A comparative study of data assimilation methods for oceanic models

Giovanni Abdelnur Abdelnur Ruggiero Ruggiero

Université Nice Sophia-Antipolis - UFR Sciences

Ecole Doctorale Sciences Fondamentales et Appliquées

T H È S E

pour obtenir le titre de

Docteur en Science

de l'Université Nice Sophia-Antipolis

Spécialité : Science de la Planète et de l'univers

présentée et soutenue par

Giovanni RUGGIERO

Une étude comparative de méthodes d'assimilation de données pour des modèles océaniques. A comparative study of Data Assimilation methods for oceanic models.

Thèse dirigée par Jacques BLUM et Yann OURMIÈRES

soutenue le 13 Mars 2014

Jury

Mark Asch	*Rapporteur*
Didier Auroux	*Examinateur*
Joaquim Ballabrera Poy	*Examinateur*
Jacques Blum	*Directeur de thèse*
Pierre Brasseur	*Rapporteur*
Emmanuel Cosme	*Examinateur*
Yann Ourmières	*Co-directeur de thèse*
Jacques Verron	*Examinateur*

Remerciements

Tout d'abord j'aimerais remercier les deux Jacques qui ont rendu possible la réalisation de cette thèse: Jacques Verron pour avoir cherché un financement pour que je puisse faire une thèse sur l'assimilation de données en France et Jacques Blum pour m'avoir accepté et m'avoir accueilli avec une telle gentillesse et attention qui ont fait disparaitre toutes les difficultés culturelles que je pouvais expérimenter comme nouveau thésard en France et surtout au LJAD.

Je remercie également Yann Ourmières, mon co-directeur de thèse, pour les réflexions et l'aide pour la rédaction de la thèse, et Emmanuel Cosme pour toute sa attention et les nombreuses discussions sur l'assimilation de données, vous avez été très importants dans la démarche de cette thèse.

Mes remerciements s'adressent aussi à Mark Asch et Pierre Brasseur pour avoir accepté aimablement de rapporter sur ma thèse.

Je remercie aussi Didier Auroux, Joaquim Ballabrera Poy, Emmanuel Cosme et Jacques Verron pour m'avoir accordé l'honneur de faire partie du jury de ma thèse.

Je remercie egalement Hélène Politano et Severine Rigot pour leur disponibilité et leur assistance à l'Ecole Doctorale.

Mes remerciements à tout le personnel du LJAD, spécialement à Isabelle De Angelis et à Jean Paul Pradere.

Un remerciements amical à Jean-Marc Lacroix et à Julien Maurin qui ont surtout assuré les moyens informatiques nécessaires à mes travaux.

Je remercie aussi Brice Eichwald pour m'avoir beaucoup aidé avec la langue française et toutes les démarches bureaucratiques. Lord Bienvenu, Nancy Abdallah et Camilo Garcia pour les bons moments passés ensembles.

Un remerciement très special à mon ami Sami Sassi qui fait partie de ma famille en France.

Je remercie mes parents pour être toujours à mon coté, grâce à eux j'en suis là !

Enfin, je remercie ma compagne Leticia qui partage sa vie avec moi depuis huit ans et sans laquelle je ne pourrais pas passer les moments de travail les plus durs. Espero poder abrir muitas cortinas contigo e com a nossa Dorinha.

Contents

List of Figures

List of Tables

Acronyms

BFN Back and Forth Nudging

BFFKF Back and Forth Fixed basis Kalman Filter

BFKF Back and Forth Kalman Filter

BLUE Best Linear Unbiased Estimate

BS Backward Smoother

DA Data Assimilation

DBFN Diffusive Back and Forth Nudging

EKF Extended Kalman Filter

EnKF Ensemble Kalman Filter

KF Kalman Filter

iBS iterative Backward Smoother

LO Luenberger Observer

LQG Layered Quasi-Geostrophic

NEMO Nucleus for European Modelling of the Ocean

OLS Ordinary Least Squares

PLS Partial Least Squares

RIP Running in Place

RM Regression Model(s)

SEEK Singular Evolutive Extended Kalman Filter

SW Shallow Water

SSH Sea Surface Height

4Dvar Four Dimensional Variational method

Abstract

Data Assimilation – DA – methods comprehend the ensemble of methods aiming to estimate the state of a system given all informations about it, i.e models and observations. Generally speaking, DA methods provide a way to weight the combination of models and observations. These methods are extensively used in the generation of accurate initial conditions for atmospheric and oceanic forecast systems, as well as to produce reanalysis fields (past state reprocessing based on future observations). Their importance relies on the possibility of accurately predicting the future evolution of atmospheric and oceanic events, which would support decisions concerning the agriculture investments and natural hazards contention plans, for instance.

This thesis developed and implemented iterative data assimilation algorithms for a primitive equation ocean model, and compared them with other well established DA methods such as the 4Dvar and the Singular Evolutive Extended Kalman (SEEK) Filter/Smoother. The numerical model used was the NEMO model. It was configured to simulate a typical subtropical double gyre circulation at an eddy permitting resolution. The new proposed iterative algorithms, similarly to the Back and Forth Nudging, are all based on a sequence of alternating forward and backward model integrations. Namely, they are the Backward Smoother (BS), which uses the backward model to freely propagate "future" observations backward in time, and the Back and Forth Kalman Filter, which also uses the backward model to propagate the observations backward in time but, at every time an observation batch is available, an update step similar to the SEEK filter step is carried out. The Bayesian formalism was used to derive these methods, which means that they may be

used with any algorithm that estimates the *a posteriori* conditional probability of the model state by means of sequential methods.

The results show that the main advantage of the methods based on the BFN is the use of the backward model to propagate the observation informations backward in time. By this way, it avoids the use of the adjoint model, needed by the 4Dvar, and of unknown temporal correlations, needed by the Kalman Smoother, to produce initial states or past model trajectories. The advantages of using the Back and Forth (BF) idea rely on the implicit use of the unstable forward subspace, which became stable when stepping backwards, that allows the errors components projecting onto this subspace to be naturally damped during the backward integration. It was shown that forecasts initialized from the Backward Smoother are 4-10% more accurate than the forecasts initialized by the SEEK Smoother using the same computational power. The BFN results are also encouraging when compared with the results produced by a 4Dvar algorithm. Using prescribed weights along with climatological correlations between observed and non- observed variables, the BFN produced results which are at least comparable with the 4Dvar results but with 3 times less computational power.

Therefore, this work shows that there is a real interest in using the BF idea for producing initial condition states for ocean forecasts due to the simplicity of implementation of the backward model and due to the possibility of using existing softwares, which implement sequential DA algorithms, along with the Back and Forth iterations. Hence, it would be very simple to implement these new methods on the existing assimilation systems.

Résumé étendu

Cette thèse a développé et mis en œuvre des algorithmes itératifs d'assimilation de données pour un modèle d'océan aux équations primitives, et les a comparés avec d'autres méthodes d'assimilation de données bien établies telles que le 4Dvar et le filtre/lisseur SEEK. Le modèle numérique utilisé est le modèle NEMO. Il a été configuré pour simuler la circulation typique subtropicale en double gyre à une résolution de mésoéchelle (voir chapitre 3). Les nouveaux algorithmes itératifs proposés, de façon similaire au Nudging direct et rétrograde, sont tout basés sur une séquence d'intégrations alternées du modèle direct et rétrograde. Dans la pratique, les méthodes diffèrent quant à la façon dont les poids donnés aux observations et au modèle sont calculés.

Nous mettons en évidence trois chapitres: l'un décrivant les méthodes d'assimilation (chap. 2) et deux autres présentant les résultats numériques (chap. 4 et 5). Dans le chapitre 2, les algorithmes ont été classés comme des algorithmes déterministes, pour lesquels le système n'a pas de terme aléatoire, et les algorithmes stochastiques. Pour les premiers, aucune condition d'optimalité n'a été utilisée pour calculer les gains, pour cette raison, nous avons continué à appeler ces méthodes le Back and Forth Nudging (BFN), alors que pour les derniers, nous avons utilisé les hypothèses de Kalman pour obtenir des gains optimaux.

Deux algorithmes, obtenus à l'aide du raisonnement de Bayes et compte tenu d'un scénario de modèle parfait, ont été proposés: le Backward Smoother (BS) et le Back and Forth Kalman Filter (BFKF). Le premier s'appuie sur la connaissance de la fonction de densité *a posteriori* finale, $p(\mathbf{x}_K|\mathbf{y}_{1:K})$, pour récupérer la probabilité conditionnelle d'un état passé étant donné les obser-

vations passées et futures, $p(\mathbf{x}_k|\mathbf{y}_{1:K})$ pour $k < K$. Dans la pratique, il intègre le modèle rétrograde en utilisant la solution finale du filtre direct comme condition initiale. Il a été montré que lorsque la transition $p(\mathbf{x}_{K-1}|\mathbf{x}_K)$ est connue et linéaire dans les variables d'état, la solution du BS obtenue au temps t_k initialisée à partir de la solution du filtre de Kalman au temps t_K est équivalente à la solution du lisseur de Kalman à t_k en utilisant les observations disponibles jusqu'à t_K. Le deuxième algorithme (BFKF) s'appuie formellement sur la décomposition des observations en p sous-ensembles indépendants, chacun d'eux étant assimilé à chaque nouveau passage du filtre. D'une manière générale, il est tout à fait semblable au BFN mais le gain $\mathbf{K}$ maximise la probabilité conditionnelle, $p(\mathbf{x}_k|\mathbf{y}^n_{1:k})$ (pour le nième passage du filtre direct) ou $p(\mathbf{x}_k|\mathbf{y}^{n+1}_{k:K})$ (pour le nième passage du filtre rétrograde), à chaque étape d'analyse. La décomposition d'observations ne fournit pas seulement un gain optimal, mais aussi un nombre optimal d'itérations. Il est clair qu'en utilisant la même décomposition d'observation pour le BS, celui-ci peut également être itéré.

Les chapitres 4 et 5 décrivent les expériences numériques réalisées avec ces méthodes. Dans le chapitre 4 l'approche déterministe a été utilisé, ce qui signifie qu'aucune condition d'optimalité n'a été considérée dans la construction des gains, tandis que dans le chapitre 5 le BS et le BFKF ont été utilisés. Indépendamment de l'optimalité du gain, un problème concernant les applications océaniques est la nature irréversible des modèles qui peuvent empêcher l'utilisation du modèle rétrograde.

Dans le chapitre 4, nous avons montré qu'il est impossible d'intégrer le modèle rétrograde sans un contrôle rigoureux et un réseau d'observation irréaliste, mais que l'inversion du signe du terme de diffusion stabilise le modèle. Toutefois, dans ce dernier cas, il a été observé une perte de précision, qui augmente avec la longueur de la fenêtre d'assimilation. On a constaté qu'il s'agissait d'un résultat de sur-diffusion, c'est à dire que les champs sont fortement lissés, ce qui équivaut à une perte de la résolution effective du modèle.

Une façon simple de réduire ce problème a été d'appliquer un coefficient

de diffusion réduit offrant les plus petites erreurs. Nous avons trouvé qu'avec un coefficient de diffusion constant (dans le temps et dans l'espace) les erreurs ont été réduites jusqu'à une limite qui est liée à la diffusion nécessaire pour stabiliser la méthode numérique. Cela a été fait de manière empirique, mais comme il a été suggéré, cela peut être fait par la construction d'un opérateur de diffusion d'ordre n qui fournit la réponse souhaitée pour chaque échelle spatiale, c'est à dire qui permet une bonne précision à moyenne échelle tout en supprimant les instabilités numériques. Il s'agit clairement d'une question qui doit être étudiée.

Deux gains ont été testés dans le chapitre 4, une matrice $\mathbf{K}$ diagonale, ce qui signifie que la covariance des erreurs sont ignorées et la dynamique du modèle lui-même propage les corrections vers les régions et les variables non observées, et un $\mathbf{K}$ consistant en deux opérations: d'abord les variables observées sont mises à jour en utilisant un poids prescrit, puis les variables non observées sont mises à jour en utilisant un modèle de régression. Comme notre objectif était de construire une méthode peu coûteuse, le modèle de régression a été maintenu fixe pendant les cycles d'assimilation. Pour ce type d'utilisation, nous avons montré qu'une méthode de régression ayant des propriétés de régularisation est préférable. Dans ce travail, cette méthode est celle des moindres carrés partiels. Elle cherche p directions dans l'espace d'observation et dans l'espace du modèle qui sont mieux corrélées, en même temps qu'elles maximisent leur variances.

Ces deux gains ont été testés avec différents réseaux d'observations: un simulant les champs de SSH maillés et un autre simulant les observations de SSH échantillonnées sur une piste typique d'un satellite altimétrique. Les conclusions que nous tirons de ce chapitre est que le modèle a une capacité limitée pour corriger les champs thermodynamiques lorsque seules les observations de SSH sont disponibles, et dans ce cas peu importe si la SSH maillée ou la SSH échantillonnée sur la trajectoire du satellite est assimilée. Plus les observations sont rares, plus important est l'utilisation de modèles de régression pour aider le modèle à diffuser les informations d'observations.

Enfin, il a été montré que, malgré la faible précision du modèle rétrograde, lorsque le modèle de régression comprend toute la partie non observée de l'espace d'état, notamment la température, le Diffusive Back and Forth Nudging (DBFN) peut donner des résultats qui sont au moins équivalents au 4Dvar que nous avons utilisé, mais avec un effort de mise en oeuvre beaucoup plus faible, car aucun code adjoint n'est nécessaire, et avec un coût de calcul près de 3 fois plus faible. Remarquablement, le DBFN atteint leur niveau d'erreur asymptotique après le premier cycle d'assimilation tandis que le 4Dvar et le Nudging direct ordinaire ont besoin de douze cycles.

Ce fait montre la puissance des itérations à réduire les erreurs lorsque l'état du modèle est très différent de l'état vrai, comme c'est le cas au début de l'expérience. Cet aspect a été discuté dans la littérature par Kalnay and Yang (2010); Yang et al. (2012); Wang et al. (2013b) qui ont souligné l'importance des itérations lorsque les non-linéarités jouent un rôle important et que les membres de l'ensemble ou de la base réduite n'engendrent pas le vrai espace d'erreur.

Dans le chapitre 5, le gain du BFN a été formellement dérivé pour être le gain de Kalman standard. Dans ce cas, le filtre de rang réduit " Singular Evolutive Extended Kalman - filtre SEEK" a été utilisé pour calculer les analyses. Le filtre SEEK et lisseur SEEK ainsi qu'une version itérative de ce lisseur (Running in Place - RIP), semblable à l'algorithme utilisé par Kalnay and Yang (2010), ont été pris comme référence.

Les résultats confirment que malgré la faible précision du modèle rétrograde, il est avantageux de l'utiliser à la place des corrélations linéaires pour mettre à jour un champ passé. Cela est dû à la correction des composantes d'erreur qui se projettent sur le sous-espace instable au cours de l'intégration rétrograde. Ce mécanisme est particulièrement important lorsque la base réduite couvre seulement un petit sous-espace instable, soit parce qu'il a un rang très réduit ou parce que la base s'étend sur les deux sous-espaces: instable et stable.

D'une part il a été montré que lorsque la matrice de covariance a un rang relativement grand (> 50) et engendre seulement un sous-espace instable,

les avantages de l'utilisation du modèle rétrograde pour améliorer la solution de filtrage sont réduits. D'autre part, les prévisions initialisées après une intégration rétrograde sont toujours améliorées par rapport à celles réalisées à partir du filtre ou lisseur SEEK, en particulier dans la gamme de 10 à 40 jours, même si les conditions initiales de ces dernières sont plus précises.

Si on est surtout intéressé par les conditions initiales pour des fins de réanalyse par exemple, une façon possible de continuer à améliorer les conditions initiales est de reproduire certains modes "instables en rétrograde/stables en direct", ce qui empêchera les erreurs qui se projettent sur ces modes de croître. Ceci implique une augmentation des coûts numériques, et donc ses bénéfices doivent être soigneusement examinés .

En ce qui concerne les versions itératives iBS, BFKF et RIP, l'amélioration des variables non observées est plus grande pour les deux premiers, tandis que pour les variables observées les trois méthodes ont montré des performances similaires. Encore une fois, cela peut révéler l'importance de l'intégration rétrograde à réduire les erreurs instables ainsi que révéler l'amélioration de la covariance entre les variables observées et non observées. En outre, les itérations ont transformé la distribution des erreurs de prévision dans une distribution plus gaussienne améliorant ainsi la performance du filtre/lisseur.

Une faiblesse détectée à propos de notre mise en œuvre de l'iBS et BFKF est la divergence observée lorsque l' inflation a été considérée, même si nous avons vu que l'inflation fournit de petites erreurs en moins d'itérations. Par conséquent, nous suggérons l'utilisation de filtres qui n'ont pas besoin d'inflation ou de schémas d'inflation adaptatifs. Cela permettrait de conserver les statistiques de filtrage cohérentes avec les statistiques de l'erreur vraie et donc d'utiliser aussi efficacement que possible une observation particulière, étant donné les limites de chaque méthode. En outre, les statistiques utilisées par l' approche adaptative peuvent être utilisées pour obtenir un critère d'arrêt plus efficace, puisque le critère de convergence que nous avons utilisé dans le chapitre 4 n'est pas capable de détecter la divergence signalée dans le chapitre 5.

Cette thèse a prouvé qu'il existe un réel intérêt pour l'utilisation du DBFN et des algorithmes basés sur les itérations du modèle direct et rétrograde pour l'assimilation des observations océaniques. Bien sûr, il est nécessaire d'étudier comment ces méthodes se comportent dans un système d'assimilation opérationnel pour lequel, en général, les modèles sont biaisés et pas parfaits. Plus de développements sont nécessaires, notamment en ce qui concerne la diffusion de l'intégration rétrograde et la stabilité à long terme des filtres/lisseurs qui peuvent être traités avec des systèmes adaptatifs.

Enfin, nous croyons que les algorithmes présentés dans cette thèse méritent d'être testés en mode opérationnel comme le sont le 4Dvar et les filtres de Kalman. En effet, compte tenu de la simplicité de mise en œuvre du modèle rétrograde, les algorithmes itératifs présentés ici peuvent être facilement mis en œuvre en tirant partie des systèmes d'assimilation existants.

Chapter 1

Introduction

1.1 Introduction (French)

La croissance de la population et la demande de plus en plus élevée de la protection sociale ont conduit à une augmentation de l'exploitation des ressources naturelles. Les zones de forte concentration de population sont très sensibles aux phénomènes naturels extrêmes. Elles sont d'une part directement affectées comme par exemple dans le cas du Tsunami qui s'est produit dans l'Océan Indien en 2004 et celui du Japon en 2013, et d'autre part indirectement affectées en conséquence des perturbations météorologiques sur la production alimentaire à grande échelle.

Sur cette base, la connaissance des composantes du système physique de la Terre tels que l'atmosphère et l'océan et des lois les régissant est devenue essentielle pour la vie moderne. La possibilité de reconstruire les événements passés et de prédire avec précision les événements futurs a changé la dynamique économique. Par exemple, les organismes d'assurance et de crédit ont eu recours à la connaissance de la météo ainsi que des prévisions météorologiques pour ajuster leurs évaluations des risques.

Au cours des cinquante dernières années, des progrès sur les prévisions météorologiques se sont produits en raison notamment de l'amélioration de la formulation des modèles, comme une conséquence du développement impressionnant des ressources informatiques, et grâce à l'amélioration des réseaux d'observation et des conditions initiales. Ce progrès dans la prévision est illustrée à la figure 1.1 par la qualité des prévisions des champs de 500hPa produites au NCEP [1]. La barre horizontale indique une augmentation d'un jour

[1] National Centers for Environmental Prediction

par décennie dans la qualité des prévisions. Pour montrer que ceci est important pour la société de nos jours, selon Ravinder Singh, chef de la physique agricole à l'Institut indien de recherche agricole, "la précision dans la prévision de la mousson peut augmenter la production agricole de 10 pour cent à 15 pour cent dans les zones pluviales ". Cela signifie que l'amélioration des capacités de prévision représenterait une croissance dans la production mondiale de nourriture.

Beaucoup moins d'attention a été accordée à l'élaboration des prévisions océaniques. Toutefois, l'océan joue un rôle clé dans les régimes météorologiques et climatiques, car il délimite la basse atmosphère. A cette interface, il y a des échanges de quantité de mouvement, vapeur d'eau et de chaleur entre l'ocean et l'atmosphère. En outre, on pense que les interactions océan-atmosphère deviennent un facteur dominant pour les prévisions à long terme. Par exemple, les variabilités atmosphériques importantes, qui touchent directement plusieurs millions de personnes, telles que la mousson de l'océan Indien et les phénomènes El Nino/La Nina du Pacifique sont fortement contrôlées par la dynamique de l'océan supérieur (Keshavamurthy and Sankar Rao, 1992; Wang et al., 2012).

En outre, l'océan et les régions côtières sont des zones riches en ressources naturelles. Cela a conduit les établissements humains à se concentrer sur les zones à proximité du littoral. A titre d'exemple, la densité populationnelle moyenne dans les régions côtières a été estimée 3 fois plus élevée que la moyenne mondiale. De plus, les commerces à travers le monde utilisent le plus souvent la mer à l'exportation et à l'importation des marchandises, et le poisson représente environ un cinquième des protéines animales dans l'alimentation humaine.

Par conséquent, l'importance de produire des prévisions fiables de l'océan et de l'atmosphère est justifiée, entre autres facteurs, par leur utilisation dans le soutien des activités économiques, ainsi que la gestion de l'environnement. Mais, quels sont les défis pour produire des prévisions fiables? En termes mathématiques, une prévision est un problème de valeur initiale. On nous donne un ensemble d'équations différentielles évolutives à partir duquel nous voulons calculer l'évolution de l'état du système, étant donné un état initial. Par conséquent, une bonne prévision dépend du degré de représentativité physique du modèle mathématique (et numérique) et du degré de précision des condi-

tions initiales.

Au début des années 60 le scientifique américain Edward Norton Lorenz a déclaré que la nature chaotique des équations régissant l'atmosphère impose une limite de prévisibilité de temps fini. Il a trouvé qu'un modèle atmosphérique simplifié initialisé avec deux conditions initiales légèrement différentes entraîne deux solutions complètement différentes après une simulation sur deux semaines. Par conséquent, beaucoup d'efforts ont été faits pour produire des conditions initiales les plus précises possible.

La science de la production de champs géophysiques précis a été nommé Assimilation de Données - DA (Bennett, 2002; Kalnay, 2003; Evensen, 2009). Ce nom vient du fait que les observations du système sont introduites (assimilées) dans les modèles pour améliorer les descriptions du système: l'observationnel et la mathématique/numérique. Cela signifie que l'assimilation vise à fournir des informations sur le système réel qui sont en même temps plus proches de la vérité que les observations et les modèles séparément.

Pour illustrer les difficultés rencontrées par l'assimilation de données océaniques, les fig. 1.2 et 1.3 montrent un réseau d'observation typique de la mer Méditerranée et deux résultats de modèle pour la même région, l'un avec une résolution spatiale de 12 km et l'autre de 3 km. Le premier aspect observé est la rareté du réseau d'observation qui révèle que les observations ne sont pas assez denses pour produire des conditions initiales pour les modèles numériques. Plus que cela, il n'y a presque pas d'information directe sur la structure synoptique du champ vertical de masse (les balles vertes dans la figure représentent les bouées Argo[2]). La plupart des données sont des mesures directes[3] de la surface de l'océan (température de surface de la mer et hauteur de la surface de la mer), même s'ils peuvent observer des propriétés associées à la dynamique de la thermocline (Wunsch, 1997).

L'Assimilation de Données a fait ses débuts sur la base de deux piliers: la théorie du contrôle (Jazwinski, 1970; Gelb, 1974) et la théorie du contrôle optimal (Lions, 1971). Ces domaines sont particulièrement intéressés par le contrôle de systèmes électriques et de trajectoire balistique et satellitaire, par exemple. De ces deux branches scientifiques ont émergé les méthodes d'assimilation de données les plus utilisées de nos jours: le Filtre

[2]Sondes robotiques océaniques déployés dans le monde entier. Les sondes flottent aussi profond que 2 km et une fois tous les 10 jours échantillonnent mesures de conductivité et de température de la colonne d'eau. Les données sont transmises aux scientifiques à terre par satellite. (source http://en.wikipedia.org/wiki/Argo_%28oceanography% 29)

[3]Cela ne signifie pas que les mesures ont été faits *in situ* , mais que les propriétés mesurées sont à la surface ou est la surface elle-même.

de Kalman (Kalman, 1960; Kalnay, 2003; Evensen, 2009) et les méthodes variationnelles en quatre dimensions (Le Dimet and Talagrand, 1986; Bennett, 2002). Toutefois, en raison de la taille très importante des problèmes et des ressources informatiques limitées, l'assimilation de données a eu son début, au milieu des années 70, avec l'utilisation de méthodes simples et peu coûteuses, parmi lesquelles se distingue le Nudging.

La méthode du Nudging est basée sur le deuxième axiome de Newton et consiste à ajouter un terme de forçage dans le second membre d'un système donné, afin de pousser doucement le modèle vers une valeur prescrite. La première apparition du Nudging dans la littérature géophysique a été en 1974 (Anthes, 1974). Dans ce travail, les auteurs ont proposé d'utiliser le Nudging pour atténuer les problèmes d'initialisation des modèles atmosphériques.

La première apparition d'une application réussie du Nudging en océanographie a été en 1992 dans une étude qui a assimilé la hauteur de la mer, issue de mesures satellitaires, dans un modèle quasi-géostrophique en couches (Verron, 1992). Depuis, la méthode a été appliquée avec succès à plusieurs problèmes océanographiques tels que les conditions aux limites (Marchesiello et al., 2001; Chen et al., 2013), downscaling (Li et al., 2012), et d'autres problèmes d'assimilation de données (Verron, 1992; Haines et al., 1993; Lewis et al., 1998; Killworth et al., 2001; Thompson et al., 2006). En ce qui concerne les applications aux problèmes d'assimilation des données, les poids attribués au modèle et aux observations, en général, ne repose sur aucune condition d'optimalité, mais sont plutôt des scalaires ou des fonctions Gaussiennes construites sur la base des hypothèses physiques. Les points forts de cette méthode sont la simplicité de mise en œuvre dans les modèles numériques complexes, la faible demande de puissance de calcul et la régularité temporelle de la solution.

La disponibilité croissante de la puissance de calcul a permis l'utilisation de méthodes d'assimilation de données plus avancées. En général, ces méthodes reposent sur l'utilisation d'informations sur les statistiques des erreurs du modèle et des observations pour pondérer la combinaison modèle-observations. Deux de ces méthodes qui sont largement utilisées par les centres de prévision sont le filtre de Kalman d'ensemble-EnKF (Evensen, 1994) et ses variations (Pham, 2001; Hunt et al., 2007), et les méthodes variationnelles en quatre dimensions-4Dvar (Le Dimet and Talagrand, 1986; Courtier et al., 1994). Pour

la première, les coûts de calcul sont dûs à la propagation de l'ensemble, généralement formé par des centaines de membres, pour calculer la prévision. Pour la seconde, les coûts sont dûs à la nécessité de minimiser une fonction coût dans un très grand espace d'état (10^8 variables). Cela peut demander plusieurs itérations de l'algorithme de minimisation, ce qui en pratique se traduit par plusieurs intégrations des modèles directs et adjoints.

Cependant, même avec l'intérêt croissant pour ces techniques complexes construites sur des arguments théoriques solides, le Nudging n'a pas été laissé de côté. Des travaux récents ont utilisé le Nudging avec des méthodes plus avancées telles que l'interpolation optimale (Clifford et al., 1997; Wang et al., 2013a), l'EnKF (Ballabrera-Poy et al., 2009; Bergemann and Reich, 2010; Lei et al., 2012; Luo and Hoteit, 2012), 4DVAR (Zou et al., 1992; Stauffer and Bao, 1993; Vidard et al., 2003; Abarbanel et al., 2010) ou les filtres particulaires (Luo and Hoteit, 2013; Lingala et al., 2013) pour extraire le meilleur de chaque méthode. Dans le cas particulier de l'hybridation avec l'EnKF proposé par Lei et al. (2012), l'algorithme résultant a l'avantage de la propagation dynamique de la matrice de covariance à partir de l'EnKF et utilise le Nudging pour atténuer les problèmes liés à l'intermittence de l'approche séquentielle, qui entre autres choses entraîne le rejet possible de certaines observations.

Récemment, Auroux and Blum (2005) ont revisité la méthode du Nudging et ont proposé un nouvel algorithme appelé Back and Forth Nudging (BFN). Le BFN consiste en une séquence d'intégrations des modèles direct et rétrograde, tous les deux en utilisant un terme de rappel aux observations, comme pour le Nudging direct. Le BFN intègre le modèle direct vers l'arrière dans le temps en évitant la construction de l'adjoint et/ou des modèles linéaires tangents (nécessaire par le 4Dvar et les filtres étendus). Par conséquent, il n'utilise que le modèle non linéaire pour propager des informations vers l'avant et vers l'arrière dans le temps. Le gain du Nudging rétrograde, qui a un signe opposé par rapport au Nudging direct, a un double rôle: pousser le modèle vers les observations et stabiliser l'intégration rétrograde, ce qui est particulièrement important lorsque le modèle n'est pas réversible.

La performance du BFN dans les applications numériques en utilisant une variété de modèles, y compris les modèles non-réversibles, tels que le modèle de Saint-Venant ou

un modèle multi-couche quasi-géostrophique, ainsi qu'un modèle atmosphérique à haute résolution, sont très encourageants (Auroux, 2009; Boilley and Mahfouf, 2012). De plus, en utilisant un gain scalaire simple, le BFN a produit des résultats comparables à ceux obtenus avec le 4Dvar, mais avec un coût de calcul inférieur (Auroux, 2009; Auroux et al., 2012).

L'objectif principal de cette thèse est de mettre en œuvre le BFN pour un modèle d'océan aux équations primitives et de comparer sa performance avec celles des filtres/lisseurs de Kalman et du 4Dvar. En outre, nous avons utilisé la théorie de l'estimation bayésienne pour construire des gains optimaux pour le BFN. En utilisant les hypothèses de Kalman (à définir) nous avons derivé le Backward Smoother et sa version itérative, ainsi que le Back and Forth Kalman filter (Cosme, personal communication) qui peut être vu comme une extension du BFN pour les systèmes stochastiques.

Le modèle numérique utilisé est NEMO (Madec (2008)), actuellement utilisé par le centre français de prévision de l'océan, MERCATOR (http://www.mercator-ocean.fr/fre), pour produire et livrer des prévisions océaniques. La configuration du double tourbillon idéalisée à une résolution "eddy permiting" est utilisé. Cette configuration présente l'avantage d'être simple du point de vue de la géométrie et du forçage, mais avec toutes les caractéristiques des processus de méso-échelle d'un ocean de latitude moyenne.

Cette thèse est organisée comme suit :

- Le chapitre 2 présente les aspects théoriques concernant les méthodes d'assimilation de données utilisées dans la thèse. L'accent est mis sur leurs différences, car il va permettre l'analyse des résultats numériques produits.

- Le chapitre 3 présente le modèle numérique et ses approximations. Les aspects de la mise en œuvre du modèle rétrograde sont également présentés. Enfin, la mise en place des expériences, en ce qui concerne ses conditions initiales et aux limites, la paramétrisation sous-maille et les champs de forçage, est décrite.

- Le chapitre 4 présente une étude approfondie du BFN appliqué à un modèle d'océan aux équations primitives et compare les résultats avec ceux produits par un algorithme 4Dvar.

- Le chapitre 5 présente une mise en œuvre de deux nouveaux lisseurs: le Backward

Smoother et le Back and Forth Kalman filter. Leur performance est comparée avec la performance d'un filtre de Kalman et d'un lisseur de Kalman.

- Enfin, le dernier chapitre présente les conclusions et les perspectives d'avenir concernant le BFN et le Back and Forth Kalman Filter.

1.2 Introduction

Over the last century, the populational growth and the increasing demand for social welfare has led to an increasing exploitation of natural resources. Large clusters of human settlements have proved to be very sensitive to extreme natural phenomenas, being on the one hand directly affected as for example in the case of 2004 Indian Ocean tsunami and 2013 Japan tsunami, and on the other hand indirectly affected as a consequence of weather perturbations on the large scale food production.

Based on this, the knowledge of the Earth's physical system components such as the atmosphere and the ocean and their governing laws has become essential for sustaining the modern way of life. The possibility of reconstructing past events and accurately predict future events has changed the economic dynamics. For example, insurance and credit agencies have been using the knowledge about the weather as well as weather predictions to adjust their risk assessments.

In the last fifty years, progress on weather forecasting occurred especially due to improvements in the models formulation, as a consequence of the impressive development of computational resources, as well as due to improved observational networks and initial conditions. This progress in forecasting is illustrated in Figure 1.1 by skill of the 500 hPa forecasts produced at the NCEP[4]. The horizontal bar indicates an one-day-per-decade increase in forecast skill. To show how important this can be for the nowadays society, according to Ravinder Singh, head of agricultural physics at the Indian Agricultural Research Institute, "Precision in the monsoon forecast can raise farm production by 10 percent to 15 percent in rainfed areas". This means that improving forecast capabilities would represent increasing the world food production.

Much less attention has been given to the development of ocean forecasts. However, the ocean plays a key role in the weather and climate regimes since it bounds the lower atmosphere. At this interface, the ocean and the atmosphere exchange momentum, heat and water vapor. Furthermore, it is believed that ocean-atmosphere interactions become a dominant factor at long forecast ranges. For example, important atmospherical variabilities, which directly affect several millions of people, such as the Indian's monsoon and the pacific El Nino/La Nina phenomena are in high degree controlled by

[4]National Centers for Environmental Prediction

the upper ocean dynamics (Keshavamurthy and Sankar Rao, 1992; Wang et al., 2012).

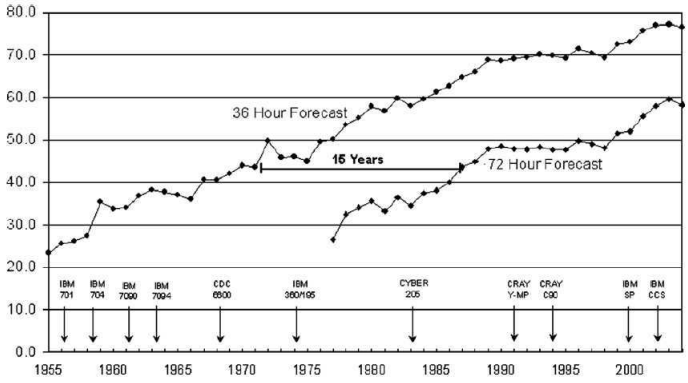


Figure 1.1: Skill of the 36 hour (1955–2004) and 72 hour (1977–2004) 500 hPa forecasts produced at NCEP. Forecast skill is expressed as a percentage of an essentially perfect forecast score. Extracted from Lynch (2008).

Moreover, ocean and coastal regions are natural resource-rich areas. This has led Human settlements to focus on near-coastal areas. As an example, the average populational density in coastal regions was estimated 3 times higher than the global average. In addition, worldwide trades mostly use the sea to export and import goods as well as fish account for roughly one fifth of all animal protein in the human diet.

Therefore, the importance of producing reliable forecasts of the ocean and the atmosphere is justified by, among other factors, their use in supporting economic activities as well as the environmental management. But, what are the challenges for producing reliable forecasts? In mathematical terms, a forecast is an initial value problem, i.e. we are given a set of evolutive differential equations from which we want to calculate the system state evolution given an initial state. Therefore, a good forecast depends on how representative of the real physics the mathematical (and numerical) model is and how accurate the initial conditions are.

In the beginning of the 60's the American scientist Edward Norton Lorenz stated that the chaotic nature of the governing equations imposes a finite time predictability limit. He founded that starting a simplified atmospheric model with slightly different initial conditions resulted in two completely different fields after two weeks simulation. Consequently, much more effort has been made for producing initial conditions as ac-

curate as possible.

The science of producing accurate geophysical fields has been named Data Assimilation - DA (Bennett, 2002; Kalnay, 2003; Evensen, 2009). This name comes from the fact that observations of the system are introduced (assimilated) into the models to improve both descriptions of the system: the observational and the mathematical/numerical. It means that assimilation aims to provide informations about the real system which are at the same time closer to the truth than observations and models separately.

To exemplify the challenges faced by ocean data assimilation, Figs.1.2 and 1.3 show a typical Mediterranean Sea observational network and two model outputs for the same region, one with a spatial resolution of 12km and the other of 3km. The first observed aspect is the observational network sparsity that reveals that the observations are not dense enough to produce initial conditions for numerical models. More than this, there is almost no direct information about the synoptic vertical mass field structure (the green bullets in the figure represent Argo[5] drifters that sample the water column every 10 days). Most of the data are direct[6] measurements of the ocean surface (sea surface temperature and sea surface height), even though they may "observe" thermocline-related motions (Wunsch, 1997).

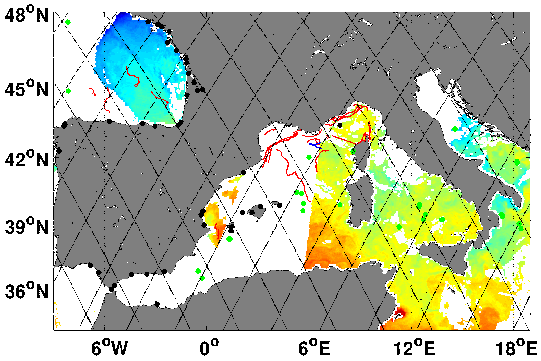


Figure 1.2: Typical Mediterranean Sea observational network for a 10day time window.

The model outputs (Fig. 1.3), in their turns, reveals that changing the model resolu-

[5]Argo consists of a large collection of small, drifting oceanic robotic probes deployed worldwide. The probes float as deep as 2 km. Once every 10 days, the probes surface, measuring conductivity and temperature profiles to the surface. The data are transmitted to scientists on shore via satellite. (source http://en.wikipedia.org/wiki/Argo_%28oceanography%29)

[6]It does not mean *in situ* measurements but that the measured properties are at the surface or is the surface itself.

tion alters the large scale features. This may be the impact of a poor representation of the sub-grid scale process, i.e. the parametrization of the non-modeled physics. However, models provide continuous fields which at some extent reproduce the underlying physics. Hence, DA methods may be seen as methods to interpolate the observations using the model as the interpolation function.

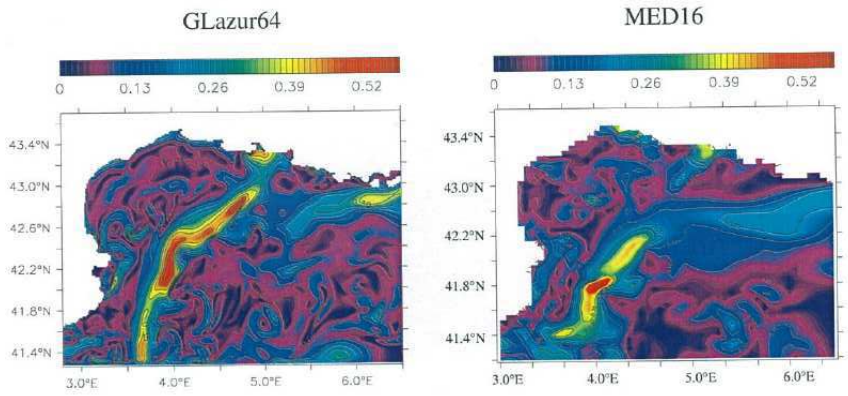


Figure 1.3: Surface velocity for experiments employing the same forcing fields but on the right using spatial resolution of 1.7km (Glazur64) and on the left of 7km (MED16). Figures adapted from Duchez (2011).

The DA science had its beginning based on two pillars: the control theory (Jazwinski, 1970; Gelb, 1974) and the optimal control theory (Lions, 1971). These domains are especially interested in the control of electric systems, ballistic and satellite trajectory, for instance. From these two scientific branches emerged the nowadays most used data assimilation methods: the Kalman Filters (Kalman, 1960; Kalnay, 2003; Evensen, 2009) and the four dimensional variational methods (Le Dimet and Talagrand, 1986; Bennett, 2002). However, due to the very huge size of the problems and the limited computational resources available, data assimilation had its beginning, in the mid-70s, with the use of simple and inexpensive methods, among which stands out the Nudging.

The well-known Nudging method is based on the Newton's second axiom and consists in adding a forcing term in the right hand side of a given system in order to gently push the model toward a prescribed value. The first appearance of Nudging in the geophysical literature was in 1974 (Anthes, 1974). In this work the authors proposed to use nudging to mitigate initialization problems in atmospheric models.

The first appearance of a successful application of Nudging to oceanographic DA was in 1992 in a work that assimilated sea surface height derived from satellite measurements into a quasi-geostrophic layered model (Verron, 1992). Since then, the method has been successfully applied to several oceanographic numerical problems such as boundary conditions (Marchesiello et al., 2001; Chen et al., 2013), downscaling (Li et al., 2012), and other DA problems (Verron, 1992; Haines et al., 1993; Lewis et al., 1998; Killworth et al., 2001; Thompson et al., 2006). With respect to applications to DA problems, the weights given to the model and the observations are in general not based on any optimality condition, but are rather scalars or Gaussian-like functions constructed based on physical assumptions. The appeals of this method are the simplicity of implementation in complex numerical models, the low computational power required and the time smoothness of the solution.

The increasing availability of computing power has allowed the use of more advanced data assimilation methods. In general, these methods rely on the use of information about the model statistics and observations errors to weight the model-observations combination. Two of these methods that are widely used by prediction centers are the ensemble Kalman filter- EnKF (Evensen, 1994) and its variations (Pham, 2001; Hunt et al., 2007), and the four dimensional variational method 4Dvar (Le Dimet and Talagrand, 1986; Courtier et al., 1994). For the first, the numerical costs are due to the propagation of the ensemble, usually formed by hundreds of members, to calculate the forecast. For the second, the costs are due to the need of minimizing a cost function in a very huge state space (10^8 variables). This may require several iterations of the minimization algorithm, which in practice requires several integrations of the direct and adjoint models.

However, even with the growing interest in these complex techniques built on solid theoretical arguments, nudging has not been left aside. Recent works have used nudging along with more advanced methods such as Optimal interpolation (Clifford et al., 1997; Wang et al., 2013a), EnKF (Ballabrera-Poy et al., 2009; Bergemann and Reich, 2010; Lei et al., 2012; Luo and Hoteit, 2012), 4Dvar (Zou et al., 1992; Stauffer and Bao, 1993; Vidard et al., 2003; Abarbanel et al., 2010) or particle filters (Luo and Hoteit, 2013; Lingala et al., 2013) to extract the best of each method. In the particular case of

the hybridization with the EnKF proposed by Lei et al. (2012), the resulting algorithm takes the advantage of the dynamical propagation of the covariance matrix from EnKF and uses nudging to mitigate problems related to the intermittence of the sequential approach, which among other things entails the possible discarding of some observations.

Recently, Auroux and Blum (2005) revisited the Nudging method and proposed a new algorithm called Back and Forth Nudging (BFN). The BFN consists in a sequence of forward and backward model integrations, both of them using a feedback term to the observations, as in the direct nudging. The BFN integrates the direct model backwards in time avoiding the construction of the adjoint and/or tangent linear models (needed by the 4DVar and extended filters). Therefore, it uses only the fully non-linear model to propagate information forward and backward in time. The nudging gain, which has an opposite sign with respect to the forward case, has a double role: push the model toward observations and stabilize the backward integration, which is especially important when the model is not reversible.

The BFN performance in numerical applications using a variety of models, including non-reversible models such as a shallow water model and a multi-layer quasi-geostrophic model as well as a high resolution atmospherical model, are very encouraging (Auroux, 2009; Boilley and Mahfouf, 2012). Moreover, using a simple scalar gain, it produced results comparable to those obtained with 4DVar but with lower computational requirements (Auroux, 2009; Auroux et al., 2012).

The main objective of this thesis is to implement the BFN for a primitive equation ocean model and to compare its performance with Kalman filters/smoother and the 4Dvar. In addition, we used the Bayesian estimation theory to construct optimal gains for the BFN. Using the Kalman's hypotheses (to be defined latter) we derived the Backward Smoother and its iterative version as well as the Back and Forth Kalman Filter (Cosme, personn. communication) that can be seen as BFN extensions for stochastic systems.

The numerical model used is NEMO (Madec (2008)), currently used by the french ocean forecast center, MERCATOR (http://www.mercator-ocean.fr/fre), to produce and deliver ocean forecasts. The well-known idealized double gyre configuration at eddy-permitting resolution is used. This configuration has the advantage of being sim-

ple from the geometry and forcing point of view but with all the mid latitude ocean mesoscale process features.

This thesis is organized as follows:

- Chapter 2 introduces the theoretical aspects regarding the data assimilation methods used in the thesis. The focus is on their differences, as it will feed the discussion of the numerical results produced.
- Chapter 3 presents the numerical model and its approximations. Aspects of the backward model implementation are also presented. Finally, the experiment set-up regarding its initial and boundary conditions, subgrid-scale parametrization and forcing fields is described.
- Chapter 4 presents a comprehensive study of the BFN applied for a primitive equation ocean model and compares the results with those produced by a 4Dvar algorithm.
- Chapter 5 presents an implementation of two "new" smoother algorithms: the Backward Smoother and the Back and Forth Kalman Filter. Their performance are compared with a Kalman Filter and a Kalman Smoother performance.
- Finally, the last chapter presents the conclusions and future prospects concerning the BFN and the Back and Forth Kalman Filter development.

Chapter 2

Data Assimilation Methods

Contents

2.1 Introduction

This chapter introduces and presents the assimilation methods used in this thesis. The main objective is to draw the similarities and differences between the methods allowing for discussions regarding the numerical results obtained.

The presented algorithms may be divided into two classes: deterministic and stochastic ones. For the deterministic algorithms, the system for which they were derived does not have random forcing, while for the stochastic ones it does. The presented deterministic algorithms are the 4Dvar (Le Dimet and Talagrand, 1986), the Nudging (Anthes, 1974) or Luenberger Observer (Luenberger, 1966) and the Back and Forth Nudging (Auroux and Blum, 2005). The stochastic algorithms are all based on the Kalman Filter and its extended versions as well as reduced rank implementations (Evensen, 2009).

First, the deterministic algorithms are presented. Then, some basics about Bayesian estimation and Kalman filtering are presented and subsequently a reinterpretation of the BFN algorithm following the Bayesian reasoning is given. In this case, it is shown that the BFN is a suboptimal iterative smoother that can be extended to be optimal when the gains are calculated using the Kalman filter formulas.

2.2 Preliminary Concepts

In this section, some preliminary concepts needed to present the assimilation methods are given. They are basically concepts of *stability* and *observability*. Let us consider the linear system given by:

$$\begin{cases} \dfrac{d\mathbf{x}(t)}{dt} = \mathbf{F}\mathbf{x}(t) \quad t \in \mathbb{R} \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \tag{2.1}$$

where $\mathbf{x} \in \mathbb{R}^n$ represents the system's state vector and $\mathbf{F} \in \mathbb{R}^{n \times n}$ a linear operator. The system solution is given by:

$$\mathbf{x}(t) = e^{\mathbf{F}t}\mathbf{x_0} \tag{2.2}$$

2.2.1. Definition. *The origin is stable in the sense of Lyapunov if for all* $\epsilon > 0$ *there*

exists $\delta > 0$ *such that for all* $\mathbf{x_0} \in N_\delta = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < \delta\}$ *and* $t > 0$

$$e^{\mathbf{F}t}\mathbf{x_0} \in N_\epsilon = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < \epsilon\}$$

Indeed, the system is said to be asymptotically stable if it is stable at the origin and if there exists $\delta > 0$ such that for all $\mathbf{x_0} \in N_\delta$

$$\lim_{t\to\infty} e^{\mathbf{F}t}\mathbf{x_0} = 0$$

2.2.2. Theorem. *The system (2.1) is asymptotically stable if and only if all eigenvalues of* $\mathbf{F}$ *have negative real part.*

Now lets consider the system:

$$\begin{cases} \dfrac{d\mathbf{x}(t)}{dt} = \mathbf{F}\mathbf{x}(t) \quad t \in \mathbb{R} \\ \mathbf{x}(0) = \mathbf{x}_0 \\ \mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) \end{cases} \tag{2.3}$$

where $\mathbf{H} \in \mathbb{R}^{n\times m}$ is the measurement model.

2.2.3. Definition. *The pair* $(\mathbf{F}, \mathbf{H})$*, or the system (2.3), is said to be observable if for an arbitrary* $\mathbf{x}_0 \in \mathbb{R}^n$*,* $\mathbf{x}(0) = \mathbf{x}_0 \neq 0$*, there exists a* $t > 0$ *such that*

$$\mathbf{y}(t) = \mathbf{H}e^{\mathbf{F}t}\mathbf{x}(0) \neq 0$$

If for a given $T > 0$ *and for arbitrary* $\mathbf{x}(0) \neq 0$ *there exists* $t \in [0, T]$ *with the above property, then the pair* (F, H) *is said to be observable at time* T*,or in other words, given* $\mathbf{y}$ *as an absolutely continuous function on* $[0, T]$ *it is possible to determine* $\mathbf{x}(0)$ *uniquely.*

In other words, the observability condition says that all the information about the dynamical behavior of the state space variables defined in (2.3) can be retrieved by using only information from the output measurements. Since the n-dimensional vector $\mathbf{x}_0$ has n unknown components, it is expected that n measurements are sufficient to determine $\mathbf{x}(0)$. For this purpose, one may take n derivative of the measurement equation to

generate the following sequence:

$$\begin{aligned} \mathbf{y}(0) &= \mathbf{H}\mathbf{x}(0) \\ \mathbf{y}^{'}(0) &= \mathbf{H}\mathbf{x}^{'}(0) = \mathbf{H}\mathbf{F}\mathbf{x}(0) \\ \mathbf{y}^{''}(0) &= \mathbf{H}\mathbf{x}^{''}(0) = \mathbf{H}\mathbf{F}^2\mathbf{x}(0) \\ &\vdots \\ \mathbf{y}^{n-1}(0) &= \mathbf{H}\mathbf{x}^{n-1}(0) = \mathbf{H}\mathbf{F}^{n-1}\mathbf{x}(0) \end{aligned} \tag{2.4}$$

Hence, a system is obtained from which $\mathbf{x}(0)$ may be uniquely calculated from the observations if and only if the matrix

$$\mathbf{O} = \begin{bmatrix} \mathbf{H} \\ \mathbf{H}\mathbf{F} \\ \vdots \\ \mathbf{H}\mathbf{F}^{n-1} \end{bmatrix}$$

is non-singular. This matrix is called observability matrix.

2.2.4. Theorem (*Kalman condition*). *The pair* $(\mathbf{F}, \mathbf{H})$ *is observable if and only if*

$$rank(\mathbf{O}) = n$$

Now lets consider a system with a control $\mathbf{B}\mathbf{u}(t)$

$$\begin{cases} \dfrac{d\mathbf{x}(t)}{dt} = \mathbf{F}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad t \in \mathbb{R} \\ \mathbf{x}(0) = \mathbf{x}_0 \\ \mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) \end{cases} \tag{2.5}$$

where $\mathbf{B} \in \mathbb{R}^{n \times m}$ and $\mathbf{u}(t) = \mathbf{K}\mathbf{x}(t)$ is the control.

