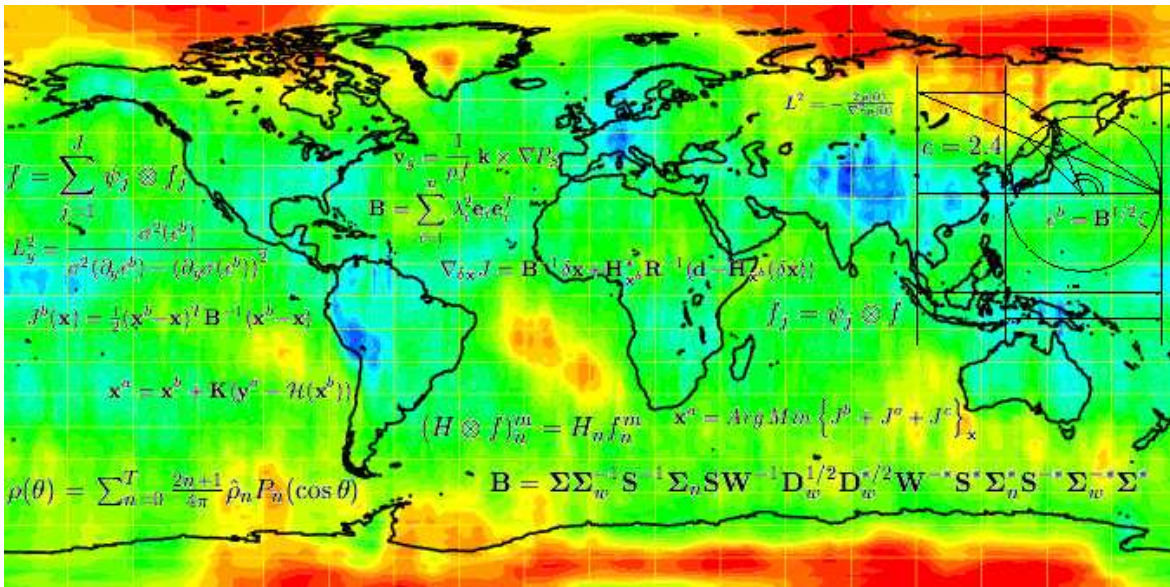


Rapport de fin d'études n° 944

Etude d'une formulation ondelette du terme de rappel à l'ébauche pour le 4D-Var ARPEGE

Olivier Pannekoucke



*Stage d'approfondissement
 réalisé au Groupe de Modélisation pour l'Assimilation et la Prévision (CNRM/GMAP),
 sous la direction de M. Loïk Berre et M. Gérald Desroziers,
 du 5 février au 13 septembre 2004.*

*Promotion d'élèves Ingénieurs de l'ENM 2001/2004,
 option Mécanique des Fluides Numérique (ENSEEIH) et
 DEA Dynamique Des Fluides (ENSEEIH-ENSAE-ENSICA-UPS)*

Planisphère reconstitué de la longueur de portée pour la température à 500hPa, et traité suivant la formulation ondelette 1D, d'après les expressions théoriques utilisées pour cette construction et la mise au point de la méthode.

Résumé

En météorologie opérationnelle, il est nécessaire de connaître l'état de l'atmosphère à un instant donné pour en déduire ses états ultérieurs. Cette étape, appelée assimilation de données, est effectuée à l'aide d'observations. Elle est capitale, avec une influence directe sur la qualité des prévisions. Le processus d'assimilation aboutit à la création d'une représentation de l'atmosphère appelée analyse, qui doit être la plus proche possible de l'état réel de l'atmosphère. Le nombre d'observations disponibles à l'heure actuelle est de l'ordre de $\mathcal{O}(10^6)$, ce qui est insuffisant pour déterminer de manière univoque l'ensemble des degrés de liberté du modèle qui est de l'ordre de $\mathcal{O}(10^7)$. Par conséquent, une prévision réalisée pour l'instant considéré est ajoutée et utilisée comme prédicteur. Le problème, ainsi fermé, est soluble de manière univoque (pour une méthode d'assimilation donnée).

La difficulté majeure vient alors de la nécessité de spécifier une matrice de covariance des erreurs de prévisions \mathbf{B} . Le rôle de cette matrice est de filtrer et de propager spatialement l'information observée. Cette matrice est trop grande pour être représentée directement ($\mathcal{O}(10^{14})$ éléments) et elle est donc plutôt modélisée sous la forme d'une cascade d'opérateurs relativement simples.

Dans la formulation actuelle, un modèle simple de cette matrice est implémenté, basé sur une formulation homogène. Ce modèle est imparfait car il ne représente pas les variations géographiques de certaines statistiques telles que celles de la longueur de portée (distance caractérisant le lien statistique entre deux points de l'atmosphère). L'objectif de ce stage est de tester une nouvelle formulation de cette matrice, en se basant sur la représentation par ondelettes sphériques. Pour simplifier l'approche et diminuer le coût numérique des expériences, une formulation ondelette sur le cercle, équivalente à celle sur la sphère, a été implémentée et étudiée.

Cette formulation simplifiée a permis de mieux comprendre la nature de cette nouvelle méthode qui combine à la fois des aspects de localisation spatiale et spectrale. L'approche ondelette a été validée en s'appuyant sur un cas analytique (donnant un cadre expérimental parfaitement contrôlé) et sur des données réelles (pression de surface et température à 500hPa). Cette méthodologie associe les perspectives théoriques et pratiques du problème. Après avoir montré la capacité des ondelettes à rendre compte des variations de longueur de portée, des étapes supplémentaires ont été introduites pour mettre au point la formulation. L'efficacité de cette modélisation a été testée dans des expériences d'assimilation de données de type 1D-Var (3D-Var pour un problème 1D), en réalisant des comparaisons avec la formulation homogène. Il apparaît que ce nouveau modèle est plus performant que le cas homogène et offre des perspectives intéressantes de filtrage. D'autre part, cette modélisation permettant une caractérisation locale de l'information, elle peut être utilisée pour rendre la matrice \mathbf{B} dépendante de l'écoulement.

Mots clés : assimilation de données, matrice \mathbf{B} , ondelettes sphériques, formulation homogène, formulation ondelette.

Remerciements

Je tiens à remercier mes encadrants Loïk Berre et Gérard Desroziers de l'équipe CNRM/GMAP (Groupe de Modélisation pour l'Assimilation et la Prévion), pour avoir proposé ce sujet stimulant et m'avoir laissé libre dans la manière de le traiter, me faisant confiance, tout en me soutenant. Ils ont su me faire partager leurs connaissances et leur goût pour la discipline.

Un grand merci à Bernard Chapnik (CNRM/GMAP) pour ses conseils avisés et son soutien.

Je tiens à remercier Anthony Weaver (CERFACS/Global Change) pour toutes les conversations très enrichissantes, échangées au cours de ces deux dernières années.

Table des matières

1	Introduction	2
2	Météorologie opérationnelle et assimilation de données	3
2.1	La prévision numérique opérationnelle sur le globe	3
2.2	Introduction à l'assimilation de données	4
2.2.1	Equations de l'analyse	4
2.2.2	Ce qu'est la matrice \mathbf{B}	5
2.2.3	Filtre de Kalman (KF), filtre de Kalman étendu (EKF)	6
2.2.4	Filtre de Kalman d'Ensemble (EnKF), méthode d'estimation de \mathbf{B}	6
2.2.5	Variance, corrélation et longueur de portée	8
2.2.6	Filtrage et propagation de l'innovation	9
2.3	Commentaire et position du problème	11
3	B dans l'atmosphère et sa modélisation	12
3.1	Physique des statistiques de l'atmosphère	12
3.2	Modélisations diagonales en points de grille et/ou spectrale	14
3.3	Modélisation hétérogène	17
3.3.1	Ondelettes sphériques	18
3.3.2	Formulation ondelette de \mathbf{B}	19
3.3.3	Problématique d'étude de la formulation	19
4	Résultats et optimisation de la modélisation ondelette	21
4.1	Formulation ondelette sur le cercle	21
4.2	Etapas de construction et normalisations introduites	22
4.2.1	Normalisation point de grille	22
4.2.2	Normalisation ondelettes	22
4.2.3	Normalisation spectrale	25
4.2.4	Impact des normalisations pour la pression de surface, validation analytique	25
4.3	Planisphère des portées méridiennes locales pour la pression de surface	28
4.4	Comparaison des modélisations dans un schéma 1D-Var	28
4.4.1	Protocole expérimental	28
4.4.2	Résultats statistiques	30
4.4.3	Conclusion	32
5	Conclusion	34
	Bibliographie	34

Chapitre 1

Introduction

Les phénomènes météorologiques extrêmes qui ont touché la France ces dernières années (en particulier celui très marquant des tempêtes de 1999), les coûts humains, matériels et économiques qu'ils engendrent conduisent les météorologistes à développer des modèles de prévision numérique pour l'atmosphère de plus en plus complexes et précis. L'objectif est de répondre au mieux à la mission de Météo-France et à l'attente de l'OMM (Organisation Mondiale Météorologique).

Pour relever le défi, il ne suffit pas d'avoir un bon modèle de prévision, mais il est nécessaire d'organiser au mieux l'ensemble des phases opérationnelles pour la prévision. En particulier, un modèle de prévision numérique nécessite la détermination de conditions initiales pour en déduire les états ultérieurs de l'atmosphère. Cette phase est l'assimilation de données. L'atmosphère étant très sensible aux conditions initiales : pour réaliser une prévision déterministe, il est nécessaire d'amorcer le cycle de prévision avec un état le plus proche de l'état vrai. La qualité de la prévision dépend fortement de cet état initial. Le problème étant sous-déterminé (pas assez d'observations pour le résoudre), on utilise une prévision en tant que prédicteur : l'ébauche.

A ce jour, l'un des points clés des algorithmes d'assimilation de données est de spécifier correctement la matrice de covariance d'erreur de prévision \mathbf{B} . C'est un enjeu actuel important et complexe à réaliser. En effet la dimension de cette matrice exclut toute représentation directe et nécessite par conséquent une modélisation. De plus, cette matrice n'est connue que par des statistiques et non de manière exacte. La recherche est très dynamique dans ce domaine où se côtoient plus que jamais physique, mathématique et informatique. Certains tentent des voies ensemblistes basées sur des statistiques de réalisation aléatoire (Evensen 1994). D'autres envisagent une modélisation basée sur l'équation de diffusion (Weaver et Courtier 2001).

Durant ce stage, une nouvelle représentation de la matrice de covariance des erreurs de prévision a été étudiée. Cette modélisation est basée sur les ondelettes sphériques radiales dont la mise en place mathématique est très récente (Freeden et Windheuser 1996, Freedon et Schreiner 1998). Cette formulation a été introduite par Fisher (2003) dans le schéma d'assimilation 4D-Var du CEPMMT. L'objectif du stage est de tester cette modélisation ondelette pour \mathbf{B} , c'est-à-dire (a) comprendre sa formulation théorique, (b) son comportement pratique et (c) l'optimiser. Pour cela, il est important d'illustrer le cadre général dans lequel ce type de besoin apparaît. Ainsi dans le chapitre 2, l'état actuel de la modélisation atmosphérique est présenté, ainsi que les contraintes liées à l'opérationnel, pour aboutir à ce qu'est l'assimilation de données et aux méthodes qui lui sont associées. Au chapitre 3, une construction pas à pas de modèle pour \mathbf{B} est présentée, ainsi que sa physique, amenant à des représentations de plus en plus raffinées, pour arriver à la modélisation ondelettes, satisfaisant ainsi l'objectif (a). Enfin au chapitre 4, la formulation ondelette est testée dans un cadre simplifié, ainsi que les améliorations apportées à la modélisation actuelle, répondant à (b) et (c).

Chapitre 2

Météorologie opérationnelle et assimilation de données

Dans ce chapitre, les bases de la météorologie opérationnelle sont présentées, ainsi que le modèle de prévision numérique mis en oeuvre à Météo-France. Le principe de l'assimilation de données est ensuite brièvement introduit.

2.1 La prévision numérique opérationnelle sur le globe

L'atmosphère est un fluide modélisé par un milieu continu. L'air est considéré comme étant un gaz parfait, mélange de deux gaz : l'air sec et la vapeur d'eau. Ce gaz est compressible et gouverné par les équations de Navier-Stokes. Dans un modèle de prévision global, l'atmosphère est décomposée en niveaux verticaux, chaque niveau étant une sphère recouvrant le niveau précédent.

Un niveau donné de la grille du modèle est alors une sphère. La sphère est un domaine périodique pour lequel les fonctions de carré intégrable admettent deux représentations classiques équivalentes :

1. La valeur de la fonction sur une grille de collocation.
2. La représentation spectrale de la fonction dans la base des harmoniques sphériques.

Les champs sont alors représentés en harmoniques sphériques. Ce sont des fonctions $Y_n^m(\lambda, \phi) = P_n^m(\sin \phi) e^{im\lambda}$, où λ est la longitude et ϕ la latitude ; les P_n^m désignent les polynômes de Legendre de première espèce. On appelle nombre d'onde total l'entier n . En notant N_T l'indice de troncature, un signal f est représenté par $f(\lambda, \phi) = \sum_{n=0}^{N_T} \sum_{m=-n}^n f_n^m Y_n^m(\lambda, \phi)$ avec (f_n^m) le spectre. Le passage d'une représentation à l'autre s'effectue par une transformation espace spectral - point de grille, équivalente à la transformation de Fourier dans le cas du cercle.

L'avantage de l'espace spectral est que les harmoniques sphériques sont fonctions propres de l'opérateur Laplacien. Ainsi, la résolution des équations de Navier-Stokes est plus simple et plus précise. Le modèle français ARPEGE¹ est un modèle pseudo-spectral dont la grille de collocation est la grille des points de Gauss sur la sphère.

Les champs physiques du modèle sont : le tourbillon relatif ζ (c'est la vorticit  verticale du vecteur vent), la divergence du vent η , la temp rature T , la pression de surface P_S et l'humidit  sp cifique² q . L' tat de l'atmosph re est ainsi mod lis  par la valeur de ces cinq param tres physiques en chaque point de la grille de Gauss et pour chaque niveau du mod le. On repr sente ainsi l'atmosph re par un vecteur de la forme $\mathbf{x} = (\zeta, \eta, T, P_S, q)$. Chaque composante de ce vecteur est un degr  de libert 

¹Action de Recherche Petite Echelle Grande Echelle

² $q = \frac{\text{masse de vapeur d'eau}}{\text{masse de vapeur d'eau} + \text{masse d'air sec}}$

dont les valeurs sont contraintes à respecter certains équilibres imposés par les équations diagnostiques (équation de continuité, relation hydrostatique,...) et dont l'évolution temporelle est donnée par les équations pronostiques (équation d'évolution du tourbillon, équation d'évolution de la divergence,...). Le vecteur d'état évolue dans l'espace du modèle.

La prévision opérationnelle se décompose en 4 cycles journaliers, un cycle est une période de 6h durant laquelle différentes opérations sont réalisées, organisées en deux phases : l'assimilation de données (1) et la prévision (2). Ainsi ces opérations s'enchaînent de la façon suivante :

- (1) Dans un premier temps, les observations de l'atmosphère sont récupérées. Comme certaines erreurs de mesure sont parfois importantes, il est nécessaire d'écarter un certain nombre de données considérées aberrantes. Ensuite, une combinaison judicieuse de ces informations avec l'ébauche est réalisée, pour en déduire l'état de l'atmosphère, initialisant la prévision. Cet état est appelé analyse.
- (2) A partir de l'analyse, une prévision est mise en place, la fenêtre temporelle calculée allant jusqu'à 4 jours (96h).

Les prévisions météorologiques sont très sensibles aux conditions initiales, comme le montre le système dynamique étudié par Lorentz (1963a, 1963b). Il est donc capital de déterminer avec le maximum de précision l'état initial de l'atmosphère pour amorcer la prévision.

2.2 Introduction à l'assimilation de données

2.2.1 Equations de l'analyse

La détermination de l'état exact \mathbf{x}^t de l'atmosphère est impossible. Il est néanmoins possible d'estimer l'état \mathbf{x}^a le plus probable, au vu des informations dont on dispose à un instant donné. L'assimilation de données a pour objectif de déterminer cet état le plus probable \mathbf{x}^a à un instant donné, en utilisant les observations \mathbf{y}^o .

De manière conventionnelle, n désigne la dimension de l'espace du modèle et p celle de l'espace des observations. Une observation est la mesure d'un paramètre physique en un point de l'atmosphère. Ainsi, ce vecteur d'observations contient des données de natures très différentes : pression, température, vitesse du vent, radiance satellitaire, etc. Pour comparer les champs modèle aux observations, il est nécessaire d'utiliser un opérateur d'observation \mathcal{H} (*a priori* non-linéaire) réalisant le passage de l'espace du modèle à l'espace des observations. Ainsi, \mathcal{H} effectue des interpolations, de la grille du modèle vers un point quelconque de la sphère, afin d'y estimer la valeur du champ physique mesuré en ce point, en fonction des valeurs modèle connues en point de grille. L'analyse est donc la solution du problème inverse $\mathbf{y}^o = \mathcal{H}(\mathbf{x})$ où \mathbf{x} est l'inconnue. La principale difficulté est que la dimension de l'espace du modèle est de l'ordre de $n = \mathcal{O}(10^7)$, tandis que celle de l'espace des observations est de $p = \mathcal{O}(10^6)$. Le problème n'est donc pas fermé. Une manière de se ramener à un problème fermé est d'utiliser une méthode de prédiction-correction. Pour cela, le modèle atmosphérique est supposé suffisamment fiable, pour qu'une prévision soit utilisée comme prédicteur. Cette prévision est l'ébauche, elle est notée \mathbf{x}^b . L'ébauche est ajustée pour rendre compte des observations. Dans le processus d'assimilation de données, l'ébauche peut-être vue comme une observation supplémentaire.

