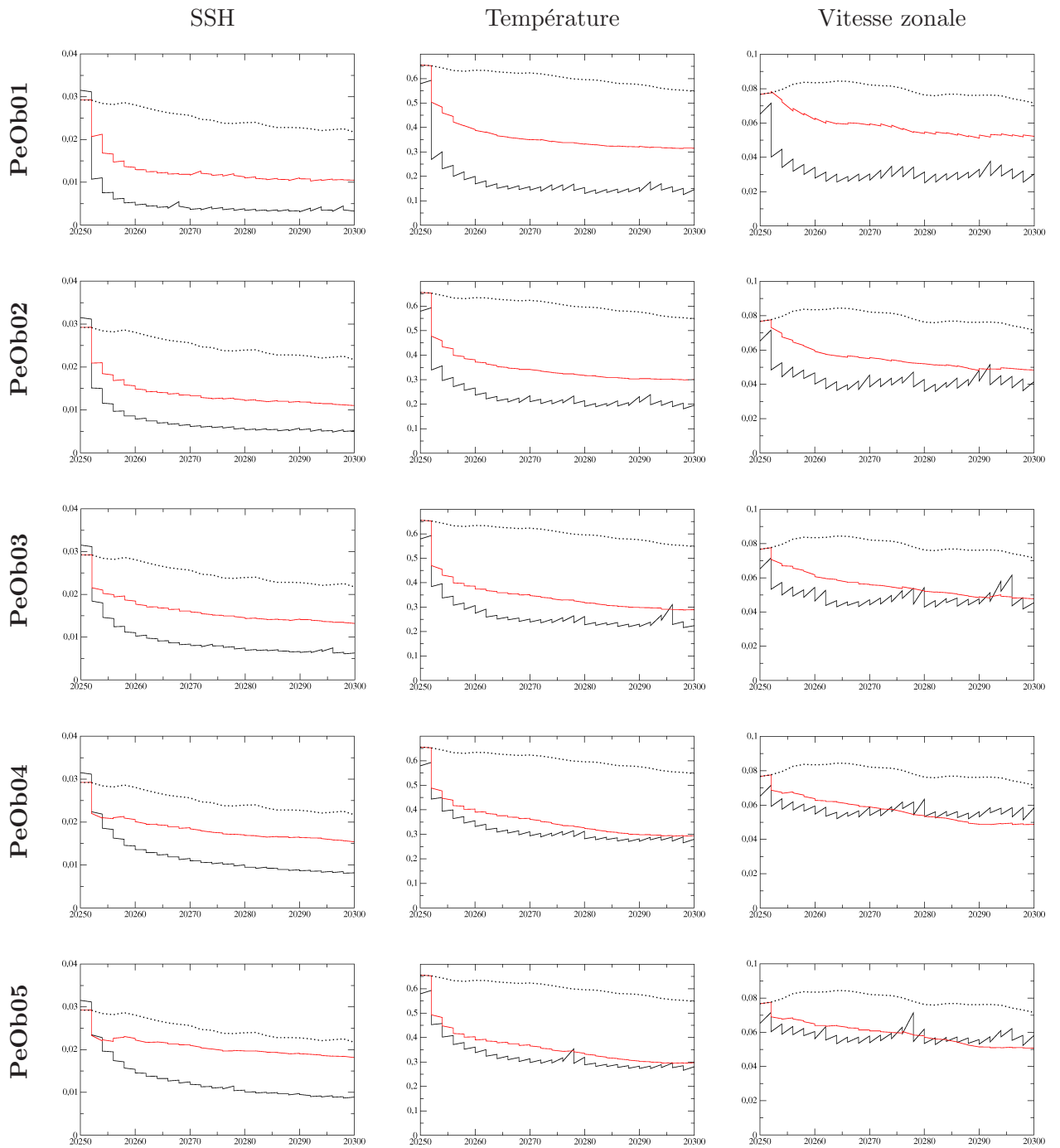


FIGURE 7.16 – Evolution des erreurs RMS réelles (courbe rouge) et estimées par le filtre (courbe noire pleine) pour les expériences PeOb01 à PeOb05, sur les variables : SSH (en m), température (en $^{\circ}\text{C}$) et vitesse zonale (en $\text{m}\cdot\text{s}^{-1}$). La courbe pointillée noire rappelle la trajectoire du run faux sans assimilation.

un niveau suffisamment élevé qui puisse inclure l'erreur de troncature, et donc permettre au filtre d'estimer correctement l'erreur réelle. Les erreurs définies pour les expériences

PeObs03 ou PeObs04 (tableau 7.4) semblent correspondre à ces attentes. Nous avons donc pu choisir trois paramètres qui nous permettront *a priori* d'avoir un filtre évolutif stable et cohérent sur le long terme tout en gardant un rapport *coût numérique / efficacité* raisonnable :

Nombre d'EOFs : 39

Zone d'influence de l'analyse locale : 30 (points de grille)

Erreurs d'observation : 0.3m (SSH), 3°C (température), 3 (salinité)

Avec ces paramètres, le filtre évolutif est stable et permet d'avoir un bon estimé (en moyenne) de l'erreur réelle. L'un des avantages de notre filtre est qu'il ne présente pas de paramétrisation sur la propagation des modes (avec l'ajout d'un facteur d'amplification par exemple). Nous avons donc réussi à rester proche de la théorie tout en travaillant dans un contexte réaliste.

Avec les résultats précédents, il reste quelques questions en suspens qui devront être associées à la mise en place du lisseur :

- D'une part, le niveau d'erreur se stabilise avec le temps à un minimum et ne semble plus converger vers la référence. Une des attentes liée au lisseur est alors de pouvoir améliorer ce minimum afin d'obtenir une solution encore plus proche de la référence.
- D'autre part, le réseau d'observation utilisé ici est plutôt réaliste mais pas forcément optimal. Un réseau plus dense ou disposé différemment pourrait éventuellement améliorer l'efficacité de l'assimilation de données dans cette région. Le lisseur pourrait permettre de compenser le manque d'observations à certains endroits pour obtenir une solution plus lisse dans l'espace.
- Notons également que la cohérence (numérique) de l'état analysé est un point important en assimilation de données. Il faut que l'état analysé reste en accord avec la dynamique du modèle. L'ajout d'observations supplémentaires par le lisseur permet *a priori* d'obtenir une solution plus continue dans l'espace, et donc éventuellement plus cohérente par rapport à la solution filtrée. Nous étudierons cet aspect grâce à un diagnostic de redémarrage du modèle au cours des chapitres suivants.

Nous allons présenter et étudier dans le chapitre 8 la mise en place du lisseur sur la base du filtre évolutif que nous avons défini dans ce chapitre. Les points énumérés ci-dessus seront au coeur des diagnostics effectués dans ce chapitre. Nous compléterons ensuite l'étude du lisseur par un dernier chapitre consacré à un cas plus éloigné de la théorie mais plus proche des systèmes opérationnels actuels.

Quatrième partie

Expériences de lissage dans un
modèle réaliste de circulation
océanique

Chapitre 8

Mise en place et évaluation du lisseur basé sur un filtre évolutif

Sommaire

8.1	Introduction	126
8.2	Contexte de l'expérience	126
8.2.1	La génération de l'erreur	126
	L'erreur initiale	126
	L'erreur après 50 jours de run libre	129
8.2.2	Les réseaux d'observation simulés	129
	Le réseau "temps-réel"	129
	Le réseau "temps-groupé"	129
8.3	Assimilation de données par filtrage et lissage	130
8.3.1	Réanalyses	130
	Corrections globales	130
	Résultats dans la région des TIWs	133
	Résultats dans la région des anneaux du Brésil	137
	Bilan et compléments sur les réanalyses effectuées avec un réseau "temps-réel"	141
	Réanalyses avec un réseau temps-groupé	143
8.3.2	"Cohérence dynamique" des solutions analysées	144
	Cas du réseau temps-groupé	147
	Bilan et compléments sur la cohérence dynamique des états analysés	148
8.4	Conclusion	149

8.1 Introduction

Dans le chapitre précédent nous avons pu déterminer quels paramètres choisir pour obtenir un filtre SEEK évolutif qui soit efficace et dont les statistiques d'erreur restent cohérentes dans le temps avec l'erreur réelle. Cette étape était nécessaire pour la mise en place du lisseur d'après la théorie formulée dans le chapitre 5 car elle a permis de préciser les covariances qui lieront les observations dans la dimension temporelle. Nous allons dans ce chapitre étudier plus en détail l'effet de l'analyse du filtre défini précédemment, et la comparer avec l'effet d'un lisseur basé sur les statistiques de ce filtre. C'est donc dans ce chapitre que la dimension 4D des observations va être introduite.

Nous commençons, dans une première partie, par définir plus en détail le cadre de l'expérience avant d'étudier les résultats obtenus dans les parties suivantes. Nous essaierons notamment de répondre à deux questions : la correction par le lisseur se rapproche-t-elle globalement plus de l'état de référence que celle du filtre (section 8.3.1), et aussi en quoi la solution lissée peut s'avérer plus cohérente avec la dynamique du modèle (section 8.3.2).

8.2 Contexte de l'expérience

Nous gardons le même protocole expérimental défini dans le chapitre précédent (figure 7.2) pour créer une trajectoire fautive et une trajectoire de référence. L'expérience d'assimilation commence donc au jour 20250 avec une première assimilation le jour 20252 (le 25 Juin 2005). Le filtre est utilisé avec les paramètres définis dans le chapitre précédent. Nous regarderons dans un premier temps la structure de l'erreur initiale que nous allons tenter de corriger pour déterminer plus précisément quelles sont les zones ou les structures qui sont le plus perturbées (section 8.2.1).

8.2.1 La génération de l'erreur

Nous allons d'abord regarder comment est distribuée l'erreur initiale, puis dans un second temps nous regarderons comment cette erreur a naturellement évolué en 50 jours (durée de l'expérience d'assimilation). Cela nous permettra ensuite de la comparer avec le run corrigé par assimilation.

L'erreur initiale

Au premier jour de l'assimilation, le 25 juin 2005, l'erreur due à la condition initiale est essentiellement concentrée dans les couches de surface au-dessus de 200 m (figure 8.1, en haut). Il existe néanmoins une erreur sur les vitesses non négligeable dans les couches profondes (de l'ordre de 2 à 4 cm.s^{-1}) alors qu'elle a tendance à tendre vers zéro en-dessous de 1000 m pour les variables thermodynamiques (température et salinité). On peut noter également que l'erreur n'est pas distribuée verticalement de la même façon selon les régions. Elle a tendance à être plus importante près de la surface (sur les 50 premiers mètres) dans la région des TIWs, au centre du bassin, excepté pour la salinité où l'on observe un fort écart dans la région des Anneaux du Brésil. Cet écart est dû à l'embouchure de l'Amazone et aux différences d'apport d'eau douce entre 2003 et 2005. En-dessous de 50m et jusqu'à 200 ou 300 m, l'erreur devient plus marquée dans la région des Anneaux du Brésil, signe que les phénomènes physiques qui ont lieu dans cette région

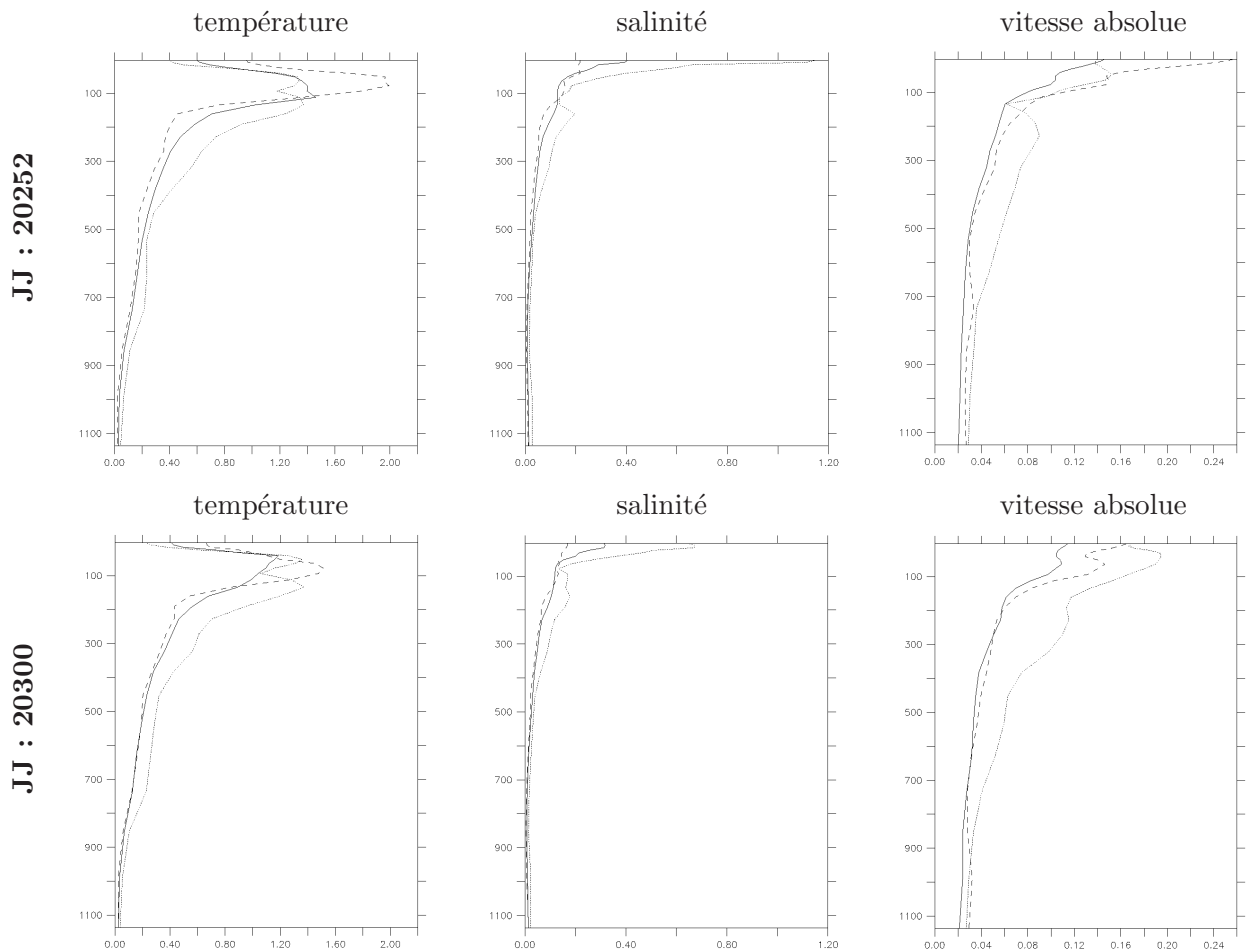


FIGURE 8.1 – Coupes verticales de l'écart RMS entre le run faux et le run de référence, moyenné horizontalement au 1er jour de l'expérience d'assimilation (20252) et après 50 jours (20300), sur tout le bassin (trait plein), dans la région des Anneaux du Brésil (trait en pointillés) et dans la région des TIWs (trait tireté) pour : la température (en $^{\circ}\text{C}$), la salinité, la vitesse absolue (en $\text{m}\cdot\text{s}^{-1}$).

ont une signature plus profonde (voir chapitre 1).

La figure 8.2 montre plus explicitement la distribution de l'erreur moyennée sur les 200 premiers mètres (l'erreur sur la vitesse méridienne n'est pas montrée mais elle est comparable à l'erreur sur la vitesse zonale). Il apparaît clairement une trace d'erreur assez forte au centre du bassin pour les vitesses. De même, une zone d'erreur marquée se trouve le long de la côte nord-est du Brésil, à la fois au niveau de la propagation du courant nord-est brésilien (et donc dans la région des Anneaux du Brésil), mais également au niveau de la réflexion du courant nord-est brésilien (40°W et 5°N). Cela laisse supposer que la dynamique est initialement perturbée sur les zones de courants marqués. Cet aspect est d'ailleurs illustré si l'on regarde les champs de vitesses du run de référence et du run perturbé sur la figure 8.3. Il apparaît alors clairement que les anneaux du Brésil ne se forment pas de la même façon dans les deux runs. De plus, au niveau du bassin central, le run de référence présente des oscillations assez marquées qui peuvent révéler la propagation des premières TIWs. Dans le run perturbé, ces oscillations sont moins développées. On

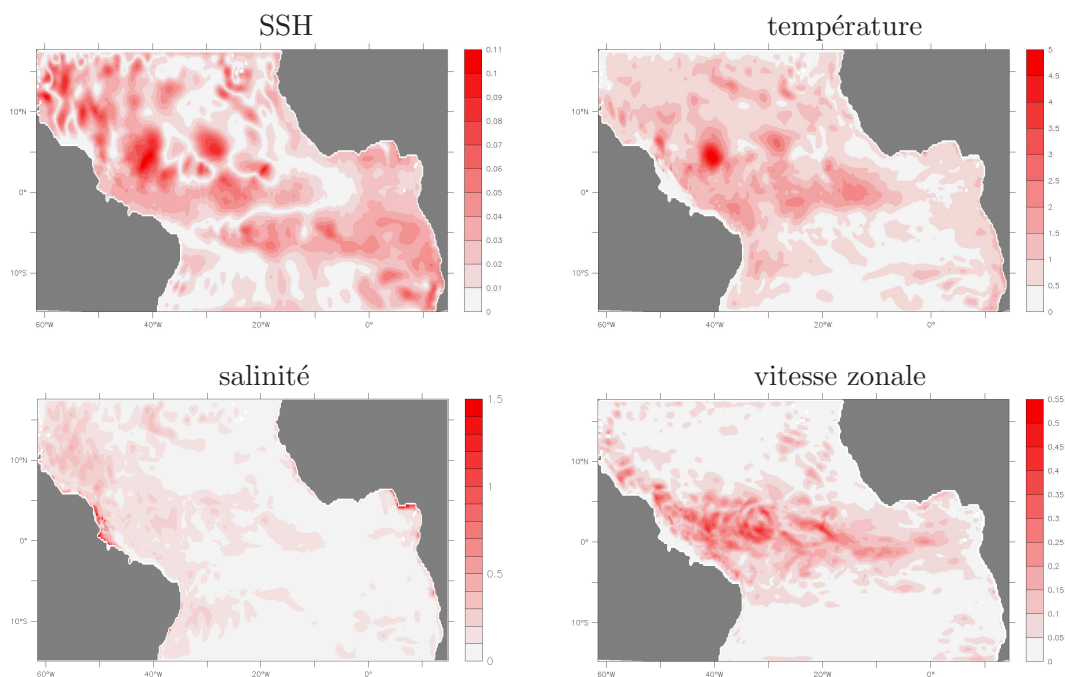


FIGURE 8.2 – Cartes de l'écart RMS entre le run faux et le run de référence, moyenné verticalement sur les 200 premiers mètres au 25 Juin 2005 pour : la SSH (en m), la température (en °C), la salinité et la vitesse zonale (en m.s⁻¹).

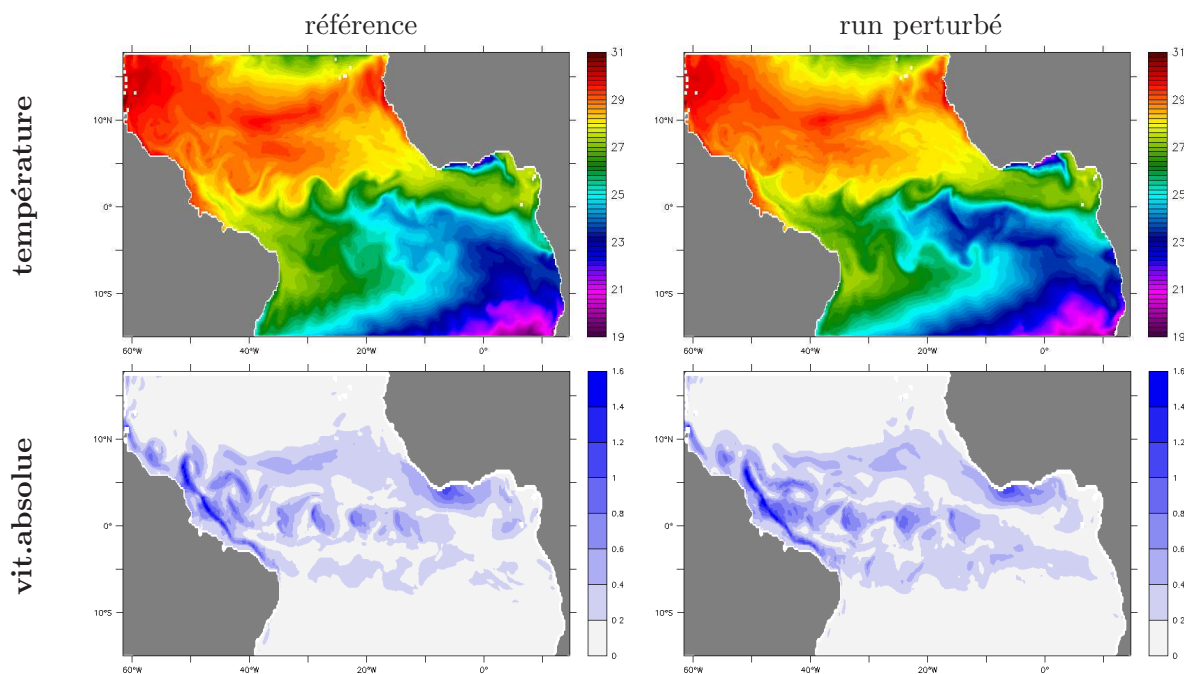


FIGURE 8.3 – Cartes des champs de température en °C (en haut) et de vitesse absolue (en m.s⁻¹) à 30m de profondeur (en bas), pour le run de référence (à gauche) et le run perturbé (à droite) après 50 jours de propagation en run libre.

peut donc supposer que la naissance des premières TIWs a du retard dans le run faux par rapport au run de référence. L'erreur sur la SSH et la température est moins bien définie (figure 8.2), bien qu'elle soit globalement plus forte dans la zone nord-ouest du bassin. Quant à l'erreur sur la salinité, elle est clairement concentrée au niveau de l'embouchure de l'Amazone.

L'erreur après 50 jours de run libre

Comme on peut le voir sur la figure 8.1 (en bas), l'erreur du run faux (propagé librement) sur la salinité et la température diminue de façon significative naturellement dans les couches de surface durant les 50 jours de notre expérience, notamment dans la région des TIWs pour la température et dans la région des anneaux du Brésil pour la salinité. L'erreur sur la vitesse absolue diminue également dans la région des TIWs (qui sont fortement affectées par les forçages), en revanche elle a tendance à augmenter dans la région des anneaux du Brésil (région moins sensible aux forçages). Cela s'explique par la formation d'anneaux qui ne sont pas bien phasés (figure 8.3) entre les deux runs. Les champs de vitesses sont alors décalés et l'erreur RMS augmente. Ces deux formes de "turbulence" (ondes ou anneaux) traduisent des mémoires différentes de l'océan en fonction des régions. En profondeur, l'amplitude de l'erreur sur les champs de vitesses reste constante à cause du courant de bord ouest, comme nous l'avons déjà évoqué dans le chapitre 7. On remarque également sur les champs de température de surface un net décalage entre les TIWs du run de référence et celles du run perturbé.

On peut donc espérer que les analyses du filtre et du lisseur permettront de corriger la part de la dynamique encore perturbée.

8.2.2 Les réseaux d'observation simulés

Afin de corriger le run faux précisé dans la section précédente, deux réseaux d'observations sont définis pour l'assimilation de données : l'un qualifié de "temps-réel" et l'autre qualifié de "temps-groupé". Chacun de ces réseaux observe les mêmes variables (température, salinité et SSH) mais les observations ne sont pas utilisées à la même fréquence par l'assimilation de données.

Le réseau "temps-réel"

Le réseau d'observations simulé est celui défini dans le chapitre précédent (section 7.3). Dans ce réseau, chaque observation est représentative de la date à laquelle elle est extraite. L'assimilation de données a donc accès à des informations à la fréquence réelle d'extraction des observations (tous les deux jours). Ce réseau étant celui utilisé pour paramétrer le filtre, c'est avec lui que nous avons effectué les principales expériences de lissage. Les résultats associés seront donc présentés de façon détaillée.

Le réseau "temps-groupé"

Ce deuxième réseau se rapproche plus des systèmes opérationnels. Il est basé sur le réseau en temps-réel, mais ici les observations sont regroupées par tranches de 10 jours. Ainsi, au lieu de considérer chaque observation à la date où elle est extraite, on considère que toutes les observations extraites sur une fenêtre de 10 jours sont représentatives d'un

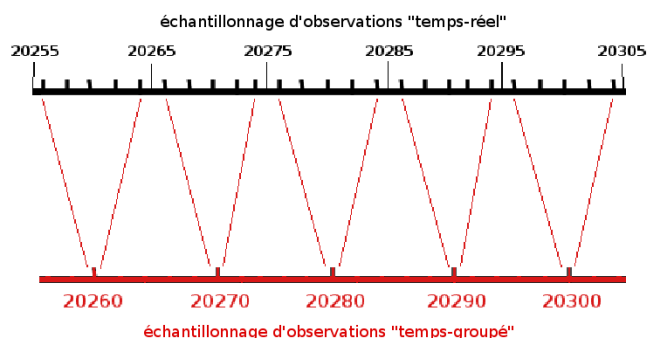


FIGURE 8.4 – Représentation de la répartition temporelle des observations dans le cadre d’un réseau temps-réel en noir et dans le cadre d’un réseau temps-goupé (en rouge).

seul et même état (situé au centre de cette fenêtre). On forme donc un jeu de données décadaire, schématisé sur la figure 8.4. En pratique, cela permet de limiter le nombre de cycles d’assimilation et donc le coût numérique. Cela permet également de disposer de plus d’observations pour une analyse. Les dates d’assimilation seront donc aux jours 20260, 20270, 20280, 20290 et 20300. Il faut néanmoins garder à l’esprit qu’il ne s’agit pas d’observations prélevées tous les dix jours. Chacune des ces dates contient des observations situées dans une fourchette temporelle à plus ou moins 5 jours, l’état corrigé au moment de l’assimilation dépendra donc de l’observation de la dynamique dans toute cette fourchette temporelle. Ce système d’observations a été introduit essentiellement pour effectuer des diagnostics supplémentaires dans un cadre un peu plus proche des systèmes opérationnels. De plus, les cycles d’assimilation étant plus éloignés dans le temps avec ce système, l’impact du lisseur peut s’avérer plus marqué que dans un système haute fréquence. Mais le filtre évolutif n’étant *a priori* pas forcément défini de façon optimale pour ce second réseau, les résultats associés seront présentés uniquement de façon synthétique et complémentaire au premier réseau. Pour une étude plus détaillée, il aurait fallu redéfinir les paramètres associés au filtre comme dans le chapitre 7 avec le nouveau réseau d’observations.

8.3 Assimilation de données par filtrage et lissage

8.3.1 Réanalyses

Nous allons maintenant regarder l’impact de l’assimilation de données, avec un filtre évolutif et le lisseur associé, sur l’erreur caractérisée dans la section 8.2.1. Nous donnerons d’abord un rapide diagnostic global des erreurs (section 8.3.1) avant de focaliser nos diagnostics sur des fenêtres plus spécifiques (tant spatiales que temporelles) dans les sections suivantes. Dans un premier temps le réseau d’observation est celui défini “temps-réel”.

Corrections globales

On vérifie tout d’abord que le filtre mis en place est stable et cohérent sur la durée de notre expérience (la figure 8.5 montre les résultats pour la température et la vitesse zonale). L’erreur vraie diminue de façon constante quelle que soit la variable, et l’erreur estimée par le filtre est du bon ordre de grandeur. Les résultats analysés par le lisseur sont

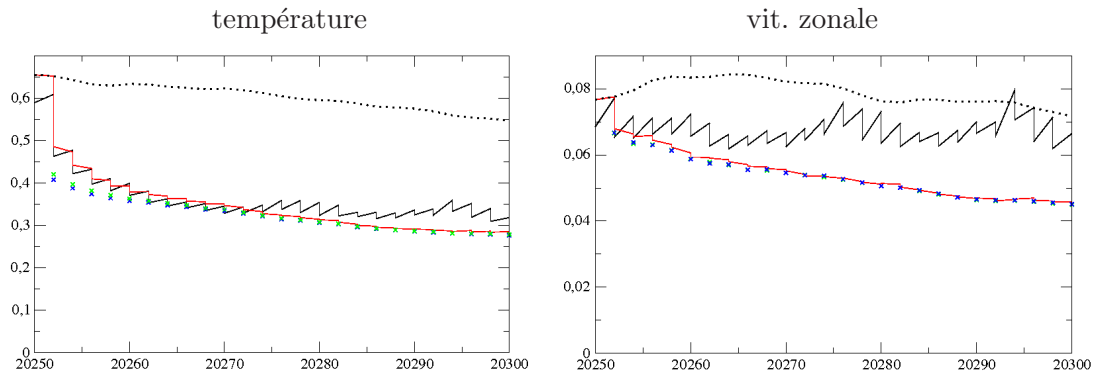


FIGURE 8.5 – Evolution de l'erreur RMS 3D estimée par le filtre (trait plein noir), pour les états analysés par filtre (trait plein rouge) et analysés par le lisseur avec 3 analyses retrospectives (croix vertes) et 5 analyses retrospectives (croix bleues), pour la température (en $^{\circ}\text{C}$) et la vitesse zonale (en m.s^{-1}). L'erreur RMS du run faux est rappelée en trait pointillé noir.