2.2.5. Definition. *The characteristic polynomial of* $\mathbf{F} \in \mathbb{R}^{n \times n}$ *is defined as:*

$$\chi_F(t) = det\,(t\mathbf{I} - \mathbf{F})$$

2.2.6. Theorem (*pole-shifting theorem*). *If* $(\mathbf{F}, \mathbf{H})$ *verifies the Kalman condition then for all unitary real polynomial* $\mathbf{P}$ *of degree* n *there exist* $\mathbf{K} \in \mathbb{R}^{m\times n}$ *such that* $\chi_{\mathbf{F}+\mathbf{BK}} = \mathbf{P}$*, i.e. the characteristic polynomial of* $\mathbf{F}+\mathbf{BK}$ *is equal* $\mathbf{P}$*.*

The theorem 2.2.6 says that if the pair $(\mathbf{F}, \mathbf{H})$ is observable it is possible to place the poles, i.e. the eigenvalues, of the system wherever we want.

2.3 Four Dimensional Variational Method - 4Dvar

The objective of the variational methods is to minimize a cost function that measures the distance between the estimated state and the available observations. Let us assume that observations are available at every instant $(t_i)_{1\leq i\leq N}$. Given a first guess $\mathbf{x}^b$ of the initial state, the 4DVar algorithm will find an optimal initial condition that minimizes the distance between the model trajectory and the observations in a given assimilation window. This optimal state is found by minimizing the following cost function:

$$\begin{aligned} J(\mathbf{x}_0) = &\frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) \\ &+ \frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))^T\mathbf{S}_i^{-1}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)) \end{aligned} \tag{2.6}$$

subject to the model constraint: $\frac{d\mathbf{x}(t)}{dt} = \mathcal{F}(\mathbf{x}(t))$ and initial condition $\mathbf{x}(0) = \mathbf{x}_0$. In this cost function, $\mathbf{B}$ and $\mathbf{S}$ are two positive definite matrices which weight the background and the observation influence on the analysed field, and $\mathcal{H}_i$ and $\mathbf{y}_i$ are the observation operator and the available observations at time t_i, respectively. The optimal initial state is found by solving:

$$\nabla J(\mathbf{x}^a(t_0)) = 0 \tag{2.7}$$

The associated Lagrangian of the system is given by:

$$\mathcal{L}(\mathbf{x}, \mathbf{x}_0, \mathbf{p}) = J(\mathbf{x}, \mathbf{x}_0) + \int_0^T \mathbf{p}(t).(\frac{d\mathbf{x}}{dt} - \mathcal{F}(\mathbf{x}(t)))\mathrm{dt} \tag{2.8}$$

where $\mathbf{p}$ is called Lagrange multiplier or adjoint variable.

The theorem 2.3.1 states the link between finding the minimum of $J(\mathbf{x}_0, \mathbf{x})$ subject to the model constraint and finding the saddle point of $\mathcal{L}(\mathbf{x}, \mathbf{x}_0, \mathbf{p})$, i.e find the point maximizing $\mathcal{L}$ with respect to $\mathbf{p}$ and minimizing $\mathcal{L}$ with respect to $(\mathbf{x}, \mathbf{x}_0)$.

2.3.1. Theorem. *If* $(\mathbf{x}, \mathbf{x}_0, \mathbf{p})$ *is a saddle point of* $\mathcal{L}$ *then* $\mathbf{x}_0$ *gives the minimum of* J *under the model constraint.*

Assuming that the functional $\mathcal{L}$ is convex with respect to $\mathbf{x}$, the saddle point is the one for which all partial derivatives are null:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{p}} = \frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \frac{\partial \mathcal{L}}{\partial \mathbf{x_0}} = 0 \tag{2.9}$$

The derivative of the Lagrangian with respect to $\mathbf{p}$ gives the model equations which should be verified since the model is a strong constraint.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{p}} = \frac{d\mathbf{x}}{dt} - \mathcal{F}(\mathbf{x}) = 0 \tag{2.10}$$

To calculate the derivative of the Lagrangian with respect to $\mathbf{x}$ it is preferable to performe an integration by parts which will provide the derivative of $\mathbf{p}$ with respect to t instead of the derivative of $\mathbf{x}$ with respect to t in the expression (2.8). The modified Lagrangian is given by:

$$\mathcal{L}(\mathbf{x}, \mathbf{x}_0, \mathbf{p}) = J(\mathbf{x}, \mathbf{x}_0) - \int_0^T \mathbf{x}(t).\frac{d\mathbf{p}}{dt}\mathrm{dt} - \int_0^T \mathbf{p}(t).(\mathcal{F}(\mathbf{x}))\mathrm{dt} + [\mathbf{p}(t)\mathbf{x}(t)]_0^T \tag{2.11}$$

Then, to obtain the condition:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 0 \tag{2.12}$$

it is required that $\mathbf{p}$ satisfies the adjoint model:

$$-\frac{d\mathbf{p}(t)}{dt} = [\mathcal{F}'(\mathbf{x}(t))]^T\mathbf{p}(t) + \sum_{i=1}^{N} \mathcal{H}_i'(\mathbf{x}_i)^T\mathbf{S}^{-1}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))\delta_{t_i}(t) \tag{2.13}$$

with final condition $\mathbf{p}(T) = 0$. In the Eq.(2.13), $\mathcal{F}'$ and $\mathcal{H}_i'$ are the tangent linear aproximation of the operators $\mathcal{F}$ and $\mathcal{H}_i$, respectively. Finally the derivative $\mathcal{L}$ with respect to $\mathbf{x_0}$ is easy to calculate:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x_0}} = \mathbf{B}^{-1}(\mathbf{x_0} - \mathbf{x^b}) - \mathbf{p}(0) \tag{2.14}$$

Therefore, if the model constraint is satisfied one has $J(\mathbf{x_0}) = \mathcal{L}(\mathbf{x}, \mathbf{x_0}, \mathbf{p})$ and hence

$$\nabla J(\mathbf{x_0}) = \frac{\partial \mathcal{L}}{\partial \mathbf{x_0}} = \mathbf{B}^{-1}(\mathbf{x_0} - \mathbf{x^b}) - \mathbf{p}(0)$$

Consequently, to calculate the gradient of J it is required to integrate the adjoint model backward in time with final conditions $\mathbf{p}(T) = 0$ to obtain $\mathbf{p}(0)$. The advantage of using the adjoint method instead of finite differences to calculate the gradient of J is the relatively lower numerical cost required by the former when the system is very large. In this case, the evaluation of the gradient using finite differences would require a number of model integrations equal to the size of the control space, while using the adjoint method only one adjoint model integration is needed.

If $\mathcal{H}$ or $\mathcal{F}$ are non-linear, the solution of the problem is not unique, i.e. the functional (2.6) may have multiple local minimum, and the minimization procedure may not stop at the global minimum. To overcome this problem, Courtier et al. (1994) proposed to solve a sequence of quadratic problems, expecting this sequence would converge to the solution of the problem given by (2.6) and (2.7). This algorithm is called the incremental 4Dvar. In this case the cost function will not be minimized with respect to the initial state but with respect to an increment $\delta\mathbf{x}_0$ defined by $\mathbf{x}_0 = \mathbf{x}^b + \delta\mathbf{x}_0$. The operators are linearized in a neighborhood of $\mathbf{x}^b$ as:

$$\mathcal{M}_{0,i}(\mathbf{x}^b + \delta\mathbf{x}_0) \approx \mathcal{M}_{0,i}(\mathbf{x}^b) + \mathcal{M}'_{0,i}(\mathbf{x}^b)\delta\mathbf{x}_0 \quad \forall i \tag{2.15}$$

$$\mathcal{H}_i(\mathbf{x}^b + \delta\mathbf{x}_0) \approx \mathcal{H}_i(\mathbf{x}^b) + \mathcal{H}'_i(\mathbf{x}^b)\delta\mathbf{x}_0 \quad \forall i \tag{2.16}$$

where $\mathcal{M}_{0,i}$ is the discrete form or the resolvent of the model $\mathcal{F}$. The new cost function may be written as:

$$\begin{aligned} J(\delta\mathbf{x}_0) = &\frac{1}{2}\delta\mathbf{x}_0^T\mathbf{B}^{-1}\delta\mathbf{x}_0 \\ &+ \frac{1}{2}\sum_{i=0}^{N}(\mathbf{d}_i - \mathbf{H}_i\mathbf{M}_{0,i}\delta\mathbf{x}_0)^T\mathbf{S}_i^{-1}(\mathbf{d}_i - \mathbf{H}_i\mathbf{M}_{0,i}\delta\mathbf{x}_0) \end{aligned} \tag{2.17}$$

where $\mathbf{d}_i = \mathbf{y}_i - \mathcal{H}_i(\mathcal{M}_{0,i}(\mathbf{x}^b))$ is called the innovation vector, and $\mathbf{M}_{0,i}$ and $\mathbf{H}_i$ are the tangent linear models. It is possible that after some iterations of the minimizer the increments become too large and a new linearization of $\mathcal{H}$ and $\mathcal{M}$ should be done. This gives rise to what is called the inner loop and outer loop iterations. The algorithm implemented in NEMO, called NEMOVAR (Mogensen et al. (2009)), uses these techniques. See below a simplified description of this algorithm:

–**Initialization** : $\mathbf{x}_0^0 = \mathbf{x}^b$
–**While** $k \leq k_{max}$ or $\|\delta\mathbf{x}_0^{a,k}\| > \epsilon$ (**Outer Loop**)
Do
•$\mathbf{d}_i^k = \mathbf{y}_i - \mathcal{H}_i(\mathcal{M}_{0,i}(\mathbf{x}_0^k))$
•Search the $\delta\mathbf{x}_0^{a,k}$ that minimizes (**Inner Loop**):

$$J(\delta\mathbf{x}_0^k) = \frac{1}{2}(\delta\mathbf{x}_0^k)^T\mathbf{B}^{-1}(\delta\mathbf{x}_0^k) + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{d}_i^k - \mathbf{H}_i\mathbf{M}_{0,i}\delta\mathbf{x}_0^k)^T\mathbf{S}_i^{-1}(\mathbf{d}_i^k - \mathbf{H}_i\mathbf{M}_{0,i}\delta\mathbf{x}_0^k)$$

•$\mathbf{x}_0^{k+1} = \mathbf{x}_0^k + \delta\mathbf{x}_0^{a,k}$

Table 2.1: Summary of the 4Dvar algorithm

2.4 The Back and Forth Nudging

2.4.1 Forward Nudging

The Forward Nudging better known simply by Nudging is based on the Newton's second axiom and consists in adding a forcing term proportional to the difference between the current state and the observations to a given system. For atmospherical sciences its interest stood in reducing initialization problems in weather forecasting (Anthes, 1974) because it permits a "soft" introduction of a new analyzed field into the model. Although not cited in the meteorological-related articles, a similar algorithm was created by Luenberger in 1966 to estimate a partially observed system governed by linear dynamics. Therefore, the general Nudging method reduces to the Luenberger Observer (LO) in the case of linear systems. Indeed, the LO is an asymptotic observer with well-known properties. A brief derivation of the LO is presented below.

Considering a system:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}\mathbf{x} \tag{2.18}$$

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{2.19}$$

where $\mathbf{x} \in \mathbb{R}^n$ represents the system's state vector, $\mathbf{F} \in \mathbb{R}^{n\times n}$ is a linear model, and $\mathbf{H} \in \mathbb{R}^{n\times m}$ is the measurement model. The goal is to create a system that reproduces the dynamics of the original system (Eq.2.18) using observations (outputs) of the system

(Eq.2.19). The observer is defined as:

$$\frac{d\hat{\mathbf{x}}}{dt} = \mathbf{F}\hat{\mathbf{x}} + \mathbf{K}(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}) \tag{2.20}$$

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{2.21}$$

where $\mathbf{K} \in \mathbb{R}^{n\times m}$ is a gain matrix such that:

$$\forall \mathbf{x}(0), \hat{\mathbf{x}}(0) \in \mathbb{R}^n \quad \hat{\mathbf{x}}(t) - \mathbf{x}(t) \xrightarrow[t\to+\infty]{} 0$$

Therefore, the method seeks to control the asymptotic error ($\mathbf{e}(t) = \hat{\mathbf{x}}(t) - \mathbf{x}(t)$) behavior. This may be done by deducing an equation for the observer error evolution. The resulting equation, obtained subtracting Eq.(2.20) from Eq.(2.18), is given by:

$$\frac{\mathrm{d}\mathbf{e}}{\mathrm{d}t} = (\mathbf{F} - \mathbf{K}\mathbf{H})\,\mathbf{e} \tag{2.22}$$

From the theorem (2.2.2) we know that

$$\mathbf{e} \to 0 \quad \forall\, \mathbf{e}(0) \iff \forall \lambda \in Spec\,(\mathbf{F} - \mathbf{K}\mathbf{H}) \quad Re\lambda < 0,$$

i.e. the LO converges if and only if $\mathbf{F} - \mathbf{K}\mathbf{H}$ is a Hurwitz matrix (Trélat, 2013). Therefore, the pole assignment theorem (Trélat, 2013) ensures that if $(\mathbf{F}, \mathbf{H})$ is observable there exists at least one matrix $\mathbf{K}$ that stabilizes the system.

For the systems in which $(\mathbf{F}, \mathbf{H})$ is not observable, it is still possible to achieve asymptotic stability. This depends on the spectrum of $\mathbf{F}$, which can be divided into observable modes and unobserved modes. If all the unobserved modes are on the left half plane, then there always exists a $\mathbf{K}$ that stabilizes the system.

Auroux and Blum (2008) also derived a variational interpretation of Nudging. They showed that the numerical solution of Eq.(2.20) using $\mathbf{K} = \mathbf{H}^T\mathbf{S}^{-1}$, where $\mathbf{S}$ is a positive definite matrix, and an implicit discretization is the solution of the following optimization problem:

$$\begin{aligned}\arg\min_{\mathbf{x}} \Bigg[&\frac{1}{2} < \mathbf{x} - \mathbf{x}^n, \mathbf{x} - \mathbf{x}^n > - \frac{\Delta t}{2} < \mathbf{F}\mathbf{x}, \mathbf{x} > \\ &+ \frac{\Delta t}{2} < \mathbf{S}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}), (\mathbf{y} - \mathbf{H}\mathbf{x}) > \Bigg]\end{aligned} \tag{2.23}$$

Hence at each time step, Eq.(2.20) produces a state that is a compromise between the minimization of the system's energy, given by the first two terms of the cost function, and the minimization of the weighted distance between data and model, given by the last term in the cost function, which is a sort of penalty term. Therefore, unlike in the 4Dvar method for which the model is a strong constraint, in the Nudging the model appears as a weak constraint.

In addition, if $\mathbf{S}$ expresses a confidence measure to have in the observations, i.e the smaller the spectral norm of $\mathbf{S}$, the greater the confidence, we see that as $\mathbf{S} \to 0$, $\mathbf{K} \to \infty$, i.e. the method becomes a direct insertion of the data into the model. Direct insertion should be avoided because of the severe disruption of the system physical balance, which may degrade the forecasts. Optimal Nudging gains were proposed by Zou et al. (1992); Stauffer and Bao (1993); Vidard et al. (2003). In these studies the Nudging gain is treated as a parameter to be estimated. This is achieved by minimizing a modified version of the cost function (2.6):

$$\begin{aligned}\underset{\mathbf{x},\mathbf{K}_i}{\arg\min} \Bigg[& \frac{1}{2} < \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_0^b), \mathbf{x}-\mathbf{x}_0^b > + \frac{1}{2}\sum_{i=0}^{n} < \mathbf{S}_i^{-1}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i), (\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i) > \\ & + \frac{1}{2}\sum_{i=1}^{n} < \mathbf{Q}_i^{-1}\mathbf{K}_i(\mathbf{y}_i - \mathbf{H}_i(\mathcal{M}_{t_i,t_{i-1}}(\mathbf{x}_{i-1}))), \mathbf{K}_i(\mathbf{y}_i - \mathbf{H}_i(\mathcal{M}_{t_i,t_{i-1}}(\mathbf{x}_{i-1}))) > \Bigg]\end{aligned} \tag{2.24}$$

In this case, the model represents a weak constraint since the Nudging term penalizes/regularizes the model-observation fitting. The difficulty of this method is the poor knowledge of the $\mathbf{Q}_i$ matrices and their size. It will be shown later that when the system given by Eq.(2.18) has Gaussian stochastic forcing terms an optimal $\mathbf{K}$ may be given by the Kalman gain (Gelb, 1974).

There also exist some theories for the construction of nonlinear observers. One possible choice involves considering linearized errors and the Lyapunov stability method (Isidori, 1995). Although stability may be proven there is no constructive procedure for determining a stabilizing gain $\mathbf{K}$. Therefore in practice, the choice of $\mathbf{K}$ is a trial and error process. This is what the geophysical community has been doing for the Nudging. Another possibility for nonlinear observer design is to use Lie-algebraic techniques (Isidori, 1995). The advantage of these techniques is that one tries to reduce the non-

linear observer problem to one that can be handled by linear techniques, similar to the one we presented above. However, they are based on nonlinear state transformations that are not trivial to be found.

In the following we restrict our presentation to the Backward Nudging and the Back and Forth Nudging in the linear case.

2.4.2 Backward Nudging

Considering the following observable system:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}\mathbf{x}(t) \quad t \in [0,T] \tag{2.25}$$

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) \quad t \in [0,T] \tag{2.26}$$

and using the following relations:

$$\mathbf{x}^b(t) = \mathbf{x}(T-t) \quad t \in [0,T] \tag{2.27}$$

$$\mathbf{y}^b(t) = \mathbf{y}(T-t) \quad t \in [0,T] \tag{2.28}$$

the corresponding backward system may be written as:

$$\frac{d\mathbf{x}(t)^b}{dt} = -\mathbf{F}\mathbf{x}^b(t) \quad t \in [0,T] \tag{2.29}$$

$$\mathbf{y}^b(t) = \mathbf{H}\mathbf{x}^b(t) \quad t \in [0,T] \tag{2.30}$$

Accordingly, the backward LO estimator is given by:

$$\frac{d\hat{\mathbf{x}}(t)^b}{dt} = -\mathbf{F}\hat{\mathbf{x}}^b(t) + \mathbf{K}^b(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}^b) \quad t \in [0,T] \tag{2.31}$$

$$\mathbf{y}^b = \mathbf{H}\mathbf{x}^b \quad t \in [0,T] \tag{2.32}$$

and the corresponding equation for the observer error evolution by:

$$\frac{\mathrm{d}\mathbf{e}^b}{\mathrm{d}t} = -\left(\mathbf{F} + \mathbf{K^b H}\right)\mathbf{e}^b \tag{2.33}$$

Thus, since the observability condition is satisfied, the pole assignment theorem ensures that there exists at least one matrix $\mathbf{K}$ that stabilizes the system; i.e. there is a matrix $\mathbf{K}$ such that $\mathbf{e}^b(t) \to 0$ when $t \to +\infty$ for all $\mathbf{e}^b(0)$.

2.4.3 Iterating the Forward and the Backward Nudging

It was shown that the Nudging is an asymptotic estimator, i.e. it converges for infinite time. For geophysical systems, however, one may be interested in finite time processes, which means that it would be interesting to derive an observer that converges for bounded time. Auroux and Blum (2005) proposed an iterative estimator substituting the infinity time condition by an infinity number of iterations within a bounded time domain. Their algorithm, called Back and Forth Nudging, sequentially solves the forward and backward nudging equations. The initial condition of the backward integration is the final state obtained after integration of the forward nudging equation.

The BFN system may be written as:

$$k \geq 1 \begin{cases} \dfrac{d\hat{\mathbf{x}}_k(t)}{dt} = \mathbf{F}\hat{\mathbf{x}}_k(t) + \mathbf{K}(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_k) & t \in [0,T] \\ \hat{\mathbf{x}}_k(0) = \hat{\mathbf{x}}^b_{k-1}(T) \end{cases}$$

$$k \geq 1 \begin{cases} \dfrac{d\hat{\mathbf{x}}_k(t)^b}{dt} = -\mathbf{F}\hat{\mathbf{x}}^b_k(t) + \mathbf{K}(\mathbf{y}^b - \mathbf{H}\hat{\mathbf{x}}^b_k) & t \in [0,T] \\ \hat{\mathbf{x}}^b_k(0) = \hat{\mathbf{x}}_k(T) \end{cases}$$

where k stands for iterations. Therefore, at the end of each iteration one obtains a new estimation of the system's initial state. According to Ahmad Ali (*personal communication, 2012*) if $\|e^{-(\mathbf{F}+\mathbf{KH})T}e^{(\mathbf{F}-\mathbf{KH})T}\| < 1$ then $\hat{\mathbf{x}}_k(0)$ converges when $k \to \infty$ and more generally if $T > 0$ for any $t \in [0,T]$, $\hat{\mathbf{x}}_k(t)$ converges when $k \to \infty$. This means that not only the initial condition but the system trajectory converges to a limit trajectory. Indeed, under the hypotheses of $\mathbf{H} = \mathbf{Id}$, i.e the system is fully observed, if $\mathbf{K} \to \infty$, which means that $min(Spec(\mathbf{K})) \to \infty$, then $\hat{\mathbf{x}}_\infty(t) \to \mathbf{x}(t)$, i.e. the estimated trajectory converges to the "true" trajectory. Moreover, Auroux and Blum (2005) observed that the function $\hat{\mathbf{x}}_\infty(t)$ is totally independent of the initial condition $\hat{\mathbf{x}}(0)$. This is especially important since with an initial condition completely different from the true state the estimated trajectory converges to the true trajectory, that is, all the informative content of the observations can be used without the need for additional information on the structure and /or statistics of the background error, as it is the case for the Kalman

Filters.

Therefore, the BFN novelty with respect to conventional nudging methods is the model integration backwards in time. This allows to recover initial conditions using "future" observations as well as to use more than once the same observations set. Consequently, the BFN may be seen as a sub-optimal iterative smoother. The iteration process makes possible the use of gains relatively weaker than those required by conventional nudging, thus enabling the conservation of physical constraints without affecting the estimator response.

Ramdani et al. (2010) extended the BFN to infinite-dimentional systems and proved the convergence for the reversible wave and Schrodinger equations. However, the backward integration is problematic when the model is diffusive or simply not reversible. Auroux and Nodet (2012) studied the BFN convergence for viscous and non-viscous linear transport and Burgers equations. They showed that for non viscous equations (both linear transport and Burgers), the convergence of the algorithm holds under observability conditions. Convergence can also be proved for viscous linear transport equations under some strong hypothesis (i.e. system should be fully observed in space, at least for a subinterval of the time period), but not for viscous Burgers' equation. Despite the theoretical convergence not proven for viscous Burgers' equation, numerical studies (Auroux et al., 2012) showed that the BFN is effective in reducing the system initial condition errors.

In the case of ocean models, there are two main aspects requiring the inclusion of diffusion: i) the control of numerical noise, and ii) the modeling of sub grid-scale processes, i.e. to parameterize the energy transfer from explicitly resolved to non-resolved scales. Indeed, diffusion naturally represents a source of uncertainty in ocean forecasts, even for the purely forward model, and has been investigated from the point of view of the optimal control theory in Leredde et al. (1999).

Still with respect to the backward model diffusion, the loss/gain of energy in the forward integration should be compensated by an increase/decrease of energy in the correct spectral band, in the backward integration, without affecting numerical stability. This would be achieved by considering anti-diffusion (backward integration) followed by the application of a spectral filter in a similar way to Large Scale Eddy Simulation, or con-

structing an order N generalized diffusion operator, $\sum_{n=1}^{N}(-i)^n\nabla^n$ that could provide the expected response for the length scale of interest and suppress numerical noise. We are currently investigating both possibilities, but this remains beyond the scope of this thesis.

Meanwhile in this thesis, the Diffusive Back and Forth Nudging (DBFN, Auroux et al., 2011) is used, for which the sign of the diffusion term remains physically consistent and only the reversible part of the model equations are really solved backward. Practical consequences of this assumption are analysed in Sect 3.6. A similar solution was proposed by Pu et al. (1997) and Kalnay et al. (2000) to stabilize their Quasi-Inverse Linear model. Kalnay et al. (2000) and Reynolds and Palmer (1998) showed that the backward model is indeed a very accurate approximation of the inverse model. All the possible mitigatory solutions just described do not take into account water masses transformations resulting from diffusion processes. Indeed, unmixing water masses in the backward integration would be quite difficult and expensive.

To describe the D-BFN algorithm, let us assume the time continuous model satisfies dynamical equations of the form:

$$\frac{\partial \mathbf{x}}{\partial t} = \mathcal{F}(\mathbf{x}) + \nu\Delta\mathbf{x}, \qquad 0 < t < T, \tag{2.34}$$

with an initial condition $\mathbf{x}(0) = \mathbf{x}_0$, where $\mathcal{F}$ denotes the nonlinear model operator without diffusive terms, ν is a diffusion coefficient and Δ represents a diffusion operator. If nudging is applied to the forward system 4.1 it gives:

$$\frac{\partial \mathbf{x}_k}{\partial t} = \mathcal{F}(\mathbf{x}_k) + \nu\Delta\mathbf{x}_k + \mathbf{K}(\mathbf{y} - \mathcal{H}(\mathbf{x}_k)) \tag{2.35a}$$

$$\mathbf{x}_k(0) = \tilde{\mathbf{x}}_{k-1}(0), \qquad 0 < t < T, \tag{2.35b}$$

where $k \in \mathbb{N}_{\geq 1}$ states for iterations. Nudging applied to the backward system with the reversed diffusion sign gives:

$$\frac{\partial \tilde{\mathbf{x}}_k}{\partial t} = \mathcal{F}(\tilde{\mathbf{x}}_k) - \nu\Delta\tilde{\mathbf{x}}_k - \mathbf{K}'(\mathbf{y} - \mathcal{H}(\tilde{\mathbf{x}}_k)) \tag{2.36a}$$

$$\tilde{\mathbf{x}}_k(T) = \mathbf{x}_k(T), \qquad T > t > 0. \tag{2.36b}$$

The system composed by Eqs.(2.35) and (2.36) are iterated until convergence. Auroux et al. (2011) have shown that by considering the variable transformation $t' = T - t$ the backward model can be written as:

$$\frac{\partial \tilde{\mathbf{x}}_k}{\partial t'} = -\mathcal{F}(\tilde{\mathbf{x}}_k) + \nu \Delta \tilde{\mathbf{x}}_k + \mathbf{K}'(\mathbf{y} - \mathcal{H}(\tilde{\mathbf{x}}_k))$$

$$\tilde{\mathbf{x}}_k(t' = 0) = \mathbf{x}_k(T)$$

Therefore, it can be solved with an initial condition and the same diffusion term as in the forward equation.

According to Yann Brenier (personal communication), if the forward and backward limite trajectory are equal, i.e $\tilde{\mathbf{x}}_\infty = \mathbf{x}_\infty$, then $\mathbf{x}_\infty$ satisfies the model equations without the Nudging and diffusion:

$$\frac{\partial \mathbf{x}_\infty}{\partial t} = \mathcal{F}(\mathbf{x}_\infty) \tag{2.37}$$

as well as the Poisson equation (2.38), which represents a smoothing process on the observations for which the degree of smoothness is given by the ratio $\frac{\nu}{\mathbf{K}}$ (Auroux et al., 2011).

$$\Delta \mathbf{x}_\infty = -\frac{\mathbf{K}}{\nu}(\mathbf{y} - \mathcal{H}(\mathbf{x}_\infty)) \tag{2.38}$$

Therefore, the DBFN provide estimations of the system which are smooth enough to avoid initialization problems related to the introduction of sparse and noisy observations into the system, at the same time they satisfy the model equation without diffusion.

All results presented in this section considered that there is no stochastic forcing neither in the system state nor in the measurement equation. In the next section we extend these methods so as to formally consider the system given by Eqs.(2.19) and (2.18) as a stochastic system. By doing so, the Kalman Filter may be seen as an extension of the LO and the Back and Forth Kalman Filter as an extension of the BFN.

2.5 Bayesian Estimation

From the probabilistic point of view, Data Assimilation can be described as a Hidden Markov Model (HMM), i.e. a statistical Markov model in which the system being

modeled is assumed to be a Markov process with unobserved (hidden) states (Cappé et al., 2005). In practice only a part of the system state space is observed (e.g some variables at discrete times) and the remaining variables and intermediary states are unobserved and must be determined. Formally, in this study we are interested in the following discrete stochastic system:

$$\mathbf{x}_k = \mathcal{M}_{k-1,k}(\mathbf{x}_{k-1}) + \eta_k \tag{2.39}$$

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \epsilon_k \tag{2.40}$$

where $\mathbf{x}_k$ is the system state at time index t_k, $\mathbf{y}_{1:k} = \{\mathbf{y}_1, ..., \mathbf{y}_k\}$ are observations from time index t_1 to t_k, and η_k and ϵ_k are random white Gaussian noise. $\mathcal{M}_{k-1,k}$ is a nonlinear model that propagates the state from time t_{k-1} to t_k and $\mathcal{H}_k$ is a map from the model space to the observation space. Therefore, the state equation (2.39) characterizes the state transition probability $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ while the measurement equation (2.40) describes the likelihood $p(\mathbf{y}_k|\mathbf{x}_k)$ which is related to the measurement noise model.

In DA one is usually interested in filtering methods. Filtering aims to estimate the posterior density $p(\mathbf{x}_k|\mathbf{y}_{0:k})$ given an initial density $p(\mathbf{x}_0)$, the transition density $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and the likelihood $p(\mathbf{y}_k|\mathbf{x}_k)$. Therefore, the objective of the filtering is to estimate the optimal current state at time t_k given the observations up to time t_k.

In the Bayes inference, priors and casual knowledge about a system are used to infer the conditional probabilities given finite observations. Formally, the conditional probability of an event $\mathbf{A}$ given that the event $\mathbf{B}$ has occurred with $P(\mathbf{B}) > 0$ is given by:

$$P(\mathbf{A}|\mathbf{B}) = \frac{P(\mathbf{A} \cap \mathbf{B})}{P(\mathbf{B})}. \tag{2.41}$$

This equation shows that to calculate the conditional probability $P(\mathbf{A}|\mathbf{B})$ the knowledge of the joint distribution of $\mathbf{A}$ and $\mathbf{B}$ and the distribution of $\mathbf{B}$ is needed. However, using the fact that $P(\mathbf{A} \cap \mathbf{B}) = P(\mathbf{B} \cap \mathbf{A})$ it is possible to derive a formula that is independent of the joint distribution of $\mathbf{A}$ and $\mathbf{B}$:

$$P(\mathbf{A}|\mathbf{B}) = \frac{P(\mathbf{B}|\mathbf{A})P(\mathbf{A})}{P(\mathbf{B})} \tag{2.42}$$

Equation (2.42) is known as the Bayes' rule. It will be extensively used in the derivation of DA methods presented here.

When using the Bayes' rule, three types of problem may be solved: filtering, smoothing and prediction. They differ with respect to the conditioning on past or future observations.

- **Filtering**: estimates the system state probability at time t_k conditioned on all past and current observations, $\mathbf{y}_{0:k}$, i.e. estimates $p(\mathbf{x}_k|\mathbf{y}_{0:k})$;
- **Prediction**: estimates the future probability of the system state given all past observations, i.e. estimates $p(\mathbf{x}_j|\mathbf{y}_{0:k}), j > k$
- **Smoothing**: estimates the system state probability at times $t_j < t_K$ given all available observations, i.e estimates $p(\mathbf{x}_j|\mathbf{y}_{0:K})$ or $p(\mathbf{x}_{0:K}|\mathbf{y}_{0:K})$.

In the following it is assumed the states follow a first-order Markov process, i.e. $p(\mathbf{x}_n|\mathbf{x}_{0:n-1}) = p(\mathbf{x}_n|\mathbf{x}_{n-1})$, and η_k is independent of η_j if $k \neq j$ and independent of ϵ_j for all j. Also, ϵ_k is independent of ϵ_j if $k \neq j$.

2.5.1 Kalman Filter

Although the Kalman Filter is well known and different text books provide detailed derivations of this algorithm, in this section the Kalman Filter is presented in detail since the results will be used to derive the Back and Forth Kalman Filter.

The linear case

In the linear case the Eqs. (2.39) and (2.40) can be written as:

$$\mathbf{x}_k = \mathbf{M}_{k-1,k}\mathbf{x}_{k-1} + \eta_k \tag{2.43}$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \epsilon_k \tag{2.44}$$

recalling that $\epsilon \sim N(0, R)$ and $\eta \sim N(0, Q)$.

The KF aims to calculate the first two statistical moments of the conditional pdf $p(\mathbf{x}_k|\mathbf{y}_{0:k})$, which in the linear Gaussian case fully determines the pdf. Due to the Markovian properties of the system and the independence of model and observation

errors, this pdf can be expressed as:

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{y}_{0:k}) &= \frac{p(\mathbf{y}_{0:k}|\mathbf{x}_k)p(\mathbf{x}_n)}{p(\mathbf{y}_{0:k})} \\ &= \frac{p(\mathbf{y}_k, \mathbf{y}_{0:k-1}|\mathbf{x}_k)p(\mathbf{x}_n)}{p(\mathbf{y}_k, \mathbf{y}_{0:k-1})} \\ &= \frac{p(\mathbf{y}_k|\mathbf{y}_{0:k-1}, \mathbf{x}_k)p(\mathbf{y}_{0:k-1}|\mathbf{x}_k)p(\mathbf{x}_n)}{p(\mathbf{y}_k|\mathbf{y}_{0:k-1})p(\mathbf{y}_{0:k-1})} \\ &= \frac{p(\mathbf{y}_k|\mathbf{y}_{0:k-1}, \mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{0:k-1})p(\mathbf{y}_{0:k-1})p(\mathbf{x}_n)}{p(\mathbf{y}_k|\mathbf{y}_{0:k-1})p(\mathbf{y}_{0:k-1})p(\mathbf{x}_n)} \\ &= \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{0:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{0:k-1})} \end{aligned} \tag{2.45}$$

Hence, the posterior pdf $p(\mathbf{x}_k|\mathbf{y}_{0:k})$ is described by the prior $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$, which defines the knowledge of the model, the likelihood $p(\mathbf{y}_k|\mathbf{x}_k)$, which determines the measurement noise in the Eq.(2.44), and a normalization factor $p(\mathbf{y}_k|\mathbf{y}_{0:k-1})$. The prior is calculated using the so-called propagation rule:

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})p(\mathbf{x}_k|\mathbf{x}_{k-1})d\mathbf{x}_{k-1} \tag{2.46}$$

where $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is called transition density. The propagation rule along with the Bayes' rule form a recursive sequential-in-time algorithm to calculate $p(\mathbf{x}_k|\mathbf{y}_{0:k})$. This is why the Kalman-like filters are often referred to as sequential methods.

The likelihood, $p(\mathbf{y}_k|\mathbf{x}_k)$, is given by:

$$p(\mathbf{y}_k|\mathbf{x}_k) = A_1 \exp -\frac{1}{2}\left(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k\right)^T \mathbf{R}^{-1}\left(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k\right) \tag{2.47}$$

where A_1 is a constant independent of $\mathbf{y}$ and $\mathbf{x}$. Concerning the prior $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$, its mean ($\mathbb{E}[\bullet]$) and covariance ($Cov[\bullet]$) are calculated according to:

$$\begin{aligned} \mathbb{E}[\mathbf{x}_k|\mathbf{y}_{0:k-1}] &= \mathbb{E}[\mathbf{M}_{k-1,k}\hat{\mathbf{x}}_{k-1} + \eta_k|\mathbf{y}_{0:k-1}] \\ &= \mathbf{M}_{k-1,k}\hat{\mathbf{x}}_{k-1} = \hat{\mathbf{x}}_{k|k-1} \end{aligned} \tag{2.48}$$

and

$$\begin{aligned} Cov[\mathbf{x}_k|\mathbf{y}_{0:k-1}] &= Cov[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}] \\ &= Cov[\hat{\mathbf{e}}_{k,k-1}]. \end{aligned} \tag{2.49}$$

Denoting the covariance of the error $\hat{\mathbf{e}}_{k,k-1}$ by $\mathbf{P}_{k,k-1}$ we obtain:

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = A_2 \exp -\frac{1}{2}\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right)^T \mathbf{P}_{k|k-1}^{-1}\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right) \tag{2.50}$$

where A_2 is a constant independent of $\mathbf{x}$.

The Eq.(2.45) can be calculated using the expressions (2.47) and (2.50):

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{y}_{0:k}) \propto A \exp &-\frac{1}{2}\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right)^T \mathbf{P}_{k|k-1}^{-1}\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right) \\ &-\frac{1}{2}\left(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k\right)^T \mathbf{R}^{-1}\left(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k\right) \end{aligned} \tag{2.51}$$

where $A = A_1A_2$ is a constant independent of $\mathbf{y}$ and $\mathbf{x}$, and the normalization term was left aside because it is also independent of $\mathbf{x}$.

The derivation of the Kalman Filter equations will follow the Maximum a Posteriori (MAP) approach that aims to find the mode of the posterior probability, which in the linear Gaussian case is similar to calculate the mean. The MAP estimative is found solving:

$$\left.\frac{\partial \ell np(\mathbf{x}_k|\mathbf{y}_{0:k})}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k=\hat{\mathbf{x}}_k} = 0 \tag{2.52}$$

Inserting the equation (2.51) into (2.52) and solving for $\hat{\mathbf{x}}_k$ gives:

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \left(\mathbf{H}_k^T\mathbf{R}^{-1}\mathbf{H}_k + \mathbf{P}_{k|k-1}^{-1}\right)^{-1}\left(\mathbf{H}_k^T\mathbf{R}^{-1}\mathbf{y}_k + \mathbf{P}_{k|k-1}^{-1}\hat{\mathbf{x}}_{k|k-1}\right) \\ &= \left(\mathbf{H}_k^T\mathbf{R}^{-1}\mathbf{H}_k + \mathbf{P}_{k|k-1}^{-1}\right)^{-1}\left[\left(\mathbf{H}_k^T\mathbf{R}^{-1}\mathbf{H}_k + \mathbf{P}_{k|k-1}^{-1}\right)\hat{\mathbf{x}}_{k|k-1} - \mathbf{H}_k^T\mathbf{R}^{-1}\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} + \mathbf{H}_k^T\mathbf{R}^{-1}\mathbf{y}_k\right] \\ &= \hat{\mathbf{x}}_{k|k-1} + \left(\mathbf{H}_k^T\mathbf{R}^{-1}\mathbf{H}_k + \mathbf{P}_{k|k-1}^{-1}\right)^{-1}\mathbf{H}_k^T\mathbf{R}^{-1}\left(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right) \end{aligned} \tag{2.53}$$

This formula requires the inversion of a matrix of size $m \times m$ where m is the size of the state space. Using the matrix inversion lemma:

$$\left(\mathbf{A} + \mathbf{UCV}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U}\right)^{-1}\mathbf{VA}^{-1}, \tag{2.54}$$

where $\mathbf{A}$ is $n \times n$, $\mathbf{U}$ is $n \times k$, $\mathbf{C}$ is $k \times k$ and $\mathbf{V}$ is $k \times n$, the Eq. (2.53) can be put into a form that requires the inversion of a matrix of size $n \times n$ where n is the size of the observation space. In most practical applications this is advantageous since the observation space is much smaller than the system state space. The new expression is similar to the one obtained when deriving the KF from the point of view of the minimum

variance estimator or the Best Linear Estimator (BLUE) (Gelb, 1974; Jazwinski, 1970):

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}\right) \tag{2.55}$$

where $\mathbf{K}_k$ is the Kalman gain defined by:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}\right)^{-1} \tag{2.56}$$

The forecast error covariance matrix, $\mathbf{P}_{k|k-1}$, can be derived observing that:

$$\begin{aligned} \hat{\mathbf{e}}_{k|k-1} &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} \\ &= \mathbf{M}_{k-1,k}\mathbf{x}_{k-1} + \eta_k - \mathbf{M}_{k-1,k}\hat{\mathbf{x}}_{k-1|k-1} \\ &= \mathbf{M}_{k-1,k}\hat{\mathbf{e}}_{k-1} + \eta_k \end{aligned} \tag{2.57}$$

Then, defining $\mathbf{P}_{k|k-1} = \mathbb{E}\left[(\hat{\mathbf{e}}_{k|k-1})(\hat{\mathbf{e}}_{k|k-1})^T\right]$ we get:

$$\mathbf{P}_{k|k-1} = \mathbf{M}_{k-1,k}\mathbf{P}_{k-1|k-1}\mathbf{M}_{k-1,k}^T + \mathbf{Q}_k \tag{2.58}$$

where $\mathbf{P}_{k-1|k-1} = \mathbb{E}\left[(\hat{\mathbf{e}}_{k-1|k-1})(\hat{\mathbf{e}}_{k-1|k-1})^T\right]$

We proceed in a similar way to derive the analysis error covariance matrix:

$$\begin{aligned} \hat{\mathbf{e}}_{k|k} &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k} \\ &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}\right) \\ &= \hat{\mathbf{e}}_{k|k-1} - \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}\right) \\ &= \hat{\mathbf{e}}_{k|k-1} + \mathbf{K}_k\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \left(\mathbf{H}_k\mathbf{x}_k + \epsilon_k\right) \\ &= \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\hat{\mathbf{e}}_{k|k-1} + \mathbf{K}_k\epsilon_k \end{aligned} \tag{2.59}$$

and hence, the analysis error covariance can be written as:

$$\begin{aligned} \mathbf{P}_{k|k} &= \mathbb{E}\left[(\hat{\mathbf{e}}_{k|k})(\hat{\mathbf{e}}_{k|k})^T\right] \\ &= \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_{k|k-1}\left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)^T + \mathbf{K}_k\mathbf{R}\mathbf{K}_k^T \end{aligned} \tag{2.60}$$

The Eq.(2.60) is valid whatever the $\mathbf{K}$ used. In the optimal case, i.e. when it is given by Eq.(2.56), the equation for $\mathbf{P}_{k|k}$ can be simplified. Noting that:

$$\mathbf{K}_k \left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}\right)\mathbf{K}_k^T = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T$$

the posterior covariance may be written as:

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\,\mathbf{P}_{k|k-1} \tag{2.61}$$

This completes the KF derivation. A summary of the Kalman Filter equations is presented in table 2.2.

Initialisation:
$\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$

Do for $k \geq 1$:

Forecast:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{M}_{k-1,k}\hat{\mathbf{x}}_{k-1|k-1} \tag{2.62}$$
$$\mathbf{P}_{k|k-1} = \mathbf{M}_{k-1,k}\mathbf{P}_{k-1|k-1}\mathbf{M}^T_{k-1,k} + \mathbf{Q}_k \tag{2.63}$$

Analysis:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}\right)^{-1} \tag{2.64}$$
$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right) \tag{2.65}$$
$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\,\mathbf{P}_{k|k-1} \tag{2.66}$$

Table 2.2: Summary of the Kalman Filter equations.

In this section the Kalman Filter was derived by seeking the MAP of the system's state conditioned on past and current observations. The KF searches for the first two statistical moments of the *a posteriori* distribution, which under Gaussian conditions fully determines the pdf. Under linearity and gaussianity the MAP solution is strictly equivalent to the minimum variance and the maximum likelihood estimations (Jazwinski, 1970).

The filter optimality relies on the correct specification of the error covariances $\mathbf{P}$, $\mathbf{R}$ and $\mathbf{Q}$. If inexact covariances are used in Kalman's equations, the filter may still give reasonable state estimates, but it will be suboptimal. Optimality, however, does not ensure stability. Asymptotic stability of the KF means that its solution will gradually become insensitive to its initial conditions, provided that the norms of the noise covariance matrices, $\mathbf{Q}_k$, $\mathbf{R}_k$ are bounded. If the system (2.43) and (2.44) with $\mathbf{x}_0$, η_k , ϵ_k , independent, is uniformly completely observable and uniformly completely controllable

and if $\mathbf{P}_0 \geq 0$ then the discrete time KF is uniformly asymptotically stable.

Therefore, two theorems summarizing the KF stability are presented by Jazwinski (1970):

2.5.1. Theorem. *If the system is uniformly completely observable and uniformly completely controllable, and if $P_0 \geq 0$ then the discrete filter, summarized in the table 2.2, is uniformly asymptotically stable.*

2.5.2. Theorem. *Let the system be uniformly completely observable and uniformly completely controllable. Suppose $P^1_{k|k}$ and $P^2_{k|k}$ are two solution of Eq.(2.63) with respective initial conditions P^1_0 and P^2_0; $P^1_0, P^2_0 \geq 0$. Let $\delta P(t_k) = P^1_{k|k} - P^2_{k|k}$. Then*

$$\|\delta P(t_k)\| \leq c_2 e^{-2c3(t_k - t_0)} \|P^1_0 - P^2_0\| \to 0 \qquad (k \to \infty)$$

$(c_2, c_3 > 0)$.

When the state and/or measurement equations are nonlinear the KF produces poor results since in general $\mathbb{E}(\mathcal{M}(\mathbf{x})) = \mathcal{M}(\mathbb{E}(\mathbf{x}))$ (Gelb, 1974). In the next section we derive an alternative solution for the KF when the operators are weakly non linear.

The nonlinear case - Extended Kalman Filter

In this section the case where the model and the observation operators are weakly nonlinear is treated. Weakly nonlinear systems refer to systems for which the linear terms tend to dominate the physics even though the nonlinear effects are still playing an essential role. In this case, to a first approximation, the system is essentially linear. The Extended Kalman Filter (EKF) idea is to linearize to the first order the non-linear model and observation operators $\mathcal{M}$ and $\mathcal{H}$:

$$\begin{aligned} \mathbf{M}_{k-1,k} &= \nabla_{\mathbf{x}} \mathcal{M}_{k-1,k}(\mathbf{x}_{k|k}) \\ \mathbf{H}_k &= \nabla_{\mathbf{x}} \mathcal{H}_k(\mathbf{x}_{k|k-1}) \end{aligned} \tag{2.67}$$

and apply the resulting tangent linear model to the KF equations. The EKF is neither optimal nor unbiased since, as it was already said, in general $\mathbb{E}(\mathcal{M}(\mathbf{x})) = \mathcal{M}(\mathbb{E}(\mathbf{x}))$. It works well when the system is weakly non-linear but provides a poor performance when the true *a posterior* is non-Gaussian (e.g. heavily skewed or multimodal).

An alternative to the EKF is the Unscented Kalman filter, which aims to approximate the prior and likelihood pdfs but not the system evolution and results in a second order estimation for the mean while the EKF is of first order. Other examples are the ensemble-like filters (Evensen, 2003) and particle filters (van Leeuwen, 2009) which are sequential Monte Carlo methods that normally do not require any assumption about the gaussianity of the probability distributions.

Regarding the EKF, even if the limits of the first order approximation are respected and the true *a posteriori* is Gaussian, the size of the state space in the case of oceanic applications are so huge that it would be impossible to explicitly compute and manage calculations with the covariance matrix. To deal with this problem some works (Toth and Kalnay, 1993; Pham, 1997; Trevisan and Palatella, 2011; Palatella et al., 2013) evoked the theory of the Lyapunov vectors.

Roughly speaking, the stability of an aperiodic orbit is studied by considering the evolution along the flow of infinitesimal perturbations. The perturbations dynamics are governed by the tangent linear equations which defines the tangent space to the nonlinear trajectory. The Lyapunov vectors are the vectors spanning the tangent space, and the associated Lyapunov exponents determines the vectors spanning the unstable (positive exponents), neutral (null) and stable (negative) subspaces (Legras and Vautard, 1996).

Based on this theory reduced rank algorithms have been investigated (Pham, 1997). The philosophy of this approach is to approximate the forecast-error covariance matrix ($\mathbf{P}$) by a reduced rank matrix representing the subspace on which the forecast errors are amplified, i.e. the unstable and neutral subspaces. The study of Trevisan and Palatella (2011) demonstrated that the subspace spanned by the Lyapunov vectors with positive and null exponents are usually much smaller than the full state space. Carrassi et al. (2007) calculated a subspace of dimension 24 for a quasi-geostrophic model whose state space has dimension 14884.

In addition, Trevisan and Palatella (2011) showed that the rank of the EKF covariance matrix converges to the rank of the unstable+neutral subspace. Their conclusion is that the full rank filter and the reduced rank filter produce, asymptotically, the same results. However, it is only true if the reduced rank matrix properly models the desired

Lyapunov vectors and if the errors evolution are linear, i.e. the errors are small. Indeed, the stability of the DA system requires the detection of the unstable structures by the observational network.

Singular Evolutive Extended Kalman filter The SEEK filter is a reduced rank approximation of the EKF that uses the square-root decomposition of $\mathbf{P}$:

$$\mathbf{P}_{k-1|k-1} = \mathbf{S}_{k-1|k-1}\mathbf{S}^T_{k-1|k-1} \tag{2.68}$$

where the dimension of $\mathbf{S}$ is $m \times r$ and $r << m = dim(\mathbf{x})$. This approximation reduces the propagation cost of the covariance matrix, Eq. (2.63), since it requires r model integrations instead of m. The EKF forecast step in the square-root formulation is given by:

$$\mathbf{x}_{k|k-1} = \mathcal{M}_{k-1,k}\mathbf{x}_{k-1|k-1} \tag{2.69}$$

$$\mathbf{S}_{k|k-1} = \mathbf{M}_{k-1,k}\mathbf{S}_{k-1|k-1} \tag{2.70}$$

where we note that the state $\mathbf{x}_{k-1|k-1}$ is propagated by the non linear model while the covariance matrix is propagated by the tangent linear model. Also, for the moment the model error is being ignored, i.e. the model is considered as perfect.

The SEEK equations are directly derived by applying the decomposition given by Eq. (2.68) to the KF equations. The Kalman gain, Eq. (2.64), is transformed to:

$$\begin{aligned}\mathbf{K}_k &= \mathbf{S}_{k|k-1}(\mathbf{H}_k\mathbf{S}_{k|k-1})^T \left((\mathbf{H}_k\mathbf{S}_{k|k-1})(\mathbf{H}_k\mathbf{S}_{k|k-1})^T + \mathbf{R}\right)^{-1} \\ &= \mathbf{S}_{k|k-1}\left(\mathbf{I} - (\mathbf{H}_k\mathbf{S}_{k|k-1})^T\mathbf{R}^{-1}(\mathbf{H}_k\mathbf{S}_{k|k-1})\right)^{-1}(\mathbf{H}_k\mathbf{S}_{k|k-1})^T\mathbf{R}^{-1}.\end{aligned} \tag{2.71}$$

Writing the gain in this form permits an interesting interpretation of the analysis increment. Assuming $\mathbf{d}_k = (\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_{k|k-1})$ and

$$\gamma_k = \left[\mathbf{I} - (\mathbf{H}_k\mathbf{S}_{k|k-1})^T\mathbf{R}^{-1}(\mathbf{H}_k\mathbf{S}_{k|k-1})\right]^{-1}(\mathbf{H}_k\mathbf{S}^f_{k|k-1})^T\mathbf{R}_k^{-1}\mathbf{d}_k$$

the KF analysis equation (2.65) takes the form (Brasseur and Verron, 2006):

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{S}_{k|k-1}\gamma_k \tag{2.72}$$

In other words, the analysis increment is a linear combination of the columns of $\mathbf{S}_{k|k-1}$. This equation explains how the filter performance depends on the structure of the covariance matrix, or more precisely on the subspace spanned by the columns of $\mathbf{S}_{k|k-1}$.

Finally, the analysis error covariance (Eq.2.66) is transformed into:

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\,\mathbf{P}_{k|k-1} = \mathbf{S}_{k|k}\mathbf{S}_{k|k}^T \tag{2.73}$$

with

$$\mathbf{S}_{k|k} = \mathbf{S}_{k|k-1}\left(\mathbf{I} - (\mathbf{H}_k\mathbf{S}_{k|k-1})^T\mathbf{R}^{-1}(\mathbf{H}_k\mathbf{S}_{k|k-1})\right)^{-1/2} \tag{2.74}$$

The table 2.4 summarizes the SEEK filter equations. The SEEK filter was derived by applying the square-root decomposition to the forecast error covariance matrix. In fact, in order to decrease the computational cost, only some eigenvectors of the full covariance matrix are retained , which motivates the name "Singular Filter". In this case the covariance error forecast is obtained by integrating a few (usually ≤ 100) model perturbations. As a consequence, it is never needed to load the full covariance matrix neither on the virtual memory nor in the hard disk.