Il existe différentes approches pour résoudre ce problème. Quand on cherche à déterminer l'état le plus probable, une approche naturelle est le maximum de vraisemblance (Lorenz 1986). L'observation est connue à une erreur d'observation ϵ^o près, suivant $\mathbf{y}^o = \mathcal{H}(\mathbf{x}^t) + \epsilon^o$. Cette erreur est modélisée par un vecteur aléatoire d'espérance $\mathbb{E}(\epsilon^o) = 0$ (ce qui signifie que la valeur moyenne de l'erreur est nulle). La densité de probabilité associée à ce vecteur est supposée Gaussienne, caractérisée par la matrice de covariance d'erreur d'observation $\mathbf{R} = \mathbb{E}(\epsilon^o \epsilon^{oT})$, où T désigne la transposée. De même,

l'ébauche est l'état vrai à une erreur de prévision près, suivant $\mathbf{x}^b = \mathbf{x}^t + \epsilon^b$. Cette erreur est modélisée par un vecteur aléatoire d'espérance $\mathbb{E}(\epsilon^b) = \mathbf{0}$. La densité de probabilité associée à ce vecteur est supposée Gaussienne, caractérisée par la matrice de covariance d'erreur de prévision $\mathbf{B} = \mathbb{E}(\epsilon^b \epsilon^{bT})$. Sous l'hypothèse que les erreurs de prévisions sont décorréelées des erreurs d'observations, il vient $\mathbb{E}(\epsilon^b \epsilon^{oT}) = \mathbf{0}$ et $\mathbb{E}(\epsilon^o \epsilon^{bT}) = \mathbf{0}$. En procédant suivant le maximum de vraisemblance, il vient que \mathbf{x}^a est l'unique solution du problème de minimisation de la fonction coût $J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}^b - \mathbf{x})^T \mathbf{B}^{-1}(\mathbf{x}^b - \mathbf{x}) + \frac{1}{2}(\mathbf{y}^o - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathcal{H}(\mathbf{x}))$. La recherche de l'analyse se ramène à un problème de minimisation. La méthode de résolution variationnelle de ce problème est le 3D-Var. Son extension temporelle est appelée 4D-Var. Dans le cas où \mathcal{H} est linéaire noté \mathbf{H} , la fonction coût est quadratique et le problème admet une solution analytique

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \text{ avec } \mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R} \quad (2.1)$$

et de plus la matrice de covariance des erreurs d'analyse est :

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} = (\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{B}. \quad (2.2)$$

Ce sont les équations de l'analyse. Une autre expression de la matrice de gain \mathbf{K} est donnée par $\mathbf{K} = \mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$.

2.2.2 Ce qu'est la matrice \mathbf{B}

La matrice \mathbf{B} représente l'incertitude sur l'état atmosphérique prévu à un instant donné, dans l'espace du modèle. Cette incertitude est la conséquence de différentes sources d'erreur : elle dépend de l'erreur d'analyse (incertitude sur l'état initial, dont la croissance temporelle entraîne la dérive vers d'autres états de l'atmosphère) et de l'erreur du modèle (mauvaise prise en compte de processus physiques, mauvaise résolution des processus pris en compte, paramétrisation approximative des processus non connus de manière explicite, forçage aléatoire non pris en compte). En supposant le modèle parfait, l'erreur de prévision se réduit à sa composante associée à l'erreur d'analyse, rendant son étude comparable à celle de la turbulence. D'ailleurs, la méthode ensembliste de Evensen (paragraphe 2.2.4) est une résolution numérique directe, que l'on peut rapprocher de la Simulation Numérique Directe (SND) pour résoudre explicitement la turbulence. Cette méthode donne accès à la matrice \mathbf{B} en réalisant des statistiques sur un ensemble d'états prévus, à partir d'un état de base perturbé de manière aléatoire. Naturellement, on sait que le modèle est imparfait. Malheureusement, l'erreur du modèle reste mal connue et son étude est difficile, aussi on modélise cette erreur suivant des hypothèses simplificatrices. Un modèle simple de l'erreur du modèle est de la considérer comme étant un processus stochastique η_k tel que $\mathbf{x}_{k+1}^t = \mathcal{M}_{k+1,k}(\mathbf{x}_k^t) + \eta_{k+1}$ (du type "marche au hasard" ou "mouvement brownien", i.e. bruit blanc), impliquant une inflation dans la distribution de probabilité. Il est à noter que dans ce cas, le système dynamique est non déterministe. L'erreur η_k est supposée être d'espérance nulle et de matrice de covariance \mathbf{Q}_k . Le problème revient alors à spécifier correctement la matrice \mathbf{Q}_k .

Un modèle non-linéaire simple à 2 degrés de liberté (modèle de Lotka-Volterra, Volterra 1926) permet de donner une intuition de ces éléments abstraits. La figure (2.1) représente le résultat de l'intégration temporelle du modèle à partir de l'état initial (0.4, 0.3) (trajectoire bleue), l'espace du modèle étant la restriction du plan aux points de coordonnées positives. L'incertitude sur l'état initial est donnée par un nuage de points localisé en A , d'aspect sphérique, issu d'une distribution Gaussienne de matrice de covariance $\mathbf{A} = 0.01 \mathbf{I}$, où \mathbf{I} représente la matrice identité. Ce nuage de points montre la correspondance entre la densité de probabilité et la dispersion d'un nuage de points qui lui est caractéristique, le volume de ce nuage étant l'entropie. En supposant le modèle parfait (fig. 2.1.a), l'erreur de prévision est caractérisée par la dispersion du nuages de points (qui évolue au cours du

temps). Le nuage de points sphérique en A se transforme en un nuage de points elliptique en B , puis en des nuages de points aux structures plus complexes en C et D . La forme elliptique du nuage de points B est associée à l'évolution linéaire des petites perturbations initiales, redonnant une distribution Gaussienne avec des directions principales d'étirement. En notant \mathbf{M} le modèle linéaire associé à l'intégration temporelle de l'état caractérisé par A jusqu'à l'état caractérisé par B , une erreur ϵ^a sur la détermination de l'état initial se transforme en une erreur $\epsilon^b = \mathbf{M}\epsilon^a$ sur l'état prévu. La matrice de covariance des erreurs de prévisions en B est donnée par $\mathbf{B}_B = \mathbb{E}(\epsilon^b \epsilon^{bT}) = \mathbf{M}\mathbf{M}^T$. Les non-linéarités entraînent une distribution de probabilité non Gaussienne et donc des nuages de points non elliptiques. D'après cet exemple, la matrice \mathbf{B} dépend de différents facteurs : l'erreur d'analyse, la durée d'intégration, le taux de croissance des erreurs. En prenant maintenant en compte l'erreur du modèle, considérée comme un processus de marche au hasard, caractérisé par une matrice de covariance \mathbf{Q} (fig. 2.1.b), il apparaît que les nuages de points sont plus dispersés. Avec les notations précédentes, il vient que $\mathbf{B}_B = \mathbf{M}\mathbf{M}^T + \mathbf{Q}$. Le fait d'ajouter la matrice \mathbf{Q} à la matrice $\mathbf{M}\mathbf{M}^T$ a pour effet d'augmenter le volume du nuage de points. A l'inverse, une analyse a tendance à diminuer le volume du nuage de points d'après $\mathbf{A} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B}$.

2.2.3 Filtre de Kalman (KF), filtre de Kalman étendu (EKF)

L'ensemble des éléments mis en place permettent de dessiner un schéma complet d'analyse-prévision dans le cas où le modèle est linéaire. Soit \mathbf{x}_{k-1}^a l'analyse au temps $k-1$, caractérisée par la matrice d'erreur d'analyse \mathbf{A}_{k-1} . De cette analyse, on réalise une prévision $\mathbf{x}_k^b = \mathbf{M}_{k,k-1}(\mathbf{x}_{k-1}^a)$, pour un état vrai $\mathbf{x}_k^t = \mathbf{M}_{k,k-1}(\mathbf{x}_{k-1}^t) + \eta_k$, en désignant par $\mathbf{M}_{k,k-1}$ l'intégration des temps $k-1$ à k du modèle et η_k l'erreur du modèle. L'erreur de prévision est caractérisée par la matrice de covariance $\mathbf{B}_k = \mathbf{M}_{k,k-1}\mathbf{A}_{k-1}\mathbf{M}_{k,k-1}^T + \mathbf{Q}_k$, où \mathbf{Q}_k est la matrice de covariance des erreurs du modèle. Soit l'observation \mathbf{y}_k^o du système à l'instant k , caractérisée par la matrice d'erreur d'observation \mathbf{R}_k , et l'opérateur d'observation (linéaire) \mathbf{H}_k . Avec ces notations, l'analyse pour cet instant est donnée par $\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{K}_k(\mathbf{y}_k^o - \mathbf{H}_k\mathbf{x}_k^b)$, où $\mathbf{K}_k = \mathbf{B}_k\mathbf{H}_k^T(\mathbf{H}_k\mathbf{B}_k\mathbf{H}_k^T + \mathbf{R}_k)^{-1}$. La matrice de covariance d'erreur d'analyse est alors donnée par $\mathbf{A}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{B}_k$. Ces différentes étapes d'analyse-prévision, forment les équations du filtre de Kalman (KF) (Kalman 1960). Ces équations ne sont valables que pour le cas d'un modèle linéaire. Une version étendue existe cependant pour le cas non-linéaire. Dans ce cas, l'étape de prévision est réalisée avec le modèle non-linéaire, tandis que dans l'expression de la matrice \mathbf{B}_k intervient le modèle linéaire tangent (modèle qui gouverne l'évolution des petites perturbations autour de l'état analysé \mathbf{x}_{k-1}^a). Cette version est le filtre de Kalman étendu (EKF). Dans la pratique, il est impossible de résoudre les équations du filtre de Kalman. En effet, cela nécessite de connaître et de faire évoluer la matrice \mathbf{B}_k . Or, la dimension de l'espace du modèle, de l'ordre de $\mathcal{O}(10^7)$, implique que la matrice \mathbf{B} est de l'ordre de $\mathcal{O}(10^{14})$, ce qui est hors de portée à l'heure actuelle pour les applications opérationnelles.

2.2.4 Filtre de Kalman d'Ensemble (EnKF), méthode d'estimation de B

Le filtre de Kalman est en réalité une version séquentielle de l'équation de Fokker-Planck ou de Kolmogorov qui régit l'évolution de la densité de probabilité dans l'espace du modèle (Cohn 1993). L'équation de Fokker-Planck est déjà très coûteuse à résoudre pour des systèmes dynamiques simples. Ainsi, elle n'est jamais résolue de manière directe mais suivant une méthode de Monte Carlo, basée sur l'évolution temporelle d'un ensemble d'états. Evensen (1994) propose de résoudre les équations du filtre de Kalman à l'aide d'un tel ensemble : cette méthode est le filtre de Kalman d'ensemble (EnKF). En particulier, cette méthode permet une estimation de la matrice \mathbf{B} . Partant d'un ensemble de N analyses $\mathbf{x}_{k-1,i}^a$ indexées sur i et valables pour l'instant $k-1$ (i.e. compatibles avec la matrice

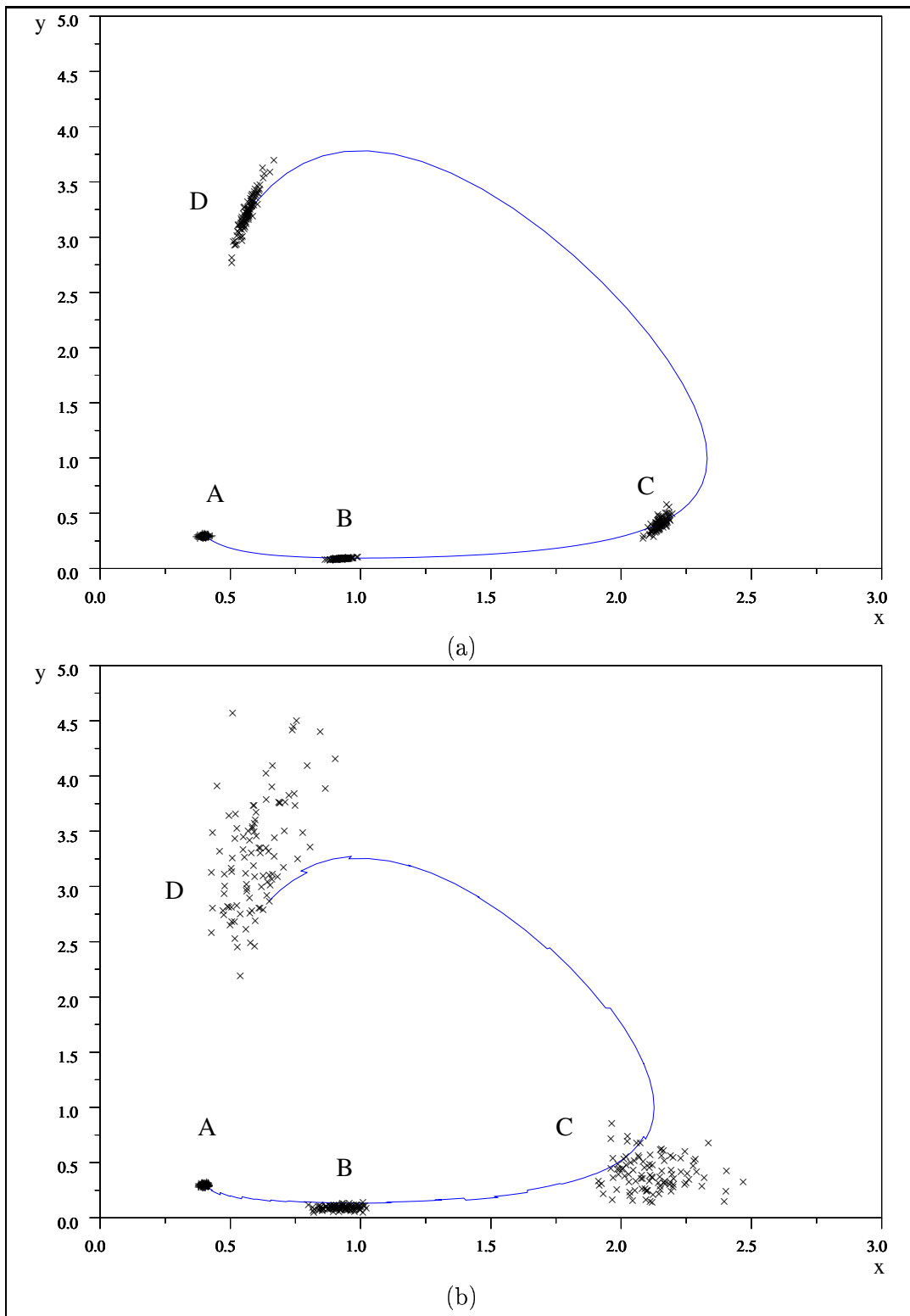


FIG. 2.1 – Représentation de l'intégration d'un nuage de points par un modèle non-linéaire à 2 degrés de liberté (ici du type Lotka-Volterra) sans perturbation du modèle (a) et avec perturbation du modèle (b).

\mathbf{A}_{k-1}), on réalise un ensemble de prévisions $\mathbf{x}_{k,i}^b = \mathcal{M}_{k,k-1}(\mathbf{x}_{k-1,i}^a) + \eta_{k,i}$, valable pour l'instant k , avec $\mathcal{M}_{k,k-1}$ l'intégration du modèle non-linéaire et $\eta_{k,i}$ une perturbation caractéristique de l'erreur du modèle. Par analogie avec les résultats obtenus avec le modèle à deux degrés de liberté, cet ensemble de prévisions est un nuage de points dans l'espace du modèle, caractéristique de la matrice de covariance d'erreur de prévision. En notant $\bar{\mathbf{x}}_k^b = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_{k,i}^b$ le centre de gravité du nuage, l'erreur de prévision est caractérisée par les différences $\epsilon_{k,i}^b = \mathbf{x}_{k,i}^b - \bar{\mathbf{x}}_k^b$. Il vient une estimation de la matrice \mathbf{B}_k par $\mathbf{B}_k^e = \frac{1}{N-1} \sum_{i=1}^N \epsilon_{k,i}^b \epsilon_{k,i}^{bT}$. A partir du vecteur d'observation \mathbf{y}_k^o , on génère un ensemble de vecteurs d'observations perturbées $\mathbf{y}_{k,i}^o = \mathbf{y}_k^o + \epsilon_{k,i}^o$ tel que $\mathbf{R}_k = \mathbb{E}(\epsilon_{k,i}^o \epsilon_{k,i}^{oT})$. Ainsi, chaque prévision $\mathbf{x}_{k,i}^b$ donne une analyse suivant $\mathbf{x}_{k,i}^a = \mathbf{x}_{k,i}^b + \mathbf{K}_k^e (\mathbf{y}_{k,i}^o - \mathbf{H} \mathbf{x}_{k,i}^b)$ où $\mathbf{K}_k^e = \mathbf{B}_k^e \mathbf{H}^T (\mathbf{H} \mathbf{B}_k^e \mathbf{H}^T + \mathbf{R}_k)^{-1}$. Dans la pratique, ces analyses ne sont pas obtenues directement avec l'équation de l'analyse, mais en résolvant le problème de minimisation associé et dans lequel \mathbf{B}_k est remplacé par son estimation \mathbf{B}_k^e .

Dans la chaîne opérationnelle, on utilise une méthode dérivant du Kalman d'ensemble, permettant d'avoir accès aux erreurs de prévision $\epsilon_{k,i}^b$ (Houtekamer et al. 1996, Fisher 2003).

2.2.5 Variance, corrélation et longueur de portée

La matrice de covariance \mathbf{B} contient des informations complexes sur les liens statistiques existant pour l'erreur de prévision en chaque degré de liberté du modèle (par exemple la valeur d'un champ physique en un point de la grille du modèle est un degré de liberté du modèle). Ces liens statistiques sont décrits par les fonctions de structure. Une fonction de structure caractérise le lien statistique existant entre un degré de liberté donné et tous les autres degrés de liberté du modèle. En particulier, cette fonction est tridimensionnelle.

La variance d'un paramètre en un point de grille quantifie la dispersion de l'erreur et donc l'amplitude de l'incertitude (en valeur absolue). Pour un degré de liberté du modèle $\alpha_{k,i}$ (valeur d'un champ α_k du modèle en un point i de la grille), la variance de ce point est la quantité $\sigma_{k,i}^2 = \mathbb{E}(\epsilon_{k,i}^b \epsilon_{k,i}^b)$. Cette variance se situe sur la diagonale de la matrice de covariance \mathbf{B} . Ainsi, la carte des variances, i.e. le champ constitué de la variance des erreurs de prévision pour chaque champ modèle et chaque point de grille, est la diagonale de la matrice \mathbf{B} .

La modélisation de la matrice \mathbf{B} passe par celle de la matrice de corrélation \mathbf{C}_B qui lui est associée. En notant Σ_B la matrice diagonale constituée des écarts types $(\sigma_{k,i})$, la matrice de corrélation est donnée par $\mathbf{C}_B = \Sigma_B^{-1} \mathbf{B} \Sigma_B^{-T}$. Chaque colonne de cette matrice \mathbf{C}_B est une fonction de corrélation associée au degré de liberté que caractérise la colonne.

La longueur de portée est la distance de corrélation significative. Cette longueur caractérise l'influence d'une erreur en un point sur les autres points. Si une erreur de prévision est commise en un point, la cohérence du modèle de l'écoulement implique une contamination spatiale de cette erreur qui est plus ou moins "locale", selon le paramètre et la situation météorologique en jeu. Le degré de localisation dépend du degré de liberté considéré : ainsi, les longueurs de portée sont plus grandes pour les champs de grande échelle (tels que la pression de surface) que pour les champs de petite échelle (tels que la température de surface).

