



**HAL**  
open science

# Régularisation de problèmes inverses à l'aide de l'équation de diffusion, avec application à l'assimilation variationnelle de données océaniques

Isabelle Mirouze

► **To cite this version:**

Isabelle Mirouze. Régularisation de problèmes inverses à l'aide de l'équation de diffusion, avec application à l'assimilation variationnelle de données océaniques. Mathématiques [math]. Université Paul Sabatier - Toulouse III, 2010. Français. NNT: . tel-00525590

**HAL Id: tel-00525590**

**<https://theses.hal.science/tel-00525590>**

Submitted on 12 Oct 2010

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

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



Université  
de Toulouse

# THÈSE

En vue de l'obtention du  
**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**

**Délivré par :**  
Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

**Discipline ou spécialité :**  
Mathématiques Appliquées

---

**Présentée et soutenue par :**  
Isabelle MIROUZE

**le :** vendredi 17 septembre 2010

**Titre :**

Régularisation de problèmes inverses à l'aide de l'équation de diffusion,  
avec application à l'assimilation variationnelle de données océaniques

---

**Ecole doctorale :**  
Mathématiques Informatique Télécommunications (MITT)

**Unité de recherche :**  
CERFACS / UMR-5219

**Directeur(s) de Thèse :**  
Directeur : Serge GRATTON  
Co-directeur : Anthony WEAVER

**Rapporteurs :**  
Yannick TREMOLET  
Eric BLAYO

**Autre(s) membre(s) du jury**  
Co-directeur : Luca AMODEI  
Président : Pierre MARECHAL  
Examineur : Arthur VIDARD  
Examineur : Kristian MOGENSEN



---

# Remerciements

Il est somme toute un peu curieux que la première page d'un mémoire soit en général la dernière à être rédigée. Mais je me plie ici volontiers à la tradition des remerciements. Abandonner ce que l'on a mis des années à construire professionnellement pour s'engager dans une voie résolument différente n'est pas un choix anodin. C'est pourtant celui que j'ai fait parce qu'aimer ce que l'on fait dans la vie, c'est essentiel. Après cinq années difficiles (deux ans de master plus 3 années de doctorat), je reste persuadée plus que jamais que j'ai fait le bon choix, même si cest à l'avenir de le confirmer. Pourtant ces années furent bien plus ardues que je ne l'avais pensé au départ, tant il m'apparaît aujourd'hui que j'ai négligé certains aspects essentiels qui font de nous ce que l'on est. Mais ces années furent aussi l'occasion exceptionnelle de découvrir un monde différent de celui que je connaissais, et d'y côtoyer des personnes formidables venant des quatre coins du globe<sup>1</sup>. A tous ceux qui m'ont soutenue, je vous suis reconnaissante et vous demande de me pardonner si je ne vous cite pas expressément dans ce qui suit.

Parce qu'il ne peut y avoir de thèse sans ceux qui acceptent de la juger, je souhaiterais tout d'abord remercier les membres de mon jury. A commencer par mes deux rapporteurs, Yannick Trémolet et Eric Blayo, qui, malgré leurs nombreuses obligations, ont accepté la charge supplémentaire d'évaluer mon travail. A Pierre Maréchal qui a bien voulu présider ce jury, et à Arthur Vidard et Kristian Mogensen qui ont fait office d'examineurs, merci. Mais il ne peut non plus y avoir de thèse sans ceux qui acceptent de l'encadrer. Alors à mon directeur Serge Gratton et mes deux co-directeurs Anthony Weaver et Luca Amodei, merci aussi, pour l'aide et le soutien que vous m'avez apporté durant ces trois années. Je vous remercie également pour le soin que vous avez pris à relire mon mémoire et pour la pertinence de vos commentaires qui m'ont permis de tant l'améliorer. Merci aussi à Claire Chauvin et Andrea Piacentini dont la relecture attentive de ce manuscrit fut d'une aide précieuse pour mettre un tant soit peu de cohérence dans l'assemblage de mots qu'il représentait.

Un grand merci également à Sophie Ricci et Nicolas Daget pour l'aide inestimable qu'ils m'ont apportés à mes débuts, ainsi qu'à Elisabeth Rémy pour toutes nos discussions. A Luc Giraud et Xavier Vasseur pour tous leurs conseils avisés. A Arnaud Fosso Pouangué pour nos pauses ti Kfé et pour son amitié, en particulier dans les moments de ras-le-bol. A Thomas Pangaud qui a partagé un bureau avec moi durant ma dernière année et a supporté avec un stoïcisme remarquable, mes monologues et comportements un tantinet insolites parfois. A Olivier Titaud, arrivé sur le tard de ma thèse, et qui s'est bien rattrapé sur la quantité de questions qu'il a pu me poser, mais m'aidant se faisant à clarifier mes idées pour ma soutenance. A mon docteur personnel Chantal Nasri, qui m'a prescrit une minute minimum de rigolade par jour, et s'est souvent assurée personnellement du bon suivi de sa prescription. Enfin à ma famille (étendue) qui ne m'a jamais questionnée sur mes décisions parfois acrobatiques.

Pour en terminer, il est trois personnes que je tiens à remercier particulièrement pour l'influence qu'ils ont eu sur mon parcours. Par ordre chronologique, je commencerai par Fabien Durand, qui m'a initiée à l'océanographie. Tuteur de mon stage de master 1, il a toujours pris la peine de répondre à mes questions, même basiques, avec beaucoup de patience et a guidé avec beaucoup de diligence mes premiers pas dans le monde de la recherche. Gérald Desroziers ensuite, que j'ai eu le privilège d'avoir comme professeur en assimilation de données. Par la qualité de ses cours, il est à l'origine de

---

<sup>1</sup>Mon prochain sujet de thèse : méthode de définition des coins d'une patatoïde

mon engouement pour cette discipline, et a su piqué ma curiosité pour cette matrice  $\mathbf{B}$  si mystérieuse. *And last but not least*, ma gratitude est toute acquise à Anthony Weaver qui m'a supporté (dans tous les sens du terme) pendant ces trois ans. Pour son enthousiasme communicatif, nos discussions sans fin, son aide et ses conseils précieux. Mais surtout pour son soutien indéfectible, y compris dans mes moments difficiles et mes crises de (mauvaise) foi. Dans ce que ce travail de thèse a de meilleur, une grande part lui en est attribuable.

Merci enfin à tous ceux que je n'ai pas nommé, mais qui ont fait parti de mon quotidien tout au long de ces années, au détour d'un couloir, d'une table de cantine ou d'une conférence, ici ou ailleurs.

---

# Résumé

## Régularisation de problèmes inverses à l'aide de l'équation de diffusion, avec application à l'assimilation variationnelle de données océaniques

En assimilation de données, les modèles de corrélation permettent de caractériser les structures d'erreurs pour les variables définies sur une grille numérique. L'équation de diffusion fournit un cadre flexible et efficace pour représenter des fonctions de corrélation pour des problèmes de grande dimension tels que ceux rencontrés en assimilation variationnelle pour l'atmosphère ou l'océan.

Dans cette thèse, une formulation implicite est d'abord étudiée en détail en dimension un (1D). On montre qu'intégrer une équation de diffusion implicite à coefficient constant sur  $M$  pas de temps est équivalent à convoluer la condition initiale à une fonction autorégressive (AR) d'ordre  $M$ . L'échelle de corrélation de la fonction AR et le facteur de normalisation requis pour générer une amplitude égale à 1 sont donnés en fonction du coefficient de diffusion et de  $M$ . Des extensions du modèle de diffusion permettant aux fonctions de corrélation de ne pas être affectées par les frontières, et tenant compte de variations des échelles sont décrites. Une approximation des facteurs de normalisation est alors proposée.

Des produits d'opérateurs de diffusion implicite 1D sont ensuite utilisés pour construire des modèles de corrélation en dimension deux et trois pour des configurations globales d'un système d'assimilation variationnelle pour le modèle océanique NEMO. Leurs performances sont comparées au modèle de diffusion explicite existant, et des exemples de structures de corrélation où les échelles sont soit paramétrées, soit issues d'une méthode d'ensemble, sont montrés. Enfin, les performances de différentes techniques de normalisation sont comparées.

Mots clés :

assimilation • variationnelle • covariances • corrélations • opérateurs • diffusion • normalisation.



# Abstract

## Regularization of inverse problems using a diffusion equation, with application to variational ocean data assimilation

Correlation models are required in data assimilation to characterize the error structures of variables defined on a numerical grid. The diffusion equation provides a flexible and efficient framework for representing correlation functions for problems of large dimension such as those encountered in variational atmospheric or ocean data assimilation.

In this thesis, an implicit formulation of the diffusion equation is first analyzed in detail for the one-dimensional (1D) case. It is shown that integrating a constant-coefficient implicit diffusion equation over  $M$  time steps is equivalent to convolving the initial condition with an  $M$ -th order autoregressive (AR) function. The correlation length scale of the AR function and the normalization factor required for generating a unit amplitude are given in terms of the diffusion coefficient and  $M$ . Extensions of the diffusion model to allow for correlation functions that are not affected by solid boundaries, and that account for varying length scales are described. An approximation of the normalization factors is then proposed.

Products of 1D implicit diffusion operators are then used for constructing two- and three-dimensional correlation models for global configurations of a variational assimilation system for the NEMO ocean model. Their efficiency are compared to the existing explicit diffusion model, and examples of correlation structures are shown, where the length scales are either parametrized or estimated using an ensemble method. Finally, the efficiency of different normalization techniques are compared.

Keywords :

assimilation • variational • covariances • correlations • operators • diffusion • normalization.



# Table des matières

<b>Introduction</b>	<b>15</b>
<b>1 Un problème inverse : l'assimilation de données</b>	<b>19</b>
1.1 Les problèmes inverses . . . . .	19
1.2 L'assimilation de données en océanographie . . . . .	21
1.2.1 Le modèle . . . . .	21
1.2.2 Les observations . . . . .	22
1.2.3 Comparaison modèle / observations . . . . .	25
1.2.4 Le problème inverse . . . . .	26
1.3 Les différentes techniques pour l'océanographie et la météorologie . . . . .	27
1.3.1 Un bref historique . . . . .	27
1.3.2 L'assimilation aujourd'hui . . . . .	29
<b>2 L'assimilation variationnelle</b>	<b>31</b>
2.1 Formalisme du problème . . . . .	31
2.1.1 Problème continu . . . . .	31
2.1.2 Problème discret . . . . .	32
2.1.3 Définition de la norme . . . . .	33
2.2 Les différentes formulations . . . . .	36
2.2.1 La formulation du 4D-Var . . . . .	36
2.2.2 La formulation incrémentale . . . . .	38
2.2.3 Le 3D-Var, une simplification du 4D-Var . . . . .	42
2.2.4 La formulation duale : 4D-PSAS et 3D-PSAS . . . . .	43
2.3 Les matrices de covariances d'erreur . . . . .	44
2.3.1 Nature des erreurs . . . . .	44
2.3.2 Les erreurs d'observations . . . . .	46
2.3.3 Les erreurs d'ébauche . . . . .	46
2.4 Résolution du problème variationnel . . . . .	47
2.4.1 Les différents espaces . . . . .	47
2.4.2 Minimisation de la fonction coût . . . . .	49
2.4.3 Le calcul du gradient en pratique . . . . .	51
<b>3 Modélisation de la matrice des covariances d'erreur d'ébauche</b>	<b>57</b>
3.1 Le problème fondamental . . . . .	57
3.1.1 Modélisation de $\mathbf{B}^{-1}$ . . . . .	57
3.1.2 Changement de variables . . . . .	60
3.2 Estimation des statistiques de $\mathbf{B}$ . . . . .	62
3.2.1 Méthodes basées sur les observations . . . . .	63
3.2.2 La méthode NMC . . . . .	65
3.2.3 La méthode d'ensemble . . . . .	65
3.3 Modélisation de $\mathbf{B}$ par une suite d'opérateurs . . . . .	68
3.3.1 D'un problème multivarié à plusieurs problèmes univariés . . . . .	68

3.3.2	Modélisation des covariances univariées . . . . .	70
3.4	Les différentes hypothèses diagonales . . . . .	70
3.5	Spécification des variances . . . . .	74
3.6	Spécification des fonctions de corrélation . . . . .	75
3.6.1	Evaluation des intégrales de convolution . . . . .	76
3.6.2	Le filtre récursif . . . . .	77
3.6.3	L'équation de diffusion . . . . .	79
3.6.4	La normalisation . . . . .	80
<b>4</b>	<b>Opérateur de diffusion implicite 1D</b>	<b>83</b>
4.1	Introduction . . . . .	87
4.2	Background theory . . . . .	89
4.2.1	Covariance functions and covariance operators . . . . .	89
4.2.2	Correlation functions and correlation operators . . . . .	91
4.3	Correlation operators and the diffusion equation . . . . .	92
4.3.1	Explicit and implicit formulations of the 1D diffusion equation . . . . .	92
4.3.2	The kernel of the implicit diffusion operator . . . . .	95
4.3.3	Characteristics of the AR functions . . . . .	97
4.3.4	Combinations of AR functions . . . . .	100
4.4	Inhomogeneous and anisotropic extensions . . . . .	103
4.4.1	Location-dependent length scales . . . . .	103
4.4.2	Boundary conditions . . . . .	107
4.5	Numerical aspects . . . . .	111
4.5.1	Curvilinear coordinates . . . . .	111
4.5.2	Discretization and solution algorithm . . . . .	111
4.5.3	Link with the recursive filter . . . . .	114
4.6	Summary and discussion . . . . .	115
4.A	Matérn and AR functions . . . . .	117
4.B	Recovering the Gaussian solution near boundaries . . . . .	118
<b>5</b>	<b>Constructions 2D et 3D à partir de l'opérateur de diffusion implicite 1D</b>	<b>121</b>
5.1	Le système NEMOVAR . . . . .	121
5.2	Construction d'opérateurs de corrélation 2 et 3×1D . . . . .	123
5.2.1	Fonctions de corrélation séparables . . . . .	123
5.2.2	Sensibilité au nombre d'itérations . . . . .	126
5.2.3	Non séparabilité près des frontières . . . . .	127
5.2.4	Non séparabilité due aux variations des échelles de corrélation . . . . .	131
5.3	Application au système NEMOVAR . . . . .	133
5.3.1	Aspects numériques . . . . .	133
5.3.2	Tests de la boucle interne . . . . .	138
5.3.3	Expérience d'assimilation . . . . .	143
5.3.4	Variation géographique des échelles . . . . .	145
5.4	Précision de la normalisation . . . . .	147
5.4.1	Les différentes méthodes de normalisation . . . . .	147
5.4.2	Simulation de frontières transparentes . . . . .	149
5.4.3	Méthode d'évaluation . . . . .	150
5.4.4	Méthodes par vecteurs aléatoires . . . . .	151
5.4.5	Méthodes par approximation du facteur théorique . . . . .	152
<b>6</b>	<b>Conclusions et perspectives</b>	<b>159</b>
	<b>Bibliographie</b>	<b>162</b>

---

<b>A</b>	<b>Analyse fonctionnelle</b>	<b>173</b>
A.1	Rappel sur les espaces . . . . .	173
A.2	Formes linéaires et espace dual . . . . .	174
A.3	Opérateurs linéaires et adjoints . . . . .	175
<b>B</b>	<b>Probabilités et statistiques</b>	<b>177</b>
B.1	Variables aléatoires . . . . .	177
B.1.1	Définitions . . . . .	177
B.1.2	Moments . . . . .	178
B.1.3	Loi de Gauss . . . . .	179
B.1.4	Probabilités conditionnelles . . . . .	180
B.2	Champs aléatoires et opérateurs statistiques . . . . .	181
<b>C</b>	<b>Estimateur de la moyenne conditionnelle</b>	<b>183</b>
<b>D</b>	<b>Minimisation par la méthode du gradient conjugué</b>	<b>187</b>
D.1	Méthode de la plus forte descente . . . . .	188
D.2	Méthode du gradient conjugué . . . . .	189
D.3	Gradient conjugué préconditionné . . . . .	192
<b>E</b>	<b>Méthode de normalisation proposée par Purser <i>et al.</i> (2003b)</b>	<b>193</b>
<b>F</b>	<b>Réorganisation des processeurs</b>	<b>197</b>
F.1	Réorganisation zonale . . . . .	198
F.2	Réorganisation méridienne . . . . .	199



# Table des figures

1.1	Modélisation d'un système physique . . . . .	20
1.2	Réseau d'observations océaniques <i>in situ</i> du programme TOGA en décembre 1994. . .	23
1.3	Déploiement de profileurs Argo au 3 janvier 2010. . . . .	24
1.4	Réseau d'observations satellite pour la météorologie et l'océanographie de la NASA. .	25
2.1	Système S. . . . .	31
2.2	Le système 4D-Var incrémental . . . . .	40
2.3	Les itérations du 4D-Var multi-incrémental . . . . .	40
2.4	Le système 3D-Var . . . . .	42
2.5	Structure des erreurs à partir de variables aléatoires . . . . .	45
2.6	Les différents espaces et leurs relations. . . . .	48
3.1	Les différentes techniques d'estimation/modélisation de $\mathbf{B}$ . . . . .	58
3.2	Estimation de $\mathbf{B}$ à partir des observations . . . . .	64
3.3	Estimation de $\mathbf{B}$ à partir d'un ensemble . . . . .	66
3.4	Modélisation de $\mathbf{B}$ près des frontières . . . . .	75
4.1	Shape of the AR functions . . . . .	99
4.2	Power spectrum of the AR functions . . . . .	99
4.3	Normalization coefficient . . . . .	100
4.4	Correlation function built from the difference of two SOAR functions . . . . .	103
4.5	Power spectrum of the correlation function built from the two SOAR functions . . . .	103
4.6	Result of a 2-step implicit diffusion operator with variable coefficients . . . . .	106
4.7	Comparison of different normalization factors for variable diffusion coefficients . . . .	107
4.8	Exact normalization near boundaries for $M = 2$ and Neumann or Dirichlet BCs . . . .	108
4.9	Exact normalization near boundaries for $M = 2$ and Robin BCs . . . . .	109
4.10	Exact normalization near boundaries for $M = 10$ and Neumann or Dirichlet BCs . . .	109
4.11	Estimation of the normalization factor near boundaries . . . . .	110
5.1	Grille "C" selon Arakawa . . . . .	122
5.2	Sensibilité au nombre d'itérations . . . . .	126
5.3	Erreur de modélisation d'une gaussienne par l'application de l'opérateur $2 \times 1D$ . . . .	127
5.4	Opérateur $2 \times 1D$ près des frontières, formulation 1 . . . . .	128
5.5	Opérateur $2 \times 1D$ près des frontières, formulations 2, 3 et 4 . . . . .	130
5.6	Exemple de variation des échelles . . . . .	132
5.7	Exemple de corrélations obtenues pour l'opérateur $3 \times 1D$ . . . . .	135
5.8	Zoom près des frontières de la figure 5.7 et comparaison avec l'explicite . . . . .	137
5.9	Comparaison des performances de l'assimilation . . . . .	144
5.10	Echelles de corrélation zonales issues d'une méthode d'ensemble (Daget, 2008) . . . .	146
5.11	Corrélations obtenues pour l'opérateur $3 \times 1D$ avec échelles géographiques . . . . .	146
5.12	Evaluation des configurations N°1 et N° 2, ORCA1, échelles paramétrées . . . . .	153
5.13	Evaluation de la configuration N° 3, ORCA1, échelles paramétrées . . . . .	154

---

5.14	Evaluation de la configuration N° 4, ORCA1, échelles paramétrées . . . . .	154
5.15	Evaluation de la configuration N° 4, ORCA2, échelles estimées . . . . .	155
5.16	Echelles de corrélation verticales issues d'une méthode d'ensemble (Daget, 2008) . . . . .	155
5.17	Effet de la variation des échelles . . . . .	157
B.1	Coefficients d'asymétrie et d'aplatissement . . . . .	179
B.2	Corrélations inhomogènes et anisotropes . . . . .	181
B.3	Symétrie de la fonction de corrélation . . . . .	182
D.1	Minimisation pour une matrice définie positive . . . . .	187
D.2	<b>A</b> -orthogonalité . . . . .	189
F.1	Décomposition par domaines horizontaux . . . . .	198
F.2	Réorganisation zonale des processeurs . . . . .	199
F.3	Réorganisation méridienne des processeurs . . . . .	200

---

## Liste des tableaux

4.1	Expression associated with the $M$ -th order AR functions . . . . .	99
4.2	Expression associated with the Gaussian function . . . . .	99
5.1	Temps d'exécution (ORCA2, <i>DELL Precision T5500</i> , $1 \times 1$ ) . . . . .	139
5.2	Temps d'exécution (ORCA2, <i>DELL Precision T5500</i> , $4 \times 2$ ) . . . . .	141
5.3	Temps CPU et mémoire (ORCA1, CGMOD, <i>IBM POWER6</i> ) . . . . .	142
5.4	Temps CPU et mémoire (ORCA1, CONGRAD/CGMOD -, <i>IBM POWER6</i> ) . . . . .	142
5.5	Détail des cinq configurations de normalisation évaluées. . . . .	151
5.6	Maximum des corrélations de la figure 5.11 . . . . .	152



---

# Introduction

La modélisation des océans permet d'établir des prévisions océaniques qui trouvent une application dans des domaines aussi variés que la recherche (support aux campagnes en mer, ...), l'industrie (ressources halieutiques, transport maritime, ...), l'environnement (estimation des risques de pollution, ...) ou même les loisirs de plaisance et les courses au large. Les océans sont également une composante essentielle de notre climat, de sorte que prévisions météorologiques et océaniques sont indissociables, et un effort important a été fait ces dernières décennies pour coupler leurs modélisations respectives. De plus, toute étude climatologique, qu'il s'agisse de paléoclimatologie ou d'estimation du climat futur, est ainsi assujettie à la connaissance de l'état océanique. Mais les processus physiques mis en œuvre dans les océans rendent leur modélisation mathématique délicate, et la grande taille du problème impose de trouver des algorithmes adaptés et de les dérouler sur des calculateurs parallèle à haute performance, complexifiant de ce fait l'aspect numérique du problème. De plus, des incertitudes, tant dans les modèles que dans les mesures physiques, peuvent produire des écarts importants entre les états prévus par le système et les états observés.

L'assimilation de données consiste alors à combiner l'état produit par le modèle (ébauche) avec des observations physiques, afin de corriger les conditions initiales et/ou forçages et/ou paramètres qui ont conduit à cette ébauche, et de produire ainsi un nouvel état (analyse) plus proche de la réalité. En océanographie, ce problème inverse se présente souvent sous la forme d'un problème de minimisation qui doit être régularisé. Compte tenu des contraintes de non linéarité, on utilise souvent un système itératif où chaque itération consiste à minimiser une fonction quadratique. Celle-ci est obtenue en linéarisant les opérateurs de la fonctionnelle non linéaire au voisinage du résultat obtenu par l'itéré précédent. La séquence est initialisée en linéarisant au voisinage de l'ébauche. Cette fonctionnelle quadratique décrit, au sens des moindres carrés, l'écart à l'ébauche pondéré par la confiance que l'on accorde à cette ébauche d'une part, et l'écart aux observations pondéré par la confiance que l'on accorde à ces observations d'autre part. Cette confiance, ou précision, s'exprime par l'inverse de la matrice des covariances de l'incertitude (erreurs) associée à l'ébauche et aux observations respectivement.

Les observations sont généralement supposées non-corrélées (bien que lorsqu'elles proviennent de satellites par exemple, cette hypothèse soit fortement discutable), et leur précision est alors obtenue en inversant simplement les variances de l'erreur liée à leur mesure. La matrice représentant les covariances d'erreur d'observations est alors diagonale. Mais concernant l'ébauche, aucun argument ne peut étayer l'hypothèse de la décorrélation des erreurs en chaque point de grille du modèle. De plus, les équations physiques résolues par le système montrent, de par leur nature, qu'il existe une relation entre les différentes variables de ce modèle. Pour résoudre le problème de minimisation, il faut donc estimer et inverser une matrice de grande taille. Dans le cas d'une grille basse résolution pour une configuration globale par exemple, et en considérant les champs en trois dimensions de la température, de la salinité et des composantes zonale (est/ouest) et méridienne (nord/sud) des courants, ainsi que le champ en deux dimensions de la hauteur d'eau, le vecteur d'état atteint facilement une taille de l'ordre de  $10^6$  éléments. Plus la résolution du modèle est haute, et plus la taille de ce vecteur augmente.

La très grande dimension de cette matrice, ainsi que la connaissance trop approximative des erreurs qui l'engendrent, empêchent sa manipulation directe, qu'il s'agisse de la créer, de la stocker ou

de l'inverser. Il est donc indispensable de développer des algorithmes efficaces permettant d'estimer ou de modéliser la matrice des covariances d'erreur d'ébauche à partir d'hypothèses raisonnables. Dans ce dernier cas, l'opérateur est donné en évaluation, *i.e.* la matrice n'est jamais accessible, mais l'évaluation correspond au produit de cette matrice par le vecteur auquel est appliqué l'opérateur. Dans un premier temps, on utilise souvent la formulation proposée par Derber et Bouttier (1999) permettant de décomposer chaque variable en une composante dite équilibrée, *i.e.* liée aux autres variables par des relations physiques d'équilibre (géostrophie par exemple), et une composante propre indépendante, donc décorrélée des autres variables. En s'affranchissant des relations qui existent entre les composantes équilibrées des différentes variables, il est alors possible de passer d'une formulation multivariée (existence de covariances d'erreur entre les variables) à plusieurs formulations univariées indépendantes (covariances d'erreur spatiales pour chaque variable). Ces covariances univariées peuvent alors s'exprimer par l'association de leurs variances et de leurs corrélations. Dans cette thèse, on se propose de poursuivre les travaux de Weaver et Courtier (2001) et Weaver et Ricci (2004) qui utilisent la résolution de l'équation de diffusion pour modéliser ces corrélations univariées.

Les corrélations univariées peuvent en effet être représentées par un opérateur de corrélation, ce qui, par définition, revient à évaluer une intégrale. Lorsque la fonction de corrélation, noyau de l'opérateur, ne dépend que d'une distance euclidienne, cette intégrale correspond à un produit de convolution. L'équation de diffusion est alors une alternative efficace à l'évaluation de tels produits pour un problème de grande taille. En effet, on montre facilement que la solution d'une équation de diffusion 1D classique est, au facteur  $1/(\sqrt{2\pi} L_g)$  prêt, le produit de convolution de la condition initiale par la fonction gaussienne  $e^{-x^2/2L_g^2}$ , où  $L_g = \sqrt{2\kappa T}$  est l'échelle de corrélation, avec  $\kappa$  le coefficient de diffusion, et  $T$  le pseudo-temps d'intégration. Appliquer l'équation de diffusion sur un pseudo-temps  $[0, T]$  à un champ scalaire, c'est donc appliquer à ce champ scalaire un opérateur de covariance de noyau gaussien. Une fois normalisé par le facteur  $\sqrt{2\pi} L_g$ , cet opérateur de covariance devient alors un opérateur de corrélation.

Weaver et Courtier (2001) exploitent cette propriété et proposent une résolution par schéma explicite d'une équation de diffusion 2D sur la sphère pour l'horizontale combinée à une équation de diffusion 1D pour la verticale. Mais si le codage explicite est aisé, ce schéma présente le gros désavantage de n'être stable que sous une condition dépendant du rapport au carré, de l'échelle de corrélation sur le pas spatial. Ainsi, plus on monte en résolution ou plus l'échelle est grande, et plus le nombre d'itérations augmente, rendant alors le coût de l'algorithme prohibitif. Weaver et Ricci (2004) suggèrent alors d'utiliser un schéma implicite, inconditionnellement stable. Mais le codage de ce schéma est plus difficile à mettre en place puisqu'il nécessite l'inversion d'une matrice de grande taille.

Dans cette thèse, on étudie plus avant le schéma implicite, en se focalisant tout d'abord sur la théorie de l'équation de diffusion 1D, et son extension à d'autres dimensions. Il est intéressant de noter que cette théorie n'est en aucune façon liée au problème de l'erreur d'ébauche et peut également servir dans d'autres contextes tels que la modélisation des corrélations de l'erreur du modèle (spatiale ou temporelle) ou même de l'erreur des observations lorsque celles-ci se présentent sous la forme d'un produit défini sur une grille. Néanmoins, pour cette thèse, l'application de ces développements théoriques seront réalisés pour l'erreur d'ébauche dans le cadre du système d'assimilation variationnelle NEMOVAR.

Mais un autre aspect important du problème, concerne la normalisation de l'opérateur de covariance représenté par l'équation de diffusion. En effet, lorsque les échelles de corrélations varient en fonction de leurs positions géographiques, ou lorsque l'on est proche des frontières, la normalisation à appliquer ne correspond plus à un simple facteur tel que  $\sqrt{2\pi} L_g$  donné plus tôt. On utilise alors des algorithmes pour calculer ces facteurs, mais ceux-ci peuvent s'avérer fort coûteux. Dans cette thèse, on s'intéresse donc également à ce problème en cherchant à évaluer différentes configurations.

La présentation de ce travail s'articule à partir des trois premiers chapitres qui permettent d'in-

roduire le contexte du sujet de cette thèse (quelques rappels mathématiques sont donnés dans les Annexes A, B et D). On expose tout d’abord la philosophie d’un problème inverse et on identifie les différentes composantes associées au cas particulier de l’assimilation de données océanographiques. Puis on détaille au chapitre 2 la technique de l’assimilation de données variationnelle (le principe du filtre de Kalman est donné en Annexe C). Enfin, on s’intéresse au chapitre 3 aux différentes techniques permettant d’estimer ou de modéliser les covariances d’erreur d’ébauche. Dans le cas de la modélisation par des opérateurs, un paragraphe est également consacré à la normalisation qui doit leur être associée, et le détail de la méthode proposée par Purser *et al.* (2003b) est donné en Annexe E. Après cette partie introductive, les deux chapitres suivants sont consacrés aux travaux réalisés au cours de cette thèse. Le chapitre 4 expose tout d’abord la théorie d’un opérateur de diffusion 1D et sa résolution par un schéma implicite. Les aspects numériques tels que les conditions aux frontières ou la variation spatiale des échelles de corrélations sont abordés. Ces opérateurs 1D sont ensuite utilisés (chapitre 5) pour construire des opérateurs 2D et 3D qui sont alors appliqués à la modélisation des corrélations univariées de l’erreur d’ébauche dans le système d’assimilation de données océaniques variationnelle NEMOVAR. Des détails techniques concernant cette construction sont donnés en Annexe F. Dans un dernier paragraphe, on évalue également différentes techniques de normalisation. Enfin, le chapitre 6 est consacré aux conclusions que l’on peut tirer de ce travail et aux perspectives qu’il ouvre.

Les techniques d’assimilation de données appliquées à l’océanographie sont très proches de celles développées en météorologie. Au cours de ce manuscrit, on trouvera donc des références liées à ces deux domaines sans qu’il soit nécessairement fait de distinction. En plus des articles et ouvrages cités expressément au cours des paragraphes suivants, les études réalisées, ainsi que l’écriture des chapitres de ce manuscrit doivent beaucoup à différentes sources d’informations :

- les cours de Master de Mathématiques Appliquées de l’Université Paul Sabatier (Toulouse III)<sup>2</sup> et notamment ceux directement liés à cette thèse et donnés par Luca Amodei, Jérémie Bigot, Abdallah Chalabi, Cécile Chouquet, Gérald Desroziers<sup>3</sup>, Pierre Maréchal, Mohamed Masmoudi, Dominikus Noll et Judith Vancostenoble ;
- les cours d’assimilation de données du CEPMMT<sup>4</sup> (Centre Européen pour les Prévisions Météorologiques à Moyen Terme) en particulier ceux donnés par Erik Andersson, Magdalena Balmaseda, Angela Benedetti, Carla Cardinali, Dick Dee, Mike Fisher, Tony MacNally et Yannick Trémolet ainsi que les notes de cours de Bouttier et Courtier (1999) ;
- les cours du CERFACS<sup>5</sup> donnés par Luc Giraud, Serge Gratton, Xavier Vasseur, et les nombreux séminaires qui y ont été dispensés (Andy Moore, Peter Oke, . . .) ;
- les différentes présentations lors des conférences “*WWRP THORPEX Workshop on 4D-Var and Ensemble Kalman Filter Inter-comparisons*”<sup>6</sup> de Buenos Aires (novembre 2008) et “*the 5th WMO Symposium on Data Assimilation*”<sup>7</sup> de Melbourne (Octobre 2009) ;
- les manuscrits de thèse de Ricci (2004), Daget (2008) et Pannekoucke (2008).

---

<sup>2</sup><http://www.ups-tlse.fr>

<sup>3</sup>CNRM : Centre National de Recherches Météorologiques : <http://www.cnrm.meteo.fr>

<sup>4</sup>ECMWF (European Centre for Medium-Range Weather Forecasts) : <http://www.ecmwf.int>

<sup>5</sup>Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique : <http://www.cerfacs.fr>

<sup>6</sup><http://4dvarenkf.cima.fcen.uba.ar>

<sup>7</sup><http://wmoda5.amos.org.au>



# Chapitre 1

## Un problème inverse : l'assimilation de données

### 1.1 Les problèmes inverses

Tout système physique, du plus simple au plus complexe, peut être décrit si l'on connaît les équations qui le régissent. Modéliser ce système, c'est donc résoudre numériquement ces équations. Ne-Zheng (1994) classe les différents modèles suivant quatre caractéristiques principales :

- déterministe / stochastique : un modèle est dit stochastique lorsqu'il fait appel à des variables aléatoires. La sortie de ce modèle n'est alors pas un état déterminé (modèle déterministe) mais une probabilité d'état ;
- linéaire / non linéaire : suivant la linéarité ou non des équations qui le régissent ;
- stationnaire / dynamique : lorsque l'état produit par le modèle évolue dans le temps (le temps fait alors partie des variables), le modèle est dit dynamique, sinon il est dit stationnaire ;
- à paramètres rassemblés / distribués : lorsque les variables sont des fonctions (continues ou discrétisées) évoluant dans l'espace, le modèle est dit à paramètres distribués.

La modélisation d'un système physique peut permettre d'estimer l'état passé, présent ou futur de ce système, et servir ainsi à analyser ou prévoir un événement dans les domaines les plus variés tels que la recherche, la sécurité, la protection, les loisirs . . . Bien sûr la performance de la modélisation, c'est à dire la pertinence de la prévision, dépend directement du degré de connaissance des équations (et de leur résolution) qui régissent le système. D'autre part, lorsque ce dernier est mal connu, sa modélisation peut alors également servir à l'étudier en comparant ses prévisions à la réalité.

De manière générale, un système à paramètres distribués est régi par une ou des équations aux dérivées partielles. Leur résolution nécessite alors les composantes suivantes (Ne-Zheng, 1994) :

- une région spatiale (paramètres distribués) et un intervalle de temps (modèle dynamique) ;
- des paramètres système : ils caractérisent la géométrie et/ou la nature physique du système
- des conditions subsidiaires : elles donnent l'état initial du système ainsi que les relations qui décrivent les échanges de masse ou d'énergie avec les systèmes voisins ;
- des variables de contrôle<sup>8</sup> : elles représentent l'excitation du modèle. Elles peuvent parfois apparaître dans les conditions subsidiaires ;
- des variables d'état : elles décrivent l'état du système.

Mais, les modèles sont soumis à plusieurs sources d'erreur : discrétisation, approximation des équations, hypothèses pas toujours valides, mauvaise connaissance des processus physiques, incertitudes sur

---

<sup>8</sup>Il s'agit ici d'une définition générale pour un modèle. Le terme "variables de contrôle" sera redéfini dans les paragraphes suivants pour le problème de l'assimilation de données.

les conditions subsidiaires, ... La solution donnée par le modèle n'est donc pas exacte, et les erreurs peuvent même s'amplifier au cours du temps (dans le cas d'un modèle dynamique). Dans ce cas, ou lorsque la solution s'avère trop imprécise, il peut être nécessaire d'asservir le modèle par toute technique permettant de le ramener vers la réalité. Toute mesure du système physique, bien qu'entachée d'erreur, est une expression de cette réalité. Il apparaît donc logique de se servir de ces observations pour effectuer cette tâche et corriger ainsi des paramètres, conditions ou variables de la modélisation comme le montre la figure 1.1.

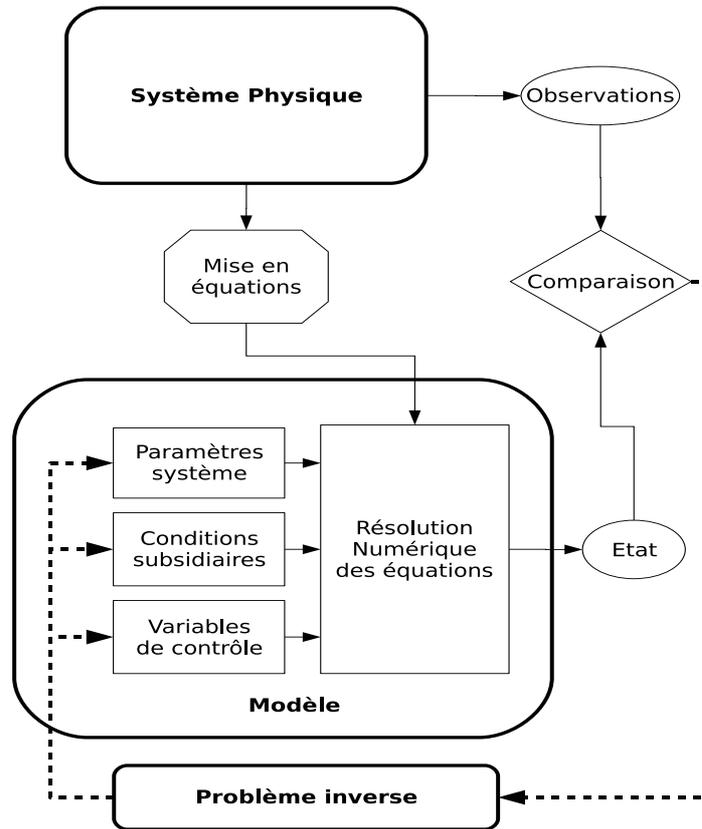


FIG. 1.1 – Modélisation d'un système physique

Le **problème inverse** (voir également Tarantola, 2005) consiste donc, à partir d'un ensemble de mesures observées, à remonter aux paramètres, conditions ou variables qui ont engendrés l'état donné par le modèle et que l'on compare à ces observations. Ainsi par exemple, en supposant que les observations sont proches de l'état exact du système physique à un instant donné, le problème inverse peut permettre de recalibrer un paramètre défaillant de la modélisation. La terminologie de problème inverse peut cependant être trompeuse comme le montre Wunsch (1996) sur l'exemple de l'équation de Poisson

$$\Delta\varphi = \psi. \quad (1.1)$$

En effet, le problème

(A) : évaluer l'équation (1.1) pour trouver  $\psi$  sachant  $\varphi$ ,  
semble simple à résoudre et d'une logique plus directe que le problème

(B) : évaluer l'équation (1.1) pour trouver  $\varphi$  sachant  $\psi$  et les conditions aux frontières,  
qui est d'un abord bien plus complexe. Pourtant c'est ce dernier problème (B) qui est qualifié de direct car c'est celui qui est le plus familièrement rencontré dans de nombreux domaines. Le problème (A), bien que plus basique, est alors qualifié d'inverse, dans le sens **inverse au problème conventionnel**.

Wunsch (1996) précise également que (A) n'est pas le seul problème inverse à (B). Un autre problème inverse pourrait être par exemple

(C) : évaluer l'équation (1.1) pour trouver les conditions aux frontières sachant  $\varphi$  et  $\psi$ .