2.5.2 Kalman Smoothers

As already said, the filtering problem is the problem of estimating the state of a dynamical system given past and current observations of the system. The filtering technique can be extended to solve smoothing problems by considering future observations to estimate the current state of a system, i.e. the smoothing solution is the probability of the system's state conditioned on past, current and future observations. Three main types of smoothing problems exist:

- fixed-point: estimates the system's state at a single particular point in time: $p(\mathbf{x}_k|\mathbf{y}_{1:K})$, for $0 < k < K$,
- fixed-interval: estimates the system states over an entire time interval using all available observations: $p(\mathbf{x}_{0:K}|\mathbf{y}_{1:K})$,
- fixed-lag: estimates the system states over an entire fixed-interval of length N that slides forward with the filters: $p(\mathbf{x}_{k-N:k}|\mathbf{y}_{1:k})$.

The Fixed-lag Kalman Smoother

In this section a fixed-lag smoother is derived based on the KF and on the augmented state vector technique (Simon, 2006). Let us suppose we want to estimate $\mathbf{x}_{k-N:k|k}$ for some fixed "time lag" N. The key idea is to define an augmented state vector and an associated linear discrete time augmented dynamical system. The new state vector $\mathbf{z} \in \mathbb{R}^{p(N+2)}$ is defined and governed by the following discrete time state equation:

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k-1} \\ \mathbf{x}_{k-2} \\ \vdots \\ \mathbf{x}_{k-N} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{k-1,k} & 0 & 0 & \cdots & 0 \\ \mathbf{I} & 0 & 0 & \cdots & 0 \\ 0 & \mathbf{I} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{x}_{k-2} \\ \mathbf{x}_{k-3} \\ \vdots \\ \mathbf{x}_{k-N-1} \end{bmatrix} + \begin{bmatrix} \eta_k \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.75}$$

while the augmented measurement equation takes the form:

$$\mathbf{y}_k = \mathbf{G}_k \mathbf{z}_k + \epsilon_k \tag{2.76}$$

where

$$\mathbf{G}_k = \begin{bmatrix} \mathbf{H}_k & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{2.77}$$

With this new system we can proceed in a similar way as deriving the KF, i.e.

$$\left. \frac{\partial \ell n p(\mathbf{z}_k | \mathbf{y}_{1:k})}{\partial \mathbf{z}_k} \right|_{\mathbf{z}_k = \hat{\mathbf{z}}_k} = 0 \tag{2.78}$$

to obtain estimatives of $\mathbf{z}_k | \mathbf{y}_{1:k}$. In this case $p(\mathbf{z}_k | \mathbf{y}_{1:k})$ takes the form:

$$\begin{aligned} p(\mathbf{z}_k | \mathbf{y}_{0:k}) \propto A \exp & -\frac{1}{2} \left(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1} \right)^T \mathbf{\Sigma}_{k|k-1}^{-1} \left(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1} \right) \\ & - \frac{1}{2} \left(\mathbf{y}_k - \mathbf{G}_k \hat{\mathbf{z}}_{k|k-1} \right)^T \mathbf{R}^{-1} \left(\mathbf{y}_k - \mathbf{G}_k \hat{\mathbf{z}}_{k|k-1} \right) \end{aligned} \tag{2.79}$$

where the augmented state covariance matrix is given by:

$$\mathbf{\Sigma}_{k|k-1} = \begin{bmatrix} \mathbf{P}_{k|k-1} & \mathbf{P}_{k,k-1|k-1} & \cdots & \mathbf{P}_{k,k-N|k-1} \\ \mathbf{P}_{k,k-1|k-1}^T & \mathbf{P}_{k-1|k-1} & \cdots & \mathbf{P}_{k-1,k-N|k-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{k,k-N|k-1}^T & \mathbf{P}_{k-1,k-N|k-1}^T & \cdots & \mathbf{P}_{k-N|k-1} \end{bmatrix} \tag{2.80}$$

and $\mathbf{P}_{k,i|k-1} = \mathbb{E}\left[(\mathbf{x}_k - \mathbf{x}_{i|k-1})(\mathbf{x}_k - \mathbf{x}_{i|k-1})^T\right]$.

The solution of the Eq. (2.78) is given by:

$$\hat{\mathbf{z}}_{k|k} = \hat{\mathbf{z}}_{k|k-1} + \mathbf{L}_k \left(\mathbf{y}_k - \mathbf{G}_k \hat{\mathbf{z}}_{k|k-1}\right) \tag{2.81}$$

where $\mathbf{L}_k$ is the Kalman gain defined by:

$$\mathbf{L}_k = \left(\mathbf{G}_k \mathbf{\Sigma}_{k|k-1}\right)^T \left(\mathbf{G}_k \mathbf{\Sigma}_{k|k-1} \mathbf{G}_k^T + \mathbf{R}\right)^{-1}$$

The form of $\mathbf{G}_k$ implies that only the first row of $\mathbf{\Sigma}_{k|k-1}$ must be calculated, which results in:

$$\mathbf{K}_{i|k} = \left(\mathbf{H}_k \mathbf{P}_{k,i|k-1}\right)^T \left(\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}\right)^{-1}$$

for $i \in \{k-N, k-N+1, ..., k\}$.

The analysis error is calculated as:

$$\mathbf{r}_{k|k} = \mathbf{z}_k - \mathbf{z}_{k|k} = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k-1} \\ \vdots \\ \mathbf{x}_{k-N} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_{k|k} \\ \mathbf{x}_{k-1|k} \\ \vdots \\ \mathbf{x}_{k-N|k} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{k|k} \\ \mathbf{e}_{k-1|k} \\ \vdots \\ \mathbf{e}_{k-N|k} \end{bmatrix} \tag{2.82}$$

Therefore, the analysis error covariance, $\mathbf{\Sigma}_{k|k} = \mathbb{E}[\mathbf{r}_{k|k}\mathbf{r}_{k|k}^T]$ is given by:

$$\mathbf{\Sigma}_{k|k} = \begin{bmatrix} \mathbb{E}[\mathbf{e}_{k|k}\mathbf{e}_{k|k}^T] & \mathbb{E}[\mathbf{e}_{k|k}\mathbf{e}_{k-1|k}^T] & \cdots & \mathbb{E}[\mathbf{e}_{k|k}\mathbf{e}_{k-N|k}^T] \\ \mathbb{E}[\mathbf{e}_{k-1|k}\mathbf{e}_{k|k}^T] & \mathbb{E}[\mathbf{e}_{k-1|k}\mathbf{e}_{k-1|k}^T] & \cdots & \mathbb{E}[\mathbf{e}_{k-1|k}\mathbf{e}_{k-N|k}^T] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[\mathbf{e}_{k-N|k}\mathbf{e}_{k|k}^T] & \mathbb{E}[\mathbf{e}_{k-N|k}\mathbf{e}_{k-1|k}^T] & \cdots & \mathbb{E}[\mathbf{e}_{k-N|k}\mathbf{e}_{k-N|k}^T] \end{bmatrix} \tag{2.83}$$

$$= \begin{bmatrix} \mathbf{P}_{k|k} & \mathbf{P}_{k,k-1|k} & \cdots & \mathbf{P}_{k,k-N|k} \\ \mathbf{P}_{k,k-1|k}^T & \mathbf{P}_{k-1|k} & \cdots & \mathbf{P}_{k-1,k-N|k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{k,k-N|k}^T & \mathbf{P}_{k-1,k-N|k}^T & \cdots & \mathbf{P}_{k-N|k} \end{bmatrix}. \tag{2.84}$$

Thus, the first entry of $\mathbf{\Sigma}_{k|k}$ represents the filter analysis error covariance while the others diagonal blocks are the smoother error covariance, which is only useful for performance diagnostics. Proceeding as for the Eq. (2.66) we obtain the expressions for $\mathbf{P}_{k,i|k}$ and

$\mathbf{P}_{i|k}$:

$$\mathbf{P}_{i|k} = \mathbf{P}_{i|k-1} - \mathbf{K}_{i|k}\mathbf{H}_k\mathbf{P}_{k,i|k-1} \tag{2.85}$$

$$\mathbf{P}_{k,i|k} = (\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_k)\mathbf{P}_{k,i|k-1} \tag{2.86}$$

The forecast error covariance, $\mathbf{P}_{k,i|k-1}$, is derived in the same way Eq. (2.63) was derived. This produces:

$$\mathbf{P}_{k,i|k-1} = \mathbf{M}_{k-1|k}\mathbf{P}_{k-1,i|k-1} \tag{2.87}$$

$$\mathbf{P}_{k|k-1} = \mathbf{M}_{k-1|k}\mathbf{P}^T_{k,k-1|k-1} + \mathbf{Q}_k. \tag{2.88}$$

The Kalman smoother derived is of the fixed-lag type, i.e. the equations are valid for $i \in \{k-N, k-N+1, ..., k\}$, where N is the lag. However, the derivation can be further generalized by considering $i \in \mathcal{I}$, where $\mathcal{I}$ is the ensemble of indexes for which the retrospective analysis is considered (Cosme et al., 2010). For $\mathcal{I} = \{0\}$ and $\mathcal{I} = \{0, 1, 2, ..., K-2, K-1, K\}$ the fixed-point smoother and the fixed-interval smoother are obtained, respectively. The Kalman smoother equations are summarized as follows:

Initialisation:
$\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$

Do for $k \geq 1$:

Forecast:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{M}_{k-1,k}\hat{\mathbf{x}}_{k-1|k-1} \tag{2.89}$$

$$\mathbf{P}_{k,i|k-1} = \mathbf{M}_{k-1|k}\mathbf{P}_{k-1,i|k-1} \tag{2.90}$$

$$\mathbf{P}_{k|k-1} = \mathbf{M}_{k-1|k}\mathbf{P}^T_{k,k-1|k-1} + \mathbf{Q}_k \tag{2.91}$$

Analysis:

$$\mathbf{K}_{i|k} = (\mathbf{H}_k\mathbf{P}_{k,i|k-1})^T \left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}\right)^{-1} \quad \forall i \in \mathcal{I} \tag{2.92}$$

$$\hat{\mathbf{x}}_{i|k} = \hat{\mathbf{x}}_{i|k-1} + \mathbf{K}_{i|k}\left(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right) \quad \forall i \in \mathcal{I} \tag{2.93}$$

$$\mathbf{P}_{i|k} = \mathbf{P}_{i|k-1} - \mathbf{K}_{i|k}\mathbf{H}_k\mathbf{P}_{k,i|k-1} \quad \forall i \in \mathcal{I} \tag{2.94}$$

$$\mathbf{P}_{k,i|k} = (\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_k)\mathbf{P}_{k,i|k-1} \quad \forall i \in \mathcal{I} \tag{2.95}$$

Table 2.3: Summary of the fixed-lag Kalman smoother equations.

The smoother presented here was derived independently of the KF equations. The

final equations include the KF equations and the associated smoothing expressions. Its peculiarity is the use of temporal correlations that allows for the use of future observations to improve past states produced by the filter. This is particularly important when the observations are sparse and the focus is on reconstructing as accurately as possible the system past evolution.

With this smoother formulation, it would be almost impractical its application to huge geophysical system. As an alternative Cosme et al. (2010) derived its square-root version which is presented in the following section.

SEEK - smoother

The SEEK smoother is a reduced rank square-root smoother based on the augmented state smoother presented in the last section. Similar to the SEEK filter in this section the perfect model hypothesis is considered. The derivation starts by noting that $\mathbf{P}_{k,i|k-1} = \mathbf{S}_{k|k-1}\mathbf{S}_{i|k-1}^T$. Applying this decomposition to the Eqs. (2.89), (2.90) and (2.91) yield the forms:

$$\mathbf{x}_{k|k-1} = \mathcal{M}_{k-1,k}\mathbf{x}_{k-1|k-1} \tag{2.96}$$

$$\mathbf{P}_{k,i|k-1} = \left(\mathbf{M}_{k-1,k}\mathbf{S}_{k-1|k-1}\right)\left(\mathbf{S}_{i|k-1}\right)^T = \mathbf{S}_{k|k-1}\left(\mathbf{S}_{i|k-1}\right)^T \tag{2.97}$$

$$\mathbf{P}_{k|k-1} = \left(\mathbf{M}_{k-1,k}\mathbf{S}_{k-1|k-1}\right)\left(\mathbf{S}_{k|k-1}\right)^T = \mathbf{S}_{k|k-1}\left(\mathbf{S}_{k|k-1}\right)^T \tag{2.98}$$

The smoother gain (Eq. 2.92) is rewritten as:

$$\begin{aligned}\mathbf{K}_{i|k} &= \mathbf{P}_{k,i|k-1}\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}\right)^{-1} \\ &= \mathbf{S}_{i|k-1}(\mathbf{H}_k\mathbf{S}_{k|k-1})^T\left((\mathbf{H}_k\mathbf{S}_{k|k-1})(\mathbf{H}_k\mathbf{S}_{k|k-1})^T + \mathbf{R}\right)^{-1} \\ &= \mathbf{S}_{i|k-1}\left(\mathbf{I} - (\mathbf{H}_k\mathbf{S}_{k|k-1})^T\mathbf{R}^{-1}(\mathbf{H}_k\mathbf{S}_{k|k-1})\right)^{-1}(\mathbf{H}_k\mathbf{S}_{k|k-1})^T\end{aligned} \tag{2.99}$$

The analysis is given by:

$$\mathbf{x}_{i|k} = \mathbf{x}_{i|k-1} + \mathbf{K}_{i|k}\mathbf{d}_k \tag{2.100}$$

and the analysis error covariances are transformed into:

$$\begin{aligned}\mathbf{P}_{k,i|k} &= \left[\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_k\right]\mathbf{S}_{k|k-1}(\mathbf{S}_{k-1|k-1})^T \\ &= \mathbf{S}_{k|k-1}\left[\mathbf{I} + \mathbf{\Gamma}_k\right]^{-1}(\mathbf{S}_{k-1|k-1})^T \qquad (2.101)\\ \mathbf{P}_{i|k} &= \mathbf{S}_{k-1|k-1}(\mathbf{S}_{k-1|k-1})^T - \mathbf{K}_{i|k}\mathbf{H}_k\mathbf{S}_{k|k-1}(\mathbf{S}_{k-1|k-1})^T \\ &= \mathbf{S}_{i|k-1}\left[\mathbf{I} + \mathbf{\Gamma}_k\right]^{-1}(\mathbf{S}_{i|k-1})^T \qquad (2.102)\end{aligned}$$

and therefore the square root formula is given by:

$$\mathbf{S}_{i|k} = \mathbf{S}_{i|k-1}\left[\mathbf{I} + \mathbf{\Gamma}_k\right]^{-1/2} \qquad (2.103)$$

Next we present a summary of the SEEK filter and smoother equations:

Initialization:
$\mathbf{x}_0$ and $\mathbf{P}^a = \mathbf{S}_0(\mathbf{S}_0)^T$

Forecast:

$$\mathbf{x}_{k|k-1} = \mathcal{M}_{k-1,k}\mathbf{x}_{k-1|k-1} \qquad (2.104)$$
$$\mathbf{S}_{k|k-1} = \mathbf{M}_{k-1,k}\mathbf{S}_{k-1|k-1} \qquad (2.105)$$

Filter Analysis:

$$\mathbf{\Gamma}_k = (\mathbf{H}_k\mathbf{S}_{k|k-1})^T\mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{S}_{k|k-1}) \qquad (2.106)$$
$$\mathbf{d}_k = (\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^f) \qquad (2.107)$$
$$\gamma_k = [\mathbf{I} + \mathbf{\Gamma}_k]^{-1}(\mathbf{H}_k\mathbf{S}_{k|k-1})^T\mathbf{R}_k^{-1}\mathbf{d}_k \qquad (2.108)$$
$$\mathbf{L}_k = [\mathbf{I} + \mathbf{\Gamma}_k]^{-1/2} \qquad (2.109)$$
$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{S}_{k|k-1}\gamma_k \qquad (2.110)$$
$$\mathbf{S}_{k|k} = \mathbf{S}_{k|k-1}\mathbf{L}_k \qquad (2.111)$$

Smoother Analysis:

$$\mathbf{x}_{i|k} = \mathbf{x}_{i|k-1} + \mathbf{S}_{i|k-1}\gamma_k \qquad \forall i \in \mathcal{I} \qquad (2.112)$$
$$\mathbf{S}_{i|k} = \mathbf{S}_{i|k-1}\mathbf{L}_k \qquad \forall i \in \mathcal{I} \qquad (2.113)$$

Table 2.4: Summary of the SEEK filter and SEEK smoother equations.

Equation (2.112) shows that the retrospective analysis is indeed a linear combination of the columns of $\mathbf{S}_{i|k-1}$ weighted by the coefficients γ_k used by the filter. Considering

a lag N, the final smoothed field is given by:

$$\mathbf{x}_{k-N|k} = \mathbf{x}_{k-N|k-N} + \sum_{i=1}^{N} \mathbf{S}_{k-N|k-N+i-1}\gamma_{k-N+i} \tag{2.114}$$

Therefore, once the filter has already passed and the covariance matrices $\mathbf{S}$ as well as the weights γ were stored the smoother is almost cost free and may be executed off-line as a reprocessing or reanalysis technique.

The Backward Smoother - BS

The Backward Smoother (BS) aims to estimate $p(\mathbf{x}_{0:K}|\mathbf{y}_{1:K})$ or $p(\mathbf{x}_k|\mathbf{y}_{1:K})$ once the filter solution at t_K, i.e. $p(\mathbf{x}_K|\mathbf{y}_{1:K})$, has already been calculated. It is based on the backward propagation of the filter analysis. Ideally it should be done using the propagation rule:

$$p(\mathbf{x}_{K-1}|\mathbf{y}_{1:K}) = \int p(\mathbf{x}_{K-1}|\mathbf{x}_K, \mathbf{y}_K)p(\mathbf{x}_K|\mathbf{y}_{1:K})d\mathbf{x}_K$$

however the estimation of $p(\mathbf{x}_{K-1}|\mathbf{x}_K, \mathbf{y}_K)$ would require the minimization of a cost function similar to that one of the weak constraint 4Dvar. In order to make the method feasible, the perfect model hypothesis, the same used by the strong constraint 4Dvar, is assumed. In this case the conditioning on $\mathbf{y}_K$ can be dropped since $p(\mathbf{x}_{k-1}|\mathbf{x}_k) = \delta(\mathbf{x}_{k-1} - \mathcal{M}_{k,k-1}(\mathbf{x}_k))$, i.e. $\mathbf{x}_{k-1}$ is entirely determined by $\mathbf{x}_k$. In this case, the propagation rule takes the form:

$$p(\mathbf{x}_{K-1}|\mathbf{y}_{1:K}) = \int p(\mathbf{x}_{K-1}|\mathbf{x}_K)p(\mathbf{x}_K|\mathbf{y}_{1:K})d\mathbf{x}_K \tag{2.115}$$

Applying the Eq.(2.115) recursively, the pdfs $p(\mathbf{x}_k|\mathbf{y}_{1:K})$ may be estimated. This means that once the filter has passed one only needs to integrate the model backward from time t_K to time t_k or t_0 to obtain the smoothed states.

Although the algorithm formulation is quite general, here we will rely on the KF for practical implementation. This means that only the first two statistical moments should be propagated backward, namely the mean and the covariance. Our algorithm has some similarities with the inverse 3Dvar proposed by Pu et al. (1997). They used a modified version of the Eq.(2.45) to calculate the conditional probability $p(\mathbf{M}_{K,0}\delta\mathbf{x}_0|\delta\mathbf{y}_K)$, where

$\delta\mathbf{x}_0 = \mathbf{x}_0 - \hat{\mathbf{x}}_{0|-1}$ and $\delta\mathbf{y}_K = \mathbf{y}_K - \mathbf{H}_K\hat{\mathbf{x}}_{K|K-1}$. This way, Eq.(2.51) is transformed into:

$$\begin{aligned} p(\mathbf{M}_{K,0}\delta\mathbf{x}_0|\delta\mathbf{y}_K) \propto A\exp -\frac{1}{2}\left(\mathbf{M}_{K,0}\delta\mathbf{x}_0\right)^T\mathbf{P}_{K|0}^{-1}\left(\mathbf{M}_{K,0}\delta\mathbf{x}_0\right) \\ -\frac{1}{2}\left(\delta\mathbf{y}_k - \mathbf{H}_k\mathbf{M}_{K,0}\delta\mathbf{x}_0\right)^T\mathbf{R}^{-1}\left(\delta\mathbf{y}_k - \mathbf{H}_k\mathbf{M}_{K,0}\delta\mathbf{x}_0\right) \end{aligned} \tag{2.116}$$

The MAP estimative is found by solving

$$\frac{\partial \ln p(\mathbf{M}_{K,0}\delta\mathbf{x}_0|\delta\mathbf{y}_K)}{\partial\mathbf{M}_{K,0}\delta\mathbf{x}_0} = 0$$

which results in:

$$\mathbf{M}_{K,0}\delta\mathbf{x}_0 = \left(\mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_k^T\mathbf{R}^{-1}\mathbf{H}_k\right)^{-1}\mathbf{H}_k^T\mathbf{R}^{-1}\delta\mathbf{y}_k \tag{2.117}$$

Since $\mathbf{M}_{K,0}^{-1}$ is available the problem may be solved directly:

$$\delta\mathbf{x}_0 = \mathbf{M}_{K,0}^{-1}\left(\mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_k^T\mathbf{R}^{-1}\mathbf{H}_k\right)^{-1}\mathbf{H}_k^T\mathbf{R}^{-1}\delta\mathbf{y}_k \tag{2.118}$$

Comparing Eqs.(2.118) and (2.53) stands out that the inverse 3Dvar propagates the KF increment at time t_K backwards to time t_0 to recover the optimal initial increment $\delta\mathbf{x}_0$. Therefore, the BS and the inverse 3Dvar are equivalent in the forward resolution but they differ in the smoothing phase because the latter propagates the assimilation increment from t_K to t_0 using the inverse tangent linear model and ours (BS) propagates the analysed field using the nonlinear model.

Auroux et al. (2012) compared the DBFN with a simplified version of the inverse 3Dvar applied to the Burgers equation. In their formulation Eq.(2.118) was reduced to:

$$\delta\mathbf{x}_0 = \mathbf{M}_{K,0}^{-1}\delta\mathbf{y}_k$$

and the system assimilated a single full unnoisy observation at the end of the DA window. They observed the "inverse 3Dvar" produces better initial conditions than the DBFN for short assimilation windows but when the assimilation window is extended, so as to consider the nonlinear regime, the DBFN outperformes the "inverse 3Dvar".

Next it is shown that in the linear and perfect model case the backward smoother is equivalent to the SEEK-smo. Let us suppose the filter analysis at time t_k is available

under the form:

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{S}_{k|k-1}\gamma_k$$

and the backward model $\tilde{\mathbf{M}}_{k-1,k} : \mathbf{x}_k \to \mathbf{x}_{k-1}$ is perfect and available. If we apply the backward model operator $\tilde{\mathbf{M}}_{k-1,k}$ to $\mathbf{x}_{k|k}$ we obtain the model state at time index t_{k-1}

$$\begin{aligned}
\tilde{\mathbf{x}}_{k-1} = \tilde{\mathbf{M}}_{k-1,k}\mathbf{x}_{k|k} &= \tilde{\mathbf{M}}_{k-1,k}\mathbf{x}_{k|k-1} + \tilde{\mathbf{M}}_{k-1,k}\mathbf{S}_{k|k-1}\gamma_k \\
&= \tilde{\mathbf{M}}_{k-1,k}\mathbf{M}_{k+1,k}\mathbf{x}_{k-1|k-1} + \tilde{\mathbf{M}}_{k-1,k}\mathbf{M}_{k,k-1}\mathbf{S}_{k-1|k-1}\gamma_k \\
&= \mathbf{x}_{k-1|k-1} + \mathbf{S}_{k-1|k-1}\gamma_k \\
&= \mathbf{x}_{k-1|k}
\end{aligned}$$

Therefore, the obtained state is nothing else than the SEEK-smo solution for a one day lag. Proceeding in a similar way, i.e. applying $\tilde{\mathbf{M}}_{k-2,k-1}$ to $\mathbf{x}_{k-1|k}$ we obtain the model state at time index t_{k-2} which is equivalent to the two day lag SEEK-smo solution.

$$\begin{aligned}
\tilde{\mathbf{x}}_{k-2} = \tilde{\mathbf{M}}_{k-2,k-1}\mathbf{x}_{k-1|k} &= \tilde{\mathbf{M}}_{k-2,k-1}\mathbf{x}_{k-1|k-1} + \tilde{\mathbf{M}}_{k-2,k-1}\mathbf{S}_{k-1|k-1}\gamma_k \\
&= \tilde{\mathbf{M}}_{k-2,k-1}\left(\mathbf{x}_{k-1|k-2} + \mathbf{S}_{k-1|k-2}\gamma_{k-1}\right) + \mathbf{S}_{k-2|k-1}\gamma_k \\
&= \mathbf{x}_{k-2|k-2} + \mathbf{S}_{k-2|k-2}\gamma_{k-1} + \mathbf{S}_{k-2|k-1}\gamma_k \\
&= \mathbf{x}_{k-2|k-2:k}
\end{aligned}$$

For a lag N the backward smoother takes the form:

$$\begin{aligned}
\mathbf{x}_{k-N|k-N:k} &= \tilde{\mathbf{M}}_{k-N,k-N+1}\cdots\tilde{\mathbf{M}}_{k-2,k-1}\tilde{\mathbf{M}}_{k-1,k}\mathbf{x}_{k|k} \\
&= \mathbf{x}_{k-N|k-N} + \sum_{i=1}^{N}\mathbf{S}_{k-N|k-N+i-1}\gamma_{k-N+i}
\end{aligned} \tag{2.119}$$

which is similar to Eq. (2.114). In addition, to obtain the covariance of the smoothed field, in the BS algorithm, it is necessary to back-propagate the reduced basis.

Therefore, in this section it was shown that under linearity and reversibility conditions the BS provides the same smoothed solution as the Kalman smoother. However, even under these hypothesis, in practice the methods may provide distinct results, since for practical purposes the Kalman smoother and especially the SEEK-smo uses techniques such as covariance localization and covariance inflation, discussed in the next section, which are avoided in the BS.

2.5.3 Iterative Kalman Smoothers

It is well-known that the use of reduced rank algorithms leads to sampling errors, i.e. underestimation of the forecast errors, and spurious covariances problems. The former implies that the background is too strongly weighted compared to the observations, and the latter results in unphysical analysis increments far from the observation location. To mitigate these problems covariance inflation and localization techniques has been successfully used (Ott et al., 2004; Hunt et al., 2007; Greybush et al., 2010).

Inflation increases the background uncertainties in the subspace spanned by the forecast perturbations. However, it does not account for new unstable directions for which the errors may project. Localization consists of limiting the spatial influence of an observation. However, as the system is usually multiscale, i.e. there are several characteristic length that interacts with each other, the choice of the influence range is quite a hard task, since it varies with the system variables, space and time. As a result, the analysed field may suffer from a lack of optimality as well as the system physical balance may be disrupted.

For oceanic and atmospheric applications, non-balanced dynamics are mainly represented by inertio-gravity waves propagation that tends to dominate the energy spectrum. Cohn et al. (1998) observed an unrealistically high ratio of divergence to vorticity as a consequence of local observation selection. This can lead to the discarding of observations by moving the DA solution back toward the balanced background state (Greybush et al., 2010).

Therefore, inflation and localization are rather ad-hoc solutions far from being optimal. The use of iterative algorithms would allow to gradually insert the observations into the model, as in the BFN, expecting the produced unbalanced dynamic to be less important than the dynamics of interest and to be eventually damped. Another important aspect is the possibility of using more than once the same data set. With this, the analysis sub-optimality, caused by the reasons mentioned above, may be mitigated. Last but not least, in the case of nonlinear models in which extended filters are used, the propagation of small successive increments are more likely to capture the linear dynamics near the truth than a large increment (Yang et al., 2012). A further aspect not analysed in this thesis is the possibility of using the iterations to adaptively estimate

some data assimilation parameters, e.g. inflation and localization parameters, as well as to estimate model and observation errors.

The theorem (2.5.2) says that the KF forgets its initial condition when "enough" observations are assimilated. As we have seen previously, the BFN limit trajectory is totally independent of the initial condition. In fact, numerical results show that the BFN error reaches its asymptotic regime after one to three assimilation cycles. Therefore, thanks to the iterations the BFN has excellent performances when the background state is quite different from the actual system's state.

With the iterative Backward Smoother (iBS) and the Back and Forth Kalman Filter (BFKF), to be presented in the following, we expect the forecast covariance matrix to converge to a matrix representing the uncertainties of the system or "the errors of the day" thanks to the iterations. This is especially important in two cases: cold start, i.e. when the assimilation system knows very little or nothing about the "real" system, and when the error structures undergoes a drastic change (Kalnay and Yang, 2010). The latter may happen with sudden changes due to strong nonlinearity, as during the initial development of a severe storm, for the atmosphere. During these transition periods, the reduced basis may not be representative of the unstable directions and extracting information from observations using them only once may not be efficient. An example of a cold start is the initialization of regional data assimilation from a global analysis obtained at coarser resolution, thus lacking features that represent the underlying mesoscale evolution (Yang et al., 2012).

Iterative Backward Smoother

Two iterative versions of the BS specialized to use the KF equations are proposed. The first algorithm, called iBS, is an iterative version of the BS algorithm presented in the last section. It is composed of three steps:

1. propagate the final filter analysis backward in time;
2. run the filter in forward mode using the KF or EKF algorithm initialized with the updated background state;
3. repeat steps 1 and 2 until convergence

A summary of the iBS algorithm using the SEEK filter is presented in the table (2.5).

Backward propagation of the filter analysis $\mathbf{x}^0_{K|K}$:

$$\mathbf{x}^1_{0|K} = \tilde{M}_{0,K}\mathbf{x}^0_{K|K} \tag{2.120}$$
$$\mathbf{S}^1_{0|K} = \tilde{M}_{0,K}\mathbf{S}^0_{K|K} \tag{2.121}$$

While i < itmax or $\|\mathbf{x}^i - \mathbf{x}^{i-1}\| > \epsilon$
Do
Forward Filter:
For k=0:number of analysis
Do
Filter Analysis:

$$\mathbf{x}^i_{k|k,K} = \mathbf{x}^i_{k|k-1,K} + \mathbf{S}^i_{k|k-1,K}\gamma^i_k \tag{2.122}$$
$$\mathbf{S}^i_{k|k,K} = \mathbf{S}^i_{k|k-1,K}\mathbf{L}^i_k \tag{2.123}$$

Forecast:

$$\mathbf{x}^i_{k+1|k,K} = \mathcal{M}_{k+1,k}\mathbf{x}^i_{k|k,K} \tag{2.124}$$
$$\mathbf{S}^i_{k+1|k,K} = \mathbf{M}_{k+1,k}\mathbf{S}^i_{k|k,K} \tag{2.125}$$

End Do

$$\mathbf{x}^{i+1}_{0|K} = \tilde{M}_{0,K}\mathbf{x}^i_{K|K} \tag{2.126}$$
$$\mathbf{S}^{i+1}_{0|K} = \tilde{M}_{0,K}\mathbf{S}^i_{K|K} \tag{2.127}$$

End Do

Table 2.5: Summary of the iBS algorithm using the SEEK filter for the analysis. The notation $\mathbf{x}^i_{k|k,K}$ indicates that we are referring to the iteration number i, that $\mathbf{x}$ is conditioned on all observations $\mathbf{y}_{1:K}$ which is related with iterations $it < i$, but for the present iterations $\mathbf{x}$ is only conditioned on $\mathbf{y}_{1:k}$. The subscript $k|k, K$ refers to a analysed state and $k|k-1, K$ to a forecast state.

The second algorithm, called iBSfU[1], aims to re-center the background state at a more accurate state, i.e. we change the $\mathbb{E}[\mathbf{x}_k]$ but not $\mathbf{P}_k$. This is done by modifying the steps given by the Eqs.(2.127) and (2.121) to $\mathbf{S}^i_{0|K} = \mathbf{I}\mathbf{S}^{i-1}_{0|0}$. At the end of step 1 the new background, obtained by the backward propagation of the final analysis, may be seen as "new educated" first guess, as observed by Jazwinski (1970).
Both algorithms have their equivalent in the literature. While the iBS may be seen as being equivalent to the "Running In Place" (RIP) method proposed by Kalnay and Yang (2010), the iBSfU is an equivalent of the Quasi-outer-loop (QOL) algorithm proposed

[1] iterative Backward Smoother with fixed Uncertainties

by Yang et al. (2012). The RIP uses a smoother similar to the SEEK-smo to perform the iterations as follows:

1. the smoother is used to go backward in time within an assimilation cycle to improve the background state and its covariance error at previous time;
2. advance with the standard filter from the updated background;
3. repeat steps 1 and 2 until the desired state is reached.

The QOL operates in the same way but for the step 2 the forward filter uses the "original" covariance matrix and not the smoothed one. Therefore, the iBS and RIP as well as the iBSfU and QOL are equivalent in the linear case, since as we have seen, the SEEK-smo and the backward model are equivalent. For the nonlinear case and in particular if the conditional mean of $\mathbf{x}_0$ given $\mathbf{x}_K$, denoted $\mathbb{E}(\mathbf{x}_0|\mathbf{x}_K)$, is not linear in $\mathbf{x}_K$, the correlation coefficient will not fully determine the form of $\mathbb{E}(\mathbf{x}_0|\mathbf{x}_K)$. In these cases, it may be advantageous using the nonlinear model to propagate the future information backward in time.

However, as discussed by Reynolds and Palmer (1998), the use of the backward model to recover initial states may be tricky, since the errors may grow in the backward integration faster than in the forward integration. This is especially important when the analysis from which the backward integration is initialized has large uncertainties.

Lastly, while the BS as well as the SEEK-smo may be executed off-line, i.e. using a assimilation run already performed, the iterative algorithms require the smoothing to be done on-line since they need the final covariance matrix to initialize the next filter/smoother cycle.

Back and Forth Kalman Filter - BFKF

Although not mentioned in the derivation of the iBS, one problem of iterating over the same set of observations is the introduction of correlated errors in the system. This would violate the independence hypothesis we have made in deriving the Kalman filters. Indeed, Yang et al. (2012) observed that using independent errors at each iteration, the RIP produces better results than using the correlated data sets. Hence, the iterative algorithms may be formally justified splitting the observational set into P subsets

according to:

$$p(\mathbf{y}_k|\mathbf{x}_k) = \prod_{p=1}^{P} p(\mathbf{y}_k|\mathbf{x}_k)^{\alpha_p} \tag{2.128}$$

with $\alpha_p > 0$ and $\sum_{p=1}^{P} \alpha_p = 1$. Similar decomposition is the basis of the well-known sequential processing of batches of observations used by KF algorithms to reduce the cost of inverting a matrix of the size of the observational space (Houtekamer and Mitchell, 2001). Proceeding like this implies that independent observational subsets are assimilated at each filter pass. Several possible solutions may be used, for instance the assimilation of different independent variables at each filter pass. However, this may not treat the balance problem.

To address these problems and considering that the observation errors are Gaussian, a decomposition like (2.128) may be achieved by multiplying the covariance matrix $\mathbf{R}$ by a scalar α_p^{-1}. With this form of decomposition, the observational error covariance is inflated, resulting in smaller increments and probably less intense unbalanced physics. Also, as already pointed, smaller increments seems to be more suitable for use along with extended filters. Indeed, this decomposition provides an optimal choice for the BFN gain and for the number of iterations.

Furthermore, the problem of underestimation of the background errors may be addressed by allowing more iterations than the optimal, given by the α_p^{-1}, but controlling the error reduction between two iterations as done for the BFN. In this case, the same batch of observation is used more than once. This can be justified by the fact that the reduced basis (the $\mathbf{S}$ matrix in the SEEK filter) may change between two iterations due to perturbations caused by the assimilation of future data and by the model dynamics; by this way the same batch of observations is assimilated more than once but using different subspaces. The iterations may have an additional advantage if after each iteration, the new estimated covariance matrix better represents the system unstable directions.

Although inspired in the BFN, the BFKF shares one similarity with the forward-backward smoother of Fraser and Potter (1969) that is the use of the backward model to back-propagate the observation content. However, differently from the forward-backward smoother, the BFKF does not rely on an information filter to back-propagate

the observation information, since the initial condition for the backward pass is simply the final condition of the forward filter.

The BFKF derivation starts from the final filter analysis obtained after processing the first batch of observations $\mathbf{y}_{1:K}^1$, i.e. we are given $p(\mathbf{x}_k|\mathbf{y}_{1:K}^1)$. The first step is to update this pdf using the observations at time t_K from the second batch $\mathbf{y}_{1:K}^2$. Using the decomposition (2.27) we can write the analysis step as:

$$p(\mathbf{x}_K|\mathbf{y}_{1:K}^1, \mathbf{y}_K^2) \propto p(\mathbf{y}_K^2|\mathbf{x}_K)p(\mathbf{x}_K|\mathbf{y}_{1:K}^1) \tag{2.129}$$

where we identify the likelihood $p(\mathbf{y}_K^2|\mathbf{x}_K)$ and the prior $p(\mathbf{x}_K|\mathbf{y}_{1:K}^1)$ which is known from the last filter analysis. Recalling this factorization is possible thanks to the errors independence and the Markovian assumption.

Then the propagation rule is used to obtain $p(\mathbf{x}_{K-1}|\mathbf{y}_{1:K}^1, \mathbf{y}_K^2)$. As already said, this can be a problematic step if the model is not considered perfect since in this case $p(\mathbf{x}_{K-1}|\mathbf{x}_K, \mathbf{y}_K)$ is different from $p(\mathbf{x}_{K-1}|\mathbf{x}_K)$. Hence, the forecast step is given by:

$$p(\mathbf{x}_{K-1}|\mathbf{y}_{1:K}^1, \mathbf{y}_K^2) = \int p(\mathbf{x}_{K-1}|\mathbf{x}_K)p(\mathbf{x}_K|\mathbf{y}_{1:K}^1, \mathbf{y}_K^2)d\mathbf{x}_K \tag{2.130}$$

which is then followed by the update step:

$$p(\mathbf{x}_{K-1}|\mathbf{y}_{1:K}^1, \mathbf{y}_K^2, \mathbf{y}_{K-1}^2) \propto p(\mathbf{y}_{K-1}^2|\mathbf{x}_{K-1})p(\mathbf{x}_{K-1}|\mathbf{y}_{1:K}^1, \mathbf{y}_K^2) \tag{2.131}$$

Therefore, the backward pass is identical to the forward pass except for the transition matrix that requires the knowledge of the backward model. Hence, the KF equations may be used in the backward pass likewise in the forward pass. The generalization of the steps just described gives rise to an iterative algorithm processing P batches of observations for a given time window $[t_0, t_K]$. At the same time this algorithm avoids using an information filter it provides a natural way to link two assimilation cycles, i.e processing the observations within $[t_0, t_K]$ and then $[t_{K+1}, t_{2K}]$, since it provides the mean and the covariance at the beginning of the assimilation window. The mean and covariance is then used in forecast mode until new observations are available.
A Bayesian description of the Back and Forth Kalman Filter is given below:

For p=1:2:P-1
Do
For k=1:K
Do

$$p(\mathbf{x}_k|\mathbf{y}_{1:K}^{1:p-1},\mathbf{y}_{1:k-1}^{p}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:K}^{1:p-1},\mathbf{y}_{1:k-1}^{p})d\mathbf{x}_{k-1} \tag{2.132}$$

$$p(\mathbf{x}_k|\mathbf{y}_{1:K}^{1:p-1},\mathbf{y}_{1:K}^{p}) = p(\mathbf{y}_k^p|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:K}^{1:p-1},\mathbf{y}_{1:k-1}^{p}) \tag{2.133}$$

End Do
The backward analysis and forecast reads:
For k=K:-1:1
Do

$$p(\mathbf{x}_k|\mathbf{y}_{1:K}^{1:p},\mathbf{y}_{k:K}^{p+1}) = p(\mathbf{y}_k^{p+1}|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:K}^{1:p},\mathbf{y}_{k+1:K}^{p+1}) \tag{2.134}$$

$$p(\mathbf{x}_{k-1}|\mathbf{y}_{1:K}^{1:p},\mathbf{y}_{k:K}^{p+1}) = \int p(\mathbf{x}_{k-1}|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:K}^{1:p},\mathbf{y}_{k:K}^{p+1})d\mathbf{x}_k \tag{2.135}$$

End Do
End Do

Table 2.6: Bayesian description of the Back and Forth Kalman Filter.

Although quite general, in this thesis the algorithm is implemented using the SEEK filter. Therefore in practice, its cost is at best twice the cost of the filter since it requires the integration of the evolutive basis forward and backward. To reduce the numerical cost it is quite reasonable to use a fixed basis within the iteration cycle, i.e. the forecast covariance matrix is simplified to $\mathbf{S}_{k|k-1} = \mathbf{I}\mathbf{S}_{k-1|k-1}$, and between two assimilation cycles the reduced basis are allowed to evolve according to the model dynamics. This is motivated by the fact that usually short assimilation windows ($5 - 20$days) are used and the errors structures may not change too much for these time windows. This makes the cost of one BFKF iteration to be the same as the SEEK-smo and the BS cost.

2.5.4 Probabilistic Four Dimensional Variational Method

The relationship between the deterministic 4Dvar, presented in section (2.3), and the probabilistic 4Dvar is made by noting that the cost function given by Eq.(2.6) may be obtained by computing the negative logarithm of the pdf $p(\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_N|\mathbf{y}_{1:N})$ under the Kalman filter hypotheses and the hypothesis of a perfect model, i.e. $p(\mathbf{x}_{k-1}|\mathbf{x}_k) = \delta(\mathbf{x}_{k-1} - \mathcal{M}_{k,k-1}(\mathbf{x}_k))$. Therefore, as noted by (Li and Navon, 2001), the 4Dvar is a

fixed-interval smoother and under the Kalman filter hypotheses is equivalent to the fixed-interval Kalman smoother. Moreover, the positive definite matrices $\mathbf{B}$ and $\mathbf{S}$ in Eq.(2.6) may be identified with the background and observation covariance matrices $\mathbf{P}$ and $\mathbf{R}$, respectively.

The incremental probabilistic 4Dvar is equivalent to the extended smoother derived from the KF. Furthermore, the RIP and QOL iterations, and hence the iBS, would be seen as a generalized outer loop as proposed by Yang et al. (2012).

2.6 Numerical Implementation

2.6.1 D-BFN

The Nudging implementation was conducted directly into the NEMO code. Several options are implemented such as the possibility to switch between forward and backward integrations with and without Nudging. Also, different forms of $\mathbf{K}$, ranging from simple scalars to reduced rank regression models, are available.

The code was constructed using C pre-processing. It means that the user may compile the same code with and without the BFN facilities. The implementation is designed to save I/O time but demands high virtual memory availability. All settings are made by configuring an input text file.

2.6.2 Kalman Filter/Smoothers

The SEEK applications used the software SESAM [2] developed and maintained by the MEOM [3] team. SESAM is a modular package written in FORTRAN95 composed by a set of operators required to perform sequential data assimilation. It is designed to perform square root or ensemble updates in their local or global form. In addition it provides facilities such as anamorphosis transformations, and the computation of truncated Gaussian estimators as well as EOF decompositions and regional RMS misfits, for instance.

[2] http://www-meom.hmg.inpg.fr/Web/Outils/SESAM/

[3] The "Modelling of large scale and mesoscale ocean flows" - MEOM - group is part of the Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), a joint research unit (UMR5183) of CNRS and UJF.

2.6.3 4Dvar

The 4Dvar implementation used in this thesis is the NEMOVAR, which is the variational data assimilation component of the NEMO model. It is developed on the basis of the VODA[4] project which is supported scientifically and financially by the Agence Nationale de la Recherche (French research agency), through the call "Conception et simulation". The experiment set-up was performed thanks to Pierre-Antoine Bouttier from the MEOM team.

2.7 Conclusions

In this chapter the linear theory of deterministic observers was presented as well as their extension to linear and weakly nonlinear stochastic systems. While the LO has the KF as a possible corresponding stochastic algorithm, the iterative deterministic observer, BFN, was extended to stochastic systems by applying the Bayesian formalism. Under the KF hypotheses, the BFN may be optimized to be an iterative KF, which is indeed a smoother algorithm. Therefore, while nudging can be seen as sub-optimal Kalman filter the BFN can be seen as a sub-optimal smoother.

The BS as well as the Back and Forth filtering were derived in a quite general way. This means that their use is not limited to the KF theory, but they may be applicable to a more general methods (e.g. particle methods) that somehow estimate $p(\mathbf{x}_k|\mathbf{y}_{1:k})$. The hypothesis made by the BF filtering and the BS is that we know how to calculate the transition pdf $p(\mathbf{x}_{k-1}|\mathbf{x}_k)$. Of course, it is necessary to have a reliable approximation of the inverse model, which can be very complicated when the system is diffusive.

Similarities and differences between the algorithms were explored. We have seen that when the system is linear and the pdfs are Gaussian, the BS composed by a forward SEEK filter followed by a backward model integration is strictly equivalent to the SEEK smoother. Also, when these hypotheses are verified and considering one assimilation cycle composed by one analysis step, the BS and the inverse 3Dvar are similar. They both use the backward model to propagate observation informations backward in time.

The BS was extended to consider iterations. Two algorithms were proposed: one is designed to improve the mean while keeping the original covariance (iBSFU) and

[4]Variational Ocean Data Assimilation for multiscale applications: http://voda.gforge.inria.fr/index.html

another is designed to improve both the mean and the covariance (iBS). It was argued that due to the equivalence between the backward integration and the SEEK smoother under linear conditions, the iBS would be similar to the RIP algorithm (Yang et al., 2012) while the iBSFU would be similar to the QOL.

Finally, we derived a stochastic version of the BFN. Under the Kalman's hypothesis the resulting algorithm is the Back and Forth Kalman Filter, i.e. the BFN but with the gains calculated thanks to the KF analysis equations. As for the KF, variants of the BFKF were also proposed, i.e. the BF filtering with a fixed basis, and the BF filtering with a semi-evolutive basis. In the latter case, the covariance matrix is not dynamically propagated within the assimilation window but it does propagate between two assimilation cycles.

Lastly, comments were made about the equivalence between the probabilistic 4Dvar and the fixed-lag Kalman smoother. It is important to note that although the methods are similar or equivalent in theoretical terms they may not produce the same results in practice. Factors as covariance localization and inflation as well as nonlinearity may strongly influence the results.

The configuration of each algorithm presented concerning the choice of the gains, the covariance matrix initialization and the observations network will be presented in the chapters on which the numerical results are presented. In chapter 4 the DBFN configured to use a scalar gain and a gain based on steady correlations is implemented. The results are compared with those produced by the 4Dvar method. In chapter 5, results produced by the SEEK filter and smoother and their iterative version "RIP" are compared with results produced by the BS, iBS and the BFKF.

Chapter 3

Ocean model

Contents

3.1 Introduction

The ocean model used in this study is the ocean component of NEMO (Nucleus for European Modeling of the Ocean; Madec, 1996). This model is able to represent a wide range of ocean motions, from basin scale up to regional scale. Currently, it has been used in operational mode at the French Mercator Océan project (http://www.mercator-ocean.fr).

In this chapter the mathematical formulation of the model as well as its main numerical approximations are presented. The chapter ends with a discussion about numerical aspects of the backward model and the model set-up used in the chapter 4 and 5 experiments.

3.2 Primitive Equation Ocean Model

Modelling the ocean requires the knowledge of five variables: density ρ, pressure P, velocity $\mathbf{u}$, potential temperature T and salinity S. The evolution of these quantities is governed by the fundamental conservation laws, i.e momentum, mass and energy conservation laws.

The conservation of momentum is governed by the Newton's equation of motion expressed as the Navier-Stokes equations for a fluid element located at (x, y, z) on the surface of our rotating planet and moving at velocity (u, v, w) relative to that surface:

$$\frac{D\mathbf{u}}{Dt} = -\frac{\nabla P}{\rho} + \mathbf{g} - 2\mathbf{\Omega} \times \mathbf{u} + \mathbf{D} + \mathbf{F}, \tag{3.1}$$

where t is the time, $\mathbf{u} = \mathbf{u}_h + w\mathbf{k}$ is the velocity vector, $\mathbf{g}$ is the gravitational acceleration, $\mathbf{\Omega}$ is the rotational vector, and $\mathbf{D}$ and $\mathbf{F}$ are the dissipation and forcings terms. The operator $\frac{D}{Dt}$ is the total derivative which includes time and local variations. It is given by:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u}.\nabla.$$

The conservation of mass is given by the continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla\left(\rho \mathbf{u}\right) = 0 \tag{3.2}$$

The equations for the conservation of salt and potential temperature are derived from the first thermodynamical law. The final equations can be written as:

$$\frac{\partial T}{\partial t} = -\nabla\left(\mathbf{u}T\right) + \mathbf{D}^T + \mathbf{F}^T \tag{3.3}$$

$$\frac{\partial S}{\partial t} = -\nabla\left(\mathbf{u}S\right) + \mathbf{D}^S + \mathbf{F}^S \tag{3.4}$$

The system is closed by a state equation linking density, temperature, salinity and pressure:

$$\rho = \rho(T, S, P) \tag{3.5}$$

NEMO also considers two physically-based hypotheses that simplify the system of equations given by Eq.(3.1) and (3.2):

Hypothesis 1: Boussinesq and incompressible ocean

The Boussinesq approximation considers that density variations are much smaller than the mean density

$$\frac{\delta\rho}{\rho} << 1$$

Therefore, the ocean density is considered constant, i.e. a mean density ρ_0 replaces ρ in the equations, except in the buoyancy term for which the density is multiplied by the gravity.

This hypothesis implies that the equation for the conservation of mass (Eq.3.2) resumes to

$$\nabla\mathbf{u} = 0 \tag{3.6}$$

and accordingly, the flow is non-divergent. Furthermore, it also implies the ocean's incompressibility. Considering the compressibility Γ as:

$$\Gamma = \frac{1}{\rho}\frac{\partial\rho}{\partial P} \tag{3.7}$$

and putting it in a differential form, we obtain:

$$\frac{\Delta\rho}{\rho} = \Gamma\Delta P << 1. \tag{3.8}$$

Since ΔP would attain $10^8 Pa$ for the ocean, $\Gamma << 1$ and therefore the Boussinesq approximation leads to the incompressibility condition.

Hypothesis 2: Hydrostatic ocean
Considering the aspect ratio $\frac{H}{L} << 1$, where L is the horizontal scale and H is the vertical scale, the equation for the evolution of w is resumed to the hydrostatic equilibrium:

$$\frac{\partial P}{\partial z} = -\rho \mathbf{g} \tag{3.9}$$

With this approximation the acceleration term, $\frac{\partial w}{\partial t}$, is neglected implying a misrepresentation of gravitational flows as well as of vertical convection processes. Indeed, the vertical component of the velocity field is a diagnostic variable calculated thanks to the continuity equation under the Boussinesq approximation:

$$\frac{\partial w}{\partial z} = -\nabla_h . \mathbf{u} \tag{3.10}$$

3.3 The Model discretization

The model's equation under their continous form have to be discretized in order to be numerically solved.

3.3.1 Spatial discretization

To discretize the model's equation, a coordinate system and a discretization grid have to be chosen. For the first the spherical earth approximation is considered. In this case, the geopotential surfaces are assumed to be spheres so that gravity (local vertical) is parallel to the Earth's radius. Since the gravitational force is so dominant in the equations of large-scale motions, it is useful to choose an orthogonal set of unit vectors (i, j, k) linked to the earth such that k is the local upward vector and (i, j) are two vectors orthogonal to k, i.e. tangent to the geopotential surfaces.