Un cadre unidimensionnel permet de représenter simplement cette information. On peut ainsi considérer un cercle muni d'une grille régulière, composée de n points. Dans ce cadre, en ne considérant qu'un seul champ (température, pression,...), l'erreur de prévision est un vecteur $\epsilon^b = (\epsilon_i^b)$. Une fonction de corrélation est une fonction périodique sur le cercle. Le cas où la fonction de corrélation est la même en chaque point correspond au cas homogène. Dans ces conditions, l'unique fonction de corrélation est notée ρ . Un cas analytique de corrélation homogène sur le cercle est donné par la corrélation homogène Gaussienne $\rho(r) = \exp(-\frac{r^2}{2L_p^2})$, avec r la géodésie séparant deux points. La figure (2.2) montre la fonction ρ dans le cas du cercle équatorial, pour une longueur de portée de $L_p = 350km$, et aussi

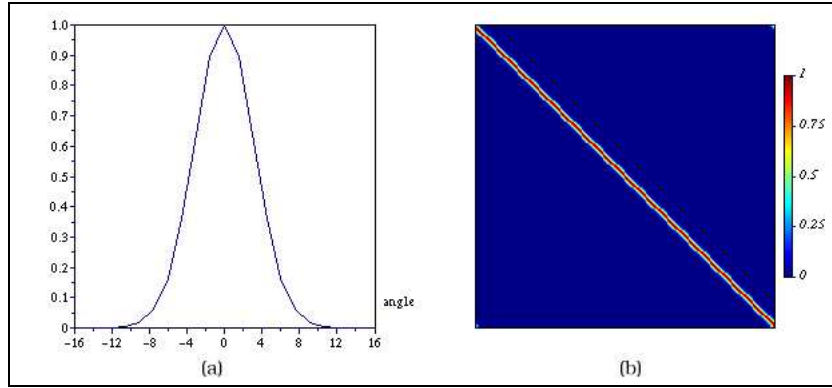


FIG. 2.2 – Représentation de la corrélation $\rho(r) = \exp(-\frac{r^2}{2L_p^2})$ dans le cas homogène (a), l'abscisse étant la séparation angulaire sur le cercle équatorial, et de la matrice de corrélation \mathbf{C}_B associée (b). Dans ce cas, on a discrétisé le cercle équatorial en 239 points, avec $L_p = 350km$.

la matrice de corrélation associée à ce cas. Il existe l'équivalent sur la sphère de telles fonctions de corrélation homogènes, à l'identique du cercle. Ce sont des fonctions qui admettent un axe de symétrie de révolution, ce sont des fonctions radiales. Une expression analytique de cette longueur est donnée par la micro-échelle turbulente de Taylor (caractéristique de l'approche corrélation en deux points en turbulence), suivant $L_p^2 = -\frac{2\rho(0)}{\nabla^2\rho(0)}$ (Daley 1991).

Tout au long de cette étude, une matrice \mathbf{B} hétérogène est utilisée, donnant un cadre analytique pour une expérimentation contrôlée. Cette matrice est obtenue à l'aide d'une transformation de Schmidt (Courtier et Geleyn 1988), utilisée comme un changement de coordonnées, un peu comme le changement de coordonnées semi-géostrophique pour les structures frontales (Desroziers 1997). Ce type de transformation permet de ramener une matrice \mathbf{B} hétérogène à une matrice \mathbf{B}' homogène; ici on réalise l'opération inverse. Dans ce cas analytique, et dès qu'il en sera question, la coordonnée sur le cercle est la pseudo-latitude qui varie de 0° à 360° . La figure (2.3.a) représente les corrélations de cette matrice en quelques points. La carte de variance sera ajustée suivant les utilisations. Cette matrice \mathbf{B} peut être vue comme un modèle schématisé d'une matrice de covariance d'erreur de prévision associée à un écoulement atmosphérique en présence d'un front unique, positionné à la pseudo-latitude 180° .

2.2.6 Filtrage et propagation de l'innovation

L'expression de \mathbf{K} sous la forme $\mathbf{K} = \mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ est fondamentale car elle rend compte de l'importance de la matrice \mathbf{B} .

En considérant une observation unique y^o , localisée en un point de la grille du modèle où l'ébauche vaut x_b , il vient que l'opérateur d'observation est une projection de l'espace du modèle sur le point de grille observé. Soit ρ la fonction de corrélation associée au point de grille, σ_b^2 la variance des erreurs de prévision en ce point et σ_o^2 la variance de l'erreur d'observation. Ainsi, l'incrément d'analyse $\delta\mathbf{x}^a = \mathbf{x}^a - \mathbf{x}^b$ s'écrit, d'après l'équation (2.1), $\delta\mathbf{x}^a = \rho(y^o - x_b)/(1 + \frac{\sigma_o^2}{\sigma_b^2})$. Dans cette expression, l'innovation $y^o - x^b$ est filtrée par $\tau = 1/(1 + \frac{\sigma_o^2}{\sigma_b^2})$, avec le facteur τ variant de 0 à 1, comme coefficient de filtrage. Ce filtrage de l'innovation produit l'incrément de correction de l'ébauche au point d'observation. La corrélation ρ propage aux autres points du modèle l'innovation filtrée, donnant la correction finale. Dans le cas où $\tau = 1$, on fait plus confiance à l'observation qu'à l'ébauche : l'analyse vérifie au point d'observation $x^a = y^o$. A l'inverse, dans le cas où $\tau = 0$, on fait davantage confiance à l'ébauche qu'à l'observation

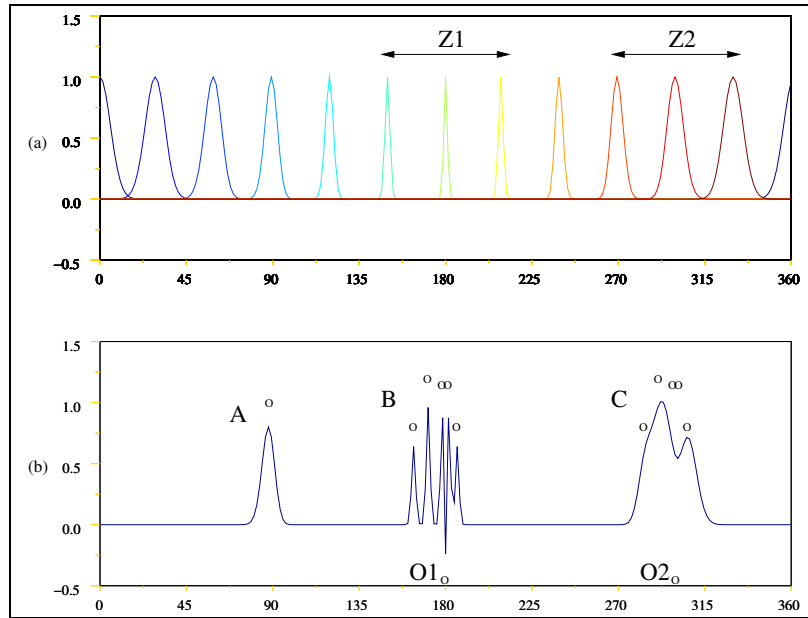


FIG. 2.3 – Représentation des corrélations (une couleur par fonction) dans un cas hétérogène sur le cercle pour les points de 1 à 239 par pas de 20 (a) et représentation de l'incrément d'analyse obtenu par l'assimilation d'observations (petits ronds) réparties de manière hétérogène dans des zones de longueurs de portée différentes.

et l'analyse vérifie au point d'observation $x^a = x^b$: l'innovation est entièrement filtrée. Dans le cas général, l'analyse vérifie au point d'observation $x^a = (1 - \tau)x^b + \tau y^o$ donnant une valeur intermédiaire entre l'ébauche et l'observation. On perçoit ainsi l'enjeu important que revêt la modélisation de \mathbf{B} : si la modélisation est erronée, la correction est dégradée. De façon générale, la matrice \mathbf{K} apparaît comme une matrice de filtrage et de propagation du vecteur d'innovation $\mathbf{d} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^b$.

L'opérateur \mathbf{KH} s'interprète comme étant une matrice (en supposant l'opérateur linéaire) variant de 0 à \mathbf{I}_n . Si $\mathbf{KH} = 0$, alors $\mathbf{x}^a = \mathbf{x}^b$: dans ce cas aucune confiance aux observations n'est faite ou bien elles ne sont pas assez nombreuses pour être prises en compte, c'est ce qui arrive dans les zones pauvres en observations (océan Pacifique, Atlantique sud, pôles). Si $\mathbf{KH} = \mathbf{I}_n$, alors dans ces régions, $\mathbf{x}^a = \mathbf{K}\mathbf{y}^o$: la contribution des observations est plus grande que celle apportée par l'ébauche. Dans la pratique, dans les zones denses en observations (Europe, Amérique du nord), $\mathbf{KH} = \mathbf{I}_n/2$: les observations et l'ébauche ont un poids semblable.

Ces quelques résultats s'illustrent simplement en considérant un exemple 1D, avec la matrice \mathbf{B} analytique hétérogène en longueur de portée, et une variance homogène unité. Les résultats sont reportés sur la figure (2.3). La figure (a) représente les fonctions de corrélation en différents points répartis de manière uniforme sur le cercle (1 à 239 par pas de 20). La variation de longueur de portée est caractérisée par des fonctions de corrélation étroite dans la zone de petites longueurs de portée (zone Z1) et par des fonctions plus larges dans la zone de grandes longueurs de portée (zone Z2). La figure (b) représente le résultat d'une expérience d'assimilation d'observations. Si les groupes d'observations sont suffisamment éloignés, alors chaque groupe peut être considéré comme indépendant des autres dans l'assimilation (conséquence de la discussion sur l'opérateur \mathbf{KH}). Dans cette expérience, on se donne un vecteur d'innovation \mathbf{d} (les petits ronds de la figure b) considéré comme un vecteur d'observations si l'on se place dans le cas $\mathbf{x}^b = 0$, et on représente l'incrément d'analyse $\delta\mathbf{x}^a = \mathbf{K}\mathbf{d}$. Les lettres A, B et C représentent différentes zones de l'expérience. Du fait de leur éloignement, elles peuvent être

considérées comme 3 expériences indépendantes. Le résultat de l'assimilation d'une seule observation est caractérisé par la lettre A . L'incrément d'analyse est alors proportionnel à la covariance au point d'observation. Les zones B et C sont le résultat de l'assimilation d'un même motif de répartition d'observations mais dans des zones de longueurs de portée différentes : $Z1$ pour B et $Z2$ pour C . La longueur de portée étant petite pour B , l'incrément d'analyse est très oscillant, comme si les observations étaient prises séparément. La correction apportée par l'incrément d'analyse est très locale et reste proportionnelle à la covariance en chaque point observé. Dans ce cas, la matrice \mathbf{K} n'a pas beaucoup filtré l'innovation (la correction descend fortement vers l'observation $O1$). Par contre, pour C , localisé dans une zone de grande longueur de portée, la correction est plus étendue et n'est plus localisée en chaque observation. En particulier, les variations rapides sont lissées (la correction descend très faiblement pas vers l'observation $O2$), ce qui est dû au filtrage apporté par \mathbf{K} .

L'influence de la carte de variance est illustrée sur la figure (3.2), à partir d'une assimilation d'observations avec une matrice \mathbf{B} dont quelques covariances sont représentées en (a1). Le résultat de l'assimilation est donné en (a2). Suivant la localisation de l'observation, l'amplitude de la correction est faible en A (zone de faible variance) et forte en B (zone de forte variance). La description de cet exemple est faite plus en détail au paragraphe (3.2).

2.3 Commentaire et position du problème

Pour déterminer l'état de l'atmosphère à un instant donné, le modèle de l'atmosphère intervient dans l'algorithme 4D-Var (Courtier et al 1994). Ainsi, le coût numérique associé à la phase d'assimilation de données est équivalent, voir prépondérant, au coût de la phase de prévision. Grâce à ces méthodes, les scores (quantification objective de la qualité des prévisions) ont été fortement améliorés. Mais il est important de spécifier correctement toutes les informations nécessaires à l'utilisation de ces méthodes. Pour le moment, les mesures sont considérées comme indépendantes les unes des autres (ce qui n'est plus vrai avec les données satellitaires, compliquant fortement leur traitement), ce qui se traduit pas une matrice \mathbf{R} diagonale, constituée des variances de chaque mesure.

En revanche, la matrice \mathbf{B} est loin d'être diagonale, sa structure est complexe et sa taille gigantesque rend impossible sa représentation complète. De plus, elle n'est pas connue de manière exacte, et ne peut être qu'estimée. La solution est de modéliser cette matrice. Différentes pistes sont explorées à l'heure actuelle. Certains préconisent une reconstitution de la matrice \mathbf{B} en réalisant des statistiques sur des ensembles d'états perturbés de manière aléatoire, en faisant évoluer les directions principales de cette matrice symétrique. D'autres modélisent \mathbf{B} en se basant sur l'équation de diffusion (Weaver et Courtier 2001, pour l'océanographie). Naturellement, une bonne méthode étant une méthode utilisable en opérationnel, il est nécessaire que la modélisation de \mathbf{B} reste la moins coûteuse possible, tant au niveau de la mémoire que des calculs : on recherche un équilibre entre représentation de la réalité et contrainte opérationnelle.

Durant ce stage, une nouvelle formulation, basée sur les ondelettes sphériques, a été étudiée. Dans le chapitre qui suit, la formulation actuelle et ce que contient la matrice \mathbf{B} en termes d'informations sont exposés. La nouvelle formulation est ensuite présentée, ainsi que les résultats obtenus dans un cadre simplifié sur cercle.

Chapitre 3

B dans l'atmosphère et sa modélisation

Dans ce chapitre, la formulation opérationnelle de la matrice \mathbf{B} est présentée, pour aboutir au formalisme classique de \mathbf{B} construite comme une cascade d'opérateurs. Un contexte simple est tout d'abord considéré, illustrant ce que représente la matrice \mathbf{B} et les informations qu'elle contient. Les avantages et les inconvénients de la formulation actuelle sont présentés, permettant d'apprécier les avancées proposées par la formulation ondelette.

3.1 Physique des statistiques de l'atmosphère

La matrice \mathbf{B} dans l'atmosphère est une matrice pleine représentée par blocs suivant les paramètres physiques que l'on considère. En notant $\mathbf{B}[\alpha; \beta]$ la matrice des covariances croisées entre les champs α et β , et $\mathbf{B}[\alpha]$ la matrice d'auto-covariance du champ α , il vient

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}[\zeta] & \mathbf{B}[\zeta; \eta] & \mathbf{B}[\zeta; (T, P_S)] & \mathbf{B}[\zeta; q] \\ \mathbf{B}[\zeta; \eta]^T & \mathbf{B}[\eta] & \mathbf{B}[\eta; (T, P_S)] & \mathbf{B}[\eta; q] \\ \mathbf{B}[\zeta; (T, P_S)]^T & \mathbf{B}[\eta; (T, P_S)]^T & \mathbf{B}[(T, P_S)] & \mathbf{B}[(T, P_S); q] \\ \mathbf{B}[\zeta; q]^T & \mathbf{B}[\eta; q]^T & \mathbf{B}[(T, P_S); q]^T & \mathbf{B}[q] \end{pmatrix}. \quad (3.1)$$

Les statistiques sur la sphère ne sont pas homogènes, comme le montre la figure (3.1).

Sur la carte de variance de l'erreur de prévision de la pression de surface, trois bandes principales apparaissent dont deux pour les pôles. Les statistiques sont calculées à partir d'une application du Kalman d'ensemble (paragraphe 2.2.4). Ce sont les phénomènes météorologiques significatifs (tempêtes, cyclone,...) qui modifient de manière importante la répartition de la variance, pour une situation météorologique donnée (dépendant de l'écoulement). On observe aussi une variance importante à l'est des Etats-Unis, c'est le rail des dépressions atlantiques : zone météorologiquement active, où les dépressions qui touchent la France se développent, par l'instabilité barocline, puis s'organisent. On retrouve une zone semblable à l'est de l'Asie.

La carte de longueur de portée présente globalement une certaine symétrie par rapport à l'équateur, avec de fortes longueurs de portée sur les pôles et à l'équateur. De vastes bandes de longueurs de portée moyenne apparaissent, couvrant les régions tropicales et tempérées. En regardant plus finement, il existe en fait une dissymétrie entre les deux hémisphères : il apparaît que les longueurs de portée sont plus grandes dans l'hémisphère sud que dans l'hémisphère nord. En particulier, il y a une vaste zone de courte longueur de portée au niveau de l'Europe. Cette disparité s'explique par le fait que l'hémisphère sud est plus pauvre en observations que l'hémisphère nord. Il est également à noter une différence entre continent et océan : les longueurs de portées sont plus faibles sur terre que sur mer. Des portées courtes sont cependant visibles sur l'Atlantique nord, ce qui correspond probablement à l'influence des

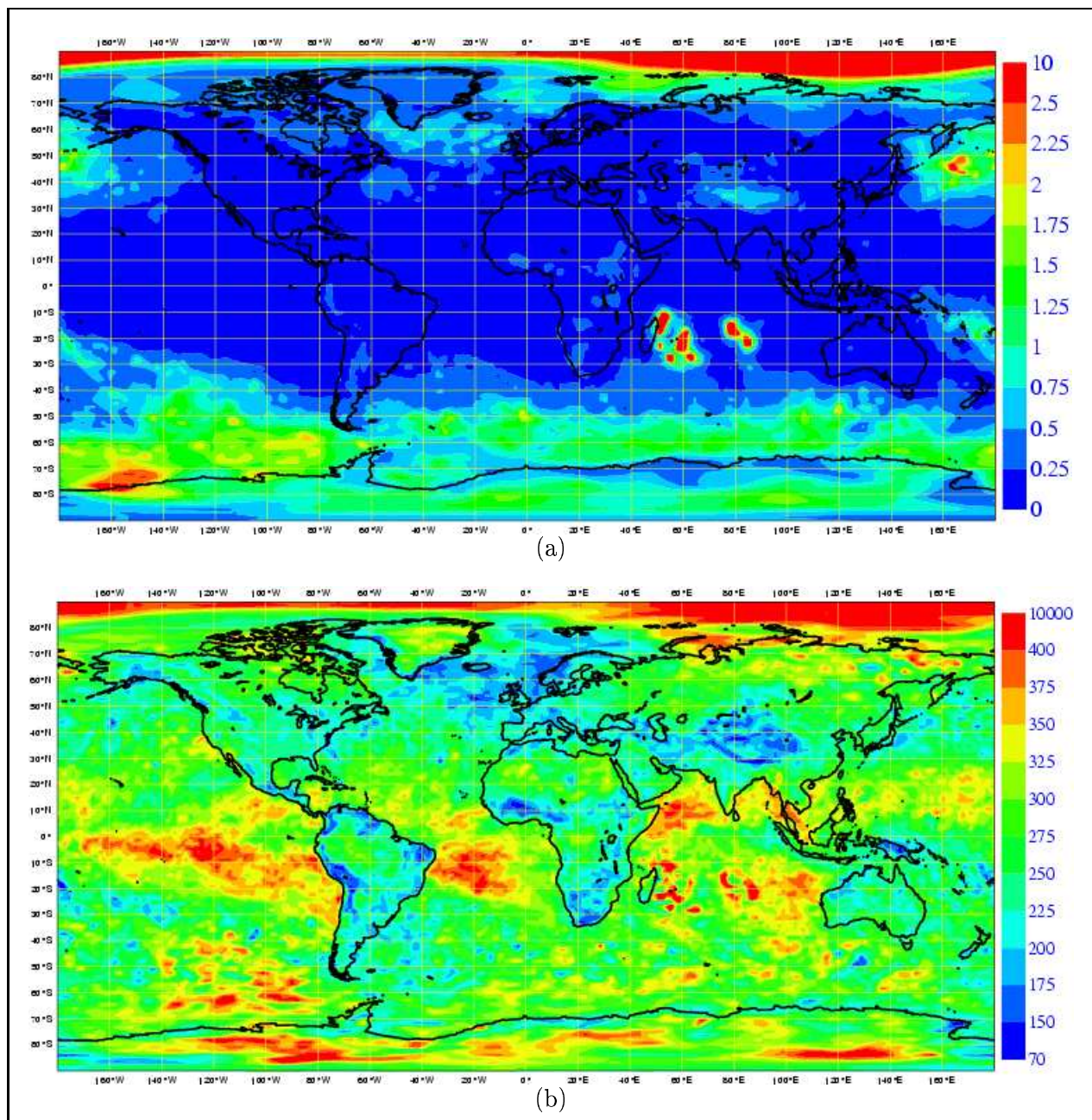


FIG. 3.1 – Représentation des cartes de variance (en hPa) (a) et de longueur de portées (en km) (b) pour la pression de surface.

structures frontales (marquées par de forts gradients) dans les régions du rail des dépressions. Le relief joue aussi un rôle puisque la chaîne Himalayenne et la Cordillère des Andes ont une influence visible.