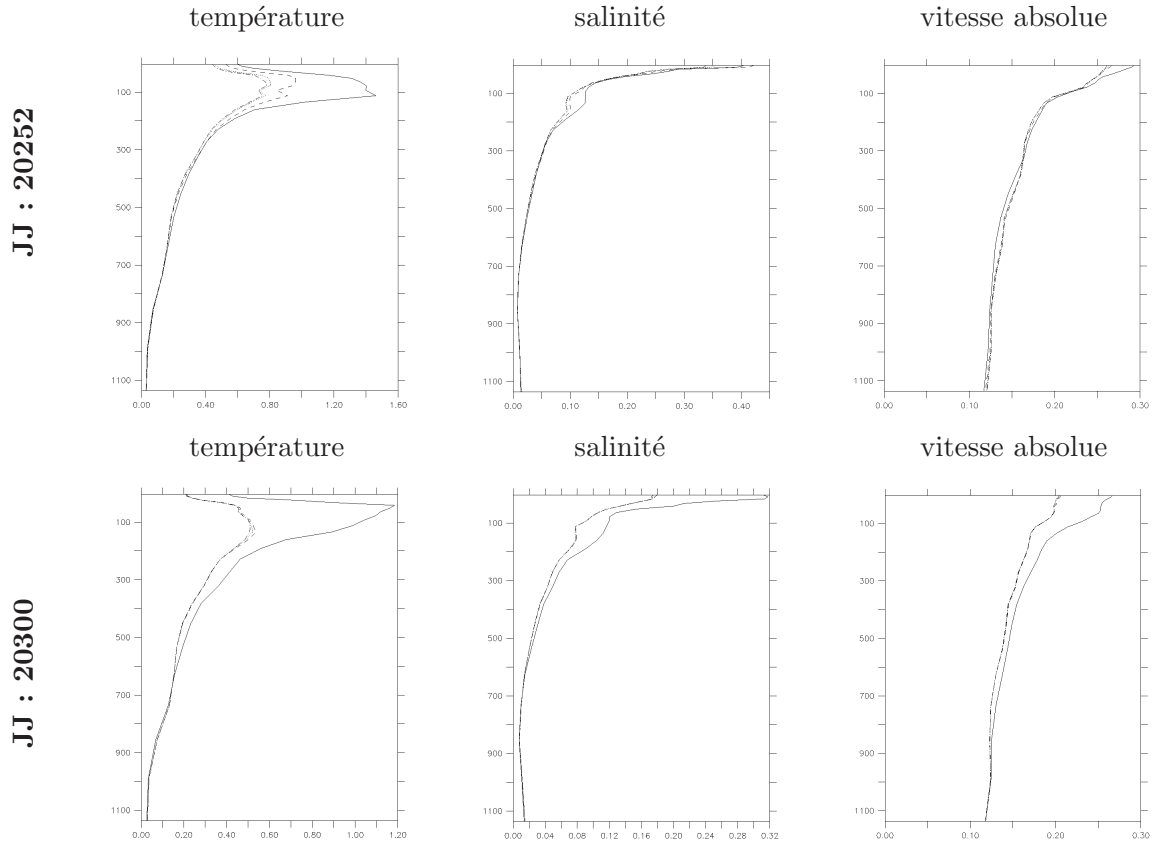


FIGURE 8.6 – Profils verticaux de l'écart RMS par rapport au run de référence moyenné horizontalement sur tout le bassin au 1er jour de l'expérience d'assimilation (20252) et au dernier jour (20300) pour : le run faux (trait plein), l'état analysé par le filtre (trait en pointillés), l'état analysé par le lisseur avec 3 analyses rétrospectives (trait en tiretés), l'état analysé par le lisseur avec 5 analyses rétrospectives (trait en point-tiretés). Les variables représentées sont : la température (en $^{\circ}\text{C}$), la salinité, la vitesse absolue (en m.s^{-1}).

représentés par les croix sur cette même figure. Au début de l'expérience, l'amélioration sur l'erreur RMS est significative. En effet, sur les variables assimilées le gain du lisseur par rapport au filtre est de l'ordre de 15 à 20% pour la température et la salinité, et de 5% pour la SSH (non montré). Ce gain devient moins important avec le temps, jusqu'à être quasiment nul vers la fin de l'expérience. On peut aussi noter que le lisseur produit des scores globalement comparables avec 3 ou 5 analyses rétrospectives (6 ou 10 jours respectivement), avec tout de même une légère amélioration avec 5 analyses. On voit ainsi la limite d'influence temporelle des observations. Au-delà d'une semaine leur impact sur l'assimilation devient plus faible. Par la suite, nous regarderons toujours les résultats du lissage sur 5 assimilations rétrospectives. Néanmoins, il faut garder à l'esprit que l'écart RMS 3D est une mesure très globale et ne permet pas à elle-seule d'affirmer qu'une analyse est équivalente à une autre même lorsque l'erreur RMS est proche. On note aussi que l'effet du lissage est moins marqué sur les vitesses, mais il ne dégrade pas le système.

Nous avons vu précédemment que l'erreur est essentiellement concentrée vers la surface, et comme le montre la figure 8.6 la correction la plus importante est également située dans ces couches (et elle est quasiment nulle en profondeur). A la fin de l'expérience, le run corrigé par le filtre est nettement amélioré sur les couches superficielles. On remarque de nouveau que si l'effet de lisseur est visible au début de l'expérience (notamment sur la température), il est quasiment nul à la fin.

Ces résultats montrent que, de façon globale, sur l'erreur 3D, le lisseur permet d'accélérer la convergence vers l'état de référence et d'atteindre plus rapidement le niveau d'erreur minimal permis par les paramètres du filtre. Lorsque ce niveau est déjà atteint par le filtre (à la fin de l'expérience), le lisseur n'a apparemment plus d'effet (ou très peu). Cela n'est pas surprenant puisqu'il s'agit d'un problème où l'erreur est située sur la condition initiale et tend à converger naturellement vers l'état de référence. Il s'agit donc de corriger au plus vite l'erreur initiale pour retrouver si possible la trajectoire de référence. Compte-tenu de notre réseau d'observation, l'effet du filtre va avoir tendance à accélérer la convergence du système jusqu'à avoir exploité au maximum les informations tirées des observations (en fonction notamment de la définition de la matrice R et de la base d'EOFs initiale). Le lisseur va permettre quant à lui d'accélérer cette convergence. Notamment, il permet d'atteindre un niveau d'erreur au jour 20252 (première assimilation) que le filtre est incapable d'atteindre. **Le lisseur peut donc extraire de l'information des observations plus rapidement et donc corriger la condition initiale plus efficacement grâce aux analyses rétrospectives.** Lorsque l'état analysé par le filtre est à son niveau minimal d'erreur, le lisseur ne peut pas l'améliorer d'avantage.

Néanmoins cette conclusion reste très globale, et il nous faut à présent regarder plus en détail l'effet du lisseur sur la correction des structures dynamiques dans les zones où l'erreur est la plus présente. Nous allons donc dans les parties suivantes regarder son effet dans la région des TIWs dans un premier temps, puis dans la région des anneaux du Brésil dans un second temps. Nous nous concentrerons surtout sur l'erreur initiale et sur les premiers cycles (puisque nous avons vu qu'une fois la convergence atteinte l'assimilation n'a plus d'impact significatif).

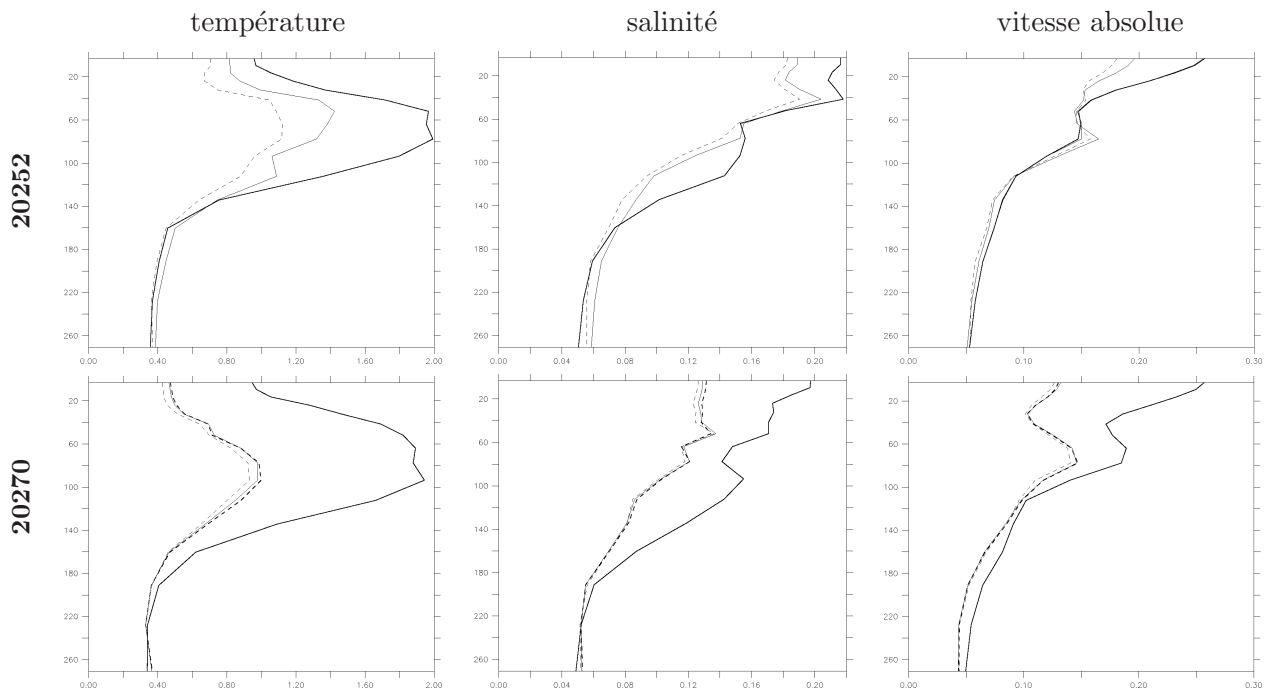


FIGURE 8.7 – Profils verticaux de l'écart RMS par rapport au run de référence, moyenné horizontalement sur la région des TIWs (entre 5°S et 5°N) au jour 20252 (en haut) et jour 20270 (en bas) pour : le run faux (trait plein gras), l'état prédit dans le run assimilé (trait gras en tiretés), l'état analysé par le filtre (trait fin) et l'état analysé par le lisseur avec 5 analyses rétrospectives (trait fin en tiretés). Les variables représentées sont la température (en $^{\circ}\text{C}$), la salinité et la vitesse absolue (en $\text{m}\cdot\text{s}^{-1}$).

Résultats dans la région des TIWs

La figure 8.7 montre les écarts RMS calculés uniquement sur le centre du bassin, entre 5°S et 5°N . On retrouve le même comportement que sur l'écart RMS global, à savoir que les analyses du filtre et du lisseur ont un impact essentiellement sur les 200 premiers mètres et dans la première moitié de l'expérience. Ce résultat s'illustre plus précisément sur la figure 8.8. Les erreurs en température pour la prévision (1ère ligne), l'analyse du filtre (2ème ligne) et l'analyse du lisseur (3ème ligne) y sont montrées pour le 1er jour de l'expérience (20252) et au jour 20270. Les isocontours 28°C sont indiqués pour l'état prédit par le modèle avant l'assimilation (pointillés) et la référence (trait plein), permettant d'illustrer le déphasage éventuel entre les TIWs. La diminution de l'erreur est nettement visible au premier jour, notamment entre l'état prédit et l'état analysé par le filtre, même si certaines zones ne sont pas corrigées malgré une erreur initiale forte (18°W , 3°N). L'impact des analyses est montré sur les 2 dernières lignes (les valeurs négatives indiquant une diminution de l'erreur grâce à l'analyse). Si le filtre est déjà très efficace dans certaines zones ($15\text{-}30^{\circ}\text{W}$, $2^{\circ}\text{S}\text{-}2^{\circ}\text{N}$), le lisseur améliore la correction dans ces zones et permet également d'atteindre des zones qui n'étaient pas ou peu corrigées. On remarque par exemple que la signature d'une TIWs à 38°W qui ne se trouve pas dans le run de référence (ce qui correspond à une advection des eaux chaudes trop au nord) va être réduite par le lisseur. Au jour 20270 on constate que les TIWs ont été efficacement rephasées par le filtre, pour être presque parfaitement en phase à la fin de l'expérience (non montré sur cette figure). Il reste des

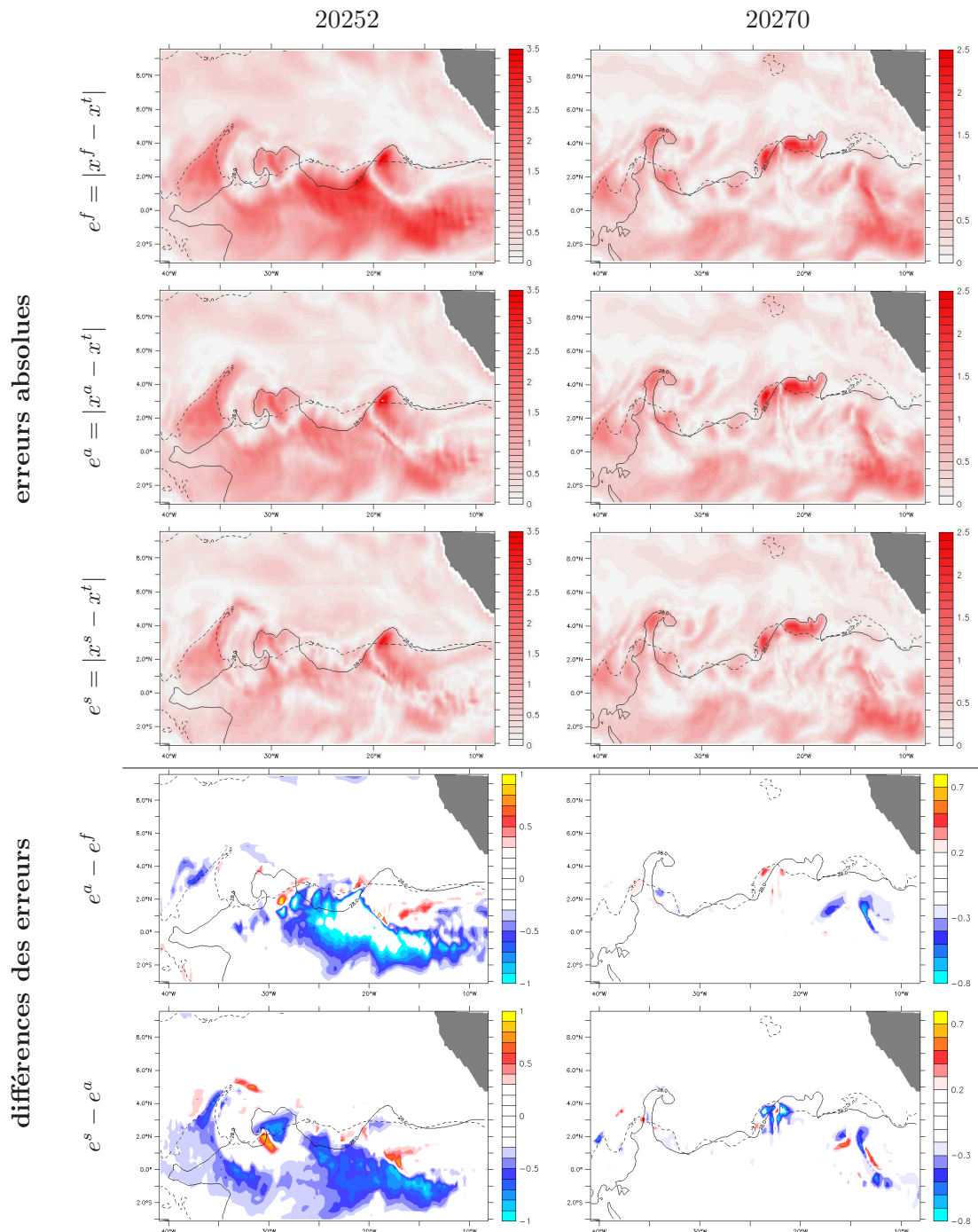


FIGURE 8.8 – Cartes des erreurs absolues en température (en °C) moyennées sur les trente premiers mètres (pour l'état prédit e^f , l'état analysé par le filtre e^a et par le lisseur e^s) et cartes de différences des erreurs absolues entre l'analyse du filtre et le forecast ($e^a - e^f$) et entre l'analyse du lisseur et l'analyse du filtre ($e^s - e^a$). Pour les cartes de différences, les valeurs négatives (respectivement positives) indiquent une amélioration (respectivement une détérioration) de l'analyse du filtre (par rapport à l'état faux) ou du lisseur (par rapport au filtre). Les cartes sont données au jour 20252 (à gauche) et au jour 20270 (à droite). Les courbes noires et pointillées noires indiquent respectivement l'iso-contour 28°C du run de référence et du run perturbé avant l'assimilation.

erreurs résiduelles dues à leurs différences en amplitudes mais qui semblent être hors de portée de l'analyse du filtre. Le lisseur permet d'améliorer à petite échelle ces erreurs (par exemple à 20°W , 3°N) mais on peut également voir très localement de légères dégradations.

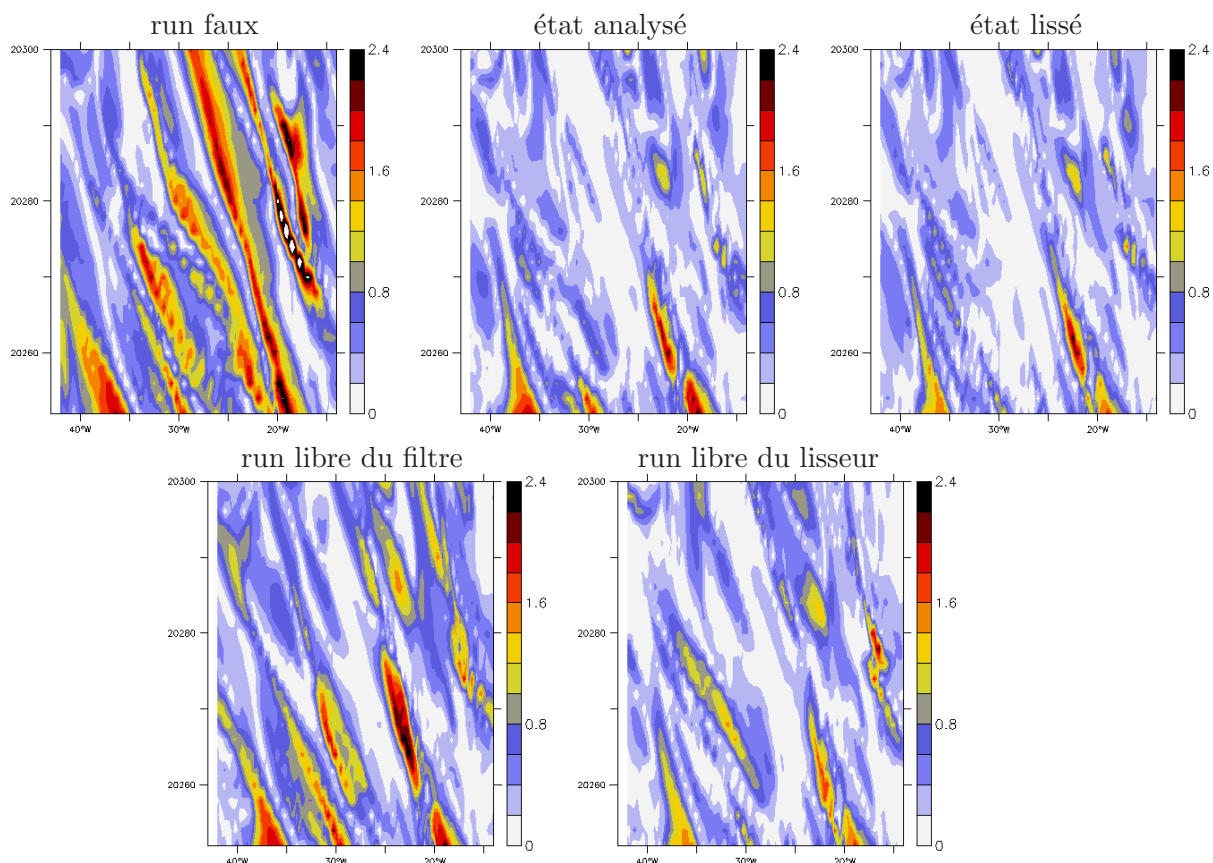


FIGURE 8.9 – Diagramme de Hovmuller montrant l'évolution de l'erreur absolue en température de surface (en $^{\circ}\text{C}$) à 3°N , entre le run de référence et le run faux, les états analysés par le filtre et le lisseur, et les runs libres issus de l'analyse (au jour 20252) du filtre et du lisseur, le long d'un transect entre 42°W et 12°W .

En regardant plus précisément la propagation des erreurs en température de surface le long d'un transect (diagramme de Hovmuller, ligne du haut de la figure 8.9), on confirme que le lissage diminue de façon significative les erreurs sur les premiers cycles (entre les jours 20252 et 20270). Passée cette période, il devient difficile de voir une différence entre les analyses du filtre et du lisseur. On note tout de même que le filtre permet de réduire fortement l'erreur sur toute la période, il est donc logique que l'effet du lisseur ne soit ensuite pas très visible. Néanmoins, le fait de corriger plus efficacement les états dès les début de l'expérience grâce au lisseur est un point important sur lequel nous reviendrons dans la section 8.3.2.

A noter que les résultats montrés sur ces cartes sont plutôt liés à la signature des TIWs en surface. Cette signature se voit clairement sur les erreurs en salinité et en vitesse (figure 8.7). En revanche, comme on le voit sur cette même figure, l'erreur en température est maximale autour de 100m. Cette erreur correspond plutôt à une différence d'épaisseur

de la thermocline qui sera fortement corrigée avec le temps dans le run assimilé.

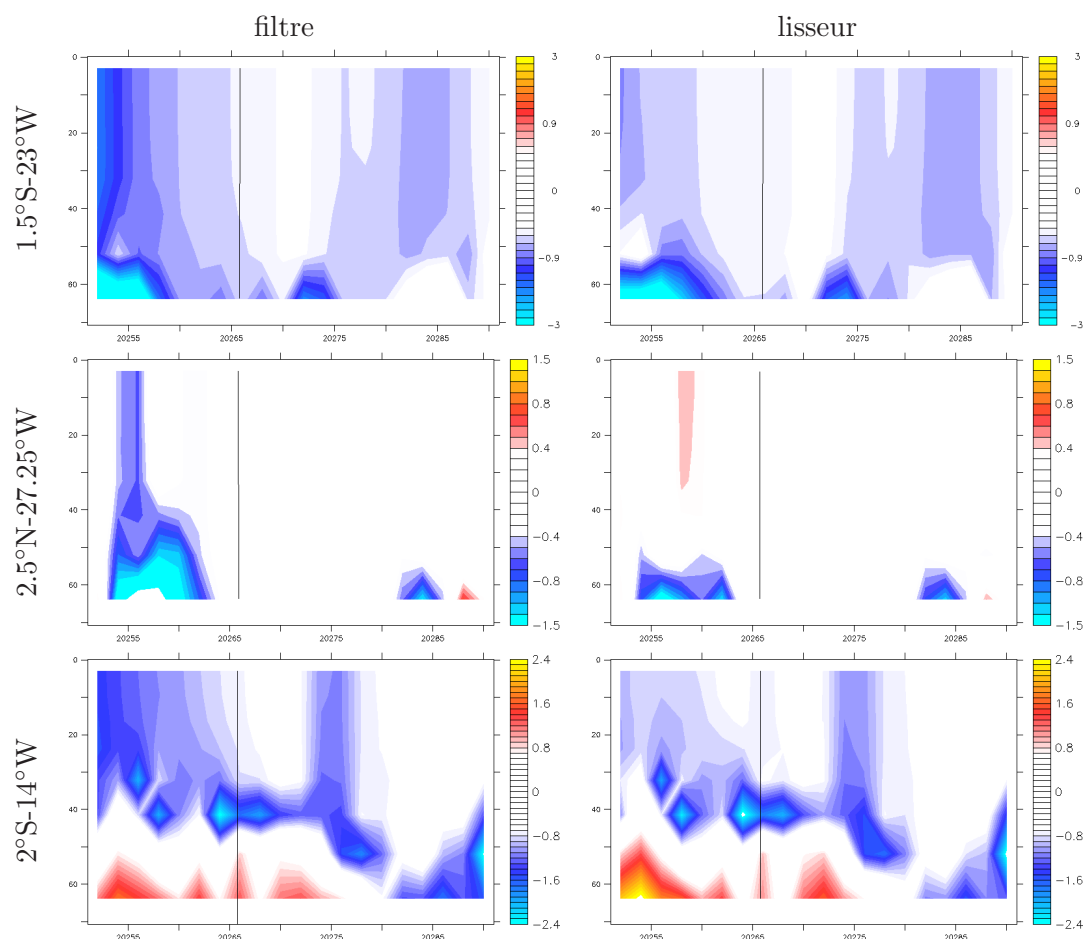


FIGURE 8.10 – Evolution de l'erreur en température (en °C), le long d'un profil vertical et en fonction du temps, pour les états analysés par le filtre (gauche) et par le lisseur (droite). Une observation est située exactement au niveau de chaque profil au jour 20266 (trait noir). Les profils sont situés aux points (de haut en bas) : 1.5°S-23°W, 2.5°N-27.25°W et 2°S-14°W

On peut également regarder plus localement le comportement du lisseur en zoomant autour d'une observation (un profil de température par exemple). Le lisseur possède un effet en subsurface nettement visible, comme le montre de façon plus locale par exemple la figure 8.10. Une observation est située exactement le long de chacun des profils au jour 20266. Le filtre ne voit l'observation qu'au jour où elle est introduite alors que le lisseur peut l'anticiper et améliorer plusieurs jours avant (comme le montrent les différences d'erreur avant la date 20266). En revanche, une fois la date de l'observation passée, les erreurs du filtre et du lisseur sont relativement comparables. Certaines zones restent tout de même hors de portée du lisseur (autour de 60m de profondeur et de la date 20255 pour le profil du haut).

L'effet du lisseur est encore plus visible sur l'erreur en un seul point de ce même profil, situé à la surface (figure 8.11) : l'erreur du lisseur est beaucoup plus faible que le filtre avant la date de l'observation (20266). Une fois la date passée, le filtre et le lisseur ont un

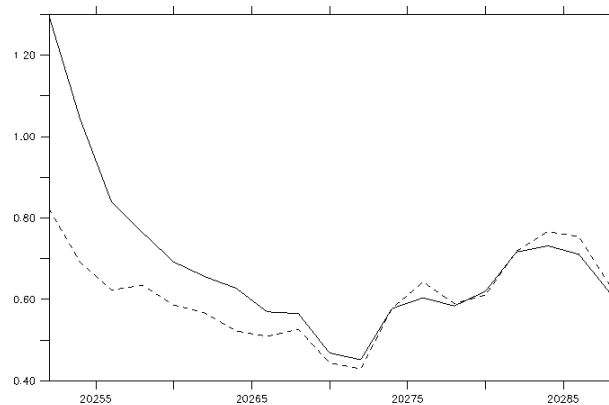


FIGURE 8.11 – Evolution de l’erreur (en valeur absolue) pour la température (en °C) en un point unique en surface et en fonction du temps, pour les états analysés par le filtre (courbe pleine) et par le lisseur (courbe pointillée). Une observation est située exactement au niveau du point de calcul des erreurs au jour 20266.

effet comparable. On note que l’erreur du filtre diminue tout de même avant la date d’observation. Cela s’explique par le fait que d’autres observations alentour peuvent influencer l’erreur au point considéré (par des corrélations spatiales) et également par le fait qu’on se trouve dans un système dynamique et donc l’advection des erreurs peut à elle-seule faire fluctuer leur valeur en un point considéré.

Nous compléterons les remarques sur l’efficacité du lisseur lors de cette expérience dans la conclusion de cette section. Nous allons avant cela regarder les résultats plus spécifiques à la région des anneaux du Brésil.