In addition, the thin-shell approximation is also considered. This approximation consists of neglecting the ocean depth from the radial coordinate. By this way, the local effect of curvature due to the topography variations is neglected in the model.

The numerical techniques used to solve the Primitive Equations are based on the

traditional centered second-order finite difference approximation. Special attention has been given to the homogeneity of the solution in the three space directions. The arrangement of variables is the same in all directions. It consists of cells centered on scalar points (T, S, p, ρ) with vector points (u, v, w) defined in the center of each face of the cells (3.1). This is the well-known "C" grid in Arakawa's classification (Mesinger and Arakawa, 1976) generalized to the three dimensional space. The relative and planetary vorticity, ζ and f , are defined in the center of each vertical edge and the barotropic stream function ϕ is defined at the horizontal points overlying the ζ and f -points.

3.3.2 Temporal discretization

The model is solved using a centered finite difference in time. For the non-diffusive terms the well-known *leapfrog* scheme is used:

$$\mathbf{x}^{t+\Delta t} = \mathbf{x}^{t-\Delta t} + \Delta t \mathcal{F}(\mathbf{x}^t) \tag{3.11}$$

where $\mathbf{x}$ denotes the state variables, the superscripts the time level and $\mathcal{F}$ represents the non-diffusive part of the equations' right hand side.

For sake of stability the diffusion terms ($\mathcal{F}_{hor\_diff}$) are discretized using non-centered schemes. For the horizontal diffusion a forward Euler scheme is used:

$$\mathbf{x}^{t+\Delta t} = \mathbf{x}^{t-\Delta t} + 2\Delta t \mathcal{F}_{hor\_diff}(\mathbf{x}^{t-\Delta t}) \tag{3.12}$$

This system is conditionally stable for a diffusion coefficient $A^h < \Delta x^2/(\pi^2 2\Delta)$. For the vertical diffusion ($\mathcal{F}_{vert\_diff}$) an implicit scheme (backward Euler) is used:

$$\mathbf{x}^{t+\Delta t} = \mathbf{x}^{t-\Delta t} + 2\Delta t \mathcal{F}_{vert\_diff}(\mathbf{x}^{t+\Delta t}) \tag{3.13}$$

This scheme is computationally expensive but necessary because of the time step constraint imposed by the explicit method.

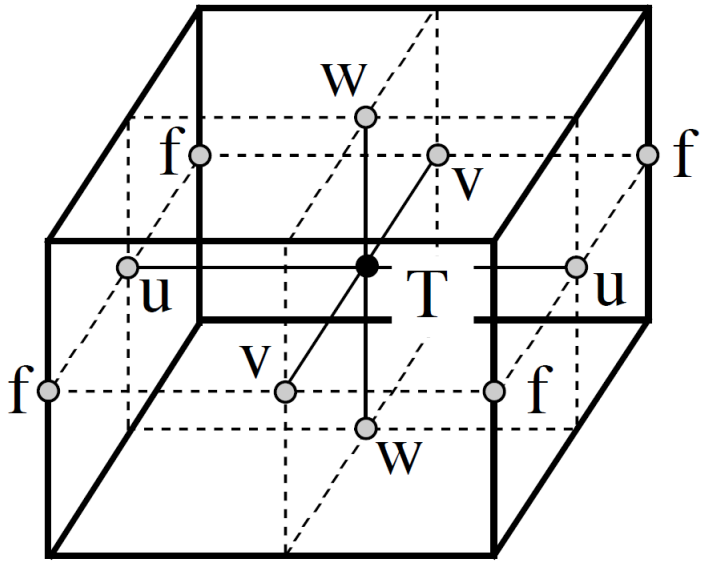


Figure 3.1: Arrangement of variables. T indicates scalar points where temperature, salinity, density, pressure and horizontal divergence are defined. (u, v, w) indicates vector points, and f indicates vorticity points where both relative and planetary vorticity are defined.

3.4 Model parameterizations

As we already saw in this chapter, the continuous equation which represents an hydrostatic and Boussinesq ocean should be further discretized to be numerically solved, since there is not analytical solution for this general problem. The discretized equations solve only a portion of the system's spectrum, since a non-infinitesimal Δx and Δt must be chosen. In the spectral space it can be seen as a low-pass filter, since it cuts-off the high wave numbers. The scales that are not explicitly represented by the model are called subgrid scales.

Due to the turbulent nature of the system, energy cascade from large scales to small scales and eventually from small to large scales. Therefore, the non-resolved portion of this cascade should be parametrized. In the direct cascade case (i.e. energy cascade from large to small scale) energy tend to accumulate in the wave number corresponding to the size of the mesh element. In the continuous case, it is the diffusion operator that acts in the molecular level transforming mechanical energy into heat. In the case of numerical models, a common practice is to consider that the turbulent fluxes depend

linearly on the large scale gradients. In this case a diffusion operator similar to the one of the continuous equation is used but with a diffusion coefficient several orders of magnitude greater than those representing the molecular diffusion. The correct specification of the diffusion coefficients is essential for the correct representation of the spectrum that is explicitly resolved by the model since it controls the energy flux from small to large scales (inverse cascade) as well as it acts directly on the resolved scales smoothing the field gradients.

The study of the subgrid scales parameterization has been the subject of many articles (Wallcraft et al., 2005; Frederiksen and O'Kane, 2008; Le Sommer et al., 2011; Kitsios et al., 2013). There is no consensus on how to address this issue and usually ad-hoc solutions are used. Some effort has been made to choose the diffusion coefficients from the point of view of the optimal control (Leredde et al., 1999) and stochastic perturbations (Frederiksen et al., 2012; Kitsios et al., 2013)

3.4.1 Horizontal physics

The horizontal diffusion and viscosity terms are usually modeled by a n order operator $(-i)^n \nu_n \nabla^n$. It is interesting to note that an order n operator has a decaying time that is inversely proportional to the wavelength raised to the power n. Therefore, the higher the operator order, the lower its effects on the large scale features. In this study, the model was configured with a fourth order operator ($n = 4$), which preserves the mesoscale features at the same time that it ensures numerical stability.

3.4.2 Vertical physics

The vertical diffusion and viscosity terms are modeled by a Laplacian operator. The vertical turbulent fluxes also depend linearly on the vertical gradients of the large scale variables:

$$\mathbf{D}^{vU} = \frac{\partial}{\partial z}\left(\nu^{v,u}\frac{\partial u_h}{\partial z}\right) \tag{3.14}$$

$$\mathbf{D}^{vT} = \frac{\partial}{\partial z}\left(\nu^{v,T}\frac{\partial T}{\partial z}\right) \tag{3.15}$$

$$\mathbf{D}^{vS} = \frac{\partial}{\partial z}\left(\nu^{v,S}\frac{\partial S}{\partial z}\right) \tag{3.16}$$

3.5 Boundary conditions

3.5.1 Lateral boundary

Over the closed lateral boundaries normal velocities are set to zero. For tangential velocities the boundary conditions may be free-slip or non-slip. For the former, the velocity at the wall is the same as off the wall, and for the latter, the velocity at the wall is zero, implying a linear decreasing of the velocity and the formation of a boundary layer.

3.5.2 Surface boundary

The surface boundary conditions are related to the exchange of properties between the ocean and the atmosphere. More specifically, they model the exchanges of momentum (Eq.3.17), heat (Eq.3.18) and water vapor (Eq.3.19).

$$\mathbf{D}^{U}_{z=0} = \left(\nu^{v,u} \frac{\partial \mathbf{u}_h}{\partial z} \right)_{z=0} = \frac{\tau}{\rho_0} \tag{3.17}$$

$$\mathbf{D}^{T}_{z=0} = \left(\nu^{v,T} \frac{\partial T}{\partial z} \right)_{z=0} = \frac{Q_{net}}{\rho_0 C_{pw}} \tag{3.18}$$

$$\mathbf{D}^{S}_{z=0} = \left(\nu^{v,S} \frac{\partial S}{\partial z} \right)_{z=0} = \gamma S|_z = 0 \tag{3.19}$$

In the above equations ρ_0 is a reference seawater density, τ is the wind stress at the ocean surface, Q_{net} is the net heat flux at the air-sea interface, C_{pw} is the sea water heat capacity and γ is the fresh water net flux at the air-sea interface.

3.5.3 Bottom boundary

It is considered that there is no flux across the bottom. However, momentum is lost by friction. Therefore, a boundary condition may be derived by using the equation for the vertical diffusion of momentum (Eq. 3.14) and a function $\mathbf{F}_h(u_b, v_b)$ that is quadratic in the bottom velocities. Hence the bottom boundary condition is written as:

$$\mathbf{D}^{U}_{z=-H} = \left(\frac{\partial \mathbf{F}_h}{\partial z} \right)_{z=-H} \tag{3.20}$$

$$\mathbf{F}_h = \left(\nu^{v,u} \frac{\partial \mathbf{u}_h}{\partial z} \right)_{z=-H} = C_d \sqrt{u_b^2 + v_b^2} \mathbf{u}_h^b \tag{3.21}$$

where C_d is friction coefficient.

3.6 The backward integration

3.6.1 Numerical aspects

One important aspect of the BFN and the other methods that uses the backward model is their simplicity of implementation. For this to be true, the same code used to integrate the model forward should be used to integrate the model backward. This is possible if the discretized backward model is built from the continuous backward equation.

In this section we discuss the difference between discretizing the continuous backward model and constructing the backward model by using the discrete direct model. The latter would provide the exact inverse model with respect to the direct model used in the simulations. We start by taking two discretization already presented: the *leapfrog* which is a three level discretization (Eq.3.11) and the Backward Euler (Eq.3.13), which will be called Euler to avoid confusion with the sense of integration. Both are centered in time but for the former the terms not involving time derivatives are evaluated at time t and for the latter at time $t + \Delta t$.

Writing the continuous forward model as:

$$\frac{dX}{dt} = F(X) \tag{3.22}$$

the *leapfrog* discretization is given by:

$$X^{t+\Delta t} = X^{t-\Delta t} + 2\Delta t F(X^t) \tag{3.23}$$

while the Euler scheme is given by:

$$X^{t+\Delta t} = X^{t-\Delta t} + 2\Delta t F(X^{t+\Delta t}). \tag{3.24}$$

The continuous backward model, in its turn, can be written as:

$$-\frac{dX}{dt} = F(X) \tag{3.25}$$

hence, the *leapfrog* is given by:

$$X^{t-\Delta t} = X^{t+\Delta t} - 2\Delta t F(X^t) \tag{3.26}$$

and the Euler scheme is given by:

$$X^{t-\Delta t} = X^{t+\Delta t} - 2\Delta t F(X^{t-\Delta t}) \tag{3.27}$$

If the backward model is constructed by using Eq.(3.23) and (3.24) instead, the following is obtained:

$$X^{t-\Delta t} = X^{t+\Delta t} - 2\Delta t F(X^t) \tag{3.28}$$

and

$$X^{t-\Delta t} = X^{t+\Delta t} - 2\Delta t F(X^{t+\Delta t}) \tag{3.29}$$

Therefore, when the *leapfrog* scheme is used, discretizing the continuous backward model (Eq.3.23) or "inversing" the forward discretized model (Eq.3.28) gives the same. This is a consequence of the method's symmetry. However when non-symmetric schemes are considered, implicit methods (Eq.3.24) become explicit (Eq.3.29) and explicit methods become implicit. In these cases, the implementation of the backward model would require a lot of programming.

As we have seen, most of the model is solved using the *leapfrog* scheme. Hence, the discrete version of the continuous backward model is quite similar to the backward version based on the discrete direct model. The only distinction between the "true" discrete backward model and our discrete backward model concerns the diffusive terms. However as already discussed, for the sake of stability the sign of the diffusive term is reversed when stepping backwards. Therefore, the only terms for which the simple change of the time-step sign would not produce the "true" backward model are already changed by a stronger hypothesis. Therefore, aside technical details regarding interpolation functions, and the diffusive/dissipative terms, the only change needed to run the model backwards is to set a negative time step.

The practical aspects of these approximations are evaluated in the chapter 4, section 4.5.

3.7 Model configuration

The double gyre configuration, extensively used for the study of jet instabilities (Chassignet and Gent, 1991; Primeau, 1998; Chang et al., 2001), meso and submeso-scale dynamics (Levy et al., 2010) and for data assimilation methods (Molcard et al., 2004; Krysta et al., 2011; Cosme et al., 2010) is used for the present study. The double gyre configuration simulates the ocean middle latitude dynamics and has the advantage of being simple, when compared to real applications, but still considering full dynamics and thermodynamics.

In our experiments we use a homogeneous horizontal grid with a 25km resolution and a vertical resolution ranging from 100m near the upper surface up to 500m near the bottom. The bottom topography is flat and the lateral boundaries are closed and frictionless. The only forcing term considered is a constant wind stress of the form $\tau = (\tau_0 cos\left(\frac{2\pi(y-y_0)}{L}\right), 0)$, where $L = 2000km$ and $\tau_0 = 0.1N/m^2$. Horizontal diffusion/viscosity are modeled by a bilaplacian operator meanwhile a laplacian operator is used in the vertical. They all use constant coefficients in time and space: $\nu_h^{u,v} = -8 \times 10^{10} m^4/s$ and $\nu_v^{u,v} = 1.2 \times 10^{-4} m^2/s$ for the momentum equations and $\nu_h^{t,s} = -4 \times 10^{11} m^4/s$ and $\nu_v^{t,s} = 1.2 \times 10^{-5} m^2/s$ for temperature and salinity.

The initial condition is similar to that used by (Chassignet and Gent, 1991) and consists of a homogeneous salinity field of 35psu and a temperature field created to provide a stratification which has a first baroclinic deformation radius of 44.7km. Velocity and pressure fields are initially set to zero. The model was integrated for 70 years, in order to reach the statistical steady state. Figure 3.2 shows the SSH field for 2 days simulation and for 70 years. Note that statistical steady state does not mean physical steady state. After 70 years it can be noted a meandering jet and eddy structures that are typical of a mid latitude ocean mesoscale features. Afterwards, ten years of free model run were performed, that are used to calculate the models statistics, when they are necessary, and then two additional years were finally performed to be used as the truth, from which the observations were extracted.

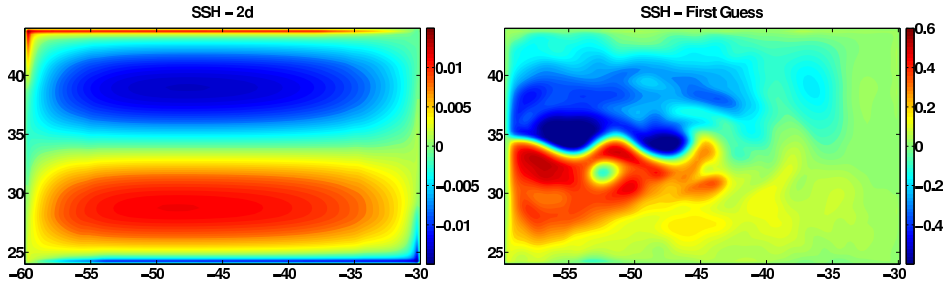


Figure 3.2: Figures show the double gyre formation and evolution. Left panel: SSH calculated after 1 day simulation. Right panel: SSH calculated after 70 years simulation. For the latter, it is observed a meandering jet and eddy structures.

This work was granted access to the HPC and visualization resources of "Centre de Calcul Interactif" hosted by "Université Nice Sophia Antipolis" and GENCI.

Chapter 4

Diffusive Back and Forth Nudging Experiments

Contents

4.1 Introduction

In this chapter we present a set of experiments performed using the Diffusive Back and Forth Nudging in its from described in the section 2.4.3. In this formulation no optimality conditions is imposed to calculate the Nudging gains. Our objective is to extend the previous studies on the BFN/DBFN applied to simplified ocean models to a primitive equation ocean model, in our case the NEMO ocean model.

Previous studies employing the BFN along with ocean models have used a Shallow Water (SW) model and a Layered Quasi-Geostrophic (LQG) model (Auroux and Blum, 2008; Auroux, 2009). For both models, significant reductions in the initial condition errors were reported. The highlighted aspects are: i) the simplicity of implementation, there is no need to implement neither the tangent linear model nor the adjoint model; ii) the fast convergence when compared with the 4Dvar; iii) the capability of providing accurate initial state estimations under sparse observational conditions using simple scalar gains; iv) low sensitivity to noisy observations and notably v) the BFN performance is comparable to the 4Dvar performance.

The BFN application to control a primitive equation ocean model represents a new challenge due to the increased model complexity. Among the differences between NEMO and the simplified oceanic models used by Auroux and Blum (2008) and Auroux (2009) stand out the more complex relationship between the variables in the former since no filtering technique is used in the derivation of the physical model (except the Boussinesq approximation which is also considered by the SW and LQG models), and the inclusion of an equation for the conservation of the thermodynamical properties. The latter requires the use of a nonlinear state equation to couple dynamical and thermodynamical variables.

Furthermore, the vertical ocean structure represented by NEMO is more complex than the vertical ocean structure represented by the SW and LQG used by Auroux and Blum (2008); Auroux (2009). This is because NEMO considers more vertical degrees of freedom, since the SW model has no vertical levels and the LQG was implemented with only 3 layers, as well as it considers vertical diffusion processes, mostly ignored by the LQG model. Vertical diffusion plays an important role in maintaining the ocean stratification and meridional overturning circulation, which is directly related to the transport

of heat in the ocean. Moreover from the practical point of view, the diffusion/viscosity required to keep the NEMO simulations stable is by far greater than for the SW or LQG at the same resolution. Thus, we are particularly concerned about the diffusive aspect of the algorithm since for oceanic simulations water masses transformation is an important issue. A reliable representation of the ocean state has to reproduce the appropriate temperature/salinity distribution along with the underlying dynamics.

In general, ocean observations are sparse and restricted to measurements of surface properties such as sea surface temperature (SST) and sea surface height (SSH). For large scale applications some vertical profiles measuring temperature and salinity are also available. However, a synoptic picture based on observations of the ocean state is impossible to construct. In this chapter we are interested in the assimilation of SSH data which is one of the most abundant, precise and informative oceanic observations. Although the SSH is an indirect measurement of the surface pressure it may "observe" thermocline-related motions. Actually, the SSH is a product of fast barotropic motions and slow baroclinic motions that depends directly on the vertical stratification. The role of DA methods is not only to correct the SSH fields but properly project these corrections onto their corresponding stratification modes. In other words, the assimilation of SSH should be able to reconstruct the principal vertical modes maintaining the mass structure consistent with the real state.

Therefore, we raise some fundamental questions concerning the Data Assimilation system NEMO-DBFN:

1. How accurate is the diffusive backward model?
2. Is the assimilation of SSH data able to control the backward integration and produce accurate estimation of the whole state space?
3. What is the role of diffusion in the method's performance?
4. Is it possible to use DA windows that are long enough to take advantage of the time-reversibility of the non-linear terms?

To answer these questions we organized this chapter as follows. In Sect 4.2 the DBFN gain configuration as well as the 4Dvar configuration are described. Section 4.3 details the experimental design. Section 4.4 discusses the physics of the Sea Surface Height

and Sect practical aspects of the backward integration. The data assimilation results are presented in Sect 4.6, for which five specific aspects are analyzed:

- sensitivity of the BFN to the length of the assimilation window (Sect 4.6.1) ;
- sensitivity to the model diffusion coefficients (Sect 4.6.1);
- effects of the number of iterations (Sect 4.6.1);
- importance of the nudging gain structure (Sect 4.6.2);
- impact of spatial and temporal observation distribution on the method performance (Sect 4.6.2 and 4.6.2).

This section ends with a comparison between the BFN and 4Dvar for the case where observations are distributed so as to simulate an altimeter satellite track.

4.2 Data Assimilation methods

To assist the reader in reading this chapter we present again the DBFN and the incremental 4Dvar.

4.2.1 Diffusive Back and Forth Nudging - DBFN

Let us assume the time continuous model satisfies dynamical equations of the form:

$$\frac{\partial \mathbf{x}}{\partial t} = \mathcal{F}(\mathbf{x}) + \nu \Delta \mathbf{x}, \qquad \text{for} \qquad 0 < t < T, \tag{4.1}$$

with an initial condition $\mathbf{x}(0) = \mathbf{x}_0$, where $\mathcal{F}$ denotes the nonlinear model operator without diffusive terms, ν is a diffusion coefficient and Δ represents a diffusion operator. If Nudging is applied to the forward system 4.1 it gives:

$$\frac{\partial \mathbf{x}_k}{\partial t} = \mathcal{F}(\mathbf{x}_k) + \nu \Delta \mathbf{x}_k + \mathbf{K}(\mathbf{y} - \mathcal{H}(\mathbf{x}_k)) \tag{4.2a}$$

$$\mathbf{x}_k(0) = \tilde{\mathbf{x}}_{k-1}(0), \qquad 0 < t < T, \tag{4.2b}$$

where $k \in \mathbb{N}_{\geq 1}$ state for iterations. Nudging applied to the backward system with the reversed diffusion sign gives:

$$\frac{\partial \tilde{\mathbf{x}}_k}{\partial t} = \mathcal{F}(\tilde{\mathbf{x}}_k) - \nu \Delta \tilde{\mathbf{x}}_k - \mathbf{K}(\mathbf{y} - \mathcal{H}(\tilde{\mathbf{x}}_k)) \tag{4.3a}$$

$$\tilde{\mathbf{x}}_k(T) = \mathbf{x}_k(T), \qquad T > t > 0. \tag{4.3b}$$

where the backward nudging gain is assumed to be equal the forward nudging gain, i.e $\mathbf{K}' = \mathbf{K}$. The system composed by equations (4.2) and (4.3) is the basis of the DBFN algorithm. They are iterated until convergence.

Therefore, one important aspect of the DBFN algorithm is the convergence criterion. Ideally, at convergence the nudging term should be null or small comparable to the other equation terms. Otherwise, when the nudging is switched off, which is the case in the forecast phase, the system may return to a state close to the background state or to a state which is not consistent to the one at convergence. The convergence is calculated as:

$$\frac{\|\mathbf{x}_k(t=0) - \mathbf{x}_{k-1}(t=0)\|}{\|\mathbf{x}_{k-1}(t=0)\|} \leq \epsilon, \tag{4.4}$$

where the choice for $\epsilon = 0.005$ is chosen based on sensitivity tests not presented in this thesis. An example of the impact of this choice is given in Sect. 4.5.1.

In this study $\mathbf{K}$ is considered as diagonal and constructed using Regression Models (RM). Our choice for a diagonal gain is based on the encouraging results found by Auroux and Blum (2008) using the BFN with simplified ocean models. If $\mathbf{K}$ is interpreted as the Kalman gain (Eq. 2.64), it is diagonal when covariances between model variables are ignored and the covariance matrix of the observation errors is diagonal, the latter being a common assumption in DA applications (Pham, 2001; Brankart et al., 2010). In this case, only the observed part of the state space is directly controlled. Nevertheless, corrections of the non-observed variables are done by the model itself. Concerning regression models, the algorithm operates in two steps: first the observed variables are updated and subsequently the other state variables are calculated using linear regression. The Partial Least Squares (PLS) regression (Tenenhaus, 1998) is used. The gain $\mathbf{K}$ is kept constant over the assimilation cycles. Our updating scheme can be seen as a rough approximation of the two steps update for EnKF presented by Anderson (2003).

Partial Least Squares regression (PLS) The PLS was first introduced by Wold (1975) to address the problem of econometric path modeling, and was subsequently adopted for regression problems in chemometric and spectrometric modeling. In the method description, $\mathbf{X} \in \mathbb{R}^{n\times M}$ is considered as the observed or predictor variables and $\mathbf{Y} \in \mathbb{R}^{n\times N}$ as the non-observed or response variables. In our notation n is the sample size and M and N are respectively the size of the state space of $\mathbf{X}$ and $\mathbf{Y}$. Besides, $\mathbf{X}$ and $\mathbf{Y}$ are centered and have the same units. The PLS regression features two steps: a dimension reduction step in which the predictors from matrix $\mathbf{X}$ are summarized in a small number of linear combinations called "PLS components". Then, that components are used as predictors in the ordinary least-squares regression.

The PLS as well as the principal component regression can be seen as methods to construct a matrix of p mutually orthogonal components $\mathbf{t}$ as linear combinations of $\mathbf{X}$:

$$\mathbf{T} = \mathbf{XW}, \tag{4.5}$$

where $\mathbf{T} \in \mathbb{R}^{n\times p}$ is the matrix of new components $\mathbf{t}_i = (t_{1i}, ..., t_{ni})^T$, for $i = 1, ..., p$, and $\mathbf{W} \in \mathbb{R}^{M\times p}$ is a weights matrix satisfying a particular optimality criterion.

The columns $\mathbf{w}_1, ..., \mathbf{w}_p$ of $\mathbf{W}$ are calculated according to the following optimization problem:

$$\mathbf{w}_i = \arg\max_{\mathbf{w}} \{cov(\mathbf{Xw}, \mathbf{Y})^2\} \tag{4.6}$$

subject to $\mathbf{w}_i^T\mathbf{w}_i = 1$ and $\mathbf{w}_i^T\mathbf{X}^T\mathbf{X}\mathbf{w}_j = 0$ for $j = 1, ..., i-1$.

The PLS estimator $\hat{\mathbf{B}}^{PLS}$ is given by:

$$\hat{\mathbf{B}}^{PLS} = \mathbf{W}(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W})^{-1}\mathbf{W}^T\mathbf{X}^T\mathbf{Y} \tag{4.7}$$

An immediate consequence of Eq. (4.7) is that when $\mathbf{W} = \mathbf{I}$ the Ordinary Least Squares (OLS) solution is obtained.

The number of components p is chosen from cross-validation. This method involves testing a model with objects that were not used to build the model. The data set is divided in two contiguous blocks; one of them is used for training and the other to validate the model. Then the number of components giving the best results in terms of

mean residual error and estimator variance is sought.

More details concerning the theoretical and practical differences between the PLS and OLS is given in the appendix 4.8.

4.2.2 Four Dimensional Variational Method - 4DVar

In this chapter the multi-incremental 4Dvar algorithm implemented in NEMO, called NEMOVAR (Mogensen et al., 2009), is used. Recalling the objective of the variational methods is to minimize a cost function that measures the distance between the estimated state and the available observations. In the incremental algorithm, the optimal initial state is found by minimizing the following cost function with respect to the increment $\delta\mathbf{x}_0$ defined by $\mathbf{x}_0 = \mathbf{x}^b + \delta\mathbf{x}_0$.:

$$\begin{aligned} J(\delta\mathbf{x}_0) = &\frac{1}{2}\delta\mathbf{x}_0^T\mathbf{B}^{-1}\delta\mathbf{x}_0 \\ &+ \frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i\mathbf{M}_{0,i}\delta\mathbf{x}_0 - \mathbf{d}_i)^T\mathbf{R}_i^{-1}(\mathbf{H}_i\mathbf{M}_{0,i}\delta\mathbf{x}_0 - \mathbf{d}_i) \end{aligned} \tag{4.8}$$

where $\mathbf{d}_i = \mathbf{y}_i - \mathcal{H}_i(\mathcal{M}_{0,i}(\mathbf{x}_0))$ is called the innovation vector, and $\mathbf{H}$ and $\mathbf{M}$ are the observation and model operators linearized in the neighborhood of the background trajectory. It is possible that after some iterations of the minimizer the increments become too large and a new linearization of $\mathcal{H}$ and $\mathcal{M}$ should be done. This gives rise to what is called the inner loop and outer loop iterations. The algorithm implemented in NEMO, called NEMOVAR (Mogensen et al., 2009), uses this technique.

The $\mathbf{B}$ matrix used in this chapter was built following Weaver et al. (2005). In this formulation the matrix is decomposed as $\mathbf{B} = \mathbf{G}\mathbf{\Lambda}^T\mathbf{C}\mathbf{\Lambda}\mathbf{G}^T$, where $\mathbf{G}$ is a multivariate balance operator, $\mathbf{\Lambda}$ is a diagonal matrix of error variance, for which the climatological variances are the entries, and $\mathbf{C}$ is a univariate correlation matrix modeled using the generalized diffusion equation. The balance operator is meant to propagate information from the observed variable to the non-observed variables. It is composed by a set of linear and non-linear relationships between the state variables such as the geostrophic balance and some temperature and salinity constraints, for example. The matrix $\mathbf{R}$ is diagonal.

4.3 Data Assimilation Experiments

This study addresses five main aspects of the DBFN: i) sensitivity to the length of the assimilation window, ii) sensitivity to the model diffusion coefficients, iii) effects of the number of iterations, iv) importance of the nudging gain structure and v) impact of spatial and temporal observations distribution on the performance.

A first set of experiments, summarized in Table 4.1 and presented in Sect 4.6.1, are designed to cover the first three main aspects. The data used are daily SSH or velocity fields available at every grid point and perturbed with a Gaussian white noise with a signal-to-noise ratio of 20%. The SSH assimilation at every grid point would be similar to assimilate gridded products such as those produced by AVISO (www.aviso.oceanobs.com) as done by Zavala-Garay et al. (2012). The difference in this case is that the observation errors may be correlated and in our case they are independent. The nudging gain is a scalar chosen to be strong enough to control the model errors but without being the dominant term of the equations.

	$K(1/s)$	$\nu_h^{u,v}(m^4/s)$	$\nu_h^{t,s}(m^4/s)$	stop criterions	Assim. Variables
ssh_xxd_dd	$1.5e^{-4}$	$-8e^{10}$	$-4e^{11}$	0.5%	SSH
ssh_xxd_rd	$1.5e^{-4}$	$-8e^{9}$	$-4e^{10}$	0.5%	SSH
uv_xxd_dd	$1.5e^{-6}$	$-8e^{10}$	$-4e^{11}$	0.5%	UV
uv_xxd_rd	$1.5e^{-6}$	$-8e^{9}$	$-4e^{10}$	0.5%	UV
ssh_xxd_dd_2it	$1.5e^{-4}$	$-8e^{10}$	$-4e^{11}$	2it	SSH
ssh_xxd_rd_2it	$1.5e^{-4}$	$-8e^{9}$	$-4e^{10}$	2it	SSH
uv_xxd_dd_2it	$1.5e^{-6}$	$-8e^{10}$	$-4e^{11}$	2it	UV
uv_xxd_rd_2it	$1.5e^{-6}$	$-8e^{9}$	$-4e^{10}$	2it	UV

Table 4.1: Summary of the experiments presented in section 4.6.1. The symbol "_xxd" states for the length of the data assimilation window in days, i.e. ssh_10d_dd refers to an experiment assimilating SSH, using a 10 days DA window and default diffusion coefficients. In the table "xx" may take the values: 2, 5, 10, 20 and 30. Two stop criterions are considered: a convergence criterion ($\epsilon = 0.5\%$), and 2 iterations.

The impact of the assimilated variable on the method performance is assessed by comparing the experiments assimilating the velocity fields and the experiments assimilating the SSH. Recalling that the SSH fields provide a vertically integrated information about

the 3-dimentional velocity fields. Therefore, this comparison may permite a discussion about the role of the model in propagating the SSH information to the 3-dimentional velocity field.

More specifically, in order to assess the BFN performance with respect to the length of the assimilation window (Sect 4.6.1) five assimilation windows were tested: 2, 5, 10, 20 and 30 days. For each configuration, two different values for the diffusion coefficients were considered: the default value and a reduced value chosen on the basis of Sect 3.6 results. The results produced with the reduced diffusion are presented in Sect 4.6.1. The effects of the number of iterations are analyzed in Sect 4.6.1 by comparing the results produced using the convergence criterion to those limiting the number of iterations to two.

In Sect 4.6.2 the last two main aspects are analysed. In Sect 4.6.2 the experiment ssh_10d_rd assimilating the SSH observations and employing a 10 days assimilation window and a reduced diffusion coefficient, is compared to an experiment employing the same configuration but with the nudging gain $\mathbf{K}$ based on the PLS regression mode. In this case small increments produced by the regression model are propagated by the non-linear model forward and backward. Afterwards in Sect 4.6.2, the impacts of the $\mathbf{K}$ structure is assessed, when observations are available every four days and subsequently, in Sect 4.6.2 it is considered that every four days a sampling similar to the Jason-1 satellite is available. In this case the results produced by the DBFN are compared to the ones produced by the 4Dvar.

When the gain $\mathbf{K}$ is diagonal and with daily observations, a linear interpolation of observations is used to make observations available at every time step. In the case where observations are available every four days, a linear weighting function that decreases to zero in two days is used to weight the gain $\mathbf{K}$ while using the last available observation. When the PLS is considered no weighting function is used.

As we only work with simulated data (extracted from a reference trajectory), the method performance is assessed by analyzing the global relative error calculated as $\frac{\|\mathbf{x}-\mathbf{x}^{true}\|}{\|\mathbf{x}^{true}\|}$. For the 3-dimensional fields the relative error for each layer is also presented. With this approach, the performance of each experiment can be analyzed with respect to the vertical structure.

4.4 Comments on the model physics

Since Sea Surface Height (SSH) is one of the most informative ocean observation and is the variable to be assimilated in our experiments, it is particularly interesting to look at the model pressure gradient formulation. As the free surface formulation is used, the pressure at a given point (x, y, Z, t) is given by $p(x, y, Z, t) = \int_0^Z g\rho(x, y, z, t)dz + \rho_0 g\eta(x, y, t)$, where η is the free surface and describes the perturbation of pressure in relation to geopotential height $Z = 0$. Thus, p is the sum of the hydrostatic pressure and the surface pressure (depth independent).

While the hydrostatic part depends directly on the density field distribution, the free surface evolves according to:

$$\begin{aligned} \frac{\partial \eta}{\partial t} &= -\nabla_h(D\bar{U}_h) + E - P \\ \bar{U}_h &= \frac{1}{D}\int_{-D}^{0} u_h dz, \end{aligned} \tag{4.9}$$

where D is the water depth, $E - P$ is the evaporation-precipitation balance, u_h is the horizontal components of the velocity field and ∇_h is the divergence operator restricted to the horizontal plan. In our experiments, the evolution of η depends only on the divergence of the vertically integrated velocities since the $E - P$ balance is set to zero. Equation (4.9) shows that resolving the vertical structure of the velocity field given observations of η is an underdetermined problem.

Analytical and numerical studies as well as direct observations have shown that most of the ocean variability corresponds to motions associated to the barotropic and the first baroclinic modes (Wunsch, 1997). This fact has been usually used to constrain DA solutions to the space spanned by these modes. This may be done by using an iterative procedure in which water masses are re-distributed on the vertical (Cooper and Haines, 1996), by ensemble methods which statistically capture these modes (Gavart and De Mey, 1997) or by explicitly restricting the minimization of cost functions similar to that given by Eq. (2.6) to the space spanned by these modes.

Therefore, one question to be addressed in this study is: if Nudging is applied to

the free surface, which corresponds to directly control the barotropic mode, are the baroclinic fields also satisfactorily corrected?

4.5 The backward integration without Nudging: Practical aspects

The backward model uses exactly the same numerical scheme as the forward model. Since most of the model is solved using centered finite differences, the backward version of the discretized model is similar to the discrete version of the backward continuous model. The only distinction between the forward and the backward model is the change in the sign of the diffusive terms when stepping backwards, this making the backward integration stable. If this is not taken into account the model blows up after a few days.

Reversing the diffusion sign in the backward model is a numerical artifact and being so its effects should be carefully analyzed. In this section, the backward integration accuracy is studied, as well as its sensitivity with respect to the choice of the diffusion coefficient. The errors are analyzed calculating the L2 error norm at the end of one forward-backward integration whithout nudging, relative to a typical one day model variation:

$$Rerror = \frac{||\mathbf{x}(0) - \tilde{\mathbf{x}}(0)||}{< ||\mathbf{x}(t + \Delta t) - \mathbf{x}(t)|| >} \tag{4.10}$$

where $\Delta t = 1$day and the brackets represent the empirical mean.

Figure 4.1 shows the *Rerror* for different window sizes. The errors grow linearly with the window size for all variables. Temperature is the most affected variable, followed by sea level and velocities. Temperature errors exceed 18 times a typical one-day variation for the 30 days experiment and 1.2 times for the 2 days. The use of reduced diffusion/viscosity coefficients reduces the errors to 6.8 and 0.16 times the one-day variation for the 30 and the 2 days experiments, respectively. Velocity errors were reduced by 50% for 30 days and 85% for 2 days, while ssh errors were reduced by 60% and 88% for 30 and 2 days, respectively.

As shown on Fig. 4.2 velocity and temperature errors are depth-dependent. Whereas for velocity they are greater at the surface and decrease with depth, for temperature

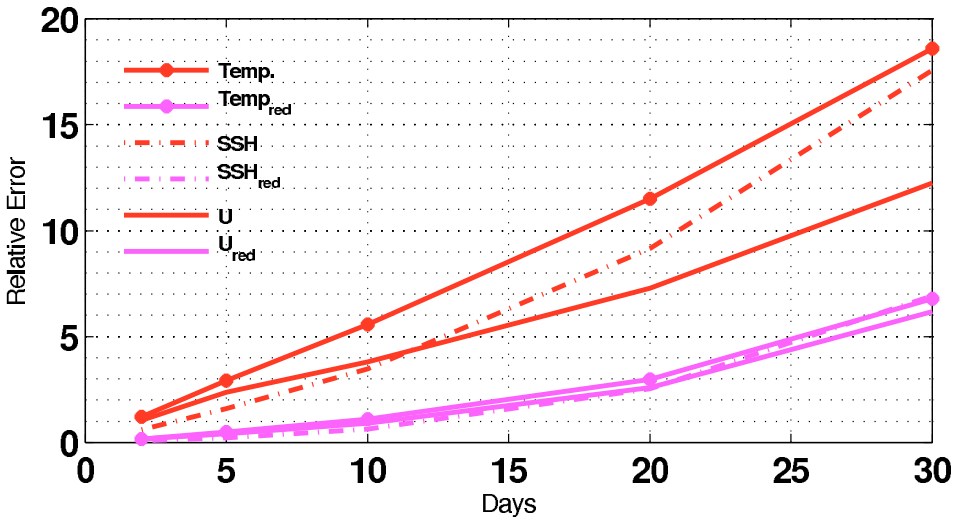


Figure 4.1: Errors of the initial condition after one forward-backward model integration perfectly initialized and without nudging. Red curves were obtained using the same diffusion coefficients used in the reference experiment ($\nu_h^{u,v} = -8 \times 10^{10} m^4/s$ and $\nu_h^{t,s} = -4 \times 10^{11} m^4/s$) and magenta curves were obtained using reduced diffusion ($\nu_h^{u,v} = -8 \times 10^{9} m^4/s$ and $\nu_h^{t,s} = -8 \times 10^{10} m^4/s$). The abscissa represents the size of the time window.

they are important in the thermocline. In the cases for which the forward-backward integrations use the same diffusion/viscosity coefficients as in the reference simulation, the temperature errors at thermocline depths exceed 3 times the typical one day variation for the 5 days experiments and reaches 15 times for the 20 days experiments. Considering the velocities, errors are proportional to 4 one-day variations for the 5 days experiment and to 8 one-day variations for the 20 days experiments. For time windows of 10, 20 and 30 days, velocities at the thermocline depths start to be influenced by the temperature errors.

Furthermore, reduction of the diffusion/viscosity coefficients greatly reduced the errors especially in the thermocline for the temperature and at the surface for the velocity. It can be noted that when the diffusion coefficient is decreased the errors converge to a limit. This limit changes with respect to the window length and should be related to the diffusion required to stabilize the numerical method, which is of second order in our case, and hence oscillatory. Therefore, there is a compromise between the errors induced by the extra diffusion and errors due to spurious oscillations.

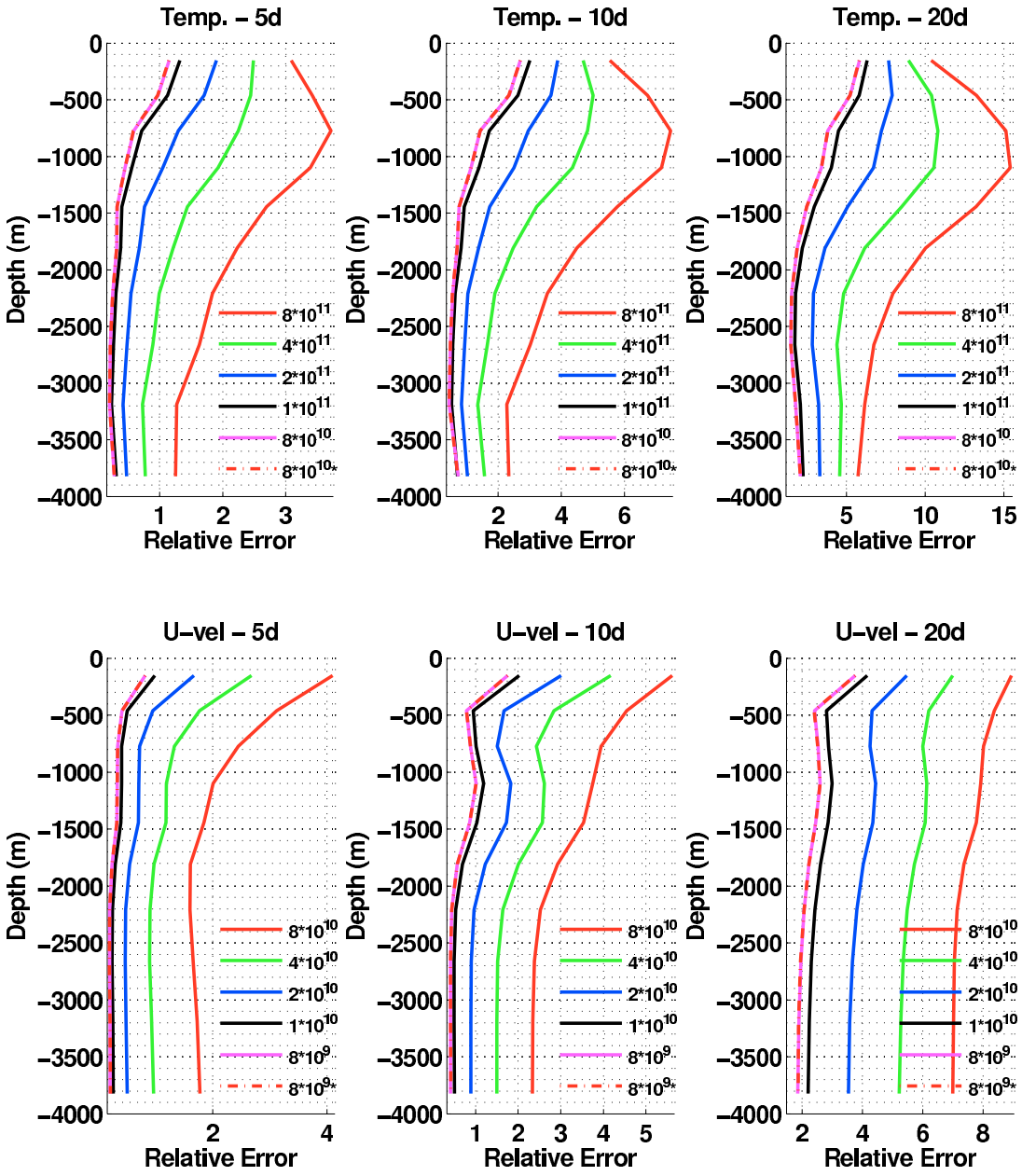


Figure 4.2: Vertical errors of the initial condition after one forward-backward model integration without nudging. Each color refers to an experiment performed using the diffusion coefficient indicated in the figure legend. The experiment represented by the dashed red curve used the same configuration as the experiment represented by the magenta curve but with a time step of 90s instead of 900s. Top panel: temperature errors; Bottom panel: zonal velocity errors. The time window is indicated in the title of each figure.

Numerical errors were assessed by changing the model time step from $900s$ to $90s$. The resulting errors do not change, suggesting that the errors induced by the diffusion

are dominant. On the one hand, this is important because the complete rewriting of the NEMO's code can be difficult, similarly to the adjoint model programming used by the 4Dvar, but on the other hand if the assimilation cannot control the diffusion errors it may represent a fundamental problem of the method when it is applied to non-reversible geophysical systems such as the ocean.

Figure 4.3 shows the spatial structures of the sea level error for the 10 days experiment. The errors are highly variable in space, being greater along the main jet axis. This is probably due to the fact that the backward integration smooths the gradients and so the greatest errors are found near the fronts. Therefore, the error structures may be of high variability in space and time since they are state dependent.

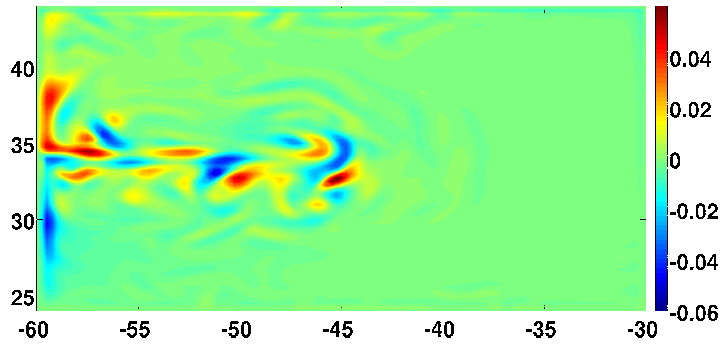


Figure 4.3: Sea level errors after one forward-backward model integration. The time window is 10 days.

Figure 4.4 shows the spectral SSH error per wave number calculated from the experiment employing the reference diffusion coefficient, and Fig. 4.5 the surface kinetic energy spectrum calculated from the experiment employing the reference diffusion coefficient and a reduced diffusion coefficient. The backward integration introduces an extra diffusion, coarsening the effective model resolution, which is defined as the portion of the spectra for which there is a changement in the spectrum slope. In the reference simulation the effective model resolution is estimated to be 190km, which is coherent with the $\approx 7 * \Delta x$ estimation of Skamarock (2004).

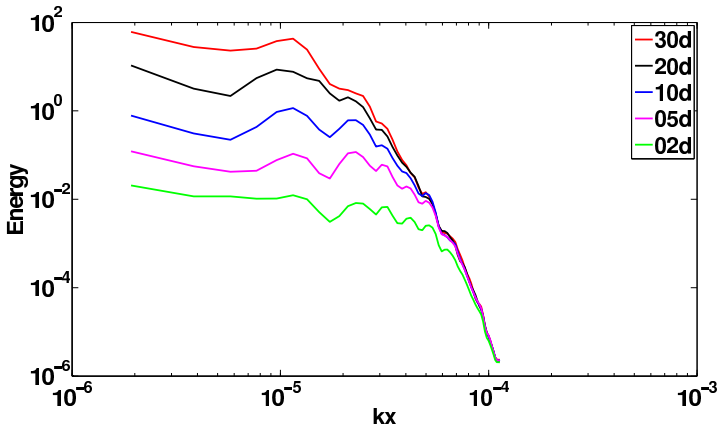


Figure 4.4: Mean power spectra energy calculated with the SSH error of one forward-backward iteration. Each color represents a DA window: 2, 5, 10, 20 and 30 days. kx stands for longitudinal wave-number.

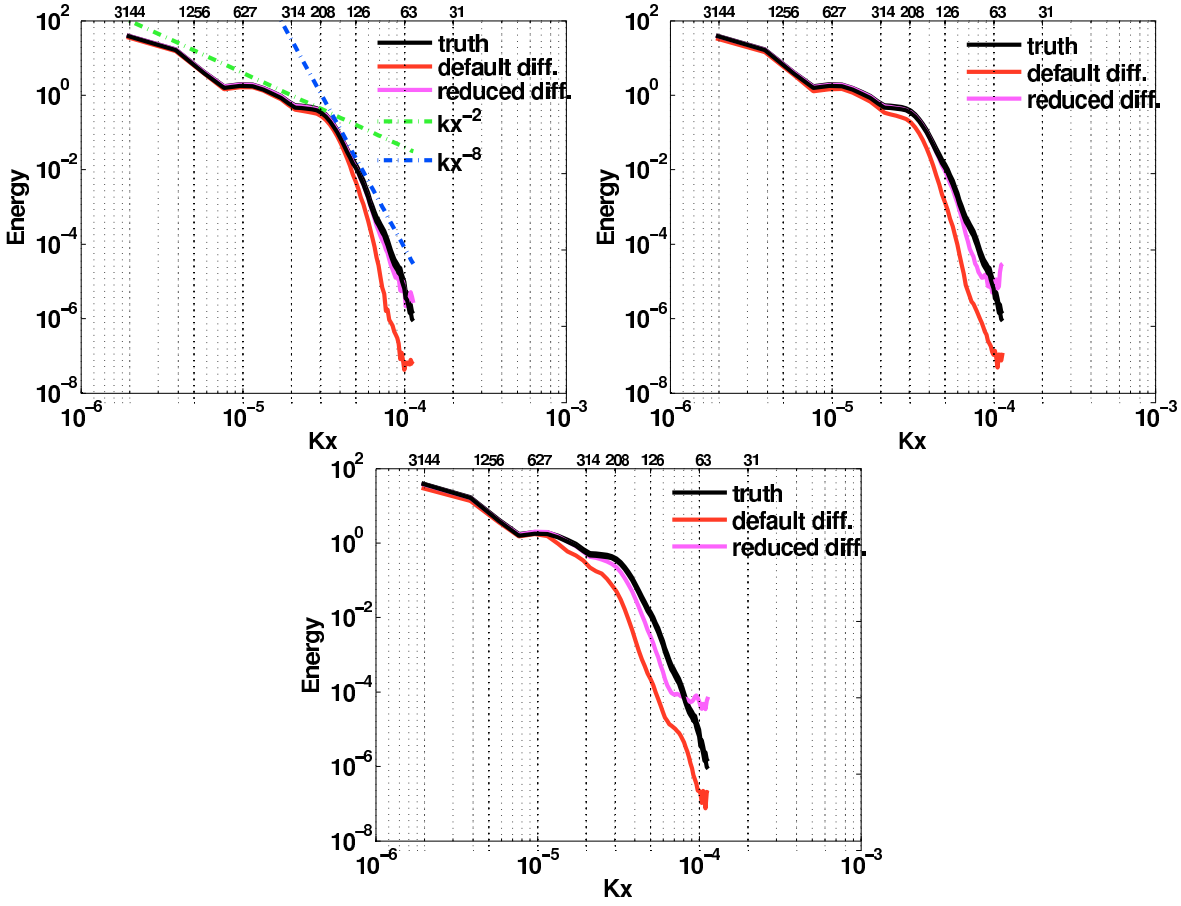


Figure 4.5: Kinetic energy mean power spectra calculated using the first layer velocity fields. Black curves represent the "true" initial condition power spectra; Red curves represent the power spectra calculated after one forward-backward iteration without the nudging term and employing the reference diffusion coefficient; Magenta curves represent the power spectra calculated after one forward-backward iteration without the nudging term and employing a reduced diffusion coefficient. Top left: 5 days assimilation window. Top right: 10 days assimilation window. Bottom: 20 days assimilation window. In the bottom abscissa the ticklabels stand for longitudinal wave-number (rad/m) while in the top abscissa the ticklabels stand for the corresponding wavelengths in km units.

The longer the time window the greater the portion of the spectra affected. For the experiment employing the reference diffusion coefficient, the divergence between the true spectra and the spectra obtained from the backward integration is observed at 126, 314 and 627km for 5, 10 and 20 days experiments, while for the experiments considering a reduced diffusion coefficient there is almost no differences for the 5 days experiment, and the divergence is observed at 126 and 314km for the 10 and 20 days experiments. If on the one hand using the reduced diffusion helps to keep the energy distribution coherent with the true distribution, on the other hand it creates noise in the range of 126km to 25km. This confirms that there is a trade-off between the errors due to the excessive smoothing and the errors due to high frequency numerical modes.