Néanmoins, et bien que l'équation de Poisson précédente ne le montre pas, une des difficultés majeures des problèmes inverses est qu'ils sont souvent mal posés au sens de Hadamard, *i.e.* au moins une des trois conditions suivantes n'est pas respectée : existence, unicité et stabilité. Le non respect de ces conditions est souvent lié à la quantité d'informations disponible pour évaluer le problème ainsi qu'aux erreurs entachant les observations. Menke (1989) cite les possibilités suivantes :

- Problèmes indéterminés (ou sous-déterminés) : lorsque le système régissant le problème inverse ne fournit pas suffisamment d'information, le nombre d'inconnues (degrés de liberté) est plus important que le nombre d'équations permettant de les déterminer. Plusieurs solutions existent alors, violant ainsi le principe d'unicité. On pourra cependant compléter le système par de l'information *a priori*, *i.e.* toute information permettant d'anticiper les caractéristiques de la solution (estimation *a priori*, signe, fourchette de valeurs, approximation de la forme, ...).
- Problèmes juste-déterminés : lorsque le système fournit très exactement la quantité d'information requise et qu'il n'y a aucune singularité, il est alors possible de définir une solution unique. Cependant, la stabilité du système n'est pas garantie. En effet, si les observations sont entachées d'erreur, deux séries d'observations distinctes donneront deux solutions différentes dont l'écart peut être important. La solution n'étant pas reproductible, le système est dit instable.
- Problèmes surdéterminés : lorsque le système fournit plus d'information que nécessaire, les erreurs liées aux observations engendrent des incohérences et il n'est plus possible de calculer une solution. Il sera alors nécessaire de déterminer des critères de sélection pour en définir une.

Beaucoup de techniques différentes existent pour **régulariser** les problèmes inverses, c'est à dire palier le fait qu'ils soient mal posés. La régularisation de Tikhonov en est une des plus célèbres et donne lieu à de multiples variantes. Pour le problème inverse que nous étudions dans les chapitres suivants, la méthode consistant à minimiser une fonction coût quadratique, correspond en fait à une régularisation de Tikhonov généralisée. Le lecteur intéressé pourra se reporter à Aster *et al.* (2005) par exemple, pour une introduction pédagogique des problèmes inverses et leur régularisation. A toutes fins utiles, on trouvera en annexe un rappel de certains outils mathématiques.

## 1.2 L'assimilation de données en océanographie

### 1.2.1 Le modèle

Les océans et le mouvement de ses masses d'eau influencent fortement un grand nombre de processus tels que la productivité biogéochimique, les populations marines, le mouvement des polluants ou encore le changement climatique et les conditions atmosphériques. Connaître et prévoir leur état est donc primordial pour un grand nombre d'applications.

Leur modélisation est cependant complexe : elle fait appel à la dynamique des fluides et doit prendre en compte les échanges et interactions intervenant avec d'autres systèmes comme l'atmosphère, les continents et le plancher océanique, la biogéochimie, les glaces de mer, le ruissellement ... Pour décrire l'océan dans son ensemble, ou plus généralement l'appareil climatique de la Terre, on fait appel à la modélisation couplée des différents systèmes. Dans cette thèse, nous ne nous intéressons qu'aux modèles de circulation générale des océans (OGCM<sup>9</sup>) pour lesquels des conditions aux frontières sont spécifiées en lieu et place des modèles couplés.

---

<sup>9</sup>Ocean General Circulation Model

Ces modèles décrivent l'état dynamique des océans par des informations (**variables d'état**) de température ( $T$ ), de salinité ( $S$ ) et de vitesse des courants (composantes zonale  $u$  et méridienne  $v$ ). Pour représenter la surface, plutôt que d'utiliser une formulation **toit rigide** qui considère la surface de l'océan comme un plan sur lequel une certaine pression est appliquée, on préfère aujourd'hui utiliser une formulation **surface libre** mettant en œuvre une nouvelle variable d'état, la hauteur d'eau en surface ( $\eta$ ). Cette variable permet de décrire les variations de l'interface air-océan avec plus de réalisme. Ces variables d'état correspondent à la discrétisation de l'océan sur une grille spatiale (modèle à paramètres distribués) et sont mises à jour selon des intervalles de temps (modèle dynamique).

Les OGCM sont basés sur la résolution des **équations primitives** sur une sphère en rotation, *i.e.* les équations de Navier-Stokes auxquelles on ajoute une équation d'état reliant les traceurs actifs (température et salinité) à la vitesse du fluide, et que l'on résout sous un certain nombre d'hypothèses (voir par exemple Haidvogel et Beckmann, 1999). Ils sont donc clairement non linéaires. Pour résoudre ces équations primitives, outre des coefficients divers (paramètres système) liés à ces équations, il est nécessaire de définir la géométrie du domaine (paramètres système), c'est à dire les lignes côtières et la bathymétrie. La **condition initiale** (conditions subsidiaires) est donnée par un état de l'océan au temps  $t_0$ . Lorsque cette condition initiale correspond à un état de repos avec une climatologie pour  $T$  et  $S$ , on intègre d'abord le modèle sur une certaine période (**spin up**) dépendant de la mémoire de l'océan, avant de l'utiliser pour une application donnée. Cette phase permet à l'océan modélisé d'"oublier" son état initial peu précis. Enfin, l'océan et ses frontières, *i.e.* la croûte terrestre, les continents, la glace et l'atmosphère, échangent des flux de chaleur, d'eau douce, de sel et de moments. Si certains de ces échanges (conditions subsidiaires) peuvent être négligés, d'autres doivent être paramétrés (par des termes de rappel par exemple) ou calculés. D'autres encore sont fournis par des modèles, que ce soit sous forme de produits (forçages, c'est souvent le cas pour les informations atmosphériques telles que tension de vent, flux de chaleur et flux d'eau douce) ou directement sous forme de couplage.

La modélisation des océans consiste, d'un point de vue grossier, en une réponse déterministe à un forçage atmosphérique. Cependant les erreurs induites tant par les incertitudes de ces forçages que par les approximations du modèle, ne peuvent pas être connues de façon déterministe. En supposant cependant que ces incertitudes soient faibles et que la modélisation soient suffisamment fidèle, ces erreurs peuvent être représentées par des perturbations stochastiques (Cohn, 1997). Bien que cela soit difficile à réaliser en pratique, Daley (1992) sépare cette erreur en deux composantes :

- l'**erreur modèle** : liée à la résolution des équations, elle inclut les approximations et simplifications de la physique, les erreurs des algorithmes numériques (discrétisation par exemple), les défauts de spécification des paramètres et des conditions aux frontières.
- l'**erreur de prévisibilité** : elle représente la façon dont le modèle fait évoluer une erreur existant sur la condition initiale.

Ainsi, si on considère le modèle comme parfait, l'erreur sur l'état qu'il prévoit est due à une mauvaise connaissance de l'état initial. Mais si on considère que l'état initial de l'océan est parfaitement connu, l'erreur sur l'état issu de la modélisation est due aux imperfections du modèle.

## 1.2.2 Les observations

Pendant longtemps, les seules observations de l'océan disponibles ont été celles des navires marchands et des marines nationales. Bien que n'offrant qu'une couverture des lignes maritimes principales, et souvent limitées à des observations de surface, ces mesures, complétées par celles de campagnes de recherche et de navires d'opportunité (volontaires), ont permis d'établir des bases de données climatiques comme celles de Levitus<sup>10</sup>. Mais ces climatologies ne permettent pas de rendre compte de la variabilité importante de l'océan sur une large gamme d'échelles spatiales et temporelles. Compte tenu de son rôle essentiel dans le changement climatique, la nécessité de mettre en place des programmes

<sup>10</sup>World Ocean Database : [http://www.nodc.noaa.gov/OC5/WOD09/pr\\_wod09.html](http://www.nodc.noaa.gov/OC5/WOD09/pr_wod09.html)

internationaux dédiés aux mesures océaniques et à l'interaction océan/atmosphère s'est alors faite sentir (Busalacchi, 1997).

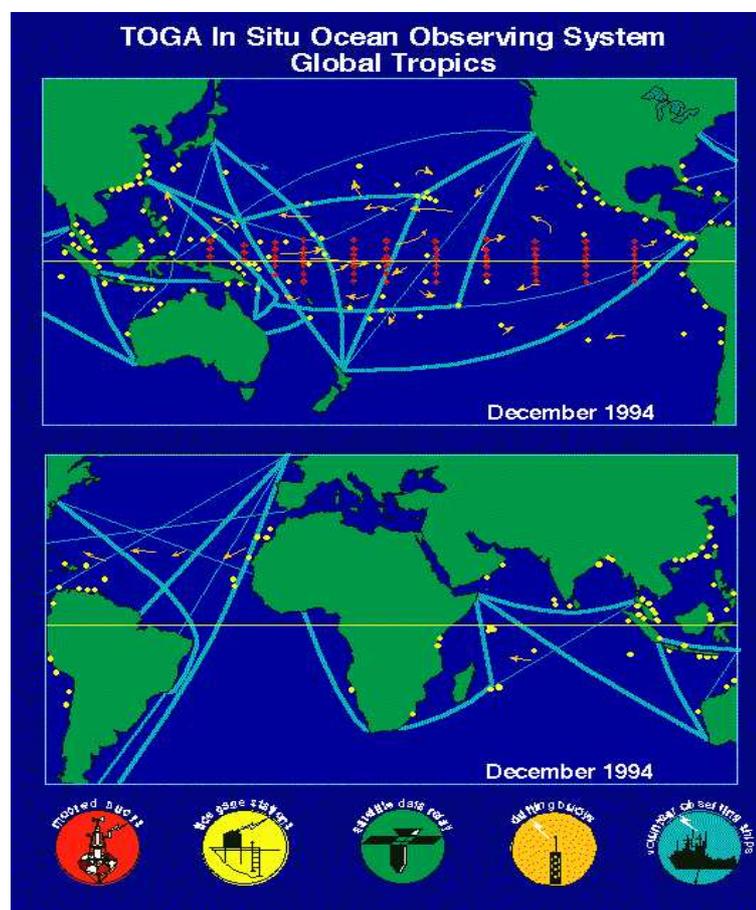


FIG. 1.2 – Réseau d'observations océaniques *in situ* du programme TOGA en décembre 1994.

En 1985, le programme TOGA<sup>11</sup> a pour objectif de décrire la dynamique particulière des océans tropicaux afin de mieux appréhender des phénomènes comme El Niño, et s'intéresse également à l'atmosphère globale afin d'étudier la possibilité de modéliser un système couplé océan-atmosphère. Il s'appuie en particulier sur le déploiement de bouées amarrées de type ATLAS<sup>12</sup> dans le Pacifique Tropical. Ce déploiement, baptisé TAO<sup>13</sup>, atteint environ 70 bouées en décembre 1994, à la fin du programme TOGA (voir figure 1.2). Dans le prolongement de ce succès, les programmes GCOS<sup>14</sup> et GOOS<sup>15</sup> sont lancés au début des années 90 afin de pérenniser le système d'observation de l'océan et de l'atmosphère. À partir de 1995, un nouveau programme voit également le jour afin d'étudier la variabilité naturelle du climat et des changements dus à l'activité humaine : le programme CLIVAR<sup>16</sup>. C'est dans ce cadre qu'un nouveau déploiement de bouées ATLAS est effectué en 1997 dans l'Atlantique Tropical sous le nom de PIRATA<sup>17</sup>. Sous la même égide, TAO prend le nom de TAO/TRITON en janvier 2000 après le déploiement dans le Pacifique Ouest de bouées amarrées TRITON<sup>18</sup>. Aujourd'hui,

<sup>11</sup>Tropical Ocean Global Atmosphere : <http://www.ncdc.noaa.gov/oa/coare/toga.html>

<sup>12</sup>Autonomous Temperature Line Acquisition System

<sup>13</sup>Tropical Atmosphere Ocean : <http://www.pmel.noaa.gov/tao>

<sup>14</sup>Global Climate Observing System : <http://www.wmo.int/pages/prog/gcos/>

<sup>15</sup>Global Ocean Observing System : <http://www.ioc-goos.org/>

<sup>16</sup>Climate Variability and Predictability : <http://www.clivar.org/>

<sup>17</sup>Pilot Research Moored Array in the Tropical Atlantic : <http://www.brest.ird.fr/pirata/index.php>

<sup>18</sup>Triangle Trans Ocean Buoy Network

un nouveau déploiement est en cours dans l'Océan Tropical Indien sous le nom de RAMA<sup>19</sup>.

Elargissant le champ des Tropiques à une perspective plus globale, le programme WOCE<sup>20</sup> de 1990 vise à améliorer la modélisation de la circulation générale des océans pour les prévisions climatiques décennales. Entre autres il met en œuvre, pour les flux de sub-surface, des profileurs (instruments de mesure sur toute ou partie d'une colonne d'eau) de type ALACE<sup>21</sup> qui donneront ensuite naissance aux profileurs Argo (voir figure 1.3). Le début des années 1990 voit aussi le lancement de programmes tels que JGOFS<sup>22</sup> qui s'intéressent davantage aux processus chimiques et biologiques des océans.

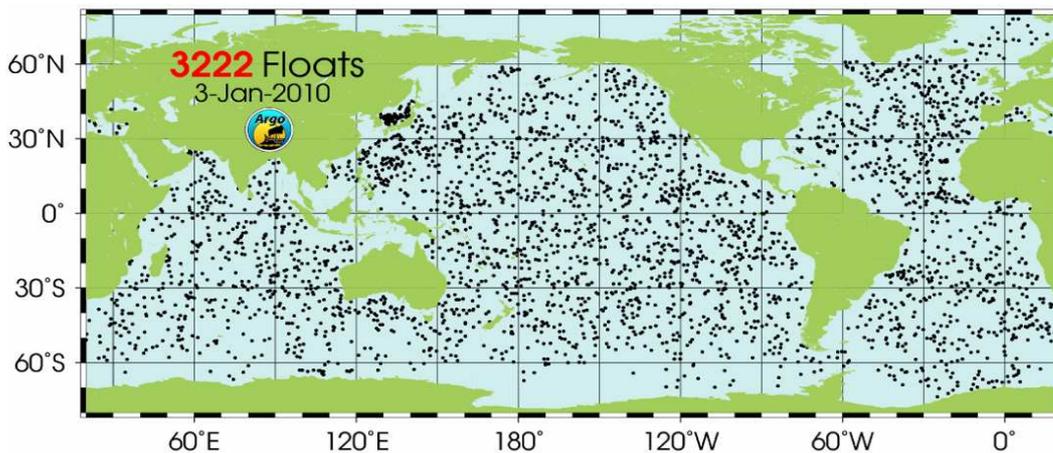


FIG. 1.3 – Déploiement de profileurs Argo au 3 janvier 2010.

Concernant les mesures elles-mêmes, on peut en distinguer deux types. Les données *in situ* consistent en des mesures en un point, ou en des profils sur toute ou une partie de la colonne d'eau, de température, salinité, vitesse et direction des courants, pression, dioxyde de carbone, . . . Elles sont effectuées par différentes sortes d'instruments tels que des thermosalinographes, CTD<sup>23</sup>, XBT<sup>24</sup>, bouées amarrées ou dérivantes, profileurs, . . . Mais parce que l'océan est un milieu particulièrement difficile à observer, ces mesures restent très hétérogènes.

Le second type de données est pourvu par les satellites. Nimbus-7 et Seasat démontraient dans les années 70 les avantages des instruments spatiaux pour une couverture homogène de la planète. Néanmoins, seule la surface des océans peut être observée, l'océan profond restant inaccessible. Aujourd'hui, une large gamme de mesures satellitaires de surface sont disponibles et permettent d'obtenir la topographie (Jason-1, ERS-2, . . .) et le géoïde associé (GRACE, GOCE), les températures de surface (instruments AVHRR, ATSR, . . .), la couleur de l'océan et donc les teneurs en chlorophylle et phytoplancton (Envisat, HY-1), la concentration de glace de mer (ICESat, CryoSat) ou encore la salinité (SMOS). En général, des centres de traitement rassemblent et éventuellement combinent plusieurs jeux de mesures de type différent pour obtenir un produit donné (Coriolis<sup>25</sup> pour des bases de données *in situ*, GHRSSST<sup>26</sup> pour des produits combinés de température de surface, . . .) et effectuent ainsi une première validation des observations. La figure 1.4 présente le réseau de satellites d'observation pour

<sup>19</sup>Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction

<sup>20</sup>World Ocean Circulation Experiment : <http://woce.nodc.noaa.gov/wdiu/index.htm>

<sup>21</sup>Autonomous Lagrangian Circulation Explorer

<sup>22</sup>Joint Global Ocean Flux Study : <http://ijgofs.whoi.edu/>

<sup>23</sup>Conductivity Temperature Depth

<sup>24</sup>Expendable Bathythermographie

<sup>25</sup><http://www.coriolis.eu.org/>

<sup>26</sup>Group for High resolution Sea Surface Temperature : <http://www.ghrsst.org/>

la météorologie et l'océanographie de la NASA<sup>27</sup>.



FIG. 1.4 – Réseau d'observations satellite pour la météorologie et l'océanographie de la NASA.

Une condition indispensable aux prévisions océaniques réalistes (et aux diverses applications qui leur sont liées) est qu'il soit possible d'exploiter un système d'observation pérenne, délivrant de façon régulière des informations suffisamment riches et diversifiées. C'est à partir de ce constat que le programme GODAE<sup>28</sup>, initié en 1997, s'efforce de donner des bases solides à une océanographie opérationnelle. A sa fin en 2008, l'expérience laisse place à des systèmes matures de prévision à méso-échelle opérationnels tels que Mercator Océan<sup>29</sup> (France), NCOF<sup>30</sup> (UK), BLUElink<sup>31</sup> (Australie), et des projets internationaux comme MyOcean<sup>32</sup> pour la mise à disposition de produits océaniques variés.

Cependant, toute observation, qu'elle soit *in situ*, satellitaire ou combinée, est sujette à une **erreur de mesure** liée à l'instrument lui-même (définition, mauvaise calibration, dérives, ...) ou à son enregistrement et sa communication (codage incorrect, défaut de transmission, ...). Dans certains cas, cette erreur peut être aberrante et l'observation doit être rejetée, dans d'autres, la mesure souffre d'un biais systématique. Il est donc indispensable de mettre en place un contrôle efficace de la qualité des observations (Kalnay, 2003). Ce contrôle n'ayant cependant pas de prise sur le caractère aléatoire des erreurs, il ne peut donc pas garantir l'intégrité complète des données.

### 1.2.3 Comparaison modèle / observations

En général, il n'est pas possible de comparer directement l'état prévu par le modèle avec les observations. En effet, il n'est pas réaliste de supposer que la grille de discrétisation (spatiale et temporelle)

<sup>27</sup>National Aeronautics and Space Administration : <http://climate.nasa.gov/>

<sup>28</sup>Global Ocean Data Assimilation Experiment : <http://www.godae.org/>

<sup>29</sup><http://www.mercator-ocean.fr/>

<sup>30</sup>National Center for Ocean Forecasting : <http://www.ncof.co.uk/>

<sup>31</sup><http://www.marine.csiro.au/bluelink/>

<sup>32</sup><http://www.myocean.eu.org/>

du modèle puisse correspondre exactement aux emplacements et aux instants de mesure des différents instruments. D'autre part, les variables d'état du modèle peuvent ne pas correspondre directement aux grandeurs mesurées mais peuvent leur être liées par des relations physiques plus ou moins complexes. Un satellite mesure par exemple la radiance émise par la Terre, et la relie à la température de surface grâce à l'équation de transfert radiatif.

Modèle et observations sont donc définis dans deux espaces différents. Pour pouvoir effectuer la comparaison, on définit un **opérateur d'observation** permettant de passer de l'**espace du modèle** à l'**espace des observations**. On peut alors calculer l'**innovation**, *i.e.* l'écart qui existe entre les observations et l'état du modèle ramené dans l'espace des observations. C'est à partir de cette innovation qu'une correction peut être calculée par le problème inverse. Cependant, l'opérateur d'observation, en fonction de sa nature, est sujet à une **erreur d'interpolation** comprenant l'aspect interpolation proprement dit mais également les aspects discrétisation et équivalence des variables (résoudre l'équation de transfert radiatif par exemple n'est pas trivial).

Le fait que le modèle et les observations soient définis dans deux espaces différents est également source d'un autre type d'erreur. En effet, les observations contiennent des informations provenant de processus physiques de différentes échelles, y compris d'échelles trop petites pour être prises en compte par la résolution du modèle. C'est cette inadéquation avec la résolution du modèle qui génère l'**erreur de représentativité** (Janjić et Cohn, 2006). Elle peut être relativement importante et même dominer les erreurs de mesure et d'interpolation lorsque le modèle utilisé est à basse résolution.

### 1.2.4 Le problème inverse

Si connaître l'état de l'océan avec exactitude à tout instant donné relève de l'utopie, il est cependant possible de s'en rapprocher suffisamment pour en donner une représentation cohérente. Pour prendre en compte l'erreur de représentativité, il convient d'augmenter les résolutions et d'améliorer la physique à petite échelle. L'assimilation de données ne peut donc pas agir sur ce type d'erreur. En revanche, elle permet de réduire l'erreur de prévisibilité et éventuellement d'estimer l'erreur modèle tout en tenant compte des erreurs liées aux observations. Ici, le problème inverse va donc corriger l'état initial pour obtenir une prévision plus proche de la réalité.

Le vecteur d'état d'un OGCM est généralement de grande taille ( $\approx 10^6$  éléments pour un modèle basse résolution). Bien que le système d'observation se soit considérablement amélioré au cours des années, il n'en reste pas moins que le nombre d'observations sur une période de comparaison donnée, est bien inférieur à la taille de ce vecteur état, ce qui fait de l'assimilation de données dans ce cas un problème sous-déterminé<sup>33</sup>. Une solution naturelle est alors de considérer que l'état donné par le modèle lui-même constitue une approximation de la réalité et peut donc servir d'information *a priori*. Dans ce cas, cet état est appelé **ébauche** tandis que l'état corrigé après application du problème inverse est appelé **analyse**. Augmenté de l'ébauche, le problème est alors surdéterminé, et compte tenu des incertitudes liées tant aux observations qu'au modèle, il n'est généralement plus possible de trouver une solution qui respecte toutes les informations fournies. Il convient alors de définir des critères permettant de déterminer malgré tout une solution.

De manière générale, les **variables de contrôle**, *i.e.* les éléments sur lesquels on agit par l'intermédiaire du problème inverse, sont les mêmes que les variables d'état. Cependant, on peut choisir de n'effectuer des corrections que sur une partie de ces variables, soit pour des raisons de coût, soit parce que certaines variables sont difficiles à prendre en compte, soit encore parce que l'on sait définir entre elles des relations de dépendance. On peut encore choisir de contrôler en plus certains paramètres

<sup>33</sup>On notera cependant que les observations n'étant homogènes ni spatialement, ni temporellement, il existe des régions localement denses en observations pour certaines variables et pour lesquelles il est donc possible d'argumenter un sur-dimensionnement du problème.

ou forçages. On distinguera donc l'espace des variables de contrôle de l'espace des variables d'état.

## 1.3 Les différentes techniques pour l'océanographie et la météorologie

Les méthodes d'assimilation de données ont, de manière générale, été développées dans le cadre de la météorologie avant d'être adaptées à l'océanographie. L'historique qui suit se concentre sur les différentes techniques et non sur les domaines d'application particuliers. C'est pourquoi, on y trouvera des références de météorologie aussi bien que d'océanographie sans distinction explicite.

### 1.3.1 Un bref historique

En assimilation de données, les météorologues sont les premiers à tenter de se servir des observations pour améliorer leurs prévisions grâce à des procédures dites d'**analyse objective**, car ne reposant pas sur le jugement (subjectif) d'un analyste. Cressman (1959) reprend la méthode de Bergthórsson et Döös (1955) et propose un algorithme simple de **corrections successives** de l'ébauche en fonction des observations. Pour cela, chaque fois qu'une observation est disponible, les valeurs des points de grille du modèle proches sont interpolées jusqu'à ce point d'observation pour en calculer l'écart. La correction appliquée en chaque point de grille est alors définie par l'innovation adéquate ramenée sur le point de grille d'origine et affectée d'un poids décroissant avec l'augmentation de la distance du point de grille à l'observation. Cette décroissance est maîtrisée par un paramètre permettant de déterminer à partir de quelle distance l'observation n'a plus d'influence. Mais un des inconvénients majeurs des méthodes de corrections successives vient du fait que la fonction poids utilisée est arbitraire et ne tient pas compte des qualités respectives de l'ébauche et des observations, risquant par exemple de dégrader une bonne estimation du modèle en y associant des mesures physiques médiocres. De plus, les corrections apportées peuvent ne pas respecter les caractéristiques physiques du système, en générant par exemple des sauts dans un champ, ou en détruisant l'équilibre entre deux variables.

Gandin (1963) propose alors une méthode où la correction à apporter en un point de grille donné est déterminée par les écarts entre ce point de grille et un ensemble d'observations autour de ce point, associés à un poids calculé en fonction de statistiques sur les différentes erreurs. Cette méthode, dite d'**interpolation statistique** ou **interpolation optimale** (OI<sup>34</sup>) et sa mise en œuvre pratique pour les problèmes à grande taille (Lorenc, 1981) a été largement utilisée pour sa simplicité et son coût raisonnable. Néanmoins, un des inconvénients de la technique vient du besoin de décomposer le domaine d'analyse en plusieurs sous-domaines. L'influence des observations est alors réduite à leur sous-domaine d'affectation et un bruit parasite en résulte aux frontières. De plus, la cohérence des petites échelles avec des échelles plus grandes n'est plus garantie. Enfin, l'utilisation d'opérateurs d'observation non linéaires s'avère problématique.

A la même époque, Kalman (1960) et Kalman et Bucy (1961) décrivent un estimateur discret pour des systèmes linéaires ou des équations différentielles ordinaires. Ce filtre sera en particulier appliqué avec succès pour le calcul de trajectoires du programme Apollo. Étudié en météorologie à partir des années 1980 (Ghil *et al.*, 1981), il permet de corriger, à chaque pas de temps, le vecteur d'état du modèle à partir du vecteur d'innovation multiplié par une matrice, appelée **matrice de gain**, formulée à partir de statistiques (covariances) sur les différentes erreurs. Si la technique est similaire à celle de l'OI, le **filtre de Kalman** (KF<sup>35</sup>) s'applique au domaine complet et permet de propager les covariances d'erreur du pas de temps précédent grâce au modèle dynamique. Mais les modèles étant

---

<sup>34</sup>Optimal Interpolation

<sup>35</sup>Kalman Filter

non linéaires, les équations doivent être linéarisées pour la propagation des erreurs, et à condition que ces non-linéarités soient faibles (Jazwinski, 1970; Evensen, 2007), la méthode, bien que perdant son optimalité, reste satisfaisante. Cette technique est connue sous le nom de **filtre de Kalman étendu** (EKF<sup>36</sup>). Néanmoins les problèmes d'assimilation de données, tant en météorologie qu'en océanographie, étant de grande dimension, construire la matrice de gain et propager les statistiques n'est pas réalisable. Des méthodes de **rang réduit** pour approximer le filtre de Kalman ont alors été développées (Fukumori et Malanotte-Rizzoli, 1995; Buehner et Malanotte-Rizzoli, 2003). Le principe est basé sur la réduction des matrices de covariances d'erreur à des matrices de rang plus faible mais restant significatif. On citera par exemple le filtre SEEK<sup>37</sup> introduit par Pham *et al.* (1998). Pour palier le problème de la propagation des covariances, on peut aussi choisir de les estimer à partir d'un ensemble d'échantillons. Pour cela, on applique l'opérateur de propagation du modèle non linéaire au même état plusieurs fois, mais en perturbant un élément différemment à chaque fois. Les résultats constituent un ensemble permettant de calculer les covariances d'erreur (Evensen, 2007) qui sont alors appliquées dans le calcul de l'analyse. Cette méthode est appelée **filtre de Kalman d'ensemble** (EnKF<sup>38</sup>) (le principe des méthodes d'ensemble sera précisé au chapitre 3).

Les approches présentées ci-dessus sont des approches d'**assimilation séquentielle**. Pour ces méthodes, la prévision peut être mise à jour dès qu'une observation est disponible. Mais cette approche ne permet aux observations que de corriger les états suivants et jamais les états précédents. La propagation de l'information contenue dans les observations ne peut donc s'effectuer que du passé vers le futur, et jamais du futur vers le passé. Si cet inconvénient n'est pas déterminant dans le cadre de prévisions puisqu'elles demandent une intégration du modèle vers le futur, il est majeur dans le cadre d'un exercice de réanalyse. Pour palier le problème, la méthode du **Kalman "smoother"** (voir par exemple Ménard et Daley, 1996) permet de revenir vers le passé après qu'une analyse vers le futur ait été faite.

Parallèlement aux méthodes exposées précédemment, Sasaki (1958) propose quant à lui, d'établir une analyse objective à partir d'une méthode variationnelle, *i.e.* de minimiser la distance aux observations tout en respectant une certaine contrainte dynamique. En particulier, Sasaki (1970) définit la notion de **contrainte forte** lorsque cette contrainte doit être satisfaite exactement, et la notion de **contrainte faible** lorsqu'elle l'est de manière approximative. Plus cohérente puisque prenant en compte des contraintes physiques, cette méthode reste néanmoins coûteuse. Le Dimet et Talagrand (1986) proposent alors des méthodes itératives pour résoudre le problème, et montrent en particulier comment l'utilisation du modèle adjoint (voir Annexe A et paragraphe 2.4.3) permet de réduire les coûts. On parle de **3D-Var** lorsque l'aspect temporel n'est pas pris en compte et de **4D-Var** sinon.

A l'inverse des méthodes séquentielles, l'**assimilation variationnelle** permet, lorsque la variable temps est incluse (4D-Var), de prendre en compte les observations sur une période donnée de façon globale. En effet, on obtient alors une correction non plus de l'état du modèle en fonction des observations disponibles à l'instant donné, mais de la **trajectoire de l'état** en fonction des observations disponibles sur l'ensemble d'une période donnée. Le détail de cette méthode sera donné au chapitre 2.

Aujourd'hui, la plupart des grands centres de météorologie utilisent un système 4D-Var pour leurs prévisions opérationnelles (Rabier *et al.*, 2000; Gauthier *et al.*, 2007; Rawlins *et al.*, 2007). En océanographie, depuis les premières tentatives d'implémentation conduites par Thacker et Long (1988), des systèmes variationnels (3D-Var ou 4D-Var) ont été mis en œuvre aussi bien pour des systèmes globaux (Stammer *et al.*, 2002; Daget *et al.*, 2009; Mogensen *et al.*, 2009) que régionaux (Weaver *et al.*, 2003; Di Lorenzo *et al.*, 2007; Dobricic et Pinardi, 2008; Li *et al.*, 2008).

---

<sup>36</sup>Extended Kalman Filter

<sup>37</sup>Singular Evolutive Extended Kalman Filter

<sup>38</sup>Ensemble Kalman Filter

Les techniques d'assimilation séquentielle et variationnelle utilisent des formulations de problèmes inverses pour calculer une correction à appliquer à l'ébauche. Néanmoins, le problème direct peut être formulé par des fonctions de densité de probabilité conditionnelles (voir paragraphe 2.1), et il est alors possible de montrer que la distribution *a posteriori* peut être calculée à partir d'un ensemble de particules que l'on fait évoluer à travers le modèle. Chaque particule représente la densité de probabilité *a priori* et est affectée d'un poids calculé *a posteriori* par rapport à la densité de probabilité des observations sachant l'état. Cette technique des **filtres particulaires** est appliquée dans le cadre d'un modèle océanique (région du courant des Aiguilles) par Van Leeuwen (2003), qui note cependant une forte tendance de l'ensemble à s'effondrer, *i.e.* à parvenir à une situation dans laquelle un seul membre de l'ensemble se voit affecter de presque tout le poids *a posteriori*. Cette tendance est confirmée par Snyder *et al.* (2008) pour les problèmes à grande dimension.

### 1.3.2 L'assimilation aujourd'hui

D'un point de vue algorithmique, on distingue deux grandes familles d'assimilation : l'assimilation séquentielle (OI et EnKF principalement) et l'assimilation variationnelle (3D-Var et 4D-Var). Mais quelque soit la méthode utilisée, les difficultés et questions auxquelles est confronté l'assimilation de données peuvent être résumées par les points suivants :

- la dimension du problème : la prise en compte des processus physiques à des échelles de plus en plus petites nécessite l'augmentation de la résolution des modèles, et donc la dimension du vecteur d'état, entraînant alors un besoin toujours plus grand tant en mémoire qu'en vitesse de calcul. L'utilisation de calculateurs à haute performance répond en partie à cette contrainte mais le facteur de réduction des temps de calculs n'est pas proportionnel au nombre de processeurs utilisés (voir par exemple Dowd et Severance, 1998). Mettre en œuvre des algorithmes de résolution efficaces reste donc une priorité.
- les contraintes d'équilibre : en océanographie comme en météorologie, des relations physiques plus ou moins complexes existent entre les différentes variables. Ces dernières ne sont donc pas indépendantes les unes des autres mais obéissent à des **contraintes d'équilibre**. On citera par exemple l'équilibre géostrophique mais aussi les relations de la thermodynamique. Il est impératif pour l'assimilation de données de prendre en compte ces contraintes afin d'obtenir une analyse physiquement cohérente (Derber et Bouttier, 1999; Weaver *et al.*, 2005).
- le contrôle de qualité : comme mentionné au paragraphe 1.2.2, il est indispensable que les observations assimilées soient d'une qualité suffisante. Leur contrôle est donc une étape indispensable à l'assimilation de données.
- la correction de biais : tout système d'assimilation est affecté par des biais réduisant la qualité de l'analyse, que ceux-ci proviennent du modèle lui-même, de ces conditions aux frontières, des observations et de leurs opérateurs, ... La difficulté de la correction d'un biais ne réside pas tant dans la méthode numérique elle-même que dans l'identification correcte de sa source (Dee, 2005; Balmaseda *et al.*, 2007).
- les non-linéarités et les distributions non-gaussiennes : les modèles ainsi que les opérateurs d'observation sont généralement non linéaires. Si ces non-linéarités ne sont pas trop sévères, il est possible de les linéariser. Mais lorsqu'elles sont trop fortes, d'autres techniques doivent être envisagées. De plus, pour les méthodes variationnelles, l'hypothèse d'une distribution gaussienne sur les densités de probabilité conditionnelles est faite, malgré son caractère restrictif (voir paragraphe 2.1.3). Prendre en compte des structures non gaussiennes et des non-linéarités, qui sont souvent liées les unes aux autres, est un des challenges de l'assimilation de données.
- la validation : rendre compte de l'impact de l'assimilation de données est malaisé puisque cela nécessite d'évaluer la précision avec laquelle l'état de l'océan (ou de l'atmosphère) est donné, précision évaluée par rapport à un état vrai qu'il n'est jamais possible de connaître parfaitement. La mise en place par exemple de diagnostics de cohérence (Desroziers *et al.*, 2005) peut aider à cette tâche.

Chaque technique d'assimilation a sa propre façon de gérer les divers points évoqués (Lorenç, 2003; Kalnay *et al.*, 2007a). Lorenç (1986) note cependant qu'aucune des méthodes couramment utilisées n'est idéale en tout point et que la préférence que l'on accorde à l'une plutôt qu'une autre dépend des caractéristiques que l'on considère comme les plus importantes. Plutôt que de comparer le 4D-Var à l'EnKF, Gustafsson (2007) suggère de s'intéresser davantage à comment combiner les différentes idées sous-jacentes de ces méthodes (voir aussi Kalnay *et al.*, 2007b). Aujourd'hui ce thème est devenu une activité de recherche très importante (Belo Pereira et Berre, 2006; Daget *et al.*, 2009; Buehner *et al.*, 2009a,b).

## Chapitre 2

# L'assimilation variationnelle

Les développements effectués dans le cadre de cette thèse s'appliquent à un système d'assimilation variationnel. C'est pourquoi, seule cette méthode est décrite en détail dans ce chapitre. Cette méthode va permettre ici de corriger la condition initiale du modèle en combinant les informations fournies d'une part par des mesures du milieu océanique, et d'autre part par l'ébauche du modèle en tant qu'information *a priori*.

### 2.1 Formalisme du problème

L'idée de Sasaki (1958, 1970) est de formuler une analyse objective qui permette de produire un champ minimisant la distance aux observations tout en satisfaisant une contrainte dynamique explicite.

#### 2.1.1 Problème continu

Prenons un système de plusieurs équations différentielles  $S(\varphi)$  où  $\varphi$  est une fonction dépendant de coordonnées spatiales et temporelle (par exemple la fonction température de l'océan). Sur un certain domaine  $\Omega$ , et étant donné la condition initiale  $\varphi_0$ , le système doit vérifier

$$S(\varphi) = 0. \quad (2.1)$$

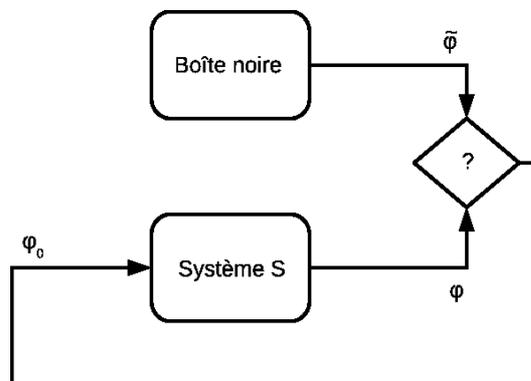


FIG. 2.1 – Système S.

Si la condition initiale  $\varphi_0$  n'est pas connue avec certitude, la résolution du système ( $S$ ) ne donne pas à l'instant  $T$  la fonction  $\varphi$  exacte. Supposons maintenant qu'il existe une fonction  $\tilde{\varphi}$  disponible et représentant une bonne estimation sur le domaine  $\Omega$  de cette fonction exacte à un instant  $t \in ]0, T[$ . Pour cet exemple simple, on considèrera que  $\varphi$  et  $\tilde{\varphi}$  sont définies dans le même espace. On va alors chercher

à réajuster  $\varphi_0$  pour réduire au maximum l'écart quadratique entre  $\varphi$  et l'estimation  $\tilde{\varphi}$  (technique des moindres carrés). Le **problème variationnel** ( $P_c$ ) s'écrit alors

$$(P_c) \begin{cases} \text{Trouver } \varphi(x, t) \text{ qui minimise la fonction } J(\varphi) = \int_{\Omega} (\varphi - \tilde{\varphi})^2 d\Omega \\ \text{Sous la contrainte } S(\varphi) = 0 \end{cases} . \quad (2.2)$$

$S(\varphi) = 0$  représente la **contrainte** du problème de minimisation tandis que  $J(\varphi)$  est appelée **fonction coût** ou **fonctionnelle**. Si on peut munir l'espace dans lequel est défini la fonction  $\varphi$  d'une norme issue du produit scalaire  $L^2$  classique,  $J(\varphi)$  est alors décrite par le carré de la norme  $L^2$

$$J(\varphi) = \|\varphi - \tilde{\varphi}\|_{L^2}^2. \quad (2.3)$$

### 2.1.2 Problème discret

Bien qu'il soit intéressant de garder en mémoire la nature du problème continu, notre application se présente en fait sous la forme d'un problème discret. Les fonctions  $\varphi$  et  $\tilde{\varphi}$  précédentes seront donc représentées sous la forme de vecteurs à valeurs dans  $\mathbb{R}$ .

L'état du modèle est donné par plusieurs champs scalaires mais sera défini sous la forme d'un seul vecteur  $\mathbf{x} = [\mathbf{x}_1^T \mathbf{x}_2^T \dots]^T$  où chaque vecteur  $\mathbf{x}_1, \mathbf{x}_2, \dots$  représente un de ces champs (température, salinité, hauteur d'eau, ...). Il en sera de même pour les observations, quelque soit leur nature. De plus, les espaces respectifs du modèle et des observations n'étant pas identiques, il faut définir un opérateur d'observation qui permettra de passer du premier au second.

On définit ainsi dans l'espace du modèle les vecteurs de dimension  $n$  :

- $\mathbf{x}^t$  : valeurs au point de grille du modèle de l'état vrai continu de l'océan que l'on cherche à approcher au plus près ;
- $\mathbf{x}^b$  : ébauche, information *a priori* fournie par le modèle sur l'état de l'océan ;
- $\mathbf{x}^a$  : analyse, ébauche corrigée après assimilation des observations.