Résultats dans la région des anneaux du Brésil

La figure 8.12 indique cette fois les écarts RMS calculés uniquement dans la zone des anneaux du Brésil. La signature des anneaux est nettement visible sur les 100 premiers mètres pour la vitesse. L’erreur maximale en température se trouve un peu plus en profondeur. Elle est une nouvelle fois liée à l’épaisseur de la thermocline qui peut être localement très perturbée par la présence d’anneaux et donc entraîner de forts écarts RMS en cas de déphasage d’anneaux entre 2 runs. La signature du lisseur (toujours sur cette figure) est une nouvelle fois visible surtout au début du run (jour 20252) sur la température près de la surface. Cet effet est positif puisqu’il tend à faire diminuer l’erreur RMS. Sur la salinité, l’effet de l’assimilation n’est visible que pour les couches très superficielles (jusqu’à 20m). On remarque, au jour 20252, une dégradation due au filtre en surface qui n’est pas répercutée par le lisseur. Au contraire, ce dernier améliore légèrement la correction. En revanche sur la vitesse l’effet du lisseur est plus mitigé. Au jour 20252, le lisseur induit à la fois une dégradation de l’analyse par rapport au filtre (entre la surface et 100m), mais aussi une amélioration (en-dessous de 100m). La complexité de la dynamique dans cette zone peut en effet rendre l’assimilation 4D plus délicate. Les corrélations multivariées (entre les variables assimilées et les vitesses) en seront d’autant plus complexes, il n’est donc pas surprenant que les corrections sur les variables non-assimilées ne soient pas nettes. Au jour 20270 le lisseur a un impact systématiquement positif, bien que très léger. On remarquera notamment que le filtre n’apporte quasiment plus de correction sur l’état prédit (différence

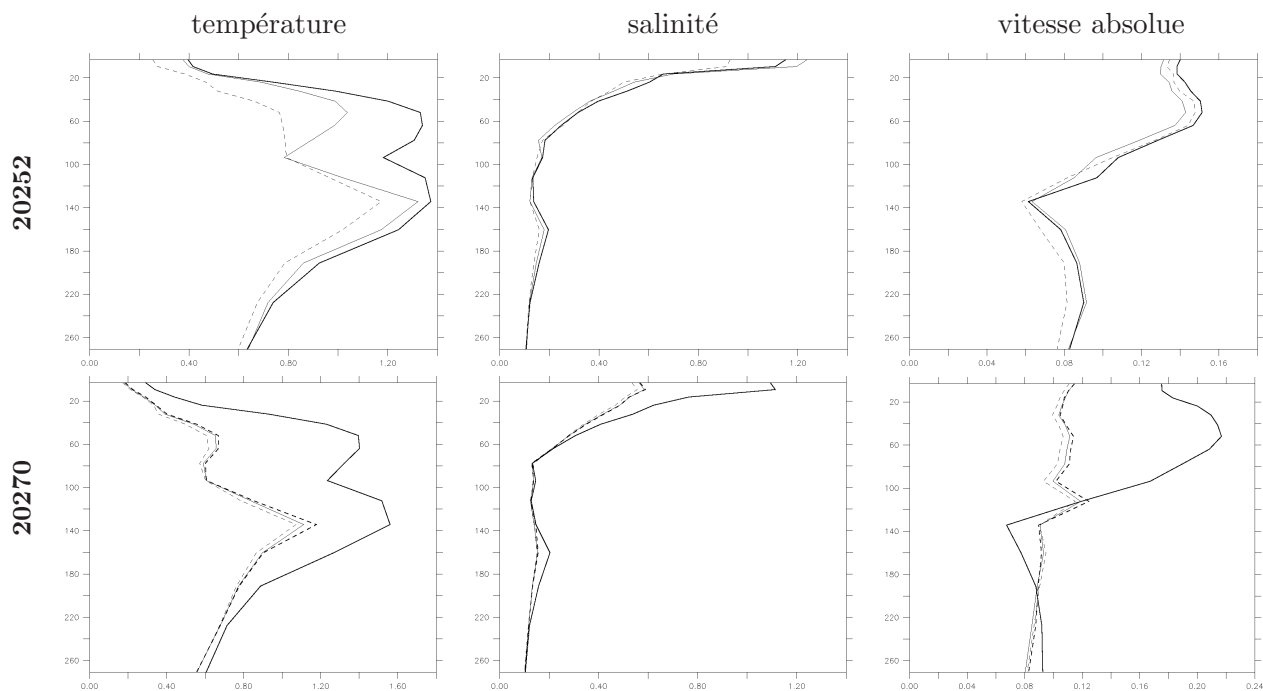


FIGURE 8.12 – Profils verticaux de l'écart RMS par rapport au run de référence, moyenné horizontalement sur la région des anneaux du Brésil au jour 20252 (en haut) et jour 20270 (en bas) pour : le run faux (trait plein gras), l'état prédit dans le run assimilé (trait gras en tiretés), l'état analysé par le filtre (trait fin) et l'état analysé par le lisseur avec 5 analyses rétrospectives (trait fin en tiretés). Les variables représentées sont la température (en $^{\circ}\text{C}$), la salinité et la vitesse absolue (en m.s^{-1}).

entre courbe en gras tiretée et courbe fine pleine) alors que le lisseur, si, notamment entre la surface et 100m.

La figure 8.13 montre, comme pour la région des TIWs, les erreurs absolues et les différences d'erreur entre les analyses du filtre et du lisseur pour les dates 20252 et 20270. Les contours sont ceux de la vitesse absolue $0,5 \text{ m.s}^{-1}$ à 30 m de profondeur, permettant cette fois de positionner les anneaux dans le run de référence (trait plein) et la prévision (trait pointillé). L'effet de l'analyse du filtre est très clair au début, avec des corrections de forte amplitude sur toute la zone. Cet effet tend à devenir plus mitigé avec le temps, induisant à la fois des corrections et des dégradations au jour 20270. On remarque que l'analyse a une influence plutôt en-dehors de la zone des anneaux au début mais qu'ensuite les corrections (bonnes ou mauvaises) se font surtout à l'intérieur des anneaux. On illustre ici l'un des effets du filtre évolutif qui est de diriger l'analyse dans les zones où il reste le plus d'erreur. Le lisseur a également un fort impact au début de l'expérience. On voit par exemple qu'il permet de compléter la correction du filtre dans certaines zones (38°W , 8°N), homogénéisant ainsi la correction sur toute la zone. Avec le temps, l'effet du lisseur continue à être visible mais il tend vers de plus petites échelles, et surtout on observe une alternance de dégradation et d'améliorations. Mais en revenant aux courbes RMS précédentes (figures 8.12), on peut dire que l'impact positif est globalement plus fort que l'impact négatif, contrairement au filtre où les bonnes corrections sont du même ordre

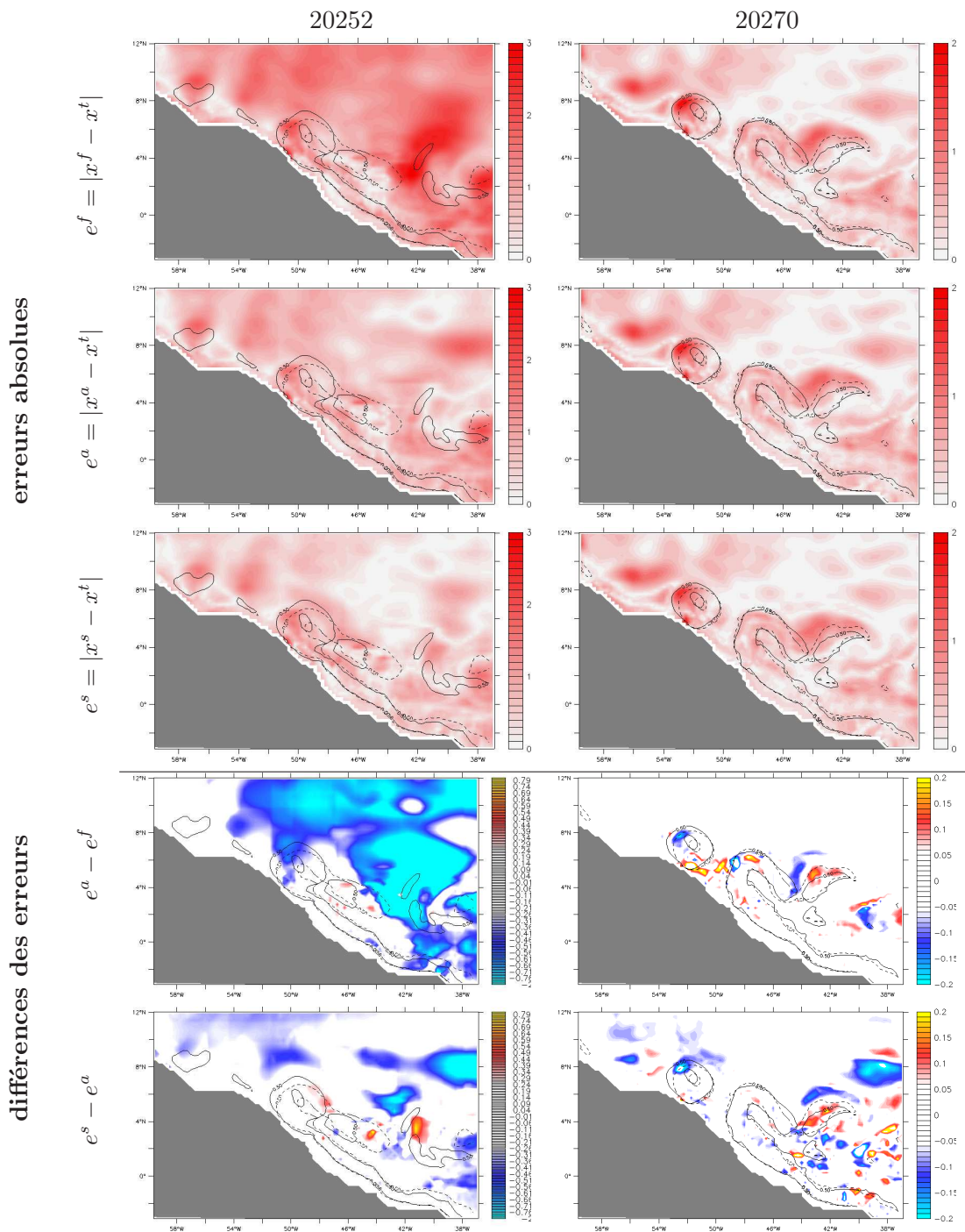


FIGURE 8.13 — Cartes des erreurs absolues en température (en °C) moyennées sur les trente premiers mètres (pour l'état prédit e^f , l'état analysé par le filtre e^a et par le lisseur e^s) et cartes de différences des erreurs absolues entre l'analyse du filtre et le forecast ($e^a - e^f$) et entre l'analyse du lisseur et l'analyse du filtre ($e^s - e^a$). Pour les cartes de différences, les valeurs négatives (respectivement positives) indiquent une amélioration (respectivement une détérioration) de l'analyse du filtre (par rapport à l'état faux) ou du lisseur (par rapport au filtre). Les cartes sont données pour le jour 20252 (à gauche) et au jour 20270 (à droite). Les courbes noires et pointillées noires indiquent respectivement l'iso-contour $0,5\text{m}\cdot\text{s}^{-1}$ du run de référence et du run perturbé avant l'assimilation.

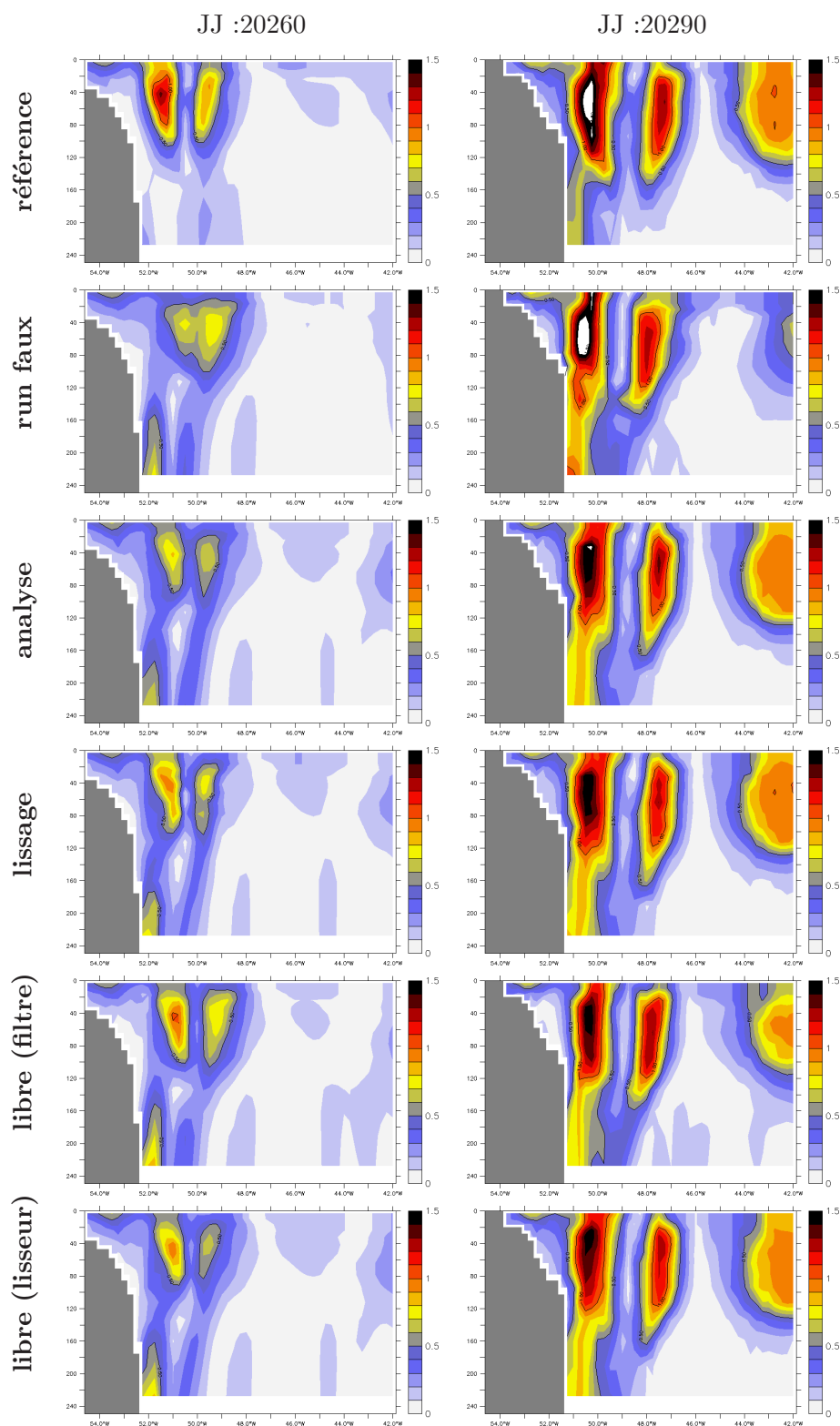


FIGURE 8.14 – Coupes de vitesse absolue (en m.s^{-1}) pour le run de référence, le run faux, le run analysé, le run lissé, et les runs libres issus de l'analyse au jour 20252 du filtre et du lisseur. Les coupes sont tracées à 6.75°N pour le jour 20260 et à 6°N pour le jour 20290.

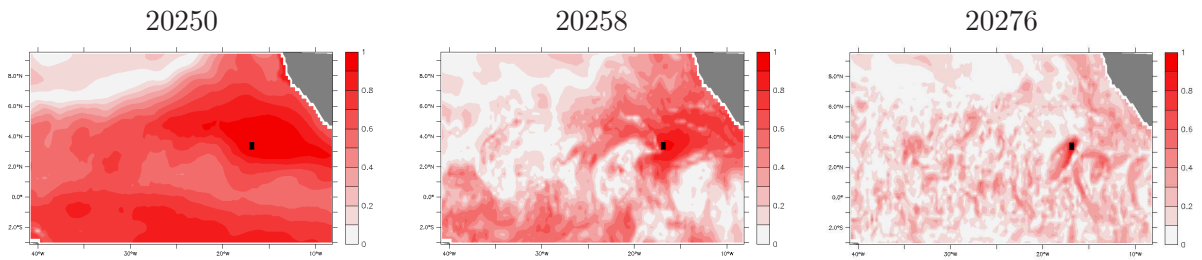


FIGURE 8.15 – Corrélations en température à une observation de température (rectangle noir), calculées avec différentes bases de covariances d’erreur : initiale (20250) et issues du filtre évolutif aux jours 20258 et 20276.

de grandeur que les mauvaises corrections (et donc l’effet sur l’erreur RMS dans la région est quasiment nul).

Si on regarde la structure verticale des anneaux (dont un exemple est donné sur la figure 8.14), on constate que les analyses du filtre et du lisseur au jour 20260 ont tendance à beaucoup plus se rapprocher de la référence que du run faux. Le lisseur permet notamment d’améliorer légèrement l’ordre de grandeur des vitesses dans la branche Ouest de l’anneau (qui est trop faible dans le run faux). Au jour 20290 en revanche l’amélioration par les analyses est moins évidente. On peut noter par contre qu’à l’Est de l’anneau, une structure de vitesse assez importante est visible dans le run de référence et non dans le run perturbé. Cette même structure se retrouve dans les runs analysés, le filtre a donc permis de corriger efficacement la dynamique de cette région. Le lisseur n’apporte pas de modification supplémentaire notable par rapport au filtre pour cette date-là.

La dynamique chaotique de la région semble donc être une limite à l’efficacité du filtre, et un peu moins du lisseur. Il faudrait sans doute des corrélations spatiales définies beaucoup plus précisément pour arriver à un effet plus tranché dans toute la région. Mais la taille de notre système (39 EOFs) limite forcément sa capacité à correctement représenter les bonnes statistiques d’erreur. Nous verrons par la suite (section 8.3.2) que les petites différences observées entre l’efficacité du filtre et du lisseur peuvent avoir un impact significatif en terme de solution modèle. Mais nous allons avant résumer quelques points sur le lisseur que nous venons de voir.

Bilan et compléments sur les réanalyses effectuées avec un réseau “temps-réel”

Les résultats observés précédemment peuvent être complétés par quelques constatations plus générales. Tout d’abord, comme on peut le voir à la fois sur l’impact du filtre et l’impact du lisseur, l’évolution des covariances d’erreur entraîne des corrections qui tendent vers des petites échelles. Cette évolution est illustrée sur la figure 8.15 où l’on peut voir les corrélations spatiales liées à une observation de température (ponctuelle) en fonction de la base de covariances d’erreur utilisée (initiale et issue du filtre évolutif aux jours 20270 et 20290). Au jour 20250, les corrélations sont de grande échelle. Mais très vite, dès le jour 20258, elles tendent vers des échelles plus fines et possèdent une empreinte de la dynamique bien visible (on peut voir la forme des TIWs au centre du bassin). Donc, avec le temps, même le lisseur deviendra moins apte à pouvoir combler les vides entre les observations

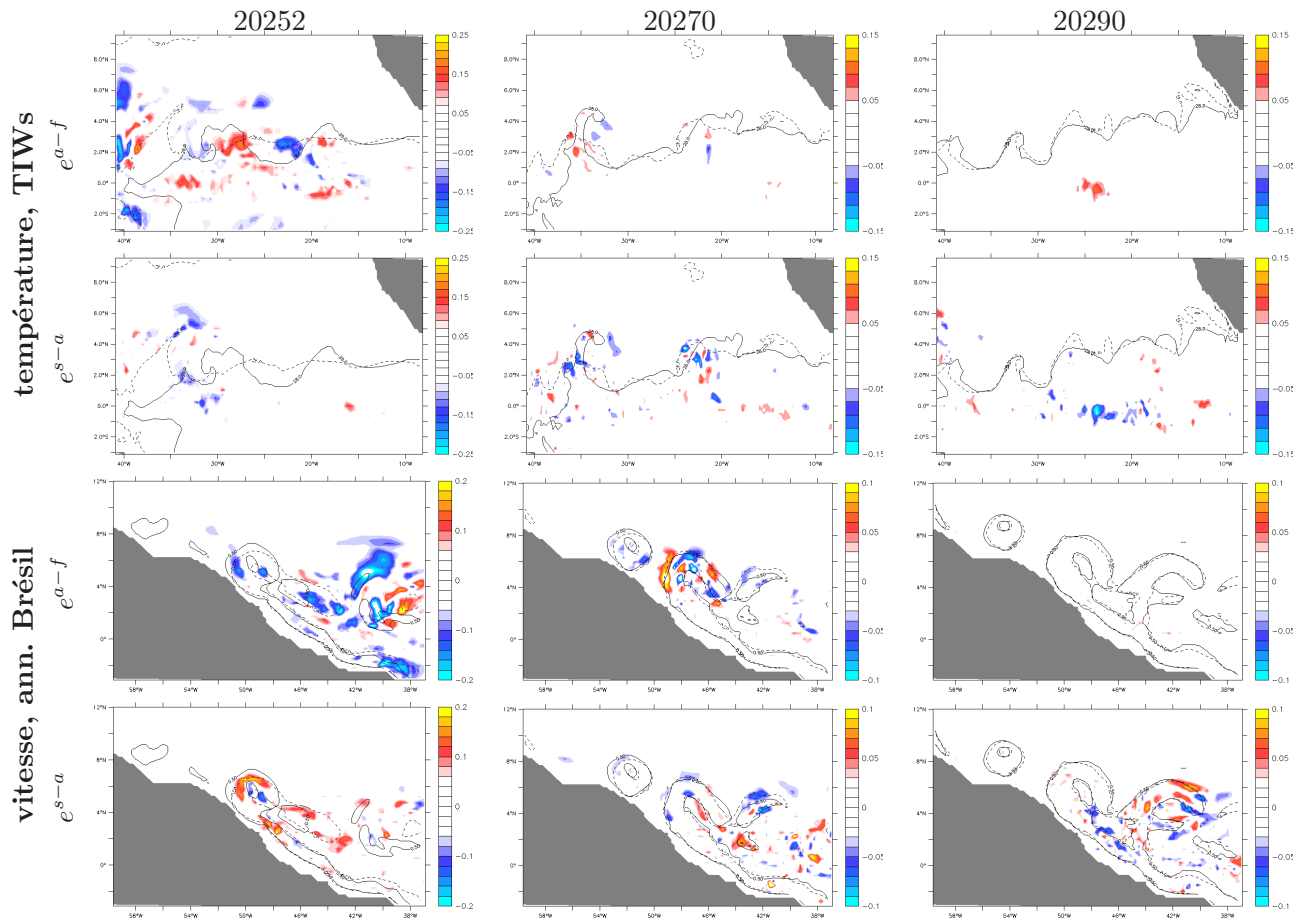


FIGURE 8.16 – Cartes de différences des erreurs absolues entre l’analyse du filtre et le forecast ($e^{a-f} = |x^a - x^t| - |x^f - x^t|$) et entre l’analyse du lisseur et l’analyse du filtre ($e^{s-a} = |x^s - x^t| - |x^a - x^t|$), pour la température moyennée sur les 30 premiers mètres (a-f) et la vitesse absolue moyennée entre 30 et 100m (g-l). Les valeurs négatives indiquent une amélioration de l’analyse du filtre par rapport au forecast ou du lisseur par rapport au filtre. Les cartes sont données pour le jour 20252 (à gauche), 20270 (au milieu) et 20290 (à droite). Toutes les valeurs sont exprimées en $^{\circ}\text{C}$ pour les températures et en m.s^{-1} pour les vitesses.

car les corrélations spatiales seront plus localisées. On peut quand même voir, à travers les cartes d’erreur précédentes, que l’effet du lisseur sur les variables assimilées reste encore visible dans l’expérience avec le temps (contrairement à ce qu’on aurait pu imaginer en regardant uniquement les écarts RMS). Mais il peut aussi introduire des dégradations locales, l’erreur RMS globale de l’état analysé par le lisseur ne sera donc pas si différente de celle de l’état analysé par le filtre.

L’influence de l’assimilation sur les vitesses (variables non assimilées) est moins facile à cerner. Si globalement l’écart RMS montre une diminution de l’erreur (figures 8.7 et 8.12), on remarque que les analyses peuvent aussi dégrader l’état, que ce soient celles du filtre ou celles du lisseur (illustré par la figure 8.16). Ces corrections se font à petite échelle spatiale dès les premières analyses. Mais dans la région des anneaux du Brésil, le lisseur montre plutôt un impact positif par rapport au filtre. L’erreur globale tend à être moins grande grâce au lisseur, malgré l’alternance de corrections bonnes et mauvaises. Il est difficile d’en

tirer des conclusions claires. Les corrélations entre les observations assimilées et les vitesses sont donc *a priori* assez limitées. On peut supposer que si les vitesses sont corrigées avec le temps c'est plutôt grâce à la correction des variables thermodynamiques qui vont à leur tour influencer la dynamique du modèle (en plus de l'ajustement du modèle aux forçages).

Du point de vue des réanalyses donc le lisseur semble apporter des améliorations non négligeables :

- Il permet d'abaisser le niveau global de l'erreur RMS par rapport à la solution filtrée, surtout au début de l'expérience. Cet effet est limité dans le temps, notamment à cause de la configuration de notre expérience (paramètres fixés en fonction de l'erreur initiale). Une fois que le niveau d'erreur minimal moyen permis par la configuration est atteint, le lisseur et le filtre présentent peu de différence.
- Le lissage permet également d'obtenir un niveau d'erreur initial (au jour 20252) impossible à atteindre pour le filtre. Cet aspect est important pour obtenir une solution analysée optimale dès les premières observations accessibles.
- Enfin, la solution lissée présente généralement des structures plus continues que la solution filtrée, grâce à l'homogénéisation spatiale des observations apportées par les analyses rétrospectives.

Ces résultats sont néanmoins issus du cas très particulier d'assimilation en temps-réel des observations. Afin d'élargir l'étude des réanalyses, nous allons utiliser dans la section suivante le cas du réseau temps-groupé (défini dans la section 8.2.2).

Réanalyses avec un réseau temps-groupé

En refaisant les mêmes expériences que dans la partie précédente, mais en utilisant cette fois un réseau "temps-groupé", on cherche à étudier le comportement du lisseur face à un séquençage plus faible du réseau d'observation.

La figure 8.17 est comparable à la figure 8.5 et permet de visualiser de façon globale les niveaux d'erreur RMS sur les variables de température et de vitesse. On note dans un premier temps que les statistiques d'erreur estimées par le filtre (courbe noire) sont proches de l'erreur réelle (courbe rouge) tout au long de l'expérience. Le filtre, bien que paramétré pour un réseau temps-réel, garde ici une bonne cohérence.

L'effet du lissage apparaît clairement dans cette expérience, surtout sur les variables assimilées (illustrées ici par la température). Il permet de corriger l'état faux bien avant l'arrivée de la première observation : dès le jour 20250 le lissage a un niveau d'erreur RMS beaucoup plus faible. Ce résultat est également valable sur les variables non-assimilées (vitesse). Le lissage présente également l'avantage de pouvoir analyser la solution de façon continue (ici, tous les deux jours), contrairement au filtre qui ne peut l'analyser qu'au moment des observations (tous les dix jours). **La solution lissée ne présente ainsi pas les sauts typiques de l'assimilation séquentielle.**

Le lissage montre donc clairement, dans un cadre de réanalyses avec un réseau d'observations peu fréquent dans le temps, des effets bénéfiques pour obtenir une solution analysée plus lisse.

Il reste maintenant à déterminer, quel que soit le réseau d'observations, si ces améliorations sont uniquement statistiques (diminution des erreurs globalement) ou si elles sont aussi liées à une meilleure correction de la dynamique (c'est-à-dire des structures de la dynamique plus proches de l'état vrai). Cet aspect est particulièrement important dans le

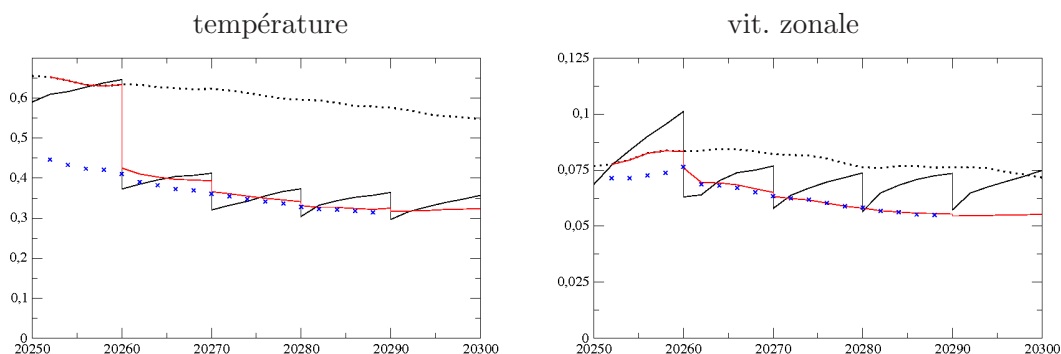


FIGURE 8.17 – Evolution de l’erreur RMS 3D estimée par le filtre (trait plein noir), pour les états analysés par filtre (trait plein rouge) et analysés par le lisseur avec 5 analyses retrospectives (croix bleues), pour la température (en $^{\circ}\text{C}$) et la vitesse zonale (en m.s^{-1}). L’erreur RMS du run faux est rappelée en trait pointillé noir.

cadre d’initialisation de runs ou pour des diagnostics climatiques. Nous allons donc nous y attarder dans la section qui suit.

8.3.2 “Cohérence dynamique” des solutions analysées

Il existe diverses façons de diagnostiquer la qualité d’une analyse (voir par exemple *Desroziers et al., 2005*). Dans notre cas, nous avons choisi un diagnostic très simple qui consiste simplement à redémarrer le modèle en mode libre avec un état issu de l’assimilation. On peut ensuite étudier qualitativement la trajectoire de ce run et déterminer dans quelle mesure l’état analysé était plus ou moins approprié pour redémarrer le modèle (et donc dans quel mesure cet état était cohérent avec la dynamique du modèle).

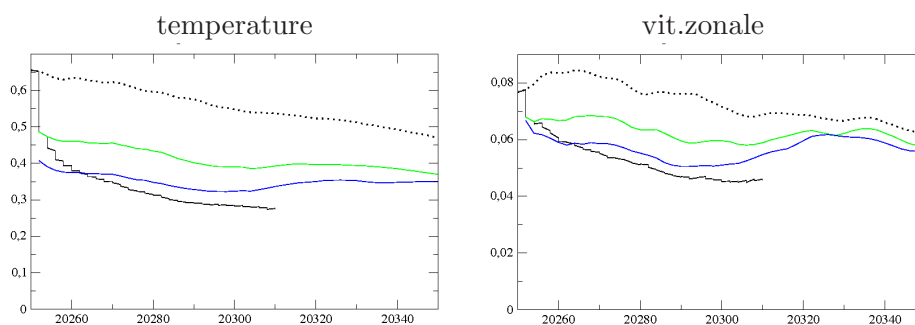


FIGURE 8.18 – Evolution de l’erreur RMS pour le run corrigé par le filtre (trait noir), et dans le cadre de runs libres redémarrant à partir d’un état analysé par le filtre (trait vert) ou d’un état analysé par le lisseur (trait bleu), pour la température (en $^{\circ}\text{C}$) et la vitesse zonale (en m.s^{-1}).