The spectral differences may be due to two things: first due to the modified energy flux between the different scales, which is verified by a changement of the spectrum slope especially at high wave number, and second due to the smoothing of the large scale gradients.

In this section we have seen that there are important backward-errors induced by over-diffusion. Therefore, short time windows with reduced diffusion coefficients would be preferable to be used in DA experiments. Two regions have to be cautiously analyzed: the surface and the thermocline. Surface layers are prone to feature errors due to their role on the wind energy dissipation while at the thermocline strong density gradients contribute to high diffusion rates.

It has been observed that diapycnal mixing (i.e. mixing between two isopycnal surfaces) are several order of magnitude smaller than mixing within an isopycnal layer. NEMO is a Z coordinate model and the diffusion operators are split into vertical and horizontal components. This means that the problem of unrealistic diapycnal mixing, due to horizontal Z layers intersecting isopycnal plans, is even more pronounced due to the iterations. Therefore, the use of diffusion operators acting along isopycnals can be of great interest along with using the DBFN.

In the next section data assimilation ability to control the errors induced by diffusion is assessed as well as the consequences of these errors on the assimilation system configuration.

4.6 Data Assimilation Results

4.6.1 Experiments with scalar nudging coefficients

Sensitivity to the DA window length

Following the results presented in the previous sections, the results concerning the sensitivity test of the DBFN to the choice of the DA window are presented, but this time including the nudging term. Thus, the considered primary aspect is whether the use of a diagonal nudging gain is sufficient to control the backward model in the presence of anti-diffusion. In this case, control was achieved using an unrealistic observation network combined with strong nudging and for short time periods ($\leq$ 20 days). Indeed, the noise induced by anti-diffusion can potentially damage the mass field since it induces large rates of spurious dyapicnal mixing.

Figures 4.6 and 4.7 show the zonal velocity and the temperature errors for the DBFN experiments assimilating the SSH and velocities, respectively. The experiments were integrated for 160 days, and for each data set five different DA windows (2, 5, 10, 20 and 30 days) were considered. All experiments used the same nudging gain (but this one changes with respect to which variable is being assimilated: SSH or velocities) and the same convergence criterion.

For all experiments the DBFN reduces the initial error by more than 60% for the dynamical variables and by more than 30% for the temperature. As expected, the errors calculated from the experiments assimilating velocities are always smaller than those from the experiments assimilating SSH, which means that the model has a limited ability to project the SSH corrections onto the velocity fields. Long assimilation windows produce better results for the beginning of the experiments (first 30 days), but using a short assimilation window is preferable at long term regardless the assimilated variable. The latter may be verified in table 4.2, which presents the mean initial and final condition errors for the experiments assimilating SSH observations. Moreover, for all cases the temperature error is not asymptotically stable, and starts to grow after 100 days. The experiment divergence is especially important for long assimilation windows and proves the difficulty of controlling the diffusion-induced errors by only assimilating SSH observations.

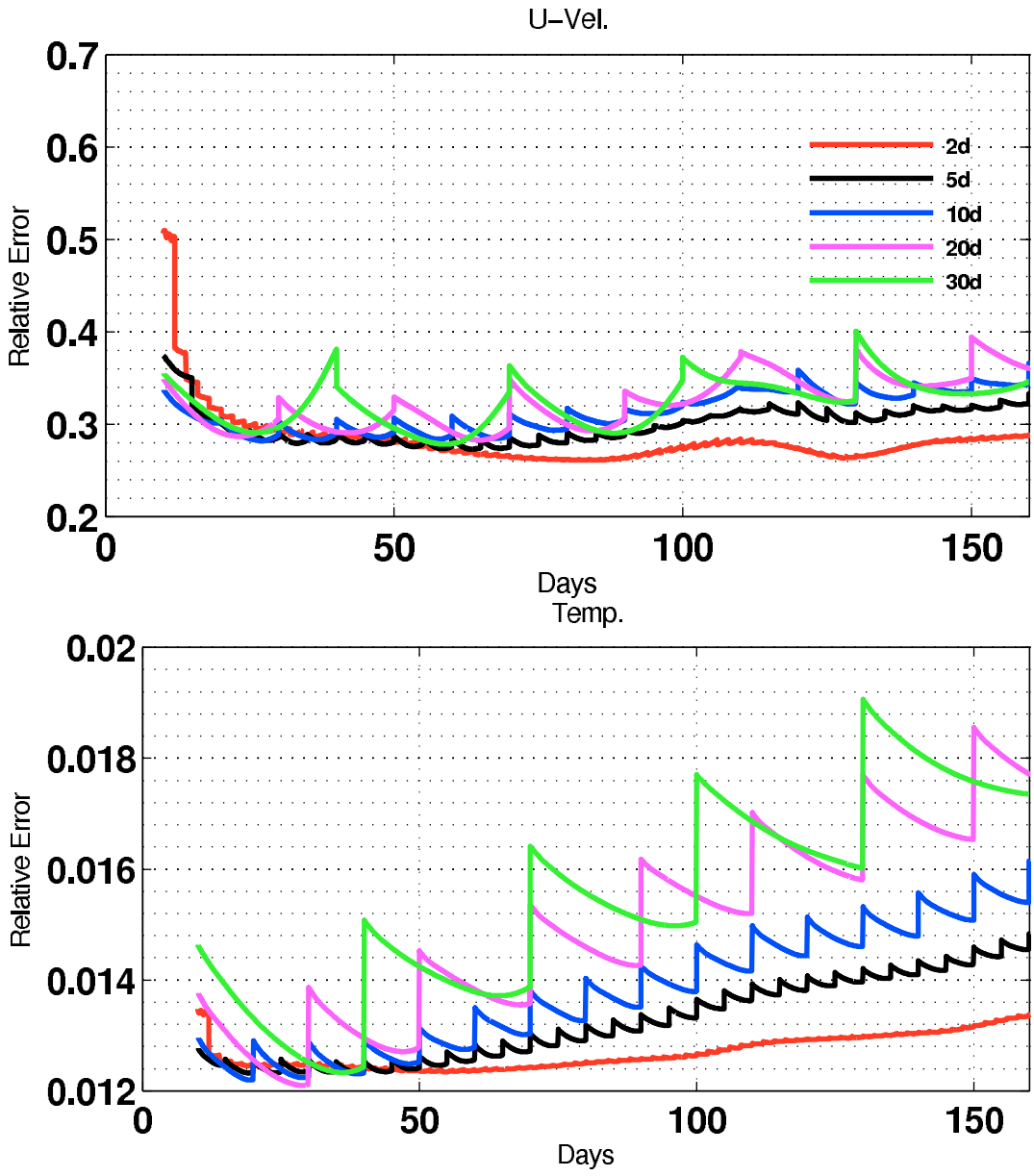


Figure 4.6: Relative errors of the zonal velocity (top) and the temperature (bottom) for the experiments listed in table 4.1 which assimilates SSH using the convergence criterion and the reference diffusion coefficient.

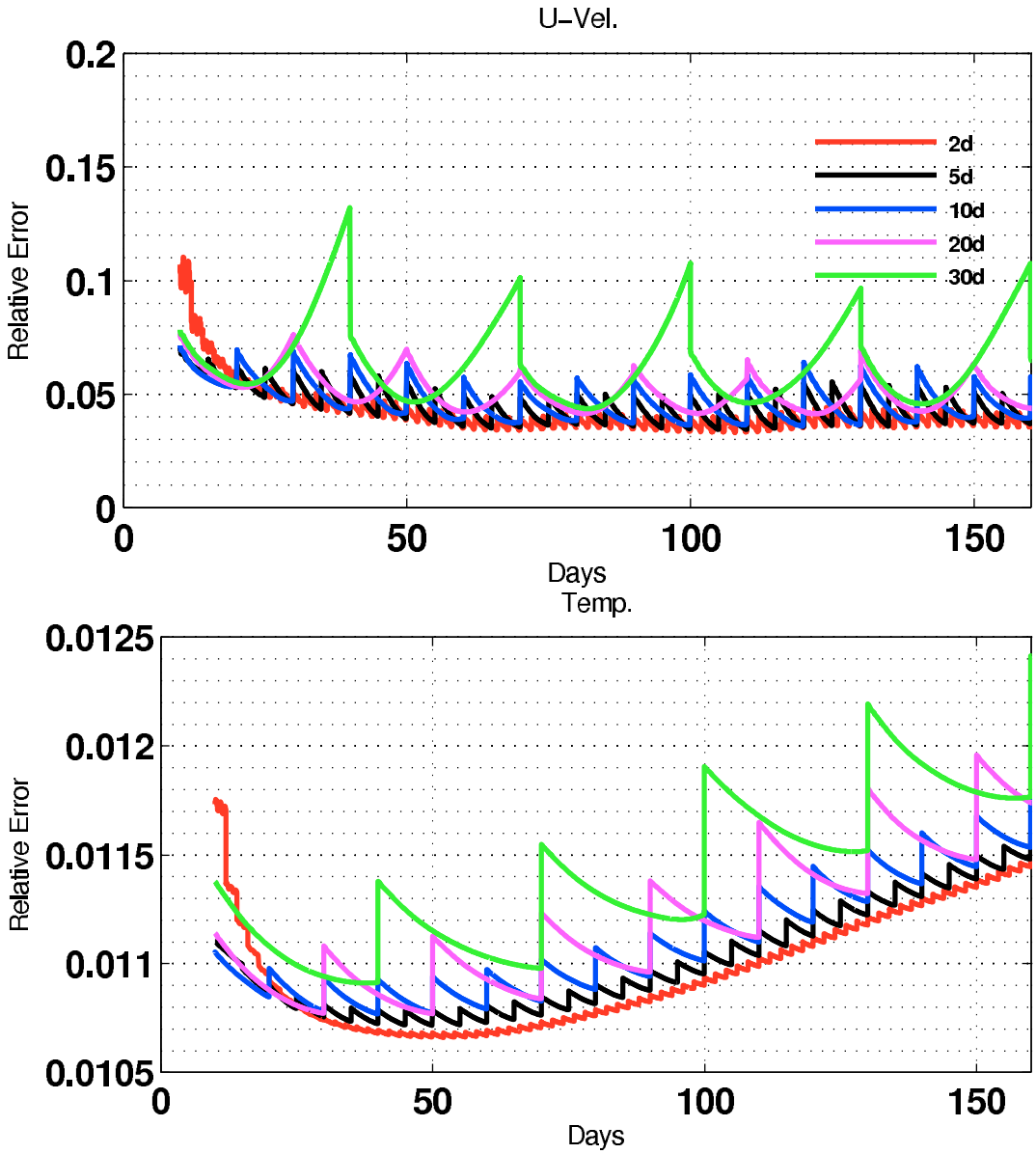


Figure 4.7: Relative errors of the zonal velocity (top) and the temperature (bottom) for the experiments listed in table 4.1 which assimilates velocity using the convergence criterion and the reference diffusion coefficient.

AW	SSH		U		V		T	
	$\mathbf{e}^0$	$\mathbf{e}^f$	$\mathbf{e}^0$	$\mathbf{e}^f$	$\mathbf{e}^0$	$\mathbf{e}^f$	$\mathbf{e}^0$	$\mathbf{e}^f$
2d	0.0341	0.0613	0.2872	0.2891	0.3974	0.3815	0.0129	0.0129
5d	0.0369	0.0574	0.3138	0.3085	0.4503	0.4014	0.0137	0.0136
10d	0.0493	0.0649	0.3368	0.3279	0.5266	0.4175	0.0147	0.0142
20d	0.0760	0.0940	0.3558	0.3298	0.6649	0.4995	0.0159	0.0147
30d	0.1025	0.1312	0.3666	0.3473	0.7472	0.5840	0.0166	0.0149

Table 4.2: Summary of the mean relative initial and final condition errors obtained from the DBFN experiments employing the reference diffusion and assimilating daily SSH observations. AW is the Assimilation window. $\mathbf{e}^0$ and $\mathbf{e}^f$ are the mean initial and final errors, respectively.

As observed by Reynolds and Palmer (1998) examining the accuracy of a backward tangent linear quasi-geostrophic model, when the initial backward condition contains uncertainties, accuracy is lost and the model may produce useless results. This is related to the backward model spectrum whithout nudging which may have higher error growth rates than the forward model. Therefore if a variable is not controllable, the iterations may produce a divergent estimation since the errors may grow indefinitely.

In real applications, the SSH is assimilated together with vertical temperature profiles and/or sea surface temperature. This is done in the next chapter using the twin experiment framework. Including the vertical profiles is enough to control the temperature drift observed in this section. Another possibility which is successfully applied in Sect.4.6.2 is to estimate the temperature from the SSH using regression models.

Sensitivity to the diffusion coefficient

In the last section it was shown that even with nudging the backward error is quite important. To study the DBFN sensitivity to the model diffusion coefficient, the experiments from Sect 4.6.1 are repeated but use the diffusion coefficients that gave the best results showed in Sect 3.6.

By comparing tables 4.2 and 4.3, it is evident that the use of a reduced diffusion improves the state estimation. This is valid for all variables and window sizes independently of the assimilated variable. For instance, Figs. 4.6 and 4.8 reveal that after 160 days of the experiments assimilating only SSH and considering an assimilation window of 10 days, the velocity error is almost 45% smaller when considering a reduced diffusion.

AW	SSH		U		V		T	
	$\mathbf{e}^0$	$\mathbf{e}^f$	$\mathbf{e}^0$	$\mathbf{e}^f$	$\mathbf{e}^0$	$\mathbf{e}^f$	$\mathbf{e}^0$	$\mathbf{e}^f$
2d	0.0505	0.0926	0.3142	0.3269	0.4272	0.4488	0.0118	0.0119
	0.0299	0.0595	0.2495	0.2590	0.3577	0.3755	0.0117	0.0117
5d	0.0412	0.0911	0.2958	0.3219	0.3993	0.4432	0.0116	0.0118
	0.0166	0.0373	0.2153	0.2194	0.2877	0.2927	0.0119	0.0119
10d	0.0401	0.0944	0.2854	0.3204	0.3904	0.4454	0.0116	0.0117
	0.0278	0.0412	0.2321	0.2243	0.3233	0.2801	0.0123	0.0122
20d	0.0608	0.1132	0.3130	0.3372	0.4644	0.4776	0.0119	0.0118
	0.0486	0.0685	0.2694	0.2517	0.4768	0.3724	0.0131	0.0126
30d	0.0815	0.1266	0.3403	0.3489	0.5643	0.5254	0.0124	0.0119
	0.0668	0.1078	0.2911	0.3035	0.5907	0.5153	0.0138	0.0130

Table 4.3: Summary of the mean relative initial and final condition errors obtained from the DBFN experiments employing a reduced diffusion and assimilating daily SSH observations. AW is the Assimilation window. For each AW the top lines represent the experiments considering only 2 iterations and the bottom line the experiments considering $\epsilon = 0.0005$. $\mathbf{e}^0$ and $\mathbf{e}^f$ are the mean initial and final errors, respectively.

Again, it seems to be preferable to consider short assimilation windows. Regarding the experiments assimilating SSH, the velocity errors are stable and smaller for 5 and 10 days windows. For the temperature, the shorter the assimilation window, the smaller the errors. They also initially decrease but start to grow after 100 days, excepted for the 2 days window. Considering the experiments assimilating velocity, the estimations from all experiments and for all variables are asymptotically stable. The best dynamical variables estimation is for the 10 days experiments while for the temperature is for the 2 days one.

Controlling the velocity is quite an effective way to control the tracer fields. This may happen thanks to the simplified thermodynamical forcing considered. In the present study, the equation for the conservation of temperature is reduced to an advection-diffusion equation. This way, controlling the velocity means controlling the advective term of the equation, which is the dominant term for short time periods. However, in real applications direct velocity observations are rare and usually restricted to coastal areas (e.g. HF ocean radar).

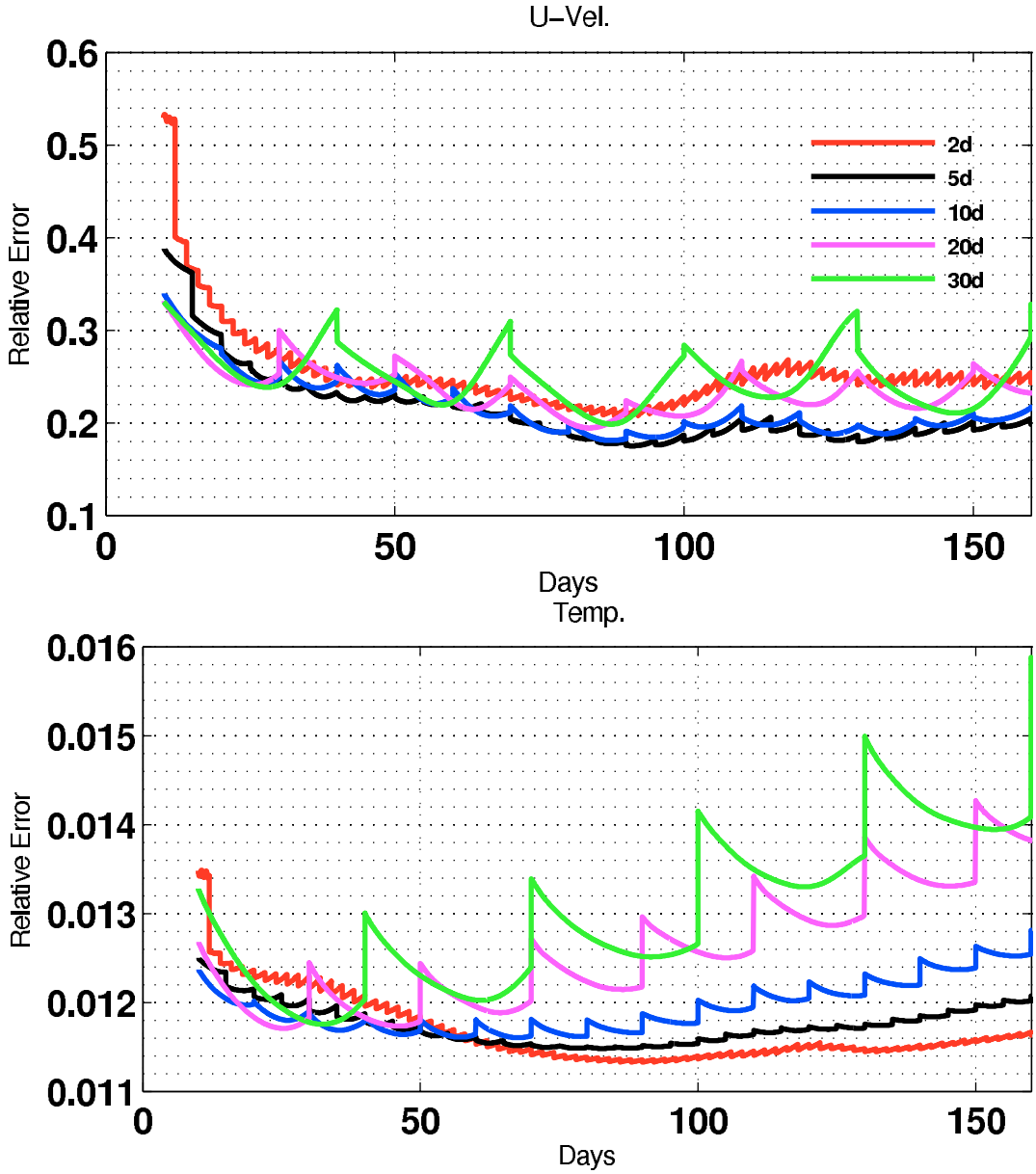


Figure 4.8: Relative errors of the zonal velocity (top) and the temperature (bottom) for the experiments listed in table 4.1 which assimilates SSH using the convergence criterion and the reduced diffusion coefficient.

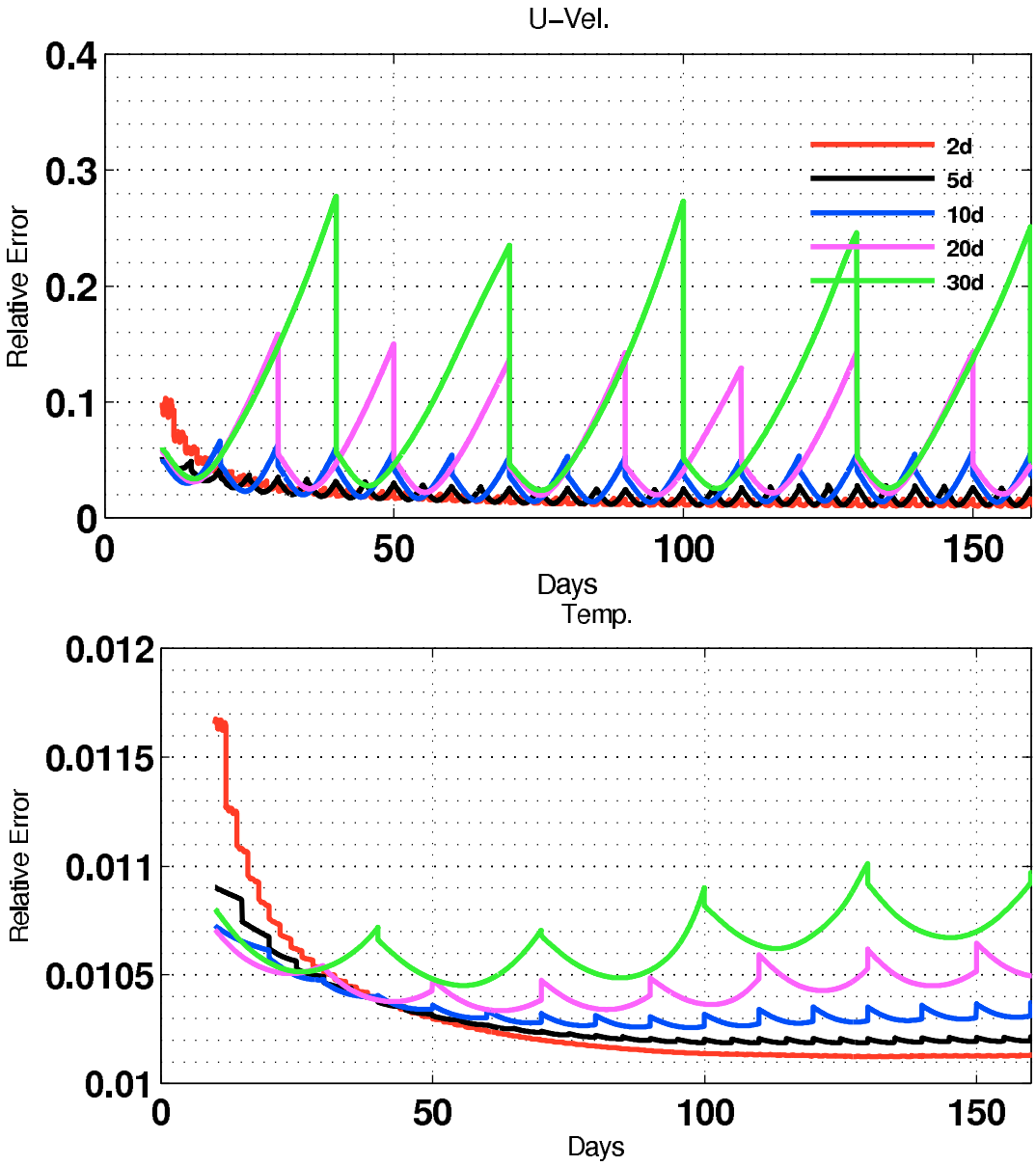


Figure 4.9: Relative errors of the zonal velocity (top) and the temperature (bottom) for the experiments listed in table 4.1 which assimilates velocity using the convergence criterion and the reduced diffusion coefficient.

Curiously, the 30 days experiment assimilating velocity with reduced diffusion produces more accurate initial conditions than the experiment employing the control diffusion but lower quality forecasts. Based on this, the table 4.4 compares the forecast error growth rate for the experiments assimilating the SSH observations with the reference and the reduced diffusion coefficients. For almost all experiments the growth rate is

smaller when using the reference diffusion coefficient.

AW	**SSH**	**U**	**V**	**T**
2d	1.3580	0.0926	-0.7921	0.0001
	1.4812	0.4711	0.8863	0.0023
5d	0.4115	-0.1055	-0.9784	-0.0036
	0.4152	0.0823	0.0985	-0.0001
10d	0.1554	-0.0884	-1.0919	-0.0050
	0.1336	-0.0782	-0.4319	-0.0014
20d	0.0901	-0.1302	-0.8269	-0.0058
	0.0998	-0.0885	-0.5218	-0.0026
30d	0.0954	-0.0644	-0.5439	-0.0054
	0.1367	0.0415	-0.2515	-0.0027

Table 4.4: Summary of the mean error growth rate obtained from the DBFN experiments assimilating daily SSH observations. AW is the Assimilation window. For each AW the top lines represent the experiments considering the reference diffusion coefficient and the bottom lines the experiments considering a reduced diffusion coefficient.

This finding supports the discussion of the last section about the backward model accuracy. If on the one hand, a reduced diffusion improves the backward model accuracy, on the other hand the error growth rate, both in forward and backward, may increase, which leads to large forecast errors. This happens because with the reduced diffusion the model spectrum must have more singular values greater than one than the model would have with the reference diffusion.

A simple solution, but not tested here, would be to use distinct diffusion coefficients in the forward and backward integrations. Furthermore, the convex character of the error curves is pronounced for long assimilation windows and reduced diffusion experiments, revealing the importance of these unstable directions for both the forward and the backward integrations. This pattern was also observed by examining the error evolution as iterations are performed (not shown).

Effects of the number of iterations and the control of the dynamics

Two new sets of experiments have been created using a reduced diffusion and assimilating the SSH. For the first, the number of iterations is limited to two while for the second the convergence criterion is more restrictive ($\epsilon = 0.001$) and iterations are limited to 50. Starting from the latter, Fig.4.10 shows that if only one assimilation cycle

is considered, more iterations results in more accurate initial condition with respect to the dynamical variables. For the temperature, the estimation starts to diverge after 8, 10 and 20 iterations for the 30, 20 and 10 days assimilation window respectively. In addition, the 10 days window has the smaller velocity error after convergence. However, as already noted in section 5.1.1 the system diverges when several assimilation cycles are considered, regardless the assimilation window used.

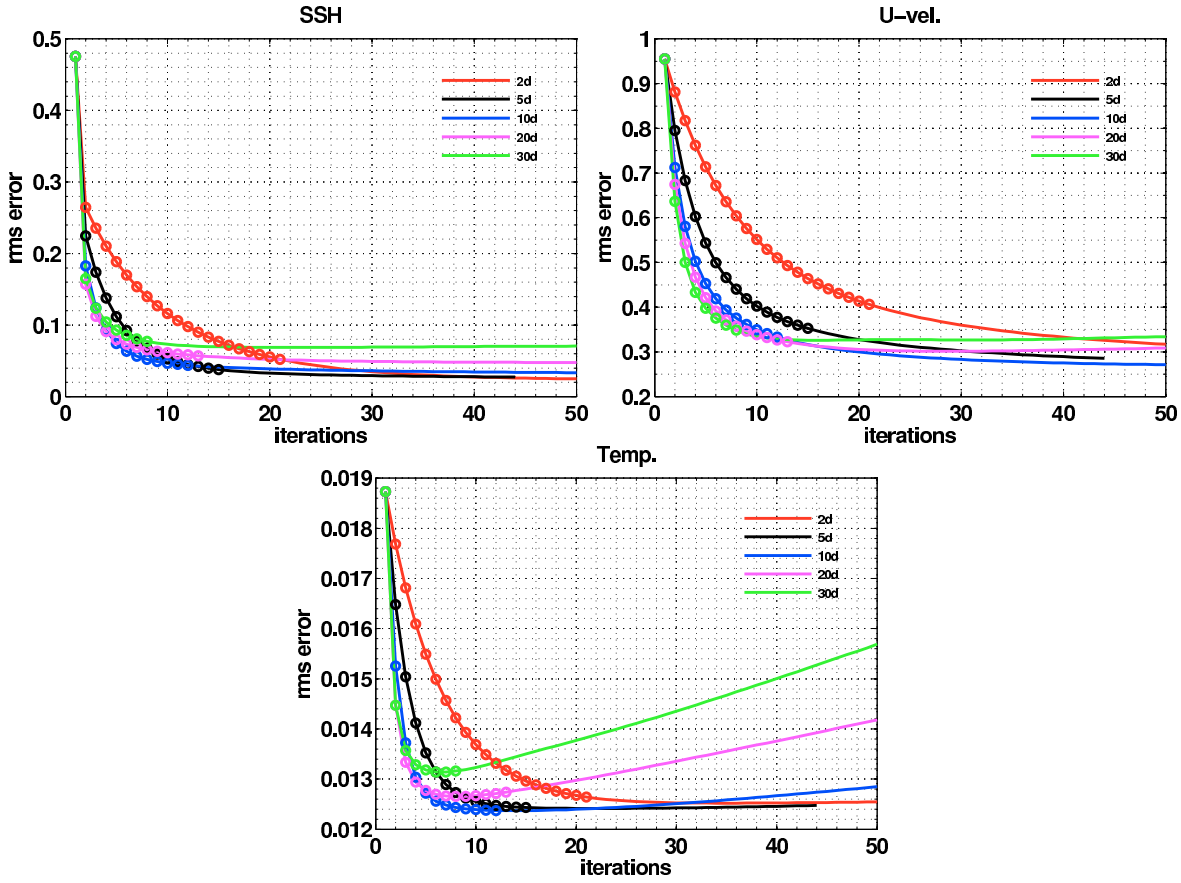


Figure 4.10: Variation of the initial condition relative errors with respect to the iterations for the experiment assimilating daily gridded SSH fields. The circles represent the results obtained using the standard convergence criterion, $\epsilon = 0.005$, and the continous lines obtained with a more restrictive criterion $\epsilon = 0.001$. Top left: SSH error; Top right: zonal velocity error; Bottom: temperature error.

Considering the convergence criterion $\epsilon = 0.005$, the DBFN converges quite fast (2-3 iterations), for all experiments and independently of which variable is being assimilated. Significant differences are observed for the first three assimilation cycles for which assimilating the SSH required much more iterations (7-20). As shown in Fig. 4.10, the number

of iterations for the first assimilation cycle depends on the length of the assimilation window. A long assimilation window leads to a faster convergence in terms of iterations. Nonetheless considering the computational cost, the short assimilation window would be preferable since the gain from reducing the number of iterations is not large enough to offset the computational cost of using longer assimilation windows. However, when convergence is reached, the errors are different and vary with the variables, e.g. for the zonal velocity the 20 days window provides the best result while for the temperature the 10 days performed better.

In the case of two iterations (Fig. 4.11), no tendency is observed for the temperature errors for windows shorter than 20 days. However, velocity errors are greater than the errors obtained using the convergence criterion. When analyzing only the first assimilation window, the best results were obtained with the 30 days experiment, although it is the 10 days window followed by the 5 and 2 days one that produced the best mean initial conditions, see table 4.3. The best results observed for the longer assimilation window is a consequence of the asymptotic character of the Nudging method. Therefore, the longer the assimilation window, the smaller the error. The same is not observed when several assimilation cycles are considered due to the diffusive aspect of the DBFN. Equation (2.38) shows that in the absence of observations the solution at convergence will be a homogenous field. This explains the observed problems in estimating temperature especially for long assimilation windows and restrictive convergence criterions.

These results suggest that there is a trade-off between considering long assimilation windows that permit the use of the non-linearities and the correction of the errors projecting onto the stable subspace, and the errors due to diffusion. The presented experiments clearly indicate that the 10 days window fulfills this criterion. Therefore, in the following sections only the 10 days window is considered.

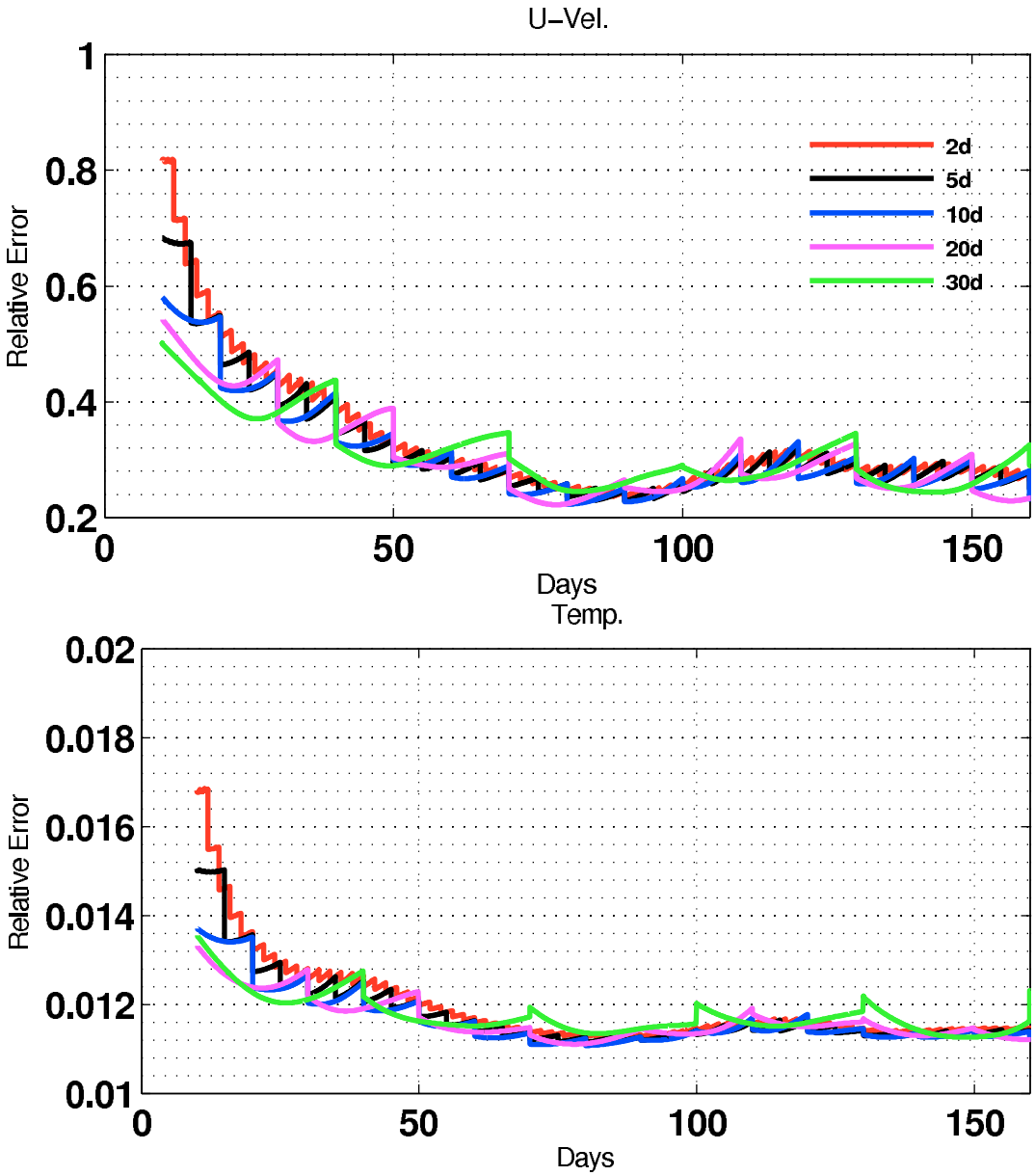


Figure 4.11: Relative errors of the zonal velocity (top) and the temperature (bottom) for the experiments listed in table 4.1 which assimilates SSH using only two iterations. All experiments have used reduced diffusion coefficients.

The vertical error structure is shown in Figs. 4.12 and 4.13 for the first and last assimilation cycle. The errors are reduced especially at the upper layers including the thermocline. This suggests that the model reproduced the appropriate dynamics since the thermocline dynamics is mainly controlled by the first baroclinic mode and recalling that assimilating the SSH is a direct control of the barotropic mode. The increase of temperature errors after 100 days is observed at depths where stratification is stronger, confirming the influence of diffusion in these errors structures. For the velocity, there is an increase of the error at depths greater than $1500m$. This may happen because at this depth velocities may be preferably baroclinic and our nudging scheme corrects the barotropic mode, and/or because the excessive smoothing of the horizontal density gradients alters the velocity fields by changing the thermal wind balance and therefore the baroclinic velocities.

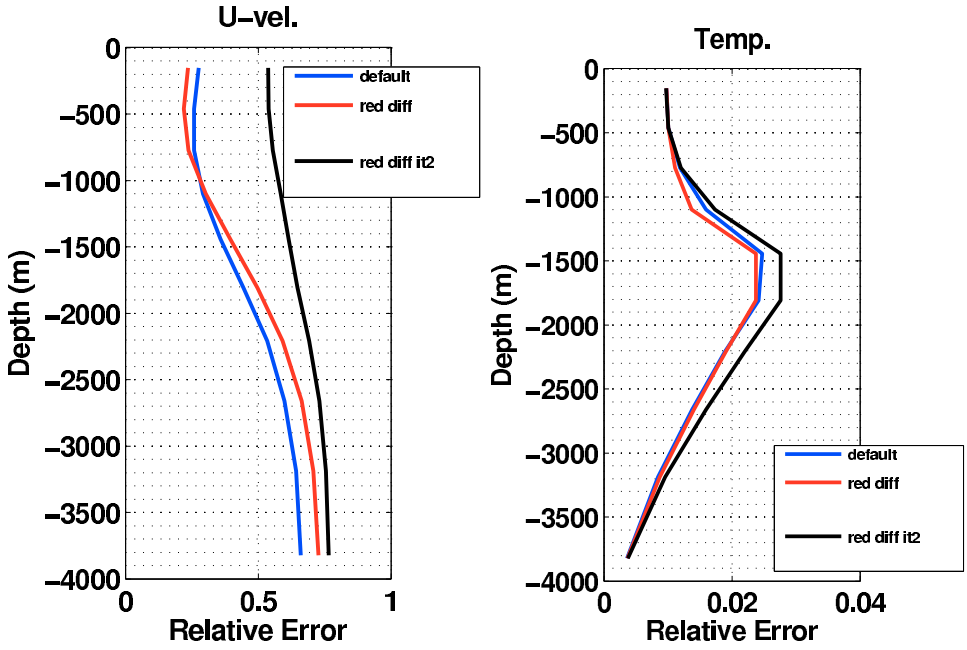


Figure 4.12: Vertical relative error of the zonal velocity (left) and the temperature (right) for the experiments ssh_10d_dd (default), ssh_10d_rd (red diff) and ssh_10d_rd_2it (red diff it2). The data refers to the identified initial conditions of the first assimilation cycle.

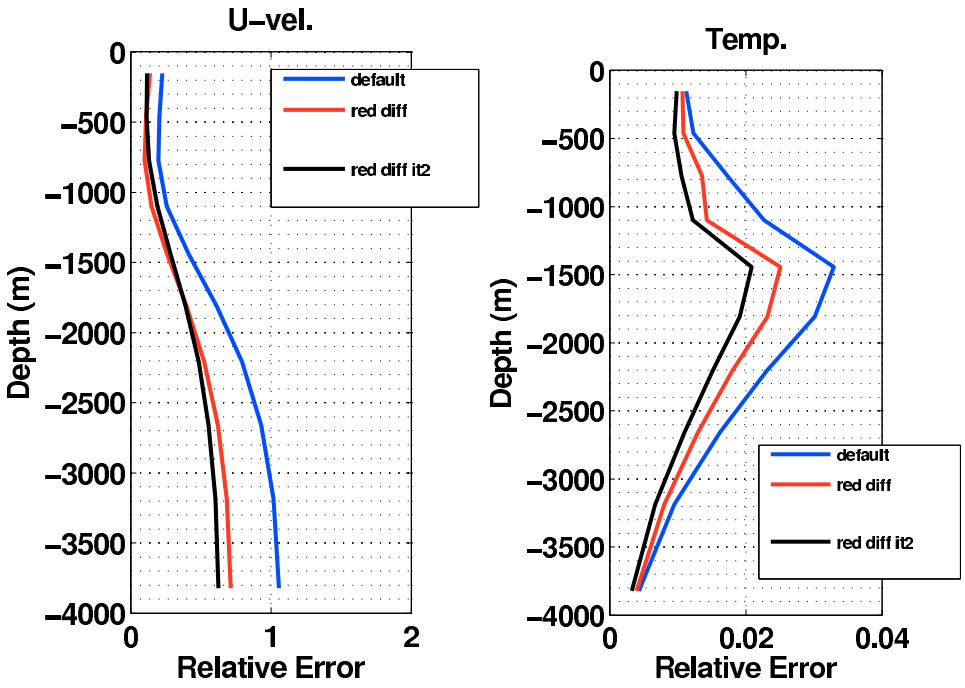


Figure 4.13: Vertical relative error of the zonal velocity (left) and the temperature (right) for the experiments ssh_10d_dd (default), ssh_10d_rd (red diff) and ssh_10d_rd_2it (red diff it2). The data refers to the identified initial conditions of the last assimilation cycle.

Conclusions

In this section we have seen that a relatively short assimilation window ($\leq$ 10 days) along with a reduced diffusion coefficient has to be preferably used with the DBFN, confirming the results presented in Sect 3.6. Concerning the velocities, Figs. 4.2 and 4.13 show that nudging the SSH reduces the errors exactly where diffusion errors are greater, i.e. at the upper ocean. Controlling the deep ocean by assimilating only SSH is quite a difficult task. At this point we cannot make conclusions about the nature of the remaining errors, i.e. if they are caused by diffusion or if the SSH is not a good predictor for the deep ocean. This discussion continues in the Sect 4.6.2 where the results produced by the regression models are discussed.

It can also be concluded that in the absence of a dense observational network, (e.g. sampling the ocean 3-dimensional structure), more complex gains, $\mathbf{K}$, which correct the non-observed variables are needed. In order to consider this aspect, the use of gains based on regression models are analyzed in the following Sect 4.6.2.

4.6.2 The Hybrid DBFN

In this section the importance of the nudging gain structure under different observation network is analyzed. This is done by comparing experiments using a diagonal **K** and a **K** constructed using the PLS regression model. The experiments are summarized in Table 4.5.

	stop criterion	Assim. Variables	Regressed Var.
ssh_10d_rd_uv	$\epsilon = 0.5\%$	daily full SSH field	UV
ssh_10d_rd_it2_uv	2it	daily full SSH field	UV
ssh4_10d_rd_uv	$\epsilon = 0.5\%$	every 4 days full SSH field	UV
ssh4_10d_rd	$\epsilon = 0.5\%$	every 4 days full SSH field	-
ssh4_10d_rd_it2_uv	2it	every 4 days full SSH field	UV
ssh4_10d_rd_it2	2it	every 4 days full SSH field	-
DBFN+PLS	$\epsilon = 0.5\%$	every 4 days Jason1-like SSH	UVT
ONDG	direct	every 4 days Jason1-like SSH	UVT

Table 4.5: Summary of the experiments presented in section 4.6.2. Two stop criterions are considered: a convergence criterion ($\epsilon = 0.5\%$), and 2 iterations.

Daily gridded SSH observations

In this section the experiments ssh_10d_rd and ssh_10d_rd_it2, that assimilates daily SSH fields with reduced diffusion coefficients and the experiments ssh_10d_rd_uv and ssh_10d_rd_it2_uv are compared. Their configuration are similar but with **K** constructed using the PLS regression model to correct the velocity with SSH increments. This choice is based on the results of the previous sections that shows that good estimates of the velocities imply good estimates of the temperature field, and on the fact that the SSH is better correlated with the velocity field than with the temperature.

The use of the regression model improves the estimation of all model variables when only two iterations were considered (Fig. 4.14). When the convergence criterion was used, the velocity errors are smaller for the first 4 DA cycles, after which the pure DBFN errors become slightly smaller than those of the experiment using the regression. The temperature estimation is also improved. This may be related to the improvement of the upper ocean velocities description.

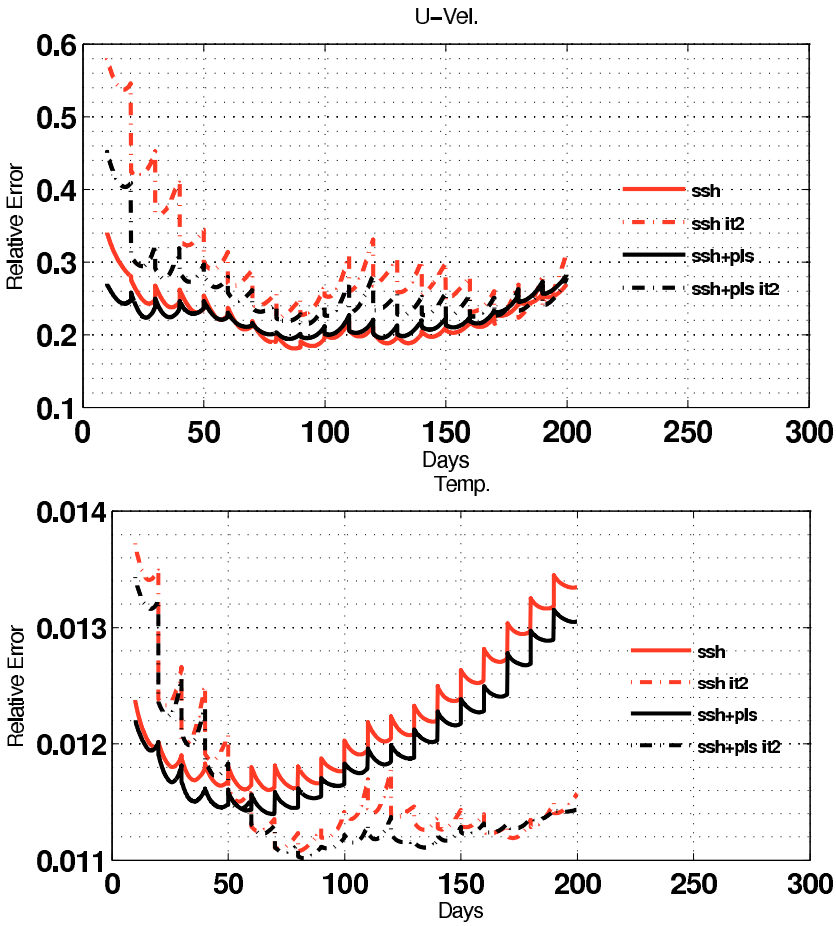


Figure 4.14: Relative errors of the zonal velocity (top) and the temperature (bottom) for the experiments ssh_10d_rd (red curve) and ssh_10d_rd_it2 (dashed red curve) and their equivalents but with K constructed using PLS regression (black curves).

The vertical structure of the velocity error for the day 1 and for the day 130 is analyzed (Fig. 4.15) by decomposing the vertical error into empirical orthogonal functions (EOF). This decomposition shows what kind of errors remains after the assimilation step. For all experiments the first mode represents an error which has the same sign over all depths. More specifically, the velocities over most of the domain are underestimated. The first mode for the day 1 shows that the error shape are changed by the iterations. After two iterations the surface velocities have higher errors than the deep ocean velocities. As iterations are performed, the surface errors decrease faster than the deep errors. This error mode accounts for 93.14%(BFN) and 97.20%(BFN+PLS) of the variability for the experiments using the convergence criterion and 88.17%(BFN)

and 79.94%(BFN+PLS) for the experiments employing only two iterations. For the day 130, the first mode of all experiments has almost the same shape as the first mode for the day 1 at convergence. It accounts for 98.56%(BFN) and 99.21%(BFN+PLS) of the variability for the experiments using the convergence criterion and 89.06%(BFN) and 93.67%(BFN+PLS) for the experiments employing only two iterations.

The results suggest that iterations help the model to correct the baroclinic field, since the residual error has a barotropic structure. In addition, the similarity of the EOF modes indicates that both methods correct the model in the same way. The PLS advantage is its efficiency in improving the upper ocean estimation with respect to the DBFN, which is especially important when less iterations are considered. However, the PLS is not so efficient as the DBFN in reducing the deep ocean errors. This proves the DBFN skill in estimating the dynamical variables when gridded observations of SSH are available. Nevertheless, the improvement of the deep ocean velocities would require direct observations of this region, since the SSH does not appears to be a good controller of this region.

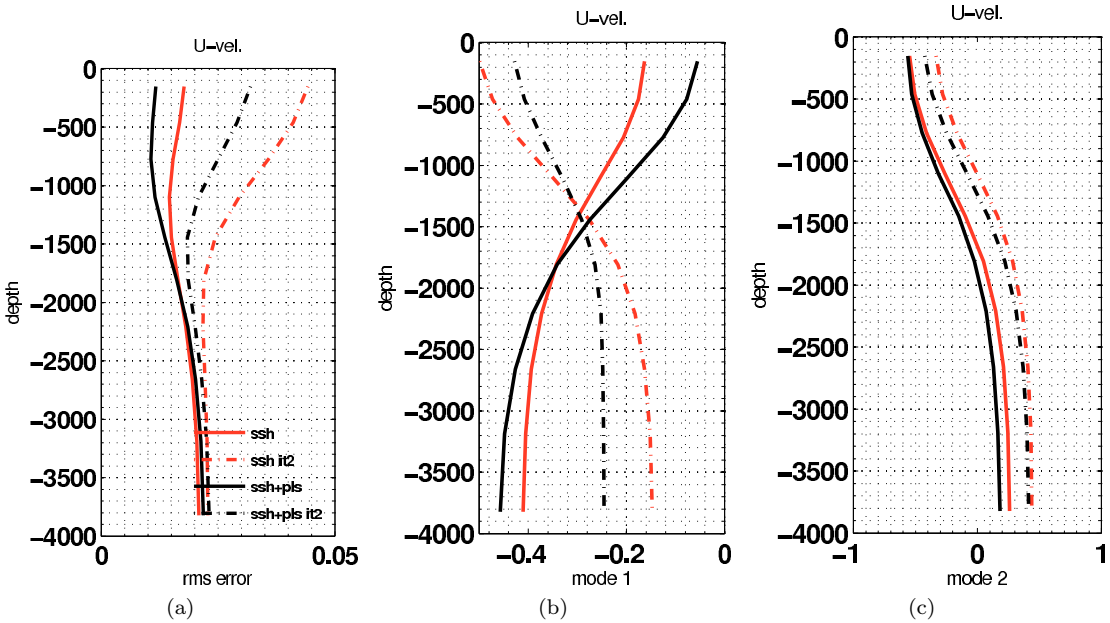
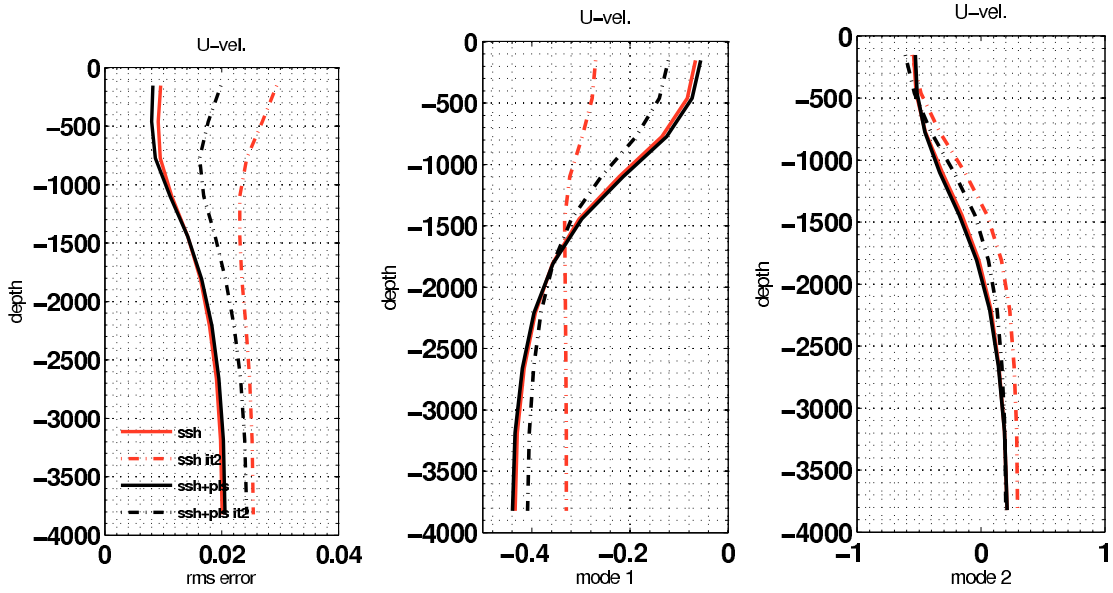


Figure 4.15



(d) (e) (f)

Figure 4.15: (**a** and **d**) Root Mean Squared errors of zonal velocity for the experiments ssh_10d_rd (red curve) and ssh_10d_rd_it2 (dashed red curve) and their equivalents but with K constructed using PLS regression (black curves). (**b** and **e**) First EOF mode and (**c** and **f**) second EOF mode calculated with the zonal velocity error. Top/bottom panel are results of the $1^o/130^o$ day of the experiment.