Ces vecteurs contiennent les variables de contrôle définies sur la grille du modèle, c'est à dire les éléments que l'on souhaite corriger grâce au problème inverse. Ils peuvent donc contenir tout ou partie des variables d'état aussi bien que des paramètres divers du système. Néanmoins, le but de notre assimilation de données étant de corriger l'erreur de prévisibilité du modèle, *i.e.* la condition initiale, il ne sera pas fait mention ici de correction de paramètres.

On suppose que l'on veut estimer l'état du système sur une période  $t_0 \leq t_i \leq t_N$  donnée. Cette période est appelée **fenêtre** ou **cycle d'assimilation**. On définit alors dans l'espace des observations le vecteur de dimension  $p$  :

$$\mathbf{y}^o = [\mathbf{y}_0^{oT} \dots \mathbf{y}_N^{oT}]^T : \text{observations de l'océan sur la période donnée,}$$

avec les vecteurs de dimension  $p_i$  :

$$\mathbf{y}_i^o : \text{observations de l'océan à l'instant } t_i.$$

Ces observations correspondent aux mesures effectuées pendant la fenêtre d'assimilation. Elles peuvent être de tout type (*in situ*, satellite ou produit), directes (variables identiques à celles du modèle) ou indirectes (variables en relation avec celles du modèle) et représentent les mesures effectuées. Le nombre de ces mesures peut être différent à l'instant  $t_i$  de celui à l'instant  $t_j$ .

On définit également les opérateurs non linéaires :

- $\mathcal{M}$  : il représente l'intégration de l'ensemble des équations du modèle sur la période ;  
 $\mathcal{M}_{i \rightarrow j}$  : restriction de  $\mathcal{M}$  à l'intégration entre  $t_i$  et  $t_j$  ;  
 $\mathcal{H}$  : opérateur d'observation, il permet de passer de l'espace du modèle à l'espace des observations ;  
 $\mathcal{H}_i$  : restriction de l'opérateur  $\mathcal{H}$  à l'instant  $t_i$ .

Pour exprimer  $\mathbf{y}$  un état intégré jusqu'à un certain temps  $t_i$  dans l'espace des observations, on définit l'opérateur  $\mathcal{G}$  et sa restriction  $\mathcal{G}_i$  par

$$\mathbf{y} = \mathcal{G}[\mathbf{x}] = \begin{pmatrix} \mathcal{G}_0 \\ \vdots \\ \mathcal{G}_i \\ \vdots \\ \mathcal{G}_N \end{pmatrix} = \begin{pmatrix} \mathcal{H}_0[\mathbf{x}] \\ \vdots \\ \mathcal{H}_i[\mathcal{M}_{0 \rightarrow i}[\mathbf{x}]] \\ \vdots \\ \mathcal{H}_N[\mathcal{M}_{0 \rightarrow N}[\mathbf{x}]] \end{pmatrix}. \quad (2.4)$$

Enfin, on définit les erreurs suivantes :

- $\boldsymbol{\varepsilon}^b = \mathbf{x}^b - \mathbf{x}^t$  : erreur d'ébauche ;  
 $\boldsymbol{\varepsilon}^o = \mathbf{y}^o - \mathcal{H}[\mathcal{M}[\mathbf{x}^t]]$  : erreur d'observation (mesure, interpolation, représentativité) ;  
 $\boldsymbol{\varepsilon}_i^o = \mathbf{y}_i^o - \mathcal{H}_i[\mathcal{M}_{0 \rightarrow i}[\mathbf{x}^t]]$  : erreur d'observation à l'instant  $t_i$ .

Cohn (1997) montre qu'il est possible de représenter ces erreurs sous la forme d'un processus stochastique (voir Annexe B). Nous supposons ces erreurs non biaisées<sup>39</sup>, avec pour covariances

$$\mathbf{B} = E[\boldsymbol{\varepsilon}^b (\boldsymbol{\varepsilon}^b)^T] : \text{matrice de covariances d'erreur d'ébauche de dimension } n \times n ;$$

$$\mathbf{R} = E[\boldsymbol{\varepsilon}^o (\boldsymbol{\varepsilon}^o)^T] : \text{matrice de covariances d'erreur d'observation de dimension } p \times p.$$

Nous prendrons de plus comme hypothèse que les erreurs sont décorrélées entre elles<sup>40</sup>.

Le problème variationnel continu ( $P_c$ ) défini précédemment par (2.2) et (2.3) consiste, une fois discrétisé, à estimer l'état vrai  $\mathbf{x}^t$  à partir de l'ébauche  $\mathbf{x}^b$  et des observations  $\mathbf{y}^o$  sous la contrainte donnée par l'opérateur  $\mathcal{G}$  (et donc de l'opérateur de modélisation  $\mathcal{M}$  et de l'opérateur d'observation  $\mathcal{H}$ )

$$(P_d) \left\{ \begin{array}{l} \text{Trouver } \mathbf{x} \text{ qui minimise la fonction } J(\mathbf{x}) = \left\| \begin{pmatrix} \mathbf{x} - \mathbf{x}^b \\ \mathbf{y} - \mathbf{y}^o \end{pmatrix} \right\|^2, \\ \text{Sous la contrainte } \mathbf{y} = \mathcal{G}[\mathbf{x}] \end{array} \right. \quad (2.5)$$

où la norme  $\|\cdot\|$  doit être déterminée pour tenir compte des précisions  $\mathbf{B}^{-1}$  et  $\mathbf{R}^{-1}$ .

## 2.1.3 Définition de la norme

### Exemple simple

Prenons tout d'abord un exemple simple (Talagrand, 1997). On souhaite estimer la quantité  $u$  à l'aide de l'ébauche  $x$  et d'une mesure  $y$

$$\begin{aligned} x &= u + \varepsilon_x \\ y &= u + \varepsilon_y, \end{aligned}$$

<sup>39</sup>On suppose du moins que la moyenne de ces erreurs est connue et peut donc être enlevée. Connaître et enlever ces biais est un des sujets de recherche actuels de l'assimilation de données, mais hors du cadre de cette thèse.

<sup>40</sup>Cette hypothèse est généralement justifiée, le modèle et le système d'observations étant indépendants. Cependant, une attention particulière doit être portée à certaines procédures de pré-traitement des observations qui utilisent l'ébauche du modèle et peuvent alors créer des corrélations artificielles.

où  $\varepsilon_x$  et  $\varepsilon_y$  sont les erreurs. Nous supposons que le modèle et l'instrument de mesure sont sans biais, d'où

$$E[\varepsilon_x] = E[\varepsilon_y] = 0,$$

et de variance d'erreur

$$\begin{aligned}\sigma_x^2 &= E[\varepsilon_x^2], \\ \sigma_y^2 &= E[\varepsilon_y^2],\end{aligned}$$

respectivement. Nous supposons de plus, que les deux erreurs sont décorréliées, le modèle et l'instrument de mesure étant indépendants

$$E[\varepsilon_x \varepsilon_y] = 0.$$

En faisant appel aux moindres carrés pondérés, on cherche un estimateur  $\tilde{u}$  qui soit proche de l'ébauche et de l'observation tout en prenant en compte leurs précisions respectives, *i.e.* l'inverse de leur variance (voir Annexe C). Dans ce cas, pour toute valeur  $\xi$ , on mesure la distance quadratique par

$$J(\xi) = \frac{1}{2} \times \frac{1}{\sigma_x^2} (\xi - x)^2 + \frac{1}{2} \times \frac{1}{\sigma_y^2} (\xi - y)^2,$$

et la distance minimale est donnée par la valeur qui en annule la dérivée

$$\left. \frac{dJ(\xi)}{d\xi} \right|_{\xi=\tilde{u}} = \frac{\tilde{u} - x}{\sigma_x^2} + \frac{\tilde{u} - y}{\sigma_y^2} = 0.$$

L'estimateur que l'on cherche est alors donné par

$$\begin{aligned}\tilde{u} &= \left( \frac{x}{\sigma_x^2} + \frac{y}{\sigma_y^2} \right) \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 + \sigma_y^2} \\ &= \frac{\sigma_y^2}{\sigma_x^2 + \sigma_y^2} x + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2} y.\end{aligned}\tag{2.6}$$

Avec ce cas simple, il est facile d'évaluer l'impact de la précision de chacun des éléments sur l'estimateur. En effet dans le cas idéaliste d'une ébauche  $x$  parfaite (modèle et condition initiale parfaits), *i.e.*  $\sigma_x^2 = 0$ , le poids affecté à l'observation est annulé et l'estimateur prend exactement la valeur de l'ébauche. Dans le cas où ébauche et observation sont de même variance, *i.e.*  $\sigma_x^2 = \sigma_y^2$ , l'estimateur est alors donné par leur moyenne  $x/2 + y/2$ .

## Généralisation

Cohn (1997) suppose que l'état à estimer est en fait donné par un modèle dynamique stochastique, et qu'il peut donc être évalué par une densité de probabilité (pdf<sup>41</sup>). Il s'agit donc de trouver un estimateur  $\mathbf{x}$  (analyse) de pdf conditionnelle  $P^a(\mathbf{x}) = P(\mathbf{x} = \mathbf{x}^t | \mathbf{y} = \mathbf{y}^o)$ , décrivant la probabilité que cet estimateur  $\mathbf{x}$  représente exactement  $\mathbf{x}^t$  étant donné les valeurs observées  $\mathbf{y}^o$ . D'après le théorème de Bayes (voir Annexe B; Wikle et Berliner, 2007), cette pdf est proportionnelle à  $P(\mathbf{y} = \mathbf{y}^o | \mathbf{x} = \mathbf{x}^t)$ , qui décrit la probabilité que les observations aient pour valeurs  $\mathbf{y}^o$  sachant que l'estimateur  $\mathbf{x}$  représente effectivement l'état vrai  $\mathbf{x}^t$ , multipliée par  $P(\mathbf{x} = \mathbf{x}^t)$ , qui quantifie l'information *a priori* que l'on peut avoir sur l'état vrai  $\mathbf{x}^t$  (climatologie, sortie du modèle, ...)

$$P^a(\mathbf{x}) = P(\mathbf{x} = \mathbf{x}^t | \mathbf{y} = \mathbf{y}^o) \propto P(\mathbf{y} = \mathbf{y}^o | \mathbf{x} = \mathbf{x}^t) P(\mathbf{x} = \mathbf{x}^t).\tag{2.7}$$

L'information *a priori*  $P(\mathbf{x} = \mathbf{x}^t)$  est liée à l'ébauche et peut être réécrite (Lorenç, 1986) en

$$P(\mathbf{x} = \mathbf{x}^t) = P^b(\mathbf{x} - \mathbf{x}^b).\tag{2.8}$$

---

<sup>41</sup>Probability Density Function

De la même manière, il est possible de réécrire la pdf des observations sachant l'état vrai comme la pdf des écarts entre l'état vrai intégré par le modèle et ramené dans l'espace des observations grâce à l'opérateur non linéaire  $\mathcal{G}$ , et les observations

$$P(\mathbf{y} = \mathbf{y}^o \mid \mathbf{x} = \mathbf{x}^t) = P^o(\mathcal{G}[\mathbf{x}] - \mathbf{y}^o). \quad (2.9)$$

L'estimateur  $\mathbf{x}$  de l'état vrai  $\mathbf{x}^t$  est donc quantifié par la pdf

$$P^a(\mathbf{x}) \propto P^b(\mathbf{x} - \mathbf{x}^b) P^o(\mathcal{G}[\mathbf{x}] - \mathbf{y}^o). \quad (2.10)$$

A partir de cette distribution, il s'agit maintenant de déterminer l'estimateur optimal, *i.e.* celui qui représente le plus certainement l'état vrai  $\mathbf{x}^t$ . Mais l'optimalité peut se définir de différentes manières, et l'estimateur optimal peut être par exemple celui de la moyenne, du mode (pic) ou encore de la médiane. Lorenc (1986) et Cohn (1997) montrent que si l'estimateur de la **moyenne conditionnelle** est à la base des méthodes séquentielles (voir Annexe C), c'est celui du **mode conditionnel** qui nous intéresse et que nous développons ici puisqu'il est à la base des méthodes variationnelles. On notera néanmoins, que sous les hypothèses de linéarité et de distributions conditionnelles gaussiennes les estimateurs de la moyenne conditionnelle et du mode conditionnel sont équivalents (Jazwinski, 1970; Lorenc, 1986; Cohn, 1997).

L'estimateur du **mode conditionnel** est celui qui correspond au pic, *i.e.* au maximum, de la pdf qui quantifie l'information *a posteriori*. Il est donc donné par

$$\mathbf{x} \text{ tel que } P^a(\mathbf{x}) \text{ est maximale.}$$

Si on prend pour hypothèse, que les pdf des équations (2.8) et (2.9) sont des distributions gaussiennes (voir paragraphe 2.3.1 pour une justification de cette hypothèse)

$$\begin{aligned} P^b(\mathbf{x} - \mathbf{x}^b) &\propto \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) \right], \\ P^o(\mathcal{G}[\mathbf{x}] - \mathbf{y}^o) &\propto \exp \left[ -\frac{1}{2} (\mathcal{G}[\mathbf{x}] - \mathbf{y}^o)^T \mathbf{R}^{-1} (\mathcal{G}[\mathbf{x}] - \mathbf{y}^o) \right], \end{aligned}$$

l'équation (2.10) devient

$$P^a(\mathbf{x}) \propto \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) - \frac{1}{2} (\mathcal{G}[\mathbf{x}] - \mathbf{y}^o)^T \mathbf{R}^{-1} (\mathcal{G}[\mathbf{x}] - \mathbf{y}^o) \right].$$

Pour l'estimation du mode conditionnel, on cherche à maximiser la pdf  $P^a(\mathbf{x})$ , ce qui est équivalent à minimiser l'expression  $-\ln P^a(\mathbf{x})$  pour obtenir le maximum de vraisemblance. Ainsi, la fonction coût à minimiser est donnée par

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

Les matrices  $\mathbf{B}$  et  $\mathbf{R}$  sont des matrices de covariances et sont donc symétriques et définies positives par définition<sup>42</sup> (voir Annexe B). Leurs inverses  $\mathbf{B}^{-1}$  et  $\mathbf{R}^{-1}$  sont donc également symétriques et définies positives. On peut alors réécrire l'équation précédente sous la forme d'une somme de produits scalaires ou de normes

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2} \left\langle \mathbf{x} - \mathbf{x}^b, \mathbf{x} - \mathbf{x}^b \right\rangle_{\mathbf{B}^{-1}} + \frac{1}{2} \left\langle \mathcal{G}[\mathbf{x}] - \mathbf{y}^o, \mathcal{G}[\mathbf{x}] - \mathbf{y}^o \right\rangle_{\mathbf{R}^{-1}} \\ &= \frac{1}{2} \left\| \mathbf{x} - \mathbf{x}^b \right\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \left\| \mathcal{G}[\mathbf{x}] - \mathbf{y}^o \right\|_{\mathbf{R}^{-1}}^2, \end{aligned}$$

<sup>42</sup>Une matrice de covariance est par définition *semi*-définie positive. Cependant, l'hypothèse est faite ici, et dans tout ce qui suit, de matrices  $\mathbf{B}$  et  $\mathbf{R}$  *strictement* définies positives et donc inversibles. On pourra se référer à Daley (1991), pp 109–111, pour une justification de cette hypothèse.

où la norme  $\|\cdot\|_{\mathbf{A}}$  dérive du produit scalaire  $\langle \cdot, \cdot \rangle_{\mathbf{A}}$ , avec  $\mathbf{A}$  correspondant à  $\mathbf{B}^{-1}$  ou  $\mathbf{R}^{-1}$ , défini par

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}} = \sum_i \sum_j x_i a_{ij} y_j = \sum_i x_i z_i \quad \text{avec} \quad z_i = \sum_j a_{ij} y_j,$$

où  $a_{ij}$  représente l'élément de la ligne  $i$  et de la colonne  $j$  de la matrice  $\mathbf{A}$ . Ce produit scalaire correspond donc à une imbrication de produits scalaires euclidiens canoniques, et peut également s'écrire comme un produit scalaire canonique pondéré par le poids de  $\mathbf{A}$

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}} = \langle \mathbf{x}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{A}\mathbf{y} \rangle. \quad (2.11)$$

On peut encore généraliser par

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2} \left\langle \left( \begin{array}{c} \mathbf{x} - \mathbf{x}^b \\ \mathcal{G}[\mathbf{x}] - \mathbf{y}^o \end{array} \right), \left( \begin{array}{c} \mathbf{x} - \mathbf{x}^b \\ \mathcal{G}[\mathbf{x}] - \mathbf{y}^o \end{array} \right) \right\rangle_{\mathbf{P}} \\ &= \frac{1}{2} \left\| \left( \begin{array}{c} \mathbf{x} - \mathbf{x}^b \\ \mathcal{G}[\mathbf{x}] - \mathbf{y}^o \end{array} \right) \right\|_{\mathbf{P}}^2, \end{aligned}$$

où  $\mathbf{P}$  est la matrice de précision de dimension  $(n+p) \times (n+p)$  donnée par

$$\mathbf{P} = \begin{pmatrix} \mathbf{B}^{-1} & 0 \\ 0 & \mathbf{R}^{-1} \end{pmatrix}.$$

On notera que pour l'estimateur du mode conditionnel, aucune assumption de linéarité n'est faite pour l'opérateur d'observation. Sous l'hypothèse de distributions gaussiennes, la fonction coût  $J(\mathbf{x})$  définie pour le problème variationnel ( $P_d$ ) de l'équation (2.5) peut également être vue sous la forme d'une norme  $L^2$  (discrétisée). Cependant, cette hypothèse de distribution gaussienne peut être relâchée et conduit alors à des normes différentes, plus complexes à manipuler.

## 2.2 Les différentes formulations

### 2.2.1 La formulation du 4D-Var

#### Modèle parfait

Dans un premier temps, nous considérons que le modèle est parfait, *i.e.* que l'on peut négliger l'erreur modèle par rapport aux erreurs d'ébauche et d'observations. **L'assimilation variationnelle 4D-Var à contrainte forte** consiste alors à minimiser la fonction coût

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2} \|\mathbf{x} - \mathbf{x}^b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \|\mathbf{y} - \mathbf{y}^o\|_{\mathbf{R}^{-1}}^2 \\ &= \underbrace{\frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^{\mathbf{T}} \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b)}_{J_b} + \underbrace{\frac{1}{2} (\mathbf{y} - \mathbf{y}^o)^{\mathbf{T}} \mathbf{R}^{-1} (\mathbf{y} - \mathbf{y}^o)}_{J_o}, \end{aligned} \quad (2.12)$$

sous la contrainte forte

$$\mathbf{y} = \mathcal{G}[\mathbf{x}], \quad (2.13)$$

où  $\mathcal{G}$  est défini par l'équation (2.4). Dans cette formulation, la fonction coût se décompose en deux termes : le terme  $J_b$  représente la distance à l'ébauche, *i.e.* à l'information *a priori*, pondérée par la confiance que l'on accorde à cette ébauche ; le terme  $J_o$  représente la distance aux observations par rapport à un terme intégré par le modèle jusqu'aux temps de ces observations, pondéré par la confiance que l'on accorde à ces observations.

Si on considère que les erreurs d'observations sont décorrélées dans le temps, et que l'on appelle  $\mathbf{R}_i$  le bloc de  $\mathbf{R}$  associé à l'instant  $t_i$ , le terme  $J_o$  peut être réécrit sous la forme d'une somme

$$J_o(\mathbf{x}) = \frac{1}{2} \sum_{i=0}^N (\mathcal{H}_i[\mathbf{x}_i] - \mathbf{y}_i^o)^\top \mathbf{R}_i^{-1} (\mathcal{H}_i[\mathbf{x}_i] - \mathbf{y}_i^o),$$

sous la contrainte

$$\forall i = 0, \dots, N \quad \mathbf{x}_i = \mathcal{M}_{0 \rightarrow i}[\mathbf{x}].$$

### Modèle avec erreur

Dans un deuxième temps, nous considérons que l'erreur associée au modèle est trop importante pour être négligée. On définit alors :

$$\begin{aligned} \boldsymbol{\varepsilon}^q &= \left[ \boldsymbol{\varepsilon}_0^q \dots \boldsymbol{\varepsilon}_N^q \right]^\top : \text{erreur totale du modèle avec} \\ \boldsymbol{\varepsilon}_i^q &= \mathbf{x}_i^t - \mathcal{M}_{i-1 \rightarrow i}[\mathbf{x}_{i-1}^t] : \text{erreur à l'instant } t_i ; \\ \mathbf{Q} &= E \left[ \boldsymbol{\varepsilon}^q (\boldsymbol{\varepsilon}^q)^\top \right] : \text{matrice de covariances d'erreur du modèle.} \end{aligned}$$

Prendre en compte cette erreur peut alors se faire de différentes manières suivant les hypothèses que l'on fait (Trémolet, 2007b). On expose ici une écriture possible en définissant :

$$\begin{aligned} \boldsymbol{\eta} &= \left[ \boldsymbol{\eta}_0^\top \dots \boldsymbol{\eta}_N^\top \right]^\top : \text{estimation de l'erreur } \boldsymbol{\varepsilon}^q \text{ avec} \\ \boldsymbol{\eta}_i &: \text{estimation de l'erreur } \boldsymbol{\varepsilon}_i^q . \end{aligned}$$

Un terme contrôlant la minimisation de  $\boldsymbol{\eta}$  est alors ajouté à la fonction coût, et l'**assimilation variationnelle 4D-Var à contrainte faible** consiste alors à minimiser

$$J(\mathbf{x}, \boldsymbol{\eta}) = \underbrace{\frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^\top \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b)}_{J_b} + \underbrace{\frac{1}{2} (\mathbf{y} - \mathbf{y}^o)^\top \mathbf{R}^{-1} (\mathbf{y} - \mathbf{y}^o)}_{J_o} + \underbrace{\frac{1}{2} \boldsymbol{\eta}^\top \mathbf{Q}^{-1} \boldsymbol{\eta}}_{J_q}, \quad (2.14)$$

sous la contrainte faible

$$\mathbf{y} = \mathcal{G}_f[\mathbf{x}],$$

avec

$$\mathcal{G}_f[\mathbf{x}] = \begin{bmatrix} \mathcal{H}_0[\mathbf{x}] \\ \mathcal{H}_1[\mathcal{M}_{0 \rightarrow 1}[\mathbf{x}] + \boldsymbol{\eta}_1] \\ \mathcal{H}_2[\mathcal{M}_{1 \rightarrow 2}[\mathcal{M}_{0 \rightarrow 1}[\mathbf{x}] + \boldsymbol{\eta}_1] + \boldsymbol{\eta}_2] \\ \vdots \end{bmatrix}.$$

Dans ce paragraphe, nous avons considéré que la contrainte était entièrement liée à la modélisation. Il est cependant possible d'envisager d'autres types de contraintes d'erreur  $\boldsymbol{\varepsilon}^c$ , et d'ajouter alors un nouveau terme  $J_c$  à la fonction coût donnée par l'équation (2.14).

La formulation contrainte faible et son implémentation est un des sujets de recherche actuels de l'assimilation de données en océanographie et météorologie, mais hors cadre de cette thèse. Nous n'irons donc pas plus avant dans sa description. Dans la suite de ce manuscrit, et sauf mention contraire, le modèle sera toujours considéré comme parfait (contrainte forte).

### Résolution

Il existe plusieurs méthodes pour résoudre un problème de minimisation sous contrainte. On peut par exemple choisir de réduire la contrainte, *i.e.* éliminer les variables que l'on sait mettre en relation avec d'autres, de manière à transformer le problème en un problème sans contrainte. La contrainte (forte) peut par exemple n'être exprimée qu'en fonction d'un état initial (Le Dimet et Talagrand, 1986;

Talagrand et Courtier, 1987). Une autre méthode consiste à associer des multiplicateurs de Lagrange au problème et à trouver les points stationnaires de la fonction Lagrangienne associée (Thacker et Long, 1988; Chua et Bennett, 2001; Bennett, 2002).

## 2.2.2 La formulation incrémentale

Le problème du 4D-Var exposé dans le paragraphe précédent est difficile à résoudre lorsque les opérateurs  $\mathcal{M}$  et  $\mathcal{H}$  (et donc  $\mathcal{G}$ ) ne sont pas linéaires. La fonction coût peut présenter plusieurs minima locaux et sa minimisation en devient d'autant plus coûteuse, sans pouvoir garantir un résultat optimal. Or pour l'océanographie comme pour la météorologie, prendre pour hypothèse la linéarité de ces opérateurs est peu réaliste compte tenu de la nature des équations du modèle et du type de mesures couramment utilisées. En supposant que cette non-linéarité n'est pas trop forte, Courtier *et al.* (1994) se basent sur la méthode de Gauss-Newton (voir par exemple Lawless *et al.*, 2005) pour transformer le problème de minimisation non-quadratique en une séquence de minimisations de fonctions quadratiques. Dans la description qui en est faite ci-après, on montre que cette méthode ne travaille plus sur les champs eux-mêmes mais sur des incréments, *i.e.* des perturbations de ces champs.

### Linéarisation des opérateurs

En premier lieu, on supposera qu'il est possible de déterminer l'état de l'océan  $\mathbf{x}_i$  à tout instant  $t_i$  de la fenêtre d'assimilation en fonction de son état à l'instant  $t_{i-1}$  précédent. Cette hypothèse se justifie par la nature causale du modèle. Si on simplifie l'écriture

$$\mathcal{M}_{i-1 \rightarrow i}[\mathbf{x}] \iff \mathcal{M}_i \mathbf{x},$$

on a alors l'expression  $\mathbf{x}_i = \mathcal{M}_i \mathbf{x}_{i-1}$ . Par récurrence, on obtient

$$\mathbf{x}_i = \mathcal{M}_i \circ \mathcal{M}_{i-1} \circ \dots \circ \mathcal{M}_1 \mathbf{x},$$

où  $\mathbf{x} = \mathbf{x}_0$  est la condition initiale et  $\mathcal{M}_0$  est représentée par la matrice identité.

On définit maintenant les opérateurs  $\widetilde{\mathcal{M}}$  et  $\widetilde{\mathcal{H}}$  en simplifiant les opérateurs  $\mathcal{M}$  et  $\mathcal{H}$  afin de les rendre linéarisables (et éventuellement moins coûteux). On s'affranchit par exemple des processus à seuil, ou encore on relativise les fortes non linéarités (voir par exemple Weaver *et al.*, 2003). Les opérateurs  $\widetilde{\mathcal{M}}$  et  $\widetilde{\mathcal{H}}$  peuvent donc être linéarisés au voisinage d'un point déterminé, grâce à un développement de Taylor du premier ordre. Au voisinage de l'état arbitraire  $\mathbf{x}^c$ , on peut alors écrire pour une perturbation  $h$  à l'instant  $t_i$

$$\begin{aligned} \widetilde{\mathcal{M}}_i[\mathbf{x}^c + h] &= \widetilde{\mathcal{M}}_i[\mathbf{x}^c] + \mathbf{M}_i h + O(\|h\|^2) \\ \widetilde{\mathcal{H}}_i[\mathbf{x}^c + h] &= \widetilde{\mathcal{H}}_i[\mathbf{x}^c] + \mathbf{H}_i h + O(\|h\|^2). \end{aligned} \quad (2.15)$$

$\mathbf{M}_i$  et  $\mathbf{H}_i$  représentent la première différentielle, ou **linéaire tangent**, des opérateurs  $\widetilde{\mathcal{M}}_i$  et  $\widetilde{\mathcal{H}}_i$  respectivement au point  $\mathbf{x}^c$

$$\mathbf{M}_i = \left. \frac{\partial \widetilde{\mathcal{M}}_i}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}^c} \quad (2.16)$$

$$\mathbf{H}_i = \left. \frac{\partial \widetilde{\mathcal{H}}_i}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}^c}. \quad (2.17)$$

Bouttier et Courtier (1999) pp. 16 et 17 discutent la validité du linéaire tangent et des hypothèses qui le soutendent.

En linéarisant l'opérateur  $\widetilde{\mathcal{M}}$  autour de l'ébauche  $\mathbf{x}^b$  pour la perturbation  $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b$ , on obtient

$$\begin{aligned}\mathbf{x}_i &= \widetilde{\mathcal{M}}_{0 \rightarrow i}[\mathbf{x}] \\ &= \widetilde{\mathcal{M}}_{0 \rightarrow i}[\mathbf{x}^b + \delta\mathbf{x}] \\ &\approx \widetilde{\mathcal{M}}_{0 \rightarrow i}[\mathbf{x}^b] + \mathbf{M}_i \mathbf{M}_{i-1} \cdots \mathbf{M}_1 \delta\mathbf{x}.\end{aligned}$$

En linéarisant l'opérateur  $\widetilde{\mathcal{H}}_i$  autour de  $\mathbf{x}_i$  à l'instant  $t_i$  pour la perturbation  $\delta\mathbf{x}_i$  on a

$$\begin{aligned}\widetilde{\mathcal{H}}_i[\mathbf{x}_i] &\approx \widetilde{\mathcal{H}}_i[\widetilde{\mathcal{M}}_{0 \rightarrow i}[\mathbf{x}^b] + \mathbf{M}_i \mathbf{M}_{i-1} \cdots \mathbf{M}_1 \delta\mathbf{x}] \\ &\approx \widetilde{\mathcal{H}}_i[\widetilde{\mathcal{M}}_{0 \rightarrow i}[\mathbf{x}^b]] + \mathbf{H}_i \mathbf{M}_i \mathbf{M}_{i-1} \cdots \mathbf{M}_1 \delta\mathbf{x}.\end{aligned}$$

### Le 4D-Var incrémental

Si on définit l'opérateur

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_0 \\ \vdots \\ \mathbf{G}_i \\ \vdots \\ \mathbf{G}_N \end{bmatrix} = \begin{bmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_i \mathbf{M}_i \cdots \mathbf{M}_1 \\ \vdots \\ \mathbf{H}_N \mathbf{M}_N \cdots \mathbf{M}_1 \end{bmatrix}, \quad (2.18)$$

la formulation du **4D-Var incrémental** consiste à minimiser sans contrainte la fonction coût

$$J(\delta\mathbf{x}) = \underbrace{\frac{1}{2} \delta\mathbf{x}^T \mathbf{B}^{-1} \delta\mathbf{x}}_{J_b} + \underbrace{\frac{1}{2} (\mathbf{G} \delta\mathbf{x} - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{G} \delta\mathbf{x} - \mathbf{d})}_{J_o}, \quad (2.19)$$

où on définit l'incrément par rapport à l'ébauche par

$$\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b,$$

et le vecteur d'innovation par

$$\mathbf{d} = \mathbf{y}^o - \mathcal{G}[\mathbf{x}^b].$$

Comme précédemment, si on considère que les erreurs d'observations sont décorrélées dans le temps, et en appelant  $\mathbf{G}_i$  la restriction de  $\mathbf{G}$  à l'instant  $t_i$ , le terme  $J_o$  peut être réécrit sous la forme d'une somme

$$J_o(\delta\mathbf{x}) = \frac{1}{2} \sum_{i=0}^N (\mathbf{G}_i \delta\mathbf{x} - \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{G}_i \delta\mathbf{x} - \mathbf{d}_i),$$

où l'innovation à l'instant  $t_i$  est donnée par

$$\mathbf{d}_i = \mathbf{y}_i^o - \mathcal{H}_i[\mathcal{M}_{0 \rightarrow i}[\mathbf{x}^b]].$$

L'incrément  $\delta\mathbf{x}^a$  optimal (voir figure 2.2) est ajouté à l'ébauche fournie par le cycle d'assimilation précédent pour obtenir l'état d'analyse optimal  $\mathbf{x}^a$

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

Lorsqu'elle est propagée par le modèle, cette analyse fournit une trajectoire correspondant à celle de l'ébauche propagée par le modèle mais ajustée par les observations disponibles.

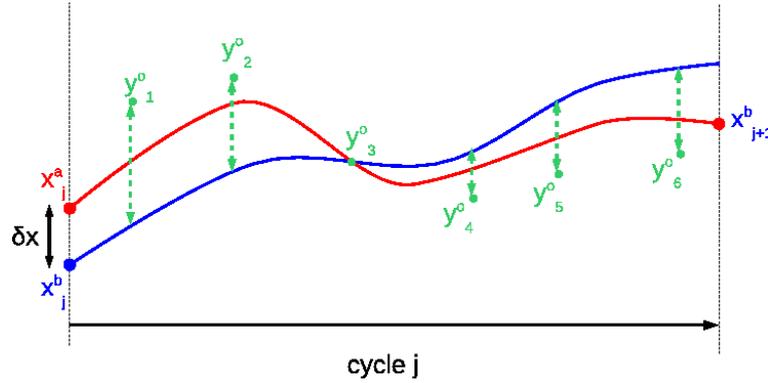


FIG. 2.2 – L'ébauche fournie par le cycle  $j - 1$  précédent est propagée par le modèle (courbe bleue) sur toute la fenêtre d'assimilation (cycle  $j$ ). À chaque pas de temps, les innovations (traits verts pointillés) sont calculées pour être comparées à l'incrément propagé par le modèle linéarisé. La correction  $\delta \mathbf{x}$  qui minimise la fonction coût est ajoutée à l'ébauche pour déterminer l'analyse qui sera à son tour propagée par le modèle (courbe rouge) et servira d'ébauche au cycle  $j + 1$  suivant.

### Le 4D-Var multi-incrémental

La Hessienne de la fonction (2.19) étant définie positive, elle ne possède qu'un minimum global. Cependant, la simplification puis la linéarisation des opérateurs représentent des approximations pouvant amener des écarts conséquents pour la solution, en particulier lorsque l'ébauche en est assez éloignée. L'algorithme de Gauss-Newton sépare alors la minimisation en deux boucles imbriquées :

- la **boucle interne** (“*inner loop*”) : son rôle est d'effectuer la minimisation de la fonction coût quadratique à l'aide d'opérateurs  $\mathbf{M}$  et  $\mathbf{H}$  issus de la linéarisation des opérateurs simplifiés  $\tilde{\mathbf{M}}$  et  $\tilde{\mathbf{H}}$ ;
- la **boucle externe** (“*outer loop*”) : elle tient compte du modèle non-linéaire complet et calcule les éléments nécessaires à l'interfaçage avec la boucle interne.

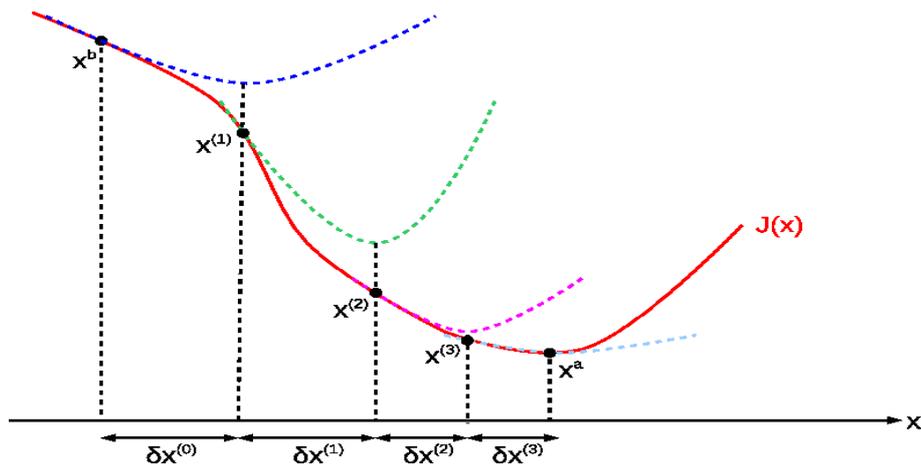


FIG. 2.3 – La courbe rouge pleine représente la fonction coût non-linéaire à minimiser. Une première linéarisation (courbe bleue pointillée) est effectuée autour de l'ébauche  $\mathbf{x}^b$ . La boucle interne permet de trouver l'incrément  $\delta \mathbf{x}^{(0)}$  qui la minimise, et qui, ajouté à l'ébauche, permet de trouver le point  $\mathbf{x}^{(1)}$ . Une nouvelle linéarisation est faite autour de ce point (courbe verte pointillée), donnant un nouvel incrément minimum et un nouveau point  $\mathbf{x}^{(2)}$ . Le processus est réitéré jusqu'à trouver le point  $\mathbf{x}^a$ , minimum de la fonction coût non linéaire.

Le problème est donc ramené à une suite itérative de minimisations quadratiques de la forme (2.19) pour évaluer le minimum de la fonction coût non-linéaire (2.12) (voir figure 2.3).



### 2.2.3 Le 3D-Var, une simplification du 4D-Var

L'algorithme du 4D-Var incrémental reste une technique qui nécessite des moyens de calculs conséquents. Une simplification possible consiste à ne pas tenir compte de la variable temps, *i.e.* à ne pas propager l'incrément par le modèle. En d'autres termes, la simplification puis la linéarisation de l'opérateur  $\mathcal{M}$  conduit à la matrice identité  $\mathbf{M}_i = \mathbf{I}$  quelque soit l'instant  $t_i$  et ainsi  $\mathbf{G} = \mathbf{H}$ . La fonction coût (2.19) devient donc pour le **3D-Var incrémental**

$$J(\delta\mathbf{x}) = \underbrace{\frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x}}_{J_b} + \underbrace{\frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})}_{J_o}, \quad (2.20)$$

où le terme  $J_o$  peut se réécrire en fonction de chaque instant  $t_i$  si les erreurs d'observations sont décorréliées dans le temps en

$$J_o(\delta\mathbf{x}) = \frac{1}{2} \sum_{i=0}^N (\mathbf{H}_i\delta\mathbf{x} - \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i\delta\mathbf{x} - \mathbf{d}_i).$$

Dans l'étape 4 de l'algorithme précédent, seul l'opérateur  $\mathcal{H}$  est simplifié et linéarisé.

Le calcul des innovations peut être effectué, comme pour le 4D-Var, en tenant compte des instants  $t_i$  des observations

$$\mathbf{d}_i = \mathbf{y}_i^o - \mathcal{H}_i[\mathcal{M}_{0 \rightarrow i}[\mathbf{x}^b]].$$

Dans ce cas, on parle de **3D-Var FGAT**<sup>44</sup>. Mais il est possible de simplifier encore, et de calculer les innovations par rapport à une ébauche non propagée ou propagée jusqu'à un certain instant  $t_c$  si on souhaite effectuer une correction à un instant différent de  $t_0$

$$\mathbf{d}_i = \mathbf{y}_i^o - \mathcal{H}_i[\mathcal{M}_{0 \rightarrow c}[\mathbf{x}^b]].$$

Pour cette formulation du **3D-Var incrémental classique**, on choisit en général d'effectuer la correction au milieu de la fenêtre d'assimilation à  $t_{N/2}$  (voir figure 2.4).

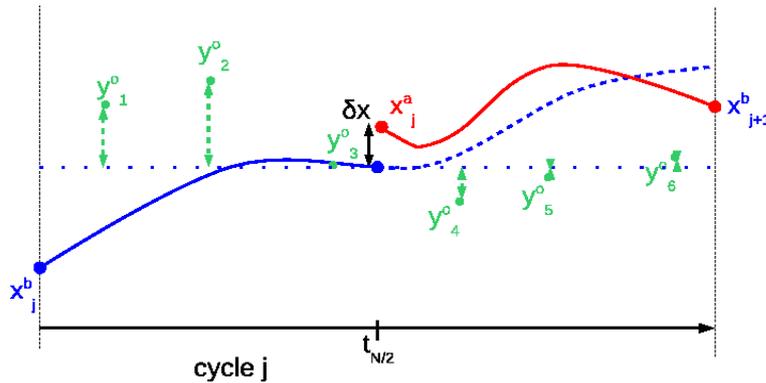


FIG. 2.4 – Dans cet exemple, l'ébauche est propagée jusqu'en milieu de fenêtre (courbe bleue pleine) avant d'être comparée aux observations (trait vert) sans tenir compte de l'instant de mesure. Un incrément est alors défini qui, ajouté à l'ébauche donne l'analyse. Cette analyse est propagée le long de la deuxième moitié de la fenêtre d'assimilation (courbe rouge) et donne l'ébauche pour le cycle suivant. La courbe en bleu pointillé correspond à l'intégration de l'ébauche sur la deuxième moitié de la fenêtre et donc à l'état sans assimilation.