Ce genre de diagnostic peut être important dans une problématique de run de prévision par exemple. En effet, l’un des principaux problèmes de l’assimilation stochastique est d’introduire parfois des chocs ou des perturbations dans le modèle à travers les analyses. Un autre aspect est également lié au positionnement de structures (par exemple un anneau dans la région nord-Brésil). L’analyse peut avoir tendance à faire une sorte de moyenne entre la position d’une structure dans le run faux et sa position dans la référence. Le

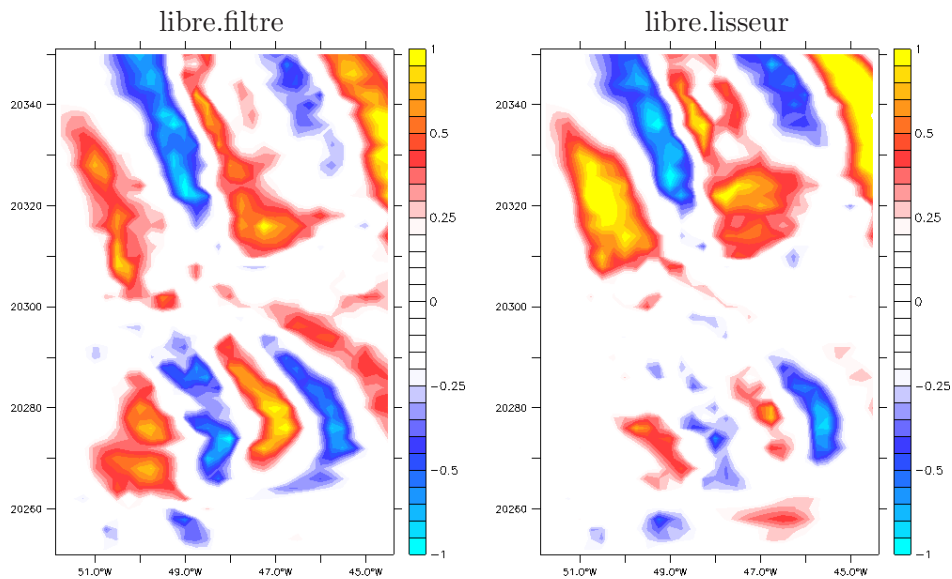


FIGURE 8.19 – Evolution de l’erreur en vitesse absolue (en $\text{m}\cdot\text{s}^{-1}$) entre le run de référence et les runs libres issus de l’analyse du filtre (à gauche) et du lisseur (à droite), le long d’un transect à 5°N (entre 40°W et 50°W), à 30 m de profondeur.

résultat ainsi obtenu peut être déformé et la structure physique s’en trouve dégradée (un anneau pourra par exemple avoir une forme de patatoïde dans la solution analysée). Pour parer à ce problème, deux approches sont envisageables : l’une consiste en une analyse non-linéaire sophistiquée (*Ravela et al.*, 2007), l’autre consiste à introduire un maximum d’observations grâce au lissage pour recouvrir au mieux toute la structure et pouvoir ainsi reconstituer son signal de manière précise. C’est évidemment ce deuxième point qui nous intéresse dans cette thèse.

Dans cette section nous allons donc tester ce que l’on pourrait appeler la “cohérence dynamique” des résultats analysés précédemment. Les redémarrages sont effectués au jour 20252 à partir de l’état analysé par le filtre et à partir de l’état analysé par le lisseur, et les runs associés sont réalisés sans nouvelle assimilation. La figure 8.18 représente les écarts RMS sur les runs issus des analyses au jour 20252 (courbes vertes et bleues). On s’aperçoit que leur trajectoire suit l’évolution naturelle du run faux (courbe pointillée noire), à savoir une tendance à la diminution de l’erreur avec le temps. L’amélioration apportée par le lisseur (courbe bleue) sur l’état initial par rapport au filtre (courbe verte) est conservée tout au long du run libre. Néanmoins, à partir du jour 20310, on commence à voir apparaître une dérive par rapport à la référence et une convergence des deux runs.

Lorsque l’on regarde où se situe l’erreur durant la durée des runs, on voit qu’elle est essentiellement située dans la zone des anneaux du Brésil, plutôt en profondeur. La figure 8.19 montre l’erreur en vitesse à 30m de profondeur le long d’un transect à 5°N . Il apparaît clairement qu’au début des expériences, la simulation issue de l’analyse de lisseur présente une erreur beaucoup plus faible que celle issue de l’analyse du filtre. La trace de ces erreurs indique clairement un déphasage entre 2 anneaux (avec alternance d’erreur positives et négatives fortes). **Le lisseur permet donc ici de rephaser plus efficacement**

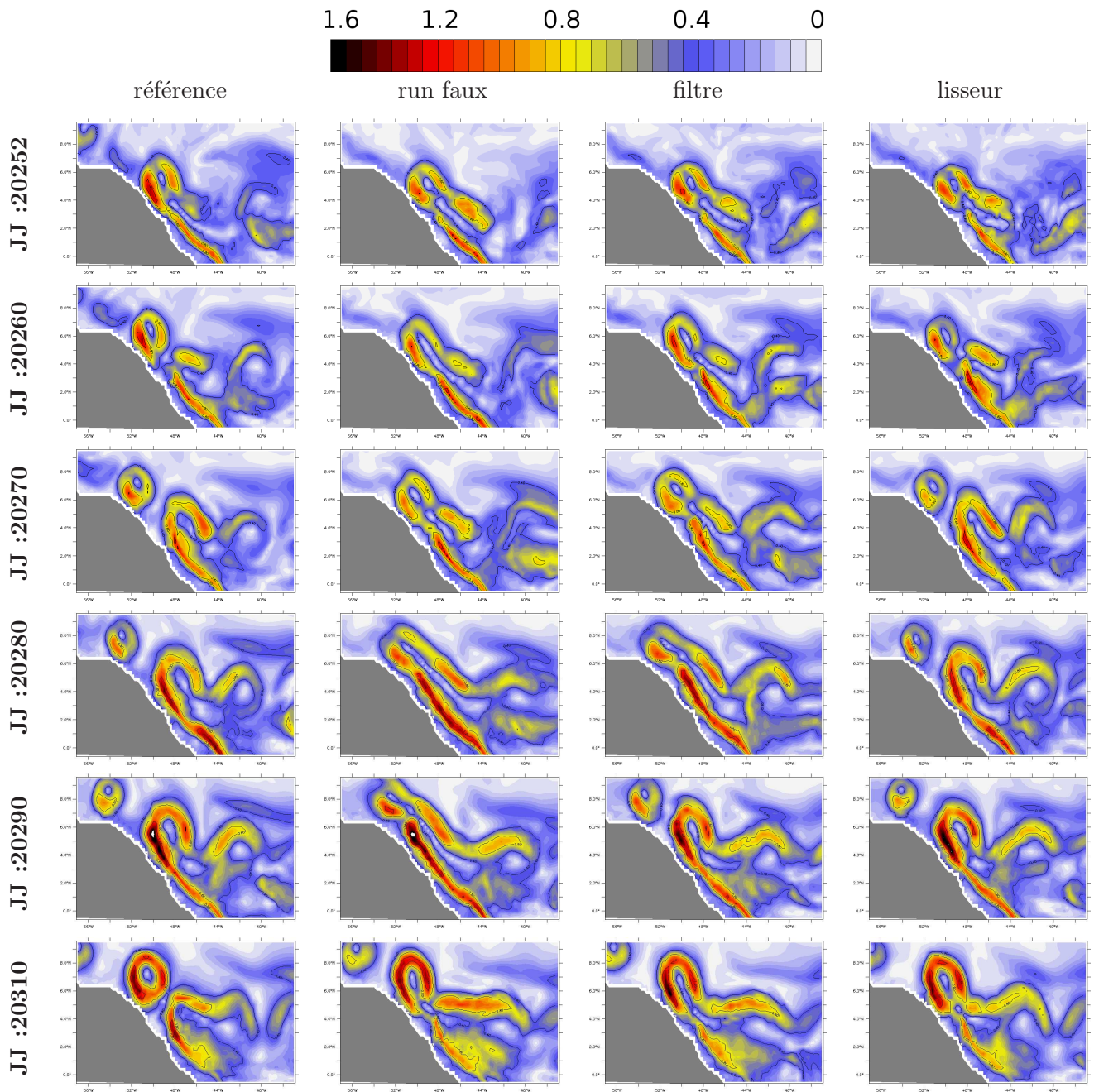


FIGURE 8.20 – Cartes de vitesse absolue (en m.s^{-1}) pour le run de référence, le run faux, le run libre issu de l'analyse du filtre au jour 20252 et le run libre issu de l'analyse du lisseur (5 analyses rétrospectives) au jour 20252. Les figures sont données aux jours 20252, 20260, 20270, 20280, 20290 et 20310.

ces structures. Néanmoins cette différence est limitée dans le temps. A partir du jour 20310, les deux runs présentent de fortes erreurs qui se ressemblent. La dynamique liée au signal faux a donc repris le dessus sur la correction liée à l'assimilation. La figure 8.20, qui montre les structures des vitesses dans chaque cas (référence, run faux, run issu de l'analyse du filtre et run issu de l'analyse du lisseur), permet de compléter ce point. Les

états initiaux des runs issus des analyses sont relativement comparables, mais très vite des différences notables sont visibles. Au jour 20270 par exemple, un anneau est correctement détaché dans le run issu du lisseur alors qu'il n'est pas complètement formé dans le run issu du filtre. Cette différence est encore plus visible au jour 20280. Par la suite, les deux runs issus des analyses présentent de nouveau les mêmes structures (jour 20310). L'effet des analyses semble donc avoir été oublié avec le temps et les états de ces deux runs tendent à se rapprocher du run faux. On retrouve ainsi l'aspect chaotique de cette région qui se rapproche des moyennes latitudes (la prévisibilité est limitée dans le temps). La structure verticale des anneaux (visible sur la figure 8.14, deux dernières lignes) montre moins de différences entre le run issu de l'analyse du filtre et celui issu de l'analyse du lisseur. On remarque tout de même que pour ce second run au jour 20290 la structure de forte vitesse à l'Est de la coupe est bien représentée, alors que dans le cas du run issu du filtre cette même structure semble s'être détériorée et avoir perdu de son intensité.

Dans la région des TIWs, l'effet du lisseur est également visible par rapport au filtre. La figure 8.9 (ligne du bas) montre l'évolution de l'erreur en SST le long d'un transect à 3°N. C'est surtout dans les amplitudes que les différences apparaissent. En effet, les positions initiales des erreurs (au jour 20252) entre les analyses du filtre ou du lisseur sont assez comparables, mais les amplitudes varient fortement. Le lissage permet de diminuer de façon très significative la trace de ces erreurs dès le début du run. Cette amélioration persistera tout au long du run libre. Cela suggère que, compte-tenu de la dynamique plutôt linéaire de la région, si l'état initial est mal corrigé cette erreur pourra se propager jusqu'à avoir traversé tout le bassin. De la même façon, un état *a priori* bien corrigé initialement aura une meilleure cohérence avec la référence durant toute la suite du run libre.

Le lisseur a donc une meilleure tendance à donner des analyses plus en accord avec la dynamique réelle, surtout au début de l'expérience. Cela nous pousse à penser que les réanalyses issues du lissage peuvent dans l'ensemble se montrer plus réalistes et plus intéressantes pour des diagnostics climatiques.

Cas du réseau temps-groupé

Dans le cas du réseau d'observations de type temps-groupé, nous effectuons les mêmes tests mais en repartant cette fois de la date de la première analyse du filtre : le jour 20260. La figure 8.21 (comparable à la figure 8.18) permet de visualiser l'évolution globale des erreurs sur la température et la vitesse zonale. Les différences sont bien moins marquées cette fois. Globalement, les deux trajectoires (bleues et vertes) restent parallèles et conservent leur écart initial sur toute la durée du run libre. On observe tout de même une convergence des deux runs sur la vitesse zonale vers la date 20310, tous les deux s'écartant de la trajectoire de référence. Cela correspond, comme dans le cas du réseau temps-réel, à un mauvais phasage des anneaux. Que ce soit dans le cas du filtre ou du lisseur, les anneaux sont mal représentés après quelques semaines de run libre (figure 8.22). On remarquera aussi que la trajectoire issue du lisseur (courbe bleue) est très proche de la trajectoire du run analysé par le filtre (courbe noire) jusqu'à la date 20300. On obtient donc dans ce cas là une trajectoire corrigée proche d'une analyse du filtre mais qui est aussi une solution modèle continue.

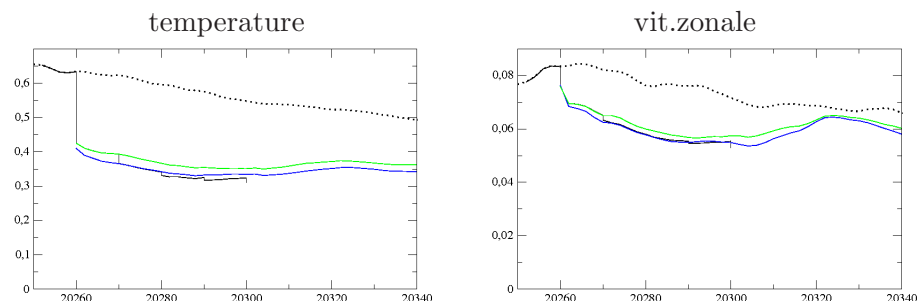


FIGURE 8.21 – Avec un réseau de type temps-goupé : évolution de l'erreur RMS pour le run corrigé par le filtre (trait noir), et dans le cadre de runs libres redémarrant à partir d'un état analysé par le filtre (trait vert) ou d'un état analysé par le lisseur (trait bleu), pour la température (en $^{\circ}\text{C}$) et la vitesse zonale (en m.s^{-1}).

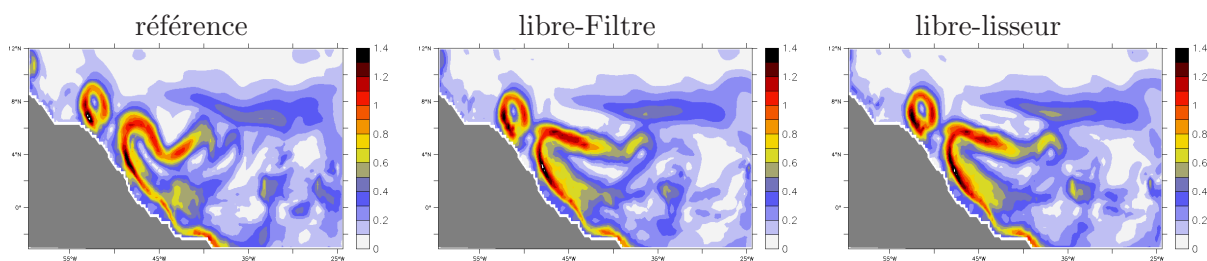


FIGURE 8.22 – Cartes de la vitesse absolue (en m.s^{-1}) à 30m de profondeur dans la région des anneaux du Brésil au jour 20330 pour le run de référence et les runs libres issus des analyses au jour 20260 du filtre et du lisseur.

Avec un réseau de type temps-goupé les différences entre le filtre et le lisseur semblent moins marquées ici (cela est probablement dû au fait qu'au jour 20260 les états filtrés ou lissés sont très proches). Les deux trajectoires ne présentent *a priori* pas de divergence marquée avec la référence sur les premières semaines. Et dans les deux cas, la dynamique chaotique dans la région du Brésil finit par reprendre le dessus vers la fin du run.

Bilan et compléments sur la cohérence dynamique des états analysés

Ces résultats suggèrent que les analyses issues du lisseur ont généralement une meilleure cohérence avec la dynamique. Cela se traduit par une nette amélioration sur les tests d'initialisation des runs libres, notamment dans les régions où la dynamique est très sensible à la condition initiale. Néanmoins, dans certaines régions (anneaux du Brésil), l'apport de l'analyse est limitée dans le temps. La dynamique du système redevient prédominante et peut introduire de nouvelles erreurs malgré une bonne analyse initiale (erreur advectée depuis une autre zone du bassin par exemple).

Enfin, on peut remarquer que dans tous les cas la trajectoire de l'écart RMS du run assimilé en continu (courbe noire) reste meilleure que celle des runs libres. Néanmoins, cette trajectoire présente de nombreux sauts dus à l'assimilation. La propagation d'un état analysé (issu du lissage notamment) peut éviter ce genre de saut et permettre d'obtenir

une trajectoire continue qui soit à la fois une solution modèle et à la fois proche des observations. Cette propagation doit néanmoins se faire sur une durée limitée en fonction de la dynamique de la région (limite de prévisibilité). Cela permet d'éviter de créer des décalages entre des structures très marquées (comme ici pour les anneaux après le jour 20310).

En plus de donner une solution plus continue dans le temps, le lisseur permet d'obtenir d'avantage de réalisme dans les analyses. Cet aspect est particulièrement important pour obtenir des réanalyses de bonne qualité.

8.4 Conclusion

Dans ce chapitre nous avons pu mettre en place un lisseur de Kalman sur la base d'un filtre évolutif (défini dans le chapitre précédent) dans une configuration réaliste de circulation océanique. Nous avons dans un premier temps pu comparer l'efficacité du lisseur et ses apports par rapport au filtre. Il s'est avéré que le lisseur permettait de diminuer significativement l'erreur RMS globale dans les 20 premiers jours de l'expérience. Une fois que la limite d'erreur minimale contrainte par les données du problème (réseau d'observation, paramétrisation de la matrice R , covariances d'erreurs) est atteinte, le lisseur n'a globalement plus d'effet qui le différencie du filtre. Il permet donc d'atteindre plus rapidement que le filtre le niveau d'erreur minimal mais pas de l'améliorer. Ce résultat est valable dans le cadre de notre expérience, c'est-à-dire avec une erreur sur la condition initiale uniquement. Un cadre différent (avec une erreur sur les forçages par exemple) mettrait sans doute en évidence d'autres aspects du lisseur.

Nous avons également pu analyser la "cohérence dynamique" des résultats, c'est-à-dire comparer l'efficacité du filtre et du lisseur pour produire un état corrigé destiné à initialiser un run libre. L'analyse du lisseur s'est révélée plus adaptée à se rapprocher de la solution de référence, notamment dans la région des anneaux du Brésil en permettant le phasage des anneaux. Cette amélioration est néanmoins limitée dans le temps. Après environ un mois les solutions des deux runs libres initialisés avec les états analysés du filtre ou du lisseur convergent. L'assimilation de données permet donc (dans le cadre de cette expérience) de corriger la dynamique de courte échelle temporelle, mais pas forcément la dynamique de longues périodes. Le lisseur accentue cette capacité et permet d'améliorer le réalisme de la prévision dans les régions sensibles à la condition initiale.

Nous garderons donc essentiellement trois points majeurs :

- D'une part, avec une expérience liée à la condition initiale, le lisseur s'avère efficace pour abaisser ce niveau d'erreur (compte-tenu d'observations sur une fourchette de temps). Cet aspect est particulièrement vrai dans le cadre d'un réseau d'observation "temps-groupé".
- D'autre part, dans le cadre d'une réanalyse, l'effet du lissage est limité dans le temps par rapport au filtrage, notamment dans le cadre d'observations temps-réel. Cela est essentiellement lié au fait que le niveau d'erreur minimal permis par les paramètres du filtrage a déjà été atteint par le filtre.
- Enfin, au-delà des réanalyses, le lissage séquentiel peut être avantageusement utilisé pour obtenir un état analysé qui soit à la fois proche des observations et conforme à la dynamique du modèle. Les tests d'initialisation de runs de

prévision ont permis de souligner ce point (dans la limite de prévisibilité liée à la dynamique).

- Enfin, un des atouts de l'algorithme de lissage est de pouvoir fournir une solution lissée avec une fréquence temporelle indépendante de la fréquence du réseau d'observations, et donc de fournir une solution corrigée plus continue dans le temps (illustré avec le réseau décadaire ici).

Le gain apporté par le lissage par rapport au filtrage est donc dans l'ensemble avantageux. Nous aurions pu nous attendre à une différence plus nette (notamment sur les erreurs RMS), mais la paramétrisation de notre expérience limite naturellement ce gain (en terme d'amplitude et aussi dans le temps, la solution du lisseur convergeant peu à peu vers celle du filtre). L'efficacité du lisseur sur les états proches de l'erreur introduite (sur la condition initiale) sont toutefois convaincants. Dans le cadre de réanalyses, nous suggérons donc l'utilisation du lisseur pour fournir des solutions plus continues (même dans le cas d'un échantillonnage temporel peu dense des observations) et plus proches de la dynamique réelle de l'océan.

Ces résultats amènent également de nouvelles pistes d'études que nous exposerons lors des conclusions générales de la thèse. L'un de ces points va néanmoins être étudié dans le chapitre suivant. En effet, comme nous l'avons vu dans le chapitre 7, la paramétrisation de l'évolution des erreurs n'est pas aisée à mettre en place dans une configuration réaliste. Bien qu'elle soit normalement indispensable au lissage (d'après la théorie exposée dans le chapitre 4), sa mise en place et son coût numérique peuvent être décourageants. Actuellement, l'utilisation de filtres évolutifs n'est d'ailleurs pas présente dans les systèmes opérationnels. De ce fait, nous nous sommes posé la question du comportement du lisseur dans un système plus simplifié et peu coûteux (donc comparable avec les contraintes opérationnelles) dans lequel les erreurs n'évoluent pas. Cette question fera l'objet du dernier chapitre de ce manuscrit.

Chapitre 9

Mise en place et évaluation du lisseur basé sur un filtre de base fixe

Sommaire

9.1	Introduction	152
9.2	Mise en place des expériences	152
9.3	Représentation 4D des covariances d’erreur	153
9.4	Lissage avec une base de covariances d’erreur “semi-fixe”	159
9.4.1	Réseau “temps-réel”	159
9.4.2	Réseau “temps-groupé”	161
9.4.3	Bilan et compléments sur les réanalyses avec une base fixe	162
9.5	“Cohérence dynamique” des analyses avec une base de covariances d’erreur fixe	162
9.5.1	Réseau “temps-réel”	162
9.5.2	Réseau “temps-groupé”	163
9.5.3	Bilan et compléments sur la “cohérence dynamique” avec un système à base fixe	164
9.6	Conclusion	165

9.1 Introduction

La propagation des matrices de covariances d'erreur dans un filtre de Kalman est théoriquement indispensable pour avoir un lisseur cohérent (chapitre 4). En effet, c'est à travers cette propagation que les corrélations temporelles entre une observation et le moment de l'analyse (antérieure) vont être prises en compte de façon optimale. Cette propagation est néanmoins coûteuse et délicate à mettre en place : dans le cas d'un filtre évolutif, il faut $1+N$ itérations modèle entre chaque étape d'assimilation (N étant le nombre d'EOFs). Dans le cadre d'un système tel que l'OI, il faut simplement 1 itération modèle à chaque étape (correspondant à la propagation de l'état analysé). C'est pourquoi généralement dans les systèmes opérationnels les filtres sont utilisés avec une paramétrisation de type "base fixe", c'est-à-dire un filtre où les covariances d'erreur ne sont pas propagées explicitement par le modèle. Bien qu'il s'agisse d'une approximation forte de la théorie (surtout dans le cadre du lissage), cette formulation présente l'avantage d'être peu coûteuse et peut fournir (dans le cadre du filtrage) de bons résultats. Nous avons donc décidé de tester l'efficacité du lisseur dans le cadre d'une telle approximation.

9.2 Mise en place des expériences

Dans ce chapitre, nous garderons exactement le même protocole expérimental que les expériences du chapitre précédent, avec les mêmes réseaux d'observations et les mêmes paramètres liés au filtre (nombre d'EOFs, matrice \mathbf{R} , taille de la zone d'influence des analyses). La différence est qu'ici le filtrage est effectué avec une matrice de covariances d'erreur fixée \mathbf{S}^f , correspondant à la base d'EOFs initiale : $\mathbf{S}^f = \mathbf{S}_0$. Ainsi, pour chaque analyse du filtre, c'est toujours la même base qui est utilisée. Il s'agit donc d'un cas d'Interpolation Optimale, OI (*Brasseur et Verron, 2006*).

Dans une telle configuration, le gain de Kalman du lissage (équation 5.32) se réécrit :

$$\mathbf{K}_{i|k} = \mathbf{S}_{i|k-1}^a [\mathbf{I} + \mathbf{\Gamma}_k]^{-1} (\mathbf{H}_k \mathbf{S}_0)^T \mathbf{R}_k^{-1} \quad (9.1)$$

où \mathbf{S}_0 est donc la base d'EOFs initiale. Le lisseur tient donc compte des analyses précédentes du filtre par la matrice \mathbf{S}^a) mais pas de la propagation dynamique des erreurs. Il est donc assez délicat d'avoir une idée de la représentation des statistiques d'erreur dans une telle formulation. Nous allons donc, dans un premier temps, examiner de façon qualitative cette représentation à travers une expérience simple décrite dans la section suivante. L'étude du lissage se fera ensuite dans les sections 9.4 et 9.5.

Dénomination	Expérience initiale	Date des observations ponctuelles
SO-F1	filtre	20258
SO-F2	filtre	20276
SO-OI1	OI	20258
S0-OI2	OI	20276

TABLE 9.1 – Dénominations des différentes expériences d'observations ponctuelles en précisant pour chacune de quelle expérience précédente sont tirés les résultats et les covariances d'erreur (expérience du filtre ou de l'OI) et à quelle date sont introduites les observations ponctuelles.

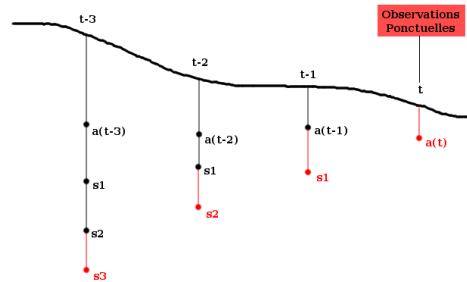


FIGURE 9.1 – Schématisation de l’expérience “d’Observations Ponctuelles”. La courbe noire symbolise un niveau d’amplitude de l’erreur RMS d’un run à corriger. Les symboles en noir font référence à l’expérience précédente du filtre. Les symboles en rouge correspondent aux étapes liées à l’ajout d’observations ponctuelles au temps t . Les notations sont : a (analyse du filtre au temps considéré) et s^x (x analyses rétrospectives du lisseur au temps considéré).

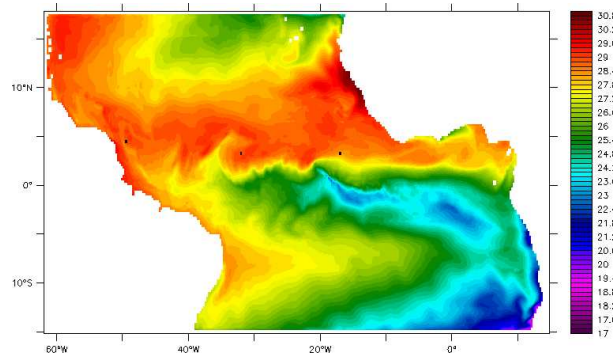


FIGURE 9.2 – Carte de SST (en °C) sur laquelle sont positionnées les observations ponctuelles (rectangles noirs) correspondant à des profils isolés spatialement de température.

9.3 Représentation 4D des covariances d'erreur

Afin d’illustrer les différences entre les représentations des covariances dans le cadre du filtre évolutif et dans le cadre de l’OI, nous construisons une expérience avec un jeu d’observations ponctuelles. Le principe est simplement de regarder quelle va être l’influence d’une observation isolée sur le système, et notamment son influence rétropropagée par le lissage. Pour se faire (voir figure 9.1 et tableau 9.1) :

- 0) nous avons d’abord mis en place deux expériences de filtrage (avec un jeu d’observations “temps-réel”) : l’une identique à celle décrite dans le chapitre précédent, et l’autre utilisant l’OI (ces expériences seront détaillées dans les sections 9.4 et 9.5). A partir du filtrage, le lisseur est utilisé jusqu’à une certaine date (20256) pour obtenir des états lissés sur les états antérieurs.

- i) un état de prévision au temps $t=20258$ (x_t^f) est extrait de l’expérience du filtre évolutif, c’est-à-dire 4 cycles après le début de l’expérience du filtre

- ii) toutes les statistiques d’analyse du filtre ou du lisseur précédant cet état sont également récupérées (c’est-à-dire, sur la figure 9.1, les états représentés en noir : \mathbf{a}_{t-1} , \mathbf{s}_1 et \mathbf{s}_2 , ainsi

que les matrices de covariances d'erreur associées)

- iii) un nouveau jeu d'observations "ponctuelles" (figure 9.2) est alors introduit pour analyser x_t^f au temps 20258 (donnant ainsi \mathbf{a}_t en rouge)
- iv) les analyses rétrospectives du lisseur sont effectuées aux temps $t-1$, $t-2$ et $t-3$ avec ces mêmes observations, donnant ainsi, sur la figure 9.1, les états représentés en rouge : \mathbf{s}_1 , \mathbf{s}_2 et \mathbf{s}_3

Pour les étapes iii) et iv), les covariances d'erreurs utilisées sont celles issues de l'expérience de filtrage initiale. On répète également la même expérience à un temps $t=20276$ (c'est-à-dire 13 cycles après le début de l'expérience du filtre). De façon résumée, cette expérience consiste à interrompre une expérience de lissage à un moment donné (effectuée avec le réseau d'observations complet) et à introduire un jeu d'observations très ponctuelles. On regarde ensuite l'influence de ces observations sur le système. Le fait que les observations soient distantes les unes des autres (spatialement) permet de dissocier plus facilement leurs influences respectives.

On recommence ensuite les 2 mêmes expériences en utilisant des résultats issus de l'OI. On obtient ainsi 4 expériences différentes (tableau 9.1) qui permettent de visualiser l'influence rétrospective des observations ponctuelles selon les dates (20258 ou 20276) et le type de filtrage (évolutif ou OI).

Les résultats sont montrés sur les figures 9.3, 9.4, 9.5 et 9.6. Les figures 9.3 et 9.4, montrent d'une part les valeurs absolues des différences en température dans la zone des TIWs et d'autre part les corrections apportées par les analyses. Chaque correction est calculée entre l'état issu de l'analyse avec les observations ponctuelles et l'état précédant cette analyse (traits rouges sur la figure figure 9.1). Des valeurs négatives indiquent que l'analyse a apporté une bonne correction en faisant baisser le niveau d'erreur, et inversement pour les valeurs positives. Les figures d'erreur montrent quant à elle l'erreur initiale avant l'ajout des observations ponctuelles, entre l'état précédent dans la nouvelle analyse et l'état de référence. Les rectangles noirs sur chaque figure indiquent la position des observations ponctuelles introduites. La même chose est montrée pour les vitesses dans la région des anneaux du Brésil sur les figures 9.5 et 9.6.