Temporal data sparsity: gridded SSH observations available every 4 days

In this section the experiments presented above have been reproduced but with observations available every 4 days (see experiments ssh4_10d_rd_uv, ssh4_10d_rd, ssh4_10d_rd_it2_uv and ssh4_10d_rd_it2 in Table 4.5). Temporal sparsity is in general a great challenge for nudging methods since they usually correct the observed variables but nothing is done with respect to the non-observed part of the state vector.

Figure 4.16 shows the results with and without the use of the regression model. In this case the experiments errors using the PLS regression are 40% smaller for the velocities, 60% for the SSH and 15% for the temperature. Once again, using only 2 iterations is beneficial to keep the temperature errors stable. However as shown in Fig. 4.15, iterations are responsible for the correction of the upper ocean velocities, resulting in better initial conditions and a more stable predictive step.

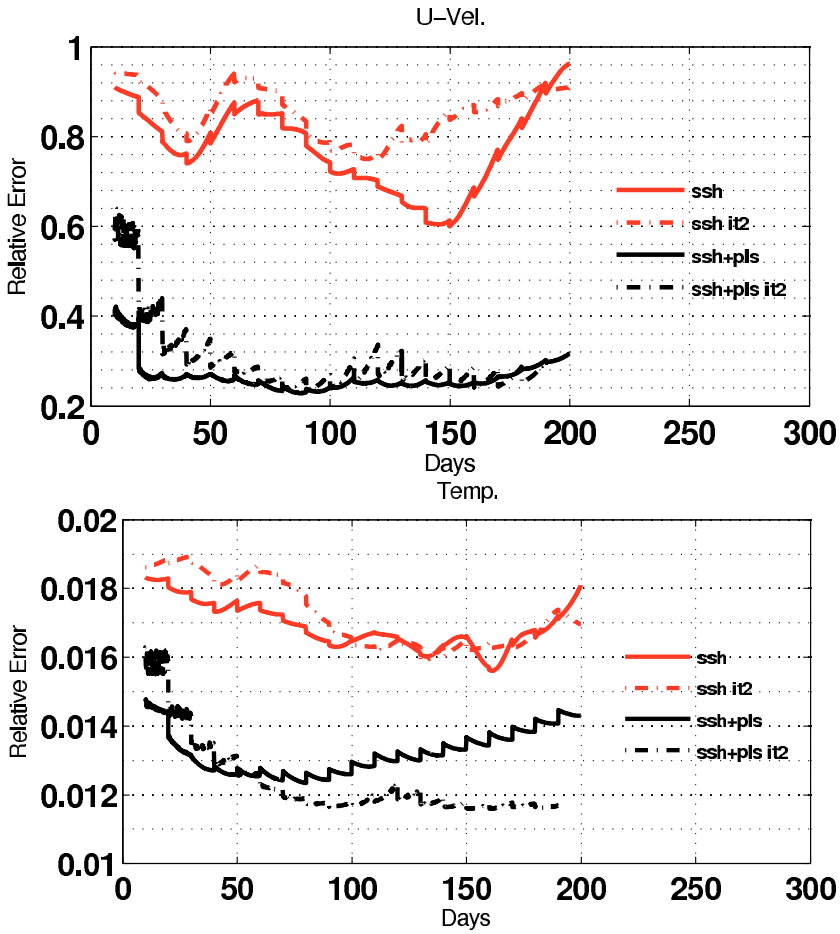


Figure 4.16: Relative errors of the zonal velocity (top) and the temperature (bottom) for the experiments ssh_10d_rd (red curve) and ssh_10d_rd_it2 (dashed red curve) and their equivalents with K constructed using PLS regression (black curves).

Therefore, we have seen that the use of **K**s accounting for corrections of the non-observed part of the state vector is mandatory in situations where the observation network is poor, as it is the case in real ocean applications.

Consequently, the results presented in the previous sections are used in the following part to configure an experiment that uses a more realistic observation network.

Intercomparisons

In this section we assume that every four days, an observation network simulating Jason-1 satellite sample is available. In addition, to validate the results produced by the DBFN, a comparison with the 4Dvar method is presented. Also, the temperature

is included in the regression model. This makes the comparison with the 4Dvar more equitable, since the non-observed variables in 4Dvar are estimated by considering the multivariate balance operator $\mathbf{G}$ (see Sect 4.2.2).

First, the minimization performance of the 4Dvar implementation is analyzed. Figure 4.17 shows the reduction of the cost function gradient for the first assimilation cycle. 4Dvar takes 26 iterations to approximately achieve the optimality condition $\nabla J = 0$. This represents 3 times the number of iterations required by the DBFN to converge, i.e., after which the errors cease to decrease. Moreover, the 4Dvar numerical cost is more than 3 times the DBFN cost since one execution of the adjoint model costs four times the cost of the direct model in terms of CPU time.

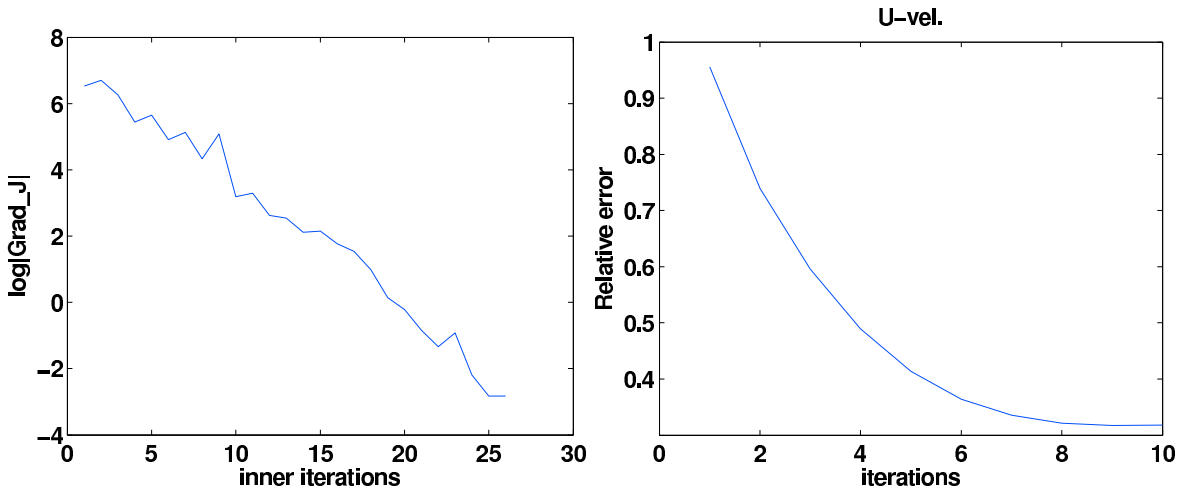
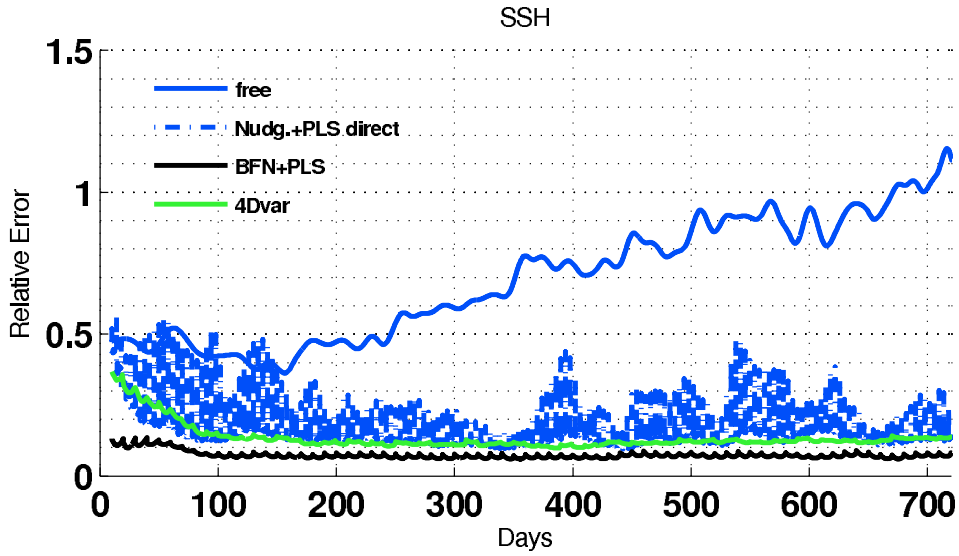


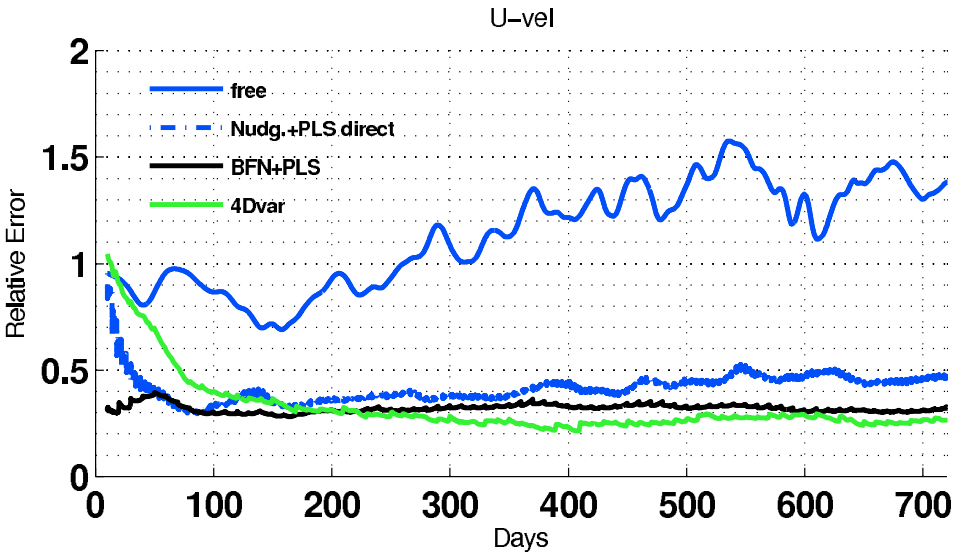
Figure 4.17: Figure shows the gradient of the cost function after each inner iteration (left) and the reduction of the relative error for the zonal velocity for experiment ssh4j_10d_rd_uvt (right).

Figure 4.18 show the relative error for the control experiment (without assimilation), an experiment using the direct nudging with PLS regression (ONDG), the DBFN+PLS experiment and a 4Dvar experiment. The DBFN+PLS experiment error is stable throughout the test while for ONDG and 4Dvar errors stop decreasing after 100 and 200 days, respectively. This is a benefit of the iterations performed by the DBFN when model and data are quite different. Among the experiments conducted, the DBFN+PLS produced the best results for all variables, except for the zonal velocity, for which the 4Dvar has slightly smaller errors. The ONDG also showed good performance, but with

mean errors greater than DBFN+PLS and 4Dvar. Also, it has initialization problems that can be identified in the Fig. 4.18 as great oscillations after each assimilation step.

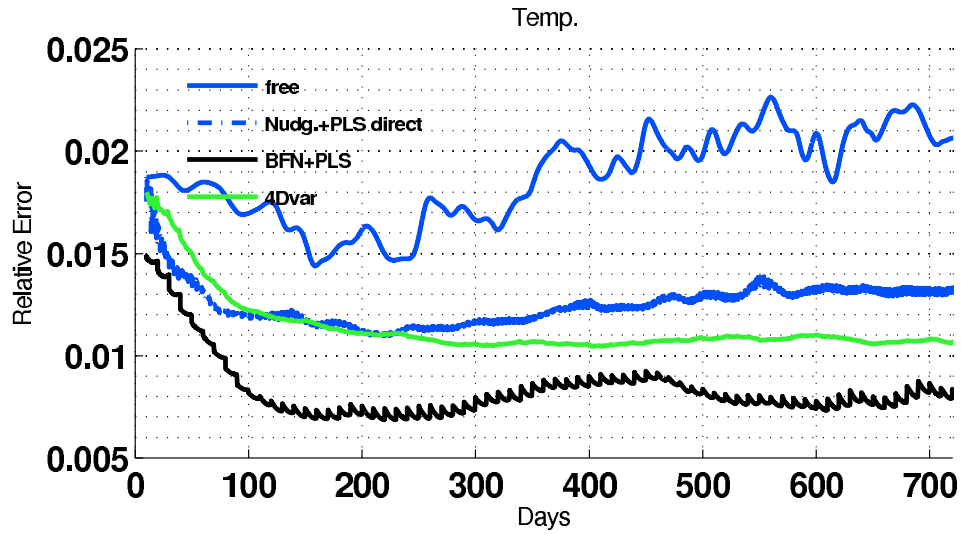


(a)



(b)

Figure 4.18



(c)

Figure 4.18: The figure shows errors of the SSH (**a**), the zonal velocity (**b**) and the temperature (**c**). Each curve correspond to a different experiment, see table 4.5 for more details.

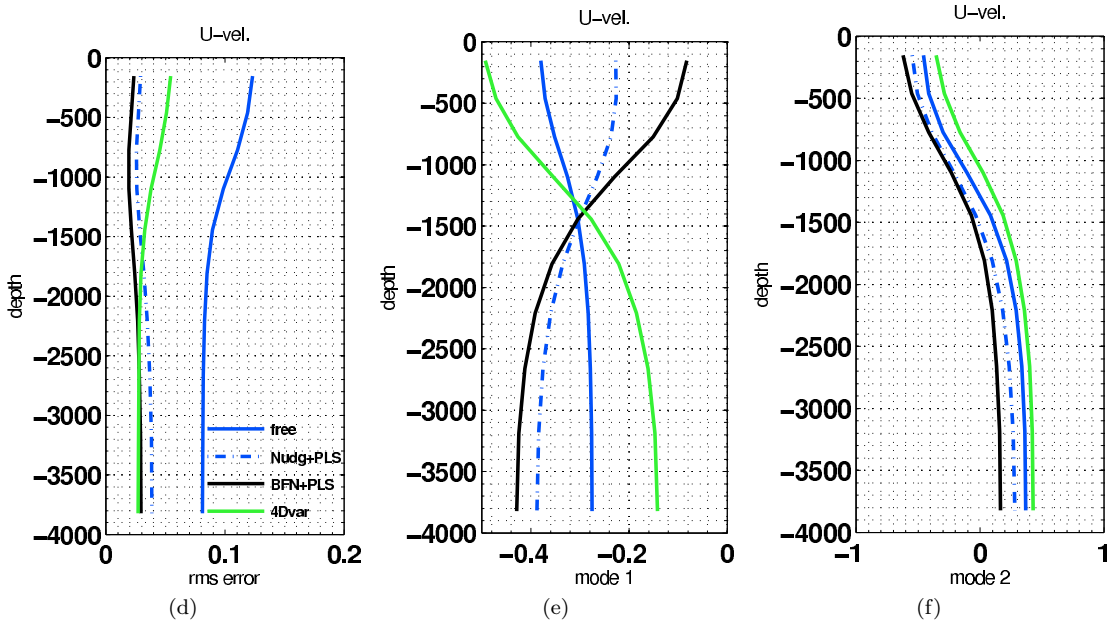
A summary of the experiments mean relative error is presented in table 4.6. The DBFN+PLS performance under a sparse observation network is comparable with the results obtained assimilating daily full SSH fieds, see table 4.2.

EXP.	SSH	U	V	T
free	0.695	1.140	2.049	0.0187
Nudg+PLS	0.141	0.416	0.549	0.0125
DBFN+PLS	0.073	0.3198	0.405	0.0084
4Dvar	0.146	0.333	0.920	0.0249

Table 4.6: Summary of the mean relative error for the control experiment (Free), the ordinary Nudging employing the PLS (Nudg+PLS), the DBFN employing the PLS (DBFN+PLS) and the 4Dvar.

In terms of vertical error (Fig. 4.19), the DBFN+PLS and the ONDG performed better for the upper ocean than 4Dvar. Clearly, the PLS also corrects the deep ocean velocity, but less accurately than 4Dvar does. The first error mode is the barotropic one and accounts for 97% of the error variability for 4Dvar, 96% and 93% for BFN+PLS and ONDG, respectively. This result confirms the iterations role in improving the baroclinic fields. Although the first mode is the barotropic one for all methods, it is important to note that the 4Dvar barotropic mode is out of phase with respect to the PLS barotropic mode. This reflects the better performance of the 4Dvar for the deep ocean and the

better performance of the PLS for the upper ocean.



(d) (e) (f)

Figure 4.19: (**a**) RMS of vertical zonal velocity and first (**b**) and second (**c**) eof error modes calculated using forecast from day 200 to day 720.

The way both methods correct the model depends on the B matrix in the 4Dvar algorithm and on the latent structures in the BFN+PLS. It means that results may be different if their calculation is changed. The main aspect of the results is that with a method which is easier to implement and cheaper to execute we can produce results that are at least equivalent to 4Dvar. Also, it is shown that iterations is an important aspect of the method: Iterations compensate for the lack of a priori information on the model errors as well as filter out noise in the observations. The latter fact must be connected to the diffusive character of the algorithm. Moreover, the iterations allow us to put information from the observations into the model, without causing initialization problems since the nudging gain is relatively weak compared to the dominant term of the equations.

4.7 Conclusions and perspectives

This chapter used the NEMO general circulation model and the double gyre configuration to investigate the Diffusive Back and Forth Nudging performance under different configurations of the data assimilation window, gains and observations network, as well as to compare the DBFN with the 4Dvar.

It has been shown that the reliability of the backward integration should be carefully examined when the BFN/DBFN is applied to non-reversible systems. This should support the choice of the assimilation window and identify whether the available observations are sufficient to control the errors induced by the non-reversible terms of the model equations. We have shown that the DBFN can be used for ocean data assimilation despite the low accuracy of the backward integration. Indeed, improving the backward integration would further improve the DBFN performance and make possible the use of longer assimilation windows.

The use of scalar gains in the DBFN requires high spatial and temporal availability of data, otherwise, the method does not bring significant improvements due to the model "inefficiency" in spreading the observations information and hence the impossibility of controlling the diffusion errors. In the case of sparse data, complex functions are necessary to propagate the information from the data to the non-observed variables and to regions of the domain that are not observed. These functions were constructed using the PLS latent structures resulting from the process of maximizing the covariance between the observed variables and the variables to be estimated. In our implementation the PLS model does not vary temporally which makes the method less computationally demanding. Notably, the iterations give a dynamic character to the solution since the increments are propagated using the nonlinear model.

Our results show that the DBFN can produce results comparable to the 4Dvar, but using lower computational power. This is because DBFN demands less iterations to converge and because one iteration of 4Dvar corresponds to one integration of the tangent linear model, one integration of the adjoint model, which costs four times more than one standard model integration, plus the cost of minimizing the cost function, while the DBFN costs twice the integration of the nonlinear model.

The twin experiment framework is favorable to the perfect model assumption used

by our 4Dvar. However, the 4Dvar had difficulties to fit the SSH observations available in one assimilation window. This is observed especially at the beginning of the experiment, when the background state is far from the observations. This can be an indicative of conditioning problem. The DBFN, however, does not suffer from this kind of problem: the errors on the dynamical variables reach their asymptotic values after the first assimilation cycle.

In the next chapter, the SSH sampled at the Jason-1 satellite track is assimilated along with vertical temperature profiles distributed so as to mimic the ARGO buoys distribution in the studied region. The assimilation methods rely on the Kalman Filter update scheme to calculate the gain $\mathbf{K}$. Therefore, the two step scheme presented in section 4.8.2 is applied but the weights given to model and observations are based on their respective uncertainties and hence the Nudging strength is calculated in a more objective way. In addition, the steady regression model is replaced by a regression model that uses the covariance matrix that evolves in time to track the errors evolution. Another aspect is that the observational set decomposition employed by the Back and Forth Kalman Filter provides in the linear Gaussian case an optimal gain and an optimal number of iterations.

4.8 Appendix

In this section the differences between the Ordinary Least Squares regression and the Partial Least Squares regression are discussed and the numerical results concerning the estimation quality are presented.

4.8.1 Ordinary Least Squares regression (OLS)

We consider $\mathbf{X} \in \mathbb{R}^{n \times M}$ as the observed or predictor variables and $\mathbf{Y} \in \mathbb{R}^{nxN}$ as the non-observed or response variables. In our notation n is the sample size and M and N are the size of the state space of $\mathbf{X}$ and $\mathbf{Y}$ respectively. Also, $\mathbf{X}$ and $\mathbf{Y}$ are centered and have the same units. Linear regression models are usually written as $\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}$, where $\mathbf{E} \in \mathbb{R}^{n \times M}$ is the residual error and $\beta \in \mathbb{R}^{m \times n}$ is the regression matrix to be estimated. It is well known that OLS solves the problem:

$$\hat{\beta} = \arg\min_{\beta} \|\mathbf{Y} - \mathbf{X}\beta\|,$$

which solution is $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}\mathbf{Y}$. Problems arise when $\mathbf{X}^T\mathbf{X}$ is poor conditioned, which is the case when variables in $\mathbf{X}$ are correlated or partially correlated, and when $\mathbf{X}$ is too large limiting the use of direct solvers. The OLS belongs to a great class of linear estimators called BLUE (Best Linear Unbiased Estimator). Its unbiased condition is verified if $\beta \in range(\mathbf{X})$.

4.8.2 Partial Least Squares regression (PLS)

The PLS was first introduced by Wold (1975) to address the problem of econometric path modeling, and was subsequently adopted for regression problems in chemometric and spectrometric modeling. As far as we know, for the first time PLS is being used and compared with traditional OLS in the oceanographic context.

The most important difference between OLS and PLS is that the later assumes that the maximum information about the response ($\mathbf{Y}$) is in those directions of the predictor ($\mathbf{X}$) space which simultaneously have the highest variance and the highest correlation

with the response. PLS decomposes $\mathbf{X}$ and $\mathbf{Y}$ as (Rosipal and Kramer (2006)):

$$\begin{aligned} \mathbf{X} &= \mathbf{T}\mathbf{P}^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{U}\mathbf{Q}^T + \mathbf{F} \end{aligned} \tag{4.11}$$

where $\mathbf{T} \in \mathbb{R}^{N\times p}$ and $\mathbf{U} \in \mathbb{R}^{M\times p}$ are matrices of p components, $\mathbf{P} \in \mathbb{R}^{M\times p}$ and $\mathbf{Q} \in \mathbb{R}^{N\times p}$ are matrices of loadings, and $\mathbf{E} \in \mathbb{R}^{nxM}$ and $\mathbf{F} \in \mathbb{R}^{nxN}$ are the residual errors. PLS as well as principal component regression can be seen as methods to construct a matrix of latent components $\mathbf{T}$ as a linear transformation of $\mathbf{X}$:

$$\mathbf{T} = \mathbf{X}\mathbf{W},$$

where $\mathbf{W} \in \mathbb{R}^{N\times p}$ is a matrix of weights. The regression can be calculated considering that there exists a linear relationship between $\mathbf{U}$ and $\mathbf{T}$ of the form:

$$\mathbf{U} = \mathbf{T}\mathbf{D} + \mathbf{H}, \tag{4.12}$$

where $\mathbf{D} \in \mathbb{R}^{p\times p}$ is a diagonal matrix and $\mathbf{H}$ is the residual error. In practice this is not a good relation because $\mathbf{T}$ and $\mathbf{U}$ are calculated independently and hence they may be weakly related. The way the algorithm we have used consider this inner relation is detailed latter in this section. Combining equation 4.11 with equation 4.12 we get an expression for $\mathbf{Y}$ which will permit explicit $\mathbf{Y}$ in terms of $\mathbf{X}$:

$$\begin{aligned} \mathbf{Y} &= \mathbf{T}\mathbf{D}\mathbf{Q}^T + (\mathbf{H}\mathbf{Q}^T + \mathbf{F}) \\ &= \mathbf{T}\mathbf{C}^T + \mathbf{F}^* \end{aligned} \tag{4.13}$$

Then, using the relation $\mathbf{T} = \mathbf{X}\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}$ (Tenenhaus (1998)) the PLS regression coefficients $\hat{\beta}^{PLS}$ is given by:

$$\hat{\beta}^{PLS} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{C}^T \tag{4.14}$$

It is readily seem that if we set $\mathbf{W} = \mathbf{I}$ we obtain the OLS solution.

The PSL components, which are uncorrelated linear combinations of all the predictors, are calculated according to following optimization problem:

$$\begin{aligned} cov(\mathbf{t},\mathbf{u})^2 &= \max_{\|\mathbf{r}\|=\|\mathbf{s}\|=1} \{cov(\mathbf{X}\mathbf{r},\mathbf{Y}\mathbf{s})^2\} \\ &= \max_{\|\mathbf{r}\|=\|\mathbf{s}\|=1} \{var(\mathbf{X}\mathbf{r})cor(\mathbf{X}\mathbf{r},\mathbf{Y}\mathbf{s})^2 var(\mathbf{Y}\mathbf{s})\} \end{aligned} \tag{4.15}$$

In other words the PLS is a compromise between the canonical analysis of $\mathbf{X}$ and $\mathbf{Y}$, $\max\{cor(\mathbf{Xr}, \mathbf{Ys})^2\}$, and the principal component analysis of $\mathbf{X}$ and $\mathbf{Y}$ given by $\max\{var(\mathbf{Xr})\}$ and $\max\{var(\mathbf{Ys})\}$ respectively (Rosipal and Kramer (2006)). The numerical algorithm used to calculate the vectors $\mathbf{t}$ and $\mathbf{u}$ is due to Wold (1975) and is briefly described in the following.

This algorithm is called PLS-W2A in the literature. It will be given a sketch of the algorithm extract from Wegelin (2000).

1. $k \leftarrow 1$
2. $\mathbf{X}^{(1)} \rightarrow \mathbf{X}$
   $\mathbf{Y}^{(1)} \rightarrow \mathbf{Y}$
3. Compute the left $\mathbf{r}_k$ and right $\mathbf{s}_k$ singular vectors associated with the first singular value of $(\mathbf{X}^{(k)})^T\mathbf{Y}^{(k)}$ thanks to the power method:
   (a) Initialize $\mathbf{r}^0$ and $\mathbf{s}^0$
   (b) Repeat until convergence:
   - $z \leftarrow z + 1$
   - Update latent vector:
     $\mathbf{t}^z \leftarrow \mathbf{X}\mathbf{r}^{z-1}$
     $\mathbf{u}^z \leftarrow \mathbf{Y}\mathbf{s}^{z-1}$
   - Update $\mathbf{r}^z$ and $\mathbf{s}^z$:
     $\mathbf{r}^z \leftarrow \mathbf{X}^T\mathbf{u}^z((\mathbf{u}^z)^T\mathbf{u}^z)^{-1}$
     $\mathbf{r}^z \leftarrow \mathbf{r}^z/\|\mathbf{r}^z\|$
     $\mathbf{s}^z \leftarrow \mathbf{Y}^T\mathbf{t}^z((\mathbf{t}^z)^T\mathbf{t}^z)^{-1}$
     $\mathbf{s}^z \leftarrow \mathbf{s}^z/\|\mathbf{s}^z\|$
4. $\mathbf{t}_k = \mathbf{X}^{(k)}\mathbf{r}_k$,
   $\mathbf{u}_k = \mathbf{Y}^{(k)}\mathbf{s}_k$,
5. Compute rank-one approximation of the data matrices by regressing $\mathbf{X}^{(k)}$ and $\mathbf{Y}^{(k)}$ on $\mathbf{t}_k$:
   $\hat{\mathbf{X}}^{(k)} = \mathbf{t}_k(\mathbf{t}_k^T\mathbf{t}_k)^{-1}\mathbf{t}_k^T\mathbf{X}^{(k)}$
   $\hat{\mathbf{Y}}^{(k)} = \mathbf{t}_k(\mathbf{t}_k^T\mathbf{t}_k)^{-1}\mathbf{t}_k^T\mathbf{Y}^{(k)}$

6. Subtract the rank-one approximation to obtain remainder matrices:
   $\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - \hat{\mathbf{X}}^{(k)}$
   $\mathbf{Y}^{(k+1)} = \mathbf{Y}^{(k)} - \hat{\mathbf{Y}}^{(k)}$

7. If $(\mathbf{X}^{(k+1)})^T\mathbf{Y}^{(k+1)} = 0$ or the model dimension should not exceed the current value of k:
   $Rank \leftarrow k$ this is the rank of the PLS model.
   Exit

8. $k \leftarrow k + 1$

9. Go to step 3.

The number of components p is chose based on cross-validation. This method involves testing a model with objects that were not used to build the model. We divided the data set in two contiguous blocks; one of them is used for training and the other to validate the model. Then we search for the number of components which gives the best results in terms of mean residual error and estimator variance.

It should be noted that while the OLS is an unbiased estimator the PLS breaks the unbiased condition of the BLUE estimators. Nevertheless, it has better predictive skills because PLS shrinks the directions of the OLS estimator that are responsible for high variance. In other words, we discharge the eigenvalues of $\mathbf{X}^T\mathbf{X}$ that are very small expecting that the increase in bias is small compared to the decrease in variance.

In section 4.8.3 we analyze the regression performance using two criterions: the Mean Squared Error of the residuals (mse) and the coefficient of determination R^2 (Tenenhaus (1998)). When these tests are applied to the training data it tells us about the fit of the model to the data set used. When applied to the validation block of data is tells us about the predictive skill of the model. The mse is calculated according to:

$$mse = \frac{1}{N}\sum(\mathbf{Y} - \hat{\mathbf{Y}})^2. \tag{4.16}$$

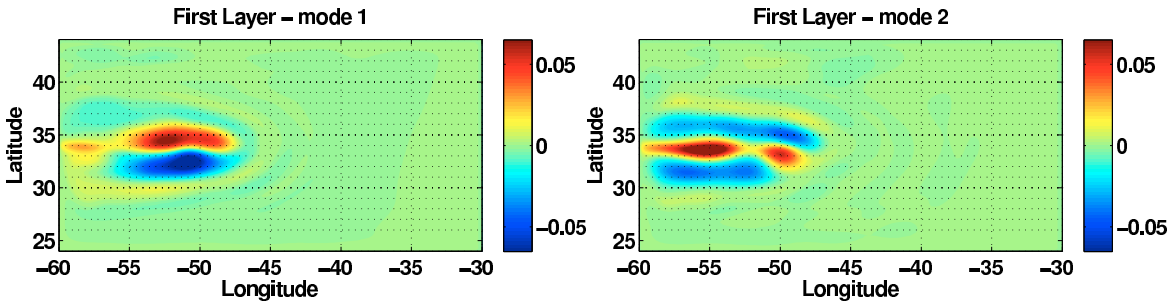
where $\hat{\mathbf{Y}}$ is the estimation of $\mathbf{Y}$. The R^2 test which compares the variance of the estimation with the variance of the data is given by:

$$R^2 = 1 - \frac{\sum(\mathbf{Y} - \hat{\mathbf{Y}})^2}{\sum(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))^2}. \tag{4.17}$$

4.8.3 Numerical Results

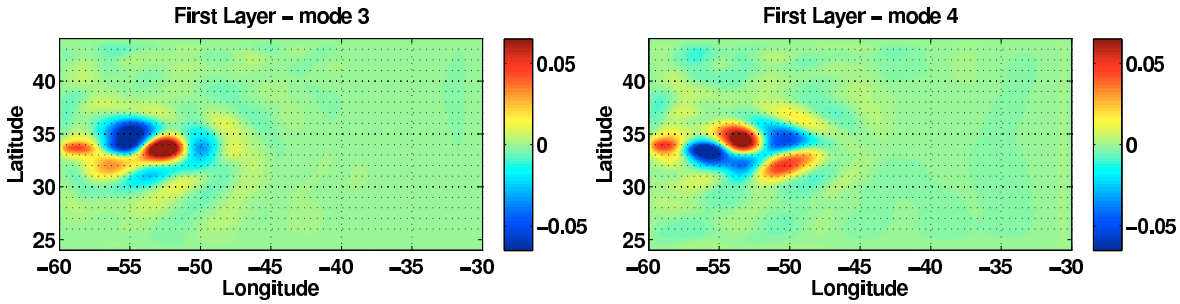
In this section, the results produced by the regression models described above are analyzed. The regression models were constructed using 10 years of a free model run sampled every 5 days. It is possible that the decorrelation time of some processes are longer than 5 days, hence the samples may be partially correlated, which is a problem for the OLS estimator. SSH was considered as the predictor variable and velocities and temperature the response variables. One regression model for each layer has been constructed, therefore differing from the approach of calculating vertical modes for each grid point used by Gavart and De Mey (1997).

Figure 4.20 shows the first four SSH modes that correlates with the surface velocity modes and Fig. 4.21 their vertical correlation with the deep layers. The first 4 modes are highly correlated with their respective modes at deeper layers. They represent the dynamics of eddies formation and westward displacement. Visual inspection revealed that their velocities components are in approximate geostrophic balance (not shown). The first two modes, which represent the largest spatial scales, are orthogonal with layers deeper than $1800m$ (mode 1) and $1200m$ (mode 2). This would be an indication that the model's SSH is governed by the first baroclinic mode, and therefore explains why assimilating only SSH by projecting it into the barotropic mode is not enough to control the deep ocean errors. Modes 3 and 4 are in phase and have high correlation, i.e. they have a barotropic signature. Higher order modes are usually numerical artifacts of the orthogonality condition imposed by the method. Therefore, it is difficult to make conclusions on their physical nature.



(a) (b)

Figure 4.20



(c) (d)

Figure 4.20: The panels represent the first four loading vectors of SSH used in the calculation of the regression model between SSH and surface velocity. More specifically, they are the first four columns of the matrix $\mathbf{P}^T$.

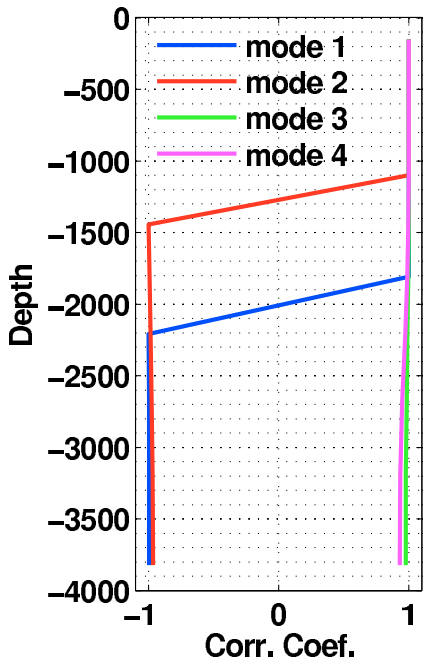


Figure 4.21: Vertical correlation of the modes presented in figure 4.20 with their respective modes at deeper layers.

Regarding the regression, OLS performed slightly better than PLS in terms of the mse and R^2 (Fig. 4.22). OLS produced null mse and R^2 equals one. This is the best result that could be expected. For the PLS the higher the number of components considered the better the fitting, but never better than OLS. This was expected since the PLS is an approximate solution of OLS. In the case for which 400 components is considered the mse and $(1 - R^2)$ is of order of 10^{-7} and 10^{-4}, respectively.

Considering a data set (two years of independent model simulations) not used in the construction of the regression models, the "optimal" number of PLS modes (components) vary with depth, see Fig. 4.23. While increasing the number of modes ameliorates the surface layers estimation for the deep layers it deteriorates both the mse and R^2. The mse of PLS is about 40% smaller than OLS one, when the number of modes is chose to be optimal. OLS fails for layers deeper than $1500m$. In this case, the R^2 is negative which means that the variance of the estimator is higher than the natural variance of the system. In other words, the mean $\mathbb{E}(\mathbf{Y})$ is a better estimative of $\mathbf{Y}$ than the $\hat{\mathbf{Y}}$ calculated by OLS. The better performance of the PLS is related to the bias-variance trade-off considered by the PLS and not by the OLS. This is a well known

result for regression models (Faber, 1999).

Results suggests that the PLS model may be favorable to be used in data assimilation experiments, since in these cases the predictor is in fact the increments of the predictor variables not used in the construction of the regression model. Results show the benefit of the shrinkage of the components applied by PLS to regularize the solution. A different application of PLS could be the reconstruction of salinity profiles from temperature and SSH data as in Nardelli and Santoleri (2005).

The vertical error structure predicted by the PLS (Fig. 4.23) is quite similar to that one produced by the DBFN when only SSH is assimilated (Figs. 4.12 and 4.13). This corroborates the fact that for the model configuration we have been using the SSH is not a good predictor variable of the deep ocean velocities. Therefore, we hypothesize that a significant part of the errors at the end of the DBFN assimilation step is not due to the backward integration but to the weak relationship between SSH and the deep ocean dynamics.

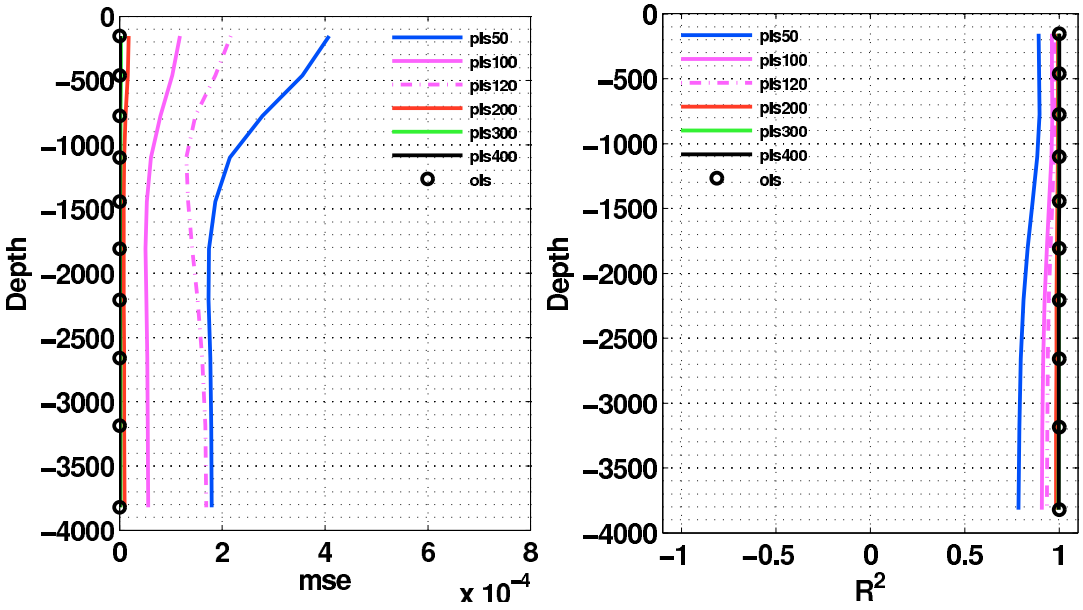


Figure 4.22: Mean Squared error of the residuals (left panel) and R^2 score (right panel) for the PLS algorithm using different number of modes, indicated in the legend, and for OLS. Results of fitting, i.e. the statistics are calculated using objects used in the construction of the regression model.

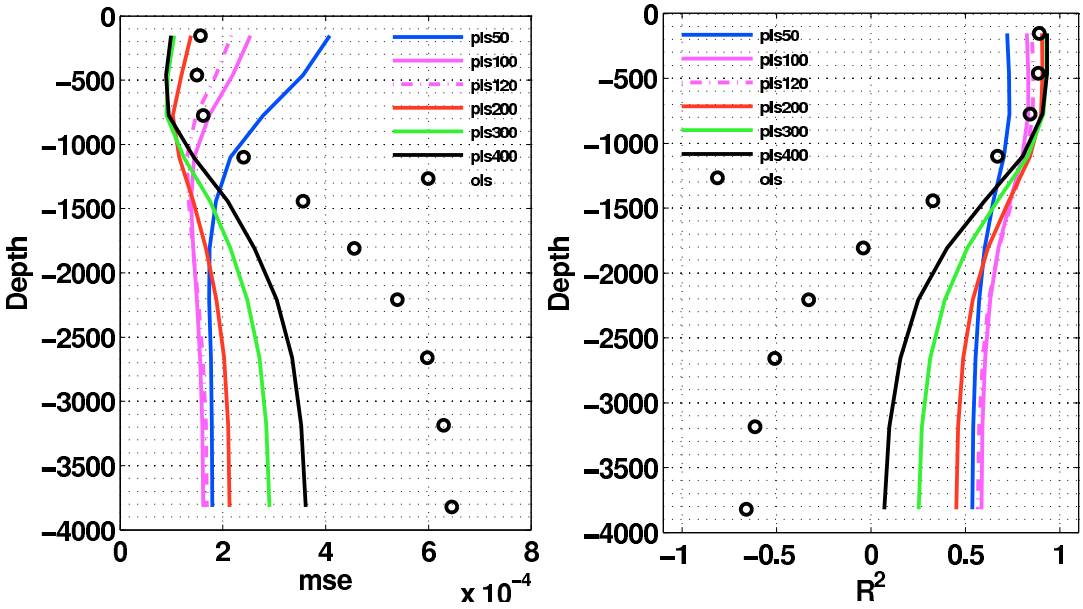


Figure 4.23: Mean Squared error of the residuals (left panel) and R^2 score (right panel) for the PLS algorithm using different number of modes, indicated in the legend, and for OLS. Results of prediction, i.e. the statistics are calculated using objects not used in the construction of the regression model.

Chapter 5

Diffusive Back and Forth Kalman Filter

Contents

5.1 Introduction

The use of smoothers in the DA context began in the latter 80's with the works of Bennett and Budgell (1989) and Stephen et al. (1994). These authors introduced a smoother based on the Kalman filter to estimate the probability of the model and observations in a given time window. Smoothers are usually used in meteorology to produce what is called "reanalysis". These products are used for a variate of purposes such as the study of physical processes (Hirata et al., 2013; von Engeln and Teixeira, 2013), boundary and initial conditions for regional models (Liston and Pielke, 2000) and to force oceanic models (Silva et al., 2009) for instance. They are known as the best available long term description of the atmosphere.

Nowadays some initiatives for producing public ocean reanalyses exists as for example: the ECCO[1], the BLUElink[2] and the ECMWF[3]. Some recent methodological development concerning oceanographic applications are the works of Cosme et al. (2010), Barth et al. (2010) and Freychet et al. (2012), all of them using a reduced order approximation.

In the chapter 2, a reinterpretation of the BFN has been presented as an iterative smoother using the perspective of the Bayesian estimation. In addition, it was shown that under the linear and perfect model assumptions, the Backward Smoother (BS) is equivalent to the SEEK-based smoother (SEEK-smoo) derived by Cosme et al. (2010). In other words, the use of the backward model to propagate the final filter analysis is equivalent to the use of linear time correlations to correct a past state.

Nevertheless, when the model is non-linear, the differences between them may become significant especially if the smoother lag is chosen so as to consider the non-linear regime. In this case, the use of linear time correlations may be no more suitable while the BS transports the observation information backwards with the non-linear model.

However some practical issues concerning the methods should be analyzed. The first one is related to sampling problems that are important when the rank of $\mathbf{S}$ is too small, i.e. it is smaller than the rank of the unstable and neutral subspace (Palatella et al., 2013). It results in two errors: spurious covariances and underestimated forecast er-

[1] Estimating the Circulation and Climate of the Ocean; http://www.ecco-group.org/about.htm
[2] http://www.csiro.au/Outcomes/Oceans/Oceans-and-climate/BLUElink.aspx
[3] European Center for Medium range Weather Forecast;
http://www.ecmwf.int/products/forecasts/d/charts/ocean/reanalysis/

rors. In the SEEK filter, the former is mitigated by using a spatial localization that cuts-off the observation influence for observations that are at a distance larger than a prescribed value, while the latter is mitigated by inflating the error statistics in the subspace spanned by the columns of $\mathbf{S}$. These *ad hoc* solutions are far from being optimal, first because the influence range of an observation may vary in time and space, and second because the subspace may be ideally inflated with new directions that take into account unstable and neutral directions.

For the SEEK-smoo, besides the spatial localization, the temporal localization plays an important role on the performance, i.e. the smoother time-lag should be carefully chosen. The optimal lag may be coincident with the doubling error time of the system, which is related to the Lyapunov exponents (Nerger *personal communication*). In other words, the nonlinearities act to reduce the degree of linear correlation between two instants of time.

Moreover, spatial and temporal localization may be at some extent correlated since as the time lag increases the spatial influence of a given observation may be extended. This is not taken into account by the SEEK-smoo because the weights used in the smoothing phase are calculated by the local filter analysis, while it is naturally held by the BS since the influence of a future observation is controlled by the model advection and diffusion. At this point we wonder if the use of the dynamical model, for which the non-linear dynamic is perfectly reversible, can improve the estimation of past states.

The second practical issue is related to the model errors. As we have seen in the chapter 4, the backward integration introduces some errors related to the diffusive aspect of the algorithm. This imposes a restriction on the length of the assimilation window to be used. Therefore, possible benefits of using the nonlinear model to smooth the solution may not be reached.

Concerning the iterative algorithms, one may expect that the use of the observations more than once may improve the filter/smoother estimation since the reduced basis usually does not span the entire unstable and neutral subspace and because sometimes the errors can be so large that they are not governed by the linear dynamics. In other words, the iterations may compensate for the lack of optimality due to the approximations considered in deriving the reduced order filter.

The iterative procedure has been shown to provide a mechanism for diversifying the subspace spanned by the reduced basis. As it was showed in section 2.4.3, the final smoother increment may be seen as a linear combination of the matrix $\mathbf{S}^i$ for i iterations. Therefore, the increment is in a space which is expected to be larger than the rank of $\mathbf{S}$. Furthermore, iterative algorithms are usually beneficial to be used along with linearized methods since at each iteration the system is linearized around a trajectory which is closer to the true trajectory (Jazwinski, 1970; Kalnay and Yang, 2010; Yang et al., 2012).

Therefore, this chapter is devoted to draw the advantages and disadvantages of using the dynamical model to smooth the filter solution as well as to verify how iterations may help to mitigate sampling problems and to handle the non-linearities.

5.2 Objectives

- To analyse whether the use of the nonlinear model to propagate the observations information backward in time is more suitable than the use of linear time correlations;
- To understand how does the optimal lag changes with respect to the methods proposed and to the ensemble size;
- To analyse the impact of iterating the smoothers;
- To apprehend the sensitivity of the ensemble size;

5.3 Data Assimilation Experiments

5.3.1 Filter and smoother initialization

An ensemble was built using 10 years of a mature run[4]. Fields were taken every 5 days. The mean was taken as the initial state and an EOF was computed to approximate the error covariance matrix. Four different rank sizes were retained to initialize the covariance matrix: 10, 20, 50 and 100. With these four different ranks it is possible

[4]mature run is a model simulation in a statistical steady state.

to analyse the impact of sampling errors on the smoother and filter solution. As our scheme, Backward Smoother, also relies on the filter solution, it is possible to analyzed the impact of back-propagating the "spurious" increments with the non-linear model.

5.3.2 Covariance localization and inflation

For all experiments, independently of the rank, the so-called domain localization (DL) is applied (Ott et al., 2004; Hunt et al., 2007). In this case, each grid point is updated independently by the filter analysis. For each local analysis, only observations within some defined cut-off radius are considered. In conjunction with the DL, the method of observation localization (OL) is used. In the OL method the inverse of the observation error covariance matrix corresponding to a local analysis domain is Schur-multiplied with a chosen localization matrix that is constructed using correlation functions. Thus, the observations weight is reduced as a function of their distance from the local analysis domain by increasing their assumed error variance. In our experiments the cut-off radius is 400km, based on the barotropic Rossby deformation radius, and the observation error covariance is multiplied by a Gaussian function with standard deviation equal to 170km.

Furthermore, forecast covariance inflation as proposed by Pham et al. (1998) is also implemented for both filter and smoother. In this technique the model error $\mathbf{Q}$ is proportional to the forecast error covariance:

$$\mathbf{Q}_{k+1} \propto \frac{1-\rho}{\rho}\mathbf{M}_{k+1,k}\left(\mathbf{M}_{k+1,k}\mathbf{P}_{k|k}\right)^T \tag{5.1}$$

which under the reduced rank approximation gives

$$\mathbf{S}_{k+1|k} = \frac{1}{\sqrt{\rho}}\mathbf{M}_{k+1,k}\mathbf{S}_{k|k} \tag{5.2}$$

As it was discussed in section 2.5.2, the SEEK-smoo relies on the calculated SEEK filter weights (Eqs.2.108 and 2.112) to estimate the smoothed states. Also, since $\mathbf{Q}$ models a white sequence, i.e. the errors are uncorrelated in time, the cross-covariance using Eq.5.2 should be written as:

$$\mathbf{P}_{k+1,i|k} = \sqrt{\rho}\mathbf{S}_{k+1|k}\mathbf{S}_{i|k} \tag{5.3}$$

where Eq.(2.113) is modified to

$$\mathbf{S}_{i|k} = \sqrt{\rho}\mathbf{S}_{i|k-1}\left[\mathbf{I} + \Gamma_k\right]^{-1/2} \tag{5.4}$$

5.3.3 Observations Network

The assimilated observations are the SSH simulated along Jason-1 satellite track and perturbed with a Gaussian white noise (0.03cm), and temperature profiles that mimics the ARGO vertical sampling and provides the same observation density sampling as the ARGO system for the modeled region. Figure (5.1) shows a typical two days observations availability and the observations accumulated after one cycle of the Jason-1 satellite, which corresponds to ten days, and one cycle of the ARGO-like system, which corresponds to eighteen days.