Les corrections calculées par les formulations du 3D-Var classique et du 3D-Var FGAT peuvent produire un choc assez brutal dans le modèle lorsqu'elles sont appliquées. Pour éviter cet inconvénient

<sup>44</sup>First Guess at Appropriate Time.

Bloom *et al.* (1996) proposent d'introduire l'incrément comme un forçage tout au long de la fenêtre d'assimilation. Cette technique, appelée IAU<sup>45</sup> peut permettre d'obtenir une intégration du modèle continue d'un cycle à l'autre.

### 2.2.4 La formulation duale : 4D-PSAS et 3D-PSAS

Dans la formulation incrémentale du 4D-Var, la fonction coût est minimisée par rapport à une variable de contrôle dans l'espace du modèle. Or, dans le paragraphe 1.3.4, nous avons vu que le nombre d'observations reste, en océanographie, inférieur à la taille d'une grille modèle même basse résolution. Il serait donc avantageux de tirer parti de ce constat pour réduire le coût de la minimisation. Le système PSAS<sup>46</sup> (Egbert *et al.*, 1994; Amodei, 1995; Cohn *et al.*, 1998) opère directement dans l'espace des observations, plutôt que dans l'espace du modèle. Courtier (1997) en formalise l'approche et définit une fonction coût par rapport à une variable de contrôle dans l'espace des observations. La dualité des formulations 3D-Var et PSAS, *i.e.* leur équivalence, y est établie en montrant que le conditionnement des problèmes est identique. L'extension à la dimension temporelle est ensuite indiquée pour les formulations 4D-Var et 4D-PSAS.

Dans l'exemple du 3D-Var incrémental, minimiser la fonction coût (2.20), c'est trouver l'incrément optimal  $\delta \mathbf{x}^a$  qui annule son gradient

$$\nabla_{\delta \mathbf{x}} J = \mathbf{B}^{-1} \delta \mathbf{x} + \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{H} \delta \mathbf{x} - \mathbf{d}),$$

et donc résoudre l'équation

$$(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \delta \mathbf{x}^a = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d} \quad \Leftrightarrow \quad \delta \mathbf{x}^a = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d}.$$

En appliquant la formule de Sherman-Morrison-Woodbury

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{D} \mathbf{A}^{-1} \mathbf{B} + \mathbf{C}^{-1})^{-1} \mathbf{D} \mathbf{A}^{-1},$$

où  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  et  $\mathbf{D}$  sont des matrices, on peut écrire

$$\begin{aligned} (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} &= \left( \mathbf{B} - \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{B} \right) \mathbf{H}^T \mathbf{R}^{-1} \\ &= \mathbf{B} \mathbf{H}^T \mathbf{R}^{-1} - \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{B} \mathbf{H}^T \mathbf{R}^{-1} \\ &= \mathbf{B} \mathbf{H}^T \left( \mathbf{I} - (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{B} \mathbf{H}^T \right) \mathbf{R}^{-1} \\ &= \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \left( (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}) - \mathbf{H} \mathbf{B} \mathbf{H}^T \right) \mathbf{R}^{-1} \\ &= \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1}, \end{aligned}$$

et on a

$$\delta \mathbf{x}^a = \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{d}. \quad (2.21)$$

Cette formulation est équivalente à la formulation du BLUE<sup>47</sup> établie en Annexe C. On considère alors  $\mathbf{z}$ , vecteur de l'espace dual des observations<sup>48</sup> et solution du problème

$$(\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}) \mathbf{z} = \mathbf{d}.$$

<sup>45</sup>Incremental Analysis Update

<sup>46</sup>Physical-space Statistical Analysis System

<sup>47</sup>Best Linear Unbiased Estimator

<sup>48</sup>L'innovation est définie dans l'espace des observations. Puisque le vecteur  $\mathbf{Cz}$ , où  $\mathbf{C} = \mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}$  est une matrice de covariance, est un élément de l'espace des observations, alors le vecteur  $\mathbf{z}$  est un élément du dual de l'espace des observations de par la définition des opérateurs de covariance (voir Annexe B).

Cette équation peut être vue comme un gradient que l'on annule, c'est à dire que l'on cherche le minimum de la fonction coût associée

$$F(\mathbf{z}) = \frac{1}{2} \mathbf{z}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}) \mathbf{z} - \mathbf{z}^T \mathbf{d}. \quad (2.22)$$

C'est la formulation du **3D-PSAS**.

De la même manière, on peut définir la fonction coût du **4D-PSAS** par

$$F(\mathbf{z}) = \frac{1}{2} \mathbf{z}^T (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}) \mathbf{z} - \mathbf{z}^T \mathbf{d}, \quad (2.23)$$

où  $\mathbf{G}$  est défini par l'équation (2.18).

## 2.3 Les matrices de covariances d'erreur

Pour prendre en compte le fait qu'une certaine incertitude existe<sup>49</sup> dans l'ébauche et les observations, il est nécessaire de définir et de modéliser les statistiques des erreurs associées  $\varepsilon^b$  et  $\varepsilon^o$  (somme des erreurs de mesure, d'interpolation et de représentativité). Dans le paragraphe 2.1.3, nous avons pris comme hypothèse que ces erreurs avaient une distribution gaussienne. Ce postulat, argumenté au paragraphe 2.3.1 suivant, nous permet de réduire les statistiques à connaître aux deux premiers moments. En ce qui concerne la moyenne, nous avons déjà supposé que ces erreurs étaient non biaisées, ou pour le moins débiaisées (voir paragraphe 2.1.2). Il nous reste donc à définir ou estimer les covariances, *i.e.* les matrices  $\mathbf{B}$  pour l'erreur d'ébauche, et  $\mathbf{R}$  pour celle des observations. Mais ces erreurs étant définies par rapport à un état vrai que l'on ne connaît pas, il n'est pas possible de calculer ces deuxièmes moments de manière explicite. Il est donc nécessaire, soit de trouver une méthode permettant de les estimer, soit d'émettre de nouvelles hypothèses sur leurs structures.

Les covariances sont le produit de deux éléments, les variances d'une part, et les corrélations d'autre part, qui influencent différemment la correction. Tout d'abord, le poids accordé à l'ébauche et aux observations dans le calcul de cette correction est directement lié à leur précision, et donc à l'inverse des variances de leurs erreurs. Les corrélations de l'erreur d'ébauche permettent ensuite de lisser la correction sur la grille du modèle.

### 2.3.1 Nature des erreurs

On prend ici l'exemple d'une modélisation de plusieurs variables sur une grille de  $P$  points de coordonnées spatiales et temporelle  $(x, y, z, t)$ .

En se plaçant en un point particulier  $(a, b, c)$  et à un certain instant  $d$ , on souhaite connaître pour la variable  $\xi$  l'erreur  $\varepsilon^\xi(a, b, c, d)$  qui existe entre la valeur donnée par la modélisation et la valeur vraie. Si on considère cette erreur comme une **variable aléatoire** (voir Annexe B), il est possible d'en calculer des statistiques (moyenne et variance par exemple) à partir d'un nombre suffisant de réalisations de l'expérience qui la génère, dans des conditions identiques (s'agissant ici d'une expérience de modélisation numérique, le caractère aléatoire sera simulé en perturbant des paramètres différemment pour chaque réalisation). En supposant que cette erreur est en fait la somme d'un grand nombre de petites contributions indépendantes, le théorème central-limite (Saporta, 2006, p. 66) nous permet de considérer que cette erreur suit une loi gaussienne

$$\varepsilon^\xi(a, b, c, d) = \xi(a, b, c, d) - \xi^t(a, b, c, d) \sim \mathcal{N}(0, \sigma^2),$$

<sup>49</sup>Le modèle sera ici considéré comme parfait, son erreur  $\varepsilon^a$  sera donc négligée.

où on considèrera ici que l'erreur est non biaisée et de variance  $\sigma^2$ . En généralisant à chaque point de la grille et à tout instant, on obtient un ensemble de variables aléatoires de distributions gaussiennes que l'on peut représenter sous la forme du vecteur

$$\boldsymbol{\varepsilon}^\xi = (\varepsilon_1^\xi, \varepsilon_2^\xi, \dots, \varepsilon_P^\xi) \quad \text{avec} \quad \varepsilon_i^\xi = \varepsilon^\xi(a_i, b_i, c_i, d_i) \sim \mathcal{N}(0, \sigma_i^2).$$

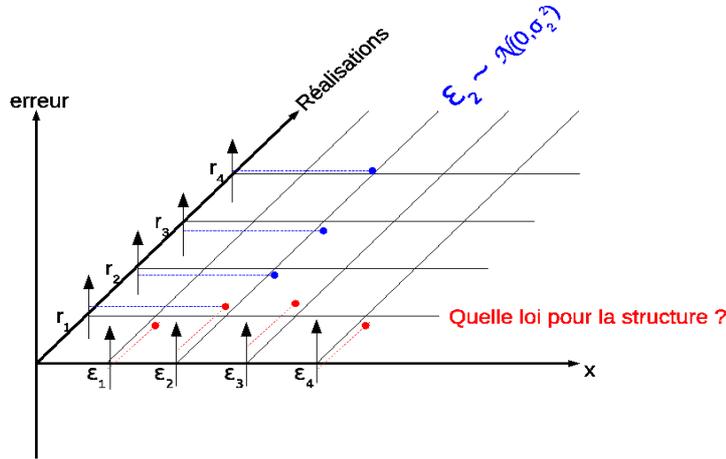


FIG. 2.5 – En fixant les coordonnées  $y, z$  et  $t$  pour la variable  $\xi$ , on illustre ici les différentes variables aléatoires  $\varepsilon_i$ ,  $i = 1, 2, \dots$ , représentant l'erreur suivant l'axe  $x$ . On estime que chacune de ces erreurs suit une loi normale  $\mathcal{N}(0, \sigma_i^2)$ . Mais à moins que ces variables aléatoires ne soient indépendantes, la loi suivie par la fonction aléatoire, constituée de l'ensemble de ces différentes variables, n'est pas connue.

Cependant, pour bien comprendre l'erreur qui affecte la variable  $\xi$ , la définition précédente n'est pas suffisante. En effet, il faut savoir également de quelle manière chacune de ces variables aléatoires agit sur les autres. Pour cela il est nécessaire de définir, non pas les statistiques de chaque variable aléatoire, mais celles de la structure complète. Il faut donc considérer l'erreur comme un **champ aléatoire** (voir figure 2.5). La question est donc maintenant de connaître la distribution de cette structure. Déterminer la loi jointe de ces variables aléatoires peut s'avérer difficile, mais si on suppose que les variables aléatoires sont indépendantes les unes des autres alors la pdf de la fonction aléatoire correspond au produit des pdfs de chaque variable aléatoire, et l'erreur suit la loi multinormale

$$\boldsymbol{\varepsilon}^\xi \sim \mathcal{N}(0, \boldsymbol{\Sigma}^\xi),$$

où  $\boldsymbol{\Sigma}^\xi$  est une matrice diagonale dont les éléments correspondent aux variances spatiales et temporelles  $E[\varepsilon_i^\xi \varepsilon_i^\xi] = \sigma_i^2$ . Lorsque les variables aléatoires ne sont pas indépendantes, la loi suivie par l'erreur peut être difficile à déterminer. L'hypothèse est donc souvent faite d'une loi gaussienne<sup>50</sup>, ce qui permet de simplifier grandement le problème comme nous l'avons vu au paragraphe 2.1.3. La matrice des covariances de l'erreur est alors donnée par

$$\boldsymbol{\Sigma}^\xi = E \left[ \boldsymbol{\varepsilon}^\xi (\boldsymbol{\varepsilon}^\xi)^T \right] = \begin{pmatrix} E[\varepsilon_1^\xi \varepsilon_1^\xi] & \dots & E[\varepsilon_1^\xi \varepsilon_P^\xi] \\ \vdots & \ddots & \vdots \\ E[\varepsilon_P^\xi \varepsilon_1^\xi] & \dots & E[\varepsilon_P^\xi \varepsilon_P^\xi] \end{pmatrix}. \quad (2.24)$$

Lorsque plusieurs variables sont impliquées et qu'il existe des relations entre elles, il faut alors également définir les covariances liant chacune de leurs erreurs respectives. On parle alors de **problème**

<sup>50</sup>La prise en compte d'une structure non-gaussienne est une des activités de recherche de l'assimilation de données.

**multivarié**, et non plus de **problème univarié**. Ainsi, si on suppose que l'on a 3 variables  $\xi$ ,  $\zeta$  et  $\chi$  de vecteurs erreurs respectifs  $\varepsilon^\xi$ ,  $\varepsilon^\zeta$  et  $\varepsilon^\chi$ , la matrice représentant les covariances de l'erreur sur l'état complet est donnée par

$$\Sigma = E[\varepsilon\varepsilon^T] = \begin{pmatrix} \Sigma^\xi & E[\varepsilon^\xi(\varepsilon^\zeta)^T] & E[\varepsilon^\xi(\varepsilon^\chi)^T] \\ E[\varepsilon^\zeta(\varepsilon^\xi)^T] & \Sigma^\zeta & E[\varepsilon^\zeta(\varepsilon^\chi)^T] \\ E[\varepsilon^\chi(\varepsilon^\xi)^T] & E[\varepsilon^\chi(\varepsilon^\zeta)^T] & \Sigma^\chi \end{pmatrix}.$$

### 2.3.2 Les erreurs d'observations

Les erreurs d'observations correspondent aux erreurs de mesure, d'interpolation et de représentativité (voir paragraphes 1.2.2 et 1.2.3). Leurs variances sont directement liées aux caractéristiques des instruments de mesure. Cependant, les erreurs de représentativité peuvent ne pas être négligeables et leurs variances devraient également être prises en compte, bien qu'en pratique cela soit difficile à réaliser.

Concernant leurs interactions, on fait généralement l'hypothèse que ces erreurs sont décorréelées entre elles et dans le temps. Bien que cela se justifie dans le cadre d'instruments de mesure distincts, cette hypothèse n'est plus satisfaisante lorsque les observations sont issues d'une même plateforme, comme pour certaines mesures satellitaires. On notera également que les pré-traitements des données tels que la conversion des mesures en variables du modèle, peuvent générer des corrélations artificielles. Enfin, les erreurs de représentativité sont corrélées par nature, en particulier lorsque les observations sont denses par rapport à la résolution du modèle.

En pratique, on essaie de minimiser ces erreurs en évitant par exemple tout traitement qui n'est pas absolument nécessaire, en réduisant par sélection le nombre d'observations dans les régions denses ("*thinning*"), ou encore en combinant plusieurs observations très proches en une seule plus précise ("*superobbing*"). On considère alors la matrice  $\mathbf{R}$ , de dimension  $p \times p$  comme diagonale

$$\mathbf{R} = E[\varepsilon^o(\varepsilon^o)^T] = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_p^2 \end{pmatrix},$$

et il est donc facile de l'inverser.

### 2.3.3 Les erreurs d'ébauche

L'erreur qui affecte l'ébauche est l'erreur de prévisibilité (voir paragraphe 1.2.1) et est due aux incertitudes sur la condition initiale. La connaissance de sa structure est primordiale compte tenu de son rôle pour l'assimilation de données.

Bannister (2008a) passe en revue l'importance de la matrice  $\mathbf{B}$  représentant les covariances de l'erreur d'ébauche, en se référant à l'expression donnée par l'équation (2.21) que l'on rappelle ici

$$\delta\mathbf{x}^a = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d}.$$

Un des premiers rôle de  $\mathbf{B}$  est de propager ou de lisser sur la grille du modèle l'information apportée par les observations. Elle permet également d'étendre l'information, quelle qu'elle soit, aux différentes variables et de leur imposer ainsi un équilibre, conformément aux relations physiques qui les lient. De plus, la matrice  $\mathbf{B}$  permet aux observations d'interférer ensemble de manière constructive et d'améliorer l'analyse de manière plus forte que si les observations étaient assimilées individuellement. Enfin, sa

position finale dans l'expression de l'incrément, oblige ce dernier à exister dans un sous-espace généré par  $\mathbf{B}$ , *i.e.* suivant les directions imposées par les vecteurs propres de  $\mathbf{B}$ .

Compte tenu de son importance, la matrice  $\mathbf{B}$  doit être définie avec le plus de précision possible. Dans la pratique cependant, cette tâche s'avère difficile. En effet, l'erreur d'ébauche est définie par rapport à un état vrai qu'il n'est jamais possible de connaître avec exactitude. Ses covariances ne peuvent alors qu'être approximées et en aucun cas considérées comme exactes. De plus il existe non seulement des covariances spatiales, mais également des covariances entre les différentes variables. Une des grandes difficultés de cette estimation réside alors dans la taille du problème. Si on prend pour exemple une grille grossière de taille  $200 \times 150$  sur 30 niveaux de profondeurs pour un vecteur de contrôle constitué de quatre champs 3D (température, salinité, vitesses zonale et méridienne des courants) et un champ 2D (hauteur de l'eau), on obtient un vecteur de contrôle de plus de 3,6 millions d'éléments et donc une matrice d'environ  $10^{13}$  éléments, ce nombre augmentant bien sûr avec la résolution. Ces tailles rendent impossible la manipulation et le stockage direct de telles matrices. Enfin, une dernière difficulté vient du fait que le manque d'informations peut conduire au calcul d'une matrice  $\mathbf{B}$  de rang réduit. Son utilisation réduit alors le sous-espace qui génère l'incrément et crée des corrélations parasites de longue portée.

Dans l'expression de la fonction coût, la matrice  $\mathbf{B}$  n'est utilisée que pour calculer la contribution de l'ébauche  $J_b$ , et aucun aspect temporel n'y est impliqué contrairement à la matrice  $\mathbf{R}$  vue précédemment. Néanmoins, un aspect essentiel de l'erreur d'ébauche est que sa structure se modifie suivant l'évolution d'évènements tels que les observations assimilées, la dynamique du modèle, son erreur, ... (dépendance à l'écoulement, "*flow-dependence*"). En d'autres termes,  $\mathbf{B}$  devrait être réestimée à chaque nouveau cycle d'assimilation pour tenir compte de l'évolution de l'état de l'océan. Mais à défaut d'une méthode d'estimation efficace, cela s'avère rapidement très coûteux.

L'estimation et/ou la modélisation des structures de l'erreur d'ébauche est un sujet complexe mais essentiel à l'assimilation de données variationnelle. Dans le chapitre 3, nous explorerons les différentes méthodes permettant de définir la matrice  $\mathbf{B}$ . Lorsque les fonctions de covariances doivent être spécifiées, on prend souvent l'hypothèse de fonctions gaussiennes. Plusieurs travaux (Julian et Thiébaux, 1975; Purser *et al.*, 2003b) montrent cependant que cette hypothèse est restrictive, et que le spectre de ces fonctions devrait donner plus d'énergie aux nombres d'onde élevés, pour mieux prendre en compte les informations de petite échelle.

## 2.4 Résolution du problème variationnel

La fonction coût quadratique à minimiser pour notre problème est de la forme

$$J(\delta\mathbf{x}) = \underbrace{\frac{1}{2}\|\delta\mathbf{x}\|_{\mathbf{B}^{-1}}^2}_{J_b} + \underbrace{\frac{1}{2}\|\mathbf{G}\delta\mathbf{x} - \mathbf{d}\|_{\mathbf{R}^{-1}}^2}_{J_o},$$

où les normes  $\|\cdot\|_{\mathbf{B}^{-1}}$  et  $\|\cdot\|_{\mathbf{R}^{-1}}$  dérivent de la norme  $L^2$  si nous faisons l'hypothèse de pdf conditionnelles gaussiennes. Pour le 4D-Var, l'opérateur  $\mathbf{G}$  est défini par la formulation (2.18), tandis qu'il correspond simplement à l'opérateur d'observation  $\mathbf{H}$  pour la simplification en 3D-Var. Pour minimiser cette fonction coût, nous suivrons ici Le Dimet et Talagrand (1986) en appliquant un algorithme de descente (gradient conjugué) utilisant les opérateurs adjoints.

### 2.4.1 Les différents espaces

On définit tout d'abord les différents espaces, suivant leurs discrétisations spatiales et temporelles particulières et les variables qu'ils comprennent :

- **Mod** : espace du modèle à pleine résolution ;
- **Ve** : espace des variables d'état du modèle ;
- **Vc** : espace des variables de contrôle du modèle ;
- **Obs** : espace des observations ;
- **Bi** : espace de la boucle interne, ou du modèle à basse résolution ;
- **Vci** : espace des variables de contrôle de la boucle interne.

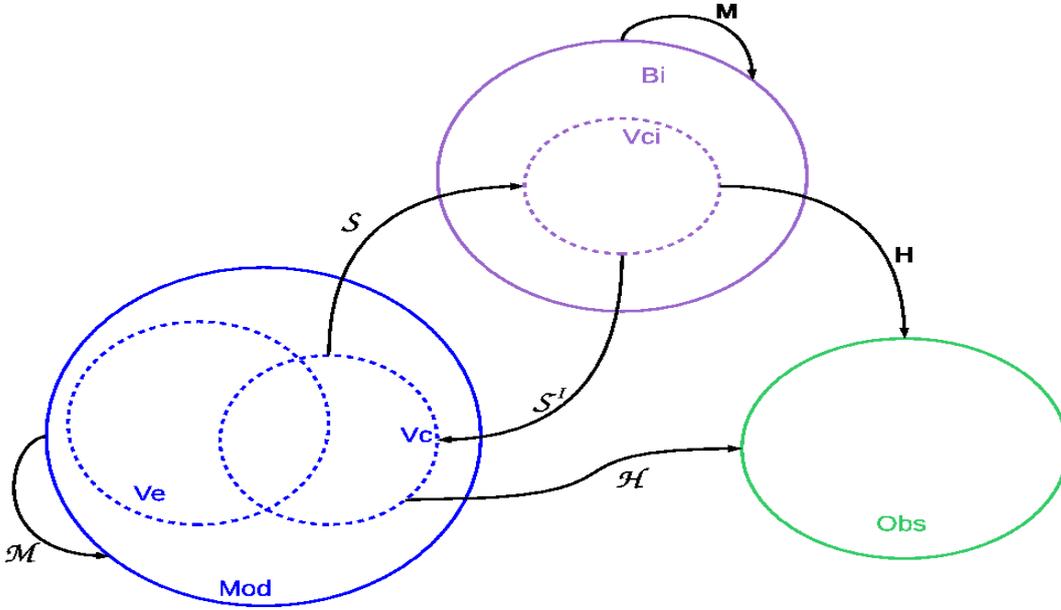


FIG. 2.6 – Les différents espaces et leurs relations.

Tous ces espaces sont en fait des sous-espace de  $\mathbb{R}^d$ , où  $d$  représente la dimension (finie). En particulier, on a  $Vci \subset \mathbb{R}^n$ , où  $n$  est le nombre de points de grille et  $Obs \subset \mathbb{R}^p$ , où  $p$  est le nombre d'observations. Munis du produit scalaire canonique, ces espaces sont des espaces euclidiens. On rappelle que dans ce cas, leurs espaces duaux respectifs peuvent leur être identifiés (voir Annexe A). La figure 2.6 représente ces différents espaces et les opérateurs qui permettent de passer de l'un à l'autre et que l'on détaille ci-dessous :

- Opérateur d'intégration du modèle : il permet de calculer l'état du modèle à l'instant  $t_j$  à partir de l'état à l'instant  $t_i$  et des différents paramètres et conditions subsidiaires du modèle. Par abus de langage, on le dit appliqué aux vecteurs de contrôle dans la formulation du 4D-Var mais il nécessite en réalité l'espace du modèle complet

$$\begin{aligned} \mathcal{M} : \text{Mod} &\rightarrow \text{Mod} \\ \mathbf{x}_i \in \text{Vc} &\mapsto \mathbf{x}_j = \mathcal{M}_{i \rightarrow j}[\mathbf{x}_i] \in \text{Vc} \quad i < j \end{aligned}$$

- Opérateur d'intégration du modèle linéarisé : il correspond à l'opérateur précédent simplifié et linéarisé pour des perturbations de l'état du modèle. Sa forme matricielle est de dimension  $n \times n$

$$\begin{aligned} \mathbf{M} : \text{Bi} &\rightarrow \text{Bi} \\ \delta \mathbf{w}_{i-1} \in \text{Vci} &\mapsto \delta \mathbf{w}_i = \mathbf{M}_i \delta \mathbf{w}_{i-1} \in \text{Vci} \end{aligned}$$

- Opérateur d'observation : il permet d'interpoler le vecteur de contrôle du modèle sur la grille des observations tout en en interprétant les variables afin de permettre la comparaison de la modélisation aux observations

$$\begin{aligned} \mathcal{H} : \text{Vc} &\rightarrow \text{Obs} \\ \mathbf{x} &\mapsto \mathbf{y} = \mathcal{H}[\mathbf{x}] \end{aligned}$$

- Opérateur d'observation linéarisé : il correspond à l'opérateur précédent simplifié et linéarisé pour des perturbations du contrôle du modèle. Sa forme matricielle est de dimension  $p \times n$

$$\begin{aligned} \mathbf{H} : \text{Vci} &\rightarrow \text{Obs} \\ \delta \mathbf{w} &\mapsto \delta \mathbf{y} = \mathbf{H} \delta \mathbf{w} \end{aligned}$$

- Opérateur de simplification : éventuellement non-linéaire, il permet de passer sur une grille modèle de résolution inférieure pour réduire le coût de la minimisation quadratique

$$\begin{aligned} \mathcal{S} : \text{Vc} &\rightarrow \text{Vci} \\ \mathbf{x} &\mapsto \mathbf{w} = \mathcal{S}[\mathbf{x}] \end{aligned}$$

- Inverse généralisé de l'opérateur de simplification : il permet de restituer le résultat de la minimisation sur la grille à pleine résolution

$$\begin{aligned} \mathcal{S}^{-I} : \text{Vci} &\rightarrow \text{Vc} \\ \mathbf{w} &\mapsto \mathbf{x} = \mathcal{S}^{-I}[\mathbf{w}] \end{aligned}$$

Les algorithmes développés dans cette thèse sont appliqués au système d'assimilation variationnelle NEMOVAR (voir chapitre 5). Pour simplifier les écritures, et en adéquation avec ce système, on considèrera pour tout ce qui suit que l'opérateur de simplification est équivalent à l'opérateur identité. On a alors  $\text{Bi} = \text{Mod}$ ,  $\text{Vci} = \text{Vc}$  et donc  $\mathbf{w} = \mathbf{x}$ ,  $\delta \mathbf{w} = \delta \mathbf{x}$ .

## 2.4.2 Minimisation de la fonction coût

Pour minimiser la fonction coût

$$J(\delta \mathbf{x}) = \underbrace{\frac{1}{2} \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x}}_{J_b} + \underbrace{\frac{1}{2} (\mathbf{G} \delta \mathbf{x} - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{G} \delta \mathbf{x} - \mathbf{d})}_{J_o}, \quad (2.25)$$

on cherche la valeur optimale qui en annule le gradient

$$\nabla_{\delta \mathbf{x}} J = \mathbf{B}^{-1} \delta \mathbf{x} + \mathbf{G}^T \mathbf{R}^{-1} (\mathbf{G} \delta \mathbf{x} - \mathbf{d}), \quad (2.26)$$

où en d'autres termes, on cherche  $\delta \mathbf{x}^a$  tel que

$$(\mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}) \delta \mathbf{x}^a = \mathbf{G}^T \mathbf{R}^{-1} \mathbf{d},$$

que l'on peut assimiler au problème général :

$$\left\{ \begin{array}{l} \text{Trouver } \delta \mathbf{x}^a \text{ tel que } \mathbf{A} \delta \mathbf{x}^a = \mathbf{b} \text{ avec} \\ \mathbf{A} = \mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \\ \mathbf{b} = \mathbf{G}^T \mathbf{R}^{-1} \mathbf{d} \end{array} \right. . \quad (2.27)$$

Compte tenu de sa grande taille, il n'est pas possible d'évaluer directement ce problème, et il faut donc faire appel aux méthodes itératives. Généralement, on utilise la méthode du gradient conjugué que l'on préconditionne, d'une part pour accélérer la convergence (voir Annexe D), et d'autre part pour assurer la cohérence des espaces (voir paragraphe 2.4.3).

Derber et Rosati (1989) proposent d'utiliser la matrice  $\mathbf{B}$  comme préconditionneur. En choisissant de plus un point de départ  $\delta \mathbf{x}_{(0)}^a = 0$ , où l'indice  $(j)$  désigne la  $j$ -ième itération de la minimisation, l'algorithme qu'ils donnent permet d'éviter l'inversion de  $\mathbf{B}$ . Son coût est principalement induit par la multiplication par  $\mathbf{B}$  nécessaire à chaque itération.

## Changement de variable

Parrish et Derber (1992) et Derber et Bouttier (1999) proposent quant à eux de factoriser la matrice des covariances d'erreur d'ébauche en définissant une nouvelle matrice  $\mathbf{U}$  telle que  $\mathbf{B} = \mathbf{U}\mathbf{U}^T$ . Utiliser cette définition pour évaluer  $\mathbf{B}$  permet en outre de garantir qu'elle sera bien symétrique et définie positive. Ils suggèrent ensuite de redéfinir la fonction coût par rapport à une nouvelle variable  $\mathbf{v}$  telle que<sup>51</sup>

$$\mathbf{v} = \mathbf{U}^{-1}\delta\mathbf{x}. \quad (2.28)$$

L'équation (2.25) devient alors

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}(\mathbf{G}\mathbf{U}\mathbf{v} - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{G}\mathbf{U}\mathbf{v} - \mathbf{d}), \quad (2.29)$$

et le gradient (2.26)

$$\nabla_{\mathbf{v}}J = \mathbf{v} + \mathbf{U}^T\mathbf{G}^T\mathbf{R}^{-1}(\mathbf{G}\mathbf{U}\mathbf{v} - \mathbf{d}). \quad (2.30)$$

En d'autres termes, on cherche  $\mathbf{v}^a$  tel que

$$(\mathbf{I}_n + \mathbf{U}^T\mathbf{G}^T\mathbf{R}^{-1}\mathbf{G}\mathbf{U})\mathbf{v}^a = \mathbf{U}^T\mathbf{G}^T\mathbf{R}^{-1}\mathbf{d},$$

où  $\mathbf{I}_n$  est la matrice identité de dimension  $n \times n$ . Cette expression peut être assimilée au problème général :

$$\left\{ \begin{array}{l} \text{Trouver } \mathbf{v}^a \text{ tel que } \mathbf{A}\mathbf{v}^a = \mathbf{b} \text{ avec} \\ \mathbf{A} = \mathbf{I}_n + \mathbf{U}^T\mathbf{G}^T\mathbf{R}^{-1}\mathbf{G}\mathbf{U} \\ \mathbf{b} = \mathbf{U}^T\mathbf{G}^T\mathbf{R}^{-1}\mathbf{d} \end{array} \right. . \quad (2.31)$$

Dans cette formulation, le spectre de  $\mathbf{A}$  (l'ensemble de ses valeurs propres) a une borne inférieure de valeur 1. En océanographie, la taille du vecteur d'état étant supérieure à celle du vecteur d'observation, la valeur propre minimale est donc donnée par  $\lambda_{\min} \geq 1$ . Le conditionnement de la matrice est alors  $\kappa(\mathbf{A}) \leq \lambda_{\max}$  (voir Annexe D). Ce changement de variable peut être considéré comme un premier niveau de préconditionnement du problème. De plus, si le point de départ de l'algorithme de minimisation est  $\delta\mathbf{x}_{(0)}^a = 0$ , alors  $\mathbf{v}_{(0)}^a = 0$ , où l'indice  $(j)$  désigne la  $j$ -ième itération de la minimisation, aucun appel à l'inverse de  $\mathbf{B}$  ou de son facteur  $\mathbf{U}$  n'est nécessaire. Après convergence de l'algorithme vers une solution  $\mathbf{v}^a$  optimale, l'incrément optimal est retrouvé en inversant la définition (2.28) :

$$\delta\mathbf{x}^a = \mathbf{U}\mathbf{v}^a.$$

L'algorithme du gradient conjugué peut être appliqué à ce problème. A chaque itération, son coût est attaché au produit matrice-vecteur impliquant  $\mathbf{A}$ . Toute technique permettant d'accélérer la convergence et permettant de s'affranchir ne serait-ce que de quelques itérations est donc la bienvenue. En pratique, on utilise souvent des préconditionneurs à mémoire limitée (Tshimanga *et al.*, 2008) ou une ré-orthogonalisation des directions de descente (Fisher, 1998). Mais ces techniques exigent le stockage de vecteurs supplémentaires, ce qui pourrait devenir un sérieux inconvénient pour les systèmes futurs d'assimilation de données (modèles à haute résolution, 4D-Var à contrainte faible).

## Minimisation dans d'autres espaces

Le coût de l'algorithme de minimisation dépend non seulement du conditionnement du problème mais également de sa taille. Très souvent, la minimisation est effectuée dans l'espace du modèle, ou plus exactement dans l'espace des variables de contrôle  $\mathbf{Vc}$  (approche primale). Or, pour l'océanographie en particulier, la taille du vecteur des observations est généralement inférieure à celle du vecteur de contrôle. Il serait donc plus intéressant d'un point de vue coût, d'effectuer la minimisation dans l'espace des observations  $\mathbf{Obs}$  suivant la formulation PSAS décrite au paragraphe 2.2.4 (approche duale).

<sup>51</sup>Suivant la nature de  $\mathbf{U}$  (matrice rectangulaire par exemple), il peut en fait s'agir de son inverse généralisée.

Amodei (1995), puis Courtier (1997), montrent que le conditionnement  $\kappa(\mathbf{A})$  des deux approches est équivalent lorsque le problème primal est préconditionné par  $\mathbf{B}$  et le dual par  $\mathbf{R}^{-1}$ , rendant donc ce dernier plus attractif puisque de taille inférieure. El Akkraoui *et al.* (2008) cherchent à comparer les deux approches et montrent en particulier que lorsque l'algorithme est stoppé avant complète convergence, ce qui est souvent le cas dans les applications de météorologie et d'océanographie pour limiter le coût, l'approximation de la solution donnée par l'approche duale peut être fortement éloignée de la solution attendue pour la fonction coût de l'équation (2.25). En fait, la fonction coût duale n'a pas de signification statistique, contrairement à la fonction coût donnée par l'équation (2.25).

Pour le gradient conjugué préconditionné appliqué au problème général défini par (2.27), les vecteurs  $\delta\mathbf{x}_{(j)}^a$ ,  $\mathbf{r}_{(j)}$ ,  $\mathbf{q}_{(j)}$ ,  $\mathbf{z}_{(j)}$  et  $\mathbf{p}_{(j)}$  (voir Annexe D) appartiennent à  $\mathbb{R}^n$ . En partant du point initial  $\delta\mathbf{x}_{(0)}^a = 0$  et en définissant une matrice  $\mathbf{F} \in \mathbb{R}^{p \times p}$  telle que

$$\mathbf{B}\mathbf{G}^T\mathbf{F} = \mathbf{M}\mathbf{G}^T,$$

où  $\mathbf{M}$  est le préconditionneur, Gratton et Tshimanga (2009) redéfinissent ces vecteurs dans  $\mathbb{R}^p$ . Ils notent de plus, qu'en choisissant la matrice  $\mathbf{B}$  comme préconditionneur, la nouvelle matrice  $\mathbf{F}$  est simplement l'identité :

$$\mathbf{B}\mathbf{G}^T\mathbf{F} = \mathbf{B}\mathbf{G}^T \Rightarrow \mathbf{F} = \mathbf{I}_p.$$

L'algorithme du gradient conjugué préconditionné restreint (RPCG<sup>52</sup>) qu'ils proposent nécessite le calcul à chaque itération (et le stockage) de deux vecteurs supplémentaires  $\mathbf{w}_i$  et  $\mathbf{t}_i$ . Néanmoins, tous les vecteurs utilisés sont de taille  $p$  comme pour le PSAS au lieu de  $n$  bien que la minimisation soit effectuée dans l'espace des vecteurs de contrôle selon le problème (2.27). Enfin, en ayant pris comme point de départ  $\delta\mathbf{x}_{(0)}^a = 0$ , aucune inversion de  $\mathbf{B}$  n'est requise. Cependant, Gratton et Tshimanga (2009) montrent qu'il est possible de relâcher cette condition, et de définir l'algorithme du RPCG pour un point de départ  $\delta\mathbf{x}_{(0)}^a$  quelconque. Dans ce cas, un nouveau vecteur  $\mathbf{s}$  doit être calculé (et stocké) au départ et tous les vecteurs utiles sont de taille  $p + 1$ .

Le RPCG présente donc les avantages du PSAS, puisque la taille des vecteurs de la minimisation est réduite à la taille des vecteurs d'observation (éventuellement +1). Néanmoins, il ne subit pas les inconvénients d'une minimisation dans un espace différent de celui du modèle, et donc le risque de solutions intermédiaires irréalistes lorsque la convergence de la minimisation n'est pas complète.

### 2.4.3 Le calcul du gradient en pratique

Pour minimiser la fonction coût (2.25) par une méthode itérative comme le gradient conjugué, il est nécessaire de calculer son gradient par rapport à  $\delta\mathbf{x}$  à chaque itération. Mais ce calcul tel que donné par l'équation (2.26) demande des produits matrice-vecteur impliquant en particulier  $\mathbf{G}$  et sa transposée. Or  $\mathbf{G}$  est en fait une formulation complexe des opérateurs linéarisés  $\mathbf{H}$  et  $\mathbf{M}$  et n'est pas accessible directement sous sa forme matricielle, pas plus que ses composantes, et par conséquent sa transposée. La question est donc : comment calculer  $\nabla_{\delta\mathbf{x}}J_o$  dans ces conditions et si possible de manière efficace ?

Une première technique consiste à déterminer numériquement le gradient par perturbation de l'état initial. Pour cela, il est nécessaire d'intégrer le modèle non linéaire sur toute la période d'assimilation autant de fois qu'il y a de composantes dans le vecteur de contrôle  $\mathbf{x}$ . Une intégration est d'abord effectuée pour le vecteur non perturbé, puis on intègre le modèle en perturbant une composante vecteur à la fois. Cette méthode ne peut pas être envisagée pour notre application compte tenu de son coût ( $n + 1$  intégrations du modèle non linéaire). Le Dimet et Talagrand (1986) proposent alors de calculer ce gradient grâce à l'utilisation des opérateurs adjoints.

---

<sup>52</sup>Restricted Preconditionned Conjugate Gradient.

Le paragraphe 2.2.2 montre que  $\mathbf{G}$  est obtenue en linéarisant l'opérateur non linéaire  $\mathcal{G}$  correspondant à l'ensemble des équations du modèle, comme le montre son expression donnée par (2.4)<sup>53</sup>. Pour l'évaluer, il faut estimer son effet sur une perturbation d'ordre 1 autour d'un point donné, comme le montrent les équations (2.16) et (2.17). En d'autres termes, il faut calculer la jacobienne de l'opérateur non linéaire, *i.e.* ses dérivées premières évaluées en ce point donné. En reprenant la formulation contrainte forte du 4D-Var (2.13), on a  $\mathbf{y} = \mathcal{G}[\mathbf{x}]$  avec  $\mathbf{y} \in \text{Obs}$  et  $\mathbf{x} \in \text{Vc}$ . A la perturbation de premier ordre  $\delta\mathbf{x}$  correspond la perturbation de premier ordre  $\delta\mathbf{y}$  telle que  $\delta\mathbf{y} = \mathbf{G}\delta\mathbf{x}$ , avec  $\mathbf{G}$  la jacobienne

$$\mathbf{G} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_p}{\partial x_1} & \cdots & \frac{\partial y_p}{\partial x_n} \end{pmatrix}.$$

En pratique, l'opérateur  $\mathcal{G}$  peut être écrit sous la forme d'une composition de  $P$  opérateurs élémentaires

$$\mathbf{y} = \mathcal{G}[\mathbf{x}] = \mathcal{G}^P \circ \cdots \circ \mathcal{G}^2 \circ \mathcal{G}^1[\mathbf{x}],$$

chaque opérateur élémentaire se présentant sous la forme d'un code logiciel de plusieurs lignes d'instructions. Sa jacobienne correspond donc à l'application successive des jacobienes de chaque opérateur élémentaire

$$\delta\mathbf{y} = \mathbf{G}\delta\mathbf{x} = \mathbf{G}^P \cdots \mathbf{G}^2 \mathbf{G}^1 \delta\mathbf{x},$$

où chaque jacobienne élémentaire est construite en évaluant l'effet de la perturbation sur chaque ligne du code non linéaire.