L'expérience où les observations sont introduites au jour 20258 correspond aux figures 9.3 et 9.5, et l'expérience où les observations sont introduites au jour 20276 correspond aux figures 9.4 et 9.6. A noter qu'il s'agit ici de regarder surtout qualitativement l'influence d'une observation sur le système. Les valeurs quantitatives des corrections ne seront donc pas étudiées en détail.

Les erreurs initiales (avant l'assimilation des observations ponctuelles) de SO-F1 et SO-OI1 sont très similaires, l'influence des observations va donc surtout dépendre de la base de covariances d'erreur utilisée. On observe, pour l'influence des observations, une forte disparité entre les deux cas (OI ou filtre) mais aussi entre les deux régions étudiées (zone de TIWs et zone des anneaux du Brésil). Dans la région de TIWs (figure 9.3), SO-F1 et SO-OI1 ont des corrections d'assez grande échelle. L'OI introduit des corrections beaucoup plus lisses que le filtre, mais dans l'ensemble leurs effets sont comparables. Dans les deux cas, on voit des zones autour des observations où les corrections sont bonnes (diminution de l'erreur) et des zones plus éloignées où l'analyse détériore le résultat. Cette détérioration semble néanmoins un peu plus limitée dans le cas du filtre (avec un ordre

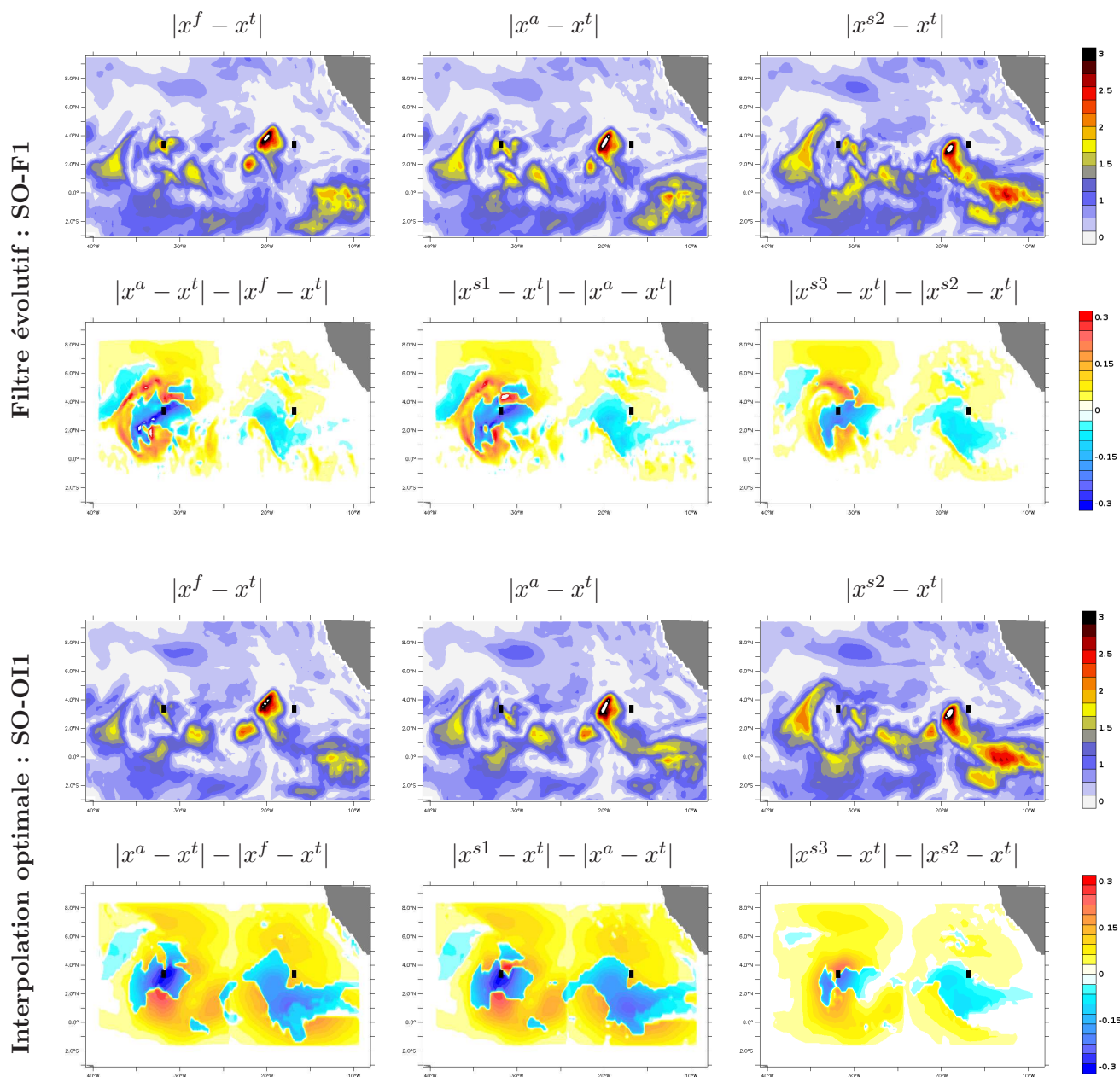


FIGURE 9.3 – Dans la région des TIWs : cartes de l'erreur absolue (première et troisième ligne) en température (en °C) entre le run de référence et chaque état précédent une analyse (du filtre ou du lisseur) avec les observations ponctuelles (introduites au jour 20258), dans le cas du filtre évolutif et dans le cas de l'interpolation optimale. Chaque carte d'erreur est associée en-dessous à une carte de correction indiquant l'effet d'une analyse suite à l'introduction d'une observation ponctuelle. Pour les cartes de correction, les valeurs négatives indiquent une bonne correction (diminution de l'erreur par l'analyse), et inversement pour les valeurs positives. Les corrections et les erreurs sont données (de gauche à droite) pour l'étape d'analyse du filtre, de la première et de la troisième analyses rétrospectives du lisseur. x^t : état de référence, x^f : état prédit avant l'analyse, x^a : état analysé par le filtre, x^{si} : état issu de la i^{eme} analyse rétrospective du lisseur. Les rectangles noirs indiquent la position des observations ponctuelles introduites.

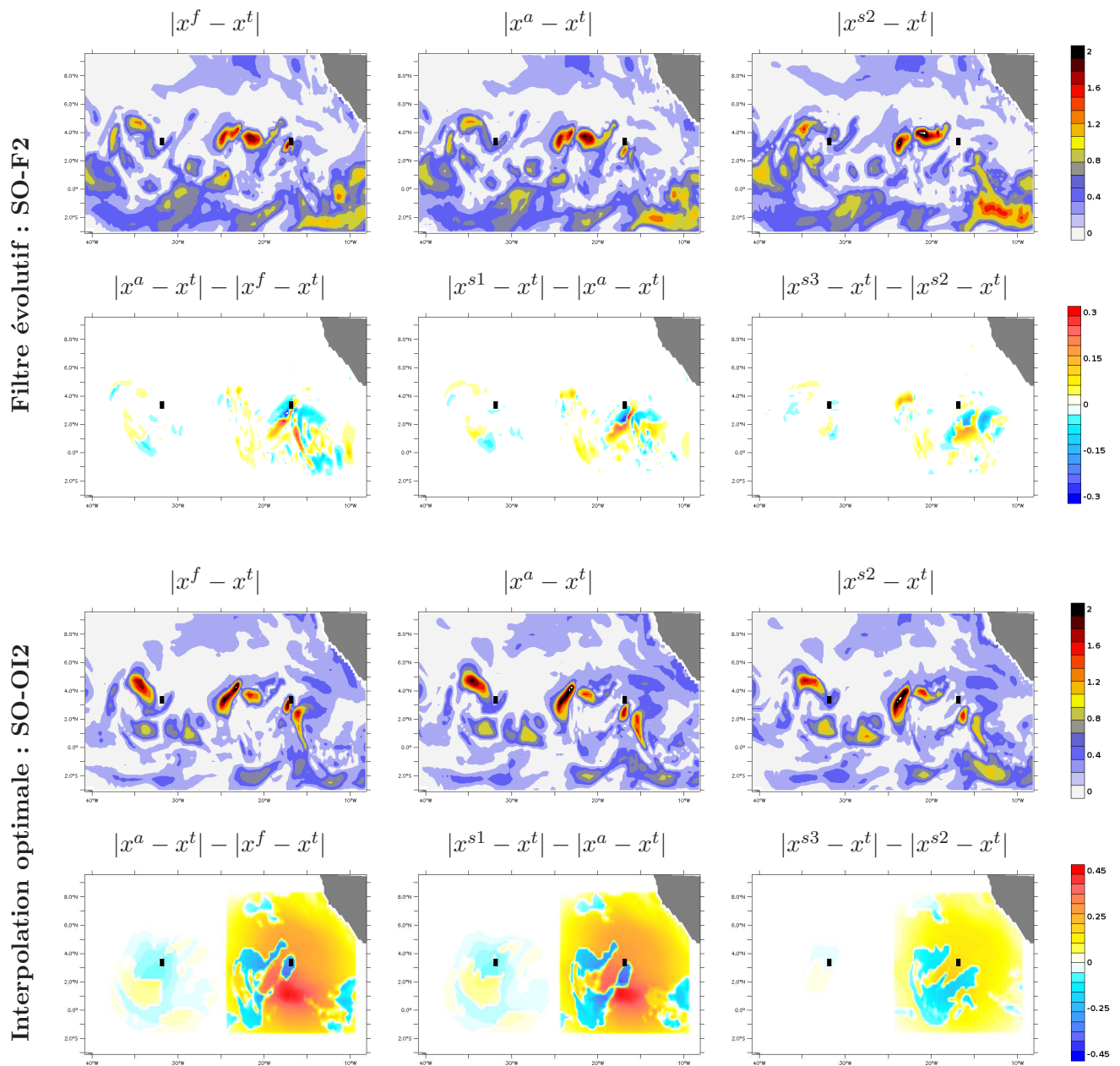


FIGURE 9.4 – Figure identique à la figure 9.3, excepté que l’observation ponctuelle est introduite ici au jour 20276.

de grandeur moins grand). On peut remarquer aussi que dans les deux cas la forme de la correction n’est pas vraiment ressemblante à la forme de l’erreur.

Dans la région des anneaux du Brésil (figure 9.5) les résultats de l’OI montrent une correction avec une forte empreinte de la dynamique des anneaux. Cela suggère que cette dynamique est déjà bien inscrite dans les EOFs initiales. Si les corrections apportent à la fois des améliorations et des détériorations, on notera que la correction de la troisième analyse rétrospective du lisseur ($|x^{s3} - x^t| - |x^{s2} - x^t|$) a plutôt tendance à dégrader la solution. Pour le filtre en revanche la correction présente des structures de plus petites

échelles dues aux analyses précédentes et à la propagation des modes. On retrouve à la fois des améliorations et des détériorations, mais on notera que les analyses du lisseur ont un effet plutôt positif.

Ces constatations supposent que les statistiques d'erreurs ne s'adaptent pas à la même vitesse selon les régions. Après 3 cycles d'assimilations, l'estimation des erreurs par le filtre est encore assez peu ciblée (elle ressemble plus à l'estimation de l'OI). En revanche dans la région des TIWs, l'estimation du filtre est déjà plus proche de l'erreur réelle.

Dans les expériences SO-F2 et SO-OI2 on retrouve les mêmes types de résultats. Cette fois néanmoins SO-F2 présente des structures de petite échelle même dans la zone des TIWs (figure 9.4). La taille de ces structures sont du même ordre de grandeur que celles observées sur les erreurs réelles, ce qui suggère que la propagation des modes joue un rôle positif sur la représentation des erreurs. Ce point est d'autant plus vrai si l'on regarde la taille des structures de la correction sur SO-OI2. L'étendue des corrections n'est pas du tout en accord avec la taille des erreurs. Cela résulte en de larges zones détériorées par les analyses. On remarquera également, notamment dans le cas du filtre, que l'observation située à 32°W se trouve dans une zone d'erreur quasiment nulle, et donc l'effet des analyses liées à cette observation en sera tout autant nulle (ou presque). Cela montre que, quel que soit le système d'assimilation utilisé, son efficacité est fortement contrainte par le réseau d'observations. Il est donc nécessaire d'avoir accès à un maximum d'observations pour limiter les chances d'une innovation nulle.

Dans la région des anneaux du Brésil (figure 9.6) les corrections apportées par SO-F2 et SO-OI2 sont chacune marquées par la dynamique de la région, bien que dans le cas de SO-F2 les échelles soient légèrement plus fines. Il est difficile de savoir si les améliorations surpassent les détériorations, mais on peut voir que dans le cas du filtre les corrections suivent de façon très parlante la trace des erreurs (les corrections apportées sur le tourbillon à $8^\circ\text{N}-54^\circ\text{W}$ par exemple, ou sur la branche de réflexion du NBC autour de $4^\circ\text{N}-44^\circ\text{W}$). L'OI, bien qu'ayant accès à la signature des anneaux, ne permet pas d'effectuer des corrections aussi ciblées.

Après quelques cycles d'assimilation on voit donc clairement le bénéfice de propager les covariances d'erreur pour avoir accès à une bonne représentation des erreurs réelles. Le lissage lié à l'OI garde une certaine bonne influence sur les analyses rétrospectives, mais les corrections ne sont pas adaptées aux tailles des structures réelles des erreurs. On notera aussi que l'influence rétrospective des observations est "artificiellement" limitée par la superposition successive des analyses du lisseur, même dans le cas de l'OI. En effet, dans notre formulation (voir section 5.2.4), le lissage tient compte des analyses précédentes (filtrage ou lissage antérieur) et limite donc de plus en plus l'amplitude des nouvelles corrections associées à des observations supplémentaires.

Puisque, d'après ces expériences, le lisseur associé à l'OI semble garder une certaine influence positive (grâce notamment à la formulation du lisseur telle que nous l'utilisons), nous allons l'utiliser pour une expérience réaliste de lissage.

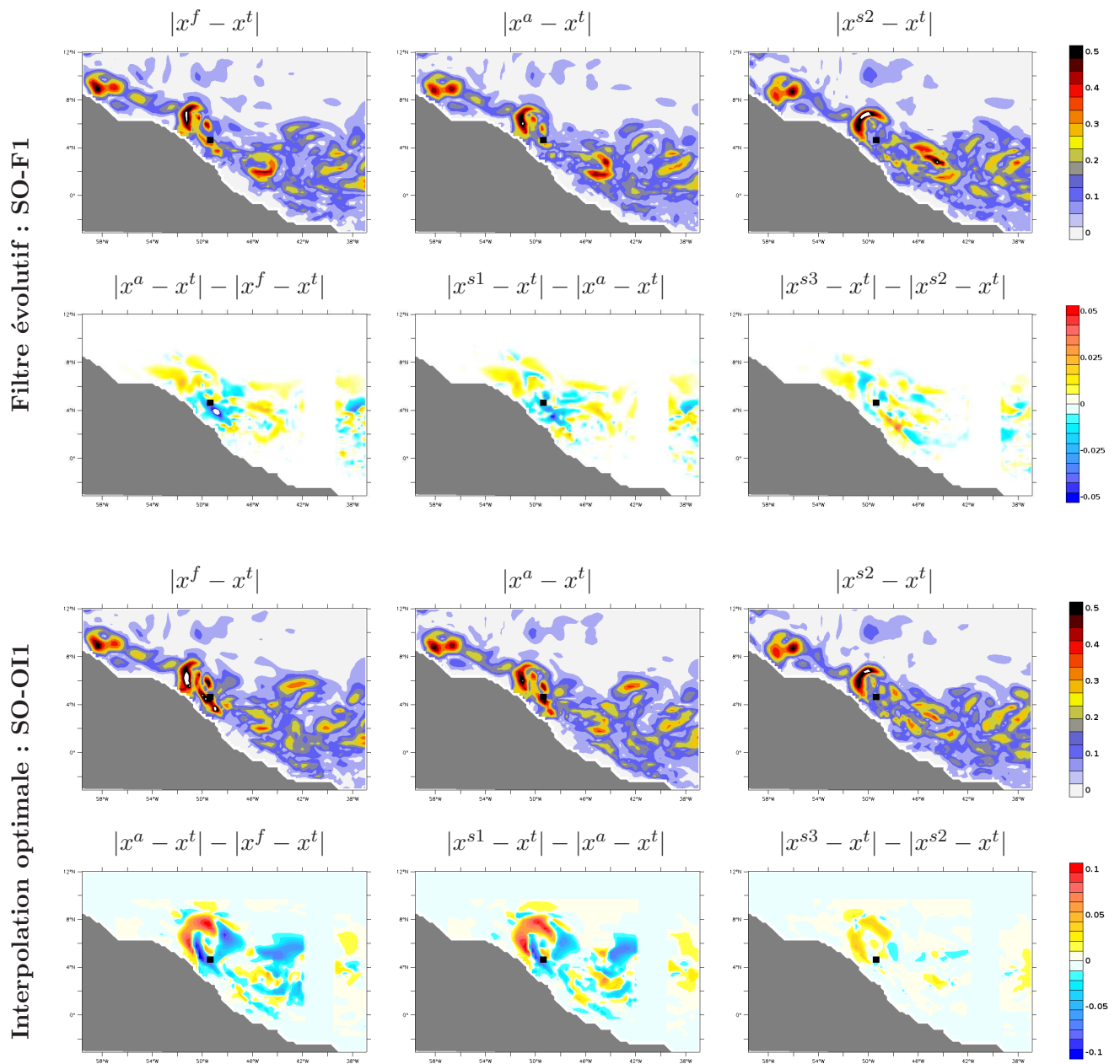


FIGURE 9.5 – Dans la région des anneaux du Brésil, cartes de l'erreur absolue en vitesse (en m.s^{-1}) entre le run de référence et chaque état précédent une analyse (du filtre ou du lisseur) avec les observations ponctuelles (introduites au jour 20258), dans le cas du filtre évolutif et dans le cas de l'interpolation optimale. Chaque carte d'erreur est associée en-dessous à une carte de correction indiquant l'effet d'une analyse suite à l'introduction d'une observation ponctuelle. Pour les cartes de correction, les valeurs négatives indiquent une bonne correction (diminution de l'erreur par l'analyse), et inversement pour les valeurs positives. Les corrections et les erreurs sont données (de gauche à droite) pour l'étape d'analyse du filtre, de la première et de la troisième analyses rétrospectives du lisseur. x^t : état de référence, x^f : état prédit avant l'analyse, x^a : état analysé par le filtre, x^{s^i} : état issu de la i^{eme} analyse rétrospective du lisseur. Les rectangles noirs indiquent la position des observations ponctuelles introduites.

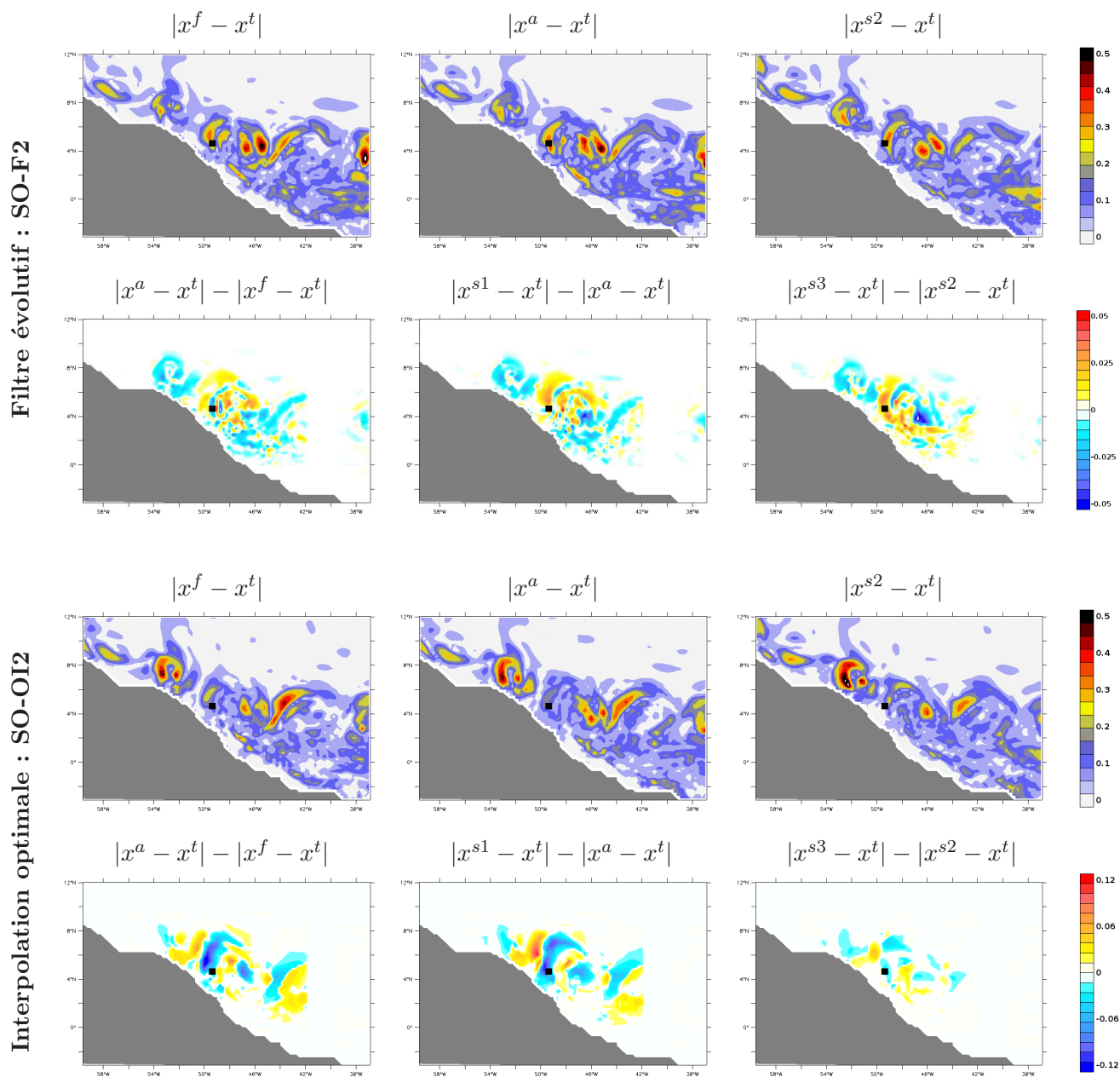


FIGURE 9.6 – Figure identique à la figure 9.5, excepté que l’observation ponctuelle est introduite ici au jour 20276.

9.4 Lissage avec une base de covariances d’erreur “semi-fixe”

9.4.1 Réseau “temps-réel”

Nous allons maintenant étudier l’efficacité des réanalyses avec l’OI et le lisseur associé (en comparant leurs résultats au filtre évolutif et lisseur associé). L’évolution de l’erreur RMS globale (figure 9.7) montre que l’OI (en bleu) et le filtre (en rouge) donnent des

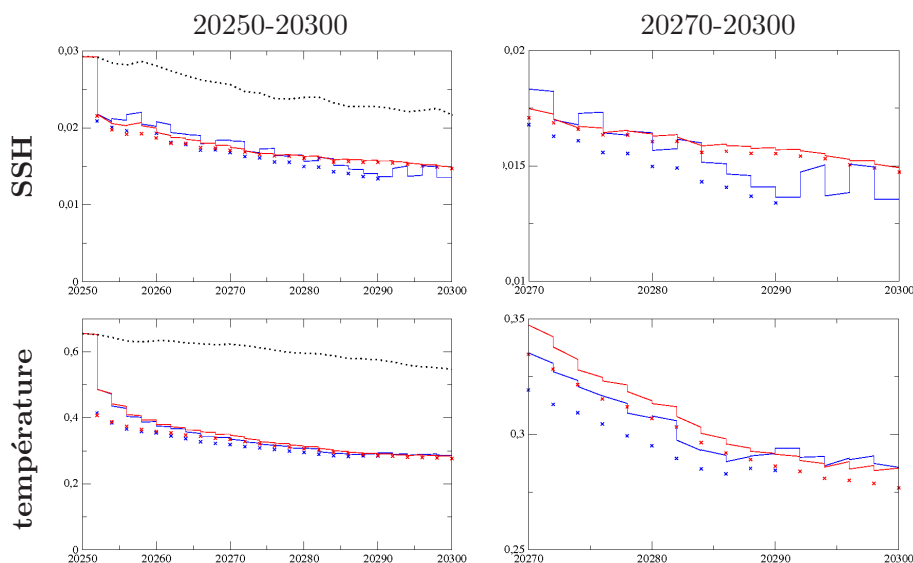


FIGURE 9.7 – Evolution des écarts RMS pour le filtre évolutif (en rouge) et l’OI (tirets bleus), pour la SSH en m (en haut) et la température en °C (milieu). Les croix rouges et bleues correspondent respectivement aux analyses du lisseur dans le cas du filtre évolutif et dans le cas de l’OI. A gauche se trouve l’évolution de l’erreur RMS sur toute la période, et à droite se trouve un zoom sur la période 20270-20300. La trajectoire du run faux est rappelée par la courbe noire pointillée.

résultats similaires au premier ordre. L’erreur RMS de l’OI diminue de façon comparable à celle du filtre évolutif (figures de gauche). L’amplitude des corrections apportées est donc plutôt contrainte par les erreurs d’observation imposées (et le système d’observations) que par la base de covariances utilisée. Néanmoins, lorsque l’on regarde un peu plus en détail le comportement de cette erreur (figures de droite), on remarque que l’OI a tendance à alterner de bonnes et de mauvaises analyses, faisant osciller l’écart RMS autour du minima permis par les conditions de l’expérience. Cet effet est surtout visible sur la SSH et la température qui sont deux variables assimilées. Le filtre évolutif a plutôt tendance à toujours faire décroître (même faiblement) le niveau d’erreur sur ces variables. Sur les variables non assimilées (vitesses), cet aspect est aussi présent mais moins flagrant (non montré). Il peut arriver que l’OI dégrade légèrement l’erreur RMS des vitesses, mais de façon générale son influence est plutôt positive. L’évolution des erreurs semble donc indispensable ici si on veut garder des analyses cohérentes avec les analyses précédentes et l’erreur réelle qui en résulte (notamment sur les variables assimilées). La base de covariances d’erreur initiale est surtout bonne pour permettre une correction globale du premier ordre.

Le lissage associé à l’OI (croix bleues) montre une influence positive sur toute la durée de l’expérience et quelles que soient les variables. Il permet de diminuer le niveau d’erreur global de l’OI, notamment sur les variables assimilées, et fournit une solution plus lisse. On n’observe pas d’alternance de bonnes ou mauvaises analyses contrairement à l’OI. En dépit de l’incohérence avec la formulation théorique, le lisseur utilisé sur une base d’OI (dans un contexte opérationnel donc) conserve un apport bénéfique pour les réanalyses. Cela peut s’expliquer notamment par le fait que nous assimilons ici des observations tous les deux jours, la dynamique n’a donc pas beaucoup le temps de changer. Si les observations peuvent rester en accord avec la dynamique sur plusieurs jours, elles ont donc

une représentativité temporelle plus étendue que la durée de nos cycles d'assimilation. Dans ce cas, le fait que le lisseur ne prenne pas explicitement en compte la propagation des erreurs ne dégrade pas son analyse. Pour les observations plus éloignées (une semaine, voire 10 jours), leur influence est limitée par la formulation de l'algorithme du lissage.

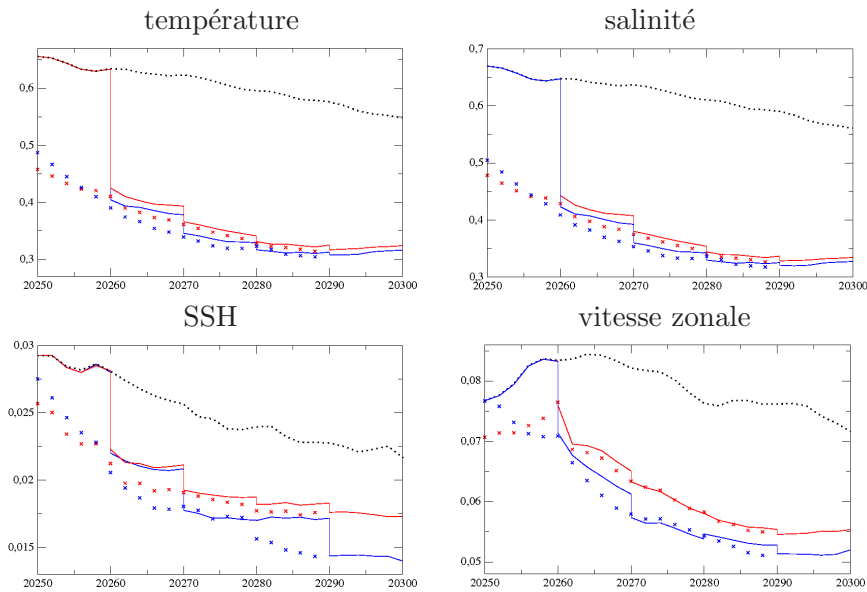


FIGURE 9.8 – Evolution des écarts RMS pour le filtre évolutif (en rouge) et l’OI (en bleu), pour la température en $^{\circ}\text{C}$, la salinité, la SSH en m et la vitesse zonale en $\text{m}\cdot\text{s}^{-1}$. Les croix rouges et bleues correspondent respectivement aux analyses du lisseur dans le cas du filtre évolutif et dans le cas de l’OI. La trajectoire du run faux est rappelée par la courbe noire pointillée.