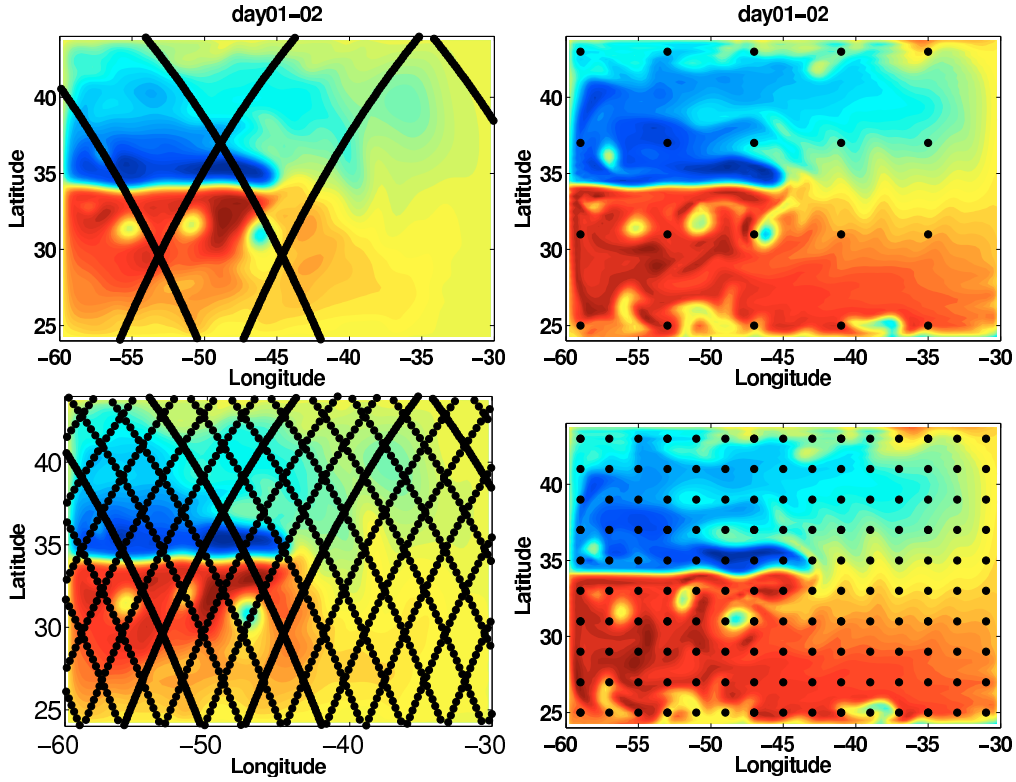


Figure 5.1: Top panels: Sea Surface Height (left) and Temperature profiles (right) observational network accumulated for a 2 days window. Bottom panels: Sea Surface Height (left) and Temperature profiles (right) observational network accumulated for 10 and 18 days window respectively

The analyses are performed every 2 days. The optimal smoother lag is analyzed in the section 5.4.1 for which we analyzed which lag gives the smaller error estimation. As observed by (Cosme et. al 2010) this is only possible thanks to the twin experiment framework we used. In the cases for which the truth is not known, one possible way to chose the lag is to apply the cross-validation technique (Tenenhaus, 1998).

For all experiments the observational error covariance, $\mathbf{R}$, is diagonal. Some sensibility tests conducted showed that the filter/smoother may not be configured with the nominal errors used to perturb the observations (0.03 for SSH and 0.2 for temperature). In this case the analyzed fields present physical inconsistencies/unbalances that lead to worthless predictions. Therefore, we chose to use a more "calibrated" observational error (0.08 for SSH and 0.5 for temperature) which provide physically consistent fields as well as a more consistent covariance spread. Therefore, considering the decomposition given by the Eq.(2.113), it would be equivalent to chose $\alpha \approx 8$. This sets the "optimal" number of iterations to 4.

5.4 Results

5.4.1 Effect of the DA window and the covariance matrix rank

In this section we analyse the effects of the DA window (i.e. the smoother lag) and the covariance matrix rank on the estimation of initial conditions for the first assimilation cycle. All experiments started from the climatological mean and covariance, which means that the system, at this phase, has much to "learn" from the observations. As observed by Cosme et al. (2010) the SEEK-smoo improves the filter estimation especially for the first few assimilation cycles.

The methods used in this section are the SEEK-smoo, the BS and the BFKF but without iterations (only one forward-backward filtering is considered). Therefore, in this case the performance of the SEEK-smoo that uses temporal correlations to obtain $p(\mathbf{x}_0|\mathbf{y}_{1:K})$ and the smoother that propagates information from K to 0 using the backward model $p(\mathbf{x}_0|\mathbf{x}_K)$ are compared. Recalling that the SEEK-smoo and the BS are equivalent only in the linear Gaussian case and considering the global analysis.

Figures (5.2), (5.3) and (5.4) show the relative error for SSH, U-velocity and tem-

perature, respectively. Except for the rank10 experiment for which the BS produces the best results, the BFKF is the one with the best performance, independently of the considered variable and rank, followed by the BS and SEEK-smoo. This is expected since the BFKF uses twice the observations. The BFKF improved in 20% the SSH estimation, 10% the velocity and 15% the temperature. The methods have almost the same behavior, i.e. a great improvement for the 2 days lag after which a more modest improvement is achieved.

The best lag varies according to the method, the rank and the variables. Considering the temperature, the SEEK-smoo for all ranks has its best lag at 18 days, which is the period of one ARGO cycle, while the BFKF at 8 days for rank greater than 10, and for the BS the best lag is at 12 days. For the SSH and the velocity the best lag is of 8 days for all methods, excepted for the SSH variable estimated by the SEEK-smoo for which the best lag is 6 days. Cosme et al. (2010) also reported 10 days as the best lag for the SSH variable using the SEEK-smoo and the same configuration used in this study, but assimilating observations extracted from a higher resolution model.

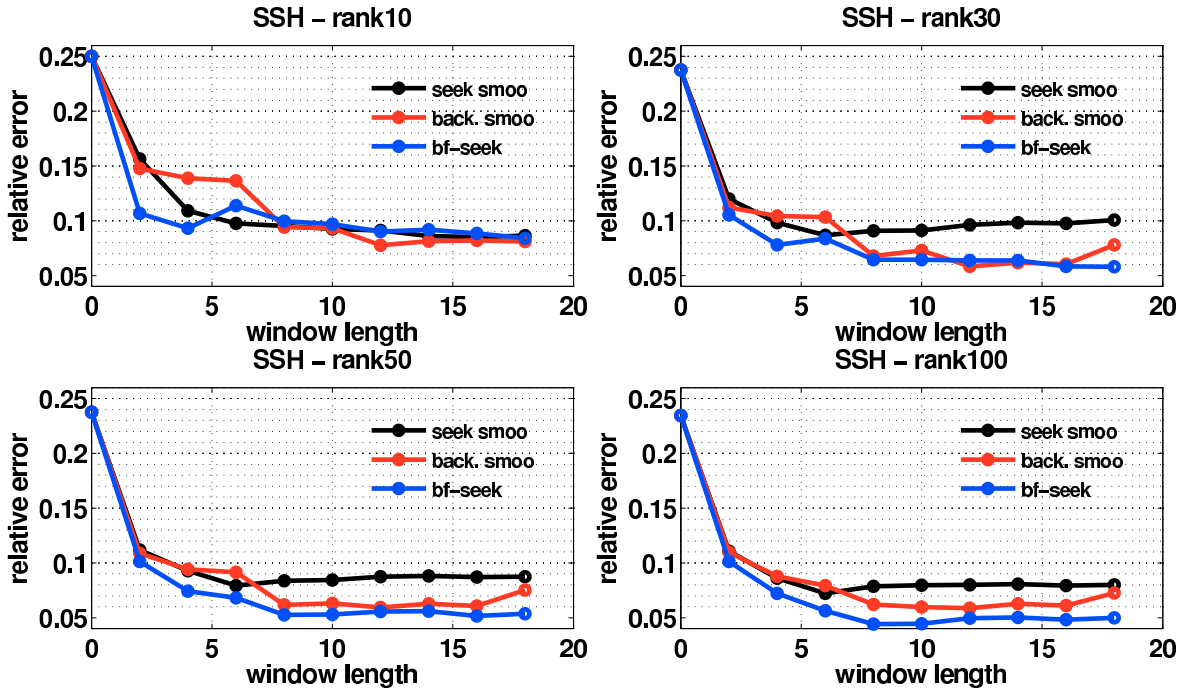


Figure 5.2: SSH relative error for different assimilation windows and covariance rank. Top left: $rank(\mathbf{S}) = 10$. Top right: $rank(\mathbf{S}) = 30$. Bottom left: $rank(\mathbf{S}) = 50$. All experiments used $\rho = 1$, i.e. no covariance inflation is considered.

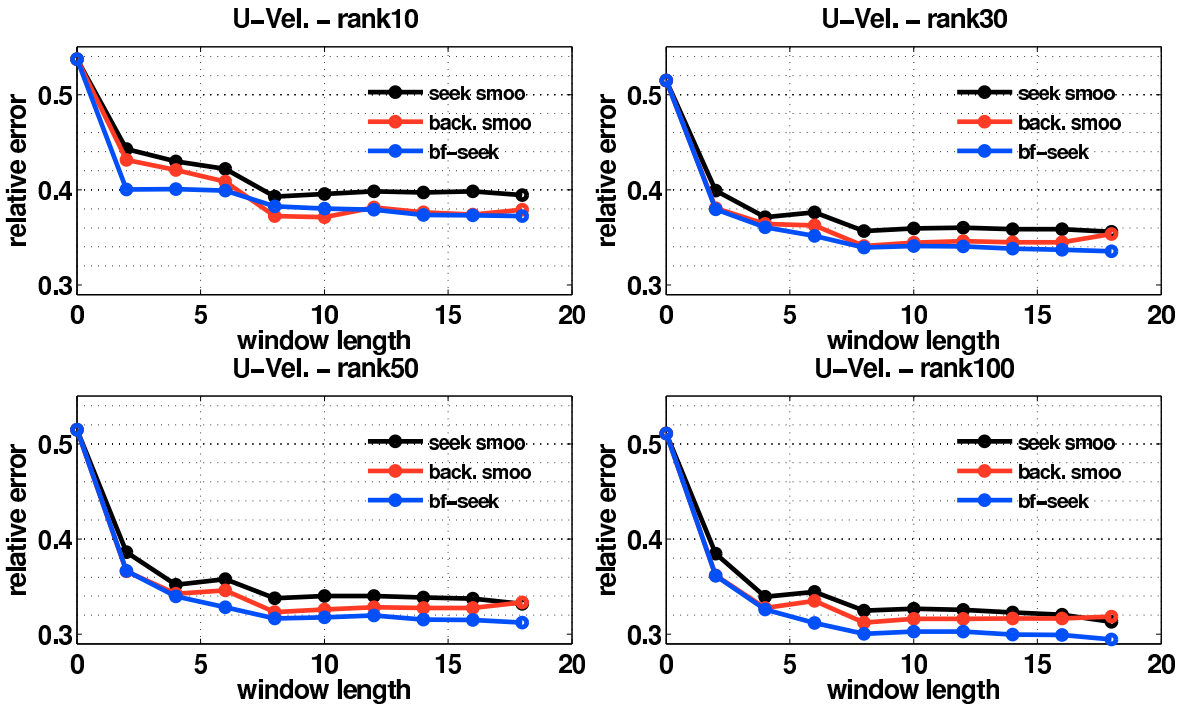


Figure 5.3: Zonal velocity relative error for different assimilation windows and covariance rank. Top left: $rank(\mathbf{S}) = 10$. Top right: $rank(\mathbf{S}) = 30$. Bottom left: $rank(\mathbf{S}) = 50$. All experiments used $\rho = 1$, i.e. no covariance inflation is considered.

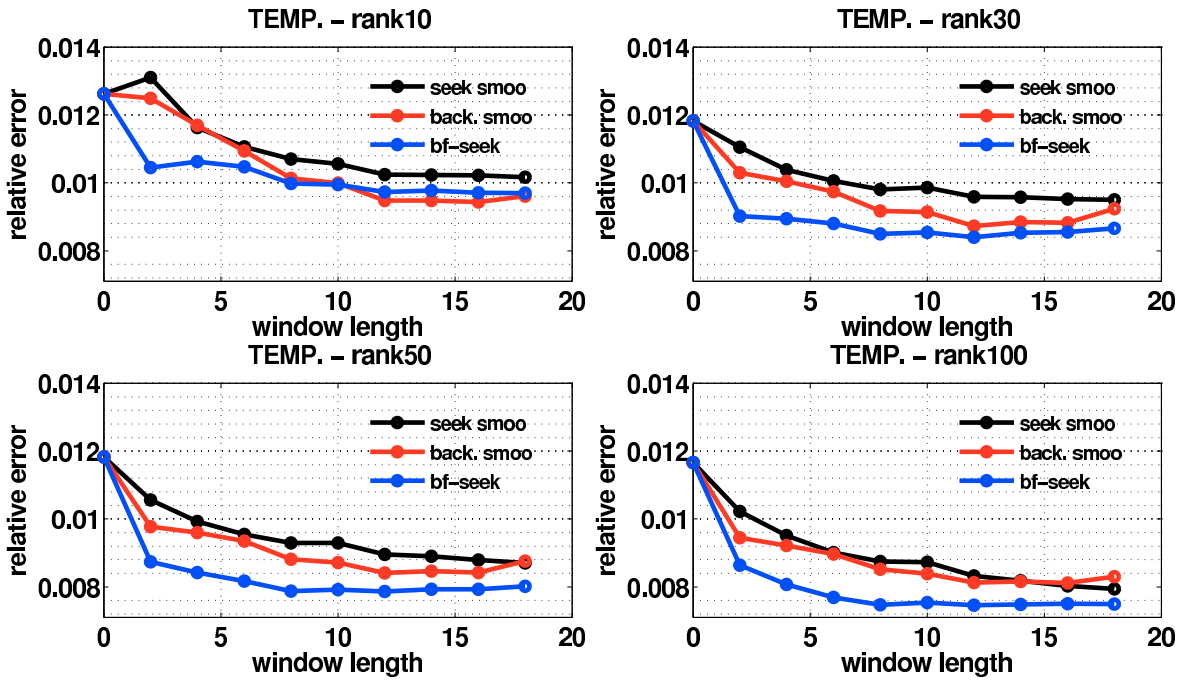


Figure 5.4: Temperature relative error for different assimilation windows and covariance rank. Top left: $rank(\mathbf{S}) = 10$. Top right: $rank(\mathbf{S}) = 30$. Bottom left: $rank(\mathbf{S}) = 50$. Bottom right: $rank(\mathbf{S}) = 100$. All experiments used $\rho = 1$, i.e. no covariance inflation is considered.

The BS solution is closer to the BFKF solution when a very reduced basis is considered, and closer to the SEEK-smoo when the rank 100 is considered. When the basis is very reduced and consequently only a small unstable subspace is used by the filter to correct the system state, the error at the end of the assimilation window still projects onto the unstable subspace. Therefore, the backward model helps to reduce these residual errors since the unstable subspace becomes stable during the backward integration, i.e. the errors are naturally damped by the backward model. This represents a great advantage for the methods based on the "Back and Forth" idea, since the backward integration is a "natural" way to correct the errors that will grow in the forecast phase at the cost of one model integration in the case of the BS algorithm.

5.4.2 Effect of iterations on the initial condition estimation and innovation statistics

Following the results of the previous section, the effects of iterations for the improvement of the initial conditions are analyzed. Two experiments are conducted: one initialized from the climatological mean and covariance, as in the last section, and another one using the mean and covariance produced by the filter after its convergence (refer to section 5.4.4 for a description of the filter results). Three iterative smoothers are tested: iBS, RIP (which is an iterative version of the SEEK-smoo), and the BFKF. Recalling iBS and RIP are equivalent in the linear Gaussian case.

For this experimental set the covariance matrix of rank 20 is considered although some results produced with the covariance matrix of rank 50 are used to discuss some aspects concerning the subspace modeled by the reduced basis. Two inflation factors are used: $\rho = 1$, which is equivalent to not considering inflation and is the same used in the previous section, and $\rho = 0.95$. The assimilation window is 10 days, which is similar to one SSH observation cycle and therefore allows correcting all the model domain.

Figures 5.5 and 5.6 show the results for the smoothers initialized from the climatological statistics and from the filter statistics, respectively. In both cases and for all experiments the iterations further decreases the filter and the smoother analysis. The BFKF produces the best results followed by the iBS and RIP. In general, improve-

ments are greater at the end of the first iteration, after which the observed variable errors almost stabilized. The non-observed variable errors, in their turn, keep decreasing reaching convergence or eventually diverging after some iterations. The BFKF and iBS produce greater improvements on the non-observed variables, relative to the improvements on the observed variables, than the RIP. This may be the consequence of improving the covariance structures as long as the iterations are performed. Furthermore, this may indicate the advantages of applying the backward model instead of using temporal covariances to update a past state.

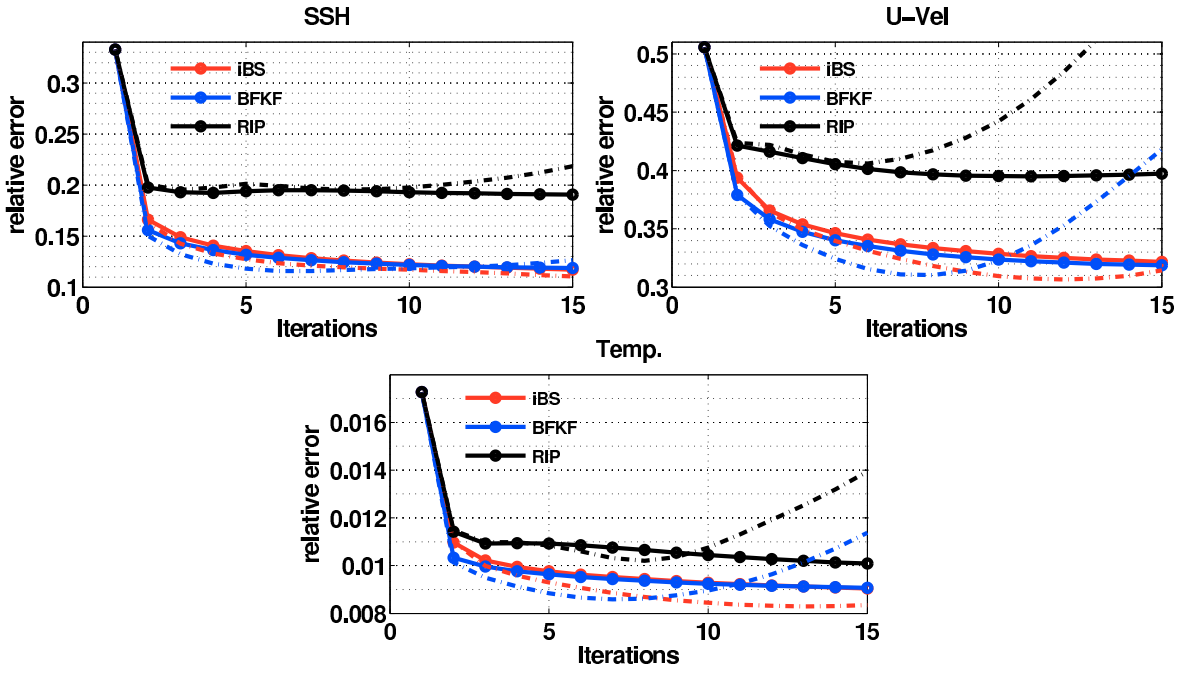


Figure 5.5: Initial condition relative error for the iterative Backward Smoother (iBS), Back and Forth Kalman Filter (BFKF) and Running in Place (RIP). All experiments were initialized from climatological mean and covariance and use $rank(\mathbf{S}) = 20$, $\mathbf{R}_{ssh} = 0.08\mathbf{I}$ and $\mathbf{R}_{temp} = 0.5\mathbf{I}$. Dots represent simulations performed using $\rho = 1$ and dashed line using $\rho = 0.95$. Top left: SSH; Top right: Zonal Velocity; Bottom: Temperature.

It seems that an optimal number of iterations exist, which varies according to the amount of information about system errors the covariance matrix represents, and the inflation factor. Too many iterations may lead to the algorithms divergence especially when inflation is considered. Nonetheless, inflation improves the BFKF and iBS estimation, leading to smaller errors in less iterations. Divergence occurs later for the

climatological case than for the filter-based initialization, indicating the importance of how representative of the real uncertainties the filter statistics are. When the filter statistics are coherent with the true system error, the iterations have a reduced effect.

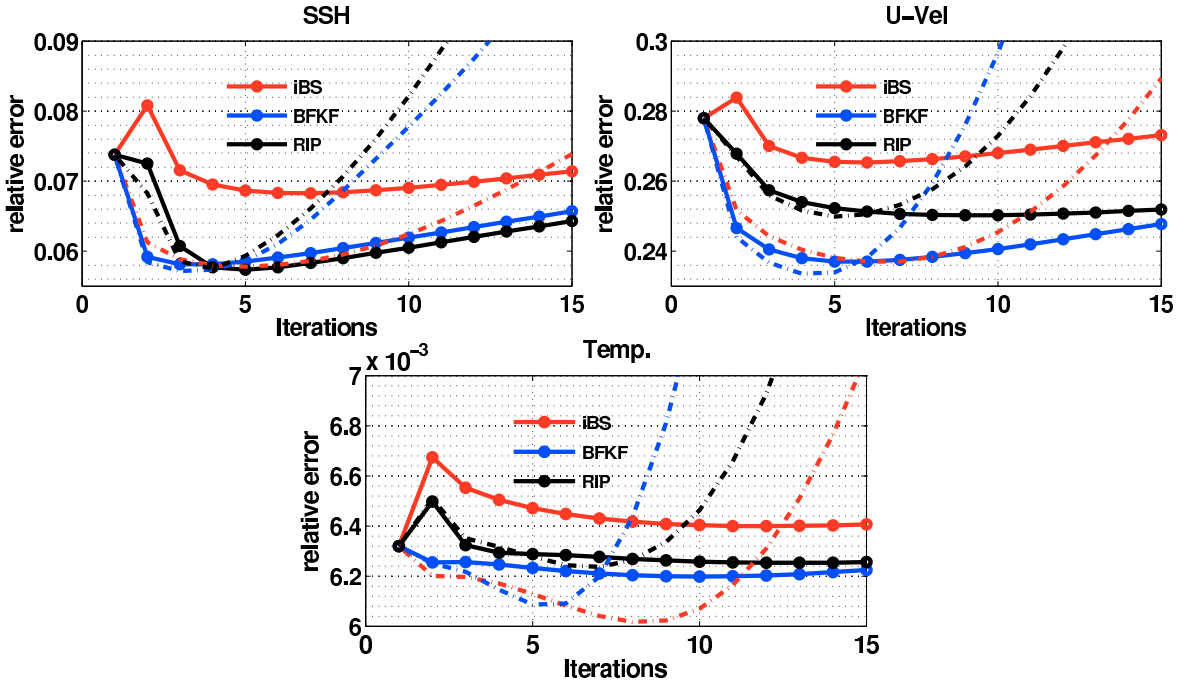


Figure 5.6: Initial condition relative error for the iterative Backward Smoother (iBS), Back and Forth Kalman Filter (BFKF) and Running in Place (RIP). All experiments were initialized from a mean and covariance calculated by the SEEK filter after its spin-up. $rank(\mathbf{S}) = 20$, $\mathbf{R}_{ssh} = 0.08\mathbf{I}$ and $\mathbf{R}_{temp} = 0.5\mathbf{I}$. Dots represent simulations performed using $\rho = 1$ and dashed line using $\rho = 0.95$. Top left: SSH; Top right: Zonal Velocity; Bottom panel: Temperature.

Figure 5.7 shows the velocity error minimization for the iterative Backward Smoother (iBS) initialized from a mean and covariance calculated by the SEEK filter after its spin-up. The initial condition error is mainly reduced by the first backward integration. It is due to the reduction of the error component that projects onto the unstable subspace which becomes stable in the backward integration. A similar effect may be seen in Fig.(5.13) by the decreasing rate of all reduced basis vector, calculated with respect to the energy norm. The decreasing behavior is especially important for the first iteration, explaining why inflation is so important for the iBS especially for the first iteration. At the iteration number 8, the decreasing rate is reduced by a half, and reach approximately zero as iterations are performed.

Although the perturbations have a smaller forward increasing rate, the basis spread increases in such a way that it no more represents the mean squared error. Therefore, this leads to the algorithm divergence.

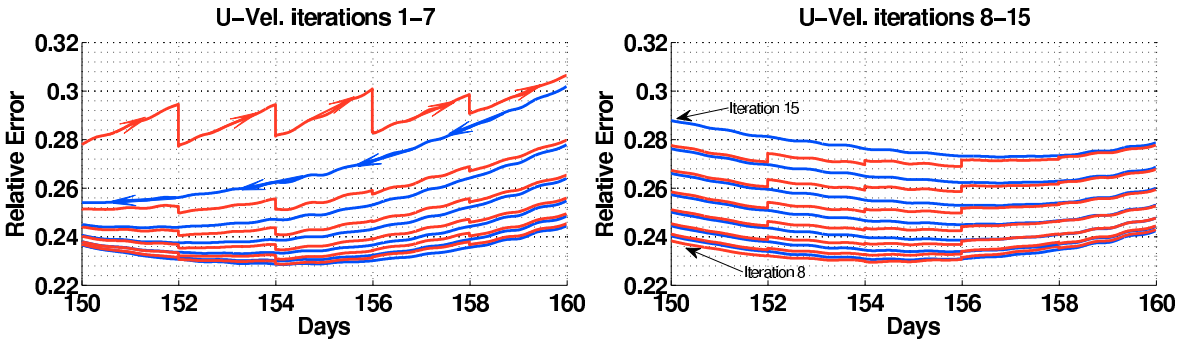


Figure 5.7: Figure shows the first seven Backward Smoother (iBS) iterations initialized from a mean and covariance calculated by the SEEK filter after its spin-up. $rank(\mathbf{S}) = 20$, $\rho = 0.95$ and $\mathbf{R}_{ssh} = 0.08\mathbf{I}$, $\mathbf{R}_{temp} = 0.5\mathbf{I}$. Left panel: iterations from 1 to 7; Right panel: iterations from 8 to 15.

Figure 5.8 is similar to Fig. 5.7 but for the rank 50 experiment. In this case the backward integration does not improve the initial condition but it does improve the final condition. The increasing initial condition error, especially for the first iterations, are due to the uncontrolled backward unstable modes. Again we refer to Fig.(5.13) to study the subspace spanned by $\mathbf{S}$. In this case, the reduced basis represents only the forward unstable subspace, hence, allowing the backward error growth.

The two solutions (Figs 5.7 and 5.8) have in common the fact that after some iterations the forward filter does not improve the solution anymore. Therefore, proceeding the iterations degrades the solution. Also, the evolution of $\mathbf{S}$ for both experiments follows the same characteristics, i.e it initially spans an unstable subspace that becomes more neutral as iterations are done. This should be one reason for the filter failure.

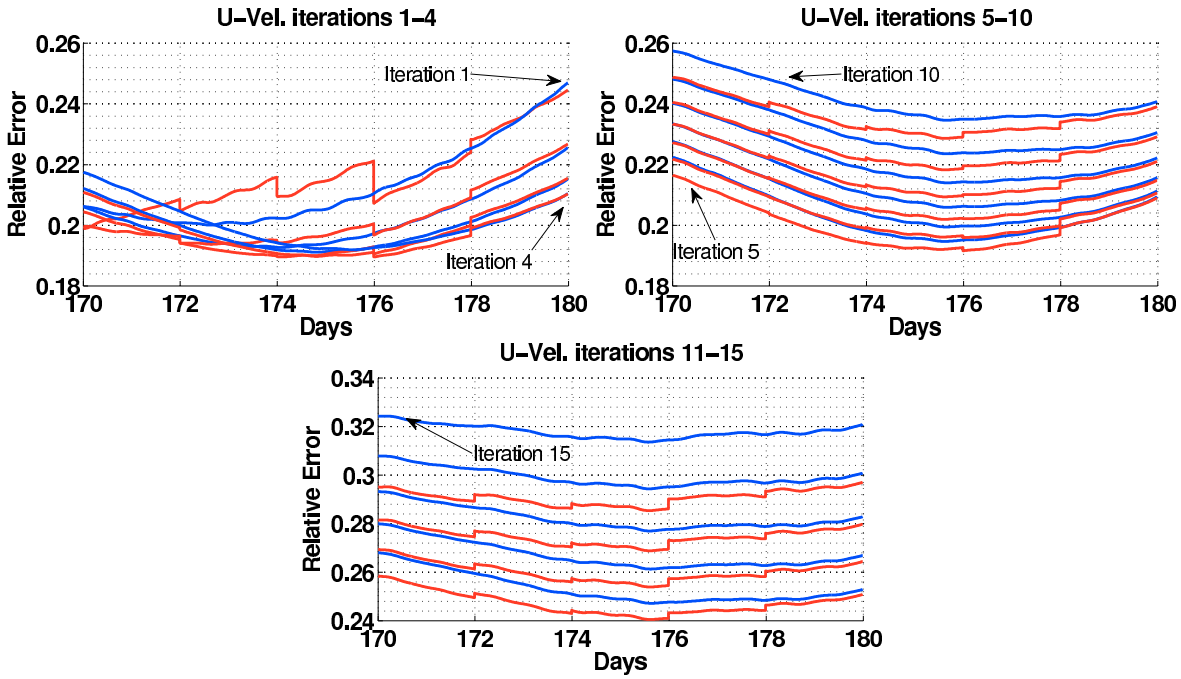


Figure 5.8: Figure shows the first seven Backward Smoother (iBS) iterations initialized from a mean and covariance calculated by the SEEK filter after its spin-up. $rank(\mathbf{S}) = 50$, $\rho = 0.95$ and $\mathbf{R}_{ssh} = 0.08\mathbf{I}$, $\mathbf{R}_{temp} = 0.5\mathbf{I}$. Top left: iterations from 1 to 4; Top right: iterations from 5 to 10; Bottom panel: iterations from 11 to 15.

To further investigate the advantages of iterating, internal consistence tests concerning the innovation statistics are performed. The innovation statistics reveals the presence of non-Gaussianity in the system, which may be a result of non-Gaussian observation or background errors or an indicative that the joint errors are non-Gaussian in the observational space (Pires et al., 2010). Indeed, the KF optimality requires Gaussian error distributions, otherwise the analysis is sub-optimal since moments of order higher than 2 are ignored.

One possible measure of Gaussianity is the negentropy, which measures the difference in entropy between a given distribution and a Gaussian distribution with the same mean and variance. The negentropy is the Kullback-Leibler divergence between the pdfs p and q_G where q_G is a Gaussian distribution. It is defined as:

$$D(p|q_G)_{KL} = \int p(x) \ln \left[\frac{p(x)}{q_G(x)} \right] dx \tag{5.5}$$

Thus, negentropy is always nonnegative and vanishes if and only if p and q_G have the

same distribution. In our case, p is the discrete innovation distribution produced by the filter and q_G is a discrete Gaussian distribution with mean and standard deviation given by the distribution of p.

Figure 5.9 present histograms and negentropy calculated with the innovation sequences available for the data assimilation window over which the methods are iterated. The negentropy values reveals that the iterations transform the innovation distribution into a more and more Gaussian-like distribution, although after fifteen iteration it is not Gaussian. The distributions become more symmetric but remain leptokurtic, i.e. they are thinner with respect to the Gaussian. The iBS results are surprising, since in this case the perturbations are integrated backward using the tangent linear model without any rescaling procedure. In this case, we would expect the basis to become non-Gaussian due to the nonlinearities acting on the assimilation window, but instead it provides the more Gaussian-like distribution.

The improvement in terms of relative errors shown in Fig.5.5 may be attributed, to some extent, to this effect of transforming the "ensemble" into a more Gaussian-like distribution. This may be a result of a reduced covariance spread and analysis increments, which helps the system to follow the nonlinear trajectory by improving the linear approximation made by the SEEK filter. This is especially important for cold initializations for which the increments are greater.

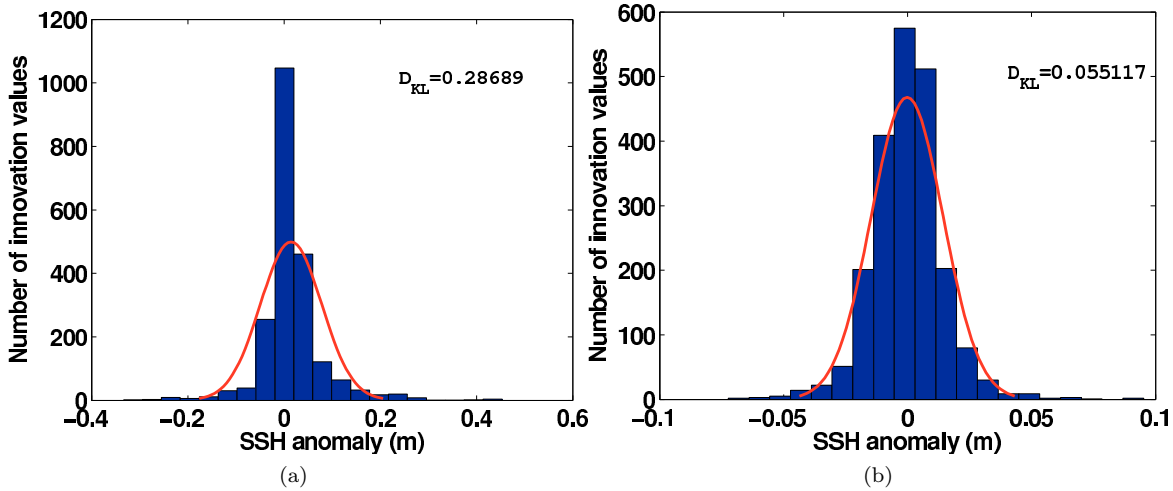
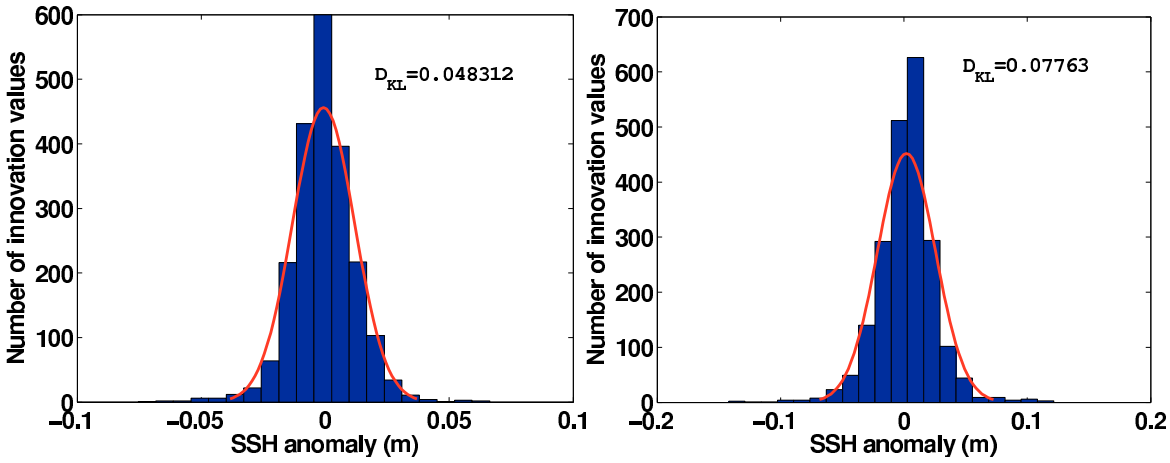


Figure 5.9



(c) (d)

Figure 5.9: Histogram and negentropy calculated from the innovation sequence available for the data assimilation window over which the methods are iterated. (**a**) filter pass. (**b**) BFKF after 15 iterations. (**c**) iBS after 15 iterations. (**d**) RIP after 15 iterations. $rank(\mathbf{S}) = 20$, $\rho = 1$ and $\mathbf{R}_{ssh} = 0.08\mathbf{I}$, $\mathbf{R}_{temp} = 0.5\mathbf{I}$.

5.4.3 Effect of iterations on the covariance matrix structure

To study the behavior of the perturbations $\mathbf{S}^i$ over the iterations i, the principal angles between two consecutive iterations $\mathbf{HS}^i$ and $\mathbf{HS}^{i+1}$ as well as the stability of the space spanned by $\mathbf{S}$ are calculated.

5.4.1. Definition (*principal angles*). *Let E be a Euclidean vector-space with inner product $<,>$ and induced norm $||.||$. Given two subspaces U and W with $dim(U) = k < dim(N) = l$ there exists a set of k angles $\theta_1, ..., \theta_k$ called principal angles. The first one is :*

$$\theta_1 = \min \left\{ \arccos \left(\frac{< u, w >}{\|u\|\|w\|} \right) | u \in U, w \in W \right\} = \angle(u_1, w_1)$$

the other principal angle and vector are then defined recursively via:

$$\theta_i = \min \left\{ \arccos \left(\frac{< u, w >}{\|u\|\|w\|} \right) | u \in U, w \in W, u \perp u_j, w \perp w_j \right.$$
$$\left. \forall j \in \{1, ..., i-1\} \right\}$$

Figure 5.10 shows the maximal and minimal principal angle for the experiments BFKF and RIP with $\rho = 1$ and $\rho = 0.95$. The spaces are quite different between the first and second iteration, which is expected since the greater improvement is due to the second iteration. The filtering phase brings most of information to the system and enrich the covariance matrix. After this, the angles converge to zero for $\rho = 1$, which means that $\mathbf{HS}^i$ and $\mathbf{HS}^{i+1}$ span the same subspace. Moreover, the covariance spread also decreases accordingly to the mean squared filter error. Therefore, the iterations converges.

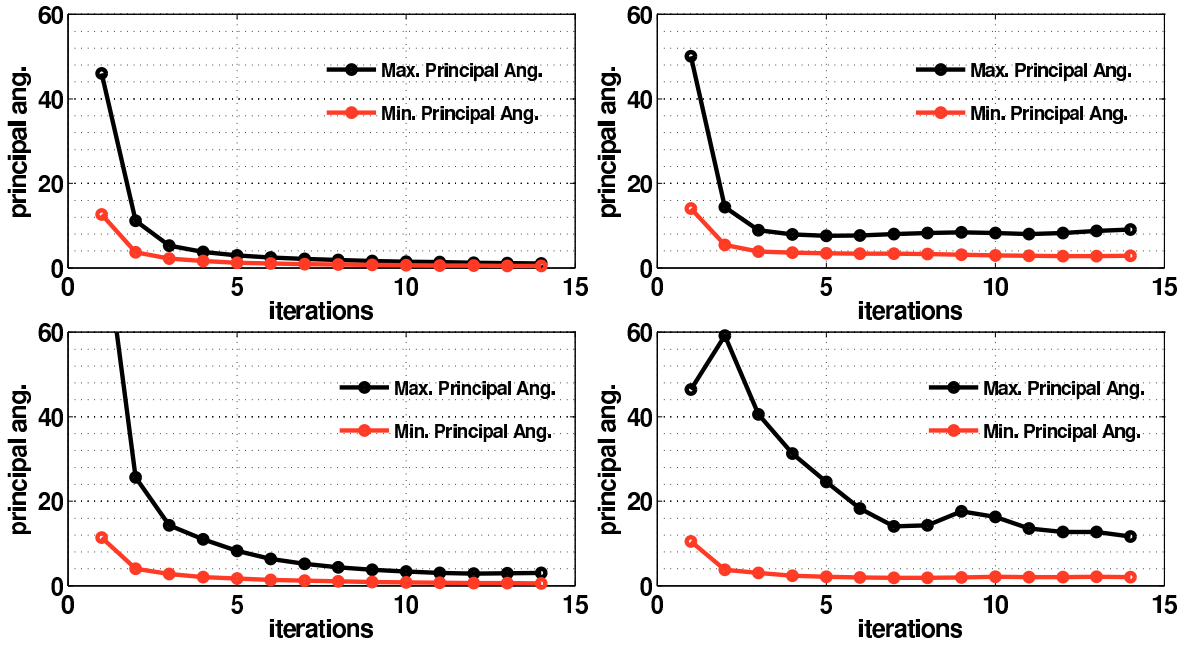


Figure 5.10: Maximal and minimal principal angles between $\mathbf{HS}^i$ and $\mathbf{HS}^{i+1}$ calculated from the climatological initialization experiment for top panels: BFKF and bottom panels: RIP. Left panels: $\rho = 1$ and right panels: $\rho = 0.95$.

For $\rho = 0.95$ the subspace spanned by $\mathbf{HS}^i$ continuously changes. In addition, the covariance spread initially decreases accordingly with the mean squared error but after five iterations, it starts to increase. Figure 5.11 shows the SSH spread for the climatological experiment and $\rho = 0.95$. Hence, it is clear that the spread takes the form of the observation network indicating an "over-inflation".

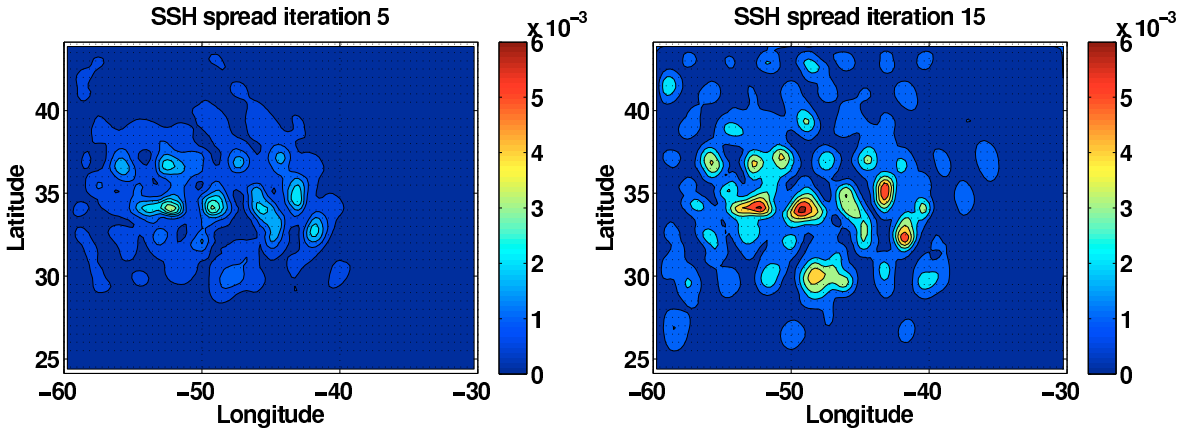


Figure 5.11: Sea Surface Height spread calculated for the iBS experiment initialized from the climatological mean and covariance and employing $\rho = 0.95$. Left panel: after 5 iterations and Right panel: after 15 iterations.

Figure 5.12 shows the principal angle for the experiments BFKF and RIP with $\rho = 0.95$ initialized from the filter statistics. In this case the estimates are not convergent and the algorithm evolution is quite similar to the case presented above for which the spread takes the form of the observational network. This explains why the first three iterations improves the estimation after which the filter diverges.

Therefore, the divergence observed in the Figs.(5.5) and (5.6) is to some extent related to an "over-inflation". In this case the covariance matrix structure keep changing as iterations are performed and the covariance spread grows non-consistently with the mean squared estimation error. The "over-inflation" in areas of sparse or variable-in-time observations may result in a unreasonably large ensemble spread in observations deficient regions. Hence a more sophisticated covariance inflation scheme such as an adaptive scheme (Miyoshi, 2010) or even the Kalman Filter formulation without the intrinsic need for inflation (Bocquet, 2011) may be useful to handle the presented inflation problem.

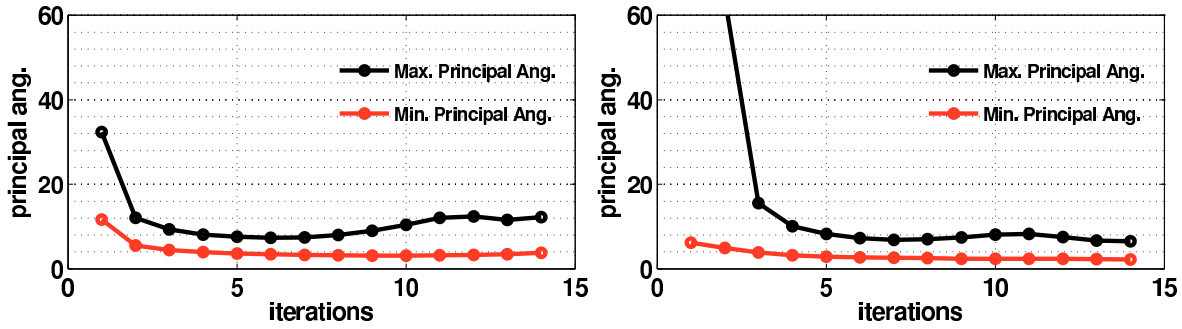


Figure 5.12: Maximal and minimal principal angles between $\mathbf{HS}^i$ and $\mathbf{HS}^{i+1}$ calculated from the filter initialization experiment for left panel: iBS and right panel: BFKF. Both using $\rho = 0.95$

Figure 5.13 shows the variation rate of the energy perturbations

$$\mathcal{E}p = \frac{1}{T}\ln\frac{\|\delta x_0\|}{\|\delta x_T\|} \tag{5.6}$$

where $\|\bullet\| = \int_\Omega [\rho(\mathbf{u}^2 + \mathbf{v}^2) + \rho g]\, \mathrm{d}\Omega$, during the backward integrations of the iBS method. For the climatological experiment, the reduced basis is represented by unstable (negative values) and stable (positive) direction. This means that the basis does not initially capture only the most unstable modes. Consequently, the algorithm corrects both unstable and stable modes. For the experiment initialized from the filter statistics, independently of the rank, the reduced basis represents only the unstable subspace. Accordingly, the backward integration does not correct the forward stable modes which are unstable backwards. The results also confirm that 20 independent unstable directions only marginally approximate the unstable subspace.

Therefore the success of iterating depends on how big the forecast errors are, on where on the attractor the system is, and on what subspace the reduced basis spans. When the system passes by unstable regimes and the reduced basis does not capture all the unstable subspace the backward integration may further improve the filter/smoother performance. However, when the system is stable but the reduced basis does not represent some of the stable directions the backward model may introduce errors in the forward stable modes which is not desired. The same situation may occur when the system is unstable but the errors projecting onto the unstable modes are already tiny regarding the observational noise. In this case, the backward integration does not improve the estimation precisely due to the stable modes and to the backward model inaccuracy.

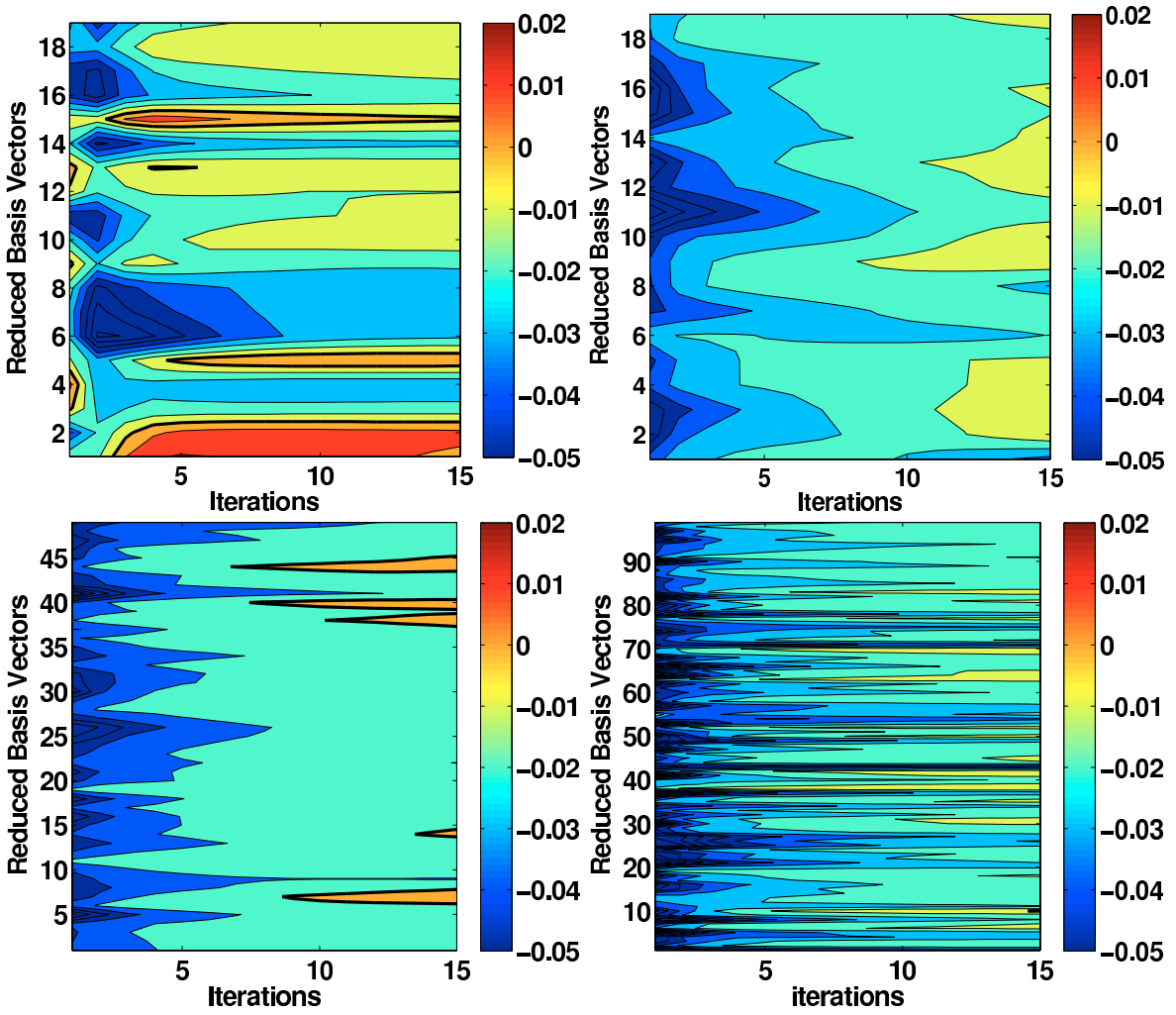


Figure 5.13: Variation rate of the energy perturbations $\mathcal{E}p = \frac{1}{T}\ln\frac{\|\delta x_0\|}{\|\delta x_T\|}$ where $\|\bullet\| = \int_\Omega \left[\rho(\mathbf{u}^2+\mathbf{v}^2)+\rho g\right] \mathrm{d}\Omega$, during the backward integrations of the iBS method. Top left: $rank(\mathbf{S}) = 20$ initialized from the climatological statistics; Top right: $rank(\mathbf{S}) = 20$ initialized from the filter statistics; Bottom left: $rank(\mathbf{S}) = 50$ initialized from the filter statistics; Bottom right: $rank(\mathbf{S}) = 100$ initialized from the filter statistics. For all experiments $\mathbf{R}_{ssh} = 0.08\mathbf{I}$, $\mathbf{R}_{temp} = 0.5\mathbf{I}$ and $\rho = 1$.

5.4.4 Stability of the assimilation system

It has been shown that the first iteration is the one providing most informations to the DA system. Therefore, this section concerns the study of the system behavior when several assimilation cycles are considered. Two covariance ranks are considered: 20 and 50, both using $\rho = 0.95$.

Figures 5.14 and 5.15 show the relative error for the experiments using rank equal

to 20 and 50 respectively, and the table 5.1 summarizes the results in terms of mean relative error. The BFKF is the most efficient in reducing the errors at the beginning of the experiment, i.e. considering a cold restart. The performance is relatively better for the rank 20 experiment confirming that the backward information propagation is especially important when very reduced basis are considered. The simplified BFKF version, namely the Back and Forth Fixed basis Kalman Filter (BFFKF), produces results that are comparable with those one produced by the BFKF for the rank20 experiment, but for the rank50, although at the beginning the performances are comparable, after 50 days the algorithm starts to diverge.