#### Exemple:

Code non linéaire permettant d'évaluer $y$ sachant la valeur de $x$ :	$y = x \times x$
Code linéarisé permettant d'évaluer $\delta y$ sachant les valeurs de $x$ et $\delta x$ :	$\delta y = 2 \times x \times \delta x$

Dans le code linéarisé, la valeur de  $x$  correspond à une valeur actualisée par le code non linéaire. Pour intégrer le modèle linéarisé, il faut donc avoir intégré auparavant le modèle non linéaire. Effectuer le calcul  $\delta\mathbf{y} = \mathbf{G}\delta\mathbf{x}$ , c'est donc appliquer ensuite le code linéarisé à la perturbation  $\delta\mathbf{x}$ . Il nous reste à déterminer à quoi correspond le calcul  $\delta\mathbf{x}^* = \mathbf{G}^T \delta\mathbf{y}^*$ , où  $\delta\mathbf{x}^*$  et  $\delta\mathbf{y}^*$  sont les variables adjointes de  $\delta\mathbf{x}$  et  $\delta\mathbf{y}$  respectivement<sup>54</sup>. On montre ici qu'il entre dans le cadre des opérateurs adjoints dont nous résumons ici la théorie générale exposée par Talagrand et Courtier (1987).

Soient  $E$  et  $F$  deux espaces euclidiens munis des produits scalaires canoniques  $(\cdot, \cdot)_E$  et  $(\cdot, \cdot)_F$  respectivement, et des produits scalaires pondérés  $\langle \cdot, \cdot \rangle_E$  de poids  $\mathbf{E}^{-1}$  et  $\langle \cdot, \cdot \rangle_F$  de poids  $\mathbf{F}^{-1}$  respectivement, où  $\mathbf{E}$  et  $\mathbf{F}$  représentent des matrices de covariances (voir Annexe A). Les relations entre produits scalaires pondérés et canoniques peuvent s'écrire

$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle_E = \mathbf{v}_1^T \mathbf{E}^{-1} \mathbf{v}_2 = (\mathbf{E}^{-1} \mathbf{v}_1, \mathbf{v}_2)_E. \quad (2.32)$$

<sup>53</sup>Dans le cas de la simplification en 3D-Var, il n'y a pas d'intégration du modèle proprement dit puisque  $\mathcal{G}$  est simplement donné par l'opérateur d'observation  $\mathcal{H}$ . Cependant, ce dernier est en général lui-même constitué d'un ensemble d'équations (modèle de transfert radiatif pour les observations satellite par exemple), et le problème peut donc se poser dans les mêmes termes.

<sup>54</sup>On peut remarquer en effet que dans l'équation du gradient (2.26),  $\mathbf{G}^T$  s'applique à une variable issue de l'application de  $\mathbf{R}^{-1}$ , et donc à une variable de l'espace dual des observations. D'autre part le résultat de l'application de  $\mathbf{G}^T$  correspond à un terme de gradient de la fonction coût par rapport à  $\delta\mathbf{x}$ , et donc, par définition du gradient, à une variable de l'espace dual du modèle. Par anticipation, nous appelons ces variables de l'espace duale les variables adjointes.

Soit  $\mathcal{K}$  un opérateur de  $E$  dans  $F$  tel que  $\mathbf{v} = \mathcal{K}[\mathbf{u}]$ . A la perturbation de premier ordre  $\delta\mathbf{u}$  correspond la perturbation de premier ordre  $\delta\mathbf{v}$  telle que  $\delta\mathbf{v} = \mathbf{K}\delta\mathbf{u}$ , avec  $\mathbf{K}$  l'opérateur linéaire tangent représentant la jacobienne. Soit  $f(\mathbf{v})$ , une fonction scalaire sur  $F$ . A la perturbation de premier ordre  $\delta\mathbf{v}$  correspond la perturbation du premier ordre  $\delta f$  telle que

$$\delta f = (\nabla_{\mathbf{v}}f, \delta\mathbf{v})_F = (\nabla_{\mathbf{v}}f, \mathbf{K}\delta\mathbf{u})_F, \quad (2.33)$$

où on a remplacé  $\delta\mathbf{v}$  par son expression en fonction de  $\delta\mathbf{u}$ . En utilisant les produits scalaires pondérés et en appliquant la définition de l'adjoint donnée en annexe A, on obtient

$$\langle \nabla_{\mathbf{v}}f, \mathbf{K}\delta\mathbf{u} \rangle_F = \langle \mathbf{K}^*\nabla_{\mathbf{v}}f, \delta\mathbf{u} \rangle_E \quad (2.34)$$

où  $\mathbf{K}^*$  existe, est unique, et est défini par l'équation (A.2) par

$$\mathbf{K}^* = \mathbf{E}\mathbf{K}^T\mathbf{F}^{-1}. \quad (2.35)$$

En utilisant l'équation (2.32), l'équation (2.34) devient

$$\langle \mathbf{K}^*\nabla_{\mathbf{v}}f, \delta\mathbf{u} \rangle_E = (\mathbf{E}^{-1}\mathbf{K}^*\nabla_{\mathbf{v}}f, \delta\mathbf{u})_E = \delta f,$$

et on a donc par définition du gradient

$$\nabla_{\mathbf{u}}f = \frac{\partial f}{\partial \mathbf{u}} = \mathbf{E}^{-1}\mathbf{K}^*\nabla_{\mathbf{v}}f.$$

Néanmoins, le but de ce calcul est de permettre d'utiliser un algorithme de descente tel que le gradient conjugué décrit en Annexe D. L'équation (D.3) montre en particulier, qu'un nouvel itéré est calculé à partir d'une fraction du résidu, et que ce résidu représente non pas l'opposé du gradient, mais la direction de la plus forte descente (l'opposé de la plus forte ascendance). En effet, puisque le résidu est employé tel quel pour définir ce nouvel itéré, il doit obligatoirement appartenir au même espace que l'itéré. Or, par définition un gradient appartient à l'espace dual de l'élément auquel il s'applique. Gradient et direction de la plus forte ascendance sont souvent employés de manière synonyme bien qu'ils représentent en réalité deux concepts différents. Mais il est possible de passer de l'un à l'autre en établissant une relation de dualité (Tarantola, 2005, chapitre 6.22). Ici, la quantité utile à l'algorithme de minimisation est donc  $\mathbf{E}\nabla_{\mathbf{u}}f$ , où  $\mathbf{E}$  permet de ramener le gradient depuis l'espace dual vers l'espace considéré. On a alors

$$\mathbf{E}\nabla_{\mathbf{u}}f = \mathbf{K}^*\nabla_{\mathbf{v}}f, \quad (2.36)$$

où on se rappelle que la formulation  $\nabla_{\mathbf{v}}f$  est défini selon le produit scalaire canonique.

On en revient maintenant à notre problème initial et au calcul de  $\nabla_{\delta\mathbf{x}}J_o$ . On a  $\delta\mathbf{y} = \mathbf{G}\delta\mathbf{x}$  avec  $\delta\mathbf{y} \in \mathbf{Obs}$  et  $\delta\mathbf{x} \in \mathbf{Vci}$ . Les produits scalaires pondérés pour ces espaces sont de poids  $\mathbf{R}^{-1}$  et  $\mathbf{B}^{-1}$  respectivement. De plus, le terme  $J_o$  de la fonction coût est défini par la norme carrée

$$J_o(\delta\mathbf{y}) = \frac{1}{2}\|\delta\mathbf{y} - \mathbf{d}\|_{\mathbf{R}^{-1}}^2,$$

où  $\mathbf{d}$  est le vecteur des innovations, et correspond bien à une fonction scalaire sur  $\mathbf{Obs}$ . Puisque les espaces  $\mathbf{Obs}$  et  $\mathbf{Vc}$  lorsqu'ils sont munis du produit scalaire canonique sont euclidiens, il existe  $\mathbf{G}^T$  opérateur adjoint du linéaire tangent  $\mathbf{G}$ , tel que le résultat donné par l'équation (2.36) s'applique

$$\mathbf{B}\nabla_{\delta\mathbf{x}}J_o = \mathbf{G}^*\nabla_{\delta\mathbf{y}}J_o,$$

avec  $\nabla_{\delta\mathbf{y}}J_o$  issu du produit scalaire canonique

$$\nabla_{\delta\mathbf{y}}J_o = \delta\mathbf{y} - \mathbf{d}.$$

et d'après l'équation (2.35)

$$\mathbf{G}^* = \mathbf{B}\mathbf{G}^T\mathbf{R}^{-1}. \quad (2.37)$$

On a donc

$$\nabla_{\delta\mathbf{x}}J_o = \mathbf{G}^T\mathbf{R}^{-1}(\delta\mathbf{y} - \mathbf{d}) = \mathbf{G}^T\mathbf{R}^{-1}(\mathbf{G}\delta\mathbf{x} - \mathbf{d}),$$

où on retrouve l'expression donnée par le gradient (2.26), la variable adjointe  $\delta\mathbf{x}^*$  étant bien définie par l'application de l'adjoint canonique  $\mathbf{G}^T$  sur la variable adjointe  $\delta\mathbf{y}^* = \mathbf{R}^{-1}\delta\mathbf{y}$ . La quantité à utiliser dans la minimisation s'exprime par

$$\mathbf{B}\nabla_{\delta\mathbf{x}}J_o = \mathbf{G}^*(\delta\mathbf{y} - \mathbf{d}),$$

avec  $\mathbf{G}^*$  défini par l'équation (2.37).

De la même manière que la jacobienne  $\mathbf{G}$  correspond à l'application de  $P$  jacobiennes élémentaires, l'opérateur adjoint  $\mathbf{G}^*$  correspond à l'application des adjoints élémentaires mais dans l'ordre inverse :

$$\delta\mathbf{x}^* = \mathbf{G}^*\delta\mathbf{y}^* = \mathbf{G}^{1*}\mathbf{G}^{2*}\dots\mathbf{G}^{P*}\delta\mathbf{y}^*.$$

Or chacun de ces adjoints doit être exprimé suivant l'équation (2.35). D'après l'équation (2.18), à l'instant  $t_i$  on a

$$\mathbf{G}_i = \mathbf{H}_i\mathbf{M}_i \dots \mathbf{M}_1.$$

L'opérateur  $\mathbf{M}$  est défini de  $\text{Vci}$  dans  $\text{Vci}$ . On a donc

$$\mathbf{M}_i^* = \mathbf{B}\mathbf{M}^T\mathbf{B}^{-1},$$

où  $\mathbf{M}^T$  est l'adjoint canonique de  $\mathbf{M}$ . L'opérateur  $\mathbf{H}$  est défini de  $\text{Vci}$  dans  $\text{Obs}$ . On a donc

$$\mathbf{H}_i^* = \mathbf{B}\mathbf{H}^T\mathbf{R}^{-1},$$

où  $\mathbf{H}^T$  est l'adjoint canonique de  $\mathbf{H}$ . L'adjoint de  $\mathbf{G}_i$  s'écrit donc

$$\begin{aligned} \mathbf{G}_i^* &= \mathbf{B}\mathbf{M}_1^T\mathbf{B}^{-1}\mathbf{B}\mathbf{M}_2^T\mathbf{B}^{-1}\dots\mathbf{B}\mathbf{M}_i^T\mathbf{B}^{-1}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1} \\ &= \mathbf{B}\mathbf{M}_1^T\mathbf{M}_2^T\dots\mathbf{M}_i^T\mathbf{H}^T\mathbf{R}^{-1}. \end{aligned}$$

On remarque que si le linéaire tangent  $\mathbf{G}$  permet d'intégrer le modèle vers le futur, son adjoint  $\mathbf{G}^*$  permet en fait d'intégrer le modèle dual vers le passé. Si on reprend l'exemple précédent, on a

**Exemple:**

Code linéarisé permettant d'évaluer $\delta y$ sachant les valeurs de $x$ et $\delta x$ :	$\delta y = 2 \times x \times \delta x$
Code adjoint canonique :	$\delta x^* = \delta x^* + 2 \times x \times \delta y^*$
	$\delta y^* = 0$

La contribution du terme d'ébauche au gradient de la fonction coût est donnée par

$$\nabla_{\delta\mathbf{x}}J_b = \mathbf{B}^{-1}\delta\mathbf{x},$$

où l'on voit l'intérêt du préconditionnement par  $\mathbf{B}$  qui permet de simplifier l'expression en

$$\mathbf{B}\nabla_{\delta\mathbf{x}}J_b = \delta\mathbf{x}.$$

En résumé, pour évaluer le gradient de la fonction coût  $J$ , il faut d'abord calculer la trajectoire établie par l'intégration du modèle non linéaire sur la fenêtre d'assimilation ( $t = t_0 \rightarrow T$ ). Cette trajectoire est stockée pour les besoins des opérateurs linéaires tangents et adjoints. En choisissant un

point de départ  $\delta\mathbf{x} = 0$ , l'intégration du modèle linéaire tangent est inutile, puisque son résultat est  $\delta\mathbf{y} = 0$ . On calcule alors la contribution du terme  $\mathbf{b}$  du problème  $\mathbf{A}\delta\mathbf{x}^a = \mathbf{b}$  (voir équation (2.27)). Pour cela, la variable adjointe  $\delta\mathbf{x}^*$  est initialisée à 0, et le modèle adjoint (canonique) est intégré à rebours ( $t = T \rightarrow t_0$ ) à partir du vecteur innovation ramené dans l'espace dual des observations par l'application de  $\mathbf{R}^{-1}$ . A l'issue de cette intégration,  $\delta\mathbf{x}^*$  est ramenée dans l'espace du modèle par l'application de  $\mathbf{B}$ . A chaque itération, la multiplication par  $\mathbf{A}$  pour le problème  $\mathbf{A}\delta\mathbf{x}^a = \mathbf{b}$  consistera à intégrer le modèle linéaire pour obtenir  $\delta\mathbf{y}$  puis, après avoir initialisé la variable adjointe  $\delta\mathbf{x}^*$  à 0, intégrer le modèle adjoint à partir de la comparaison entre ce  $\delta\mathbf{y}$  et le vecteur innovation, multipliée au préalable par  $\mathbf{R}^{-1}$ . Le résultat est ensuite ramené dans l'espace du modèle par application de  $\mathbf{B}$  et la contribution de l'ébauche est ajoutée.

D'après l'expression de  $\mathbf{G}$  donnée par l'équation (2.18), on a

$$\mathbf{G}^T = [\mathbf{G}_0^T \quad \dots \quad \mathbf{G}_i^T \quad \dots \quad \mathbf{G}_N^T]. \quad (2.38)$$

Mais puisque le problème est linéaire, on peut écrire

$$\begin{aligned} \nabla_{\delta\mathbf{x}} J_o &= \mathbf{G}^T \mathbf{R}^{-1} (\delta\mathbf{y} - \mathbf{d}) \\ &= \mathbf{M}_0^T \dots \mathbf{M}_N^T \mathbf{H}_N^T \mathbf{R}_N^{-1} (\delta\mathbf{y}_N - \mathbf{d}_N) \\ &\quad + \mathbf{M}_0^T \dots \mathbf{M}_{N-1}^T \mathbf{H}_{N-1}^T \mathbf{R}_{N-1}^{-1} (\delta\mathbf{y}_{N-1} - \mathbf{d}_{N-1}) \\ &\quad + \dots \\ &\quad + \mathbf{H}_0^T \mathbf{R}_0^{-1} (\delta\mathbf{y}_0 - \mathbf{d}_0). \end{aligned}$$

D'après cette expression, on peut également écrire les variables adjointes à chaque instant :

$$\begin{aligned} \delta\mathbf{x}_N^* &= \mathbf{H}_N^T \mathbf{R}_N^{-1} (\delta\mathbf{y}_N - \mathbf{d}_N), \\ \delta\mathbf{x}_{N-1}^* &= \mathbf{M}_N^T \delta\mathbf{x}_N^* + \mathbf{H}_{N-1}^T \mathbf{R}_{N-1}^{-1} (\delta\mathbf{y}_{N-1} - \mathbf{d}_{N-1}), \\ &\vdots \\ \delta\mathbf{x}_0^* &= \mathbf{M}_1^T \delta\mathbf{x}_1^* + \mathbf{H}_0^T \mathbf{R}_0^{-1} (\delta\mathbf{y}_0 - \mathbf{d}_0), \end{aligned}$$

où  $\delta\mathbf{x}_0^*$  correspond à  $\nabla_{\delta\mathbf{x}} J_o$ . Contrairement à ce que laisse croire l'expression de  $\mathbf{G}^T$  donnée par l'équation (2.38), une seule intégration du modèle adjoint est donc nécessaire pour calculer le gradient de  $J_o$ , chaque contribution des observations étant ajoutée à l'instant adéquat.

L'algorithme du 4D-Var multi-incrémental proposé au paragraphe 2.2.2 est modifié en pratique de la manière donnée en page suivante (on ne tient pas compte ici de l'opérateur de simplification).

Compte tenu de la définition des espaces concernés (euclidiens), l'opérateur adjoint existe avec certitude. Son codage ne pose pas de difficultés particulières même s'il demande une somme de travail importante. Il exige cependant de respecter certaines règles de codage que l'on pourra trouver par exemple en Annexe B de Bouttier et Courtier (1999). Néanmoins, le point important de cette technique est que l'existence de l'opérateur adjoint est conditionnée par l'existence de l'opérateur linéaire tangent. Et c'est dans le codage de ce dernier que réside toute la difficulté, car il nécessite en général de s'appuyer sur un code non linéaire simplifié, et en particulier affranchi de tout processus non linéarisable, comme les processus à seuil par exemple. Or, ces simplifications doivent être réalisées avec discernement pour ne pas détériorer la convergence de la minimisation non-linéaire (boucle externe).

La méthode des adjoints rend abordable le coût de la minimisation d'une fonctionnelle d'assimilation variationnelle. Cependant, ce coût reste très élevé et la mise en place de toute technique permettant d'améliorer l'efficacité de l'algorithme de minimisation (préconditionnement, ré-orthogonalisation des directions de descente, ...) est à rechercher.

## 4D-Var multi-incrémental en pratique

- 
- $\mathbf{x}_0^{(0)} = \mathbf{x}^b$
  - Pour  $k = 1$  à  $k_{\max}$ 

<i>Bcl Externe</i>	1 : Calcul de la trajectoire $\mathbf{x}^{(k-1)} = \mathcal{M} [\mathbf{x}_0^{(k-1)}]$
<i>Bcl Interne</i>	2 : Calcul des innovations $\mathbf{d}^{(k)} = \mathbf{y}^o - \mathcal{H} [\mathbf{x}^{(k-1)}]$
<i>Minimisation</i>	<ul style="list-style-type: none"> <li>i : Initialisation <math>\delta \mathbf{x}_0^{[0]} = 0</math> et <math>j = 1</math></li> <li>ii : Tant que le critère d'arrêt de la minimisation n'est pas satisfait <ul style="list-style-type: none"> <li>a : Calcul d'une direction de descente <ul style="list-style-type: none"> <li>a.1 : Intégration du modèle linéaire tangent</li> <li>a.2 : Intégration du modèle adjoint</li> <li>a.3 : Contribution du terme d'ébauche</li> </ul> </li> <li>b : Descente vers le nouveau point <math>\delta \mathbf{x}_0^{[j]}</math></li> <li>c : <math>j = j + 1</math></li> </ul> </li> <li>iv : Fin tant que, <math>j = j_{\text{final}}</math></li> </ul>
	6 : Mise à jour $\mathbf{x}_0^{(k)} = \mathbf{x}_0^{(k-1)} + \delta \mathbf{x}_0^{[j_{\text{final}}]}$
  - Fin pour
  - Mise à jour de l'analyse  $\mathbf{x}^a = \mathbf{x}_0^{(k_{\max})}$
  - Propagation de l'analyse pour calcul de l'ébauche du cycle suivant  $\mathcal{M} [\mathbf{x}^a]$
-

## Chapitre 3

# Modélisation de la matrice des covariances d'erreur d'ébauche

Compte tenu de son importance, l'estimation de la matrice des covariances d'erreur d'ébauche est une des tâches essentielles de l'assimilation de données variationnelle. Le paragraphe 2.3.3 en résume le rôle et en décrit les principales difficultés pour les domaines de l'océanographie et de la météorologie. Dans ce chapitre, on s'appuie en particulier sur Bannister (2008a,b) pour exposer de manière non exhaustive les différentes techniques couramment utilisées pour estimer ou modéliser  $\mathbf{B}$ .

### 3.1 Le problème fondamental

Dans la formulation incrémentale, le terme lié à l'ébauche pour la fonction coût (2.25) à minimiser est donné par

$$J_b(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x},$$

et son gradient canonique (2.26) par

$$\nabla_{\delta\mathbf{x}}J_b = \mathbf{B}^{-1}\delta\mathbf{x}.$$

Ces équations montrent qu'il peut être nécessaire de déterminer  $\mathbf{B}^{-1}$ , la précision de l'ébauche. L'estimation des covariances de l'erreur d'ébauche  $\mathbf{B}$  est déjà un problème délicat en soi compte tenu des difficultés inhérentes. Inverser en outre une telle matrice et stocker le résultat est un sujet techniquement inabordable lorsque le système est de grande taille. Bien qu'il soit possible de modéliser directement la précision  $\mathbf{B}^{-1}$  appliquée à un vecteur quelconque (le paragraphe 3.1.1 décrit la méthode proposée par Xu (2005) utilisant des opérateurs différentiels), on utilise généralement un changement de variable permettant de s'affranchir de l'inversion comme il est décrit au paragraphe 3.1.2.

Ce chapitre constitue une revue des différentes techniques couramment employées en météorologie et océanographie pour estimer ou modéliser  $\mathbf{B}$ . On ne prétend pas ici être exhaustif et on se limite aux techniques employées en assimilation variationnelle. La figure 3.1 de la page suivante présente les techniques abordées dans ce chapitre. L'objet de cette thèse, la modélisation à l'aide d'une équation de diffusion généralisée, est présentée au paragraphe 3.6.3.

#### 3.1.1 Modélisation de $\mathbf{B}^{-1}$

Dans ce paragraphe, on résume le principe de la méthode proposée par Xu (2005) et permettant de modéliser directement  $\mathbf{B}^{-1}$  à partir d'opérateurs différentiels. Si  $C(z)$  est une fonction homogène et isotrope (voir Annexe B) sur le domaine  $\Omega$ , on peut définir son inverse  $C^{-1}(z)$  telle que (voir le

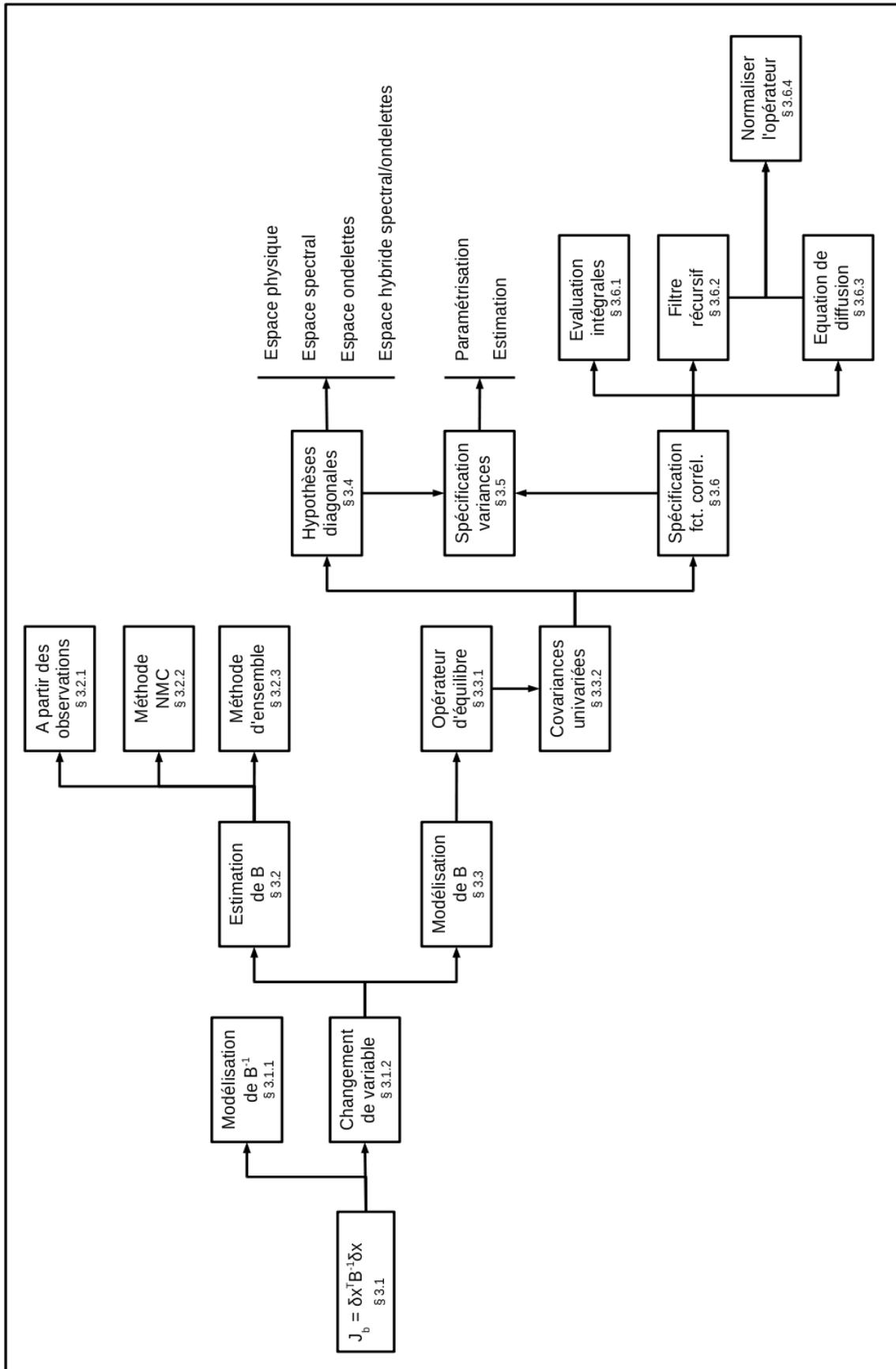


FIG. 3.1 – Les différentes techniques couramment employées pour estimer ou modéliser la matrice des covariances d'erreur d'ébauche.

paragraphe 4.2 avec  $z'' = 0$ )

$$\int_{\Omega} C(z - z')C^{-1}(z') dz' = \delta(z),$$

avec  $\delta(z)$  la distribution de Dirac. Cette expression représentant un produit de convolution, sa transformée de Fourier (voir la définition donnée au paragraphe 4.2) donne le produit

$$\widehat{C}(\hat{z})\widehat{C^{-1}}(\hat{z}) = 1. \quad (3.1)$$

En supposant que  $\widehat{C^{-1}}(\hat{z})$  soit développable en série entière (et notamment qu'elle soit indéfiniment dérivable au voisinage de 0), on peut écrire

$$\widehat{C^{-1}}(\hat{z}) = \sum_{j=0}^J \alpha_j \hat{z}^{2j} = \sum_{j=0}^J (-1)^j \alpha_j (i\hat{z})^{2j},$$

puisque les termes en  $\hat{z}^j$  avec  $j$  impair sont nuls. Connaissant la transformée de Fourier de la fonction de corrélation  $\widehat{C}(\hat{z})$ , il est également possible de l'exprimer sous la forme d'une série de Taylor où les coefficients sont déterminés par ses dérivées. Les  $\alpha_j$  peuvent alors leur être identifiés grâce à la relation donnée par l'équation (3.1). En appliquant maintenant la propriété (Jones, 1982, p. 74) de la transformation de Fourier inverse

$$\mathfrak{F}^{-1}[(i\hat{z})^n] = \frac{1}{2\pi} \int_{-\infty}^{\infty} (i\hat{z})^n e^{i\hat{z}z} d\hat{z} = \delta^{(n)}(z),$$

on obtient

$$C^{-1}(z) = \sum_{j=0}^J (-1)^j \alpha_j \delta^{(2j)}(z). \quad (3.2)$$

D'autre part, le terme  $J_b$  peut s'écrire sous la forme du carré d'une norme  $L^2$  :

$$J_b = \frac{1}{2} \|\delta\mathbf{x}\|_{\mathbf{B}^{-1}}^2.$$

En définissant  $\varphi(z)$  comme la fonction continue associée au vecteur incrément  $\delta\mathbf{x}$ , et en supposant que la variance  $\sigma^2$  est constante (on traite donc bien des covariances univariées et non pas des corrélations uniquement ; le cas de variances non-constantes est également traité par Xu (2005)), on peut écrire en continu

$$J_b = \frac{\sigma^2}{2} \int_{\Omega} \int_{\Omega} \varphi(z)C^{-1}(z - z')\varphi(z') dz' dz.$$

En appliquant le résultat de l'équation (3.2), on a

$$\begin{aligned} J_b &= \frac{\sigma^2}{2} \int_{\Omega} \int_{\Omega} \varphi(z) \sum_{j=0}^J (-1)^j \alpha_j \delta^{(2j)}(z - z') \varphi(z') dz' dz \\ &= \frac{\sigma^2}{2} \int_{\Omega} \varphi(z) \sum_{j=0}^J (-1)^j \alpha_j \varphi^{(2j)}(z) dz, \end{aligned}$$

où on a utilisé la propriété (Jones, 1982, p. 178)

$$\int_{-\infty}^{\infty} \delta^{(n)}(z - z') \varphi(z') dz' = \varphi^{(n)}(z).$$

En développant le terme de somme dans l'expression de  $J_b$ , on obtient

$$J_b = \frac{\sigma^2}{2} \int_{\Omega} \alpha_0 \varphi^2(z) dz - \underbrace{\frac{\sigma^2}{2} \int_{\Omega} \alpha_1 \varphi(z) \frac{\partial^2 \varphi(z)}{\partial z^2} dz}_{I_1} + \underbrace{\frac{\sigma^2}{2} \int_{\Omega} \alpha_2 \varphi(z) \frac{\partial^4 \varphi(z)}{\partial z^4} dz}_{I_2} - \dots,$$

où chaque intégrale peut faire l'objet d'une ou plusieurs intégrations par partie successives. On prend par exemple le cas de  $I_2$  :

$$\begin{aligned}
I_2 &= \int_{\Omega} \alpha_2 \varphi(z) \frac{\partial^4 \varphi(z)}{\partial z^4} dz \\
&= \alpha_2 \left[ \varphi(z) \frac{\partial^3 \varphi(z)}{\partial z^3} \right]^{\Gamma} - \int_{\Omega} \alpha_2 \frac{\partial \varphi(z)}{\partial z} \frac{\partial^3 \varphi(z)}{\partial z^3} dz \\
&= -\alpha_2 \left[ \frac{\partial \varphi(z)}{\partial z} \frac{\partial^2 \varphi(z)}{\partial z^2} \right]^{\Gamma} + \int_{\Omega} \alpha_2 \frac{\partial^2 \varphi(z)}{\partial z^2} \frac{\partial^2 \varphi(z)}{\partial z^2} dz \\
&= \int_{\Omega} \alpha_2 \left( \frac{\partial^2 \varphi(z)}{\partial z^2} \right)^2 dz,
\end{aligned}$$

en considérant que les conditions sur la frontière  $\Gamma$  du domaine sont telles que la fonction incrément ou ses dérivées sont nulles sur cette frontière. On peut donc bien réécrire le terme  $J_b$  comme le carré de la norme  $L^2$  d'un vecteur d'opérateurs différentiels  $\mathcal{D}$  appliqué à la fonction incrément :

$$J_b = \frac{\sigma^2}{2} \int_{\Omega} (\mathcal{D}[\varphi(z)])^2 dz,$$

avec<sup>55</sup>

$$\mathcal{D} = \left( \sqrt{\alpha_0} \quad \sqrt{\alpha_1} \frac{\partial}{\partial z} \quad \cdots \quad \sqrt{\alpha_J} \frac{\partial^J}{\partial z^J} \right)^{\text{T}}.$$

Cette méthode peut être utilisée pour modéliser en particulier l'inverse d'une fonction autorégressive (l'opérateur est alors une somme finie ; voir chapitre 4) ou d'une gaussienne (la somme est infinie et doit être tronquée en pratique). Xu (2005) en donne des exemples en une et plusieurs dimensions. En exprimant  $J_b$  comme l'intégrale de la somme des différentes dérivées de  $\varphi(z)$ , on note l'effet de lissage qui en résulte en pénalisant les "petites" structures.

Bien que présentant un intérêt certain, cette méthode peut s'avérer difficile à mettre en œuvre en pratique. En effet, la théorie exposée ci-dessus s'appuie sur des hypothèses d'homogénéité et d'isotropie, bien qu'en exploitant des changements de variables ou de coordonnées appropriés, il soit possible d'étendre cette technique à des covariances inhomogènes, anisotropes et même multivariées. De plus, nous verrons dans les paragraphes suivants qu'il existe des techniques permettant d'estimer certaines caractéristiques de  $\mathbf{B}$ . Mais ces paramètres peuvent être difficiles à exploiter dans le cadre de l'inverse  $\mathbf{B}^{-1}$ . Enfin, la fonction coût est en général minimisée par une méthode de descente (voir paragraphe 2.4.2). Or un préconditionnement est nécessaire, qu'il soit direct ou par l'intermédiaire d'un changement de variable, d'une part pour mieux conditionner le problème, et d'autre part pour assurer une cohérence des espaces. Ce préconditionnement est souvent choisi de telle manière à ce que l'inversion de  $\mathbf{B}$  ne soit plus nécessaire.

### 3.1.2 Changement de variables

Le principe de ce changement est de définir une relation entre l'incrément  $\delta \mathbf{x}$  et une nouvelle variable  $\chi$  telle que

$$\delta \mathbf{x} = \mathbf{T} \chi,$$

où  $\mathbf{T}$  est une matrice de transformation valide, et telle qu'il ne soit plus nécessaire de définir l'inverse de  $\mathbf{B}$  lors de la minimisation de la fonction coût<sup>56</sup>. Il existe plusieurs changements de variables différents

<sup>55</sup>D'après Xu (2005),  $\mathcal{D}$  est affecté d'un facteur  $(2\pi)^{1/4}$ . Nous pensons que ce facteur est une erreur due à la transformée de Fourier de type "symétrique" (directe et inverse affectées du poids  $1/\sqrt{2\pi}$ ). En effet, l'utilisation de cette transformée nécessite de tenir compte également d'un poids  $1/\sqrt{2\pi}$  dans la définition de la convolution, poids qui semble avoir été oublié lors d'une manipulation d'équation.

<sup>56</sup>La minimisation étant effectuée par un algorithme de descente, on cherche plutôt ici à accélérer sa convergence. S'affranchir de l'utilisation de l'inverse de  $\mathbf{B}$  en est une conséquence utile.

suivant l'espace dans lequel on les définit (il s'agit souvent des espaces duals, voir à ce sujet la note 48). Outre celui des observations, il peut également s'agir de celui du modèle soit dans l'espace physique, soit dans l'espace spectral (suite à une transformation de Fourier). L'espace spectral est souvent employé en météorologie, contrairement à l'océanographie où la nécessité de gérer des frontières complexes rend malaisée son utilisation. On ne s'attardera donc pas sur cette approche dans ce paragraphe.

### Définition de $\chi$ dans l'espace (dual) des observations : $\mathbf{T} = \mathbf{B}\mathbf{G}^T$

La fonction coût incrémentale (2.25) devient

$$\begin{aligned} J(\chi) &= \frac{1}{2}\chi^T \mathbf{G}\mathbf{B}^T \mathbf{B}^{-1} \mathbf{B}\mathbf{G}^T \chi + \frac{1}{2}(\mathbf{G}\mathbf{B}\mathbf{G}^T \chi - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{G}\mathbf{B}\mathbf{G}^T \chi - \mathbf{d}) \\ &= \frac{1}{2}(\mathbf{G}^T \chi)^T \mathbf{B}\mathbf{G}^T \chi + \frac{1}{2}(\mathbf{G}\mathbf{B}\mathbf{G}^T \chi - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{G}\mathbf{B}\mathbf{G}^T \chi - \mathbf{d}), \end{aligned}$$

où on applique la définition des matrices inverses  $\mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$  et la symétrie d'une matrice de covariance  $\mathbf{B}^T = \mathbf{B}$ . Le gradient (2.26) est donné par

$$\nabla_{\chi} J = \mathbf{G}\mathbf{B}\mathbf{G}^T \chi + \mathbf{G}\mathbf{B}\mathbf{G}^T \mathbf{R}^{-1} (\mathbf{G}\mathbf{B}\mathbf{G}^T \chi - \mathbf{d}).$$

Le problème  $\nabla_{\chi} J = 0$  à résoudre s'écrit alors, trouver  $\chi^a$  tel que

$$\begin{aligned} \mathbf{G}\mathbf{B}\mathbf{G}^T (\mathbf{I} + \mathbf{R}^{-1} \mathbf{G}\mathbf{B}\mathbf{G}^T) \chi^a &= \mathbf{G}\mathbf{B}\mathbf{G}^T \mathbf{R}^{-1} \mathbf{d} \\ \mathbf{R}^{-1} (\mathbf{R} + \mathbf{G}\mathbf{B}\mathbf{G}^T) \chi^a &= \mathbf{R}^{-1} \mathbf{d} \\ (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}) \chi^a &= \mathbf{d}, \end{aligned}$$

en supposant que  $\mathbf{G}\mathbf{B}\mathbf{G}^T$  soit de rang égal au nombre total d'observations. On retrouve alors le problème du 4D-PSAS décrit au paragraphe 2.2.4.

### Définition de $\chi$ dans l'espace (dual) du modèle physique : $\mathbf{T} = \mathbf{B}$

La fonction coût incrémentale (2.25) devient

$$\begin{aligned} J(\chi) &= \frac{1}{2}\chi^T \mathbf{B}^T \mathbf{B}^{-1} \mathbf{B}\chi + \frac{1}{2}(\mathbf{B}\chi - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{B}\chi - \mathbf{d}) \\ &= \frac{1}{2}\chi^T \mathbf{B}\chi + \frac{1}{2}(\mathbf{B}\chi - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{B}\chi - \mathbf{d}), \end{aligned}$$

où on applique la définition des matrices inverses  $\mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$  et la symétrie d'une matrice de covariance  $\mathbf{B}^T = \mathbf{B}$ . Le gradient (2.26) est donné par

$$\nabla_{\chi} J = \mathbf{B}\chi + \mathbf{B}\mathbf{G}^T \mathbf{R}^{-1} (\mathbf{B}\chi - \mathbf{d}).$$

Le problème  $\nabla_{\chi} J = 0$  à résoudre s'écrit alors, trouver  $\chi^a$  tel que

$$\begin{aligned} (\mathbf{B} + \mathbf{B}\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}\mathbf{B}) \chi^a &= \mathbf{B}\mathbf{G}^T \mathbf{R}^{-1} \mathbf{d} \\ (\mathbf{I} + \mathbf{B}\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}) \mathbf{B}\chi^a &= \mathbf{B}\mathbf{G}^T \mathbf{R}^{-1} \mathbf{d}. \end{aligned}$$

Cette approche est similaire à celle de Derber et Rosati (1989) qui proposent d'utiliser  $\mathbf{B}$  comme préconditionneur du problème classique (voir paragraphe 2.4.2).