9.4.2 Réseau “temps-groupé”

Si maintenant on considère le cas d’un réseau décadaire (figure 9.8), les résultats sont un peu différents. L’OI produit de nouveau des résultats globalement similaires au filtre au premier ordre, mais cette fois on observe moins de dégradations au niveau des analyses de l’OI. Toutes les variables semblent plutôt corrigées dans la bonne direction. En revanche, lorsque l’on considère les analyses associées au lissage de l’OI, on constate qu’elles ne sont pas toujours effectuées de façon optimales. Sur les variables assimilées, le lisseur de l’OI est moins efficace pour améliorer la condition initiale (20250). De plus, sur certains tronçons, les analyses du lisseurs ont tendance à détériorer légèrement la solution (entre 20270 et 20280 pour la SSH par exemple). L’évolution des erreurs semble limiter ces problèmes et donner une solution (en rouge) plus perspicace. Enfin, pour les variables non assimilées (vitesse), on retrouve un peu les mêmes comportements mais de façon amplifiés (notamment sur l’amélioration de la condition initiale qui n’est pas du tout optimale dans le cas du lissage de l’OI).

Avec un réseau de type décadaire, où les observations introduisent des erreurs de représentativité temporelle, l’évolution des erreurs semble donc nécessaire pour garder une bonne influence du lissage sur toute la période. On notera également que même avant la première date observée (20260), l’évolution des erreurs permet de faire remonter plus efficacement l’in-

formation des observations dans le temps, indépendamment du fait qu'il y ait eu ou non des analyses antérieures. Cet aspect s'illustre surtout sur les vitesses.

9.4.3 Bilan et compléments sur les réanalyses avec une base fixe

Avec les résultats obtenus dans cette section nous avons pu constater que le lisseur peut avoir un apport positif pour les réanalyses, et ce, même avec un système à base fixe (OI). Ce résultat est vrai particulièrement avec un système d'assimilation de type "temps-réel", où les observations ne sont pas trop décorréélées entre la date d'assimilation et la date de mesure. En revanche, comme il l'a été illustré avec un réseau décadaire, le lisseur associé à l'OI devient moins efficace lorsque l'on utilise un réseau de type "temps-groupé". Cela est sans doute lié au fait que les fenêtres temporelles entre les observations sont assez grandes (10 jours) et donc l'assimilation rétrospective est plus délicate si on ne prend pas en compte le temps qui sépare une analyse des observations. **L'algorithme du lisseur que nous utilisons est particulièrement bien indiqué pour un cas simplifié d'OI car sa formulation (introduisant séquentiellement les analyses) permet de limiter peu à peu l'influence des observations avec le temps.** Par contre l'empreinte de la dynamique (qui est normalement prise en compte par l'évolution des covariances) devrait également être considérée pour optimiser un peu mieux ce système (en faisant par exemple évoluer uniquement les premières EOFs principales). Mais ce genre d'expérience ne sera pas effectuée dans le cadre de nos travaux.

Nous allons compléter ces résultats avec en regardant la "cohérence dynamique" des analyses, tout comme nous l'avons fait dans le chapitre précédent.

9.5 "Cohérence dynamique" des analyses avec une base de covariances d'erreur fixe

9.5.1 Réseau "temps-réel"

Pour comparer les résultats de "cohérence dynamique" entre les analyses liées au filtre et celles liées à l'OI nous effectuons des simulations initialisées à partir des états de deux dates : au jour 20252 et au jour 20270 (figure 9.9).

Dans le premier cas, il s'agit de la première date d'analyse, les statistiques d'erreurs ont eu peu de temps pour évoluer (entre le début de l'expérience, au jour 20250, et la date de la première analyse, au jour 20252). Les analyses du filtre et de l'OI sont donc très proches. En revanche le lissage est effectué avec des observations allant jusqu'au jour 20260, avec des statistiques ayant évolué sur toute la période 20252-20260. C'est donc surtout les différences entre les deux lissages au jour 20252 qui nous intéressent. On observe une différence significative sur les premières semaines de run libre (jusqu'au jour 20300). Le lisseur issu de l'OI (en pointillés bleus) donne une erreur plus importante, que ce soit sur les variables assimilées (comme ici la température) ou non (les vitesses). L'évolution des covariances d'erreur semble donc avoir un impact non négligeable au début pour donner une analyse plus en accord avec la dynamique du modèle.

Au jour 20270, les statistiques d'erreur ont eu le temps d'évoluer et de donner de réelles différences entre le filtre et l'OI. Ces différences se traduisent par un abaissement du niveau global d'erreur RMS pour l'OI (et le lisseur associé) sur la température. Le lisseur associé à l'OI donne lieu à un run libre qui limite fortement la dérive de l'erreur RMS vers le

jour 20320. Cet aspect est notamment illustré sur les vitesses : tous les runs sauf celui en pointillés bleus dérivent de la même façon. Cette dérive est liée au déphasage d’anneaux dans la région du Brésil (comme nous avons déjà pu le voir dans le chapitre précédent).

Le lissage semble donc surprenamment efficace avec l’OI, surtout après quelques jours

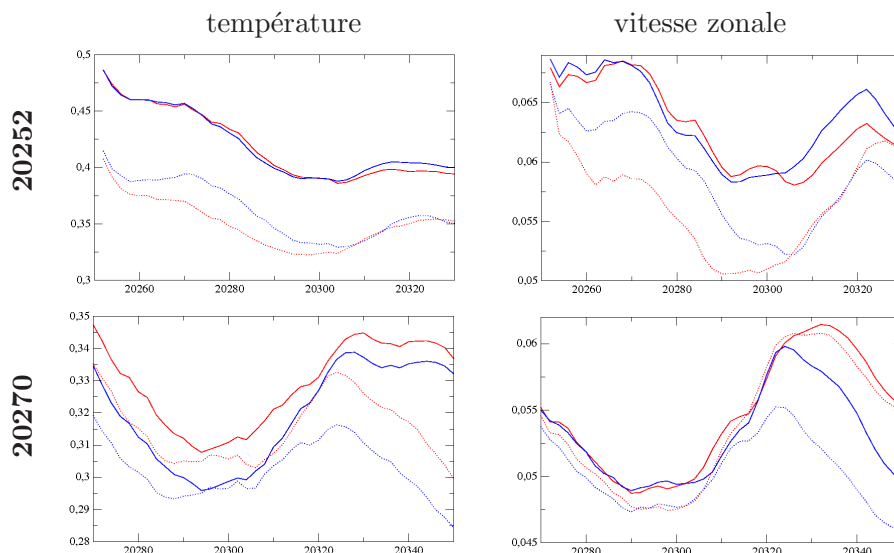


FIGURE 9.9 – Evolution de l’erreur RMS vraie pour le run libre redémarrant à partir d’un état analysé par le filtre (trait plein rouge) ou d’un état analysé par l’OI (trait plein bleu), et pour les runs libres redémarrant à partir du lisseur associé au filtre (trait pointillé rouge) ou associé à l’OI (trait pointillé bleu). Les courbes sont données pour la température (en $^{\circ}\text{C}$) et la vitesse zonale (en m.s^{-1}), avec des redémarrages effectués aux jours 20252 (en haut) et 20270 (en bas).

d’évolution (jour 20270). L’évolution des erreurs semble plus pertinente au début de l’expérience, et perdre en efficacité avec le temps. Néanmoins dans le cas de la correction d’une erreur sur l’état initial, le fait de faire évoluer les erreurs sur quelques cycles permet d’obtenir un état initial lissé plus efficace.

9.5.2 Réseau “temps-goupé”

Avec un réseau d’observations de type “temps-goupé” nous effectuons des redémarrages à la date de la première analyse (20260) et dix jours plus tard (20270). A noter que les covariances d’erreur commencent à évoluer à travers la dynamique dès le jour 20250 (début de l’expérience). La figure 9.10 montre les résultats pour les écarts RMS en température et sur la vitesse zonale.

Lors de la première analyse, au jour 20260, les états du filtre et de l’OI semblent assez proche. Par contre les analyses issues du lissage diffèrent fortement. On observe peu de différence à court terme (courbes pointillées rouge et bleue), mais ensuite, autour du jour 20320 les deux runs divergent. Le lissage lié au filtre évolutif présente une trajectoire proche de celle du filtre. Le lisseur lié à l’OI au contraire montre une bonne aptitude à conserver une trajectoire qui tend vers la référence (et permet d’éviter la dérive observée dans les autres cas).

Les résultats après quelques jours d’évolution (au jour 20270) sont très différents. Les runs issus de l’analyse du filtre et du lisseur associé présentent des trajectoires proches.

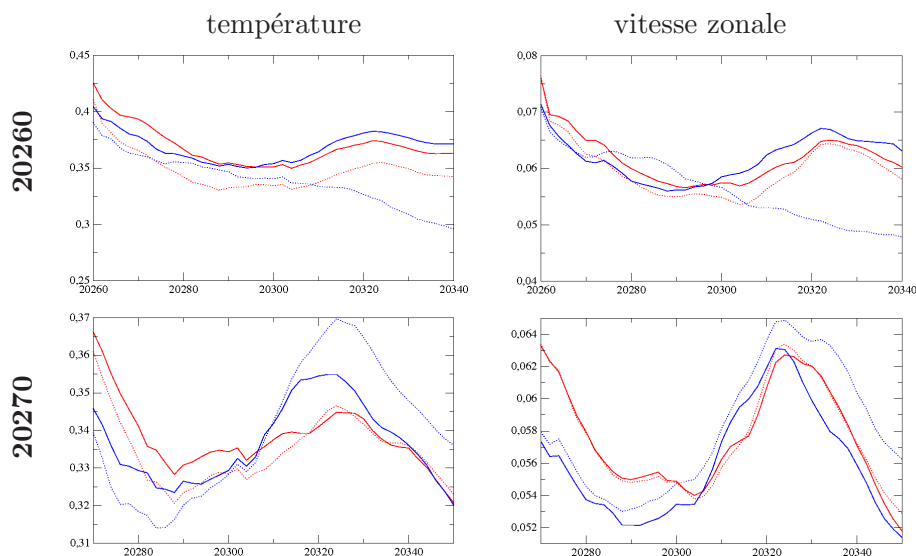


FIGURE 9.10 – Dans le cadre d’un réseau de type “temps-groué” : évolution de l’erreur RMS vraie pour le run libre redémarrant à partir d’un état analysé par le filtre (trait plein rouge) ou d’un état analysé par l’OI (trait plein bleu), et pour les runs libres redémarrant à partir du lisseur associé au filtre (trait pointillé rouge) ou associé à l’OI (trait pointillé bleu). Les courbes sont données pour la température (en °C) et la vitesse zonale (en m.s^{-1}), avec des redémarrages effectués aux jours 20260 (en haut) et 20270 (en bas).

Le lissage n’améliore donc pas vraiment l’analyse pour cet état. Pour le cas de l’OI en revanche, la trajectoire issue du lissage est globalement moins bonne, notamment après la date 20300. Le lisseur a tendance à accentuer la dérive déjà observée pour l’OI, suggérant ainsi qu’il a amplifié le déphasage des anneaux dans la région du Brésil. Cet aspect a été prévenu par l’évolution des covariances d’erreur dans le cas du filtre.

Les constatations faites avec un réseau décadaire sont donc plutôt contradictoire que celles faites avec le réseau “temps-réel”. Ici, le lisseur associé à l’OI semble plus efficace en début d’expérience. Après quelques cycles, le lissage ne semble plus être en mesure de prendre en compte correctement les observations sans qu’il y ait d’évolution des covariances d’erreur.

9.5.3 Bilan et compléments sur la “cohérence dynamique” avec un système à base fixe

Nous avons pu le voir dans cette section, les analyses issues de l’OI et du lissage associé sont relativement bien cohérentes. Le lisseur est capable, même dans un système à base de covariances d’erreur fixes, d’améliorer la représentation de la dynamique. Ce résultat n’est néanmoins plus valide dans certaines conditions, comme nous avons par exemple pu le constater avec un système d’observations “temps-groué” après plusieurs jours de propagations. En réalité il est assez délicat dans ce cas de vouloir généraliser les résultats à tous les cas. **On peut simplement suggérer que la solution lissée reste meilleure que la solution filtrée (pour l’OI) dans le cas où les observations ne sont pas trop décorréllées avec l’état analysé (ou bien dans le cas où la dynamique évolue peu).** Si les observations sont trop distantes en temps ou si la dynamique est trop chaotique, l’influence du lisseur est plus aléatoire.

9.6 Conclusion

Dans ce chapitre, nous nous sommes orientés sur un problème contradictoire avec la théorie, à savoir utiliser un lisseur de Kalman avec un système d'assimilation de données où les statistiques d'erreur sont fixes. Cette étude était motivée par le fait que les systèmes évolutifs sont peu utilisés en pratiques (dans les centres opérationnels), et que l'OI présente l'avantage d'être peu coûteuse numériquement par rapport au filtre évolutif.

Les résultats du lissage obtenus avec l'OI se sont montrés plutôt efficaces. Les réanalyses effectuées dans ce cadre par le lisseur montrent une nette diminution des erreurs et une cohérence dynamique assez bonne. Nous n'avons pas observé de détériorations marquantes comme nous aurions pu nous y attendre. Cela vient du fait que les assimilations rétrospectives sont limitées à 10 jours (la dynamique évolue donc peu), et dans notre formulation du lissage l'influence des observations est automatiquement réduite avec les assimilations successives. Ainsi, même avec une base fixe, les observations lointaines ont moins d'influence et ne peuvent pas dégrader la solution. Certains cas peuvent tout de même conduire à des états lissés moins cohérents (comme nous avons pu le voir dans la section 9.5). Le lisseur associé à un filtre évolutif permet dans tous les cas d'obtenir des états lissés au moins aussi cohérent (voire meilleurs) que les états analysés par le filtre. Le côté plus "aléatoire" du lissage avec de l'OI est sans doute dû à une mauvaise représentation des erreurs réelles, comme nous avons pu l'illustrer dans la section 9.3. Nous retiendrons donc à l'issue de ce chapitre :

- Bien qu'inconsistant sur le plan théorique, le lissage associé à l'OI est efficace au premier ordre avec la formulation séquentielle du lisseur telle que nous l'utilisons.
- Les corrections apportées par le lissage dans le cadre de l'OI présentent le désavantage d'être peu en accord avec les structures réelles des erreurs. Ce type de lissage sera donc limité pour corriger des erreurs nécessitant plus de précision.
- De façon surprenante, les états lissés restent néanmoins cohérents avec la référence, ce qui est encourageant dans l'idée de la mise en place d'un lisseur basé sur les systèmes actuels des centres opérationnels d'océanographie.

Ce chapitre marque la fin de ce travail de thèse, que nous allons maintenant conclure en revenant sur les résultats clés.

Conclusions et perspectives

Parmi les deux branches principales des méthodes d'assimilation (le contrôle optimal et l'estimation optimale), le filtre de Kalman, appartenant à la deuxième branche, a été largement diffusé dans la communauté océanographique. Dans sa formulation théorique et évolutive, le filtre met en balance les erreurs estimées pour chaque sources d'informations (modèle et observations) afin de fournir un état analysé optimal. Le filtre est ainsi évolutif mais uniquement dans le sens du présent vers le futur. Or, dans un système dynamique tel que l'océan, il existe de fortes corrélations temporelles. Il est donc tout à fait possible *a priori* d'utiliser de façon rétrograde l'information d'une observation pour corriger un état antérieur. Cet aspect n'est pas pris en compte par le filtre, c'est pourquoi une extension de cette méthode a été développée, le lisseur de Kalman, qui permet d'étendre les méthodes séquentielles à une dimension 4D. Cette méthode a été testée dans divers cas idéalisés (par exemples *van Leeuwen, 1999; van Leeuwen, 2001*) ou en rang réduit (*Cosme et al., 2010*) et s'est avérée efficace.

Nous avons, dans le cadre de ce travail de thèse, mis en place le lisseur SEEK de type "fixed-lag" (*a priori* adapté pour les réanalyses) dans un problème réaliste (tant par la modélisation que par les réseaux d'observations simulés) et nous avons pu étudier son comportement par des expériences jumelles. Les objectifs étaient essentiellement de tester l'efficacité de cet outil dans le cadre de réanalyses, et de définir les atouts ou difficultés de sa mise en place dans un système proche des systèmes réels.

L'implémentation de l'algorithme de lissage s'avère relativement aisée lorsque le filtre évolutif est déjà en place. Le coût numérique supplémentaire du lisseur par rapport au filtre est quasiment négligeable puisqu'il ne nécessite aucune itération modèle supplémentaire. Le stockage des matrices de covariances peut éventuellement poser problème sur plan numérique, mais le fait de travailler en rang réduit et la localisation temporelle introduite par la forme "fixed-lag" du lisseur permettent de limiter le coût de stockage. La mise en place de la version évolutive du filtre de Kalman (théoriquement nécessaire au lisseur) est une phase qui peut être relativement délicate. Cet aspect a été l'une des difficultés pointée par le passage d'une configuration idéalisée (*Cosme et al., 2010*) à une configuration réaliste.

Divers points et conclusions ont pu être soulevés tout au long de nos travaux, que nous allons résumer ci-dessous. Nous pouvons orienter ces résultats selon trois principales thématiques : la stabilisation d'un filtre SEEK évolutif, l'apport d'un lisseur SEEK sur l'optimisation de l'utilisation des observations, et l'intérêt d'utiliser des covariances d'erreur propagatives dans le cadre du lissage.

La propagation dynamique des modes d'erreur et la prise en compte de l'erreur modèle sont nécessaires dans le cadre du lissage pour représenter correctement les corrélations temporelles entre une observation et un état antérieur (chapitre 4). Ces deux aspects ont été étudiés dans la 4^{ème} partie de ce manuscrit. Nous avons pu retenir :

- D'une part, la mise en place d'un filtre évolutif (chapitre 7) dans un contexte réaliste (modèle haute résolution de l'Atlantique Tropical) s'est avérée délicate. Même dans un cadre d'implémentation relativement bien contrôlé, dans lequel seule la condition initiale est inconnue, l'erreur de représentativité à introduire pour corriger l'effet de troncature de l'espace d'erreur domine d'un facteur 10

l'erreur de mesure classiquement considérée pour les observations altimétriques et par profileurs. Notons tout de même que le filtre tel que nous l'avons défini est stable et cohérent statistiquement sans avoir utilisé d'artifices pour prévenir sa dérive (facteurs d'amplifications des modes...).

- D'autre part, tenir compte des incertitudes de certains paramètres physiques du modèle (dans notre cas, du schéma TKE) par une méthode de Monte Carlo (chapitre 6) s'est ici avéré particulièrement délicate pour les statistiques des ensembles générés. Les statistiques de ces ensembles ne comportaient que du signal de très petite échelle, plus proche d'un bruit que d'un vrai signal statistique. Et ceci, malgré le fait que les perturbations sur les paramètres étaient globales.

Une fois le filtre évolutif défini de façon optimale pour notre expérience, nous avons pu implémenter ce filtre et le lisseur associé dans une expérience de réanalyses de plusieurs semaines.

L'implémentation du lisseur s'est avérée relativement aisée à la suite du filtre. L'impact du lisseur de Kalman a été présenté dans le chapitre 8, notamment en comparaison des résultats du filtrage. L'utilisation du lissage a montré plusieurs points intéressants. Son effet peut être regardé selon deux points de vue : d'une part sur l'aspect réanalyses des états (avec les écarts RMS), et d'autre part sur la représentation de la dynamique dans les états analysés et leur cohérence avec le modèle. Suite à ces expériences, nous retiendrons essentiellement :

- Dans une expérience liée à une perturbation sur la condition initiale, le lisseur s'avère efficace pour abaisser le niveau d'erreur initiale en utilisant les informations issues des observations ultérieures à cet état. La trajectoire des réanalyses par lissage est ainsi plus proche de la réalité et plus continue (on évite les sauts de trajectoire généralement introduits par l'assimilation séquentielle). Cet aspect est particulièrement vrai dans le cas de réseaux d'observations basse fréquence (décadaires).
- Les analyses du lisseur se sont également avérées de façon générale plus en adéquation avec la réalité et le modèle (ce que nous avons appelé la "cohérence dynamique"), comme l'ont montré les tests de redémarrage effectués.
- Enfin, nous avons pu constater que l'effet du lissage, tout comme le filtrage, est limité dans le temps dans le cadre de notre expérience (avec une convergence, après quelques semaines, des solutions filtrée et lissée vers un minima d'erreur).

Dans un cadre réaliste et complexe, le lisseur a donc montré une forte capacité à améliorer les résultats issus d'un simple filtrage 3D, grâce à la prise en compte des corrélations 4D des observations. Ce gain est limité dans le temps, sans doute à cause de la définition de notre expérience où l'erreur est liée à la condition initiale. Nous avons également souhaité terminer ce travail de thèse en nous rapprochant d'un cadre plus opérationnel (moins coûteux numériquement) où l'évolution des erreurs n'est plus prise en compte explicitement.

Une méthode du type interpolation optimale (OI) n'est normalement pas en accord avec la théorie du lissage. Néanmoins, cette méthode, plus facile à mettre en place et

numériquement moins coûteuse que le filtre de Kalman, est souvent utilisée dans les centres opérationnels, c'est pourquoi nous nous sommes intéressés à un cas de lissage avec une base de covariances d'erreur fixe, définissant un nouvel algorithme de lissage de base "semi-fixe" (chapitre 9). Les expériences menées dans ce chapitre nous ont permis de retenir plusieurs points :

- Bien qu'inconsistant sur le plan théorique, l'implémentation du schéma de lissage séquentiel (présenté et utilisé dans cette thèse) avec de l'interpolation optimale peut s'avérer bénéfique. Certaines précautions s'imposent cependant pour éviter des analyses du lisseur contre-productives, en limitant notamment fortement la taille des fenêtres d'assimilation rétrospectives.
- Les corrections apportées par le lissage dans le cadre de l'OI présentent le désavantage d'être peu en accord avec les structures réelles des erreurs. Dans un système où les corrections nécessiteraient plus de précision, on peut supposer que le lissage lié à l'OI aurait un effet néfaste.
- Les états lissés restent tout de même surprenamment cohérents avec la référence.
- Un des atouts de l'algorithme de lissage (illustré aussi bien dans le cas de l'OI que du filtre évolutif) est de pouvoir fournir une solution lissée avec une fréquence temporelle indépendante de la fréquence du réseau d'observations, et donc de fournir une solution corrigée plus continue dans le temps.

Le lisseur séquentiel a donc montré une étonnante capacité d'analyse dans un système simplifié similaire à de l'interpolation optimale. Ces derniers résultats sont encourageants pour étendre par exemple l'usage du lissage à des systèmes opérationnels, ce que nous allons évoquer dans les perspectives de nos travaux.

L'ensemble de ces résultats suggère que, d'une façon générale, la représentation quadri-dimensionnelle des observations à travers le lisseur a un effet positif indéniable pour améliorer de la qualité des réanalyses. Le lisseur séquentiel en rang réduit, tel qu'il a été défini et utilisé dans cette thèse, présente le fort avantage d'être facilement mis en place sur la base d'un filtre préalablement défini (sous condition que ce filtre soit déjà optimal). Il peut ainsi être utilisé sur des expériences de filtrage ayant déjà été effectuées sans nécessiter de nouvelles itérations du modèle. Même dans un cas d'OI, la solution lissée reste meilleure que la solution filtrée (pour l'OI) dans le cas où les observations ne sont pas trop décorréélées avec l'état analysé (ou bien dans le cas où la dynamique évolue peu). La mise en place des méthodes stochastiques 4D est donc plus que jamais justifiée pour une utilisation optimale des observations.

Nous devons tout de même souligner que l'étude menée dans cette thèse est limitée à un cas particulier d'expérience (avec une erreur sur la condition initiale). De nouvelles perspectives peuvent être envisagées à la suite de cette thèse, que nous allons évoquer ci-dessous.

Perspectives

Ce travail a permis de soulever certaines questions qui demandent encore à être résolues. Bien que le sujet soit vaste, les perspectives à l'issue de nos travaux peuvent s'orienter prin-

ci-palement sous deux axes :

- l'évolution et l'adaptativité des erreurs pour une meilleure efficacité du filtre et du lisseur
- la mise en place du filtrage dans un contexte réel

La paramétrisation des covariances d'erreur s'est révélée être l'un des points sensibles soulevés par nos résultats. Nous avons pu montrer que si l'évolution des erreurs apporte des bénéfices quant à l'estimation des erreurs réelles, elle peut également limiter l'efficacité du filtre et du lisseur avec le temps. De même, le fait d'avoir une matrice d'erreur d'observation fixe n'est pas optimal dans un cadre évolutif. Pour parer à ces effets, une solution est de prendre en compte l'adaptativité des erreurs. Un filtre adaptatif (*Brankart et al.*, 2010) permettrait par exemple d'optimiser la représentation des erreurs. Le principe d'un tel filtre est de réajuster les niveaux d'estimation des erreurs (P et R) en fonction des résultats de chaque analyse en appliquant un facteur d'amplification à P et R . Le filtre adaptatif est alors maintenu à un niveau optimal, de même que le lisseur associé. Cette forme adaptative doit encore être étudiée dans un contexte réaliste, mais elle semble indispensable pour optimiser la paramétrisation de la représentation des erreurs.

D'autres pistes peuvent également s'avérer pertinentes pour maintenir une bonne représentation des erreurs. Par exemple, le ré-échantillonnage de la base des covariances d'erreur d'un cycle à l'autre (en remplaçant certains des modes évolués par les modes initiaux) pourrait permettre de garder à la fois l'information sur la variabilité de grande échelle initiale tout en considérant l'évolution des covariances (et donc la représentation temporelle des observations). Ce ré-échantillonnage peut se faire en remplaçant les modes les plus faibles par exemple, ou simplement par tirage aléatoire d'un certain nombre de modes. Le but étant toujours de réajuster correctement le niveau de représentation des erreurs. Ce genre de filtre "semi évolutif" a d'ailleurs déjà été en partie testé, par exemple par *Hoteit et al.* (2001).

Divers pistes sont donc possibles pour conserver une meilleure représentation des matrices de covariances d'erreur. Le lisseur étant très sensible à cette représentation, une meilleure maîtrise de la propagation des modes ne pourrait que lui être bénéfique. De façon plus générale, une paramétrisation plus juste des modes d'erreur et de leur propagation est nécessaire pour une prise en compte optimale des observations dans une dimension temporelle (et spatiale). Cette piste est donc la plus importante dans le contexte d'assimilation 4D.

En parallèle de l'aspect évolutif, la définition des covariances d'erreur initiales peut aussi varier. On pourrait envisager de définir une erreur sur les forçages par exemple, ou sur la résolution du modèle. Ces différentes sources d'erreur rendent forcément l'influence du lissage différente du cas de notre expérience. Il est donc nécessaire de les étudier aussi car, dans un contexte réaliste, les sources d'erreur regroupent toutes les possibilités (sur la condition initiale, les forçages, les paramètres du modèle...).

L'extension du lissage à une gamme d'expériences variées s'avère nécessaire pour avoir une vision plus générale de l'assimilation 4D, le but étant toujours de se rapprocher d'un contexte de plus en plus réel.

Enfin, comme nous l'avons évoqué dans le chapitre 5, la formulation du lissage décrite dans cette thèse peut facilement être appliquée à n'importe quel cas de filtrage (en théorie évolutif) ayant déjà été effectué. Il serait donc pertinent d'utiliser l'assimilation 4D pour des projets réels déjà mis en place (par exemple les réanalyses GLORYS). L'investissement serait faible par rapport aux apports du lissage. Cela permettrait également d'étudier le lisseur dans un contexte réel (tant par les sources d'erreur que par les réseaux d'observation). La contrainte de l'évolutivité des erreurs liée à la théorie du lisseur peut poser certaines limites. Néanmoins, nous avons pu le voir dans le chapitre 9, le lisseur peut montrer des résultats encourageants dans un cas d'interpolation optimale. Son implémentation avec des covariances d'erreur fixes pourrait donc donner des résultats satisfaisants. Une autre piste à explorer serait également la reconstitution de réanalyses continues à travers des runs redémarrés à partir des états lissés (qui se sont avérés plutôt efficace pour l'initialisation de runs libres). On obtiendrait ainsi des réanalyses de meilleure qualité et correspondant à une vraie solution modèle. Cette méthode pourrait également être effectuée de manière séquentielle sur des périodes suffisamment courtes pour prévenir la dérive du modèle et conserver l'apport de l'état initial lissé.

D'une façon plus générale, l'assimilation séquentielle 4D semble avoir une place plus que jamais justifiée dans le panorama des méthodes d'assimilation. A l'heure où les observations sont de plus en plus nombreuses et les modèles de plus en plus précis, l'assimilation se doit de pouvoir utiliser au mieux les informations tirées des observations, et cela tant spatialement que temporellement. Ce problème est d'autant plus intéressant qu'il demeure complexe et peut être abordé sous de nombreux angles.

S'il est encore loin le jour où nous pourrons voyager dans le temps, la transmission des informations vers le passé est quant à elle plus que jamais à portée de main.

Bibliographie

- Aceituno, P., On the functioning of the southern oscillation in the south american sector : Part i : Surface climate, *Mon. Wea. Rev.*, *97*, 505–524, 1988.
- Alexander, M., et J. Scott, The influence of enso on air-sea interaction in the atlantic, *Geophys. Res. Lett.*, *29*, doi :10.1029/2001GL014,347, 2002.
- Allen, M. R., S. P. Lawrence, M. J. Murray, C. T. Mutlow, T. N. Stockdale, D. T. Llewellynjones, et D. L. T. Anderson, Control of tropical instability waves in the pacific, *Geophys. Res. Lett.*, *22*, 2581–2584, 1995.
- Amodei, L., Solution approchée pour un problème d’assimilation de données météorologiques avec prise en compte de l’erreur modèle., *C.R. Acad. Sci. Paris, t. 321, série II*, 1087–1094, 1995.
- Anderson, B. D. O., et J. B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.
- Andersson, E., The ecmwf implementation of the three-dimensional variational assimilation (3d-var), iii : Experimental results., *Q. J. R. Meteorol. Soc.*, *124*, 1831–1860, 1998.
- Arakawa, A., et V. Lamb, Computational design of the basic dynamical processes of the ucla general circulation model, *Methods in Comput. Phys.*, *17*, 173–265, 1977.
- Argo Steering Team, t., On the design and implementation of argo - an initial plan for the global array of profiling floats, *International CLIVAR Project Office Report*, *32*, 1998.
- Asselin, R., Frequency filter for time integrations, *Monthly Weather Review*, *100*, 487–490, 1972.
- Ballabrera-Poy, J., P. Brasseur, et J. Verron, Dynamical evolution of the error statistics with the seek filter to assimilate altimetric data in eddy-resolving ocean models, *Q. J. R. Meteorol. Soc.*, *127*, 233–253, 2001.
- Barnier, B., L. Brodeau, J. L. Sommer, J. Molines, T. Penduff, S. Theetten, A. M. Treguier, G. Madec, A. Biastoch, C. Böning, J. Dengg, S. Gulev, R. Bourdallé, J. Chanut, G. Garric, S. Alderson, A. Coward, B. de Cuevas, K. Haines, G. Smith, S. Drijfhout, W. Hazeleger, C. Severijns, et P. Myers, Eddy permitting ocean circulation hindcasts of past decades, *CLIVAR Exchanges*, *12*, 8–10, 2008.
- Baturin, N., et P. P. Niiler, Effects of instability waves in the mixed layer of the equatorial pacific, *J. Geophys Res.*, *102*, 27,771–27,793, 1997.