D'autre part, des interactions existent entre les différents champs physiques. Dans l'atmosphère, des équilibres ou balances de grande échelle relient les champs entre eux. L'équilibre géostrophique ou balance géostrophique est l'équilibre qui à nos latitudes relie le gradient du champ de pression avec le vent géostrophique \mathbf{v}_g : \mathbf{v}_g est le vent horizontal moyen de grande échelle, suivant $\mathbf{v}_g = \frac{1}{\rho f} \mathbf{k} \times \nabla P_s$ avec $f = 2\Omega \sin(\phi)$ le paramètre de Coriolis, ρ la masse volumique, Ω la vitesse angulaire de la terre, ϕ la latitude et \mathbf{k} le vecteur normal à la surface du globe au point considéré. Il vient qu'une erreur sur le champ de pression P_s se répercute sur le vent géostrophique, et donc sur le tourbillon relatif ζ et sur la divergence η . Cet exemple montre que la physique, de par les équations de balance, met en relation les erreurs de prévision entre elles. Ceci illustre la complexité de la matrice \mathbf{B} , mais suggère une modélisation par la mise en relation des champs entre eux, basée sur ce type d'équilibre. Ce type de modélisation est utilisé en opérationnel pour rendre multivariée la matrice \mathbf{B} (Derber et Bouttier 1999).

D'après le paragraphe (2.2.2) la matrice \mathbf{B} dépend de l'état de l'atmosphère, de son évolution et de sa connaissance initiale. Ainsi \mathbf{B} dépend en réalité de la situation du jour. En opérationnel, on considère une version statique de \mathbf{B} , indépendante de la situation, ce qui est pratique mais imparfait. En effet, les variances sont susceptibles d'être plus fortes dans les zones de fort gradient de l'ébauche, tandis que dans ces mêmes zones, la longueur de portée est plus petite.

Ainsi, une bonne modélisation de la matrice \mathbf{B} doit prendre en compte de manière explicite ou implicite les différents points soulevés précédemment. Cependant, la taille énorme de la matrice ne permet pas une description complète. Il est nécessaire de réaliser des hypothèses supplémentaires pour arriver à modéliser cette matrice.

3.2 Modélisations diagonales en points de grille et/ou spectrale

La matrice \mathbf{B} est en réalité une entité intrinsèque à l'espace du modèle à un instant donné. C'est un tenseur doublement covariant. Par conséquent, il est plus approprié de parler d'opérateur \mathbf{B} . Ainsi, cet opérateur peut être représenté de manière matricielle dans différentes bases de l'espace du modèle. L'espace du modèle est l'espace des fonctions représentables de manière exacte sur la grille de collocation du modèle (théorème de Shannon sur l'échantillonnage). Ainsi, une base n'est autre qu'une base de fonctions. Dans le cas de la sphère, les bases fonctionnelles classiques sont la base des Diracs en point de grille (espace point de grille), ainsi que la base des harmoniques sphériques (espace spectral). Le cas du cercle est identique, avec cette fois la base de Fourier. Si \mathbf{B} est représentée par la matrice \mathbf{B}_e dans la base \mathbf{e} , alors sa représentation \mathbf{B}_f dans une base \mathbf{f} est donnée par $\mathbf{B}_f = \mathbf{P}_{fe} \mathbf{B}_e \mathbf{P}_{fe}^T$, avec \mathbf{P}_{fe} la matrice de passage (ou transformation) d'une représentation à l'autre. Dans la suite, la représentation point de grille est privilégiée car les résultats sont facilement représentables. Dans ARPEGE, on privilégie d'autres représentations (spectrales). La structure de \mathbf{B} étant complexe, on la simplifie en supposant que la matrice représentant \mathbf{B} dans une base donnée est diagonale. Ainsi, en considérant la base \mathbf{e} formée des vecteurs (\mathbf{e}_i) , l'hypothèse diagonale dans cette base se traduit par une représentation de \mathbf{B} sous la forme d'une somme de projections $\mathbf{B} = \sum_i \lambda_i \mathbf{e}_i \mathbf{e}_i^T$, avec $\lambda_i > 0$. Par cette écriture, on peut factoriser \mathbf{B} sous la forme $\mathbf{B} = \mathbf{B}^{1/2} \mathbf{B}^{T/2}$ avec $\mathbf{B}^{1/2} = \sum_i \lambda_i^{1/2} \mathbf{e}_i \mathbf{e}_i^T$. Ainsi, $\mathbf{B}^{1/2} \mathbf{e}_i = \lambda_i^{1/2} \mathbf{e}_i$. Finalement, l'hypothèse diagonale n'est autre que la modélisation des directions principales de la matrice par les vecteurs de la base (quand celle-ci est une base orthogonale).

Un exemple 1D est utilisé pour visualiser l'impact de l'hypothèse suivant le choix de la base. La figure (3.2) représente en (a1) les covariances d'une matrice \mathbf{B} modélisant de manière schématique les variations physiques associés à un front météorologique positionné à la pseudo-latitude 180° , caractérisé

par une faible variance dans la zone de grande longueur de portée (zone Z1) et une variance importante dans la zone de petite longueur de portée (zone Z2). En (a2) est représenté l'incrément d'analyse pour deux observations localisées aux points A et B . Le point A est caractéristique de la zone Z1, tandis que B est caractéristique de Z2. Dans cette expérience, l'écart type de l'erreur d'observation est constant, pris égale à $\sigma_o = 0.5$. En A , dans la zone stable de l'écoulement, la variance de l'erreur d'ébauche étant faible par rapport à celle de l'erreur d'observation, l'incrément d'analyse a une amplitude plus faible. D'autre part il est très étalé, propageant la correction à grande échelle. A l'inverse, en B , dans la zone instable de l'écoulement, la variance de l'erreur d'ébauche est grande par rapport à celle de l'observation : l'incrément d'analyse tend à déformer fortement l'ébauche pour coller davantage à l'observation. Par contre, la propagation de la correction reste très localisée au voisinage de l'observation. Ces résultats correspondent à la référence que l'on cherche à obtenir par l'application de l'hypothèse diagonale dans une base bien choisie.

Hypothèse diagonale en points de grille

En se plaçant dans l'espace point de grille, les vecteurs de la base constituée de Diracs sont notés Δ_i avec i l'indice de point de grille. La figure (3.3.a) montre la représentation des fonctions de base sur une grille régulière, pour le cercle. Chaque niveau correspond à la représentation d'une fonction. L'hypothèse diagonale s'exprime suivant $\mathbf{B} = \sum_i \sigma_i^2 \Delta_i \Delta_i^T = \text{Diag}(\sigma_i^2)$. Dans ce cas, σ_i^2 représente la variance de l'erreur de prévision associée au point de grille considéré. En notant $\Sigma = \text{Diag}(\sigma_i)$, il vient $\mathbf{B} = \Sigma \mathbf{I} \Sigma^T$. La matrice de corrélation associée à \mathbf{B} est dans ce cas l'identité. Ainsi, les fonctions de corrélation sont des Diracs, ce qui signifie qu'une erreur en un point n'implique de contribution en aucun autre point. En particulier, les longueurs de portée sont nulles. En considérant l'expérience 1D figure (3.2), les covariances modélisées dans ce cas sont représentées en (b1). Dans ce modèle, la carte de variance est correctement représentée, mais pas la longueur de portée ni a fortiori sa variation géographique. En particulier, d'après (b2) il n'y a pas de propagation de la correction, alors qu'elle existe en (a2). Par contre, la modulation de l'amplitude de la correction est exacte (pour ce cas à une observation) égal à celle de (a2).

L'hypothèse diagonale dans l'espace point de grille permet donc une bonne représentation de la variance locale, mais elle n'est pas adaptée pour la propagation de la correction.

Hypothèse diagonale dans l'espace spectral

En se plaçant maintenant dans l'espace spectral sur la sphère, les vecteurs de la base spectrale sont notés $\mathbf{Y}_{n,m}$. La figure (3.3.b) montre la représentation des fonctions de base dans l'espace physique et pour le cercle. Chaque niveau correspond à la représentation par nombre d'onde. Pour cet espace, l'hypothèse diagonale s'exprime suivant $\mathbf{B} = \sum_{n,m} \lambda_n^2 \mathbf{Y}_{n,m} \mathbf{Y}_{n,m}^T$. Dans ce cas, il est possible de montrer que la covariance est homogène dans l'espace point de grille. On retrouve ce résultat dans le cas du cercle, ce qui permet de l'illustrer simplement. En considérant l'expérience 1D figure (3.2), les covariances modélisées dans ce cas sont représentées en (c1). Dans ce modèle, la longueur de portée n'est pas nulle. Par contre elle ne varie pas. De plus, la carte de variance homogène ne représente pas les variations réelles, mais caractérise la variance moyenne. En particulier, d'après (b2) il y a propagation de la correction, mais cette propagation est trop courte en A et pas assez localisée en B . De plus, comme la variance est uniforme, l'amplitude de la correction n'est pas modulée comme en (a2) ou (b2).

L'hypothèse diagonale dans l'espace spectral permet donc de représenter une propagation moyenne de la correction, mais elle n'est pas adaptée pour la représentation de la variance locale. La propagation moyenne de l'information n'est par ailleurs pas satisfaisante localement.

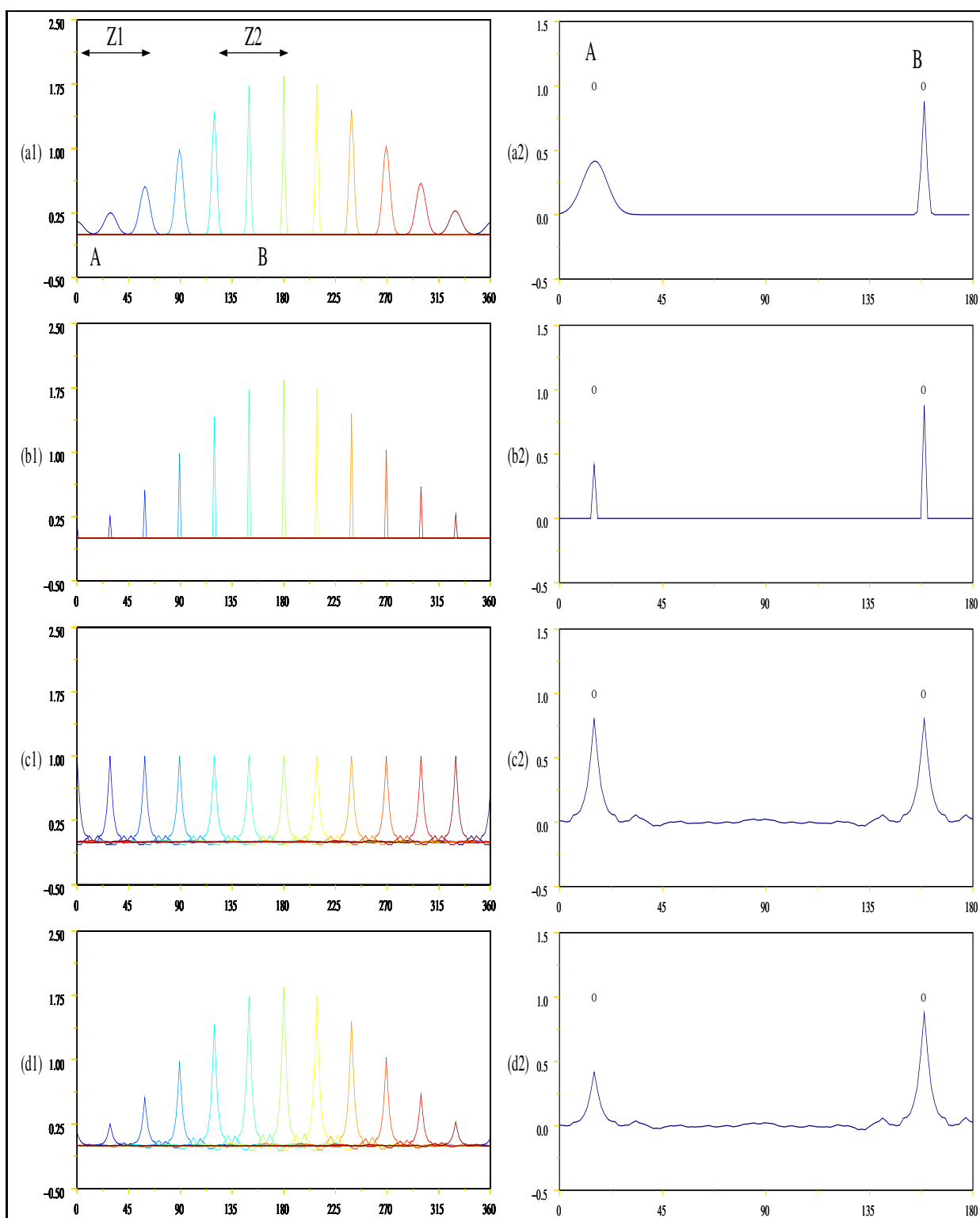


FIG. 3.2 – Expérience d'assimilation comparée à partir d'une matrice de référence exacte, hétérogène. La lettre pour chaque figure représente une matrice de covariance associée dont on représente quelques fonctions de covariance en (a1,b1,c1,d1) et l'incrément d'analyse obtenu avec cette matrice en (a2,b2,c2,d2). La matrice de covariance de référence **B** est en (a1,a2), sa modélisation diagonale point de grille en (b1,b2), sa modélisation diagonale spectrale (homogène au sens des corrélations) en (c1,c2) et enfin sa modélisation diagonale hybride (points de grille / spectral) en (d1,d2).

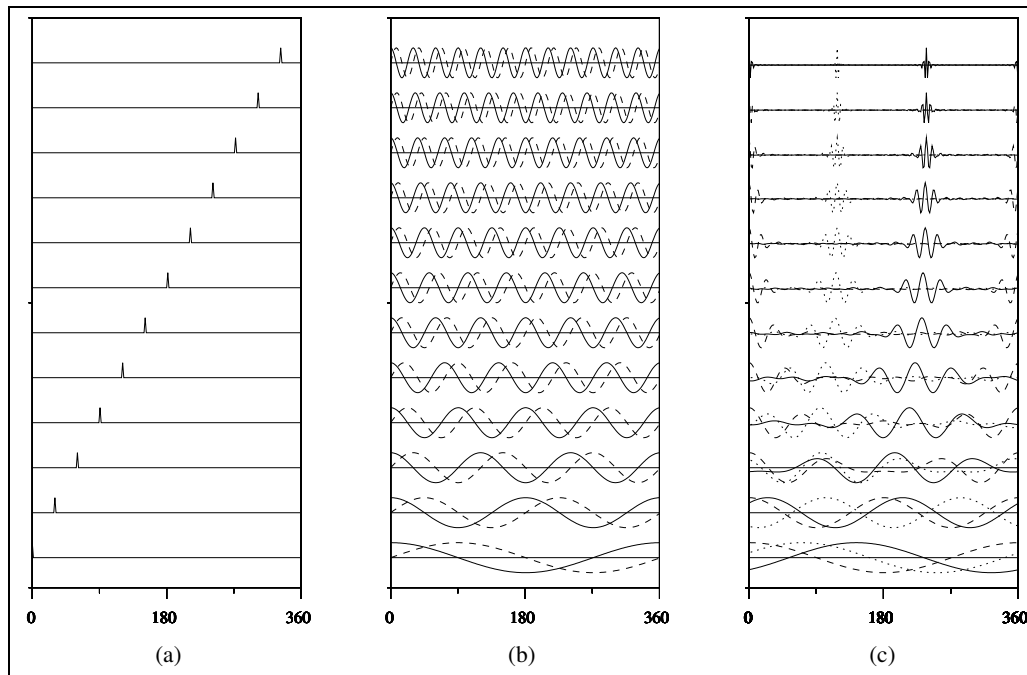


FIG. 3.3 – Représentation de la base orthogonale formée de Diracs (a), de la base orthogonale de Fourier (b), et de quelques fonctions ondelettes formant une famille génératrice non-orthogonale (c).

Formulation hybride

Les deux modélisations précédentes permettent de modéliser certains aspects du signal réel. On envisage maintenant une formulation dans laquelle on tente de profiter au mieux de leurs points forts respectifs. Ainsi, on construit une formulation hybride pour \mathbf{B} sous la forme $\mathbf{B} = \mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}^T$ où $\mathbf{\Sigma}$ est la carte d'écart type exacte (hypothèse diagonale point de grille) et \mathbf{C} est la matrice de corrélation modélisée par l'hypothèse diagonale spectrale. Ainsi, $\mathbf{C} = \mathbf{S}^{-1}\mathbf{D}\mathbf{S}^{-T}$ où \mathbf{D} est une matrice diagonale et \mathbf{S} la matrice de passage de l'espace point de grille à l'espace spectral. En considérant l'expérience 1D de la figure (3.2), les covariances modélisées dans ce cas sont représentées en (d1). Dans ce modèle, la carte de variance est la carte exacte, la longueur de portée n'est pas nulle, par contre elle reste homogène. D'après (d2), les déficiences de la formulation homogène relevées dans la partie précédente sont résolues pour la modulation de l'amplitude de la correction, mais pas pour la modulation de la propagation de la correction. Cette formulation de $\mathbf{B} = \mathbf{\Sigma}\mathbf{S}^{-1}\mathbf{D}\mathbf{S}^{-T}\mathbf{\Sigma}^T$ sous la forme d'un produit d'opérateurs en cascade est classique. Comme on n'est pas capable de spécifier directement des structures complexes, on passe par des étapes intermédiaires plus simples aboutissant globalement à une structure complexe.