**Définition de  $\chi$  dans l'espace (dual) du modèle physique :  $\mathbf{T} = \mathbf{U}$  tel que  $\mathbf{B} = \mathbf{U}\mathbf{U}^T$**

Proposé par Parrish et Derber (1992), ce changement de variable est le plus communément utilisé actuellement. Outre ce rôle, il fait également office de préconditionnement de premier niveau pour un algorithme de descente, comme exposé au paragraphe 2.4.2. On rappelle ici la transformation de la fonction coût en

$$\begin{aligned} J(\chi) &= \frac{1}{2}\chi^T \mathbf{U}^T \mathbf{B}^{-1} \mathbf{U} \chi + \frac{1}{2}(\mathbf{G}\mathbf{U}\chi - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{G}\mathbf{U}\chi - \mathbf{d}) \\ &= \frac{1}{2}\chi^T \chi + \frac{1}{2}(\mathbf{G}\mathbf{U}\chi - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{G}\mathbf{U}\chi - \mathbf{d}), \end{aligned}$$

où on utilise le fait que

$$\mathbf{U}^T \mathbf{B}^{-1} \mathbf{U} = \mathbf{U}^T (\mathbf{U}\mathbf{U}^T)^{-1} \mathbf{U} = \mathbf{U}^T (\mathbf{U}^T)^{-1} \mathbf{U}^{-1} \mathbf{U} = \mathbf{I}.$$

Le gradient (2.26) est donné par

$$\nabla_{\chi} J = \chi + \mathbf{U}^T \mathbf{G}^T \mathbf{R}^{-1} (\mathbf{G}\mathbf{U}\chi - \mathbf{d}).$$

Le problème  $\nabla_{\chi} J = 0$  à résoudre s'écrit alors, trouver  $\chi^a$  tel que

$$(\mathbf{I} + \mathbf{U}^T \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G}\mathbf{U}) \chi^a = \mathbf{U}^T \mathbf{G}^T \mathbf{R}^{-1} \mathbf{d}.$$

### Le nouveau problème

Le changement de variable permet de s'affranchir de l'inversion de  $\mathbf{B}$ . Néanmoins, déterminer et manipuler cette matrice reste un problème difficile bien qu'essentiel. Plusieurs techniques ont été développées pour estimer ces covariances d'erreur d'ébauche. Le principe des principales méthodes est décrit au paragraphe 3.2. Cependant, les matrices résultant de ces approches présentent souvent une déficience de rang. Or, les formulations précédentes montrent que l'exploitation qui est faite de la matrice  $\mathbf{B}$  en assimilation variationnelle, nécessite en fait de déterminer l'effet de son application sur un champ donné. En d'autres termes, le besoin n'est pas tant de construire une estimation de la matrice elle-même que de calculer le produit matrice-vecteur  $\mathbf{B}\zeta$  où  $\zeta$  est un champ scalaire quelconque. Dans ce cas, on peut alors choisir de construire  $\mathbf{B}$  par un modèle dont l'entrée est le champ  $\zeta$ , et la sortie le produit  $\mathbf{B}\zeta$ . Cette approche est décrite au paragraphe 3.3.

## 3.2 Estimation des statistiques de $\mathbf{B}$

Estimer  $\mathbf{B}$  consiste à évaluer ses principales caractéristiques, ce qui reste difficile puisque l'état vrai permettant de les définir n'est pas accessible. Néanmoins, il est possible d'utiliser soit des informations disponibles telles que les innovations, soit des quantités ayant des statistiques similaires, pour mesurer ces caractéristiques. L'hypothèse d'ergodicité<sup>57</sup> est alors souvent utilisée pour calculer les espérances impliquées dans ces statistiques, à partir d'échantillons temporels plutôt que spatiaux. On décrit ci-après les principales méthodes de cette estimation. On notera qu'une autre possibilité, que l'on n'abordera pas ici, consiste à utiliser une décomposition à rang réduit basée par exemple sur des EOF<sup>58</sup> (voir par exemple Robert *et al.*, 2006a,b) ou des ensembles (Lorenç, 2003).

<sup>57</sup>Un processus stochastique est dit ergodique si les caractéristiques statistiques déduites d'un ensemble de ses réalisations à un instant donné sont équivalentes à celles déduites de l'ensemble des valeurs successives dans le temps d'une de ces réalisations.

<sup>58</sup>Empirical Orthogonal Functions.

### 3.2.1 Méthodes basées sur les observations

Les innovations, *i.e.* l'écart entre les observations et l'ébauche ramenée dans l'espace des observations, sont calculées pour chaque cycle d'assimilation. Ces informations sont donc systématiquement disponibles et peuvent être utilisées pour estimer les caractéristiques de  $\mathbf{B}$  (Hollingsworth et Lönnberg, 1986; Lönnberg et Hollingsworth, 1986; Järvinen, 2001). En effet, avec  $\mathcal{G}$  défini suivant l'équation (2.4) le vecteur observation s'écrit

$$\mathbf{y}^o = \mathcal{G} [\mathbf{x}^t] + \boldsymbol{\varepsilon}^o = \mathcal{G} [\mathbf{x}^b - \boldsymbol{\varepsilon}^b] + \boldsymbol{\varepsilon}^o,$$

que l'on peut linéariser sous la forme

$$\mathbf{y}^o \approx \mathcal{G} [\mathbf{x}^b] - \mathbf{G}\boldsymbol{\varepsilon}^b + \boldsymbol{\varepsilon}^o,$$

où  $\mathbf{G}$  est défini suivant l'équation (2.18). Le vecteur innovation s'écrit alors

$$\mathbf{d} = \mathbf{y}^o - \mathcal{G} [\mathbf{x}^b] \approx \boldsymbol{\varepsilon}^o - \mathbf{G}\boldsymbol{\varepsilon}^b.$$

Le calcul des covariances de l'innovation peut apporter un grand nombre d'informations sur la structure de l'erreur d'ébauche. En effet, on a

$$\begin{aligned} E [\mathbf{d}\mathbf{d}^T] &= E \left[ \left( \boldsymbol{\varepsilon}^o - \mathbf{G}\boldsymbol{\varepsilon}^b \right) \left( \boldsymbol{\varepsilon}^o - \mathbf{G}\boldsymbol{\varepsilon}^b \right)^T \right] \\ &= E \left[ \boldsymbol{\varepsilon}^o \left( \boldsymbol{\varepsilon}^o \right)^T \right] + \mathbf{G} E \left[ \boldsymbol{\varepsilon}^b \left( \boldsymbol{\varepsilon}^b \right)^T \right] \mathbf{G}^T \\ &= \mathbf{R} + \mathbf{G}\mathbf{B}\mathbf{G}^T, \end{aligned}$$

où on a supposé que les erreurs d'observations et d'ébauche étaient décorréliées, soit  $E \left[ \boldsymbol{\varepsilon}^o \left( \boldsymbol{\varepsilon}^b \right)^T \right] = 0$ . Pour deux vecteurs d'observations aux points  $i$  et  $j$ ,  $\mathbf{y}_i^o$  et  $\mathbf{y}_j^o$ , on a

$$E [\mathbf{d}_i \mathbf{d}_j^T] = \mathbf{R}_{ij} + \mathbf{G}_i \mathbf{B} \mathbf{G}_j^T.$$

Si  $i = j$ , la variance de l'innovation est alors la somme des variances de l'erreur d'observations et de l'erreur d'ébauche ramenée dans l'espace des observations :

$$\text{Var} (\mathbf{d}_i) = \text{Var} (\boldsymbol{\varepsilon}_i^o) + \text{Var} \left( \mathbf{G}_i \boldsymbol{\varepsilon}^b \right). \quad (3.3)$$

Si  $i \neq j$ , et en supposant que les observations sont décorréliées entre elles ( $\mathbf{R}_{ij} = 0$ ), les covariances de l'innovation représentent en fait les covariances de l'erreur d'ébauche ramenée dans l'espace des observations :

$$\text{Cov} (\mathbf{d}_i, \mathbf{d}_j) = \text{Cov} \left( \mathbf{G}_i \boldsymbol{\varepsilon}^b, \mathbf{G}_j \boldsymbol{\varepsilon}^b \right).$$

Ces covariances peuvent être représentées en fonction de la distance séparant les observations deux à deux (voir figure 3.2). Lorsque cette distance est suffisamment petite, la covariance tend en fait vers la variance de l'erreur d'ébauche (puisque  $\mathbf{R}_{ij} = 0$ ). En comparant cette limite à l'estimation obtenue par l'équation (3.3), on obtient une information sur la variance de l'erreur d'observation. L'hypothèse d'ergodicité permet de calculer ces covariances à partir de plusieurs cycles d'assimilation si on suppose qu'à chaque point d'observation, des mesures sont effectuées à plusieurs reprises dans le temps.

Cette méthode permet d'obtenir une estimation directe des covariances de l'erreur d'ébauche à condition de disposer d'un réseau d'observations indépendantes suffisant et relativement uniforme (hypothèses de décorrélation des observations entre elles). Cette technique n'est donc adaptée qu'à des régions densément observées par des instruments indépendants (*in situ*). Pour relâcher cette hypothèse, Desroziers *et al.* (2005) proposent d'établir le même type de diagnostics mais à partir de différentes

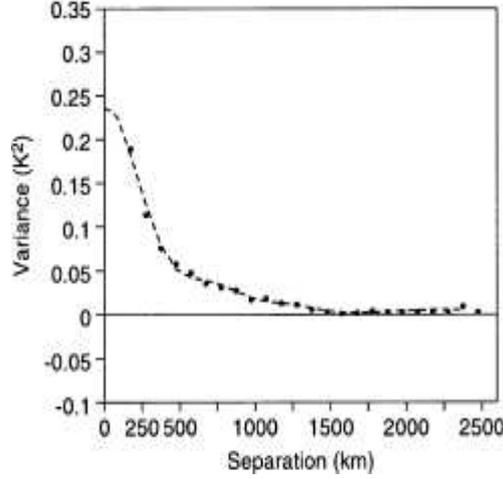


FIG. 3.2 – Covariances issues des innovations pour des températures AIREP (mesures atmosphériques effectuées par des avions). Figure 1.(b) de Järvinen (2001).

combinaisons. Dans ce qui suit, on appellera  $\mathbf{d}_b^o$  le vecteur innovation  $\mathbf{d}$  précédent représentant l'écart entre observations et ébauche ramenée dans l'espace des observations.

La formulation incrémentale présentée au paragraphe 2.2.2 permet d'obtenir une analyse en ajoutant un incrément à l'ébauche. Cet incrément est obtenu en multipliant le vecteur innovation par une matrice de gain comme décrit par l'équation (2.21), où on a étendu l'opérateur  $\mathbf{H}$  à l'opérateur  $\mathbf{G}$ . On peut donc écrire

$$\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x}^a = \mathbf{x}^b + \mathbf{B}\mathbf{G}^T (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R})^{-1} \mathbf{d}_b^o.$$

Si on définit  $\mathbf{d}_b^a$  l'écart, toujours dans l'espace des observations, entre l'analyse et l'ébauche, on a

$$\begin{aligned} \mathbf{d}_b^a &= \mathcal{G}[\mathbf{x}^a] - \mathcal{G}[\mathbf{x}^b] = \mathcal{G}[\mathbf{x}^b + \delta\mathbf{x}^a] - \mathcal{G}[\mathbf{x}^b] \\ &\approx \mathbf{G}\mathbf{B}\mathbf{G}^T (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R})^{-1} \mathbf{d}_b^o. \end{aligned}$$

On peut alors calculer

$$\begin{aligned} E[\mathbf{d}_b^a (\mathbf{d}_b^a)^T] &= \mathbf{G}\mathbf{B}\mathbf{G}^T (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R})^{-1} E[\mathbf{d}_b^o (\mathbf{d}_b^o)^T] \\ &= \mathbf{G}\mathbf{B}\mathbf{G}^T (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R})^{-1} (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}) \\ &= \mathbf{G}\mathbf{B}\mathbf{G}^T. \end{aligned}$$

Cette formulation permet donc d'accéder à une estimation de la matrice  $\mathbf{B}$  dans l'espace des observations. De la même manière, on peut estimer la matrice  $\mathbf{R}$  en définissant  $\mathbf{d}_a^o$ , l'écart entre observations et analyse ramenée dans l'espace des observations. En effet, on a

$$\begin{aligned} \mathbf{d}_a^o &= \mathbf{y}^o - \mathcal{G}[\mathbf{x}^a] = \mathbf{d}_b^o + \mathcal{G}[\mathbf{x}^b] - \mathcal{G}[\mathbf{x}^b + \delta\mathbf{x}^a] \\ &\approx (\mathbf{I} - \mathbf{G}\mathbf{B}\mathbf{G}^T (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R})^{-1}) \mathbf{d}_b^o \\ &= (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R} - \mathbf{G}\mathbf{B}\mathbf{G}^T) (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R})^{-1} \mathbf{d}_b^o \\ &= \mathbf{R} (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R})^{-1} \mathbf{d}_b^o, \end{aligned}$$

et on peut calculer

$$\begin{aligned} E[\mathbf{d}_a^o (\mathbf{d}_a^o)^T] &= \mathbf{R} (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R})^{-1} E[\mathbf{d}_b^o (\mathbf{d}_b^o)^T] \\ &= \mathbf{R} (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R})^{-1} (\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}) \\ &= \mathbf{R}. \end{aligned}$$

Les méthodes décrites ci-dessus reposent sur des prévisions et des observations non biaisées et permettent d'établir une estimation de  $\mathbf{B}$  (et de  $\mathbf{R}$ ) dans l'espace des observations. La méthode de Desroziers *et al.* (2005) est aujourd'hui souvent employée pour établir des diagnostics *a posteriori* permettant de régler des paramètres comme par exemple les variances des erreurs d'ébauche et d'observations.

### 3.2.2 La méthode NMC

Lorsque les innovations ne peuvent pas être utilisées efficacement par la méthode précédente, on cherche alors à exploiter une quantité supposée avoir les mêmes caractéristiques statistiques que l'erreur d'ébauche. La méthode NMC<sup>59</sup> proposée par Parrish et Derber (1992) met en œuvre des prévisions à différentes échéances mais valides au même instant. Supposons par exemple que l'on dispose de deux prévisions d'échéances différentes (il peut s'agir par exemple de prévisions à 10 et 20 jours en océanographie ou à 24 et 48 heures en météorologie). La première prévision (échéance la plus longue) est effectuée à partir d'une analyse à  $t_0$  tandis que la deuxième (échéance la plus courte) est effectuée à partir d'une analyse à l'instant  $t_1$ . Les deux prévisions sont alors valides au même instant  $t_2$  et peuvent être écrites comme l'état vrai entaché d'une erreur aléatoire  $\boldsymbol{\varepsilon}$  et d'un biais  $\mathbf{b}$  respectifs

$$\begin{aligned}\mathbf{x}^1 &= \mathcal{M}_{t_0 \rightarrow t_2}[\mathbf{x}^a] = \mathbf{x}^t + \boldsymbol{\varepsilon}^1 + \mathbf{b}^1 \\ \mathbf{x}^2 &= \mathcal{M}_{t_1 \rightarrow t_2}[\mathbf{x}^a] = \mathbf{x}^t + \boldsymbol{\varepsilon}^2 + \mathbf{b}^2.\end{aligned}$$

En supposant que les biais sont équivalents  $\mathbf{b}^1 = \mathbf{b}^2$ , la différence des deux prévisions donne

$$\Delta \mathbf{x} = \mathbf{x}^2 - \mathbf{x}^1 = \boldsymbol{\varepsilon}^2 - \boldsymbol{\varepsilon}^1.$$

Si de plus, les deux erreurs sont décorréelées et qu'elles possèdent les même covariances que l'erreur d'ébauche, on obtient

$$\begin{aligned}E[\Delta \mathbf{x} (\Delta \mathbf{x})^T] &= E[(\boldsymbol{\varepsilon}^2 - \boldsymbol{\varepsilon}^1) (\boldsymbol{\varepsilon}^2 - \boldsymbol{\varepsilon}^1)^T] \\ &= E[\boldsymbol{\varepsilon}^2 (\boldsymbol{\varepsilon}^2)^T] + E[\boldsymbol{\varepsilon}^1 (\boldsymbol{\varepsilon}^1)^T] \\ &\approx 2E[\boldsymbol{\varepsilon}^b (\boldsymbol{\varepsilon}^b)^T] \\ &= 2\mathbf{B}.\end{aligned}$$

Ainsi cette méthode permet d'estimer  $\mathbf{B}$  dans l'espace du modèle à partir de la différence de paires de prévisions à différentes échéances

$$\mathbf{B} \approx \frac{1}{2} E[(\mathbf{x}^2 - \mathbf{x}^1) (\mathbf{x}^2 - \mathbf{x}^1)^T].$$

Cependant, elle ne permet pas de différencier l'erreur d'ébauche proprement dite de l'erreur du modèle et cette dernière influence donc les statistiques obtenues. De plus, la méthode NMC fonctionne mal dans les régions peu observées. En effet, les différences entre les prévisions y sont faibles et une sous-estimation des variances en résulte. Enfin, pour les grandes échelles, les propriétés inhérentes à la méthode produisent des covariances sur-estimées (Berre *et al.*, 2006).

### 3.2.3 La méthode d'ensemble

La méthode d'ensemble est une méthode de type Monte Carlo appliquée à l'origine dans le cadre du filtre de Kalman (Evensen, 1994). Elle consiste à générer un ensemble de prévisions dont la dispersion est supposée représenter la dispersion de l'erreur d'ébauche, et à réalimenter le système prévisionnel

<sup>59</sup>National Meteorological Center. Ancienne dénomination du NCEP (National Center for Environmental Prediction).

avec les membres de cet ensemble à chaque cycle d'assimilation. Pour ce faire, les paramètres tels que l'ébauche, les forçages atmosphériques, les observations, l'erreur du modèle, . . . , sont perturbés de manière indépendante pour chaque membre et à chaque cycle d'assimilation. Si chaque perturbation est définie suivant la distribution de l'erreur du paramètre auquel elles est appliquée, la perturbation de l'analyse est alors elle-même définie par rapport à la distribution de l'erreur d'analyse (Fisher, 2004). Il est possible de ne pas perturber expressément l'ébauche, chaque membre de l'ensemble ayant naturellement une ébauche différentes des autres membres après le premier cycle. Dans ce cas, il convient néanmoins de réaliser un certain nombre de cycles afin que les conditions initiales soient "oubliées".

La figure 3.3 montre la procédure de cyclage<sup>60</sup> d'un ensemble pour deux membres  $p$  et  $q$  dans un système océanique. Avant chaque nouvelle fenêtre d'assimilation, les statistiques de la matrice  $\mathbf{B}$  sont estimées à partir de la prévision issue du cycle précédent. Pour chaque membre, la prévision du cycle  $c-1$  précédent constitue l'ébauche du nouveau cycle  $c$  entachée de son erreur  $\varepsilon^b$  propre. Cette ébauche est propagée par le modèle non linéaire et comparée aux observations  $\mathbf{y}^o$  entachées d'une erreur  $\varepsilon^o$  particulière à chaque membre. L'analyse ainsi fournie permet d'alimenter une prévision où les forçages atmosphériques  $\mathbf{f}$  sont perturbés par une erreur  $\varepsilon^f$  propre à chaque membre. A partir de l'ensemble de prévisions, on calcule les statistiques de  $\mathbf{B}$  pour le prochain cycle  $c+1$ .

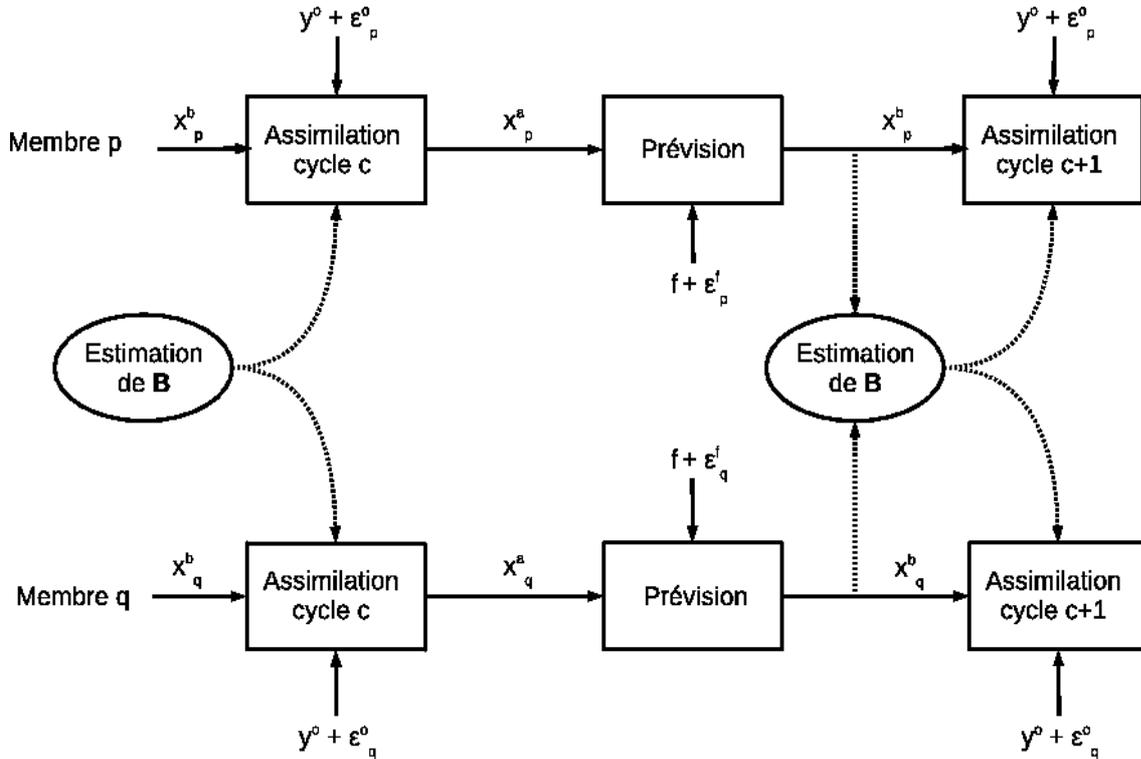


FIG. 3.3 – Procédure de cyclage d'un ensemble. Sur ce schéma, seuls les membres  $p$  et  $q$  de l'ensemble sont représentés. Voir le texte pour les commentaires.

Les statistiques de l'erreur d'ébauche peuvent être estimées à partir de la différence de l'ébauche de chacun des  $P-1$  membres perturbés par rapport à la moyenne des membres  $\bar{\mathbf{x}}^b$

$$\mathbf{B} = E \left[ \left( \mathbf{x}_p^b - \bar{\mathbf{x}}^b \right) \left( \mathbf{x}_p^b - \bar{\mathbf{x}}^b \right)^T \right] \approx \frac{1}{P-1} \sum_{p=1}^{P-1} \left( \mathbf{x}_p^b - \bar{\mathbf{x}}^b \right) \left( \mathbf{x}_p^b - \bar{\mathbf{x}}^b \right)^T .$$

Une formulation alternative consiste à moyenniser les différences des  $P$  paires de membres de l'ensemble

<sup>60</sup>Il existe différentes procédures de cyclage. On ne montre ici qu'un exemple possible pour l'océanographie.

(Fisher, 2004; Berre *et al.*, 2006; Daget *et al.*, 2009)

$$\mathbf{B} \approx \frac{1}{2}E \left[ \left( \mathbf{x}_p^b - \mathbf{x}_q^b \right) \left( \mathbf{x}_p^b - \mathbf{x}_q^b \right)^T \right] \approx \frac{1}{2(P-1)} \sum_{p=0}^{P-1} \left( \mathbf{x}_p^b - \mathbf{x}_{p+1}^b \right) \left( \mathbf{x}_p^b - \mathbf{x}_{p+1}^b \right)^T,$$

où  $\mathbf{x}_0^b = \mathbf{x}_P^b$  est l'ébauche non perturbée. Le facteur  $1/2(P-1)$  de cette formulation provient de la même démonstration que pour la méthode NMC précédente. En effet, lorsque l'espérance est calculée à partir de la différence entre deux membres perturbés, la somme obtenue est de  $2\mathbf{B}$ . Mais lorsque l'espérance est calculée à partir de la différence entre un membre perturbé et l'ébauche non perturbée, la somme obtenue correspond à  $\mathbf{B}$  comme dans la formulation précédente. Ici, il y a  $P-2$  paires de membres perturbés et 2 paires de membre perturbé/ébauche, soit une somme équivalente à  $(P-2) \times 2\mathbf{B} + 2 \times \mathbf{B} = 2(P-1) \times \mathbf{B}$ .

La méthode d'ensemble présente de nombreux avantages par rapport aux méthodes précédentes. En effet, elle permet d'estimer les statistiques du système en cours de manière plus réaliste et surtout sans être tributaire des régions peu observées. En calculant les espérances sur plusieurs cycles on peut obtenir des statistiques climatologiques (Fisher et Andersson, 2001). En les calculant sur un seul cycle, on peut introduire dans le système variationnel des statistiques dépendantes de l'écoulement, une caractéristique souhaitable pour l'amélioration des prévisions (Buehner, 2005). Néanmoins, cette méthode nécessite de définir les différentes perturbations utilisées à partir des covariances adéquates qu'il faut donc connaître. Pour les observations, définir ces statistiques est un problème en soi. Quant aux forçages atmosphériques, il est possible de définir l'erreur à partir de la différence de plusieurs produits (Daget *et al.*, 2009). Enfin, l'erreur du modèle devrait également être simulée, ce qui s'avère être une tâche difficile.

Un point important concerne la taille de l'ensemble. Compte tenu du coût induit, on estime aujourd'hui qu'il est possible de mettre en place un ensemble dont la taille est comprise entre une dizaine et une centaine de membres (Bannister, 2008a). Mais cela reste largement inférieur à la taille du système et la matrice ainsi estimée n'est pas de rang plein. En effet, pour un ensemble de  $P$  membres, la matrice estimée est au plus de rang  $P-1$  et l'incrément ainsi projeté (voir paragraphe 2.3.3) est généré dans un espace sous-optimal et ne peut probablement pas permettre au modèle de représenter l'état vrai. Pour pallier à ce problème, plusieurs stratégies peuvent être mises en place afin de réduire le nombre de paramètres indépendants. Une première possibilité consiste à faire l'hypothèse restrictive et peu réaliste de corrélations stationnaires, globalement homogènes et isotropes. Houtekamer et Mitchell (1998) montrent quant à eux dans le contexte d'un Filtre de Kalman d'Ensemble, qu'il est possible et cohérent de définir un rayon de coupure fonction de la taille de l'ensemble, et d'éliminer ainsi les corrélations parasites de longue portée. Evensen (2003) et Kepert (2009) remarquent néanmoins que cette localisation risque d'introduire des modes non dynamiques dans le modèle.

Une alternative à l'utilisation d'une matrice de faible rang (Houtekamer et Mitchell, 2001; Buehner et Charron, 2007; Oke, 2007) est de construire une estimation hybride en combinant une matrice statique de rang plein (avec hypothèse d'homogénéité et d'isotropie) d'une part, et une matrice issue de la localisation d'un ensemble d'autre part. Cette approche présente une grande flexibilité et semble donner de bons résultats (Hamill et Snyder, 2000; Buehner, 2005; Wang *et al.*, 2008). Mais l'utilisation d'une somme de deux matrices dans le cadre de l'assimilation variationnelle nécessite de doubler la taille du vecteur de contrôle lors de la minimisation, rendant ainsi l'algorithme moins efficace.

Une dernière alternative consiste non plus à estimer directement  $\mathbf{B}$  à partir de l'ensemble mais d'en déterminer les paramètres nécessaires à sa modélisation (Fisher, 2004; Belo Pereira et Berre, 2006; Berre *et al.*, 2006; Daget *et al.*, 2009). Cette technique est abordée dans le paragraphe 3.3 suivant.

### 3.3 Modélisation de $\mathbf{B}$ par une suite d'opérateurs

La méthode d'ensemble permettant d'estimer les statistiques de  $\mathbf{B}$  présente de nombreux avantages par rapport aux méthodes NMC ou celles basées sur les observations. Cependant, elle ne permet généralement pas d'obtenir une représentation de rang plein de la matrice puisque le nombre de membres de l'ensemble est très inférieur à la taille du vecteur d'état. Mais comme nous l'avons vu, ce n'est pas tant la matrice elle-même qui nous intéresse que son produit avec un vecteur  $\boldsymbol{\zeta}$  quelconque. Dans ce cas, il est possible d'obtenir une représentation de rang plein à partir d'un "modèle", *i.e.* de formuler  $\mathbf{B}$  comme un opérateur de covariance à partir d'une suite de  $L$  opérateurs linéaires (ou linéarisés) appliqués à ce vecteur, ce qui donne sous forme matricielle

$$\mathbf{B}\boldsymbol{\zeta} = \mathbf{A}_L \mathbf{A}_{L-1} \dots \mathbf{A}_1 \boldsymbol{\zeta}.$$

Cependant, suivant la nature des différents opérateurs, une telle formulation ne garantit pas d'obtenir une matrice  $\mathbf{B}$  symétrique, définie (semi-)positive, propriétés essentielles à un opérateur de covariance (voir Annexe B). Pour garantir ces propriétés, on préfère alors le formuler sous la forme (Gaspari et Cohn, 1999)

$$\mathbf{B} = \mathbf{U}\mathbf{U}^T,$$

où  $\mathbf{U}$  est un opérateur représentant une "racine carrée". Quant à  $\mathbf{U}^T$ , il représente en fait la forme matricielle de l'opérateur adjoint de  $\mathbf{U}$  (voir Annexe A).

Le paragraphe 2.3.1 montre que la matrice  $\mathbf{B}$  représente non seulement les covariances spatiales de chaque variable mais également les covariances qui existent entre chacune de ces variables. Derber et Bouttier (1999) proposent alors d'effectuer un changement de variables afin de transformer ce problème **multivarié** en plusieurs problèmes **univariés** indépendants. Pour cela il s'agit de définir un **opérateur d'équilibre**  $\mathbf{K}$  qui caractérise les relations physiques existant entre les différentes variables. La matrice  $\mathbf{B}$  s'écrit alors :

$$\mathbf{B} = \mathbf{K}\mathbf{B}_U\mathbf{K}^T,$$

où  $\mathbf{B}_U$  est la matrice bloc-diagonale des covariances spatiales de chaque variable. Par définition, chacune de ces covariances peut s'écrire comme un produit variances-corrélations. Ainsi, en définissant  $\mathbf{D}$  comme la matrice diagonale des **variances** et  $\mathbf{C}$  comme la matrice bloc-diagonale des **corrélations spatiales**, on peut écrire

$$\mathbf{B} = \mathbf{K}\mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2}\mathbf{K}^T.$$

La définition de chacun de ces opérateurs est décrite dans les paragraphes suivants.

#### 3.3.1 D'un problème multivarié à plusieurs problèmes univariés

L'idée fondamentale de Derber et Bouttier (1999) est de simplifier la spécification des covariances d'erreur d'ébauche en transformant l'espace du modèle, où les variables sont fortement corrélées, en un espace où les variables peuvent être considérées comme décorréelées.

Pour notre application à un modèle océanique, les variables du modèle sont la température  $T$ , la salinité  $S$ , la hauteur d'eau  $\eta$ , et les composantes horizontales de la vitesse du courant  $u$  et  $v$ . La matrice  $\mathbf{B}$  peut s'écrire sous la forme de la matrice bloc (voir paragraphe 2.3.1)

$$\mathbf{B} = \begin{pmatrix} E[\boldsymbol{\varepsilon}_T(\boldsymbol{\varepsilon}_T)^T] & E[\boldsymbol{\varepsilon}_T(\boldsymbol{\varepsilon}_S)^T] & E[\boldsymbol{\varepsilon}_T(\boldsymbol{\varepsilon}_\eta)^T] & E[\boldsymbol{\varepsilon}_T(\boldsymbol{\varepsilon}_u)^T] & E[\boldsymbol{\varepsilon}_T(\boldsymbol{\varepsilon}_v)^T] \\ E[\boldsymbol{\varepsilon}_S(\boldsymbol{\varepsilon}_T)^T] & E[\boldsymbol{\varepsilon}_S(\boldsymbol{\varepsilon}_S)^T] & E[\boldsymbol{\varepsilon}_S(\boldsymbol{\varepsilon}_\eta)^T] & E[\boldsymbol{\varepsilon}_S(\boldsymbol{\varepsilon}_u)^T] & E[\boldsymbol{\varepsilon}_S(\boldsymbol{\varepsilon}_v)^T] \\ E[\boldsymbol{\varepsilon}_\eta(\boldsymbol{\varepsilon}_T)^T] & E[\boldsymbol{\varepsilon}_\eta(\boldsymbol{\varepsilon}_S)^T] & E[\boldsymbol{\varepsilon}_\eta(\boldsymbol{\varepsilon}_\eta)^T] & E[\boldsymbol{\varepsilon}_\eta(\boldsymbol{\varepsilon}_u)^T] & E[\boldsymbol{\varepsilon}_\eta(\boldsymbol{\varepsilon}_v)^T] \\ E[\boldsymbol{\varepsilon}_u(\boldsymbol{\varepsilon}_T)^T] & E[\boldsymbol{\varepsilon}_u(\boldsymbol{\varepsilon}_S)^T] & E[\boldsymbol{\varepsilon}_u(\boldsymbol{\varepsilon}_\eta)^T] & E[\boldsymbol{\varepsilon}_u(\boldsymbol{\varepsilon}_u)^T] & E[\boldsymbol{\varepsilon}_u(\boldsymbol{\varepsilon}_v)^T] \\ E[\boldsymbol{\varepsilon}_v(\boldsymbol{\varepsilon}_T)^T] & E[\boldsymbol{\varepsilon}_v(\boldsymbol{\varepsilon}_S)^T] & E[\boldsymbol{\varepsilon}_v(\boldsymbol{\varepsilon}_\eta)^T] & E[\boldsymbol{\varepsilon}_v(\boldsymbol{\varepsilon}_u)^T] & E[\boldsymbol{\varepsilon}_v(\boldsymbol{\varepsilon}_v)^T] \end{pmatrix},$$

où chaque bloc représente une matrice de covariances spatiales uni- (sur la diagonale) ou multivariée (hors diagonale). Dans les applications océaniques, les covariances entre variables (termes hors diagonaux) ont souvent été négligées (voir par exemple Weaver *et al.*, 2003). Cette approximation se justifie dans le cadre d'une assimilation 4D-Var, puisque le modèle est considéré comme la contrainte du problème, et que par conséquent, ce dernier contient alors une composante multivariée. Mais dans le cadre d'un 3D-Var, et donc sans cette contrainte, l'approximation devient contestable. De plus, et quelque soit le schéma d'assimilation, des études montrent que lorsque l'on n'assimile qu'une ou partie des variables d'état, des structures parasites peuvent apparaître sur les variables non assimilées (voir par exemple Vialard *et al.*, 2003).

Derber et Bouttier (1999) considèrent donc qu'il est possible de séparer chacune des variables en composantes **équilibrées** ("*balanced*") et **non-équilibrées** ("*unbalanced*"), à l'exception d'une qui sera traitée en totalité et servira de point de départ à la définition de la partie équilibrée des autres variables. Chaque variable peut donc être exprimée par rapport à la variable totale grâce à des équations d'équilibre dynamique et de conservation de masse d'eau. Ces relations ne seront pas décrites ici, on se référera par exemple à Ricci *et al.* (2005) et Weaver *et al.* (2005) pour plus de détail.

En choisissant la température  $T$  comme variable totale, chacune des autres variables peut être décrite comme la somme de sa composante non-équilibrée (indice  $U$ ) et de sa composante équilibrée (indice  $B$ ), cette dernière correspondant à la transformation  $K_{xy}$  de la variable  $y$  en la variable  $x$ . On a donc les relations suivantes :

$$\begin{cases} T = T, \\ S = S_B + S_U = K_{ST}[T] + S_U, \\ \eta = \eta_B + \eta_U = K_{\eta\rho}[\rho] + \eta_U, \\ u = u_B + u_U = K_{up}[p] + u_U, \\ v = v_B + v_U = K_{vp}[p] + v_U, \end{cases} \quad (3.4)$$

avec

$$\begin{cases} \rho = K_{\rho T}[T] + K_{\rho S}[S], \\ p = K_{p\rho}[\rho] + K_{p\eta}[\eta], \end{cases} \quad (3.5)$$

où  $\rho$  est la densité et  $p$  la pression. En linéarisant ces opérateurs de transformation, on définit l'**opérateur d'équilibre** ("*balance operator*")  $\mathbf{K}$  tel que

$$\mathbf{B} = \mathbf{K}\mathbf{B}_U\mathbf{K}^T,$$

avec  $\mathbf{B}_U$  est la matrice bloc-diagonale des covariances univariées

$$\mathbf{B}_U = \begin{pmatrix} E[\boldsymbol{\varepsilon}_T(\boldsymbol{\varepsilon}_T)^T] & 0 & 0 & 0 & 0 \\ 0 & E[\boldsymbol{\varepsilon}_{S_U}(\boldsymbol{\varepsilon}_{S_U})^T] & 0 & 0 & 0 \\ 0 & 0 & E[\boldsymbol{\varepsilon}_{\eta_U}(\boldsymbol{\varepsilon}_{\eta_U})^T] & 0 & 0 \\ 0 & 0 & 0 & E[\boldsymbol{\varepsilon}_{u_U}(\boldsymbol{\varepsilon}_{u_U})^T] & 0 \\ 0 & 0 & 0 & 0 & E[\boldsymbol{\varepsilon}_{v_U}(\boldsymbol{\varepsilon}_{v_U})^T] \end{pmatrix}. \quad (3.6)$$

On notera que dans le cas où la vitesse du courant est considéré comme un vecteur  $(u, v)^T$  plutôt que comme deux champs scalaires distincts, les covariances  $E[\boldsymbol{\varepsilon}_{u_U}(\boldsymbol{\varepsilon}_{v_U})^T]$  et  $E[\boldsymbol{\varepsilon}_{v_U}(\boldsymbol{\varepsilon}_{u_U})^T]$  ne sont pas nulles. L'opérateur  $\mathbf{K}$  se présente sous la forme de la matrice bloc triangulaire inférieure

$$\begin{pmatrix} \mathbf{I} & 0 & 0 & 0 & 0 \\ \mathbf{K}_{ST} & \mathbf{I} & 0 & 0 & 0 \\ \mathbf{K}_{\eta T} & \mathbf{K}_{\eta S} & \mathbf{I} & 0 & 0 \\ \mathbf{K}_{uT} & \mathbf{K}_{uS} & \mathbf{K}_{u\eta} & \mathbf{I} & 0 \\ \mathbf{K}_{vT} & \mathbf{K}_{vS} & \mathbf{K}_{v\eta} & 0 & \mathbf{I} \end{pmatrix},$$

où, d'après les équations (3.4) et (3.5) lorsqu'elles sont linéarisées,

$$\begin{aligned}
\mathbf{K}_{\eta T} &= \mathbf{K}_{\eta\rho} (\mathbf{K}_{\rho T} + \mathbf{K}_{\rho S} \mathbf{K}_{ST}), \\
\mathbf{K}_{uT} &= \mathbf{K}_{up} (\mathbf{K}_{p\rho} + \mathbf{K}_{p\eta} \mathbf{K}_{\eta\rho}) (\mathbf{K}_{\rho T} + \mathbf{K}_{\rho S} \mathbf{K}_{ST}), \\
\mathbf{K}_{vT} &= \mathbf{K}_{vp} (\mathbf{K}_{p\rho} + \mathbf{K}_{p\eta} \mathbf{K}_{\eta\rho}) (\mathbf{K}_{\rho T} + \mathbf{K}_{\rho S} \mathbf{K}_{ST}), \\
\mathbf{K}_{\eta S} &= \mathbf{K}_{\eta\rho} \mathbf{K}_{\rho S}, \\
\mathbf{K}_{uS} &= \mathbf{K}_{up} (\mathbf{K}_{p\rho} + \mathbf{K}_{p\eta} \mathbf{K}_{\eta\rho}) \mathbf{K}_{\rho S}, \\
\mathbf{K}_{vS} &= \mathbf{K}_{vp} (\mathbf{K}_{p\rho} + \mathbf{K}_{p\eta} \mathbf{K}_{\eta\rho}) \mathbf{K}_{\rho S}, \\
\mathbf{K}_{u\eta} &= \mathbf{K}_{up} \mathbf{K}_{p\eta}, \\
\mathbf{K}_{v\eta} &= \mathbf{K}_{vp} \mathbf{K}_{p\eta}.
\end{aligned}$$

### 3.3.2 Modélisation des covariances univariées

L'opérateur d'équilibre permet de traiter les covariances d'erreur d'ébauche qui existent entre les différentes variables en définissant les contraintes qui relient ces dernières entre elles. Mais chaque variable possède une composante qui lui est propre et donc une erreur associée qui, par définition, est indépendante des autres, mais pour laquelle il faut définir des covariances spatiales. Ce sont ces covariances spatiales pour chaque variable qui constituent les blocs diagonaux de la matrice  $\mathbf{B}_U$  donnée par l'équation (3.6). La taille de la matrice à estimer ou modéliser se trouve donc grandement réduite. Néanmoins, et bien que chacun de ces blocs puisse être traité individuellement, le problème reste d'une taille conséquente ( $\mathcal{O}(10^{12})$  pour chaque bloc sur une grille grossière), et ne peut donc pas être abordé de façon explicite. Par définition, ces covariances spatiales peuvent s'écrire en séparant les variances des corrélations

$$\mathbf{B}_U = \mathbf{D}^{1/2} \mathbf{C} \mathbf{D}^{1/2},$$

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

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_T & 0 & 0 & 0 & 0 \\ 0 & \mathbf{C}_{S_U} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{C}_{\eta_U} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{C}_{u_U} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{C}_{v_U} \end{pmatrix}.$$

Ainsi, pour chaque variable, deux problèmes sont à considérer : la définition des variances d'une part (voir paragraphe 3.5), et celle des corrélations d'autre part. De façon générale, les techniques employées pour mener à bien cette dernière tâche dépend du domaine dans lequel on travaille. En météorologie, on utilise souvent une approche où il s'agit de trouver un espace dans lequel la matrice  $\mathbf{B}_U$  est supposée diagonale (voir une description sommaire donnée au paragraphe 3.4). Mais cette approche permet difficilement de manipuler des frontières complexes, problème essentiel à l'océanographie. Dans ce cas, on préfère alors spécifier des fonctions de corrélation que l'on cherche à modéliser par un algorithme efficace (voir paragraphe 3.6).