- Bearman, G. (Ed.), *Ocean circulation*, The Open University. Oceanography Course Team, 1989.
- Bennett, A., *Inverse methods in physical oceanography.*, Cambridge University Press, Cambridge, 1992.
- Bennett, A., *Inverse Modeling of the Ocean and Atmosphere.*, Cambridge University Press, Cambridge, 2002.
- Bergthorsson, P., et B. Döös, Numerical weather map analysis, *Tellus*, 5, 329–340, 1955.
- Blanke, B., et 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.
- Blayo, E., J. Verron, et J. M. Molines, Assimilation of topex/poseidon altimeter data into a circulation model of the north atlantic, *J. Geophys. Res.*, 99, 24,691–24,705, 1994.
- Bonekamp, H., G. van Oldenborgh, et G. Burges, Variational assimilation of tropical atmosphere-ocean and expendable bathythermograph data in the hamburg ocean primitive ocean general circulation model, adjusting the surface fluxes in the tropical ocean, *J. Geophys. Res.*, 106, 16,693–16,709, 2001.
- Brankart, J.-M., C. E. Testut, et L. Parent, An integrated system of sequential assimilation modules : sesam reference manual, *Tech. rep.*, Tech. Rep. Office Note. LEGI/MEOM, Grenoble, France, 2002.
- Brankart, J.-M., C.-E. Testut, P. Brasseur, et J. Verron, Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models : Application to a 1993-96 hindcast of the north atlantic ocean circulation, *J. Geophys. Res.*, 108, 1–20, 2003.
- Brankart, J.-M., C. Ubelmann, C.-E. Testut, E. Cosme, P. Brasseur, et J. Verron, Efficient parametrization of the observation error covariance matrix for square-root or ensemble kalman filters : application to ocean altimetry, *Mon. Wea. Rev.*, 137, 1908–1927, 2009.
- Brankart, J.-M., E. Cosme, C.-E. Testut, E. Cosme, P. Brasseur, et J. Verron, Efficient adaptive error parameterizations for square root or ensemble kalman filters : application to the control of ocean mesoscale signals, *Mon. Wea. Rev.*, 138(3), 932–950, 2010.
- Brasseur, P., et J. Verron, The seek filter method for data assimilation in oceanography : a synthesis, *Ocean Dynamics*, pp. doi :10.1007/s10,236–006–0080–3, 2006.
- Brodeau, L., B. Barnier, A. M. Tréguier, T. Penduff, et S. Gulev, An era40-based atmospheric forcing for global ocean circulation models, *Ocean Modelling*, 31, 88–104, 2010.
- Bryden, H., et E. C. Brady, Eddy momentum and heat fluxes and their effect on the circulation of the equatorial pacific ocean, *J. Mar. Res.*, 47, 55–79, 1989.
- Busalacchi, A., et J. Picaut, Seasonal variability from a model of the tropical atlantic ocean, *J. Phys. Oceanogr.*, 13, 1564–1588, 1983.

- Cailleau, S., Validation de méthodes de contrainte aux frontières d'un modèle océanique : application à un modèle hauturier de l'atlantique nord et à un modèle régional du golfe de gascogne, Ph.D. thesis, Université Joseph Fourier, 2004.
- Caltabiano, A. C. V., I. S. Robinson, et L. P. Pezzi, Multi-year satellite observations of instability waves in the tropical atlantic ocean, *Ocean Science*, *1*, 97–112, 2005.
- Cane, M. A., A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, et A. J. Busalacchi, Mapping tropical pacific sea level : Data assimilation via a reduced state space kalman filter, *J. Geophys. Res.*, *101*, 22,599–22,617, 1996.
- Castruccio, F., Apports des données gravimétriques grace pour l'assimilation de données altimétriques et in-situ dans un modèle de l'océan pacifique tropical, Ph.D. thesis, Université Joseph Fourier, 2006.
- Chang, P., R. Saravanan, L. Ji, et G. C. Hegerl, The effects of local sea surface temperatures on atmospheric circulation over the tropical atlantic sector, *J. Climate*, *13*, 2195–2216, 2000.
- Chassignet, E. P., et J. Verron, *Ocean modelling and parametrization*, vol. 516 of *Mathematical and Physical Sciences*, Kluwer Academic Publishers, The Netherlands, 1998.
- Cohn, S. E., N. S. Sivakumaran, et R. Todling, A fixed-lag kalman smoother for retrospective data assimilation, *Mon. Wea. Rev.*, *122*, 2838–2867, 1994.
- Cosme, E., J.-M. Brankart, J. Verron, P. Brasseur, et M. Krysta, Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation, *Ocean Modelling*, *33*, 87–100, 2010.
- Cosme, E., J. Verron, P. Brasseur, J. Blum, et D. Auroux, Smoothing problems in a bayesian framework and their linear gaussian solutions, *Month. Weath. Rev.*, in press.
- Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, et M. Fisher., The ecmwf implementation of the three-dimensional variational assimilation (3d-var), i : Formulation., *Q. R. J. Meteorology Society*, *124*, 1783–1807, 1998.
- Covey, D. L., et S. Hastenrath, The pacific el niño phenomenon and the atlantic circulation, *Mon. Wea. Rev.*, *106*, 1280–1287, 1978.
- Cressman, G. P., An operational objective analysis scheme, *Mon. Wea. Rev.*, *87*, 367–374, 1959.
- Desroziers, G., L. Berre, B. Chapnik, et P. Poli, Diagnosis of observation, background and analysis-error statistics in observation space, *Q. J. R. Meteorol. Soc.*, *131*, 3385–3396, 2005.
- Didden, N., et F. Schott, Eddies in the north brazil current retroflection region observed by geosat altimetry, *J. Geophys. Res.*, *98*, 121–131, 1993.
- Douglas, B. C., et R. E. Cheney, Geosat : Beginning a new era in satellite oceanography, *J. Geophys. Res.*, *95*, 2833–2836, 1990.

- Du Penhoat, Y., et A. M. Tréguier, The seasonal linear response of the tropical atlantic ocean, *J. Phys. Oceanogr.*, *15*, 316–329, 1985.
- Dussin, R., A. Treguier, J. Molines, B. Barnier, T. Penduff, L. Brodeau, et G. Madec, Definition of the interannual experiment orca025-b83, *Tech. rep.*, 2009.
- Eliassen, G., Provisional report on calculation of spatial covariance and autocorrelation of the pressure field, *Tech. rep.*, Report no 5, Videnskaps-Akademiets Institut for Vaer-Og Klimaforskning, Oslo, Norway, 12 pp., 1954.
- Evensen, G., Sequential data assimilation with a non linear quasigeostrophic model using monte carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*,(C5), 10,143–10,162, 1994.
- Evensen, G., The ensemble kalman filter : Theoretical formulation and practical implementation, *Ocean Dynamics*, *53*(4), 343–367, 2003.
- Evensen, G., et P. J. van Leeuwen, An ensemble kalman smoother for nonlinear dynamics, *Mon. Wea. Rev.*, *128*, 1852–1867, 2000.
- Fraser, D. C., et J. E. Potter, The optimum linear smoother as a combination of two optimum linear filters, *IEEE Trans. Autom. Control*, *14*, 387–390, 1969.
- Fratantoni, D. M., et D. A. Glickson, North brazil current ring generation and evolution observed with seawifs, *J. Phys. Oceanogr.*, *32*, 1058–1074, 2002.
- Fratantoni, D. M., W. E. Johns, et T. L. Townsend, Rings of the north brazil current : Their structure and behavior inferred from observations and a numerical simulation, *J. Geophys. Res.*, *10*, 633–654, 1995.
- Gandin, L. S., Objective analysis of meteorological fields, *Gidrometeor. Izd., Leningrad (in Russian)*, 1963.
- Garraffo, Z. D., W. E. Johns, E. P. Chassignet, et G. J. Goni, North brazil current rings and transport of southern waters in a high resolution numerical simulation of the north atlantic, *Elsevier B. V.*, *68*, 375–409, 2003.
- Garzoli, S., et E. Katz, The forced annual reversal of the north atlantic equatorial counter-current, *J. Phys. Oceanogr.*, *13*, 2082–2090, 1993.
- Gent, P. R., et J. C. McWilliams, Isopycnal mixing in ocean circulation models, *J. Phys. Oceanogr.*, *20*, 150–155, 1990.
- Giannini, A., Y. Kushnir, et M. A. Cane, Interannual variability of caribbean rainfall, enso, and the atlantic ocean, *J. Climate*, *13*, 297–311, 2000.
- Goni, G., et W. Johns, A census of north brazil current rings observed from t/p altimetry : 1992-1998, *Geophys. Res. Lett.*, *28*, 1–4, 2001.
- Gourdeau, L., Assimilation par filtrage de kalman de données altimétriques dans un modèle linéaire de l’océan atlantique tropical, Ph.D. thesis, Université de Paris 06, Paris, FRANCE, 1991.

- Gouriou, Y., C. Andrie, B. Bourles, S. Freudenthal, S. Arnault, A. Aman, G. Eldin, Y. D. Penhoat, F. Baurand, F. Gallois, et R. Chuchla, Deep circulation in the equatorial atlantic ocean, *Geophysical Research Letters*, *28*(5), 819–822, 2001.
- Grodsky, S. A., J. A. Carton, C. Provost, et J. Servain, Tropical instability waves at 0n, 23w in the atlantic : A case study using pilot research moored array in the tropical atlantic (pirata) mooring data, *J. Geophys. Res.*, *110*, C08,010.1–C08,010.12, 2005.
- Halpern, D., R. A. Knox, et D. S. Luther, Observation of 20-day period meridional current oscillations in the upper ocean along the pacific equator, *J. Phys. Oceanogr.*, *18*, 1514–1534, 1988.
- Hamill, T. M., J. S. Whitaker, et C. Snyder, Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter, *Mon. Wea. Rev.*, *129*, 2776–2790, 2001.
- Hansen, D., et C. Paul, Genesis and the effect of long waves in the equatorial pacific, *J. Geophys. Res.*, *89*, 10,431–10,440, 1984.
- Hayes, S. P., M. J. McPhaden, et J. M. Wallace, The influence of sea-surface temperature on surface wind in the eastern equatorial pacific - weekly to monthly variability, *J. Clim.*, *2*, 1500–1506, 1989.
- Hisard, P., Observation de réponse du type "el niño" dans l'atlantique tropical oriental - golfe de guinée, *Ocean. Acta.*, *3*, 69–78, 1980.
- Hoke, J., et R. Anthes, The initialization of numerical models by a dynamic initialization technique, *Monthly Weather Review*, *104*, 1551–1556, 1976.
- Hoteit, I., D.-T. Pham, et J. Blum, A semi-evolutive partially local filter for data assimilation, *Marine Pollution Bulletin*, *43*, 164–174, 2001.
- Houghton, R., Influence of local and remote wind forcing in the gulf of guinea, *J. Phys. Oceanogr.*, *94*, 4816–4828, 1989.
- Houtekamer, P. L., et H. L. Mitchell, Data assimilation using an ensemble kalman filter technique, *Mon. Wea. Rev.*, *126*, 796–811, 1998.
- Houtekamer, P. L., et H. L. Mitchell, A sequential ensemble kalman filter for atmospheric data assimilation, *Mon. Wea. Rev.*, *129*, 123–137, 2001.
- Hunt, B. R., E. J. Kostelich, et I. Szunyogh, Efficient data assimilation for spatiotemporal chaos : a local ensemble transform kalman filter, *Physica D*, *230*, 112–126, 2007.
- Ihler, A. T., S. Kirshner, M. Ghil, A. W. Ribertson, et P. Smyth, Graphical models for statistical inference and data assimilation, *Physica D*, *230*(1-2), 72–87, 2007.
- Jochum, M., P. Malanotte-Rizzoli, et A. Busalacchi, Tropical instability waves in the atlantic ocean, *Ocean Modelling*, *7*, 145–163, 2004.
- Johns, W., T. Lee, F. Schott, R. Zantopp, et R. Evans, The north brazil current retroflection : Seasonal structure and eddy variability, *J. Geophys. Res.*, *95*, 22,103–22,120, 1990.

- Juza, M., Modélisation numérique et observations de l'océan global : développement des interfaces, évaluation de simulations et de réseaux d'observations, investigations dynamiques., Ph.D. thesis, Université Joseph Fourier - Grenoble I, 2011.
- Kalman, R. E., A new approach to linear filter and prediction problems, *J. Basic. Eng.*, *82*, 35–45, 1960.
- Khare, S. P., J. L. Anderson, T. J. Hoar, et D. Nychka, An investigation into the application of an ensemble kalman smoother to high-dimensional geophysical systems, *Tellus A*, *60*, 97–112, 2008.
- Langlais, C., Etude de la variabilité interannuelle des échanges côte-large : simulation haute résolution de la dynamique du golfe du lion, Ph.D. thesis, Université du Sud Toulon Var, 2007.
- Large, W. G., G. Danabasoglu, S. C. Doney, et J. C. McWilliams, Sensitivity to surface forcing and boundary layer in a global ocean model : Annual-mean climatology, *J. Phys. Oceanogr.*, *27*, 2418–2447, 1997.
- Lecointre, A., Variabilité interannuelle à décennale et atlantique nord et mers nordiques, études conjointe d'observations, de simulation numériques et de réanalyses, Ph.D. thesis, Université Joseph Fourier, 2009.
- Legeckis, R., Long waves in the eastern equatorial pacific ocean : a view from a geostationary satellite, *Science*, *197*, 1179–1181, 1977.
- Lermusiaux, P. F. J., Data assimilation via error subspace statistical estimation. part ii : middle atlantic bight shelfbreak front simulation and esse validation, *Mon. Wea. Rev.*, *127*, 1408–1432, 1999a.
- Lermusiaux, P. F. J., Estimation and study of mesoscale variability in the strait of sicily, *Dynamics of Atmospheres and Oceans*, *29*, 255–303, 1999b.
- Lermusiaux, P. F. J., et A. R. Robinson, Data assimilation via error subspace statistical estimation. part i : Theory and schemes, *Mon. Wea. Rev.*, *127*, 1385–1407, 1999.
- Lermusiaux, P. F. J., A. R. Robinson, P. J. H. Haley, et W. G. Leslie, Advanced interdisciplinary data assimilation : filtering and smoothing via error subspace statistical estimation., in *The OCEANS*, edited by H. P. MTS/IEEE Conference. IEEE, pp. pp. 795–802, 2002.
- Lions, J. L., *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles.*, Dunod, Gauthier-Villars, Paris, 1968, 426 pp.
- Lions, J. L., O. P. Mankey, R. Temam, et S. Wang, Physical interpretation of the attractor dimension for the primitive equations of atmospheric circulation, *J. Atmo. Sci.*, *54*, 1137–1143, 1997.
- Luther, J. M., et E. S. Johnson, Eddy energetics in the upper equatorial pacific during the hawaii-to-tahiti shuttle experiment, *J. Phys. Oceanogr.*, *20*, 913–944, 1990.

- Madec, G., P. Delecluse, M. Imbard, et C. Levy, Opa 8.1 general circulation model reference manual, *Tech. Rep. 11*, Notes de l'IPSL, Université P. et M. Curie, B102 T15-E5, 4 place Jussieu, Paris cedex 5, France, 1998.
- Masina, S., Instabilities of continuously stratified zonal equatorial jets in a periodic channel model, *Annal. Geophys*, *20*, 729–740, 2002.
- Masina, S., et S. G. H. Philander, An analysis of tropical instability waves in a numerical model of the pacific ocean - 1. spatial variability of the waves, *J. Geophys. Res.-Oceans*, *104*, 29,613–29,635, 1999.
- Masina, S., S. G. H. Philander, et A. Bush, An analysis of tropical instability waves in a numerical model of the pacific ocean, 2. generation and energetics of the waves, *J. Geophys. Res.-Oceans*, *104*, 29,637–29,661, 1999.
- McPhaden, M. J., Toga-tao and the 1991-93 el niño-southern oscillation event, *Oceanography*, *6*, 36–44, 1993.
- McPhaden, M. J., A. J. Busalacchi, R. Cheney, J.-R. Donguy, K. S. Gage, D. Halpern, M. Ji, P. Julian, G. Meyers, G. T. Mitchum, P. P. Niiler, J. Picaut, R. W. Reynolds, N. Smith, et K. Takeuchi, The tropical ocean-global atmosphere observing system : A decade of progress, *J. Geophys. Res.*, *103*, 14,169–14,240, 1998.
- Merle, J., Variabilité thermique et interannuelle de l'océan atlantique équatorial est. l'hypothèse d'un "el niño" atlantique, *Oceanol. Acta.*, *3*, 209–220, 1980.
- Mestas-Nunez, A. M., et D. B. Enfield, Eastern equatorial pacific sst variability : Enso and non-enso components and their climatic associations, *J. Climate*, *14*, 391–402, 2001.
- Oke, P. R., et P. Sakov, Representation error of oceanic observations for data assimilation, *Journal of Atmospheric and Oceanic Technology*, *25*, 1004–1017, 2008.
- Okumura, Y., S.-P. Xie, A. Numaguti, et Y. Tanimoto, Tropical atlantic air-sea interaction and its influence on the nao, *Geophys. Res. Lett.*, *28*, 1507–1510, 2001.
- Patil, D. J., B. Hunt, E. Kalnay, J. A. Yorke, et E. Otte, Local low dimensionality at atmospheric dynamics, *Phys. Rev. Lett.*, *86*, 5878–5881, 2001.
- Pham, D. T., J. Verron, et M. C. Roubaud, A singular evolutive extended kalman filter for data assimilation in oceanography, *J. Mar. Sys.*, *16*, 323–340, 1998.
- Philander, S., et R. C. Pacanowski, The mass and heat budget in a model of the tropical atlantic ocean, *J. Phys. Oceanogr*, *14*, 212–220, 1986.
- Philander, S. G. H., Unusual conditions in the tropical atlantic ocean in 1984, *Nature*, *222*, 236–238, 1986.
- Philander, S. G. H., W. J. Hurlin, et R. C. Pacanowski, Properties of long equatorial waves in models of the seasonal cycle in the tropical atlantic and pacific oceans, *J. Geophys. Res.-Oceans*, *91*, 14,207–14,211, 1986.

- Pinardi, N., K. Miyakoda, A. Rosati, et R. Gudgel, A global ocean assimilation system for hydrographic and satellite altimeter data., in *The Global Ocean*, vol. 1, Proceedings of Oceanology International 1994, 1995.
- Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Uden, A. Hollingsworth, et F. Bouttier., The ecmwf implementation of the three-dimensional variational assimilation (3d-var), ii : Structure functions., *Q. J. R. Meteorol. Soc.*, *124*, 1809–1829, 1998.
- Radon, J., Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten, berichte sächsische akademie der wissenschaften, leipzig, *Math.-Phy*, *69*, 262–267, 1917.
- Rauch, H. E., F. Tung, et C. T. Striebel, Maximum likelihood estimates of linear dynamic systems, *AIAA Journal*, *3(8)*, 1445–1450, 1965.
- Ravela, S., K. Emanuel, et D. McLaughlin, Data assimilation by field alignment, *Physica D : Nonlinear Phenomena, Volume 230, Issues 1-2*, 127–145, 2007.
- Richardson, P., et D. Walsh, Mapping climatological seasonal variations of surface currents in the tropical atlantic using ship drifts, *J. Geophys. Res.*, *91*, 10,537–10,550, 1986.
- Richardson, P. L., G. Hufford, R. Limeburner, et W. Brown, North brazil current retroflection eddies, *J. Geophys. Res.*, *99*, 5081–5093, 1994.
- Richman, J. G., R. N. Miller, et Y. H. Spitz, Error estimates for assimilation of satellite sea surface temperature data in ocean climate models, *Geophys. Res. Lett.*, *32*, L18,608, doi :10.1029/2005GL023,591, 2005.
- Roulet, G., et G. Madec, Salt conservation, free surface, and varying levels : a new formulation for ocean general circulation models, *J. Geophys. Res.*, *105*, 23,927–23,942, 2000.
- Rozier, D., E. Cosme, F. Birol, P. Brasseur, J.-M. Brankart, et J. Verron, A reduced-order kalman filter for data assimilation in physical oceanography, *SIAM Review*, *49(3)*, 449–465, 2007.
- Sasaki, Y., A fundamental study of the numerical prediction based on the variational principle., *Journal of the Meteorological Society of Japan*, *33*, 262–275, 1955.
- Sasaki, Y., An objective analysis based on the variational method., *Journal of the Meteorological Society of Japan*, *II(36)*, 77–88, 1958.
- Servain, J., A. Busalacchi, A. Moura, M. McPhaden, G. Reverdin, M. Vianna, et S. Zebiak, Pilot research moored array in the tropical atlantic "pirata", *The Bulletin of American Meteorological Society*, *79*, 2019–2031, 1998.
- Simon, D., *Optimal State Estimation*, Wiley & sons, 2006.
- Skachko, S., J.-M. Brankart, F. Castruccio, P. Brasseur, et J. Verron, Improved turbulent air-sea flux bulk parameters for the control of the ocean mixed layer : a sequential data assimilation approach., *Journal of Atmospheric and Oceanic Technologies*, *26(3)*, 538–555, 2009.

- Stammer, D., K. Ueyoshi, A. Köhl, W. G. Large, S. A. Josey, et C. Wunsch, Estimating air-sea fluxes of heat, freshwater, and momentum through global ocean data assimilation, *J. Geophys. Res.*, *109*, C05,023, doi :10.1029/2003JC002,082, 2004.
- Stramma, L., et F. Schott, The mean flow field of the tropical atlantic ocean, *Deep-Sea Research II*, *46*, 279–303, 1999.
- Sutton, R. T., S. P. Jewson, et D. P. Rowell, The elements of climate variability in the tropical atlantic region, *J. Climate*, *13*, 3261–3284, 2000.
- Swenson, M. S., et D. V. Hansen, Tropical pacific ocean mixed layer heat budget : The pacific cold tongue, *J. Phys. Oceanogr.*, *29*, 69–81, 1999.
- Terray, L., et C. Cassou, Tropical atlantic sea surface temperature forcing of quasi-decadal climate variability over the north atlantic-european region, *J. Climate*, *15*, 3170–3187, 2002.
- Testut, C., P. Brasseur, J. Brankart, et J. Verron, Assimilation of sea-surface temperature and altimetric observations during 1992–1993 into an eddy permitting primitive equation model of the north atlantic ocean, *J. Mar. Sys.*, *40–41*, 291–316, 2003.
- Testut, C. E., Assimilation de données satellites avec un filtre de kalman de rang réduit dans un modèle aux équations primitives de l’océan atlantique, Ph.D. thesis, Université Joseph Fourier, 2000.
- Ubelmann, C., Etude de scénarios d’altimétrie satellitaire pour le contrôle de la circulation océanique dans l’océan atlantique tropical par assimilation de données, Ph.D. thesis, Université Joseph Fourier, 2009.
- Ubelmann, C., J. Verron, J. Brankart, P. Brasseur, et E. Cosme, Impact of data from upcoming altimetric missions on the control of the three dimensional circulation in the tropical atlantic ocean, *J. Operational Oceanography*, *2*, 2009.
- van Leeuwen, P. J., The time mean circulation in the agulhas region determined with the ensemble smoother, *J. Geophys. Res.*, *104*, 1393–1404, 1999.
- van Leeuwen, P. J., Data assimilation and inverse methods in terms of probabilistic formulation, *Mon. Wea. Rev.*, *124*, 2898–2913, 2001.
- van Leeuwen, P. J., et G. Evensen, Data assimilation and inverse methods in terms of a probabilistic formulation, *Monthly Weather Review*, *124*, 2898–2913, 1996.
- Verlaan, M., et A. W. Heemink, Tidal flow forecasting using reduced-rank square root filter, *Stochastic Hydrology and Hydraulics*, *11*, 349–368, 1997.
- Verron, J., et W. R. Holland, Impacts de données d’altimétrie satellitaire sur les simulations numériques des circulations océaniques aux latitudes moyennes, *Annales Geophysicae*, *71*, 31–46, 1989.
- Verron, J., J. M. Molines, et E. Blayo, Assimilation of geosat data into a quasigeostrophic model of the north atlantic between 20⁰ n and 50⁰ n : preliminary results, *Oceanologica Acta*, *15*, 575–583, 1992.

- Vialard, J., C. Menkes, J.-P. Boulanger, P. Delecluse, E. Guilyardi, M. J. McPhaden, et G. Madec, A model study of oceanic mechanisms affecting equatorial pacific sea surface temperature during 1997-98 el niño, *J. Phys. Oceanogr.*, *31*, 1649–1675, 2001.
- Wang, W., et M. J. McPhaden, The surface layer heat balance in the equatorial pacific ocean, part i : Mean seasonal cycle, *J. Phys. Oceanogr.*, *29*, 1812–1831, 1999.
- Weingartner, T. J., et R. H. Weisberg, A description of the annual cycle in sea-surface temperature and upper ocean heat in the equatorial atlantic, *J. Phys. Oceanogr.*, *21*, 83–96, 1991.
- Wikle, C. K., et L. M. Berliner, A bayesian tutorial for data assimilation, *Physica D*, *230*, 1–16, 2007.
- Xie, S.-P., et J. A. Carton, Tropical atlantic variability : patterns, mechanisms and impacts, *AGU Geogr. Monogr. Ser.*, *147*, 121–142, 2004.