La formulation hybride apparaît comme une alternative simple pour modéliser de manière précise la carte de variance, tout en ayant une longueur de portée non nulle. Cependant, on n'a pas résolu le problème de la sur/sous-estimation de la propagation. Maintenant, comment faire varier la longueur de portée?

3.3 Modélisation hétérogène

Dans le cas de la base de Dirac en point de grille, on a une excellente localisation spatiale mais aucune localisation spectrale, puisque dans ce cas le spectre est uniformément égal à 1. A l'opposé, avec

la base spectrale, on a une excellente localisation spectrale mais aucune localisation spatiale d'après (3.3.b). Ces déficiences, intrinsèques aux bases considérées, ne peuvent être résolues avec l'hypothèse diagonale relative à ces espaces. Il est à noter que ce problème de localisation spatiale-spectrale pour une fonction est à rapprocher du principe d'incertitude d'Heisenberg. Si on disposait d'une base (plus largement d'une famille génératrice de l'espace fonctionnel représentable) dans laquelle il serait possible de spécifier localement en espace une information localisée en spectre, alors la réalisation de l'hypothèse diagonale dans cette base rendrait possible la dépendance géographique des structures de corrélation.

Une formulation ondelette est un outil mathématique puissant dans l'étude micro-locale des signaux, construite autour du principe d'Heisenberg. Elle permet de caractériser d'une part l'information spatiale d'un signal, tout en prenant en compte d'autre part une localisation spectrale. La figure (3.3.c) représente une famille d'ondelettes dans l'espace points de grille. Ces fonctions seront étudiées au paragraphe (4.1). Les ondelettes de grande échelle (sur la figure en bas) sont proches de la base spectrale car globales et les ondelettes de plus petite échelle sont proches de la base de Dirac en point de grille. Le spectre des fonctions utilisées est reporté en figure (3.4). Cette figure représente le spectre en cosinus des ondelettes représentées en (3.3.c). Leur spectre est une dilatation du spectre en gras.

Pour modéliser l'hétérogénéité des corrélations, Fisher (2003) a ainsi introduit une décomposition de \mathbf{B} sur la sphère, telle que $\mathbf{B} = \mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}^T$ en modélisant \mathbf{C} par une hypothèse diagonale dans une base d'ondelettes sphériques radiales à bande limitée. Cette formulation est maintenant étudiée après avoir présenté les ondelettes sphériques utilisées.

3.3.1 Ondelettes sphériques

Freeden et Windheuser (1996) ont construit une formulation ondelette basée sur le produit de convolution sur la sphère. Il est à noter que la construction d'ondelettes sur la sphère est complexe et reste très récente. En particulier, on ne dispose pas de tout le savoir-faire intervenant dans le traitement des signaux sur la droite, ou des espaces plats (\mathbb{R}^n). Une formulation ondelette sur la droite réelle peut être reformulée en terme de produit de convolution. L'idée de Freeden et Windheuser est d'utiliser ce type de formulation pour la sphère. Pour définir un produit de convolution sur la sphère, on fait agir le groupe des rotations. Dans ces conditions, le produit de convolution d'une fonction H radiale (fonction sur la sphère qui possède une symétrie de révolution et s'exprime uniquement à l'aide de son spectre de Legendre H_n) par un champ f est une fonction $H \otimes f$ dont le spectre vérifie $(H \otimes f)_n^m = H_n f_n^m$. Une base d'ondelettes sphériques radiales est une famille de J fonctions radiales ψ_j , telle que $\sum_{j=1}^J \hat{\psi}_j(n)^2 = 1$ pour n appartenant à la troncature et avec $\hat{\psi}_j(n)$ le spectre de Legendre de ψ_j .

Dans ces conditions, les coefficients d'ondelettes sont les fonctions $f_j = \psi_j \otimes f$ et la formule de reconstitution est donnée par $f = \sum_{j=1}^J \psi_j \otimes f_j$. Pour $\underline{\eta}$ un point sur la sphère, $f_j(\underline{\eta})$ représente le coefficient de ressemblance entre la fonction f et une rotation de la fonction radiale ψ_j centrée au point $\underline{\eta}$. Il existe un grand nombre d'autres ondelettes définies sur le même principe et obtenues par un procédé systématique très général (Freeden et Windheuser 1996, Freeden et Shreiner 1998). Fisher a choisi de considérer des ondelettes à bandes limitées, c'est-à-dire dont le spectre de Legendre des fonctions radiales ψ_j est nul en-dehors d'une bande N_{j-1} à N_{j+1} , avec (N_j) une suite d'entiers strictement croissante. Le spectre des fonctions utilisées est représenté sur la figure (3.4). Ainsi, ces fonctions sont localisées en spectre, avec une structure de passe-bande (spectrale). Elles sont de plus localisées en espace comme le montre la figure (3.5), ce qui leur confère une structure de filtre spatial. Cette dernière figure montre l'effet de la dilatation caractéristique d'une famille d'ondelettes qui a tendance à recentrer la fonction autour d'un point. Suivant l'indice d'échelle j , les ondelettes recouvrent la sphère (j petit) ou sont très localisées autour d'une région (j grand). Dans la suite on parle de base d'ondelettes. En réalité nous avons une famille génératrice car le nombre de fonctions utilisées est supérieur à la dimension de l'espace de la grille; cependant cela ne pose pas de difficulté (si ce n'est

que le famille n'est pas orthogonale, ce qui induit des couplages que l'on ne détaille pas ici). Il existe aussi une formulation ondelette non-radiale (Torresani 1995).

3.3.2 Formulation ondelette de \mathbf{B}

La modélisation ondelette de la matrice \mathbf{B} passe par la modélisation de sa matrice de corrélation en utilisant une hypothèse diagonale dans l'espace des ondelettes. En notant $\Psi_{j,k}$ les vecteurs de la base d'ondelettes, l'application de l'hypothèse donne $\mathbf{B} = \sum_{j,k} \lambda_{j,k} \Psi_{j,k} \Psi_{j,k}^T$. En notant $\mathbf{D}_w = \text{Diag}(\lambda_{j,k})$ la matrice diagonale des coefficients $\lambda_{j,k}$ dans l'espace des ondelettes, et \mathbf{W} la matrice représentant la transformation "point de grille - espace des ondelettes", la décomposition suivante est obtenue : $\mathbf{B} = \mathbf{S}\mathbf{\Sigma}\mathbf{W}^{-1}\mathbf{D}_w\mathbf{W}^{-T}\mathbf{\Sigma}^T\mathbf{S}^T$.

Dans le cas particulier de la formulation ondelette radiale à bande limitée, il existe une factorisation explicite supplémentaire. Soit Ψ_j la matrice des coefficients de Legendre de ψ_j et \mathbf{f} le vecteur représentant un signal f en point de grille. Dans ces conditions, le produit de convolution $\psi_j \otimes f$ est le vecteur $\mathbf{S}^{-1}\Psi_j\mathbf{S}\mathbf{f}$. Comme ψ_j est à bande limitée, le calcul est effectué en passant par un espace spectral de plus petite dimension, ce qui est équivalent à considérer une grille de résolution plus faible. Soit \mathbf{S}_j la matrice de transformation de l'espace de la grille à l'espace spectral associé à une grille de résolution adaptée. Dans ces conditions, le produit de convolution est le vecteur $\mathbf{S}_j^{-1}\Psi_j\mathbf{S}_j\mathbf{f}$. Ainsi, $\mathbf{W}\mathbf{f} = (\mathbf{f}_1, \dots, \mathbf{f}_J)$ avec $\mathbf{f}_j = \mathbf{S}_j^{-1}\Psi_j\mathbf{S}_j\mathbf{f}$ et $\mathbf{f} = \mathbf{W}^{-1}(\mathbf{f}_1, \dots, \mathbf{f}_J) = \mathbf{S}_J^{-1} \sum_{j=1}^J \Psi_j\mathbf{S}_j\mathbf{f}_j$. En notant \mathbf{C}_j la matrice $\text{Diag}(\lambda_{j,k})$ associée à l'indice j , la formulation devient $\mathbf{B} = \mathbf{S}\mathbf{\Sigma}\mathbf{S}_J^{-1} \left(\sum_{j=1}^J \Psi_j\mathbf{S}_j\mathbf{C}_j\mathbf{S}_j^T\Psi_j^T \right) \mathbf{S}_J^{-T}\mathbf{\Sigma}^T\mathbf{S}^T$.

3.3.3 Problématique d'étude de la formulation

Avec cette formulation, on ne retrouve pas la carte de variance exacte (Fisher 2004, communication personnelle), à l'encontre de ce qui est attendu. D'où provient cet écart, et comment le corriger? Existe-t-il des méthodes efficaces pour y arriver?

Peut-on utiliser des propriétés intrinsèques à la formulation ondelette? Les ondelettes sur la sphère sont relativement complexes à mettre en oeuvre et pour bien comprendre leur phénoménologie, il serait plus simple de les approcher par une formulation comparable sur un domaine simplifié, mais une telle formulation existe-t-elle? Quel est l'apport de la méthode comparée aux approches hybride ou ensembliste? Obtient-on une variation des longueurs de portée comme on l'espère? Peut-on en estimer l'impact?

Pour résoudre cette difficulté et mieux comprendre le comportement de la formulation, un cadre simplifié est mis en place et étudié au chapitre suivant.

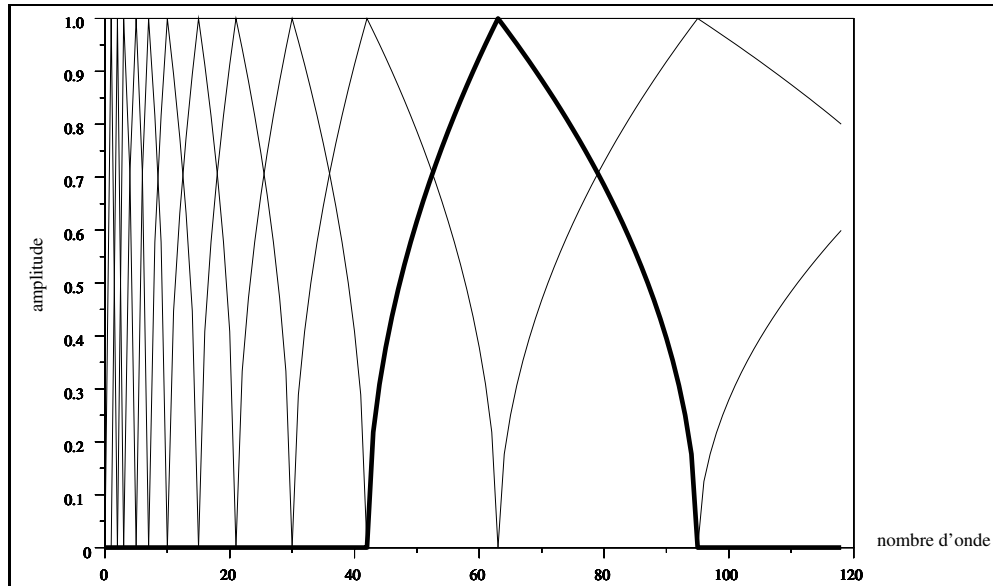


FIG. 3.4 – Spectre de Legendre des fonctions ψ_j utilisées dans la formulation de Fisher pour la sphère. C'est aussi le spectre en cosinus des fonctions ψ_j sur le cercle (étudiées au paragraphe 4.1). La troncature est de $N_T = 119$. En gras, le spectre d'une fonction ψ_j .

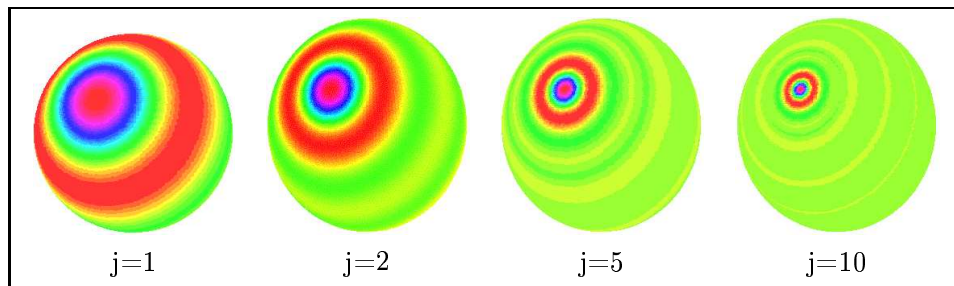


FIG. 3.5 – Représentation graphique des ondelettes sur la sphère. L'entier j caractérise la localisation de l'ondelette sur la sphère : quand j est petit, l'ondelette intéresse toute la sphère, au contraire, quand j est grand, l'ondelette est fortement localisée autour d'un centre.

Chapitre 4

Résultats et optimisation de la modélisation ondelette

L'étude de la formulation ondelette générale dans l'atmosphère est complexe car elle nécessite la prise en compte de l'ensemble des champs du modèle ainsi que de l'ensemble des niveaux. Par conséquent, la mise au point et la compréhension de l'outil ondelette ont été réalisées sur le cercle dans un contexte simplifié. Dans tout ce qui suit, un champ unique est considéré, défini sur un cercle méridien (cercle passant par les pôles du globe). Cette étude est basée aussi bien sur des champs fictifs dans une configuration modélisant schématiquement un front météorologique (pour la validation analytique), que sur des champs réels comme la pression de surface et la température à 500 hPa, utilisant des statistiques réelles de l'atmosphère suivant une méthode similaire à celles présentée au paragraphe (2.2.4), sur la période 4 février- 15 mars 2002. Le domaine considéré est un cercle méridien dont la grille correspond à une troncature $N_T = 119$. Dans les expériences réelles, la coordonnée utilisée est la latitude réelle qui varie de -90° sud à $+90^\circ$ nord, sur un demi-cercle méridien.

Cette démarche a permis de dessiner de manière progressive une formulation adaptée et complexe. Les essais successifs nous ont mis sur la voie de la compréhension et de la maîtrise de cette formulation. Les outils mis en place ont tous été implémentés avec Scilab¹. L'ensemble des opérateurs adjoints sont issus de la "méthode d'adjointisation" des codes directs (Rostaing 1993).

4.1 Formulation ondelette sur le cercle

Dans un premier temps, une étude théorique préliminaire a montré qu'il était impossible d'appliquer directement au cercle la formulation ondelette définie sur la sphère. En effet, les polynômes de Legendre trigonométriques $P_n(\cos \theta)$ ont un spectre en base cosinus étalé quelque soit l'ordre, ceci étant dû à la distribution très hétérogène des points de Gauss (zéros du polynôme P_n). La conséquence est que l'on n'obtient pas de structure de type passe bande comme celle sur la sphère, du moins pas avec la formulation des ondelettes sphériques. Cet échec nous a amenés à construire une formulation ondelette sur le cercle très proche de celle sur la sphère, en faisant jouer au spectre de Fourier sur le cercle le rôle des harmoniques sphériques sur la sphère. De même dans ce contexte, les fonctions radiales sont les fonctions paires et le spectre de Legendre pour les fonctions radiales sur la sphère est remplacé par le spectre en cosinus pour les fonctions radiales sur le cercle.

Cette transformation en ondelettes sur le cercle est également basée sur le produit de convolution et reste au niveau formel identique à celle présentée sur la sphère. En particulier, les fonctions radiales ψ_j admettent un spectre en base de cosinus à bande limité. Ainsi, le produit de convolution d'une fonction

¹Scilab est un logiciel de calcul numérique vectoriel proche de Matlab et développé par l'Inria.

radiale H (de spectre en cosinus H_n) avec un champ f (de spectre de Fourier f_p) est un champ sur le cercle ($H \otimes f$) dont le spectre est $(H \otimes f)_p = H|_p|f_p$. Une formulation ondelette est obtenue en considérant à nouveau une famille de J fonctions radiales ψ_j vérifiant $\sum_j \psi_{j,n}^2 = 1$ pour n appartenant à la troncature et avec $\psi_{j,n}$ le spectre en cosinus de ψ_j . L'ensemble du formalisme sur la sphère est alors retrouvé sur le cercle. On obtient bien une formulation ondelette. Les ondelettes utilisées sont à bande limitée, leur spectre est représenté sur la figure (3.4) et leur forme dans l'espace physique sur la figure (3.3.c).

Dans ces conditions, la matrice de covariance des erreurs de prévision en point de grille et dans la formulation de Fisher est donnée par $\mathbf{B} = \mathbf{\Sigma} \mathbf{S}_J^{-1} \left(\sum_{j=1}^J \mathbf{\Psi}_j \mathbf{S}_j \mathbf{C}_j \mathbf{S}_j^T \mathbf{\Psi}_j^T \right) \mathbf{S}_J^{-T} \mathbf{\Sigma}^T$. Ceci peut s'écrire formellement $\mathbf{B} = \mathbf{\Sigma} \mathbf{W}^{-1} \mathbf{D}_w \mathbf{W}^{-T} \mathbf{\Sigma}^T$, avec $\mathbf{\Sigma}$ la matrice diagonale des écarts types en point de grille, \mathbf{W} la matrice de la transformation en ondelettes et \mathbf{D}_w la matrice diagonale des variances en ondelettes. Cette matrice peut donc s'écrire $\mathbf{D}_w^{1/2} \mathbf{D}_w^{T/2}$.

Dans la suite, la longueur de portée en un point du cercle méridien est calculée et définie par $L_y^2 = \frac{\sigma^2(\epsilon^b)}{\sigma^2(\partial_y \epsilon^b) - (\partial_y \sigma(\epsilon^b))^2}$ (Belo-Pereira et al. 2002) avec $\sigma(\alpha)$ l'écart type associé à la variable aléatoire α , en un point du cercle ; y désigne ici la latitude sur le cercle, c'est la coordonnée méridienne dans un repère local sur la sphère.

4.2 Etapes de construction et normalisations introduites

4.2.1 Normalisation point de grille

Dans la pratique, la matrice de covariance \mathbf{B} est obtenue en modélisant sa matrice de corrélation associée. Par conséquent, une première normalisation en points de grille est réalisée. En écrivant \mathbf{B} sous la forme $\mathbf{B} = \mathbf{\Sigma} \mathbf{C}_B \mathbf{\Sigma}^T$, la matrice de corrélation s'écrit $\mathbf{C}_B = \mathbf{\Sigma}^{-1} \mathbf{B} \mathbf{\Sigma}^{-T}$. Ainsi, dans la formulation établie par Fisher, $\mathbf{C}_B = \mathbf{W}^{-1} \mathbf{D}_w^{1/2} \mathbf{D}_w^{T/2} \mathbf{W}^{-T}$. Cela donne un premier modèle pour \mathbf{B} .

4.2.2 Normalisation ondelettes

Après avoir réalisé une normalisation en points de grille, la matrice de corrélation \mathbf{C}_B est modélisée. En particulier, cette matrice est de variance unité (c'est-à-dire que sa diagonale ne comporte que des 1). Pour comprendre le comportement de la formulation ondelette, le cas analytique où la longueur de portée est variable sur le cercle a été considéré.