## 3.4 Les différentes hypothèses diagonales

Pour une variable  $\chi$  donnée, il est possible de formuler ses covariances spatiales sous la forme d'une factorisation utilisant les variances d'une part ( $\mathbf{D}_\chi$ ), et les corrélations ( $\mathbf{C}_\chi$ ) d'autre part :

$$\mathbf{B}_\chi = (\mathbf{D}_\chi)^{1/2} \mathbf{C}_\chi (\mathbf{D}_\chi)^{1/2}.$$

Afin de simplifier les notations, l'indice  $\chi$ , n'est pas utilisé dans ce qui suit. On se rappellera néanmoins, qu'il s'agit de covariances, variances et corrélations spatiales pour une variable donnée.

On notera que les corrélations sont généralement séparées en corrélations horizontales ( $\mathbf{C}_h$ ) et verticales ( $\mathbf{C}_v$ ). En général, le nombre de niveaux verticaux d'un modèle est très faible comparé aux nombre de points horizontaux. Le problème de taille concerne donc plutôt les corrélations horizontales que verticales. Afin d'introduire une certaine dépendance horizontale / verticale (non-séparabilité) et afin de respecter la nécessité de formuler  $\mathbf{B}$  sous une forme factorisée, on écrit

$$\mathbf{C} = \mathbf{C}_h^{1/2} \mathbf{C}_v^{1/2} \left( \mathbf{C}_v^{1/2} \right)^T \left( \mathbf{C}_h^{1/2} \right)^T. \quad (3.7)$$

Pour toute matrice symétrique réelle, on peut définir une matrice réelle orthogonale qui permette de la diagonaliser. En d'autres termes, il existe une base orthonormale dans laquelle la matrice peut être représentée sous forme diagonale. Or, par définition, la matrice des covariances d'erreur d'ébauche est une matrice symétrique réelle. S'il était possible de la connaître et de la stocker, et si on savait définir l'opérateur  $\mathbf{T}$  permettant le passage entre cette base, dite de Karhunen-Loève, et la base d'origine, on saurait alors réduire  $\mathbf{B}$  à une matrice diagonale  $\hat{\mathbf{B}}$ . La base même de Karhunen-Loève n'étant généralement pas connue, le principe de cette méthode que l'on décrit ci-après (d'après Deckmyn et Berre, 2005), est donc de trouver un espace approprié (et sa base) pour lequel l'**hypothèse diagonale** est suffisamment valide. En supposant que l'on sache l'estimer, cette diagonale peut être manipulée et stockée. Modéliser  $\mathbf{B}$  consiste alors à appliquer l'opérateur de passage  $\mathbf{T}$  à cette diagonale.

Si  $\boldsymbol{\varepsilon}$  est l'erreur dans l'espace physique (point de grille), on peut définir  $\tilde{\boldsymbol{\varepsilon}}$  comme l'erreur normalisée par les écarts-type en point de grille

$$\tilde{\boldsymbol{\varepsilon}} = \mathbf{D}_g^{-1/2} \boldsymbol{\varepsilon},$$

de telle manière que

$$\mathbf{B} = E [\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T] = \mathbf{D}_g^{1/2} E [\tilde{\boldsymbol{\varepsilon}} \tilde{\boldsymbol{\varepsilon}}^T] \mathbf{D}_g^{1/2} = \mathbf{D}_g^{1/2} \mathbf{C} \mathbf{D}_g^{1/2}.$$

En appliquant l'opérateur  $\mathbf{T}$  à cette erreur normalisée, on définit dans l'espace approprié l'erreur  $\boldsymbol{\varepsilon}_t$  telle que

$$\boldsymbol{\varepsilon}_t = \mathbf{T} \tilde{\boldsymbol{\varepsilon}},$$

et on a ainsi

$$\begin{aligned} \mathbf{B} &= \mathbf{D}_g^{1/2} E [\tilde{\boldsymbol{\varepsilon}} \tilde{\boldsymbol{\varepsilon}}^T] \mathbf{D}_g^{1/2} \\ &= \mathbf{D}_g^{1/2} \mathbf{T}^{-1} E [\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T] (\mathbf{T}^{-1})^T \mathbf{D}_g^{1/2} \\ &= \mathbf{D}_g^{1/2} \mathbf{T}^{-1} \mathbf{B}_t (\mathbf{T}^{-1})^T \mathbf{D}_g^{1/2}, \end{aligned}$$

où  $\mathbf{B}_t$  est maintenant une matrice diagonale. Comme il s'agit d'une matrice de covariances, on peut aussi l'exprimer en définissant une matrice diagonale des variances  $\mathbf{D}_t$  dans ce nouvel espace, tel que la matrice diagonale des corrélations  $\mathbf{C}_t$  correspond simplement à la matrice identité, puisque la diagonale de toute matrice de corrélations est composée de 1

$$\mathbf{B}_t = \mathbf{D}_t^{1/2} \mathbf{C}_t \mathbf{D}_t^{1/2} = \mathbf{D}_t^{1/2} \mathbf{I} \mathbf{D}_t^{1/2}.$$

### Espace physique : $\mathbf{T} = \mathbf{I}$

Une première hypothèse consiste à considérer que la matrice  $\mathbf{B}$  est diagonale dans l'espace physique et donc d'imposer  $\mathbf{T} = \mathbf{I}$ , l'identité. On a alors

$$\mathbf{B} = \mathbf{D}_g^{1/2} \mathbf{B}_i \mathbf{D}_g^{1/2} = \mathbf{D}_g^{1/2} \mathbf{I} \mathbf{D}_g^{1/2}, \quad (3.8)$$

où la matrice diagonale  $\mathbf{B}_i$  est équivalente à la matrice identité puisqu'il n'y a pas réellement de transformation d'espace.

Pour cette hypothèse, l'équation (3.8) montre qu'il est possible de définir une variance différente pour chaque point de grille (erreur locale) mais que l'on considère qu'il n'existe aucune corrélation

spatiale entre ces différents points. Si elle permet d'accéder à des calculs simples, cette assumption est cependant peu réaliste, les erreurs étant généralement structurées. Pour la rendre plus vériste, on peut alors définir des corrélations verticales différentes en fonction de la position horizontale. La matrice ainsi définie n'est plus tout à fait diagonale mais reste manipulable compte tenu du faible nombre de niveaux verticaux par rapport au nombre de points horizontaux. Néanmoins, l'hypothèse reste très restrictive.

### Espace spectral : $\mathbf{T} = \mathbf{F}$

Beaucoup de modèles météorologiques utilisent l'espace spectral. Cet espace correspond à une transformation de Fourier généralisée (ou Fourier-Legendre) de l'espace physique. Sur la sphère en particulier, il s'agit de la décomposition en harmoniques sphériques (voir par exemple Arfken, 1966) de tout champ scalaire  $\zeta(\lambda, \phi, z)$

$$\zeta(\lambda, \phi, z) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \zeta_l^m(z) Y_l^m(\lambda, \phi),$$

où  $\lambda$  est la longitude ( $0 \leq \lambda \leq 2\pi$ ),  $\phi$  la latitude ( $-\pi/2 \leq \phi \leq \pi/2$ ) et  $z$  la profondeur.  $Y_l^m(\lambda, \phi)$  représente les harmoniques sphériques avec  $m$  le nombre d'onde zonal et  $l$  le nombre d'onde total. Ainsi, si dans l'espace physique, un signal peut être vu comme l'évolution dans le temps de sa valeur à une position géographique donnée, dans l'espace spectral, il représente l'évolution dans le temps du coefficient associé à un nombre d'onde donné. On appellera  $\mathbf{F}$ , l'opérateur permettant de passer de l'espace physique à l'espace spectral. Les mérites et inconvénients de l'utilisation de l'espace spectral ne seront pas discutés ici.

Courtier *et al.* (1998) montrent que si on considère que les corrélations sont homogènes et isotropes dans l'espace physique, alors les harmoniques sphériques sont orthogonales et forment ainsi une base (voir aussi Berre (2000) dans le cadre d'un modèle à aire limitée). Ainsi, dans cette base spectrale, la matrice  $\mathbf{B}$  peut être représentée par une diagonale. En imposant  $\mathbf{T} = \mathbf{F}$ , on a alors

$$\mathbf{B} = \mathbf{D}_g^{1/2} \mathbf{F}^{-1} \mathbf{D}_f^{1/2} \mathbf{D}_f^{1/2} (\mathbf{F}^{-1})^T \mathbf{D}_g^{1/2}. \quad (3.9)$$

Ici, les variances spectrales de la matrice  $\mathbf{D}_f$  de l'équation (3.9) représentent l'amplitude de la contribution de chacun des différents nombres d'onde. L'hypothèse diagonale consiste à considérer qu'il n'existe pas de corrélations entre, non plus les différents points de grille comme précédemment, mais entre ces différentes échelles. Un point important à noter ici est que, si la transformation  $\mathbf{F}^{-1}$  était appliquée aux covariances plutôt qu'aux corrélations ( $\varepsilon_f = \mathbf{F} \varepsilon$  au lieu de  $\varepsilon_f = \mathbf{F} \tilde{\varepsilon}$ ), on aurait

$$\mathbf{B} = \mathbf{F}^{-1} \mathbf{D}_f^{1/2} \mathbf{D}_f^{1/2} (\mathbf{F}^{-1})^T,$$

où  $\mathbf{D}_f^{1/2} \neq \mathbf{D}_f^{1/2}$ . Dans l'espace physique, les variances de  $\mathbf{B}$  seraient alors uniformes, résultat global des différentes contributions spectrales. Ainsi, les covariances en point de grille seraient non seulement homogènes et isotropes mais également de variance constante, ce qui est peu réaliste. Il est donc important de pouvoir appliquer des variances en point de grille et, par conséquence, d'appliquer l'opérateur  $\mathbf{F}$  aux corrélations, et non pas aux covariances.

Considérer que les corrélations sont homogènes et isotropes dans l'espace physique reste cependant une hypothèse restrictive. Pour introduire de l'anisotropie, Desroziers (1997) propose par exemple d'effectuer un changement de coordonnées de telle manière que les corrélations soient isotropes dans cette nouvelle base mais anisotrope dans la base originelle. Enfin, il est également possible ici de définir des corrélations verticales et de les faire varier en fonction du nombre d'onde horizontal.

### Espace des ondelettes : $\mathbf{T} = \mathbf{Z}$

Si l'hypothèse diagonale dans l'espace spectral est attractive, elle n'en reste pas moins limitée à des corrélations homogènes et isotropes dans l'espace physique. Une approche naturelle consiste alors à utiliser une base d'ondelettes, puisque celle-ci possède à la fois des composantes spatiales et spectrales. On n'entrera pas ici dans la description des différentes ondelettes et de leurs caractéristiques (voir Mallat, 2009). On considèrera simplement qu'un champ scalaire  $\zeta(\lambda, \phi, z)$  peut s'écrire

$$\zeta(\lambda, \phi, z) = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \zeta_{j,k}(z) \psi_{j,k}(\lambda, \phi),$$

où  $\lambda$  est la longitude ( $0 \leq \lambda \leq 2\pi$ ),  $\phi$  la latitude ( $-\pi/2 \leq \phi \leq \pi/2$ ) et  $z$  la profondeur.  $\psi_{j,k}(\lambda, \phi)$  représente les fonctions ondelettes avec  $j$  l'échelle (nombre d'onde),  $k$  représentant une distance entre les fonctions pour une même échelle. On appellera  $\mathbf{Z}$  l'opérateur permettant de passer de l'espace physique à l'espace des ondelettes.

En imposant  $\mathbf{T} = \mathbf{Z}$ , on a alors

$$\mathbf{B} = \mathbf{D}_g^{1/2} \mathbf{Z}^{-1} \mathbf{D}_w^{1/2} \mathbf{D}_w^{1/2} (\mathbf{Z}^{-1})^T \mathbf{D}_g^{1/2}. \quad (3.10)$$

Cependant Deckmyn et Berre (2005) illustrent comment cette approche donnée par l'équation (3.10) conduit à des problèmes de sous-échantillonnage de la base d'ondelettes et induit certaines limitations dans l'hypothèse diagonale.

### Espace hybride spectral/ondelettes : $\mathbf{T}_1 = \mathbf{F}$ et $\mathbf{T}_2 = \mathbf{ZF}^{-1}$

Pour pallier aux problèmes précédents, il est possible de combiner les approches de l'espace spectral et de l'espace ondelettes en effectuant non plus une, mais deux transformations. Dans l'espace spectral, nous avons défini l'erreur  $\boldsymbol{\varepsilon}_f$  par rapport à l'erreur normalisée en point de grille  $\tilde{\boldsymbol{\varepsilon}}$  par

$$\boldsymbol{\varepsilon}_f = \mathbf{F} \tilde{\boldsymbol{\varepsilon}},$$

ce qui nous a conduit à l'équation (3.9), où nous avons considéré que la matrice des corrélations  $\mathbf{C}_f$  était équivalente à l'identité. En supposant que ce ne soit pas le cas, *i.e.* que l'on considère qu'il existe des corrélations entre les différents nombre d'ondes, relâchant alors l'hypothèse d'homogénéité et d'isotropie dans l'espace physique, on aurait alors

$$\mathbf{B} = \mathbf{D}_g^{1/2} \mathbf{F}^{-1} \mathbf{D}_f^{1/2} \mathbf{C}_f \mathbf{D}_f^{1/2} (\mathbf{F}^{-1})^T \mathbf{D}_g^{1/2},$$

avec  $\mathbf{C}_f \neq \mathbf{I}$ . Cette matrice étant également réelle et symétrique, il est donc possible de la diagonaliser en appliquant une nouvelle transformation. Pour cela, on peut normaliser l'erreur spectrale par les variances spectrales

$$\tilde{\boldsymbol{\varepsilon}}_f = \mathbf{D}_f^{-1/2} \boldsymbol{\varepsilon}_f,$$

et la ramener dans l'espace des ondelettes en définissant l'erreur

$$\boldsymbol{\varepsilon}_w = \mathbf{ZF}^{-1} \tilde{\boldsymbol{\varepsilon}}_f.$$

On notera l'utilisation de l'opérateur  $\mathbf{F}^{-1}$ , étape intermédiaire pour ramener l'erreur de l'espace spectral à l'espace physique avant d'appliquer l'opérateur  $\mathbf{Z}$  permettant de passer de l'espace physique à l'espace spectral. On obtient alors

$$\mathbf{B} = \mathbf{D}_g^{1/2} \mathbf{F}^{-1} \mathbf{D}_f^{1/2} \mathbf{F} \mathbf{Z}^{-1} \mathbf{D}_w^{1/2} \mathbf{D}_w^{1/2} (\mathbf{Z}^{-1})^T \mathbf{F}^T \mathbf{D}_f^{1/2} (\mathbf{F}^{-1})^T \mathbf{D}_g^{1/2}, \quad (3.11)$$

où on a considéré cette fois que la matrice résultante  $\mathbf{B}_w = \mathbf{D}_w^{1/2} \mathbf{I} \mathbf{D}_w^{1/2}$  était diagonale. On notera que lorsque le modèle travaille dans l'espace spectral, la matrice  $\mathbf{B}$  doit être définie dans cet espace. Il faut

alors rajouter en début et fin de l'expression donnée par l'équation (3.11) une transformation  $\mathbf{F}$  et  $\mathbf{F}^T$  respectivement.

Fisher (2004) propose de se limiter à un certain nombre d'ondelettes en bornant les échelles ( $j = 0, \dots, K$ ) et de définir pour chacune de ces échelles, des corrélations verticales en chaque point de grille horizontal  $\mathbf{C}_j(\lambda, \phi)$ . Les différentes ondelettes peuvent être vues comme des filtres passe-bande, c'est à dire qu'elles permettent de représenter différents niveaux de détail du signal qu'elles transforment : du détail grossier de la grande échelle ( $j$  petit) au détail fin de la petite échelle ( $j$  grand). Ainsi pour une position horizontale donnée, les différents  $\mathbf{C}_j$  associés représentent des niveaux de détails différents de la fonction de corrélation verticale totale en ce point.

Appliquer une matrice  $\mathbf{B}$  issue d'une hypothèse diagonale nécessite néanmoins de définir les différentes matrices des variances, et éventuellement les corrélations verticales (Pannekoucke (2009) propose de modéliser ces corrélations à partir d'un opérateur de diffusion - voir paragraphe 3.6.3). Pour cela, il est par exemple possible d'utiliser une technique d'ensemble (voir paragraphe 3.2) en ramenant les erreurs de chaque membre dans l'espace adéquat avant d'en calculer les statistiques (on revient sur les variances en point de grille au paragraphe 3.5). Mais comme nous l'avons vu, la méthode d'ensemble est une technique coûteuse et il est nécessaire d'en limiter le nombre de membres. Mais un ensemble de petite taille génère un bruit d'échantillonnage qu'il est nécessaire de filtrer. Buehner et Charron (2007) comparent les bénéfices respectifs d'une localisation spectrale ou spatiale. Ces méthodes sont basées sur l'hypothèse que les corrélations tendent vers 0 lorsque la distance entre les nombre d'ondes (localisation spectrale), ou la distance entre deux points spatiaux (localisation spatiale), s'accroît. Une fonction de localisation adéquate (égale à 0 à partir d'une certaine distance) est alors appliquée dans l'espace idoine. Pannekoucke *et al.* (2007) montrent quant à eux, que la transformation en ondelettes utilisée dans l'approche diagonale hybride permet également de filtrer naturellement ce bruit d'échantillonnage.

## 3.5 Spécification des variances

Comme nous l'avons vu dans le paragraphe précédent, les variances doivent être définies dans l'espace physique même lorsque l'on utilise l'hypothèse diagonale dans un espace différent.

### Paramétrisation

En océanographie, une première façon de spécifier les variances est de les paramétrer en fonction de connaissances physiques et de diverses hypothèses. Weaver *et al.* (2003) illustrent par exemple comment cette simple technique permet de capturer certains effets dynamiques pris en compte naturellement par un 4D-Var. Les variances pour un champ donné peuvent par exemple être calculées par rapport au gradient vertical de ce champ et bornées par des variances maximales que l'on définit suivant que l'on est dans la couche de mélange, la thermocline<sup>61</sup> ou l'océan profond (Weaver *et al.*, 2005; Cummings, 2005; Balsmeda *et al.*, 2008). En recalculant cette paramétrisation pour chaque cycle, cette technique permet d'introduire une dépendance à l'écoulement.

### Estimation

Une autre façon de spécifier les variances est de les estimer à partir d'une des méthodes décrites au paragraphe 3.2. Depuis quelques années, il est fait de plus en plus souvent appel à la méthode d'ensemble, qui présente de multiples avantages par rapport aux techniques basées sur les observations ou à la méthode NMC. Mais son coût n'est pas négligeable et n'est abordable, en particulier dans le cadre d'une dépendance à l'écoulement, que si le nombre de membres est limité. Or, utiliser un ensemble de

---

<sup>61</sup>Couche de transition thermique rapide séparant la couche de mélange (eaux superficielles) chauffée par le soleil, de l'océan profond.

petite taille génère un bruit d'échantillonnage qu'il est nécessaire de filtrer pour ne pas dégrader les statistiques calculées.

Ce bruit d'échantillonnage est de petite échelle et fortement lié aux caractéristiques des corrélations (Raynaud *et al.*, 2008). A partir des travaux de Berre *et al.* (2007), Raynaud *et al.* (2009) proposent un algorithme de filtrage basé sur le rapport signal à bruit des variances. Appliquée dans le cadre d'un modèle météorologique spectral, cette technique est peu coûteuse à mettre en place, puisqu'elle tire parti de l'opérateur de transformation de l'espace physique à l'espace spectral déjà existant. Un système d'ensemble permettant d'estimer des variances pour une assimilation variationnelle a également été développé en océanographie (Daget *et al.*, 2009).

## 3.6 Spécification des fonctions de corrélation

Un des problèmes principaux en océanographie est la gestion de frontières complexes, ce qui rend difficile l'application de techniques liées à d'autres espaces que l'espace point de grille. Or, parmi les différentes hypothèses diagonales décrites au paragraphe 3.4, celle de l'espace physique est la plus irréaliste. On peut alors imaginer utiliser l'hypothèse diagonale dans un autre espace, puis, une fois revenu dans l'espace physique, appliquer un masque représentant les différentes côtes et la bathymétrie. Mais une telle utilisation a des effets entraînant des incohérences physiques comme le montre la figure 3.4. Dans cet exemple, on trace sur la grille du modèle la zone d'influence (corrélation) d'une observation lorsqu'elle est assimilée. Après avoir appliqué un masque éliminant les mailles de terre, on s'aperçoit que cette technique permet à une observation située dans l'Océan Pacifique près des côtes du Panama d'avoir une influence sur une zone de la Mer des Caraïbes, ce qui est physiquement incohérent.

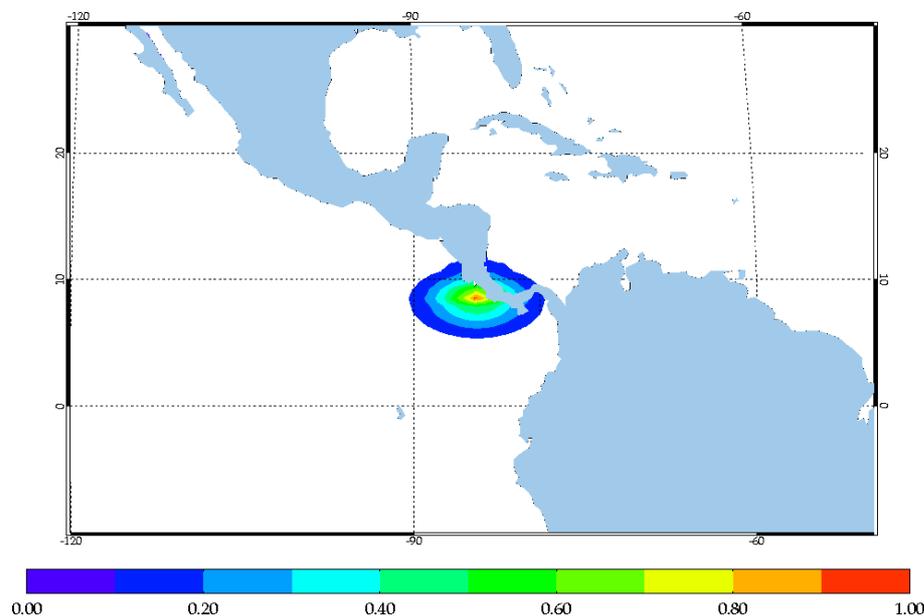


FIG. 3.4 – On assimile une observation située dans l'Océan Pacifique près de la côte du Panama. Sa zone d'influence (corrélation) est calculée sur la grille du modèle, puis un masque représentant les mailles de terre est appliqué. L'observation étant donc lissée sur la grille du modèle sans tenir compte des côtes, on obtient ici des corrélations irréalistes qui traversent le continent entre l'Océan Pacifique et la Mer des Caraïbes. Ces corrélations peuvent aller jusqu'à 0,4.

L'hypothèse diagonale étant difficilement applicable, il convient alors de trouver d'autres postulats permettant de gérer au mieux les frontières physiques. Ici, plutôt que d'émettre une hypothèse sur la forme de la matrice de corrélations, on présume de la forme des fonctions de corrélation qui la

composent. Ainsi, effectuer le produit de la matrice des corrélations avec un vecteur champ  $\zeta$ , c'est appliquer à ce champ un opérateur de corrélation  $\mathcal{C}$  dont le noyau est la fonction de corrélation présumée  $c(z, z')$  (voir Annexe B)

$$\mathcal{C}[\zeta(z)] = \int_{-\infty}^{\infty} c(z, z')\zeta(z') dz'.$$

Lorsqu'une fonction de corrélation est homogène et isotrope, elle ne dépend que de la distance euclidienne entre les points et on peut écrire  $c(z, z') = c(z - z')$ . Un opérateur représenté par une telle fonction est un opérateur de convolution

$$\mathcal{C}[\zeta(z)] = \int_{-\infty}^{\infty} c(z - z')\zeta(z') dz' = (c * \zeta)(z). \quad (3.12)$$

Appliquer un tel opérateur, c'est donc évaluer des intégrales de convolution. Le paragraphe 3.6.1 montre comment l'utilisation de fonctions à support compact permet de réduire le coût de ce calcul. Néanmoins, l'évaluation de ces intégrales peut s'avérer difficile et limitée à certaines fonctions de corrélation. C'est pourquoi on préfère souvent trouver une méthode alternative, *i.e* une façon d'évaluer ces intégrales sans les calculer de manière explicite. Deux méthodes alternatives sont décrites ici : le filtre récursif (paragraphe 3.6.2) et l'équation de diffusion (paragraphe 3.6.3).

On montrera également qu'il est possible d'étendre ces méthodes à des fonctions de corrélations inhomogènes et anisotropes. Mais ces techniques permettent en général d'évaluer ces intégrales à un facteur prêt. Il est donc nécessaire de normaliser le résultat qu'elles fournissent afin que les corrélations ainsi définies présentent bien un maximum à 1 (paragraphe 3.6.4).

### 3.6.1 Evaluation des intégrales de convolution

En océanographie comme en météorologie, il est communément admis que les fonctions de corrélations impliquées s'annulent à partir d'une certaine distance. Il semble alors naturel de s'intéresser aux fonctions à support compact pour les construire de façon efficace. Ainsi, l'intégrale de convolution donnée par l'équation (3.12) ne doit plus être évaluée sur le domaine complet mais uniquement sur le support de la fonction de corrélation, ce qui réduit généralement le coût du calcul de manière significative.

Gaspari et Cohn (1999) montrent par exemple comment obtenir une fonction de corrélation auto-régressive homogène et isotrope (famille de Matérn ; voir chapitre 4) en auto-convolant des fonctions exponentielles ou certaines fonctions continues et linéaires par morceaux. Dans ce dernier cas, un paramètre supplémentaire permet même d'obtenir des lobes négatifs, une caractéristique désirable pour certains champs physiques (voir par exemple Julian et Thiébaux, 1975). On notera que si le support de la fonction de base est  $[-l, l]$ , le support de la fonction de corrélation issue du produit de convolution est  $[-2l, 2l]$ .

Gneiting (2002) décrit plusieurs méthodes existant dans la littérature et propose ses propres constructions. Il montre en particulier qu'il est possible d'obtenir des fonctions de la classe exponentielle à une certaine puissance à partir de fonctions puissance tronquées, et des fonctions de la famille de Matérn à partir d'un produit de différentes fonctions.

Enfin, Gaspari *et al.* (2006) notent qu'il est possible d'obtenir des fonctions inhomogènes et/ou anisotropes en appliquant une transformation adéquate à la fonction homogène et isotrope. Un des exemples les plus courants sur lequel nous reviendrons au chapitre 5, consiste à étirer ou contracter les coordonnées.

### 3.6.2 Le filtre récursif

Par définition (Raymond et Garder, 1991, et les références citées), un filtre récursif est un filtre dont la sortie  $s$  dépend non seulement des quantités d'entrée  $p$  mais également des quantités filtrées précédentes

$$s_n = f(p_0, p_1, \dots, p_K, s_0, s_1, \dots, s_{n-1}),$$

tandis qu'un filtre non-récursif ne dépend que des quantités en entrée. L'ordre du filtre indique le nombre des sorties précédentes à utiliser. Par exemple, pour un filtre de second ordre, les quantités filtrées  $s_{n-1}$  et  $s_{n-2}$  sont nécessaires au calcul de  $s_n$ .

Suivant Hayden et Purser (1995), prenons un champ scalaire  $\mathbf{p}$  sur un domaine 1D infini, auquel on applique le filtre spatial de premier ordre suivant, afin d'obtenir le champ filtré  $\mathbf{q}$

$$q_i = \alpha q_{i-1} + (1 - \alpha) p_i, \quad \text{avec } 0 < \alpha < 1,$$

où l'indice  $i$  représente le  $i$ -ème élément du champ. En développant de façon répétitive l'expression précédente, il est possible de n'exprimer un élément du champ filtré  $q_i$  qu'à partir des éléments du champ d'entrée  $\mathbf{p}$  précédant cette position :

$$\begin{aligned} q_i &= \alpha q_{i-1} + (1 - \alpha) p_i \\ &= \alpha (\alpha q_{i-2} + (1 - \alpha) p_{i-1}) + (1 - \alpha) p_i \\ &= \alpha^2 (\alpha q_{i-3} + (1 - \alpha) p_{i-2}) + (1 - \alpha) (p_i + \alpha p_{i-1}) \\ &\vdots \\ &= (1 - \alpha) (p_i + \alpha p_{i-1} + \dots + \alpha^j p_{i-j} + \dots) \\ &= \sum_{j=0}^{\infty} (1 - \alpha) \alpha^j p_{i-j}. \end{aligned} \tag{3.13}$$

Or, la définition d'un produit de convolution dans un domaine discret entre un noyau  $\mathbf{E}$  et un champ  $\mathbf{p}$  s'écrit pour la  $i$ -ème composante

$$(\mathbf{E} * \mathbf{p})_i = \sum_j E_j p_{i-j}.$$

L'application du filtre récursif décrite par l'équation (3.13) constitue donc le produit de convolution

$$\mathbf{q} = \mathbf{E}^+ * \mathbf{p}, \quad \text{avec } \mathbf{E}_j^+ = \begin{cases} (1 - \alpha) \alpha^j, & j \geq 0 \\ 0, & j < 0 \end{cases}.$$

Mais l'application unidirectionnelle de ce filtre correspond à une formulation qui introduit un déphasage dans la solution (Raymond et Garder, 1991). Pour palier ce problème, on définit alors un "filtre adjoint", qui, appliqué dans la direction opposée au champ  $\mathbf{q}$  précédent, permet de symétriser le filtre total, et d'obtenir une solution  $\mathbf{s}$  qui n'est plus déphasée. Comme pour l'équation (3.13), on peut écrire

$$\begin{aligned} s_i &= \alpha s_{i+1} + (1 - \alpha) q_i \\ &= \alpha (\alpha s_{i+2} + (1 - \alpha) q_{i+1}) + (1 - \alpha) q_i \\ &\vdots \\ &= (1 - \alpha) (q_i + \alpha q_{i+1} + \dots + \alpha^j q_{i+j} + \dots) \\ &= \sum_{j=0}^{\infty} (1 - \alpha) \alpha^j q_{i+j} \\ &= \sum_{j=0}^{-\infty} (1 - \alpha) \alpha^{-j} q_{i-j}. \end{aligned}$$

que l'on peut, de la même manière que précédemment associer au produit de convolution

$$\mathbf{s} = \mathbf{E}^- * \mathbf{q}, \quad \text{avec } \mathbf{E}_j^- = \mathbf{E}_{-j}^+ = \begin{cases} 0, & j > 0 \\ (1 - \alpha)\alpha^{-j}, & j \leq 0 \end{cases}.$$

Au final, le filtre récursif symétrique est constitué de deux passes de directions opposées

$$\begin{aligned} q_i &= \alpha q_{i-1} + (1 - \alpha) p_i, & i = 1, \dots, n \\ s_i &= \alpha s_{i+1} + (1 - \alpha) q_i, & i = n, \dots, 1 \end{aligned}$$

où  $\mathbf{p}$  est le champ à filtrer,  $\mathbf{s}$  le champ filtré, et  $\mathbf{q}$  un champ intermédiaire. Il est équivalent au produit de convolution

$$\mathbf{s} = \mathbf{E}^- * \mathbf{q} = \mathbf{E}^- * (\mathbf{E}^+ * \mathbf{p}) = (\mathbf{E}^- * \mathbf{E}^+) * \mathbf{p} = \mathbf{F} * \mathbf{p}.$$

Lorsque le filtre récursif symétrique n'est appliqué qu'une seule fois, le noyau  $\mathbf{F}$  est équivalent à une fonction exponentielle (Lorenç, 1992). Lorsqu'il est appliqué  $M$  fois, le noyau est représenté par  $M$  produits de convolution de  $\mathbf{F}$  avec lui-même, et correspond alors à une fonction autorégressive d'ordre  $M$  (voir chapitre 4). Hayden et Purser (1995) notent qu'il est possible de montrer que plus  $M$  est grand et plus le noyau se rapproche de la forme d'une fonction gaussienne. Purser *et al.* (2003a) indiquent qu'une meilleure approximation de la gaussienne est obtenue avec une seule application d'un filtre récursif symétrique d'ordre plus élevé.

Le coefficient  $\alpha$  doit être ajusté pour représenter l'échelle de corrélation que l'on souhaite obtenir. Et pour être plus réaliste, Purser *et al.* (2003b) montrent qu'il est possible de prendre en compte des échelles variant géographiquement. De plus, ils soulignent qu'en exploitant une métrique (tenseur), le filtre récursif symétrique peut être appliqué à des grilles irrégulières.

Un point délicat de cette méthode concerne la gestion des frontières. En effet, les coefficients permettant de traiter ces conditions limites sont différents pour chaque application et leur calcul se complexifie à chaque nouvelle passe (Purser *et al.*, 2003a). Dans une expérience sur la Méditerranée, Dobricic et Pinardi (2008) propose d'ajouter des points d'océan fictifs à toutes les côtes. Lors de l'application du filtre, le traitement des frontières est alors reporté sur ces points fictifs qu'il suffit de retirer après le filtrage. Le même principe est utilisé par Liu *et al.* (2009) dans un modèle côtier.

Le filtre récursif à une dimension peut être généralisé à plusieurs dimensions en l'appliquant sur chaque axe (Wu *et al.*, 2002; Purser *et al.*, 2003a). Néanmoins, des effets parasites apparaissent lorsque le nombre d'application et l'ordre du filtre sont faibles. De plus, en séparant les axes, il n'est pas possible de modéliser des fonctions de corrélation réellement anisotropes, *i.e.* dont les déformations suivent un axe différent de la base. Pour y remédier, Purser *et al.* (2003b) (voir également Derber *et al.*, 2004) proposent de mettre en place des triades (2D) ou des hexades (3D) permettant de définir et d'utiliser des tenseurs en chaque point de grille. Le filtre récursif symétrique peut également être étendu à la sphère, mais dans ce cas des singularités apparaissent au niveau des pôles qui doivent être traitées de manière adéquate. Enfin, on notera que le filtre récursif symétrique étant construit à partir d'un "filtre direct" et d'un "filtre adjoint", il est donc possible d'en formuler l'opérateur sous une forme factorisée en "racines carrées".

Raymond et Garder (1991) montrent que filtre récursif et filtre implicite (tel que  $\mathbf{C}\mathbf{s} = \mathbf{p}$ , avec  $\mathbf{C}$  une matrice inversible,  $\mathbf{p}$  le champ à filtrer et  $\mathbf{s}$  le champ filtré) sont équivalents bien que de formulations différentes. Dans le paragraphe suivant, l'équation de diffusion est brièvement décrite comme une seconde alternative à l'évaluation des intégrales de convolution. Sa résolution par une factorisation de Cholesky pour le schéma implicite 1D, et donc son équivalence avec le filtre récursif, est donnée au chapitre 4. On indique également dans ce même chapitre, comment cette méthode permet de gérer les frontières de manière plus simple qu'avec le filtre récursif et la relation directe qui existe entre le coefficient de diffusion et l'échelle de corrélation. Au chapitre 5, l'opérateur 1D est étendu à plusieurs

dimensions de la même manière que pour le filtre récursif et s'avère soumis aux mêmes contraintes et difficultés.

### 3.6.3 L'équation de diffusion

Pour modéliser l'action des covariances horizontales de l'erreur d'ébauche sur un champ scalaire, Derber et Rosati (1989) proposent d'appliquer itérativement un filtre lisseur constitué d'un opérateur Laplacien. Egbert *et al.* (1994) décrivent une variante proche s'apercevant que la solution d'une équation de diffusion est en fait, à un facteur  $1/\lambda$  prêt, le produit de convolution d'une fonction de corrélation gaussienne avec la condition initiale. En effet, si on prend l'équation de diffusion 1D

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

où  $\eta = \eta(z, t)$  est un champ scalaire et  $\kappa$  un coefficient de diffusion, sa solution dans un domaine infini est donnée par (voir le détail au paragraphe 4.3.1)

$$\eta(z, t) = \frac{1}{\sqrt{4\pi\kappa t}} \int_{-\infty}^{\infty} e^{-(z-z')^2/4\kappa t} \eta_0(z') dz' = \frac{1}{\lambda} (g * \eta_0)(z). \quad (3.14)$$

Ainsi, la résolution de l'équation de diffusion et sa normalisation par le facteur  $\lambda$  représente un opérateur de corrélation dont le noyau est donné à  $t$  fixé par

$$g(r) = e^{-r^2/4\kappa t},$$

où  $r = |z - z'|$ . Ici,  $t$  est un paramètre ("pseudo-temps") qui permet, avec le coefficient de diffusion, de déterminer l'échelle de corrélation de la gaussienne  $L = \sqrt{2\kappa t}$ .

Weaver et Courtier (2001) généralisent la méthode et l'appliquent par un schéma explicite (Euler) sur une grille irrégulière, pour modéliser des fonctions de corrélation en dimension 2 (sur la sphère) ou 3. L'opérateur est accessible sous une forme factorisée, la "racine carrée" résultant d'une intégration sur la moitié du pseudo-temps. Ils montrent également que les frontières sont aisément prises en compte en appliquant des conditions limites de Neumann (dérivées nulles sur les bords) lors de la résolution du schéma. De plus, en définissant un coefficient de diffusion comme une fonction des coordonnées spatiales (équation de diffusion inhomogène), il est possible de faire varier géographiquement les échelles de corrélation. Enfin, ils illustrent la possibilité de modéliser des fonctions de corrélation anisotropes sur deux exemples. Le premier provient d'un simple étirement des coordonnées, tandis que l'autre est basé sur une rotation des coordonnées (géopotentielle vers isopycnales par exemple).

Cependant, la résolution de l'équation de diffusion par un schéma explicite ne permet de modéliser que des fonctions de corrélation gaussiennes. Pour élargir les possibilités, Weaver et Courtier (2001) considèrent l'équation de diffusion généralisée

$$\frac{\partial \eta}{\partial t} + \sum_{p=1}^P \kappa_p (-\nabla^2)^p \eta = 0,$$

où les coefficients  $\kappa_p$  sont positifs ou nuls. Cette formulation permet de représenter des fonctions avec oscillations, caractéristique intéressante pour certains champs. Mais l'ajout de Laplaciens d'ordre plus élevé, outre le coût plus élevé, se traduit au niveau spectral par une pente plus abrupte aux grands nombres d'onde, caractéristique moins désirable pour des champs géophysiques.