Table des figures

1	L'avenir de Nicolas Freychet tel qu'il est prédit par différents experts : cas I : "Le Trièves : la harpe et la perruque blonde" par Quiquet, 2010 ; cas II : "Le Trièves bis : la poterie au bord de la route" par Quiquet, 2011 ; cas III : "Témoignage chez Delarue : retour aux origines" par Ronie-be, 2010 ; cas IV : "Nikeula Freycher : un expert peebler comme on n'en fait plus" par Varin, 2012, basé sur une étude de Buchonnet, 2009 et antérieur.	4
1.1	Représentation de la bathymétrie dans l'océan Atlantique. Le rectangle noir indique la zone définie comme étant l'Atlantique tropical. Les rectangles rouge et jaune indiquent respectivement les zones définies comme la région des TIWs et la région des anneaux du Brésil.	11
1.2	Figure d'après <i>Stramma et Schott</i> (1999). Circulation moyenne dans les 100 premiers mètres, au printemps (a) et à l'automne (b).	13
1.3	Figures d'après <i>Jochum et al.</i> (2004). Image de la couleur de l'eau prise par le capteur SeaWiFS où l'on peut distinguer le passage des TIWs (par le contraste entre les zones vertes et bleues).	15
1.4	Figures d'après <i>Caltabiano et al.</i> (2005). Diagrammes de Hovmöller de la température de surface issue des données du satellite TMI (Tropical Rainfall Measuring Mission Microwaves Imager), entre janvier 1998 et décembre 2001, tracés (de gauche à droite) : à l'équateur, à 1°N, à 2°N, à 3°N, à 4°N et à 5°N.	16
1.5	Figures d'après <i>Fratantoni et Glickson</i> (2002). a) Schématisation des principaux courants et de la formation des anneaux du Brésil. b) Image de la chlorophylle prise par le satellite SeaWiFS où l'on peut distinguer un anneau.	17
2.1	Figures d'après <i>Juza</i> (2011). Evolution du nombre de profils <i>in situ</i> annuels disponibles pour différents types d'observations entre 1958 et 2007.	21
2.2	Figure tirée du site http://www.pmel.noaa.gov/tao/global/global.html . Représentation des différents réseaux de mouillages dans la bande tropicale.	23
2.3	Figure tirée du site http://www.argo.net . Représentation de la répartition des flotteurs du réseau ARGO qui ont émis un signal dans les 30 derniers jours au 5 mai 2011.	24
2.4	Figure d'après <i>Castruccio</i> (2006). Principe de la mesure altimétrique et décomposition des différentes grandeurs liées à cette mesure. Notations : SSH (Sea Surface Height), MSSH (Mean Sea Surface Height), DT (Dynamic Topography), MDT (Mean Dynamic Topography), SLA (Sea Level Anomaly), R (distance entre le satellite et la surface de l'océan), H (distance entre le satellite et l'ellipsoïde de référence). . . .	27

3.1	Figure d'après <i>Castruccio</i> (2006). Disposition des variables du modèle NEMO sur la grille C. T indique le point où les grandeurs scalaires sont calculées. u , v et w indiquent les points de calcul des grandeurs vectorielles (vitesses).	39
3.2	Cartes de la température de surface instantanée en °C (haut) et de la SSH instantanée en m (bas) au 25 juillet 2005 pour le run libre de référence dans la configuration TATL4.	42
3.3	Cartes de la vitesse absolue instantanée en $m.s^{-1}$: à 30 m de profondeur (haut) et moyennée entre 1500 et 2500 m (bas) au 25 juillet 2005 pour le run libre de référence dans la configuration TATL4. Le trait noir sur la figure du haut indique la position des coupes tracées sur la figure 3.5.	44
3.4	Profils de température moyennée horizontalement sur tout le bassin, pour la configuration TATL4 (trait plein) et le domaine équivalent dans B83 (en pointillés). Les profils sont moyennés sur les années 2001 (à gauche), 2003 (au milieu) et 2005 (à droite).	45
3.5	Coupes de la vitesse absolue instantanée en $m.s^{-1}$ à 5°N : entre la surface et 250 m de profondeur (gauche) et entre 500 et 5000 m de profondeur (droite) au 25 juillet 2005 pour le run libre de référence dans la configuration TATL4. La position des coupes est rappelée sur la figure 3.3 (bas).	46
3.6	Représentation des vecteurs vitesses de sub-surface (entre 10m et 110m) moyennées sur le printemps 2005 (à gauche) et l'automne 2005 (à droite).	46
3.7	Evolution de la SSH en m (en haut) et de la température en °C (en bas) pour le run effectué avec des forçages en vent haute fréquent (en trait plein) et celui effectué avec des forçages lissés sur 5 jours (en pointillés). La SSH est donnée en un point et la température est moyennée sur tout le domaine.	48
4.1	Schématisation des différentes représentations d'une observation dans l'espace et le temps. Pour plus de lisibilité la dimension spatiale est représentée sur un seul axe mais il s'agit en réalité d'une dimension 3D.	55
4.2	Représentation schématique du lissage séquentiel de type <i>fixed-interval</i> ou <i>fixed-lag</i> . Les notations sont : y^o (triangles) pour les observations, x^a (ronds) pour les états analysés par le filtre, x^s (ronds) pour les états analysés par le lisseur et x^f (ronds noirs) pour les états prédits par le modèle. L'axe des abscisses représente différentes dates, avec k la date de la dernière observation disponible. Etape 0 : au temps $k - 2$ un état modèle de background et une observation y^o (en vert) sont disponibles; Etape 1 : le filtre de Kalman détermine un état analysé x^a au temps $k - 2$; Etape 2 : l'état analysé précédemment, ainsi que les statistiques issues de l'analyse, sont propagés par le modèle jusqu'au temps $k - 1$ où une nouvelle observation y^o (en bleu) est disponible; Etape 3 : le filtre effectue une analyse au temps $k - 1$, et le lisseur peut utiliser l'observation y^o (en bleu) pour réanalyser l'état x^a au temps $k - 2$; Etape 4 : l'état analysé par le filtre au temps $k - 1$, ainsi que les statistiques issues de l'analyse, sont propagés par le modèle jusqu'au temps k où une nouvelle observation y^o (en rouge) est disponible; Etape 5 : le filtre effectue une analyse au temps k , et le lisseur peut utiliser l'observation y^o (en rouge) pour réanalyser les états x^a et x^{s1} respectivement aux temps $k - 1$ et $k - 2$	59

- 4.3 Représentation schématique de l'*ensemble smoother*, dans une version séquentielle. Les notations sont les mêmes que pour la figure 4.2 excepté que tous les états analysés sont notés x^a . **Etape 1** : une expérience d'ensemble (sans assimilation) est menée sur toute la fenêtre temporelle de $k - 2$ à k ; **Etape 2** : la première observation disponible au temps $k - 2$ est utilisée pour analyser les états sur toute la durée de l'expérience (les états x^a sont donc des états issus du lissage et non pas du filtrage); **Etape 3** : la deuxième observation disponible au temps $k - 1$ est à son tour utilisée pour analyser les états sur toute la durée de l'expérience; **Etape 4** : les étapes d'analyses sont répétées jusqu'à la dernière observation disponible (au temps k). 61
- 4.4 Représentation schématique du *forward-backward smoother*. Les notations sont les mêmes que pour la figure 4.2. **Etape 1** : un filtrage est effectué sur toute la fenêtre temporelle, du temps $k - 2$ jusqu'à la dernière observation disponible au temps k ; **Etape 2** : le dernier état analysé par le filtre peut être considéré comme le premier état lissé ($x^a = x^s$ au temps k), le lisseur rétropropage alors l'information extraite de la dernière analyse (d'après les états prédit x^f et analysé x^s) et l'utilise pour réanalyser l'état filtré précédent (en bleu), donnant ainsi x^s au temps $k - 1$; **Etape 3** : le lissage rétropropage de nouveau l'information en utilisant les états prédit x^f et lissé x^s au temps $k - 1$ pour réanalyser l'état filtré précédent (en vert), donnant ainsi x^s au temps $k - 2$ 62
- 4.5 Représentation schématique de la *two-filter smoother*. Les notations sont les mêmes que pour la figure 4.2, avec en plus les états \mathbf{z} correspondant au filtrage rétrograde. **Etape 1** : un filtrage est effectué sur toute la fenêtre temporelle, du temps $k - 2$ jusqu'à la dernière observation disponible au temps k ; **Etape 2** : au temps k , on utilise $\mathbf{P} = +\infty$ (non informatif), et l'on définit $\mathbf{N} = \mathbf{P}^{-1}$ et $\mathbf{z} = \mathbf{P}^{-1}x$, qui permettent d'initialiser le filtre rétrograde avec $\mathbf{N} = 0$ et $\mathbf{z} = 0$; d'après un état z^f contenant une information non-descriptive ou nulle; **Etape 3** : la première analyse du filtre rétrograde (z^a) est effectuée d'après l'état z^f et l'observation y^o au temps k ; **Etape 4** : l'état z^a est propagé par le modèle adjoint jusqu'au temps précédent $k - 1$, donnant ainsi z^f ; **Etape 5** : une analyse est effectuée par le filtre rétrograde d'après l'état analysé par le filtre direct au temps $k - 1$ et l'état z^f au même temps, donnant ainsi un état lissé a^s ; **Etape 6** : le même état z^f est également analysé directement avec les observations du temps $k - 1$, donnant un nouvel état analysé du filtre rétrograde z^a ; **Etape 7** : cet état z^a est de nouveau propagé par le modèle adjoint jusqu'au temps $k - 2$; **Etape 8** : de nouveau, des analyses permettent d'obtenir un état lissé x^s et un état filtré z^a au temps $k - 2$. 63
- 4.6 Récapitulatif des applications privilégiées propres à chaque lisseur. 65
- 5.1 Schématisation du principe de filtrage séquentiel (en noir) et du lissage (en rouge), compte-tenu de divers observations (croix bleues) disponibles aux dates t_i . Les points x^a désignent les analyses du filtre, les points x^f désignent les prévisions, les points x^{si} désigne les i^{eme} analyses rétrospectives du lisseur. 74

5.2	Schématisation des différents types de filtrage en fonction de la propagation où non des covariances d'erreur : a) pleinement propagée ($\mathbf{S}^f = \mathbf{M}\mathbf{S}^a\mathbf{M}^T + \mathbf{Q}$) ; b) non propagée explicitement par le modèle mais l'information est tout de même prise en compte d'une étape à l'autre ($\mathbf{S}^f = \mathbf{S}^a + \mathbf{Q}$) ; c) non propagée, les covariances d'erreur de prévision sont toujours identiques et égale à la matrice de covariances initiale ($\mathbf{S}^f = \mathbf{S}_0$).	77
5.3	Schématisation du principe de localisation de l'analyse. Les points noirs indiquent les lieux des analyses du filtre, les croix indiquent la position des différentes observations disponibles et les cercles pointillés gris indiquent la limite des zones dans lesquelles les observations sont prises en compte pour les analyses.	80
6.1	Représentation schématique du principe de la transformée de Radon.	90
6.2	Profils de vitesse (en abscisse, en cm.s^{-1}) calculés par transformée de Radon sur la SSH entre les jours 20170 et 20435 en fonction de la latitude (en ordonnée, en $^{\circ}\text{N}$). Les dénominations ES1 à ES6 font référence aux runs perturbés (voir tableau 6.1). Les courbes bleues indiquent les vitesses calculées pour chaque run perturbé, et les courbes pointillées rouges indiquent le profil des vitesses dans le run de référence.	91
6.3	Diagrammes de Hovmöller sur la SLA (en m) pour les différents runs perturbés (voir tableau 6.1).	92
6.4	Cartes de température de surface (en $^{\circ}\text{C}$) instantannée au jour 20351 (4 octobre 2005) pour les différents runs perturbés (en couleur), et pour le run de référence (en contours).	93
6.5	Cartes de température de surface (en $^{\circ}\text{C}$) instantannée au jour 20435 (27 décembre 2005) pour les différents runs perturbés (en couleur), et pour le run de référence (en contours).	94
6.6	Coupes à l'équateur des différences de vitesses zonales (en m.s^{-1}) au jour 20435 (27 décembre 2005) entre le run de référence et les différents runs perturbés.	95
6.7	Densités de répartition des valeurs tirées aléatoirement pour chaque paramètre de diffusion. Le total des tirages de chaque paramètre est de 100. L'axe des abscisse indique la valeur des tirages (et le facteur d'échelle associé), l'axe des ordonnées indique le nombre de tirages par tranches de $\frac{1}{3}$. Les croix rouges indiquent la position de la valeur de référence de chaque paramètre.	97
6.8	Représentation des valeurs tirées aléatoirement de chaque paramètre de diffusion en fonction des autres : A_0^{vm} en fonction de A^{vm} (en haut à gauche), A^{hm} en fonction de A^{vm} (en haut à droite), A_0^{vm} en fonction de A^{hm} (en bas à gauche), A^{hm} en fonction de A^{vm} et de A_0^{vm} (en bas à droite). Les croix rouges (x) indiquent la position des valeurs de référence de chaque couple ou triplet de paramètres. Les + indiquent la position des couples extrêmes.	98
6.9	Ecart RMS global entre les runs d'ensemble et le run de référence (en vert), sur la température en $^{\circ}\text{C}$ (à gauche) et la vitesse zonale en m.s^{-1} (à droite). La courbe noire indique l'écart RMS entre le run perturbé à corriger et le run de référence.	99
6.10	Représentation de la répartition de la variance (en pourcentage de la variance totale) sur une décomposition en EOFs des runs d'ensemble.	99
6.11	Carte de la température de surface (en $^{\circ}\text{C}$) de la 1ere EOF calculée à partir de l'ensemble perturbé (à gauche) et d'un run interannuel (à droite).	100

7.1	Evolution du run interannuel REF de 1995 à 2005, chaque variable étant moyennée sur les 100 premiers mètres (de haut en bas : SSH (en m), température (en °C), salinité, vitesse zonale (en m.s ⁻¹), vitesse méridienne (en m.s ⁻¹)).	103
7.2	Schématisation de l'expérience d'assimilation. L'état de référence (année 2005) est indiquée en trait noir épais ; l'état perturbé par la condition initiale est indiqué par un trait bleu fin ; enfin l'état corrigé par assimilation est indiqué en trait rouge fin.	104
7.3	Evolution de l'écart RMS sur le champ 3D (sur toute la profondeur) entre le run faux et le run de référence durant les premiers mois du run libre faux (de haut en bas : SSH (en m), température (en °C), salinité, vitesse zonale (en m.s ⁻¹), vitesse méridienne (en m.s ⁻¹)).	106
7.4	Cartes de l'écart absolu en SSH (en m) entre le run faux et le run de référence au 26 Mai 2005 (a) et au 30 novembre 2005 (b).	107
7.5	Profils verticaux de l'écart RMS entre le run faux et le run de référence, moyenné horizontalement au 26 Mai 2005 (trait plein) et au 30 novembre 2005 (trait en pointillés) pour : a) la température (en °C), b) la salinité, c) la vitesse zonale (en m.s ⁻¹) et d) la vitesse méridienne (m.s ⁻¹).	108
7.6	Erreur RMS au 26 mai 2005 (en haut) et au 30 novembre 2005 (en bas) entre le run faux et le run de référence pour le champ de vitesse absolue, moyennée sur la vecticale entre 1000m et 4000m (à gauche) et en coupe E-W à 3°N au niveau de la côte du Brésil (à droite).	109
7.7	Pourcentage de la variance totale en fonction du nombre d'EOFs.	109
7.8	Représentation des modes de variabilité par les EOFs (à gauche : 1ère EOF, à droite : 25ème EOF, en haut : la SSH (en m), au milieu : la température (en °C), en bas : la vitesse zonale (en m.s ⁻¹)).	110
7.9	Représentation du système d'observation de SSH (en haut) et de T/S (en bas), pour la couverture sur 10 jours (à droite) ou la couverture partiellement disponible pour une analyse, c'est-à-dire tous les deux jours (à gauche).	111
7.10	Erreur RMS résiduelle sur le champ 3D après une analyse pour les différentes variables d'état, en fonction de la taille de la zone d'influence du filtre SEEK. Le point d'abscisse 0 correspond à l'erreur initiale avant l'analyse, et le point d'abscisse 150 correspond à une analyse globale.	113
7.11	Erreur RMS résiduelle sur le champ 3D après une analyse pour les différentes variables d'état, en fonction du nombre d'EOFs utilisées (la valeur à 0 EOF correspond à l'erreur initiale avant l'analyse).	113
7.12	Corrélation entre la vitesse méridienne en un point (symbolisé par le cercle noir) et la température sur le reste du bassin (en °C).	114
7.13	Ecart RMS sur les variables de SSH (en m), température (en °C) et vitesse zonale (en m.s ⁻¹), en fonction des différentes variables assimilées de façon indépendantes (SSH, température(TEM) et salinité(SAL)), dans le cas où l'on assimile un réseau d'observations partiel (haut) ou un champ d'observations 3D complet (bas). La valeur de l'écart RMS au point d'abscisse 0 indique l'erreur initiale avant l'analyse.	115
7.14	Evolution des erreurs RMS réelles (courbe rouge) et estimées par le filtre (courbe noire pleine) pour les expériences PeZL-A01 à PeZL-B02, sur les variables : SSH (en m), température (en °C) et vitesse zonale (en m.s ⁻¹). La courbe pointillée noire rappelle la trajectoire du run faux sans assimilation.	117

7.15	Evolution des erreurs RMS réelles (courbe rouge) et estimées par le filtre (courbe noire pleine) pour les expériences PeEOF01 à PeEOF03, sur les variables : SSH (en m), température (en °C) et vitesse zonale (en m.s ⁻¹). La courbe pointillée noire rappelle la trajectoire du run faux sans assimilation.	118
7.16	Evolution des erreurs RMS réelles (courbe rouge) et estimées par le filtre (courbe noire pleine) pour les expériences PeOb01 à PeOb05, sur les variables : SSH (en m), température (en °C) et vitesse zonale (en m.s ⁻¹). La courbe pointillée noire rappelle la trajectoire du run faux sans assimilation.	120
8.1	Coupes verticales de l'écart RMS entre le run faux et le run de référence, moyenné horizontalement au 1er jour de l'expérience d'assimilation (20252) et après 50 jours (20300), sur tout le bassin (trait plein), dans la région des Anneaux du Brésil (trait en pointillés) et dans la région des TIWs (trait tireté) pour : la température (en °C), la salinité, la vitesse absolue (en m.s ⁻¹).	127
8.2	Cartes de l'écart RMS entre le run faux et le run de référence, moyenné verticalement sur les 200 premiers mètres au 25 Juin 2005 pour : la SSH (en m), la température (en °C), la salinité et la vitesse zonale (en m.s ⁻¹).	128
8.3	Cartes des champs de température en °C (en haut) et de vitesse absolue (en m.s ⁻¹) à 30m de profondeur (en bas), pour le run de référence (à gauche) et le run perturbé (à droite) après 50 jours de propagation en run libre.	128
8.4	Représentation de la répartition temporelle des observations dans le cadre d'un réseau temps-réel en noir et dans le cadre d'un réseau temps-groupé (en rouge). . .	130
8.5	Evolution de l'erreur RMS 3D estimée par le filtre (trait plein noir), pour les états analysés par filtre (trait plein rouge) et analysés par le lisseur avec 3 analyses retrospectives (croix vertes) et 5 analyses retrospectives (croix bleues), pour la température (en °C) et la vitesse zonale (en m.s ⁻¹). L'erreur RMS du run faux est rappelée en trait pointillé noir.	131
8.6	Profils verticaux de l'écart RMS par rapport au run de référence moyenné horizontalement sur tout le bassin au 1er jour de l'expérience d'assimilation (20252) et au dernier jour (20300) pour : le run faux (trait plein), l'état analysé par le filtre (trait en pointillés), l'état analysé par le lisseur avec 3 analyses rétrospectives (trait en tiretés), l'état analysé par le lisseur avec 5 analyses rétrospectives (trait en point-tiretés). Les variables représentées sont : la température (en °C), la salinité, la vitesse absolue (en m.s ⁻¹).	131
8.7	Profils verticaux de l'écart RMS par rapport au run de référence, moyenné horizontalement sur la région des TIWs (entre 5°S et 5°N) au jour 20252 (en haut) et jour 20270 (en bas) pour : le run faux (trait plein gras), l'état prédit dans le run assimilé (trait gras en tiretés), l'état analysé par le filtre (trait fin) et l'état analysé par le lisseur avec 5 analyses rétrospectives (trait fin en tiretés). Les variables représentées sont la température (en °C), la salinité et la vitesse absolue (en m.s ⁻¹).	133

8.8	Cartes des erreurs absolues en température (en °C) moyennées sur les trente premiers mètres (pour l'état prédit e^f , l'état analysé par le filtre e^a et par le lisseur e^s) et cartes de différences des erreurs absolues entre l'analyse du filtre et le forecast ($e^a - e^f$) et entre l'analyse du lisseur et l'analyse du filtre ($e^s - e^a$). Pour les cartes de différences, les valeurs négatives (respectivement positives) indiquent une amélioration (respectivement une détérioration) de l'analyse du filtre (par rapport à l'état faux) ou du lisseur (par rapport au filtre). Les cartes sont données au jour 20252 (à gauche) et au jour 20270 (à droite). Les courbes noires et pointillées noires indiquent respectivement l'iso-contour 28°C du run de référence et du run perturbé avant l'assimilation.	134
8.9	Diagramme de Hovmuller montrant l'évolution de l'erreur absolue en température de surface (en °C) à 3°N, entre le run de référence et le run faux, les états analysés par le filtre et le lisseur, et les runs libres issus de l'analyse (au jour 20252) du filtre et du lisseur, le long d'un transect entre 42°W et 12°W.	135
8.10	Evolution de l'erreur en température (en °C), le long d'un profil vertical et en fonction du temps, pour les états analysés par le filtre (gauche) et par le lisseur (droite). Une observation est située exactement au niveau de chaque profil au jour 20266 (trait noir). Les profils sont situés aux points (de haut en bas) : 1.5°S-23°W, 2.5°N-27.25°W et 2°S-14°W	136
8.11	Evolution de l'erreur (en valeur absolue) pour la température (en °C) en un point unique en surface et en fonction du temps, pour les états analysés par le filtre (courbe pleine) et par le lisseur (courbe pointillée). Une observation est située exactement au niveau du point de calcul des erreurs au jour 20266.	137
8.12	Profils verticaux de l'écart RMS par rapport au run de référence, moyenné horizontalement sur la région des anneaux du Brésil au jour 20252 (en haut) et jour 20270 (en bas) pour : le run faux (trait plein gras), l'état prédit dans le run assimilé (trait gras en tiretés), l'état analysé par le filtre (trait fin) et l'état analysé par le lisseur avec 5 analyses rétrospectives (trait fin en tiretés). Les variables représentées sont la température (en °C), la salinité et la vitesse absolue (en m.s ⁻¹).	138
8.13	Cartes des erreurs absolues en température (en °C) moyennées sur les trente premiers mètres (pour l'état prédit e^f , l'état analysé par le filtre e^a et par le lisseur e^s) et cartes de différences des erreurs absolues entre l'analyse du filtre et le forecast ($e^a - e^f$) et entre l'analyse du lisseur et l'analyse du filtre ($e^s - e^a$). Pour les cartes de différences, les valeurs négatives (respectivement positives) indiquent une amélioration (respectivement une détérioration) de l'analyse du filtre (par rapport à l'état faux) ou du lisseur (par rapport au filtre). Les cartes sont données pour le jour 20252 (à gauche) et au jour 20270 (à droite). Les courbes noires et pointillées noires indiquent respectivement l'iso-contour 0,5m.s ⁻¹ du run de référence et du run perturbé avant l'assimilation.	139
8.14	Coupes de vitesse absolue (en m.s ⁻¹) pour le run de référence, le run faux, le run analysé, le run lissé, et les runs libres issus de l'analyse au jour 20252 du filtre et du lisseur. Les coupes sont tracées à 6.75°N pour le jour 20260 et à 6°N pour le jour 20290.	140
8.15	Corrélations en température à une observation de température (rectangle noir), calculées avec différentes bases de covariances d'erreur : initiale (20250) et issues du filtre évolutif aux jours 20258 et 20276.	141

8.16	Cartes de différences des erreurs absolues entre l'analyse du filtre et le forecast ($e^{a-f} = x^a - x^t - x^f - x^t $) et entre l'analyse du lisseur et l'analyse du filtre ($e^{s-a} = x^s - x^t - x^a - x^t $), pour la température moyennée sur les 30 premiers mètres (a-f) et la vitesse absolue moyennée entre 30 et 100m (g-l). Les valeurs négatives indiquent une amélioration de l'analyse du filtre par rapport au forecast ou du lisseur par rapport au filtre. Les cartes sont données pour le jour 20252 (à gauche), 20270 (au milieu) et 20290 (à droite). Toutes les valeurs sont exprimées en °C pour les températures et en $m.s^{-1}$ pour les vitesses.	142
8.17	Evolution de l'erreur RMS 3D estimée par le filtre (trait plein noir), pour les états analysés par filtre (trait plein rouge) et analysés par le lisseur avec 5 analyses retrospectives (croix bleues), pour la température (en °C) et la vitesse zonale (en $m.s^{-1}$). L'erreur RMS du run faux est rappelée en trait pointillé noir.	144
8.18	Evolution de l'erreur RMS pour le run corrigé par le filtre (trait noir), et dans le cadre de runs libres redémarrant à partir d'un état analysé par le filtre (trait vert) ou d'un état analysé par le lisseur (trait bleu), pour la température (en °C) et la vitesse zonale (en $m.s^{-1}$).	144
8.19	Evolution de l'erreur en vitesse absolue (en $m.s^{-1}$) entre le run de référence et les runs libres issus de l'analyse du filtre (à gauche) et du lisseur (à droite), le long d'un transect à 5°N (entre 40°W et 50°W), à 30 m de profondeur.	145
8.20	Cartes de vitesse absolue (en $m.s^{-1}$) pour le run de référence, le run faux, le run libre issu de l'analyse du filtre au jour 20252 et le run libre issu de l'analyse du lisseur (5 analyses rétrospectives) au jour 20252. Les figures sont données aux jours 20252, 20260, 20270, 20280, 20290 et 20310.	146
8.21	Avec un réseau de type temps-groupé : évolution de l'erreur RMS pour le run corrigé par le filtre (trait noir), et dans le cadre de runs libres redémarrant à partir d'un état analysé par le filtre (trait vert) ou d'un état analysé par le lisseur (trait bleu), pour la température (en °C) et la vitesse zonale (en $m.s^{-1}$).	148
8.22	Cartes de la vitesse absolue (en $m.s^{-1}$) à 30m de profondeur dans la région des anneaux du Brésil au jour 20330 pour le run de référence et les runs libres issus des analyses au jour 20260 du filtre et du lisseur.	148
9.1	Schématisation de l'expérience "d'Observations Ponctuelles". La courbe noire symbolise un niveau d'amplitude de l'erreur RMS d'un run à corriger. Les symboles en noir font référence à l'expérience précédente du filtre. Les symboles en rouge correspondent aux étapes liées à l'ajout d'observations ponctuelles au temps t. Les notations sont : a(analyse du filtre au temps considéré) et s'x'(x analyses rétrospectives du lisseur au temps considéré).	153
9.2	Carte de SST (en °C) sur laquelle sont positionnées les observations ponctuelles (rectangles noirs) correspondant à des profils isolés spatialement de température.	153

9.3	<p>Dans la région des TIWs : cartes de l'erreur absolue (première et troisième ligne) en température (en °C) entre le run de référence et chaque état précédent une analyse (du filtre ou du lisseur) avec les observations ponctuelles (introduites au jour 20258), dans le cas du filtre évolutif et dans le cas de l'interpolation optimale. Chaque carte d'erreur est associée en-dessous à une carte de correction indiquant l'effet d'une analyse suite à l'introduction d'une observation ponctuelle. Pour les cartes de correction, les valeurs négatives indiquent une bonne correction (diminution de l'erreur par l'analyse), et inversement pour les valeurs positives. Les corrections et les erreurs sont données (de gauche à droite) pour l'étape d'analyse du filtre, de la première et de la troisième analyses rétrospectives du lisseur. x^t : état de référence, x^f : état prédit avant l'analyse, x^a : état analysé par le filtre, x^{si} : état issu de la i^{eme} analyse rétrospective du lisseur. Les rectangles noirs indiquent la position des observations ponctuelles introduites.</p>	155
9.4	<p>Figure identique à la figure 9.3, excepté que l'observation ponctuelle est introduite ici au jour 20276.</p>	156
9.5	<p>Dans la région des anneaux du Brésil, cartes de l'erreur absolue en vitesse (en $m.s^{-1}$) entre le run de référence et chaque état précédent une analyse (du filtre ou du lisseur) avec les observations ponctuelles (introduites au jour 20258), dans le cas du filtre évolutif et dans le cas de l'interpolation optimale. Chaque carte d'erreur est associée en-dessous à une carte de correction indiquant l'effet d'une analyse suite à l'introduction d'une observation ponctuelle. Pour les cartes de correction, les valeurs négatives indiquent une bonne correction (diminution de l'erreur par l'analyse), et inversement pour les valeurs positives. Les corrections et les erreurs sont données (de gauche à droite) pour l'étape d'analyse du filtre, de la première et de la troisième analyses rétrospectives du lisseur. x^t : état de référence, x^f : état prédit avant l'analyse, x^a : état analysé par le filtre, x^{si} : état issu de la i^{eme} analyse rétrospective du lisseur. Les rectangles noirs indiquent la position des observations ponctuelles introduites.</p>	158
9.6	<p>Figure identique à la figure 9.5, excepté que l'observation ponctuelle est introduite ici au jour 20276.</p>	159
9.7	<p>Evolution des écarts RMS pour le filtre évolutif (en rouge) et l'OI (tirets bleus), pour la SSH en m (en haut) et la température en °C (milieu). Les croix rouges et bleues correspondent respectivement aux analyses du lisseur dans le cas du filtre évolutif et dans le cas de l'OI. A gauche se trouve l'évolution de l'erreur RMS sur toute la période, et à droite se trouve un zoom sur la période 20270-20300. La trajectoire du run faux est rappelée par la courbe noire pointillée.</p>	160
9.8	<p>Evolution des écarts RMS pour le filtre évolutif (en rouge) et l'OI (en bleu), pour la température en °C, la salinité, la SSH en m et la vitesse zonale en $m.s^{-1}$. Les croix rouges et bleues correspondent respectivement aux analyses du lisseur dans le cas du filtre évolutif et dans le cas de l'OI. La trajectoire du run faux est rappelée par la courbe noire pointillée.</p>	161
9.9	<p>Evolution de l'erreur RMS vraie pour le run libre redémarrant à partir d'un état analysé par le filtre (trait plein rouge) ou d'un état analysé par l'OI (trait plein bleu), et pour les runs libres redémarrant à partir du lisseur associé au filtre (trait pointillé rouge) ou associé à l'OI (trait pointillé bleu). Les courbes sont données pour la température (en °C) et la vitesse zonale (en $m.s^{-1}$), avec des redémarrages effectués aux jours 20252 (en haut) et 20270 (en bas).</p>	163

9.10 Dans le cadre d'un réseau de type "temps-groupé" : évolution de l'erreur RMS vraie pour le run libre redémarrant à partir d'un état analysé par le filtre (trait plein rouge) ou d'un état analysé par l'OI (trait plein bleu), et pour les runs libres redémarrant à partir du lisseur associé au filtre (trait pointillé rouge) ou associé à l'OI (trait pointillé bleu). Les courbes sont données pour la température (en °C) et la vitesse zonale (en m.s⁻¹), avec des redémarrages effectués aux jours 20260 (en haut) et 20270 (en bas). 164

Liste des tableaux

3.1	<i>Profondeurs (en mètres) des niveaux verticaux dans la configuration TATL4.</i>	43
3.2	<i>Caractéristiques principales de la configuration TATL4.</i>	47
4.1	<i>Les différentes étapes du filtre de Kalman.</i>	57
4.2	<i>Equations du lisseur de Kalman.</i>	66
5.1	<i>Les différentes étapes du filtre de Kalman.</i>	75
6.1	Tableau récapitulatif des valeurs des différents paramètres de diffusion dynamique (diffusion horizontale A^{hm} , diffusion verticale en surface 1×10^{-4} et diffusion verticale en profondeur A^{vm}) pour le run de référence REF-ES et les run perturbés ES1 à ES6. Les valeurs perturbées par rapport aux valeurs de référence sont indiquées en gras.	90
7.1	Dénominations et tailles des différentes zones d'influence utilisées (en nombre de points de grille). Le premier chiffre indique la zone de coupure (c'est-à-dire la zone au-delà de laquelle les observations n'ont plus d'influence) et le chiffre entre parenthèses indique la zone d'influence maximale (c'est-à-dire la zone au-delà de laquelle le poids des observations diminue pour atteindre 0).	112
7.2	Zones d'influences et nombre d'EOFs associés à chaque expérience de sensibilité.	116
7.3	Nombre d'EOFs utilisé pour les expériences de sensibilité à la troncature de la base d'erreur.	118
7.4	Paramètres d'erreur d'observation sur les variables assimilées : température (en °C.), salinité et SSH (en m), pour les différents tests de sensibilité de la propagation des statistiques d'erreur.	119
9.1	Dénominations des différentes expériences d'observations ponctuelles en précisant pour chacune de quelle expérience précédente sont tirés les résultats et les covariances d'erreur (expérience du filtre ou de l'OI) et à quelle date sont introduites les observations ponctuelles.	152