Cette matrice a été modélisée suivant la formulation ondelette avec le modèle du paragraphe précédent. La figure (4.1) montre la capacité de la formulation ondelette à restituer une variation des longueurs de portée, mais illustre aussi la difficulté à retrouver une carte de variance unité. En analysant le résultat de manière plus fine, il apparaît que la formulation ondelette a lissé les variations de longueur de portée. En effet, dans la zone de grande longueur de portée $Z1$, la courbe verte (ondelette) est inférieure à la courbe bleue (analytique). Ce comportement est inversé dans la zone de courte de longueur de portée $Z2$. Ainsi, la formulation ondelette lisse les variations de longueurs de portée. D'autre part, la carte d'écarts types n'est pas constante égale à 1 : les valeurs sont inférieures à l'unité, il y a donc une sous-estimation des écarts types. De plus, l'écart type admet des variations anticorrélées à la carte de longueurs de portée. Dans la zone de grande longueur de portée, l'écart type est faible, et il devient plus important dans la zone de courte longueur de portée. Pour fixer les idées, le comportement de la carte de variance (modélisée) en fonction de celle de longueur de portée (modélisée) peut s'intuiter suivant la relation

$$\text{Longueur de portée} \times \text{Ecart-type} \approx C^{ste}. \quad (4.1)$$

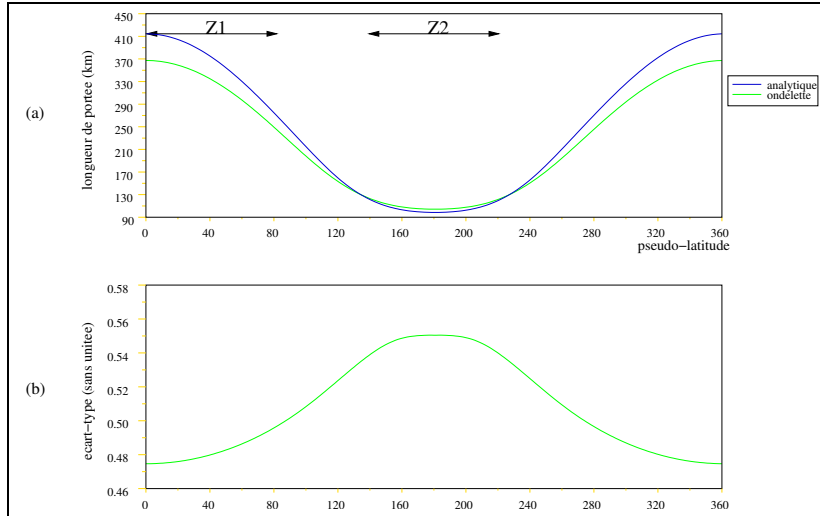


FIG. 4.1 – Représentation de la carte de longueurs de portée (a) ; en vert la carte modélisée avec les ondelettes, en bleu celle de départ. En (b) la carte d'écart types modélisée avec les ondelettes est représentée. Le résultat attendu est une valeur constante égale à 1 (variance de départ).

Cette relation ne se vérifie pas numériquement, mais elle donne une idée qualitative de la dépendance entre ces deux grandeurs. Ainsi, la formulation ondelette sur le cercle entraîne une dépendance entre les longueurs de portée et les écarts types. En particulier, ceci montre qu'il n'est pas pensable de chercher à vouloir modéliser directement la matrice de covariance : dans ce cas les ondelettes cherchent à caractériser à la fois les variations d'écart type et de longueur de portée. Ce résultat a été interprété comme étant une conséquence de la non-orthogonalité qui introduit des termes de couplage et donc une mauvaise répartition de l'information, entraînant une perte de la variance totale.

Ainsi, pour forcer la modélisation d'une matrice de corrélation, une étape de normalisation supplémentaire a été introduite. En notant Σ_w la matrice diagonale constituée par la carte d'écart type, on écrit $\mathbf{C}_B = \Sigma_w^{-1} \mathbf{W}^{-1} \mathbf{D}_w^{1/2} \mathbf{D}_w^{T/2} \mathbf{W}^{-T} \Sigma_w^{-T}$. Nous appelons normalisation ondelette la carte Σ_w de normalisation ondelette. Après cette normalisation, il en résulte bien une matrice de corrélation.

Ce cas analytique est très informatif car il préfigure ce qui apparaît dans les données réelles, telles que pour la pression de surface. En effet, la figure (4.2) montre une dépendance entre variation de longueur de portée et coefficients de normalisation ondelette.

Pour réaliser cette normalisation, il est nécessaire de connaître la carte de variance Σ_w^2 . D'après la figure (4.2.b), l'écart type utilisé pour la normalisation est quasiment constant, sa valeur est ici 0.54 ± 0.03 . Le coefficient de normalisation (inverse de l'écart type) est donc 1.85 ± 0.1 , ce qui implique une erreur de 10% sur la variance finale, à ajouter à l'erreur d'estimation de cette variance. Si cette précision n'est pas suffisante, plusieurs possibilités sont envisageables :

1. **Normalisation analytique.** Cette méthode consiste en l'évaluation explicite de la variance de la matrice $\mathbf{C}_w = \mathbf{W}^{-1} \mathbf{D}_w^{1/2} \mathbf{D}_w^{T/2} \mathbf{W}^{-T}$, ce qui revient à calculer explicitement la diagonale de la matrice \mathbf{C}_w en calculant son action sur des Diracs en point de grille. Cette méthode est difficilement envisageable en opérationnel où le nombre de degrés de liberté est de l'ordre de $\mathcal{O}(10^6)$.
2. **Normalisation analytique à basse résolution.** Cette méthode dérive de la première en utilisant une grille plus lâche adaptée à chaque champ et en effectuant une interpolation implicite par retour à une grille à pleine résolution.

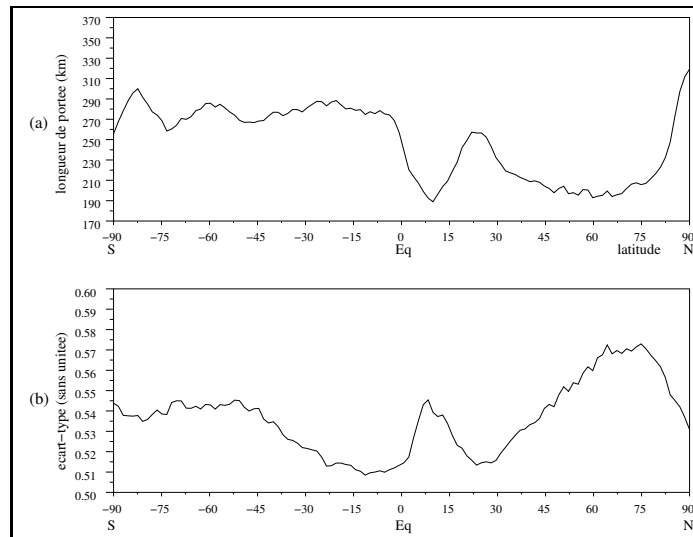


FIG. 4.2 – Représentation de l’influence de la longueur de portée (a) sur la normalisation ondelette (b), pour la pression de surface sur le méridien 0° .

3. **Normalisation statistique.** Il est possible de générer des réalisations aléatoires de champs compatibles avec une distribution de probabilité a priori, avec des vecteurs $\epsilon = \mathbf{W}^{-1} \mathbf{D}_w^{1/2} \zeta$, ζ suivant une loi normale $\mathcal{N}(0, \mathbf{I}_N)$. La normalisation Σ_w^2 n’est autre que la carte de variance des réalisations ainsi générées. Cette méthode, introduite par Fisher et Courtier (1995), est utilisée actuellement par Weaver et Ricci (2003) dans la chaîne opérationnelle du modèle d’océan OPA ².
4. **Normalisation analytique régionalisée.** Cette méthode est basée sur la dépendance entre longueur de portée et écart type modélisé. Elle consiste à appliquer la méthode de normalisation analytique en seulement quelques points caractéristiques de l’écoulement, suivant un découpage de la sphère en grandes zones de longueurs de portée équivalentes.
5. **Normalisation par modèle statistique.** En se basant à nouveau sur la dépendance entre longueur de portée et écart type, la longueur de portée est utilisée comme prédicteur de la normalisation à l’aide d’un modèle statistique simple (régression linéaire, réseaux de neurones,...).
6. **Normalisation par modèle statistique direct.** Cette méthode est semblable à la précédente mais avec comme prédicteur les variances d’ondelettes. En effet, les variances d’ondelettes correspondent à des échelles excitées et par conséquent sont caractéristiques des longueurs de portée.

Les trois dernières méthodes ne peuvent être envisagées que s’il existe sur la sphère une dépendance entre la carte de longueur de portée et celle de variance. D’autre part, il se trouve que dans les zones actives, cette dépendance existe de manière physique. Il serait ainsi envisageable de profiter de la dépendance ondelette pour caractériser à moindre coût la dépendance physique. En effet, comme il est nécessaire de réaliser une normalisation ondelette pour retrouver la carte exacte, il est possible d’imaginer que cette carte soit déduite des variances d’ondelettes (utilisées comme prédicteur) à l’aide d’un modèle statistique simple. Cette méthode permettrait d’obtenir un \mathbf{B} dépendant de l’écoulement.

²Océan PARallélisé.

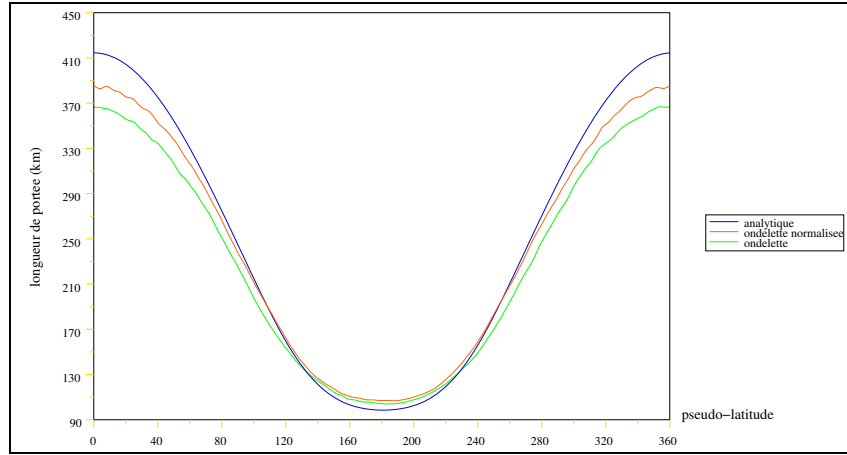


FIG. 4.3 – Visualisation de l'impact de la normalisation spectrale dans un cas analytique. La variation de longueurs de portée est représentée pour le signal analytique, le signal modélisé en ondelettes avec et sans normalisation spectrale.

4.2.3 Normalisation spectrale

Deckmyn et Berre (2004) ont introduit une normalisation supplémentaire pour optimiser le travail des ondelettes. En effet, en procédant à une normalisation spectrale, il est possible de donner plus de poids à certaines échelles en vue de la représentation, par les ondelettes, des variations géographiques associées. Cela peut améliorer la modélisation de la longueur de portée. En notant Σ_n une matrice diagonale de valeurs strictement positives, et \mathbf{S} la matrice de passage de l'espace de la grille du modèle à l'espace spectral, nous avons la formulation $\mathbf{C}_B = \Sigma_w^{-1} \mathbf{S}^{-1} \Sigma_n \mathbf{S} \mathbf{W}^{-1} \mathbf{D}_w^{1/2} \mathbf{D}_w^{T/2} \mathbf{W}^{-T} \mathbf{S}^T \Sigma_n^T \mathbf{S}^{-T} \Sigma_w^{-T}$. Cette normalisation est appelée normalisation spectrale. Il est possible d'imaginer toutes sortes de normalisations spectrales, en se donnant un spectre particulier à l'avance, ou bien en exploitant le spectre issu des statistiques d'erreurs elles-mêmes. La normalisation étudiée, basée sur les statistiques d'erreur de prévision, a pour effet de se ramener à un spectre "blanc". La figure (4.3) montre l'impact de la normalisation spectrale dans un cas analytique, et pour une normalisation à spectre "blanc". L'effet de cette normalisation est de mieux caractériser les grandes longueurs de portée.

4.2.4 Impact des normalisations pour la pression de surface, validation analytique

La formulation de la matrice \mathbf{B} , avec les normalisations introduites est donnée par

$$\mathbf{B} = \Sigma \Sigma_w^{-1} \mathbf{S}^{-1} \Sigma_n \mathbf{S} \mathbf{W}^{-1} \mathbf{D}_w^{1/2} \mathbf{D}_w^{T/2} \mathbf{W}^{-T} \mathbf{S}^T \Sigma_n^T \mathbf{S}^{-T} \Sigma_w^{-T} \Sigma^T. \quad (4.2)$$

Pour illustrer l'apport de chaque normalisation dans la modélisation ondelette, différents tests ont été réalisés. La figure (4.4) montre les effets de l'utilisation progressive des normalisations sur le signal de longueur de portée et sur celui d'écart type.

Le signal en bleu est obtenu en modélisant directement la matrice \mathbf{B} sans utiliser aucune normalisation, ce qui correspond à la formulation $\mathbf{B} = \mathbf{W}^{-1} \mathbf{D}_w \mathbf{W}^{-T}$. Il apparaît que cette formulation capte une variation de longueur de portée et de variance, mais elle le fait très mal. Plus précisément, une sur-estimation de la longueur de portée est observée dans les régions où la variance (estimée par les données brutes) est forte, ainsi qu'une sous-estimation de la longueur de portée dans les régions où la variance (brute) est faible. Cette dépendance est bridée en normalisant en point de grille, comme le montre le signal vert. Dans ce cas, les variations de longueur de portées se superposent à celles de

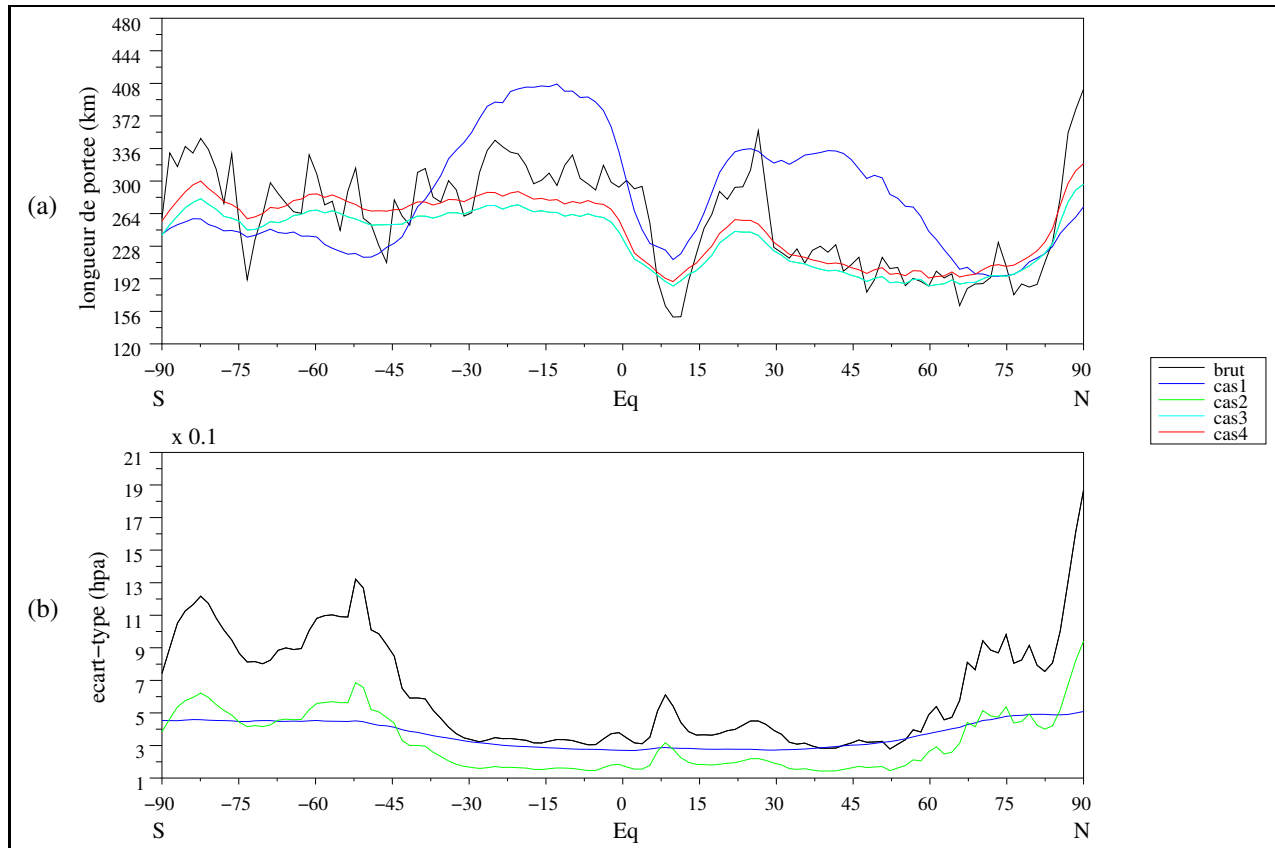


FIG. 4.4 – Visualisation de l’effet des différentes normalisations pour la pression de surface. La longueur de portée pour le méridien 0° est représentée en (a), ainsi que la carte d’écart type pour ce méridien (b). Différentes modélisations sont superposées au signal brut (noir). Cas1 : aucune normalisation (bleu). Cas2 : normalisation point de grille (vert). Cas3 : normalisation point de grille et ondelette (cyan). Cas4 : normalisation point de grille, ondelette et spectrale (rouge). Pour cas2 et cas3, les cartes de longueurs de portée modélisées se superposent. Pour cas3 et cas4, les cartes des écarts types bruts et modélisés se superposent.