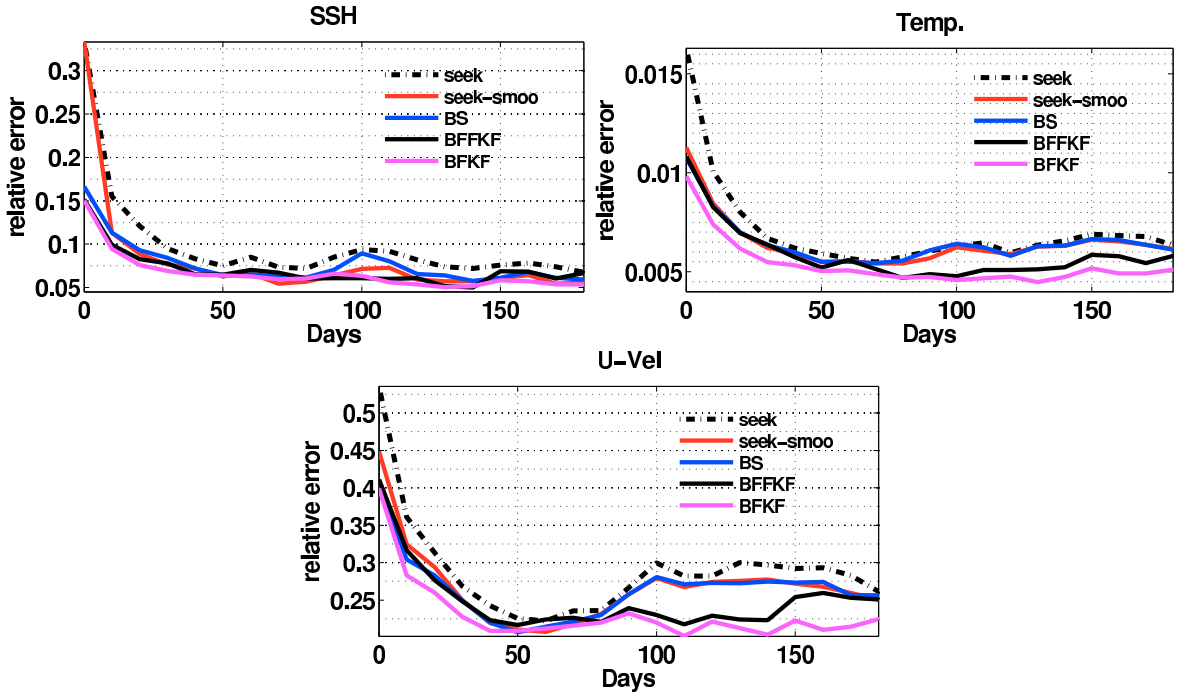


Figure 5.14: Initial condition relative error for the SEEK filter and smoother, the Backward Smoother (BS), the Back and Forth Kalman Filter (BFKF) and the BFKF but using a fixed basis (BFFKF) within the assimilation window. Top left: SSH error. Top right: zonal velocity error. Bottom: temperature error. All experiments were initialized from climatological mean and covariance and the filter parameters are $rank(\mathbf{S}) = 20$, $\rho = 0.95$ and $\mathbf{R}_{ssh} = 0.08\mathbf{I}$, $\mathbf{R}_{temp} = 0.5\mathbf{I}$.

The BS performance is almost similar to the SEEK-smoo in the case of rank20 experiments but for the rank50 case the SEEK-smoo provides slightly better results. While for the rank20 experiment the BFKF exhibits the best performance for all variables

remarkably after 100 days, for the rank50 it is initially better but after 50 days the SEEK-smoo and the BS are more accurate.

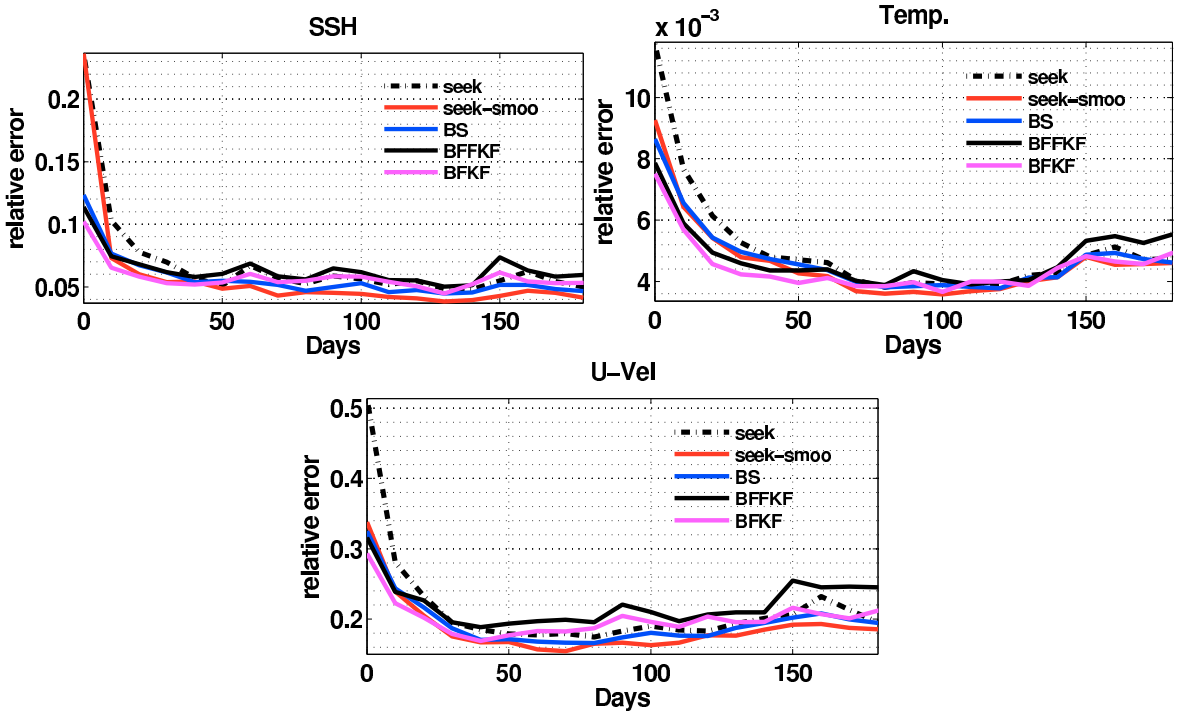


Figure 5.15: Initial condition relative error for the SEEK filter and smoother, the Backward Smoother (BS), the Back and Forth Kalman Filter (BFKF) and the BFKF but using a fixed basis (BFFKF) within the assimilation window. Top left: SSH error. Top right: zonal velocity error. Bottom: temperature error. All experiments were initialized from climatological mean and covariance and the filter parameters are $rank(\mathbf{S}) = 50$, $\rho = 0.95$ and $\mathbf{R}_{ssh} = 0.08\mathbf{I}$, $\mathbf{R}_{temp} = 0.5\mathbf{I}$.

Therefore, the results confirm that the iterations are especially important for "cold" initialization and when the system undergoes drastic changes. In the latter case, the system may diverge or present large oscillations (e.g. Fig.5.14) if some error components project onto an unstable subspace not spanned by the columns of $\mathbf{S}$. In this case the backward integration may provide a method to reduce these error components. However, when the system already models the most unstable subspace, i.e. the rank of $\mathbf{S}$ is large enough to span the most unstable subspace and the filter statistics are coherent, the iterations may not further improve the initial condition estimation. Nevertheless, as showed in Fig.5.8, the final conditions (i.e. the forecasts) may be improved by the iterations.

Methods	rank20			rank50		
	SSH	T	U	SSH	T	U
SEEK	0.1759	0.0103	0.3696	0.1218	0.0077	0.3054
	0.0797	0.0062	0.2683	0.0552	0.0043	0.1916
SEEK-smo	0.1859	0.0082	0.3295	0.1385	0.0065	0.2404
	0.0634	0.0060	0.2525	0.0460	0.0041	0.1730
BS	0.1144	0.0081	0.3115	0.0824	0.0064	0.2451
	0.0697	0.0061	0.2549	0.0501	0.0043	0.1816
BFFKF	0.1028	0.0081	0.3132	0.0795	0.0058	0.2442
	0.0621	0.0053	0.2314	0.0596	0.0044	0.2124
BFKF	0.0975	0.0072	0.2923	0.0697	0.0055	0.2243
	0.0588	0.0049	0.2145	0.0544	0.0041	0.1934

Table 5.1: Summary of the mean relative initial condition errors obtained by the algorithms: SEEK filter, SEEK smoother, Backward Smoother (BS), Back and Forth Kalman Filter with Fixed basis (BFFKF) and the Back and Forth Kalman Filter (BFKF). Two reduced basis approximations are exploited one with rank 20 and another with rank 50. For each method the top lines are the mean relative errors calculated from day 1 to day 50, and the bottom lines the mean relative errors calculated from day 51 to 180.

5.4.5 The forecast performance

Motivated by the fact that the backward integration may not improve the initial condition but the forecast (as shown in Fig.5.8), in this section the initial conditions identified in the last section are used to initialize 180 days forecasts. This allows to test if a better initial condition, in terms of global error, generates a more accurate forecast. Only the rank 50 experiment is analyzed, since for this case the benefits of using the backward model is less evident for the long range experiment. Only the identified initial conditions produced by the SEEK-smoo and BS are used to initialize the forecasts. To avoid the spin-up effects in our analysis, which may overestimate the BS performance, the studied period covers the days 50 to 180.

The results confirm what is presented in Fig.5.8. Although less accurate initial conditions are produced by the BS (see Fig. 5.16), the forecasts are about $4-10\%$ more accurate especially in the range of 10 to 40 days. At the end of the assimilation window (10 days, represented by the black vertical line in the figure) the BS forecast is already more accurate than the SEEK-smoo forecasts. The gain in accuracy for the BS at the end of the assimilation window is similar to the improvement of the filter solution due to the smoother restrospective analysis, which means that the improvement is small but significant.

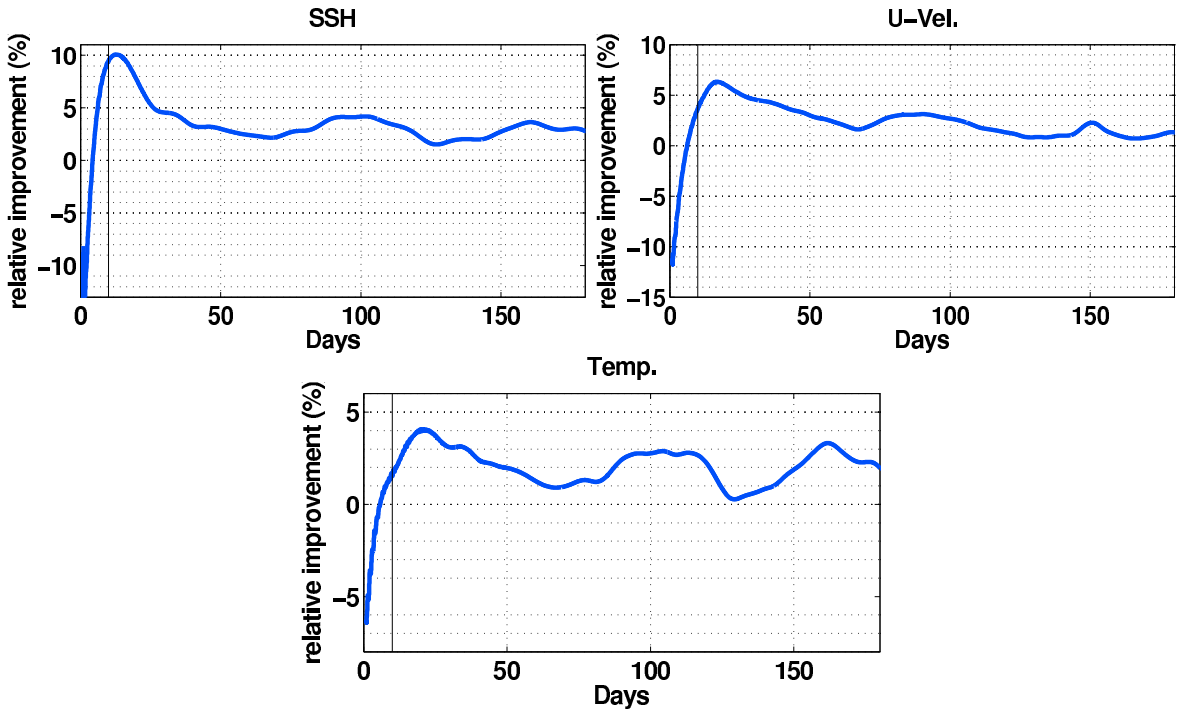


Figure 5.16: Mean forecast improvement calculated from the forecasts initialized from the BS and SEEK-smoo. Top left: SSH relative improvement. Top right: zonal velocity relative improvement. Bottom: temperature relative improvement. Positive (negative) values indicates the BS (SEEK-smoo) has better forecast performance.

It may be concluded that the backward integration corrects the errors projecting onto the "fast" growing modes. Auroux (2009) also reported the same result for a comparison between the BFN and a 4Dvar implementation, i.e. the 4Dvar initial condition is more accurate but the BFN produces more accurate forecasts. Additionally, this may be a result of a more dynamically coherent fields since the dynamical model is used to back-propagate future information.

5.5 Conclusions

In this chapter the SEEK filter and three smoothers based on the SEEK filter were implemented, namely the SEEK-smoo, the Backward Smoother (BS) and the Backward Kalman Filter (BKF). The smoothers were implemented along with their iterative versions: iterative SEEK-smoo (RIP), iterative BS (iBS) and Back and Forth Kalman

Filter (BFKF). All smoothers improved the filter estimation and all iterative smoothers further improved the smoother estimation.

Among all the presented algorithms the BFKF is the one that produced the best estimates. The BS, which is equivalent to the SEEK-smoo in the linear Gaussian case, produced results that are comparable to the results produced by the SEEK-smoo. Regarding their iterative version, the iBS performance is very close to the BFKF which is better than the SEEK-smoo. The reason for this difference is the use of the backward model by the iBS and BFKF instead of temporal correlations. This improves the results in two folds: first by reducing the forward errors that project onto the unstable subspace and second by improving the variables cross-covariance, given a more dynamical character to it.

It has been shown that the iterations play an important role for cold restarts and when the system passes by unstable periods. They are responsible for transforming the innovation statistics into a more Gaussian-like distribution, thus favoring the KF-based algorithms. This is due to the reduced increments and perturbations which helps the extended filter to follow the nonlinear trajectory.

For the methods using the backward model, the iterations are responsible for reducing the errors that project onto the unstable subspace. In addition, the iterations change the subspace spanned by the covariance matrix. It was shown that when the initial covariance matrix represents the unstable subspace the iterations leads the basis vectors to span a more neutral subspace. When this is coherent with the true error, i.e. the basis spans the space containing the forecast error, the system is stable. However as it was observed, the filter statistics may not follow the real system error leading to the filter divergence.

An observed problem of iterations is related to an uncontrollable perturbation growth in the non-observed areas or variables. This problem is further accentuated when a constant inflation parameter is considered. As discussed, this may be avoided by considering adaptive techniques or methods that avoid the need for inflation.

Finally, although the iterative methods here studied are able to mitigate sampling problems concerning the covariance inflation and localization techniques, these topics were not formally investigated. Therefore, we think these iterative procedures may be

further explored to take advantage of adaptive procedures which may also provide a natural stopping criterion.

Chapter 6

Conclusions

6.1 Conclusions (French)

Cette thèse a développé et mis en œuvre des algorithmes itératifs d'assimilation de données pour un modèle d'océan aux équations primitives, et les a comparé avec d'autres méthodes d'assimilation de données bien établies telles que le 4Dvar et le filtre/lisseur SEEK. Le modèle numérique utilisé est le modèle NEMO. Il a été configuré pour simuler la circulation typique subtropicale en double gyre à une résolution de mésoéchelle. Les nouveaux algorithmes itératifs proposés, de façon similaire au Nudging direct et rétrograde, sont tous basés sur une séquence d'intégrations alternées du modèle direct et rétrograde. Dans la pratique, les méthodes diffèrent quant à la façon dont les poids donnés aux observations et au modèle sont calculés.

6.1.1 Principaux résultats

Le résultat principal de cette thèse est que les itérations du modèle direct et rétrograde sont un outil puissant pour être utilisé dans l'assimilation des données d'observations océaniques. Nous avons vu que, indépendamment du gain utilisé pour pondérer la combinaison entre le modèle et les observations, les itérations ont toujours amélioré la solution "originale".

En outre, il a été montré que les nouvelles méthodes itératives produisent de meilleurs résultats que les méthodes existantes. Le DBFN est capable de produire des champs d'analyse de la même qualité que le 4DVAR mais avec une utilisation 3 fois moindre de

la puissance de calcul, et l'iBS et le BFKF sont supérieurs au RIP (Running in Place), lorsque les trois méthodes utilisent une matrice de covariance de même rang.

L'utilisation du modèle rétrograde est préférable à l'utilisation de covariances temporelles mal connues qui sont sujettes à des erreurs d'échantillonnage. Elle permet d'éviter les problèmes liés à l'accouplement de localisation spatiale et temporelle, puisque c'est le modèle lui-même qui propage les informations d'observation vers le passé. L'intérêt de l'utilisation du modèle rétrograde est encore amplifié lorsque le modèle est non linéaire. Dans ce cas, les corrélations linéaires, utilisées par les lisseurs de Kalman, ne peuvent pas déterminer complètement un état passé conditionné à un état futur.

Un autre aspect important concernant l'utilisation du modèle rétrograde, est qu'il permet une utilisation implicite du sous-espace instable du modèle direct pour corriger les conditions initiales. Cela peut assurer la stabilité du système d'assimilation en particulier lorsque la base réduite utilisée est connue pour avoir un rang beaucoup plus petit que le rang du véritable sous-espace instable+neutre.

6.1.2 Questions ouvertes

Quelques points importants ont été soulevés mais pas entièrement résolus dans cette thèse. Le plus important concerne la précision du modèle rétrograde et son spectre de Lyapunov. Dans ce cas, le point ouvert implique la recherche d'un compromis entre la précision du modèle rétrograde, ce qui implique un spectre plus instable pour lequel des gains du Nudging plus importants sont nécessaires, et un modèle plus diffusif, ce qui augmente la stabilité du modèle, au risque de produire des champs lissés irréalistes en l'absence d'observations.

Une autre question, qui est dans une certaine mesure liée au problème présenté ci-dessus, est la dégradation des conditions initiales identifiées lorsque la condition initiale de l'intégration rétrograde contient des erreurs qui se projettent surtout sur le sous-espace stable du modèle direct et que ce sous-espace n'est pas modélisé par la base réduite. Dans ce cas, si l'on est surtout intéressé par les conditions initiales pour des fins de réanalyse par exemple, une façon possible de continuer à améliorer les conditions initiales est de reproduire certains modes "instables en rétrograde/stables en direct", ce

qui empêchera les erreurs qui se projettent sur ces modes de croître. Ceci implique une augmentation des coûts numériques, et donc ses bénéfices doivent être soigneusement examinés .

Une faiblesse détectée à propos de notre mise en œuvre de l'iBS et BFKF est la divergence observée lorsque l'inflation a été considérée, même si nous avons vu que l'inflation fournit de petites erreurs en moins d'itérations. Par conséquent, nous suggérons l'utilisation de filtres qui n'ont pas besoin d'inflation ou de schémas d'inflation adaptatifs. Cela permettrait de conserver les statistiques de filtrage cohérentes avec les statistiques de l'erreur vraie et donc d'utiliser aussi efficacement que possible une observation particulière, étant donné les limites de chaque méthode. En outre, les statistiques utilisées par l' approche adaptative peuvent être utilisées pour obtenir un critère d'arrêt plus efficace, puisque le critère de convergence que nous avons utilisé dans le chapitre 4 n'est pas capable de détecter la divergence signalée dans le chapitre 5.

6.1.3 Perspectives

Du point de vue théorique, la convergence de l'iBS en utilisant des filtres étendus peut être prouvée en considérant que la première estimation (ébauche) est dans un voisinage de la vraie trajectoire. Toujours du point de vue théorique, deux points pourraient être abordés: l'estimation des paramètres, du modèle et de la méthode d'assimilation, et l'inclusion des erreurs du modèle dans la formulation bayésienne de l'iBS et BFKF.

D'un point de vue plus pratique, un point à explorer consiste à profiter des itérations pour assimiler des observations de haute densité (par exemple les données radar, des images de couleur de l'océan, la salinité de surface mesurée par satellite et les mesures de température) avec un gain du type Nudging, cependant que toutes les données plus classiques et clairsemées sont assimilées en utilisant la mise à jour du filtre de Kalman par exemple. Cela permettrait une assimilation multi-échelle, puisque les données de SSH satellitaires échantillonnent principalement le premier mode barocline et les processus sous-inertiels alors que le radar et les données de couleur de l'océan peuvent capturer les mouvements super-inertiels et inertiels ainsi que la sous-mésoéchelle régissant la dynamique de la couche limite de surface .

6.1.4 Remarques finales

Cette thèse a prouvé qu'il existe un réel intérêt pour l'utilisation du DBFN et des algorithmes basés sur les itérations du modèle direct et rétrograde pour l'assimilation des observations océaniques. Bien sûr, il est nécessaire d'étudier comment ces méthodes se comportent dans un système d'assimilation opérationnel pour lequel, en général, les modèles sont biaisés et pas parfaits. Plus de développements sont nécessaires, notamment en ce qui concerne la diffusion de l'intégration rétrograde et la stabilité à long terme des filtres/lisseurs qui peuvent être traités avec des systèmes adaptatifs.

Enfin, nous croyons que les algorithmes présentés dans cette thèse méritent d'être testés en mode opérationnel comme le sont le 4Dvar et les filtres de Kalman. En effet, compte tenu de la simplicité de mise en œuvre du modèle rétrograde, les algorithmes itératifs présentés ici peuvent être facilement mis en œuvre en tirant partie des systèmes d'assimilation existants.

6.2 Conclusions

This thesis developed and implemented iterative data assimilation algorithms for a primitive equation ocean model, and compared them with other well established DA methods such as the 4Dvar and the SEEK Filter/Smoother. The numerical model used was the NEMO model. It was configured to simulate a typical subtropical double gyre circulation at an eddy permitting resolution. The new proposed iterative algorithms, similarly to the Back and Forth Nudging, are all based on a sequence of alternating forward and backward model integrations. In practice, the methods differ with respect to how the weights given to the observations and the model are calculated.

6.2.1 Main findings

The main result of this thesis is that the iterations of the forward and backward model are a powerful tool to be used in data assimilation of oceanic observations. We have seen that independently of the gain used to weight the combination of model and observations, the iterations always improved the "original" solution.

Furthermore, it was shown that the new iterative methods performed better than the existing methods. The DBFN is able to produce analysis fields of the same quality as the 4Dvar but using 3 times less computational power, and the iBS and BFKF are superior than the RIP, when all three methods use a covariance matrix of the same rank.

The use of the backward model is preferable compared to the use of poorly known temporal covariances that are subject to sampling errors. It avoids problems related to the coupling of spatial and temporal localization, since it is the model itself that propagates the observation informations backward in time. The interest in the use of the backward model is further amplified when the model is non-linear. In this case, the linear correlations, used by the Kalman-based smoothers, may not fully determine a past state conditioned on a future state.

Another important aspect concerning the use of the backward model is that it allows for an implicit use of the unstable forward model subspace to correct the initial conditions. This may provide stability for the assimilation system especially when the

used reduced basis is known to have a rank much smaller than the rank of the true unstable+neutral subspace.

6.2.2 Open questions

Some important points were raised up but not entirely solved in this thesis. The most important one concerns the accuracy of the backward model and its Lyapunov's spectrum. In this case, the open point involves the search for a compromise between the backward model accuracy, which implies a more unstable spectrum for which stronger Nudging gains are needed, and a more diffusive model, which increases the model stability at the risk of producing unrealistic smoothed fields in the absence of observations.

Another issue, which is at some extent related to the problem presented above, is the degradation of the identified initial conditions when the initial condition of the backward integration contains errors that project mostly onto the forward stable subspace and this subspace is not modeled by the reduced basis. In this case, if one is mainly interested in the initial conditions for reanalysis purposes for example, one possible way to keep improving the initial conditions is to breed some "backward unstable/forward stable" modes, which will prevent the errors projecting onto these modes to grow. This would further increase the numerical costs and hence its benefits should be carefully examined.

A detected weakness concerning our implementation of the iBS and BFKF is the divergence observed when the inflation was considered, although it was seen that inflation provides smaller errors in less iterations. Therefore, we suggest the use of inflation free filters or adaptive inflation schemes. This would permit to keep the filter statistics coherent with the true error statistics and therefore to use as efficiently as possible a particular observation, given the limits of each methodology. Moreover, the statistics used by the adaptive approach may be used to derive a more effective stopping criterion, since the convergence criterion we have been using in chapter 4 would not detect the divergence reported in the Chapter 5.

6.2.3 Perspectives

From the theoretical point of view the convergence of the iBS using extended filters may be proved considering that the first guess is in a neighborhood of the true trajectory. Still from the theoretical point of view, two points might be addressed: the estimation of parameters, from the model and from the assimilation method, and the inclusion of the model errors in the Bayesian formulation of the iBS and BFKF.

From a more practical point of view, a point to be explored is to avail the iterations to assimilate high density observations (e.g. radar data, ocean color images, satellite-based surface salinity and temperature measurements) with a Nudging-like gain while the more conventional and sparse data are assimilated using the standard Kalman Filter update for example. This would allow a multi-scale assimilation, since SSH data derived from satellite samples preferably the first baroclinic mode and sub-inertial features while the radar and ocean color data may capture super-inertial and near inertial movements as well as the submesoscale governing the dynamics of the surface boundary layer.

6.2.4 Concluding Remarks

This thesis proved that there is a real interest in the use of the DBFN and the algorithms based on the iterations of the forward and backward model for the assimilation of ocean observations. Of course, it is necessary to further investigate how these methods would perform in an operational assimilation system for which in general the models are biased and not perfect. Further development is required especially regarding the diffusion in the backward integration and the long term stability of the filters/smoothers which may be treated with adaptive schemes.

Finally, we believe that the algorithms presented in this thesis deserve to be tested in operational mode as the 4Dvar and the Kalman filters are. Indeed, considering the simplicity of implementation of the backward model, the presented iterative algorithms may be easily implemented taking advantage of the existing assimilation systems.

Bibliography

H. D. I. Abarbanel, M. Kostuk, and W. Whartenby. Data assimilation with regularized nonlinear instabilities. *Quarterly Journal of the Royal Meteorological Society*, 136(648):769–783, 2010. ISSN 1477-870X. doi: 10.1002/qj.600. URL http://dx.doi.org/10.1002/qj.600.

J. L. Anderson. A local least squares framework for ensemble filtering. *Monthly Weather Review*, (131):634–642, 2003.

R. A. Anthes. Data assimilation and initialization of hurricane prediction models. *J. Atmos. Sci.*, 31:702–719, 1974. doi: 10.1175/1520-0469(1974). URL http://dx.doi.org/10.1175/1520-0469(1974).

D. Auroux. The back and forth nudging algorithm applied to a shallow water model, comparison and hybridization with the 4D-VAR. *Int. J. Numer. Methods Fluids*, 61 (8):911–929, 2009.

D. Auroux and J. Blum. Back and forth nudging algorithm for data assimilation problems. *C. R. Acad. Sci. Paris, Ser. I*, 340:873–878, 2005.

D. Auroux and J. Blum. A nudging-based data assimilation method for oceanographic problems: the back and forth nudging (bfn) algorithm. *Nonlin. Proc. Geophys.*, 15: 305–319, 2008.

D. Auroux and M. Nodet. The back and forth nudging algorithm for data assimilation problems: theoretical results on transport equations. *ESAIM Control Optim. Calc. Var.*, 18(2):318–342, 2012.

D. Auroux, J. Blum, and M. Nodet. Diffusive back and forth nudging algorithm for data assimilation. *C. R. Acad. Sci. Paris, Ser. I*, 349:849–854, 2011.

D. Auroux, P. Bansart, and J. Blum. An evolution of the back and forth nudging for geophysical data assimilation: application to burgers equation and comparisons. *Inv. Prob. Sci. Eng.*, 21(3):399–419, 2012.

J. Ballabrera-Poy, E. Kalnay, and S-C. Yang. Data assimilation in a system with two scales—combining two initialization techniques. *Tellus A*, 61(4), 2009. ISSN 1600-0870. URL http://www.tellusa.net/index.php/tellusa/article/view/15608.

A. Barth, A. Alvera-Azcárate, K.-W. Gurgel, J. Staneva, A. Port, J.-M. Beckers, and E. V. Stanev. Ensemble perturbation smoother for optimizing tidal boundary conditions by assimilation of high-frequency radar surface currents – application to the german bight. *Ocean Science*, 6(1):161–178, 2010. doi: 10.5194/os-6-161-2010. URL http://www.ocean-sci.net/6/161/2010/.

A. F. Bennett. *Inverse Modeling of the Ocean and Atmosphere*. Cambridge, Cambridge, first edition, 2002.

A. F. Bennett and W. P. Budgell. The kalman smoother for a linear quasi-geostrophic model of ocean circulation. *Dynamics of Atmospheres and Oceans*, 13(3–4):219 – 267, 1989. ISSN 0377-0265. doi: http://dx.doi.org/10.1016/0377-0265(89)90041-9. URL http://www.sciencedirect.com/science/article/pii/0377026589900419.

K. Bergemann and S. Reich. A mollified ensemble kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 136(651):1636–1643, 2010. ISSN 1477-870X. doi: 10.1002/qj.672. URL http://dx.doi.org/10.1002/qj.672.

M. Bocquet. Ensemble kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 18(5):735–750, 2011. doi: 10.5194/npg-18-735-2011. URL http://www.nonlin-processes-geophys.net/18/735/2011/.

A. Boilley and J.-F. Mahfouf. Assimilation of low-level wind in a high-resolution mesoscale model using the back and forth nudging algorithm. *Tellus A*, 64(0), 2012. ISSN 1600-0870. URL http://www.tellusa.net/index.php/tellusa/article/view/18697.

J. M. Brankart, E. Cosme, C. E. Testut, P. Brasseur, and J. Verron. Efficient local error parameterizations for square root or ensemble kalman filters: Application to

a basin-scale ocean turbulent flow. *Mon. Wea. Rev.*, 139(2):0027–0644, 2010. doi: 10.1175/2010MWR3310.1.

P. Brasseur and J. Verron. The seek filter method for data assimilation in oceanography: a synthesis. *Ocean Dynamics*, 56(5-6):650–661, 2006. ISSN 1616-7341. doi: 10.1007/s10236-006-0080-3. URL `http://dx.doi.org/10.1007/s10236-006-0080-3`.

O. Cappé, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer, 2005. ISBN 9780387402642. URL `http://books.google.fr/books?id=-3_A3_l1yssC`.

A. Carrassi, A. Trevisan, and F. Uboldi. Adaptive observations and assimilation in the unstable subspace by breeding on the data-assimilation system. *Tellus A*, 59(1), 2007. ISSN 1600-0870. URL `http://www.tellusa.net/index.php/tellusa/article/view/14855`.

K. I. Chang, M. Ghil, K. Ide, and C. C. A. Lai. Transition to aperiodic variability in a wind-driven double-gyre circulationmodel. *J. Phys. Oceanography*, 31:1260–1286, 2001.

E. P. Chassignet and P. R. Gent. The influence of boundary conditions on midlatitude jet separation in ocean numerical models. *J. Phys. Oceanography*, 21:1290–1299, 1991.

X. Chen, C. Liu, K. O'Driscoll, B. Mayer, J. Su, and T. Pohlmann. On the nudging terms at open boundaries in regional ocean models. *Ocean Modelling*, 66 (0):14 – 25, 2013. ISSN 1463-5003. doi: 10.1016/j.ocemod.2013.02.006. URL `http://www.sciencedirect.com/science/article/pii/S1463500313000401`.

M. Clifford, C. Horton, J. Schmitz, and L. H. Kantha. An oceanographic nowcast/forecast system for the red sea. *Journal of Geophysical Research: Oceans*, 102(C11):25101–25122, 1997. ISSN 2156-2202. doi: 10.1029/97JC01919. URL `http://dx.doi.org/10.1029/97JC01919`.

S. E Cohn, A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich. Assessing the effects of data selection with the dao physical-space statistical analysis system. *Mon. Wea. Rev*, 1998.

M. Cooper and K. Haines. Altimetric assimilation with water property conservation. *Journal of Geophysical Research: Oceans*, 101(C1):1059–1077, 1996. ISSN 2156-2202. doi: 10.1029/95JC02902. URL http://dx.doi.org/10.1029/95JC02902.

E. Cosme, J.-M Brankart, J. Verron, P. Brasseur, and M. Krysta. Implementation of a reduced-rank, square-root smoother for ocean data assimilation. *Ocean Modelling*, 33:87–100, 2010.

P. Courtier, J. N. Thepaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Q. J. R. Meteorol. Soc.*, 123: 1367–1387, 1994.

A. Duchez. *Contrôle du Courant Nord Méditerranéen dans le golfe du Lion: une approche par simulation du système d'observation*. PhD thesis, Université du Sud Toulon Var, 2011.

G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994. ISSN 2156-2202. doi: 10.1029/94JC00572. URL http://dx.doi.org/10.1029/94JC00572.

G. Evensen. The ensemble kalman filter: Theoretical formulation and pratical implementation. *Ocean Dynamics*, 53:343–367, 2003.

G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2009. ISBN 9783642037115. URL http://books.google.fr/books?id=2_zaTb_O1AkC.

N. M. Faber. A closer look at the bias–variance trade-off in multivariate calibration. *Journal of Chemometrics*, 13(2):185–192, 1999. ISSN 1099-128X. doi: 10.1002/(SICI)1099-128X(199903/04)13:2¡185::AID-CEM538¿3.0.CO;2-N. URL http://dx.doi.org/10.1002/(SICI)1099-128X(199903/04)13:2<185::AID-CEM538>3.0.CO;2-N.

D. Fraser and J. Potter. The optimum linear smoother as a combination of two optimum linear filters. *Automatic Control, IEEE Transactions on*, 14(4):387–390, 1969. ISSN 0018-9286. doi: 10.1109/TAC.1969.1099196.

J. S. Frederiksen and T. J. O'Kane. Entropy, closures and subgrid modeling. *Entropy*, 10(4):635–683, 2008. ISSN 1099-4300. doi: 10.3390/e10040635. URL `http://www.mdpi.com/1099-4300/10/4/635`.

J. S. Frederiksen, T. J. O'Kane, and M. J. Zidikheri. Stochastic subgrid parameterizations for atmospheric and oceanic flows. *Physica Scripta*, 85(6):068202, 2012. URL `http://stacks.iop.org/1402-4896/85/i=6/a=068202`.

N. Freychet, E. Cosme, P. Brasseur, J.-M. Brankart, and E. Kpemlie. Obstacles and benefits of the implementation of a reduced-rank smoother with a high resolution model of the tropical atlantic ocean. *Ocean Science*, 8(5):797–811, 2012. doi: 10.5194/os-8-797-2012. URL `http://www.ocean-sci.net/8/797/2012/`.

M. Gavart and P. De Mey. Isopycnal eofs in the azores current region: A statistical tool fordynamical analysis and data assimilation. *J. Phys. Oceanogr.*, 27(10):2146–2157, 1997.

A. Gelb. *Applied optimal estimation.* MIT Press, 1974.

S. J. Greybush, E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt. Balance and ensemble kalman filter localization techniques. *Mon. Wea. Rev.*, 2010.

K. Haines, P. Malanotte-Rizzoli, R. E. Young, and W. R. Holland. A comparison of two methods for the assimilation of altimeter data into a shallow-water model. *Dynamics of Atmospheres and Oceans*, 17(2–3):89 – 133, 1993. ISSN 0377-0265. doi: 10.1016/0377-0265(93)90014-X. URL `http://www.sciencedirect.com/science/article/pii/037702659390014X`.

F. E. Hirata, P. J. Webster, and V. E. Toma. Distinct manifestations of austral summer tropical intraseasonal oscillations. *Geophysical Research Letters*, 40(12):3337–3341, 2013. ISSN 1944-8007. doi: 10.1002/grl.50632. URL `http://dx.doi.org/10.1002/grl.50632`.

P. L. Houtekamer and H. L. Mitchell. A sequential ensemble kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, 129:123–137, 2001.

B. R. Hunt, E. J. Kostelich, and I. Szunyogh. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D: Nonlinear Phenomena*,

230(1–2):112 – 126, 2007. ISSN 0167-2789. doi: 10.1016/j.physd.2006.11.008. URL http://www.sciencedirect.com/science/article/pii/S0167278906004647. ¡ce:title¿Data Assimilation¡/ce:title¿.

A. Isidori. *Nonlinear Control Systems*. Number vol. 1 in Communications and Control Engineering. Springer, 1995. ISBN 9783540199168. URL http://books.google.fr/books?id=fPGzHK_pto4C.

Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, April 1970. ISBN 0123815509. URL http://www.worldcat.org/isbn/0123815509.

R. E. Kalman. A new approach to linear filter and prediction problems. *J. Basic.Eng.*, 82:35–45, 1960.

E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2003. ISBN 9780521796293. URL http://books.google.fr/books?id=Uqc7zC7NULMC.

E. Kalnay and S-C. Yang. Accelerating the spin-up of ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 136(651):1644–1651, 2010. ISSN 1477-870X. doi: 10.1002/qj.652. URL http://dx.doi.org/10.1002/qj.652.

E. Kalnay, S Ki Park, ZX Pu, and J. Gao. Application of the quasi-inverse method to data assimilation. *Month. Weather Rev.*, 128:864–875, 2000.

R. N. Keshavamurthy and M. Sankar Rao. *The Physics of Monsoons*. South Asia Books, 1992.

P. D. Killworth, C. Dieterich, C. Le Provost, A. Oschlies, and J. Willebrand. Assimilation of altimetric data and mean sea surface height into an eddy-permitting model of the north atlantic. *Progress in Oceanography*, 48(2–3): 313 – 335, 2001. ISSN 0079-6611. doi: 10.1016/S0079-6611(01)00009-X. URL http://www.sciencedirect.com/science/article/pii/S007966110100009X. ¡ce:title¿Dynamics of the North Atlantic Circulation: Simulation and Assimilation with High-Resolution Models (DYNAMO)¡/ce:title¿.

V. Kitsios, J.S. Frederiksen, and M.J. Zidikheri. Scaling laws for parameterisations of subgrid eddy–eddy interactions in simulations of

oceanic circulations. *Ocean Modelling*, 68(0):88 – 105, 2013. ISSN 1463-5003. doi: http://dx.doi.org/10.1016/j.ocemod.2013.05.001. URL http://www.sciencedirect.com/science/article/pii/S1463500313000772.

M. Krysta, Blayo E., E. Cosme, and J. Verron. A consistent hybrid variational-smoothing data assimilation method: Application to a simple shallow-water model of the turbulent mid-latitude ocean. *Month. Weath. Rev.*, 139:3333–3347, 2011.

F. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations. *Tellus*, 38A:97–110, 1986.

J. Le Sommer, F. d'Ovidio, and G. Madec. Parameterization of sub-grid stirring in eddy resolving ocean models. part 1: Theory and diagnostics. *Ocean Modelling*, 39(1–2):154 – 169, 2011. ISSN 1463-5003. doi: http://dx.doi.org/10.1016/j.ocemod.2011.03.007. URL http://www.sciencedirect.com/science/article/pii/S1463500311000655. ¡ce:title¿Modelling and Understanding the Ocean Mesoscale and Submesoscale¡/ce:title¿.

B. Legras and R. Vautard. A guide to lyapunov vectors. In *Predictability vol I, ECMWF proceedings*, 1996.

L. Lei, D. Stauffe, S. E. Haupt, and G. Young. A hybrid nudging-ensemble kalman filter approach to data assimilation. part i: application in the lorenz system. *Tellus A*, 64(0), 2012. ISSN 1600-0870. URL http://www.tellusa.net/index.php/tellusa/article/view/18484.

Y. Leredde, J.-L. Devenon, and I. Dekeyser. Turbulent viscosity optimized by data assimilation. *Annales Geophysicae*, 17(11):1463–1477, 1999. doi: 10.1007/s00585-999-1463-9. URL http://www.ann-geophys.net/17/1463/1999/.

M. M. Levy, P. Klein, A.-M. Tréguier, D. Iovino, G. Madec, S. Masson, and S. K. Takahashi. Modifications of gyre circulation by sub-mesoscale physics. *Ocean Modelling*, 34(1–2):1 – 15, 2010. ISSN 1463-5003. doi: 10.1016/j.ocemod.2010.04.001. URL http://www.sciencedirect.com/science/article/pii/S1463500310000582.

J. K. Lewis, I. Shulman, and A. F. Blumberg. Assimilation of doppler radar current data into numerical ocean models. *Continental Shelf Research*, 18(5): 541 – 559, 1998. ISSN 0278-4343. doi: 10.1016/S0278-4343(98)00006-5. URL http://www.sciencedirect.com/science/article/pii/S0278434398000065.

H. Li, M. Kanamitsu, and S.-Y. Hong. California reanalysis downscaling at 10 km using an ocean-atmosphere coupled regional model system. *Journal of Geophysical Research*, in press, 2012.

Z. Li and I. M. Navon. Optimality of variational data assimilation and its relationship with the kalman filter and smoother. *Quarterly Journal of the Royal Meteorological Society*, 127(572):661–683, 2001. ISSN 1477-870X. doi: 10.1002/qj.49712757220. URL http://dx.doi.org/10.1002/qj.49712757220.

N. Lingala, N. Sri Namachchivaya, N. Perkowski, and H.C. Yeong. Optimal nudging in particle filters. *Procedia* {*IUTAM*}, 6(0):18 – 30, 2013. ISSN 2210-9838. doi: 10.1016/j.piutam.2013.01.002. URL http://www.sciencedirect.com/science/article/pii/S2210983813000035. ¡ce:title¿IUTAM Symposium on Multiscale Problems in Stochastic Mechanics 2012¡/ce:title¿.

J. L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations.* Springer-Verlag, Berlin, Federal Replubic of Germany, first edition, 1971.

G. E. Liston and R. A. Pielke. A climate version of the regional atmospheric modeling system. *Theor. Appl. Climatol.*, 2000.

D. G.. Luenberger. Observers for multivariable systems. *IEEE Transactions on Automatic Control*, AC-11, 1966.

X. Luo and I. Hoteit. Ensemble kalman filtering with residual nudging. *Tellus A*, 64(0), 2012. ISSN 1600-0870. URL http://www.tellusa.net/index.php/tellusa/article/view/17130.

X. Luo and I. Hoteit. Efficient particle filtering through residual nudging. *Quarterly Journal of the Royal Meteorological Society*, pages n/a–n/a, 2013. ISSN 1477-870X. doi: 10.1002/qj.2152. URL http://dx.doi.org/10.1002/qj.2152.

P. Lynch. The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, 227(7):3431 – 3444, 2008. ISSN 0021-9991. doi: http://dx.doi.org/10.1016/j.jcp.2007.02.034. URL http://www.sciencedirect.com/science/article/pii/S0021999107000952. ¡ce:title¿Predicting weather, climate and extreme events¡/ce:title¿.

G. Madec. *NEMO ocean engine*. Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), France, 27 edition, 2008.

P. Marchesiello, J. C. McWilliams, and A. Shchepetkin. Open boundary conditions for long-term integration of regional oceanic models. *Ocean Modelling*, 3 (1–2):1 – 20, 2001. ISSN 1463-5003. doi: 10.1016/S1463-5003(00)00013-5. URL http://www.sciencedirect.com/science/article/pii/S1463500300000135.

F. Mesinger and A. Arakawa. Numerical methods used in atmospheric models. *GARP Publication Series*, Volume I:64, 1976.

T. Miyoshi. The gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform kalman filter. *Mon. Wea. Rev.*, 2010.

K. Mogensen, M. A. Balmaseda, A. T. Weaver, Martin M., and A. Vidard. Nemovar: A variational data assimilation system for the nemo ocean model. *ECMWF Newsletter*, 2009.

A. Molcard, A. Griffa, and T. M. Ozgokmen. Lagrangian data assimilation in multilayer primitive equation ocean models. *J. Atmos. and Ocean Tech.*, 22:70–83, 2004.

B. B. Nardelli and R. Santoleri. Methods for the reconstruction of vertical profiles from surface data: Multivariate analyses, residual gem, and variable temporal signals in the north pacific ocean. *Journal of Atmospheric and Oceanic Technology*, 22(11): 1762–1781, 2005.

E. Ott, B. R.. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke. A local ensemble kalman filter for atmospheric data assimilation. *Tellus A*, 56(5):415–428, 2004. ISSN 1600-0870. doi: 10.1111/j.1600-0870.2004.00076.x. URL http://dx.doi.org/10.1111/j.1600-0870.2004.00076.x.

L. Palatella, A. Carrassi, and A. Trevisan. Lyapunov vectors and assimilation in the unstable subspace: theory and applications. *Journal of Physics A: Mathematical and Theoretical*, 46(25):254020, 2013. URL http://stacks.iop.org/1751-8121/46/i=25/a=254020.

D. T. Pham. Dimension, predictibility and reduced rank kalman filtering in data assimilation. In *State University*, 1997.

D. T. Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Weather Rev.*, 129:1494–1207, 2001.

D. T. Pham, J. Verron, and M. C. Roubaud. A singular evolutive extended kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16(3–4):323 – 340, 1998. ISSN 0924-7963. doi: http://dx.doi.org/10.1016/S0924-7963(97)00109-7. URL http://www.sciencedirect.com/science/article/pii/S0924796397001097.

C. A. Pires, O. Talagrand, and M. Bocquet. Diagnosis and impacts of non-gaussianity of innovations in data assimilation. *Physica D: Nonlinear Phenomena*, 239(17):1701–1717, 2010.

F. W. Primeau. Multiple equilibria of a double-gyre ocean model with super-slip boundary conditions. *J. Phys. Oceanography*, 28:2130–2147, 1998.

ZX Pu, E. Kalnay, J. Sela, and I. Szunyogh. Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method. *Month. Weather Rev.*, 125:2479–2503, 1997.

K. Ramdani, M. Tucsnak, and G. Weiss. Recovering the initial state of an infinite-dimensional system using observer. *Automatica*, 2010.

C. A. Reynolds and T. N Palmer. Decaying singular vectors and their impact on analysis and forecast correction. *J. Atmos. Sci.*, 55(19), 1998.

R. Rosipal and N. Kramer. Overview and recent advances in partial least squares. In C. Saunders, M. Grobelnik, S. Gunn, and J. Shawe-Taylor, editors, *Subspace, Latent Structure and Feature Selection Techniques*. Springer, 2006.

M. Silva, M. Araujo, J. Servain, P. Penven, and C. A. D. Lentini. High-resolution regional ocean dynamics simulation in the southwestern tropical atlantic. *Ocean Modelling*, 30(4):256 – 269, 2009. ISSN 1463-5003. doi: http://dx.doi.org/10.1016/j.ocemod.2009.07.002. URL http://www.sciencedirect.com/science/article/pii/S1463500309001413.

Dan Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, 1 edition, August 2006. ISBN 0471708585. URL http://www.worldcat.org/isbn/0471708585.

W. C. Skamarock. Evaluating mesoscale nwp models using kinetic energy spectra. *Mon. Wea. Rev.*, 2004.

D. Stauffer and J.-W. Bao. Optimal determination of nudging coefficients using the adjoint equations. *Tellus A*, 45(5), 1993. ISSN 1600-0870. URL http://www.tellusa.net/index.php/tellusa/article/view/14899.

E. C. Stephen, N.S. Sivakumaran, and R. Todling. A fixed-lag kalman smoother for retrospective data assimilation. *Mon. Wea. Rev*, 122:2838–2867, 1994.

M. Tenenhaus. *La régression PLS : Théorie et Pratique.* éditions Technip, Paris, France, first edition, 1998.

K. R. Thompson, D. G. Wright, Y. Lu, and E. Demirov. A simple method for reducing seasonal bias and drift in eddy resolving ocean models. *Ocean Modelling*, 13(2):109 – 125, 2006. ISSN 1463-5003. doi: 10.1016/j.ocemod.2005.11.003. URL http://www.sciencedirect.com/science/article/pii/S1463500305000910.

Z. Toth and E. Kalnay. Ensemble Forecasting at NMC: The Generation of Perturbations. *Bull Amer Met Soc*, 74(12):2317, 1993. doi: 10.1175/1520-0477(1993)074%253C2317:EFANTG%253E2.0.CO;2. URL http://dx.doi.org/10.1175/1520-0477(1993)074%253C2317:EFANTG%253E2.0.CO;2.

A. Trevisan and L. Palatella. On the kalman filter error covariance collapse into the unstable subspace. *Nonlinear Processes in Geophysics*, 18(2):243–250, 2011. doi: 10.5194/npg-18-243-2011. URL http://www.nonlin-processes-geophys.net/18/243/2011/.

E. Trélat. *Contrôle optimal : théorie & applications.* Vuibert, Collection Mathématiques Concrètes, https://www.ljll.math.upmc.fr/ trelat/fichiers/livreopt.pdf, electronic version only edition, 2013.

P. J. van Leeuwen. Particle Filtering in Geophysical Systems. *Mon. Wea. Rev.*, 137(12):4089–4114, December 2009. doi: 10.1175/2009mwr2835.1. URL http://dx.doi.org/10.1175/2009mwr2835.1.

J. Verron. Nudging satellite altimeter data into quasi-geostrophic ocean models. *Journal of Geophysical Research: Oceans*, 97(C5):7479–7491, 1992. ISSN 2156-2202. doi: 10.1029/92JC00200. URL http://dx.doi.org/10.1029/92JC00200.

P. A. Vidard, F.-X. Le Dimet, and A. Piacentini. Determination of optimal nudging coefficients. *Tellus A*, 55(1):1–15, 2003. ISSN 1600-0870. doi: 10.1034/j.1600-0870.2003.201317.x. URL http://dx.doi.org/10.1034/j.1600-0870.2003.201317.x.

A. von Engeln and J. Teixeira. A planetary boundary layer height climatology derived from ecmwf reanalysis data. *J. Climate*, 2013.

A. J. Wallcraft, A. B. Kara, and H. E. Hurlburt. Convergence of laplacian diffusion versus resolution of an ocean model. *Geophysical Research Letters*, 32(7):n/a–n/a, 2005. ISSN 1944-8007. doi: 10.1029/2005GL022514. URL http://dx.doi.org/10.1029/2005GL022514.

C. Wang, C. Deser, J.-Y. Yu, P. DiNezio, and A. Clement. *El Nino and Southern Oscillation (ENSO): A Review. Coral Reefs of the Eastern Pacific.* 2012.

K. Wang, J. Debernard, A. K. Sperrevik, E. Isachsen, and T. Lavergne. A combined optimal interpolation and nudging scheme to assimilate osisaf sea-ice concentration into roms. *Annals of Glaciology*, 54(62):8–12, 2013a.

S. Wang, M. Xue, A.D. Schenkman, and J. Min. An iterative ensemble square root filter and tests with simulated radar data for storm-scale data assimilation. *Quarterly Journal of the Royal Meteorological Society*, pages n/a–n/a, 2013b. ISSN 1477-870X. doi: 10.1002/qj.2077. URL http://dx.doi.org/10.1002/qj.2077.

A. T. Weaver, C. Deltel, E. Machu, S. Ricci, and N. Daget. A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc*, 131:3605–3625, 2005.

J. A. Wegelin. A survey of partial least squares (pls) methods, with emphasis on the two-block case. Technical report, University of Washington, Department of Statistics, 2000.

H. Wold. Soft modelling by latent variables: the non-linear iterative partial least squares (nipals) approach. *Perspectives in probability and statistics*, pages 117–142, 1975.

C. Wunsch. The vertical partition of oceanic horizontal kinetic energy. *J. Phys. Oceanogr.*, 27(8):1770–1794, 1997.

S.-C. Yang, E. Kalnay, and B. Hunt. Handling nonlinearity in an ensemble kalman filter: Experiments with the three-variable lorenz model. *Mon. Wea. Rev*, pages 2628–2646, 2012.

J. Zavala-Garay, J. L. Wilkin, and H. G. Arango. Predictability of mesoscale variability in the east australian current given strong-constraint data assimilation. *J. Phys. Oceanogr.*, 2012.

X. Zou, I. M. Navon, and F. X. Ledimet. An Optimal Nudging Data Assimilation Scheme Using Parameter Estimation. *Quarterly Journal of the Royal Meteorological Society*, 118(508):1163–1186, 1992. ISSN 1477-870X. doi: 10.1002/qj.49711850808. URL `http://dx.doi.org/10.1002/qj.49711850808`.