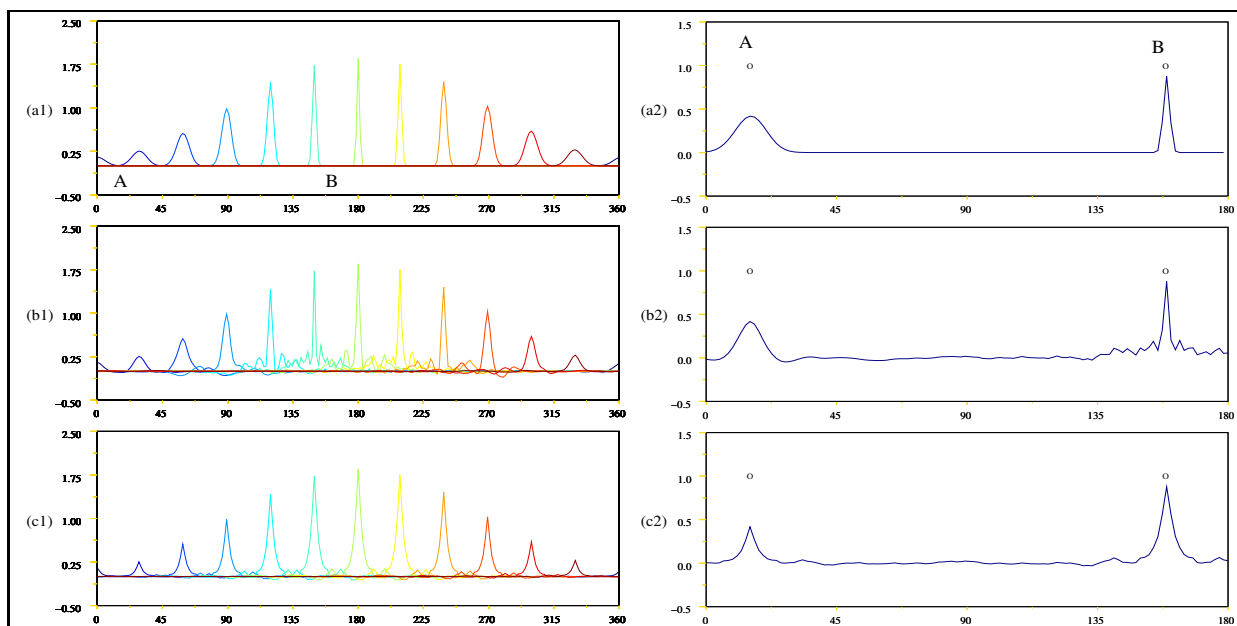


FIG. 4.5 – Suite des tests comparatifs, présentés sur la figure (3.2). La lettre pour chaque figure représente une matrice de covariance associée, dont on représente quelques fonctions de covariance en (a1,b1,c1) et l'incrément d'analyse obtenu avec cette matrice en (a2,b2,c2). La matrice de covariance de référence **B** correspond à (a1,a2), sa modélisation ondelette à (b1,b2) et enfin la modélisation hybride à (c1,c2).

la courbe cyan en (a), mais l'écart type en (b) est sous-estimé d'un facteur proche de 2 par rapport à la courbe brute noire. Ainsi, en fixant la carte de variance par la normalisation point de grille, les ondelettes n'ont dû caractériser que la variation en longueur de portée. Les ondelettes ne peuvent visiblement pas modéliser directement les 2 degrés de liberté et en fixant la variance à l'unité par la normalisation, la formulation ondelette caractérise correctement la variation en longueur de portée. Pour retrouver la carte de variance brute, il faut procéder à la normalisation ondelette, ce qui donne la courbe cyan. Cette fois, on retrouve bien la carte de variance, puisqu'en (b), les courbes noire et cyan se superposent. En observant la figure (a), il apparaît que cette dernière formule sous-estime légèrement la longueur de portée. Ce biais est corrigé en partie en utilisant la normalisation spectrale représentée par le signal en rouge.

Finalement, les conclusions de cette expérience sont très encourageantes : la formulation ondelette pour **B** caractérise correctement les variations de longueur de portée, avec un lissage de la carte brute cohérent avec les contrastes de grande échelle, tout en conservant une carte de variance exacte. C'est une première validation de la méthode. Pour obtenir une validation plus fine par rapport à la problématique exposée aux paragraphes (3.2,3.3), le cas analytique exposé au paragraphe (3.2) est à nouveau considéré. L'assimilation de 2 observations dans des régions météorologiquement différentes est représentée sur la figure (4.5). La référence est donnée par la matrice analytique dont les covariances sont données en (a1) et pour laquelle l'incrément d'analyse est donné en (a2). En (b1,b2) sont représentés les résultats obtenus avec les ondelettes, et en (c1,c2) sont reportés les résultats obtenus précédemment avec la formulation hybride. D'après (b1,b2), la carte de variance est exacte, et la formulation ondelette est effectivement capable d'intégrer la variation de longueur de portée, ce qui se traduit par une propagation grande échelle pour *A* et plus petite échelle pour *B*. Cette variation est inexistante en (c2). La formulation ondelette est par conséquent un succès.

4.3 Planisphère des portées méridiennes locales pour la pression de surface

Cette étude a considéré deux champs : la pression de surface et la température à 500hPa. Les résultats présentés sont ceux de la pression, les conclusions étant les mêmes dans le cas de la température. En utilisant la modélisation ondelette de \mathbf{B} sur le cercle, un planisphère a été reconstitué à partir des différents cercles méridiens en calculant la longueur de portée brute et celle modélisée. Les cartes de longueurs de portée résultantes sont rapportés sur la figure (4.6). Sur la carte issue de la modélisation ondelette, il apparaît des variations de longueurs de portée. L'aspect "vertical" (nord/sud) est dû au fait que les calculs ont été effectués sur des cercles et non directement sur la sphère : il s'agit ici d'une recomposition du signal sur la sphère ; de plus, les cercles ont été considérés indépendamment les uns des autres. La carte modélisée (b) est plus lisse que la carte brute (a) : les ondelettes ont filtré le bruit. Cet exemple montre le potentiel de la méthode pour le cas de la sphère.

4.4 Comparaison des modélisations dans un schéma 1D-Var

Pour comparer finement les différents modèles pour \mathbf{B} , les données réelles ne sont pas adaptées, aussi l'exemple analytique modélisant la présence d'un front dans l'écoulement est à nouveau utilisé dans la suite.

4.4.1 Protocole expérimental

L'impact de la formulation ondelette dans le cas simplifié du cercle a été estimé. Différentes expériences d'assimilation ont été mises en place, sous la forme d'une analyse 1D-Var sur le cercle. Pour des observations \mathbf{y}^o , un opérateur d'observations \mathbf{H} associé, les matrices de statistiques \mathbf{R} et \mathbf{B} et enfin une ébauche \mathbf{x}^b , l'analyse \mathbf{x}^a est obtenue suivant (2.1). La principale difficulté est que dans la réalité, il est impossible d'avoir accès à la matrice \mathbf{B} exacte mais uniquement à une estimation de cette matrice basée sur des statistiques $\mathcal{E} = (\epsilon_k^b)_{k \in [1, N]}$, suivant l'estimateur $\mathbf{B}^e = \frac{1}{N-1} \sum_{k=1}^N (\epsilon_k^b - \bar{\epsilon}_k^b)(\epsilon_k^b - \bar{\epsilon}_k^b)^T$, où $\bar{\epsilon}_k^b = \frac{1}{N} \sum_{k=1}^N \epsilon_k^b$. Ainsi $\mathbf{B} \approx \mathbf{B}^e$. La précision de l'estimation dépend du nombre d'éléments N de l'ensemble. La précision de l'estimation est en $1/\sqrt{N}$ dans le cas Gaussien.

On se place dans le cas simplifié où $\mathbf{x}^t = 0$, ainsi $\mathbf{y}^o = \epsilon^o$, $\mathbf{x}^b = \epsilon^b$ et $\mathbf{x}^a = \epsilon^a$, donnant ainsi la relation $\epsilon^a = \epsilon^b + \mathbf{K}^e (\epsilon^o - \mathbf{H}(\epsilon^b))$ avec $\mathbf{K}^e = \mathbf{B}^e \mathbf{H}^T (\mathbf{H} \mathbf{B}^e \mathbf{H}^T + \mathbf{R})^{-1}$. Pour obtenir une perturbation ϵ^b compatible avec la matrice \mathbf{B} , un vecteur est généré suivant $\epsilon^b = \mathbf{B}^{1/2} \zeta$ où ζ est une réalisation aléatoire d'un vecteur aléatoire suivant la loi Gaussienne $\mathcal{N}(0, \mathbf{I}_n)$ (n étant le nombre de points de grille). Naturellement, cette génération n'est valable que dans le cas où la matrice \mathbf{B} est connue (cas analytique). En pratique, on ne dispose pas de \mathbf{B} mais uniquement de \mathbf{B}^e . Ainsi il n'est pas possible d'utiliser les données réelles pour réaliser cette comparaison des méthodes. Dans la réalité "grandeur nature", les comparaisons peuvent être réalisées en examinant les scores (norme de la différence "prévision-analyse" pour certains niveaux caractéristiques) des prévisions obtenues à partir de l'analyse issue de différentes formulations pour \mathbf{B} .

De même, des observations sont générées avec $\epsilon^o = \mathbf{R}^{1/2} \eta$, où η est une réalisation aléatoire d'un vecteur aléatoire suivant la loi Gaussienne $\mathcal{N}(0, \mathbf{I}_p)$ (p étant le nombre d'observations), avec $\mathbf{R} = \sigma_o^2 \mathbf{I}_p$ et où σ_o est l'écart type de la mesure relative au paramètre assimilé. En particulier, les erreurs d'observation sont supposées être décorréliées.

Différentes expériences d'assimilation de données ont été réalisées pour estimer l'impact de la formulation ondelette. Le cadre théorique pour tester les formulations est le cas analytique. Dans les tests réalisés, la matrice \mathbf{B} utilisée comme matrice vraie de référence est celle déjà utilisée au cours

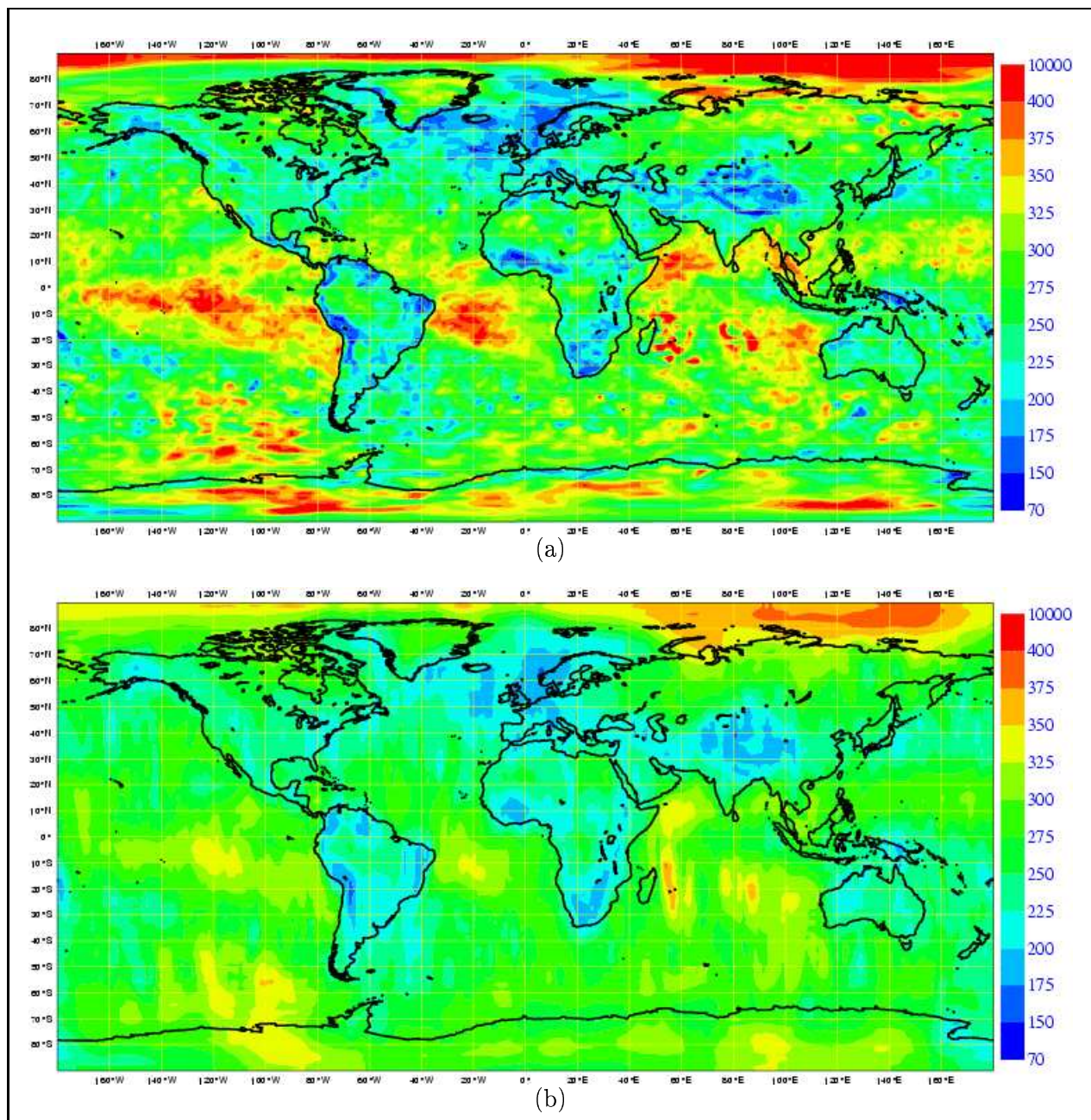


FIG. 4.6 – Carte des longueurs de portée (en km) pour la pression de surface, reconstituée à partir des cartes des cercles méridiens. Carte brute (a) et carte modélisée ondelette (b).

de ce chapitre, modélisant schématiquement la matrice de corrélation associée à un phénomène frontal localisé à la pseudo-latitude 180° . Cette matrice est notée \mathbf{B}_{ana} . Ce cas analytique est fondamental car il donne un cadre expérimental entièrement maîtrisé pour lequel il est effectivement possible de comparer les différentes formulations entre elles. A partir de \mathbf{B}_{ana} , un ensemble de 100 erreurs $\mathcal{E} = (\epsilon_k^b)$ a été généré. De cet ensemble, les matrices \mathbf{B}_w , \mathbf{B}_h et \mathbf{B}^e ont été calculées. Pour mieux comprendre l'impact de chaque formulation et effectuer des comparaisons avec la référence analytique, différentes assimilations d'un même motif d'observations sont reportées sur la figure (4.7). Cet exemple est dans la continuité de celui présenté au paragraphe (2.2.6) faisant référence à la figure (2.3). D'après la comparaison (a2)-(b2), le modèle point de grille est bon dans la zone de très petite longueur de portée $Z1$ relative à l'expérience B , mais est très mauvais dans la zone de grande longueur de portée $Z2$ relative à l'expérience C : il y est incapable de filtrer et de propager correctement l'information. Le modèle homogène (c2) effectue un filtrage trop important en B et pas assez marqué ni propagé en C . La formulation ondelette (d2) est meilleure que la précédente en filtrant de manière différente suivant les zones. Il y a moins de filtrage dans l'expérience B et plus dans l'expérience C . Les résultats restent très proches du résultat analytique de référence donné en (a2). Enfin, l'expérience avec l'estimation ensembliste \mathbf{B}^e (e2) montre une similitude avec le cas ondelettes mais avec un bruit très important en B , qu'il est impossible de filtrer simplement. Ce dernier cas est comparable à ce qu'obtiennent les méthodes d'assimilation relatives au filtre de Kalman d'ensemble.

4.4.2 Résultats statistiques

En se donnant une ébauche (compatible avec \mathbf{B}) et des observations (compatible avec \mathbf{R}), les résultats issus des différents modèles sont comparés en réalisant des statistiques sur l'erreur quadratique $(\epsilon^a)^2$. La figure (4.8) représente l'analyse obtenue par l'assimilation d'un jeu de 48 observations régulièrement réparties (environ une observation tous les 850km). Ce cas illustre simplement l'influence de la variation de longueur de portée, tant sur la structure de l'ébauche que sur celle de l'analyse. En effet, le lien entre l'ébauche et la longueur de portée apparaît, avec des variations de plus haute fréquence dans la zone de courte longueur de portée et de basse fréquence dans la zone de grande longueur de portée. En particulier, ce cas montre l'importance de générer une ébauche de manière compatible avec la matrice \mathbf{B} réelle. Pour l'analyse, dans les zones où la longueur de portée est petite (vers 180 degrés) la correction apportée à l'ébauche est d'extension faible au sens où la correction est très locale, autour de la région d'observation. A l'inverse, dans la zone où la longueur de portée est grande (vers 0 degré), la correction se répartit sur un domaine plus important. D'autre part, les principales déficiences de la modélisation homogène (cyan) se retrouvent.

Différentes expériences ont été réalisées, dont les statistiques sont visualisées sur la figure (4.9). Pour ne pas influencer les résultats, un grand nombre de réalisations a été considéré, tant sur l'ébauche (50 réalisations par expérience) que sur les observations (50 réalisations par ébauche). Dans le cas (a), l'écart type de l'erreur de prévision est fixé à $\sigma_b = 1$, et l'écart type de l'erreur d'observation est fixé à $\sigma_o = 0.95$. Ainsi, les observations sont supposées avoir une précision semblable à celle de l'ébauche, ce qui est le cas dans la réalité. Sur la figure (a), il apparaît que la formulation ondelette est meilleure que la formulation homogène (ceci est dû à la variation de longueur de portée qui n'est pas vue dans la formulation homogène) et proche de la formulation basée sur l'ensemble. Dans tous les cas, ces deux dernières formulations sont très proches du cas analytique, ce qui montre leur efficacité. A ce stade, il n'est cependant pas nécessaire de quantifier davantage le gain de chaque méthode. En effet, c'est le cas de l'atmosphère qui est intéressant et non celui du cercle analytique.

Une autre expérience a été menée : elle simule le cas où l'on augmente le nombre d'observations pour que la formulation de \mathbf{B} devienne moins cruciale. Assimiler un grand nombre de N observations d'erreurs de mesure d'écart type σ_o , et suivant une loi non nécessairement Gaussienne, c'est assimiler

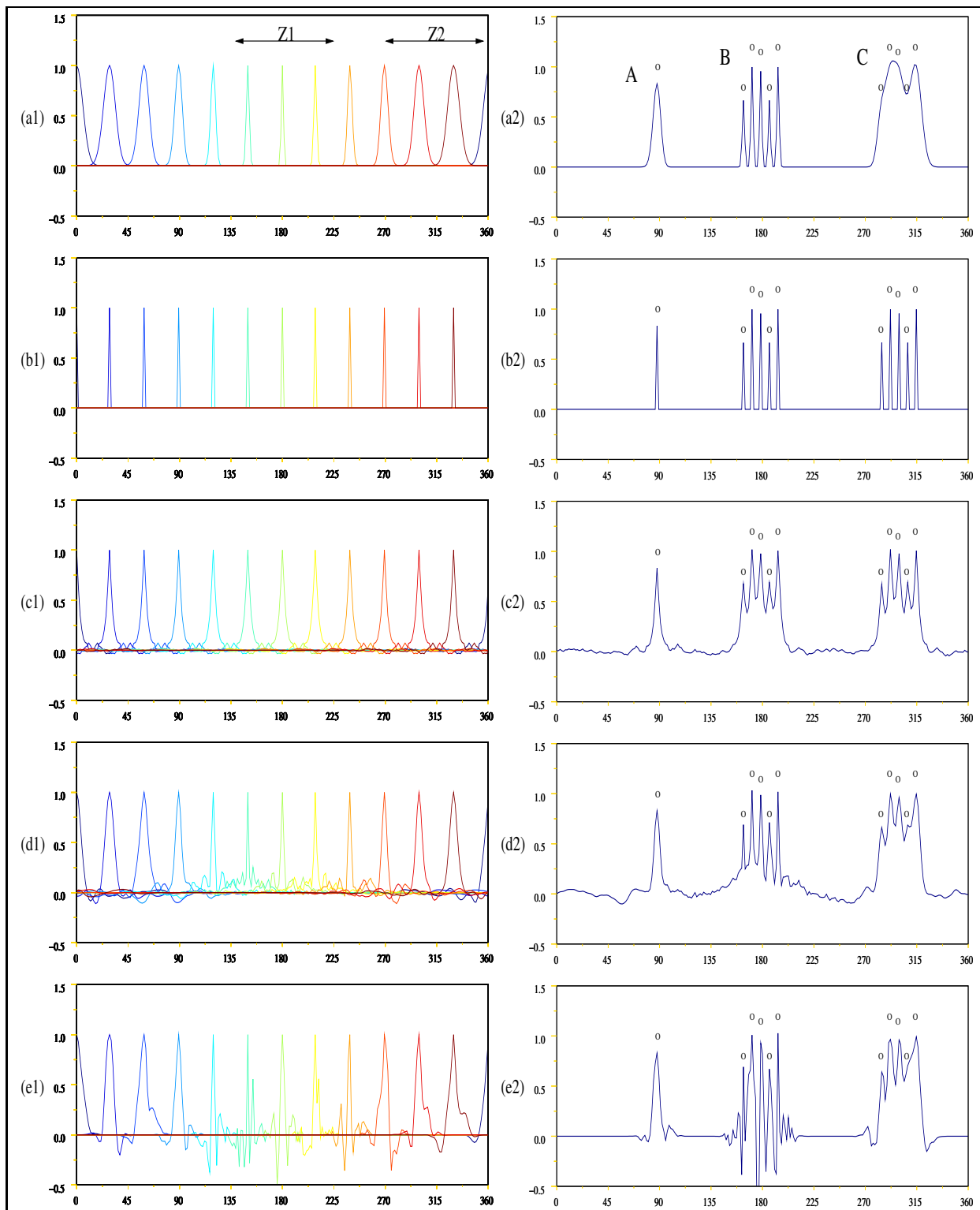


FIG. 4.7 – Assimilation 1D-Var pour un motif d'observations donné. La lettre pour chaque figure représente une matrice de covariance associée dont on représente quelques fonctions de covariance en (a1,b1,c1,d1,e1) et l'incrément d'analyse obtenu avec cette matrice en (a2,b2,c2,d2,e2). La matrice de covariance de référence \mathbf{B}_{ana} est en (a1,a2), sa modélisation point de grille en (b1,b2), sa modélisation homogène en (c1,c2), sa modélisation ondelette en (d1,d2) et son estimation ensembliste \mathbf{B}^e en (e1,e2).

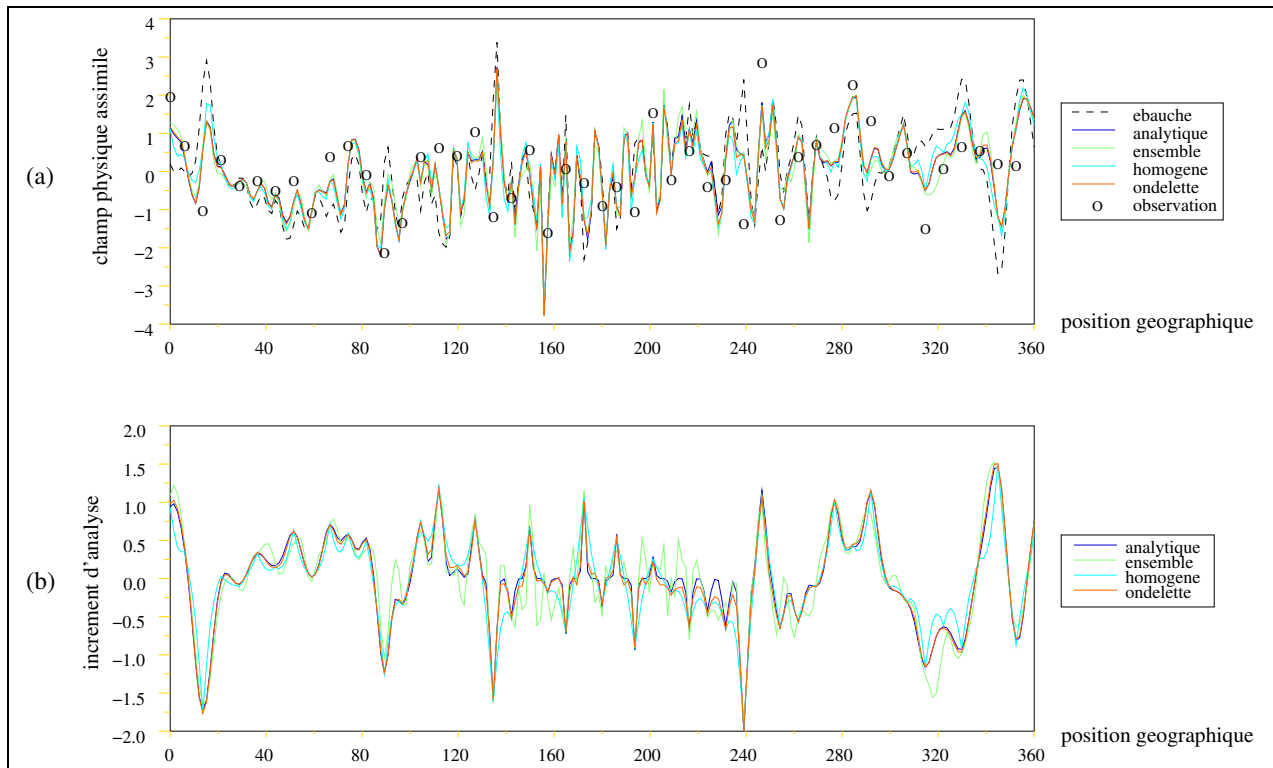


FIG. 4.8 – Exemple d’une assimilation 1D-Var (type 3D-Var pour un problème 1D). L’ébauche est représentée en tiretés en (a), ainsi que les différentes analyses obtenues en assimilant 48 observations réparties régulièrement sur le cercle. En (b), l’incrément d’analyse (différence $\mathbf{x}^a - \mathbf{x}^b$) est reporté pour les différents modèles utilisés.

la valeur moyenne de ces observations qui admet cette fois-ci une distribution Gaussienne (d’après le théorème de la limite centrale), dont l’écart type varie en $\frac{1}{\sqrt{N}}\sigma_o$. Ainsi, diminuer σ_o dans la première expérience, c’est simuler un grand nombre d’observations. Le résultat de ces simulations est donné sur la figure (b) pour $\sigma_o = 0.1$ (équivalent à $N = 100$ d’après $\frac{1}{\sqrt{N}}\sigma_o = 0.1 \times 0.95 \approx 0.1$). On constate qu’il faut un très grand nombre d’observations (environ 225 au total pour $\sigma_o = 0.1$, ce qui correspond à 100 observations par point de grille pour $\sigma_o \approx 1$) pour que le modèle de \mathbf{B} n’influence plus l’analyse, et ne soit donc plus utile.

4.4.3 Conclusion

D’après les résultats précédents, il vient que la méthode ondelette est très performante, proche des cas analytique et ensembliste. La formulation homogène quand à elle montre des faiblesses qui ne disparaissent que quand le nombre d’observations est très grand. Naturellement, la formulation ondelette prend en compte plus d’informations et admet un coût mémoire supérieur à la formulation homogène. En effet, la formulation ondelette sur le cercle est 10 fois plus coûteuse que la formulation homogène. Ce coût supplémentaire est accessible, la preuve en est l’implémentation d’une première version de la méthode dans le code ARPEGE/IFS et réalisée par Fisher. Tout en s’intégrant dans une chaîne 4D-Var, la formulation ondelette permet de caractériser des structures d’erreur variables géographiquement sur la sphère et pouvant évoluer dans le temps de manière cohérente avec la physique. Ce modèle pour la matrice \mathbf{B} semble donc très prometteur.

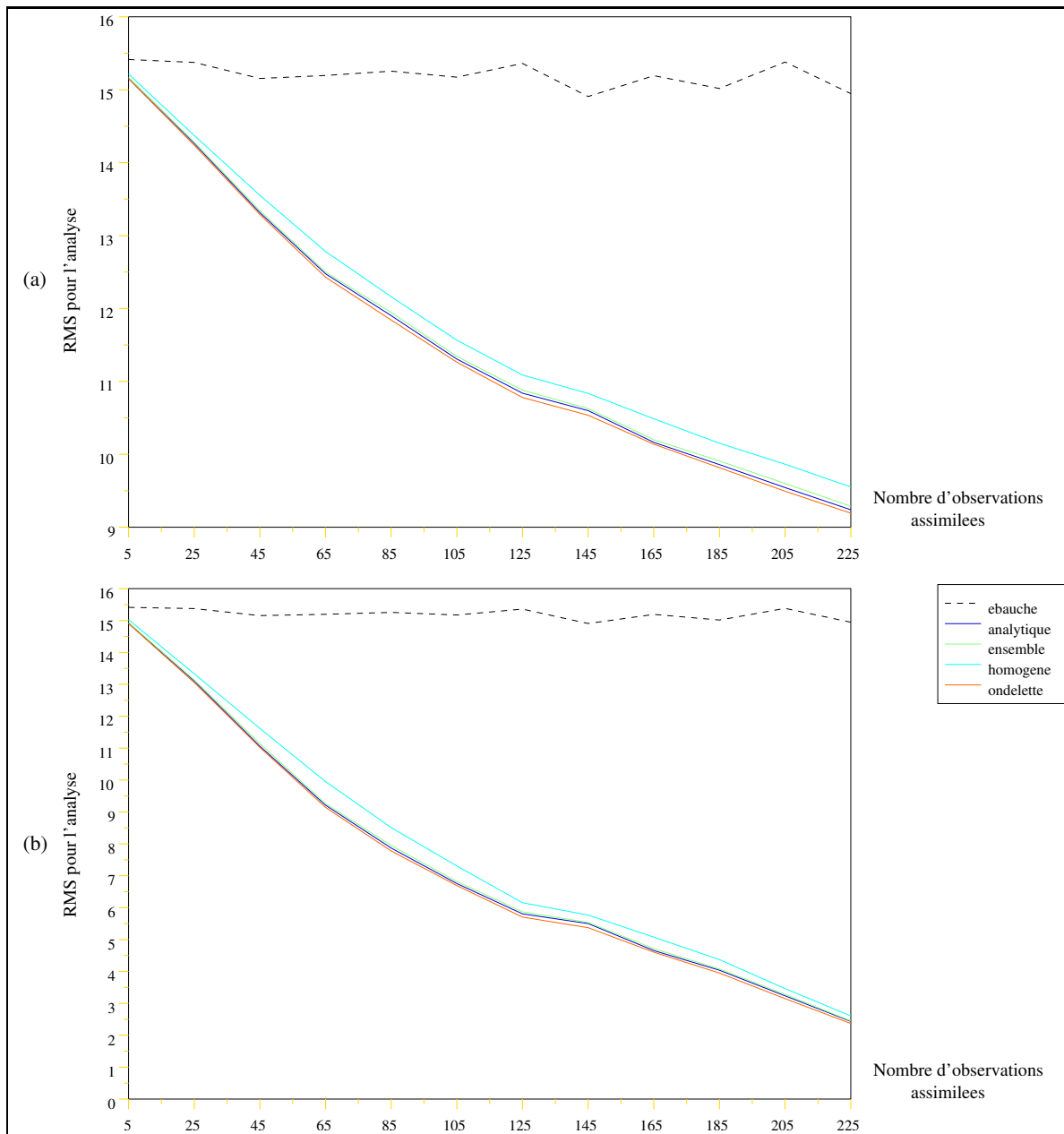


FIG. 4.9 – Les graphiques (a) et (b) représentent les erreurs quadratiques moyennes comparatives des assimilations obtenues avec les différentes méthodes. Ces résultats sont obtenus en calculant la norme de ϵ^a en moyenne sur 50 réalisations aléatoires de l'ébauche (générées à l'aide du $\mathbf{B}^{1/2}$ analytique) et 50 réalisations aléatoires d'un jeu d'observations par ébauche. Le nombre d'observations assimilé varie de 5 à 225 par pas de 20. Sur cette figure est représentée la moyenne de la norme de l'erreur ϵ^a pour l'ensemble. Dans ces expériences, l'écart type des erreurs de prévision est constant et vaut $\sigma_b = 1$. En (a), l'écart type des erreurs d'observation est fixé à $\sigma_o = 0.95$. En (b), l'écart type est de $\sigma_o = 0.1$, ce qui revient à considérer que l'on a assimilé $N = 100$ observations par point de grille, là où n'y en avait qu'une en (a), avec un écart type de 0.95.

Chapitre 5

Conclusion

L'objectif de ce stage de fin d'études/DEA était de tester une nouvelle modélisation de la matrice de covariance d'erreur de prévision. La formulation des ondelettes sur la sphère a été étudiée en détail ainsi que son utilisation dans la formulation proposée par Fisher. Cette première phase a constitué l'approche théorique nécessaire à la bonne maîtrise de la formulation ainsi qu'à sa compréhension détaillée. Par la suite, en se plaçant dans un cadre simplifié, une formulation équivalente sur le cercle a été étudiée, permettant des simulations rapides et peu coûteuses, idéales pour optimiser un travail à la fois théorique et appliqué.

Un certain nombre de difficultés ont été rencontrées, amenant à introduire des éléments supplémentaires dans la formulation d'origine afin de résoudre ces problèmes. Ainsi, une progression parallèle, à la fois sur la compréhension théorique et sur l'efficacité pratique, a permis d'aboutir à une version avancée de la formulation.

A l'aide d'un certain nombre de normalisations, la modélisation ondelette de \mathbf{B} permet d'obtenir une formulation capable de capter les hétérogénéités de grande échelle des structures, améliorant sensiblement les scores par rapport à la formulation homogène actuelle (d'après les résultats analytiques obtenus sur le cercle).

Désormais, et grâce à cette étude, une méthode efficace pour la modélisation de la matrice \mathbf{B} est potentiellement disponible. L'une des difficultés est maintenant de réaliser en pratique la normalisation ondelette introduite. Pour cela, il est important de reconstituer ces expériences dans le cas de la sphère, et d'y tester les méthodes envisagées dans le cas du cercle. D'autre part, un certain nombre de propriétés des ondelettes ont été mises en évidence, en particulier celle du filtrage de l'information utile. Ce point fait l'objet d'une publication indépendante en cours. Enfin, il est à noter que l'une des applications de cette formulation ondelette peut être de rendre la matrice \mathbf{B} dépendante de l'écoulement en spécifiant de manière appropriée les variances dans l'espace des ondelettes. Ainsi, une matrice caractérisant de manière fine les structures locales de plus ou moins grande échelle, telles que les perturbations atlantiques (pour un jour donné ou même un type de temps), serait disponible.

D'autre part, la formulation ondelette dans sa version actuelle est adaptative et pourrait être utilisée à des degrés de précision échelonnés. Il est possible d'imaginer une spécification de \mathbf{B} très locale dans l'hémisphère nord et moins locale dans l'hémisphère sud (avec un coût mémoire identique). Ceci peut permettre d'une part d'optimiser les prévisions sur l'Europe fournies par le modèle global ARPEGE, et d'autre part de rendre plus précise l'initialisation des grandes échelles des modèles à aire limitée pour la très haute résolution. En outre, une telle approche peut contribuer à l'étude des perturbations atlantiques dans un système d'assimilation adaptative enrichie, s'intégrant dans un contexte général pour l'étude des systèmes frontaux.

Cette formulation ondelette offre de nouvelles perspectives dans la modélisation de \mathbf{B} , et plus largement dans la prévision des phénomènes météorologiques, ultime préoccupation du prévisionniste.

Bibliographie

- [1] M. Belo-Pereira, L. Berre, and G. Desroziers. Estimation et études des covariances d'erreur de prévision d'arpege/aladin. *Atelier de modélisation de l'atmosphère, Météo-France/CNRM*, pages 17–20, 2002.
- [2] S.E. Cohn. Dynamics of short-term univariate forecast error covariance. *Monthly Weather Review*, 121 :3030–3038, november 1993.
- [3] P. Courtier and J.F. Geleyn. A global numerical weather prediction model with variable resolution : Application to the shallow-water equations. *Q.J.R Meteorol. Soc.*, 114 :1321–1346, 1988.
- [4] P. Courtier, J.-N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Q.J.R Meteorol. Soc.*, 120 :1367–1387, 1994.
- [5] R. Daley. *Atmospheric Data Analysis*. Cambridge University Press, 1991.
- [6] A. Deckmyn and L. Berre. A wavelet approach to representing background error covariances in a LAM. *Monthly Weather Review*, submitted, 2004.
- [7] J. Derber and F. Bouttier. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, 5 :195–221, 1999.
- [8] G. Desroziers. A coordinate change for data assimilation in spherical geometry of frontal structure. *Monthly Weather Review*, 125 :3030–3038, november 1997.
- [9] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error. *J. Geophys. Res.*, 99, 1994.
- [10] M. Fisher. Background error covariance modelling. *ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean*, pages 45–63, 8-12 september 2003.
- [11] M. Fisher and P. Courtier. Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *Technical Memorandum*, 220, 1995.
- [12] W. Freeden and S. Schreiner. Orthogonal and non-orthogonal multiresolution analysis, scale discrete and exact fully discrete wavelet transform on the sphere. *Constructive Approximation*, 14 :493–515, 1998.
- [13] W. Freeden and U. Windheuser. Spherical wavelet transform and its discretization. *Advanced in Computational Mathematics*, 5 :51–94, 1996.
- [14] P.L. Houtekamer and al. A system simulation approach to ensemble prediction. *Monthly Weather Review*, 124 :1225–1242, 1996.
- [15] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME Ser. D, J. Basic Eng.*, 82D :35–45, 1960.
- [16] E.N. Lorentz. Deterministic non-periodic flow. *J. Atmos. Sci.*, 20 :130–141, 1963a.
- [17] E.N. Lorentz. The predictability of hydrodynamic flow. *Trans. NY Acad. Sci. Series II*, 25 :409–432, 1963b.

- [18] N. Rostaing-Schmidt. *Différentiation automatique : Application à un problème d'optimisation en météorologie*. PhD thesis, Université de Nice-Sophia Antipolis, décembre 1993. disponible INRIA.
- [19] B. Torresani. Position-frequency analysis for signals defined on spheres. *Signal Processing*, 43 :341–346, 1995.
- [20] V. Volterra. Fluctuations in the abundance of species considered mathematically. *Nature*, 118 :558–560, 1926.
- [21] A. Weaver and P. Courtier. Correlation modelling on the sphere using a generalized diffusion equation. *Q.J.R Meteorol. Soc*, 127 :1815–1846, 2001.
- [22] A. Weaver and S. Ricci. Constructing a background-error correlation model using generalized diffusion operators. *Workshop*, 2003.