La modélisation des corrélations par une méthode explicite a été implémentée avec succès dans différents modèles océanographiques en global (Mogensen *et al.*, 2009) ou en régional (Weaver *et al.*, 2003; Di Lorenzo *et al.*, 2007; Powell *et al.*, 2008) ainsi qu'en chimie atmosphérique (Elbern *et al.*,

2010). Cependant, le nombre d'itérations est soumis à un critère de stabilité qui peut rendre l'algorithme coûteux pour des modèles à haute résolution ou lorsque les échelles de corrélation sont grandes. Weaver et Ricci (2004) suggèrent alors d'utiliser un schéma implicite. Inconditionnellement stable, il est cependant plus difficile à mettre en œuvre puisqu'il nécessite la résolution d'un système linéaire à chaque pas de temps. La définition d'un algorithme efficace pour ce schéma implicite est le sujet de cette thèse. Dans le chapitre 4, la formulation explicite en dimension 1 est détaillée avant d'examiner les aspects théoriques se rattachant à celle de l'implicite. Une extension en dimensions 2 et 3 est ensuite proposée au chapitre 5.

### 3.6.4 La normalisation

L'utilisation de méthodes alternatives telles que le filtre récursif symétrique ou l'équation de diffusion permettent d'évaluer les intégrales de convolution à un facteur prêt. Pour correspondre à des opérateurs de corrélation et non de covariance, il est donc nécessaire de les normaliser. Lorsque les fonctions de corrélation à modéliser sont homogènes et isotropes, cette normalisation est simple à calculer loin des frontières. Dans le cas de l'équation de diffusion 1D par exemple, l'équation (3.14) permet de déduire le facteur de normalisation

$$\lambda = \sqrt{4\pi\kappa t} = \sqrt{2\pi} L, \quad (3.15)$$

où  $L$  est l'échelle de corrélation. Néanmoins, lorsque cette échelle varie en fonction de sa position géographique, et/ou lorsque l'on s'approche des frontières, ce facteur n'est plus utilisable. Des algorithmes plus complexes tels que la méthode exacte ou la méthode par vecteurs aléatoires doivent alors être envisagés et sont décrit ci-après. Ces méthodes permettent de déterminer, non plus un facteur, mais une matrice diagonale des facteurs de normalisation  $\mathbf{\Lambda} = \text{diag}(\dots \Lambda_i^{-1} \dots)$ . Pour respecter la symétrie de l'opérateur de corrélation  $\mathbf{C}$ , la normalisation est appliquée de la façon suivante :

$$\mathbf{C} = \mathbf{\Lambda}^{1/2} \mathbf{F}^{1/2} \left( \mathbf{F}^{1/2} \right)^T \mathbf{\Lambda}^{1/2},$$

avec  $\mathbf{F}^{1/2} \left( \mathbf{F}^{1/2} \right)^T$  l'évaluation de l'intégrale de convolution par application du filtre récursif symétrique ou de l'équation de diffusion.

La méthode dite exacte consiste à évaluer la fonction de Green (ou réponse impulsionnelle) de l'opérateur en un point de grille, *i.e.* à appliquer l'opérateur sur une fonction de Dirac en ce point (toutes les valeurs du champ scalaire  $\mathbf{e}_i$  sont mises à 0 exceptée pour le point de grille traité où elle est initialisée à 1). La solution en ce point correspond au maximum (variance) de l'opérateur de covariance. Pour transformer ce dernier en opérateur de corrélation, il suffit donc de multiplier la solution par  $\mathbf{\Lambda}_{ii}^{-1}$ , l'inverse de cette variance, où

$$\Lambda_i^{-1} = \mathbf{e}_i^T \mathbf{F}^{1/2} \left( \mathbf{F}^{1/2} \right)^T \mathbf{e}_i.$$

---

#### Méthode de normalisation exacte

---

- Pour  $i = 1$  à  $N$ , le nombre de points de grille

$$\left| \begin{array}{l} \bullet \mathbf{e}_i = (0 \dots 0 \underbrace{1}_{i\text{-ème}} 0 \dots 0)^T \\ \bullet \mathbf{s}_i = \left( \mathbf{F}^{1/2} \right)^T \mathbf{e}_i \\ \bullet \Lambda_i^{-1} = \mathbf{s}_i^T \mathbf{s}_i \end{array} \right.$$

- Point de grille suivant
- 

En déterminant de cette manière les variances de chacun des points de la grille du modèle, on obtient en les inversant la matrice diagonale des facteurs de normalisation à appliquer. Cette méthode garantit

que le maximum de la fonction de corrélation modélisée est à 1. Néanmoins, suivant le nombre de points de grille et le coût de l'opérateur lui-même, cette méthode peut être extrêmement coûteuse.

Pour réduire le coût élevé de la méthode exacte, la méthode par vecteur aléatoire procède du même principe mais s'applique à un vecteur aléatoire plutôt qu'à un point de grille particulier (voir Weaver et Courtier, 2001, et les références citées). Le vecteur aléatoire  $\mathbf{v}$  utilisé est défini de telle manière que sa moyenne soit nulle ( $E[\mathbf{v}] = 0$ ) et que sa matrice de covariance soit l'identité ( $E[\mathbf{v}\mathbf{v}^T] = \mathbf{I}$ ). Lorsque l'opérateur est appliqué à un nombre  $P$  suffisant de ces vecteurs, la moyenne de l'ensemble converge vers le résultat de la méthode exacte.

---

Méthode de normalisation par vecteur aléatoire				
<ul style="list-style-type: none"> <li>• Pour <math>k = 1</math> à <math>P</math>, le nombre de membre           <div style="margin-left: 20px;"> <table style="border-left: 1px solid black; border-right: 1px solid black; border-collapse: collapse;"> <tr> <td style="padding: 0 10px;">•</td> <td>Initialisation du vecteur aléatoire <math>\mathbf{v}_k</math></td> </tr> <tr> <td style="padding: 0 10px;">•</td> <td><math>\mathbf{w}_k = \mathbf{F}^{1/2}\mathbf{v}_k</math></td> </tr> </table> </div> </li> <li>• Membre suivant</li> <li>• <math>diag(\mathbf{\Lambda}^{-1}) \approx diag\left(\frac{1}{P-1} \sum_{k=1}^P E[\mathbf{w}_k\mathbf{w}_k^T]\right)</math></li> </ul>	•	Initialisation du vecteur aléatoire $\mathbf{v}_k$	•	$\mathbf{w}_k = \mathbf{F}^{1/2}\mathbf{v}_k$
•	Initialisation du vecteur aléatoire $\mathbf{v}_k$			
•	$\mathbf{w}_k = \mathbf{F}^{1/2}\mathbf{v}_k$			

---

On a en effet

$$E[\mathbf{w}_k\mathbf{w}_k^T] = \mathbf{F}^{1/2} E[\mathbf{v}\mathbf{v}^T] (\mathbf{F}^{1/2})^T = \mathbf{F}^{1/2} (\mathbf{F}^{1/2})^T.$$

Le coût de la méthode est donné par le nombre de membres multiplié par le coût de l'application de  $\mathbf{F}^{1/2}$ . En supposant qu'il soit possible de générer une erreur réellement aléatoire, l'erreur d'approximation sur les coefficients de normalisation peut être évaluée à  $1/\sqrt{2P}$ , soit par exemple 10% pour  $P = 50$  membres. Mais en pratique, il n'est jamais possible de générer des nombres purement aléatoires, et pour une même erreur, l'approximation nécessite davantage de membres. Et si le coût de l'application de  $\mathbf{F}^{1/2}$  est important, la méthode peut s'avérer onéreuse.

Dans le chapitre 4, on montre que ces méthodes, outre leur coût potentiel, ne sont pas sans présenter quelques défauts, en particulier une déformation de l'échelle de corrélation près des frontières. Une généralisation intuitive (hors frontière) de l'équation (3.15) consiste à tenir directement compte de la variation de l'échelle de corrélation dans l'expression précédente (Pannekoucke et Massart, 2008, dans le cas d'une fonction gaussienne dans  $\mathbb{R}^2$ ) :

$$\tilde{\lambda}(z) \approx \sqrt{4\pi\overline{\kappa(z)}t} = \sqrt{2\pi} L(z).$$

Si cette formulation permet de réduire le coût de la normalisation à celui d'un simple calcul analytique, l'approximation se détériore lorsque cette variation est rapide. On peut alors utiliser la méthode proposée par Purser *et al.* (2003b) dans le cadre du filtre récursif, et qui consiste ici à filtrer le coefficient de diffusion  $\kappa(z)$  utilisé dans la formulation par la "racine carré" de l'opérateur de diffusion (voir la justification en Annexe E)

$$\bar{\lambda}(z) \approx \sqrt{4\pi\overline{\kappa(z)}t} = \sqrt{2\pi(\overline{L(z)})^2}, \quad \text{avec} \quad \overline{\kappa(z)} = \mathbf{F}^{1/2}\kappa(z).$$

Ces différentes méthodes sont comparées au chapitre 4.

Le calcul de la normalisation est un point important lorsque l'on choisit d'utiliser le filtre récursif ou l'équation de diffusion pour modéliser les fonctions de corrélation. En particulier, envisager que ces dernières dépendent de l'écoulement nécessite de recalculer cette normalisation à chaque cycle. Cela ne peut se concevoir que si ce calcul découle d'un algorithme efficace.



## Chapitre 4

# Opérateur de diffusion implicite 1D

L'estimation ou la modélisation de la matrice des covariances d'erreur d'ébauche est une tâche cruciale pour l'assimilation de données variationnelle. Le chapitre 3 en a décrit les différentes méthodes qui existent pour ce faire, et en particulier l'utilisation de l'équation de diffusion comme méthode permettant d'évaluer efficacement les intégrales de convolution requises par un opérateur de covariance. Cette technique, combinée à une procédure de normalisation, a été appliquée avec succès pour modéliser les corrélations spatiales univariées de la matrice  $\mathbf{B}$ . Néanmoins, résolue numériquement par une discrétisation temporelle explicite, elle peut s'avérer coûteuse en raison du critère de stabilité qui lui est associé. L'utilisation d'un tel schéma rend difficile son application dans des modèles à haute résolution et restreint la définition des échelles de corrélation à des contraintes plus numériques que physiques ou statistiques. Une première partie de cette thèse consiste donc à explorer en dimension 1, les possibilités offertes par un schéma implicite pour surmonter ces difficultés. On notera bien que cette étude n'est pas liée à l'erreur d'ébauche en particulier, et que la technique exposée ici peut également s'appliquer aux erreurs d'observation ou du modèle, qu'il s'agisse d'une structure spatiale ou temporelle. Dans ce chapitre, un bref résumé de cette étude est d'abord proposé en français. Les détails sont ensuite donnés en anglais sous la forme d'une publication (Mirouze et Weaver, 2010) parue au *Quarterly Journal of the Royal Meteorological Society*.

### Résumé en français

Lorsqu'elle est discrétisée en temps suivant un schéma implicite (Euler), pour une intégration entre  $t_{m-1}$  et  $t_m$ , l'équation de diffusion 1D

$$\frac{\partial \eta}{\partial t} - \kappa \frac{\partial^2 \eta}{\partial z^2} = 0, \quad (4.1)$$

avec  $\kappa$  le coefficient de diffusion et  $\eta = \eta(z, t)$  un champ scalaire quelconque, conduit à l'inverse de l'opérateur de diffusion implicite  $\mathcal{L}_i^{-1}$

$$\eta(z, t_{m-1}) = \left[ 1 - \kappa \Delta t \frac{\partial^2}{\partial z^2} \right] \eta(z, t_m) \equiv \mathcal{L}_i^{-1} [\eta(z, t_m)]. \quad (4.2)$$

En intégrant l'équation de diffusion (4.1) sur le pseudo-temps  $[0, T]$ , on obtient

$$\eta(z, 0) = \left[ 1 - \kappa \Delta t \frac{\partial^2}{\partial z^2} \right]^M \eta(z, T),$$

où  $\Delta t$  est le pas de temps et  $M$  le nombre d'itérations tel que  $T = M\Delta t$ . Ainsi, le champ  $\eta(z, T)$  est le résultat de  $M$  applications de l'opérateur implicite  $\mathcal{L}_i$  sur la condition initiale  $\eta(z, 0)$  :

$$\eta(z, T) \equiv \mathcal{L}_i^M [\eta(z, 0)]. \quad (4.3)$$

En normalisant l'équation (4.3) par le facteur

$$\lambda_M = \nu_M L, \quad \text{avec} \quad \nu_M = \frac{2^{2M-1}[(M-1)!]^2}{(2M-2)!}, \quad (4.4)$$

on obtient l'opérateur de corrélation

$$\mathcal{F}_M[\eta(z, 0)] = \lambda_M \mathcal{L}_i^M[\eta(z, 0)], \quad (4.5)$$

dont le noyau est défini par la fonction autorégressive (AR) d'ordre  $M$

$$f_M(r) = \sum_{j=0}^{M-1} \beta_j \left(\frac{r}{L}\right)^j e^{-\frac{r}{L}},$$

que l'on écrit souvent sous la forme

$$f_M(r) = \left[ 1 + \sum_{j=1}^{M-1} \beta_j \left(\frac{r}{L}\right)^j \right] e^{-\frac{r}{L}}, \quad (4.6)$$

lorsque  $M > 1$ .  $r = |z - z'|$  est la distance de corrélation,  $L = \sqrt{\kappa \Delta t}$  le paramètre d'échelle et  $\beta_j$  un coefficient tel que

$$\beta_j = \frac{2^j (M-1)! (2M-j-2)!}{j! (M-j-1)! (2M-2)!}.$$

Les fonctions AR sont un cas particulier de la famille de Matérn (voir Stein, 1999; Guttorp et Gneiting, 2006, ainsi que les paragraphes 4.3.3 et 4.A) et correspondent aux fonctions d'autocorrélation des processus de Markov ou processus autorégressifs d'où elles tirent leur nom (Gelb, 1974; Priestley, 1981; Gneiting, 1999). Si l'échelle de corrélation  $L_{AR}$  de la fonction (4.6) est déterminée pour  $M > 1$  suivant Daley (1991) par

$$L_{AR} = \sqrt{-\frac{f_M(r)}{d^2 f_M(r)/dr^2} \Big|_{r=0}},$$

le paramètre d'échelle  $L$  est alors donné par la relation

$$L_{AR} = \sqrt{2M-3} L.$$

Pour  $L_{AR}$  fixé, l'opérateur de corrélation implicite défini par l'équation (4.5) tend, lorsque  $M$  est grand, vers l'opérateur de corrélation explicite, tant au niveau du noyau de covariance (lorsque  $M$  est grand, la fonction AR tend vers la fonction gaussienne, voir Stein, 1999, p. 33) que de la normalisation utilisée ( $\lambda_\infty \rightarrow \sqrt{2\pi} L_{AR}$ ).

La classe de fonctions de corrélation modélisées par l'équation de diffusion implicite peut être élargie en utilisant une combinaison linéaire de fonctions AR, telle que

$$f_{LC}(r) = \frac{1}{\bar{\gamma}} \sum_{p=1}^P \gamma_p f_{M_p}(r),$$

où  $M_p$  est l'ordre de la  $p$ -ième fonction et  $\gamma_p$  son poids. Le facteur  $1/\bar{\gamma}$ , avec  $\bar{\gamma} = \sum_{p=1}^P \gamma_p$ , permet de maintenir le maximum à 1. Il est alors possible de prendre en compte différentes sources d'erreur (Martin *et al*, 2007), où même de créer des lobes négatifs, caractéristique utile pour certains champs géophysiques. L'opérateur de corrélation est alors donné par :

$$\mathcal{F}_{LC} = \frac{1}{\bar{\gamma}} \sum_{p=1}^P \gamma_p \lambda_{M_p} \mathcal{L}_{i, M_p}^M.$$

Dans un contexte réaliste, les échelles de corrélation ne sont cependant pas constantes, mais varient en fonction de leur position géographique. Pour prendre en compte ces variations, on utilise l'équation de diffusion inhomogène

$$\frac{\partial \eta}{\partial t}(z, t) - \frac{\partial}{\partial z} \left( \kappa(z) \frac{\partial}{\partial z} \right) \eta(z, t) = 0, \quad (4.7)$$

où  $\kappa(z)$  représente un coefficient de diffusion variant en fonction de  $z$  tel que  $L(z) = \sqrt{\kappa(z)\Delta t}$ . L'opérateur de corrélation de l'équation (4.5) est modifié de la façon suivante

$$\mathcal{F}_M[\eta(z, 0)] = \sqrt{\lambda_M(z)} \mathcal{L}_i^M \left[ \sqrt{\lambda_M(z)} \eta(z, 0) \right],$$

où  $\lambda_M(z)$ , le facteur de normalisation à appliquer en fonction de  $z$ , est divisé en racines carrées pour préserver la symétrie. Lorsque ces variations ne sont pas trop rapides, une bonne approximation du coefficient de normalisation est donnée par

$$\lambda_M(z) \approx \nu_M L(z) \equiv \tilde{\lambda}_M(z), \quad (4.8)$$

où  $\nu_M$  est défini par l'équation (4.4). La méthode de Purser *et al.* (2003b) (voir Annexe E) semble même donner de meilleurs résultats lorsque ces variations deviennent plus importantes. Quant au noyau de corrélation, on peut le supposer modifié comme suit :

$$f_M(z, z') = \frac{\sqrt{L(z)}\sqrt{L(z')}}{\sqrt{\overline{L^2(z, z')}}} \left[ 1 + \sum_{j=1}^{M-1} \beta_j \left( \frac{|z - z'|}{\sqrt{\overline{L^2(z, z')}}} \right)^j \right] \exp \left( \frac{-|z - z'|}{\sqrt{\overline{L^2(z, z')}}} \right), \quad (4.9)$$

où  $\overline{L^2(z, z')}$  représente la moyenne des échelles en  $z$  et  $z'$ .

D'un point de vue pratique, le système à résoudre à chaque itération, est donné par  $\mathbf{W}\boldsymbol{\eta}^{m-1} = \mathbf{W}\mathbf{A}\boldsymbol{\eta}^m$ , où  $\mathbf{A}$  est la forme matricielle (discrétisation spatiale) de l'opérateur  $\mathcal{L}_i^{-1}$  donné par l'équation (4.2).  $\mathbf{W}$  la matrice diagonale de la métrique de la grille, permet de symétriser la formulation (contrairement à  $\mathbf{A}$ , la matrice  $\mathbf{W}\mathbf{A}$  est symétrique et définie positive puisque à diagonale strictement dominante). Ce système peut donc être résolu grâce, par exemple, à la décomposition de Cholesky  $\mathbf{W}\mathbf{A} = \mathbf{G}\mathbf{G}^T$ , avec  $\mathbf{G}$  une matrice triangulaire inférieure.

Ainsi, la formulation implicite permet de résoudre le même type de problèmes que la formulation explicite, mais à moindre coût puisqu'elle n'est assujettie à aucun critère de stabilité. En libérant le nombre d'itérations  $M$  de toute contrainte, elle permet non seulement de modéliser une plus large classe de fonctions de corrélation que la simple fonction gaussienne, mais surtout d'augmenter la résolution du problème et les échelles de corrélation sans surcoût. On reviendra sur ces points dans le cadre du système d'assimilation variationnelle NEMOVAR au chapitre 5.2.

Néanmoins, les approximations données par les équations (4.8) et (4.9) ne sont plus valides lorsque l'échelle de corrélation varie trop rapidement ou à l'approche des frontières. En effet, la représentation de l'équation (4.7) par un schéma aux différences finies centré permet de prendre facilement en compte les frontières, que les conditions s'y appliquant soient de Neumann ou de Dirichlet. Mais ces conditions ont une répercussion sur la forme du noyau de corrélation. C'est pour palier ces problèmes que l'on utilise généralement un algorithme de normalisation (tel que la méthode exacte ou la méthode par vecteur aléatoire, voir paragraphe 3.6.4) plutôt que l'approximation  $\tilde{\lambda}_M$  donnée par l'équation (4.8). Néanmoins, cela est au prix d'un coût supplémentaire, et n'est pas sans effet sur l'échelle de corrélation près des frontières. Lorsque les paramètres ne varient pas dans le temps, la procédure de normalisation peut être effectuée au préalable et stockée dans un fichier. Néanmoins, une façon d'introduire une dépendance à l'écoulement consiste par exemple à réestimer les échelles de corrélation. Dans ce cas, la procédure de normalisation doit être appliquée chaque fois que ces paramètres sont modifiés. Pouvoir utiliser l'approximation définie par l'équation (4.8) s'avère alors attractif. Pour ce faire, il est tout

d'abord essentiel de gérer les frontières de manière transparente afin que ces dernières n'aient plus d'influence sur la fonction modélisée. En étudiant l'équation de diffusion à coefficient constant (4.1) par la méthode de séparation des variables, la fonction de Green (ou réponse impulsionnelle, *i.e.* la réponse à une fonction de Dirac en  $z_0$ ) pour des conditions aux frontières de Neumann, est donnée par

$$\eta_N(z, t) = \frac{1}{2l} + \sum_{n=1}^{\infty} \frac{1}{l} e^{-\left(\frac{n\pi}{l}\right)^2 \kappa t} \cos\left(\frac{n\pi}{l}(z-z_0)\right) + \frac{1}{2l} + \sum_{n=1}^{\infty} \frac{1}{l} e^{-\left(\frac{n\pi}{l}\right)^2 \kappa t} \cos\left(\frac{n\pi}{l}(z+z_0)\right). \quad (4.10)$$

Lorsque les conditions aux frontières de Dirichlet sont utilisées, la solution devient

$$\eta_D(z, t) = \frac{1}{2l} + \sum_{n=1}^{\infty} \frac{1}{l} e^{-\left(\frac{n\pi}{l}\right)^2 \kappa t} \cos\left(\frac{n\pi}{l}(z-z_0)\right) - \frac{1}{2l} - \sum_{n=1}^{\infty} \frac{1}{l} e^{-\left(\frac{n\pi}{l}\right)^2 \kappa t} \cos\left(\frac{n\pi}{l}(z+z_0)\right). \quad (4.11)$$

Or, en considérant le développement d'une fonction gaussienne en série de Fourier, on s'aperçoit que la solution (4.10) est en fait la somme d'une fonction gaussienne centrée en  $z_0$  et de la même fonction gaussienne centrée en  $-z_0$ . De même, la solution (4.11) est la différence entre les deux fonctions gaussiennes pré-citées. La corruption du noyau de covariance près des frontières est donc lié à la fonction centrée en  $-z_0$ . Pour s'en affranchir, on définit donc l'opérateur de diffusion comme la moyenne de deux problèmes de diffusion, l'un avec des conditions aux frontières de Neumann ( $\mathcal{L}_N$ ), et l'autre avec des conditions aux frontières de Dirichlet ( $\mathcal{L}_D$ ) :

$$\mathcal{L}^M = \frac{1}{2} (\mathcal{L}_N^M + \mathcal{L}_D^M).$$

Si cette formulation permet de simuler des frontières transparentes, elle est au prix de l'application d'une équation de diffusion supplémentaire. Cette méthode sera appliquée au chapitre 5.4 dans le cadre du système NEMOVAR.

# Representation of correlation functions in variational assimilation using an implicit diffusion operator

I. Mirouze<sup>a</sup> and A. T. Weaver<sup>b</sup>

<sup>a</sup> CERFACS and CNRS-UMR 5219, Toulouse, France

<sup>b</sup> CERFACS / SUC URA 1875, Toulouse, France

**Abstract :** Correlation models are required in data assimilation to characterize the error structures of variables defined on a numerical grid. Previous studies have shown that the diffusion equation can provide a flexible and computationally efficient framework for representing grid-point correlation functions for problems of large dimension such as those encountered in atmospheric or ocean variational data assimilation. In this paper, an implicit formulation of the diffusion-based correlation model is presented as an alternative to the traditional explicit formulation. The implicit formulation is analyzed in detail for the one-dimensional (1D) problem and shown to be closely related to the first-order recursive filter. Integrating a 1D implicit diffusion equation, with constant coefficient, over  $M$  steps is shown to be equivalent to convolving the initial condition with an  $M$ -th order auto-regressive (AR) function. Expressions for both the length scale of the AR function and the normalization factor required to generate unit-amplitude (correlation) functions are given in terms of  $M$  and the diffusion coefficient. For a fixed length scale, the Gaussian function, which is the only function that can be represented using an explicit formulation of the constant-coefficient diffusion equation, is the limiting case as  $M \rightarrow \infty$  of the AR functions generated by the implicit diffusion equation. Generalizations of the diffusion model are discussed to allow for different shapes in the correlation function and spatial variations in the length scale. An important consequence of employing spatially varying length scales is that the normalization factors are no longer constant. Approximate expressions for the normalization factors are evaluated in terms of their effectiveness to provide viable alternatives to estimates produced using expensive algorithms such as randomization. Boundary conditions can distort the correlation functions near the boundaries and significantly degrade the accuracy of the analytical expressions for the normalization factors. These problems can be avoided through a straightforward extension of the diffusion model that makes the boundaries effectively transparent, although the solution comes at the expense of an extra application of the diffusion equation. Extensions of the method to construct two- and three-dimensional correlation models are discussed.

## 4.1 Introduction

Variational methods are widely used in meteorology and oceanography to combine observations and a model background state into a more accurate estimate of the atmospheric or oceanic state (Bennett, 2002; Rabier, 2005). In variational data assimilation, a cost function based on the square of a weighted  $L_2$  norm is used to quantify the fit of the model to the assimilated information (observations and background state). The inverse of an estimate of the error covariance matrix of this information defines the weighting matrix of the  $L_2$  norm. The minimum of the square of the  $L_2$  norm corresponds to the conditional mode estimate of the state under the assumption that the information errors have a Gaussian distribution (Cohn, 1997).

The quality of the optimal estimate is largely determined by the quality of the estimate of the error covariance matrix of the assimilated information. Under the reasonable assumption that the observations are uncorrelated with the background state, the cost function can be split into the sum of two terms : an observation term measuring the fit to the observations, and a background term measuring the fit to the background state. The inverse of the error covariance matrix of the observations and that of the background state define the respective weighting matrix for each term. In weak-constraint formulations of four-dimensional variational assimilation, a third term is included to measure the degree to which the time-evolving model state should satisfy the prognostic model equations (Bennett,

2002; Trémolet, 2007b). The weighting matrix for this term is given by the inverse of an estimate of the so-called model-error covariance matrix.

For high dimensional problems such as those encountered in atmospheric and ocean data assimilation, the practical procedure for minimizing the cost function is iterative. In the widely-used incremental formulation of variational assimilation (Courtier *et al.*, 1994), an approximate solution of the underlying nonlinear (nonquadratic) minimization problem is obtained by solving a sequence of linearized (quadratic) minimization problems. Each quadratic problem can be solved efficiently using a conjugate gradient or Lanczos algorithm (Golub et van Loan, 1996). An important feature of the iterative approach to solving the variational assimilation problem is that the error covariance matrices, or the inverse of these matrices, never need to be known explicitly. On a given iteration of the minimization algorithm, only the result of the product of the covariance matrix (or its inverse) with a vector is required. This can be achieved using an operator-based covariance model. Given the size of the matrices involved (on the order of  $10^{12}$ – $10^{13}$  elements), this feature constitutes a major algorithmic advantage over alternative assimilation methods such as the Kalman filter that require an explicit representation of the covariance matrices. Apart from the restrictive case of uncorrelated errors, which may be justified for certain observation errors but is totally unrealistic for describing background or model errors, explicit representations of covariance matrices can only be achieved by approximating the full matrix by a low-rank matrix.

In practical applications, the true covariances of background, observation and model error are never known to great accuracy since there is insufficient information to compute them. Techniques for representing the covariances tend to be based on a mixture of physical insight, statistical calibration and computational expediency. For background errors, useful covariance information can be extracted from an adequately perturbed ensemble data assimilation system (Buehner, 2005; Berre *et al.*, 2006; Daget *et al.*, 2009). In variational assimilation, the challenge is then how to synthesize this information effectively in a covariance model that can be applied efficiently on each iteration of the minimization algorithm. Ideally, the covariance model should be designed to capture robust features in the available covariance estimates, such as multivariate relationships, geographical and wavenumber dependencies of the variances and correlation length scales, and anisotropic variations which often occur near regions of pronounced density or topographic gradients. Bannister (2008a,b) provides a comprehensive review of methods for representing background-error covariances in atmospheric variational assimilation.

Multivariate relationships between variables give rise to a cross-variable component in the covariances. In variational assimilation, it is common to specify the multivariate component of the covariances through a balance operator that captures known physical or statistical relationships between model variables (Derber et Bouttier, 1999; Weaver *et al.*, 2005). The inverse of the balance operator can be interpreted as a transformation to a new set of variables whose cross-variable covariances are much weaker than those of the original set of variables. The cross-covariances of the transformed variables are usually considered sufficiently small that they can be neglected. By doing so, the full multivariate covariance matrix can be decomposed into a sum of simpler univariate covariance matrices acting on each of the transformed variables. Each univariate covariance matrix can in turn be factored into a symmetric product of a correlation matrix and a diagonal matrix of standard deviations. The remaining challenge is how to represent adequately and evaluate efficiently the product of a generally full-rank, non-diagonal correlation matrix with an arbitrary vector.

The continuous analogue of a correlation matrix–vector product is an integral operator whose kernel is a symmetric and positive definite correlation function. When evaluated at discrete points, the correlation function defines a (full-rank) correlation matrix although this definition is of limited practical interest for large matrix problems. The evaluation of integral correlation operators, especially those associated with background- or model-error correlations, are among the most costly steps in the variational assimilation algorithm.

Several methods have been developed in meteorology and oceanography for representing univariate correlation operators (Bannister, 2008b). Methods based on spectral or wavelet transforms have been developed extensively for atmospheric data assimilation (Deckmyn et Berre, 2005; Bannister, 2007; Pannekoucke *et al.*, 2007). Correlation operators based on physical-space models have also been developed. Physical-space formulations are more convenient than spectral or wavelet formulations in complex-boundary domains such as those encountered in ocean data assimilation. Gaspari et Cohn (1999) and Gneiting (2002) derive families of parameterized homogeneous and isotropic correlation functions with the important property of compact support so that they can be integrated efficiently in grid-point space. Gaspari *et al.* (2006) describe extensions to these functions to account for spatially varying length scales. Correlation integrals can also be represented efficiently using grid-point filters. The recursive filter (Purser *et al.*, 2003a,b, hereafter P03a and P03b) and the diffusion equation (Weaver et Courtier, 2001, hereafter WC01) fall into this class of correlation model.

This paper describes recent developments to the diffusion approach to correlation modelling which build on the results of WC01. The focus is on the one-dimensional (1D) diffusion problem which is analyzed in detail for the special case of an implicit “time”-stepping scheme. New insight is given into the problem of determining the normalization factors required to generate correlation functions with correct (unit) amplitude and to the related problem of specifying appropriate conditions at solid boundaries. 1D correlation operators are needed for modelling temporal correlations of model error (Bennett *et al.*, 1996, 1997, 1998; Ngodock, 2005; Trémolet, 2006). Furthermore, 1D correlation operators can be used as the basis for constructing two-dimensional (2D) and three-dimensional (3D) spatial correlation operators (Wu *et al.*, 2002; Chua et Bennett, 2001; Martin *et al.*, 2007; Dobricic et Pinardi, 2008, P03a,b). The results from this paper are therefore relevant for applications in higher dimensions.

The structure of the paper is as follows. Section 4.2 provides a review of key properties from covariance and correlation theory which are necessary for understanding the developments in the rest of the paper. The relationship between the class of auto-regressive (AR) correlation operators and the diffusion equation is established in Section 4.3 by exploiting results in Appendix 4.A that relate AR functions to the much larger class of correlation functions known as the Matérn family. Section 4.4 describes generalizations of the diffusion algorithm to account for spatially varying length scales and boundary conditions. An important modification of the correlation model is proposed in the presence of boundaries, following a careful analysis of the solution of the diffusion equation subject to different boundary conditions. The mathematical details are given in Appendix 4.B. Numerical aspects of the algorithm including its relationship with the first-order recursive filter are presented in Section 4.5. A summary and discussion are given in Section 4.6.

## 4.2 Background theory

Before discussing the relationship between correlation operators and the diffusion equation, it is helpful to review some basic definitions and properties of covariance (correlation) functions and covariance (correlation) operators. For the purpose of this study, we focus on covariance (correlation) theory related to the one-dimensional Euclidean space  $\mathbb{R}$ . A more comprehensive review that includes extensions of the theory to higher dimensions and to the sphere can be found in Gaspari et Cohn (1999).

### 4.2.1 Covariance functions and covariance operators

The covariance function  $b(z, z')$  of a random field  $X$  is the statistical relationship between random variables  $X(z)$  and  $X(z')$  at arbitrary points  $z$  and  $z'$  in  $\mathbb{R}$  defined by

$$b(z, z') = E [(X(z) - E[X(z)])(X(z') - E[X(z')])] \quad (4.12)$$

where  $E[\cdot]$  denotes mathematical expectation of the random variable  $X(z)$ . In estimation theory (Cohn, 1997),  $X(z)$  can be interpreted as the difference between the background value of a model variable and the true value of that variable at a particular point  $z$ . From Equation (4.12), the covariance function is clearly symmetric,  $b(z, z') = b(z', z)$ .

Let  $L^2(\mathbb{R})$  denote the space of square integrable functions on  $\mathbb{R}$ . The covariance operator  $\mathcal{B}$  associated with  $b(z, z') \in L^2(\mathbb{R} \times \mathbb{R})$  is the integral operator mapping the functions  $\psi(z) \in L^2(\mathbb{R})$  into  $\varphi(z) \in L^2(\mathbb{R})$  :

$$\varphi(z) = \mathcal{B}[\psi(z)] \equiv \int_{-\infty}^{\infty} b(z, z') \psi(z') dz'. \quad (4.13)$$

The covariance operator  $\mathcal{B}$  is symmetric and positive semi-definite ; i.e., for (assumed real-valued) functions  $\psi_1$  and  $\psi_2$  in  $L^2(\mathbb{R})$ ,

$$\langle \psi_1(z), \mathcal{B}[\psi_2(z)] \rangle = \langle \mathcal{B}[\psi_1(z)], \psi_2(z) \rangle, \quad (4.14)$$

$$\text{and } \langle \psi_1(z), \mathcal{B}[\psi_1(z)] \rangle \geq 0, \quad (4.15)$$

where the inner product  $\langle \cdot, \cdot \rangle$  is given by

$$\langle \psi(z), \varphi(z) \rangle = \int_{-\infty}^{\infty} \psi(z) \varphi(z) dz.$$

Property (4.14) follows directly from the symmetry of the kernel  $b(z, z')$ , while property (4.15) follows from the linearity of the expectation operator (Gaspari et Cohn, 1999, p. 728). Integral equations of the form (4.13) must be evaluated numerically in data assimilation algorithms when multiplication of a vector by the background-error covariance matrix is required. These computations are among the most costly in the assimilation algorithm.

The inverse of the covariance operator (4.13),  $\mathcal{B}^{-1}$ , is defined as the integral operator

$$\psi(z) = \mathcal{B}^{-1}[\varphi(z)] \equiv \int_{-\infty}^{\infty} b^{-1}(z, z') \varphi(z') dz' \quad (4.16)$$

where the kernel  $b^{-1}(z, z') \in L^2(\mathbb{R} \times \mathbb{R})$  denotes the inverse of the covariance function. Substituting Equation (4.16) in (4.13) gives the formal relation

$$\int_{-\infty}^{\infty} b(z, z') b^{-1}(z', z'') dz' = \delta(z - z'') \quad (4.17)$$

where  $\delta(z)$  is the Dirac delta function :

$$\psi(z) = \int_{-\infty}^{\infty} \delta(z - z') \psi(z') dz'.$$

In general, the inverse of a covariance function is not an ordinary function, but a generalized function involving terms like the Dirac delta function and its derivatives (Jones, 1982). This is the case for the inverse correlation functions derived in Section 4.3.2. In analogy with Equations (4.14) and (4.15), the inverse covariance operator  $\mathcal{B}^{-1}$  satisfies

$$\langle \varphi_1(z), \mathcal{B}^{-1}[\varphi_2(z)] \rangle = \langle \mathcal{B}^{-1}[\varphi_1(z)], \varphi_2(z) \rangle, \quad (4.18)$$

$$\text{and } \langle \varphi_1(z), \mathcal{B}^{-1}[\varphi_1(z)] \rangle \geq 0,$$

for functions  $\varphi_1$  and  $\varphi_2$  in  $L^2(\mathbb{R})$ . Equation (4.18) defines an inner product in which the  $\mathcal{B}^{-1}$  operator is the weighting metric. In variational assimilation, the norm associated with this inner product is used to define the background term of the cost function.

### 4.2.2 Correlation functions and correlation operators

In this study we are concerned with the construction of (unit-amplitude) correlation functions  $c(z, z')$  and their associated correlation operators  $\mathcal{C}$ . To isolate the correlation function, we introduce the symmetry-preserving factorization of the covariance function,

$$b(z, z') = \sigma(z) \sigma(z') c(z, z') \quad (4.19)$$

where  $\sigma(z) = \sqrt{b(z, z)}$  is the standard deviation and  $c(z, z')$  satisfies  $c(z, z) = 1$  and  $c(z, z') = c(z', z)$ . The associated correlation operator  $\mathcal{C}$  is obtained by substituting Equation (4.19) in Equation (4.13) and introducing new variables  $\bar{\psi}(z) = \sigma(z)\psi(z)$  and  $\bar{\varphi}(z) = \varphi(z)/\sigma(z)$  :

$$\bar{\varphi}(z) = \mathcal{C}[\bar{\psi}(z)] \equiv \int_{-\infty}^{\infty} c(z, z') \bar{\psi}(z') dz'$$

The covariance operator can be recovered simply by post-multiplying the correlation operator by  $\sigma(z)$ . The inverse of the covariance function can be written as

$$b^{-1}(z, z') = \frac{c^{-1}(z, z')}{\sigma(z) \sigma(z')}$$

where  $c^{-1}(z, z')$  is the inverse of the correlation function. The inverse of the correlation operator,  $\mathcal{C}^{-1}$ , is then

$$\bar{\psi}(z) = \mathcal{C}^{-1}[\bar{\varphi}(z)] \equiv \int_{-\infty}^{\infty} c^{-1}(z, z') \bar{\varphi}(z') dz'$$

where from Equation (4.17)

$$\int_{-\infty}^{\infty} c(z, z') c^{-1}(z', z'') dz' = \delta(z - z''). \quad (4.20)$$

For clarity of notation, the overbar is hereafter dropped from the normalized variables  $\bar{\psi}(z)$  and  $\bar{\varphi}(z)$ .

An important class of correlation functions is the class of *homogeneous* and *isotropic* correlation functions. Homogeneous and isotropic correlation functions on  $\mathbb{R} \times \mathbb{R}$  can be written as  $c(z, z') = f(z - z') = f(r)$  where  $r = |z - z'|$  is Euclidean distance (Gaspari et Cohn, 1999, Theorem 2.12). The function  $f(r)$  is said to represent the correlation function  $c(z, z')$ . A correlation operator  $\mathcal{C}$  represented by a homogeneous and isotropic correlation function  $f(r)$  is a *convolution* operator :

$$\varphi(z) = \int_{-\infty}^{\infty} f(z - z') \psi(z') dz' \equiv (f * \psi)(z). \quad (4.21)$$

Likewise, for the inverse correlation operator  $\mathcal{C}^{-1}$ ,

$$\psi(z) = \int_{-\infty}^{\infty} f^{-1}(z - z') \varphi(z') dz' \equiv (f^{-1} * \varphi)(z) \quad (4.22)$$

where  $f^{-1}(r)$  denotes the inverse of the correlation function.

The inverse of a homogeneous and isotropic correlation function can be derived using the Fourier transform (FT), assuming now that  $\varphi(z)$  and  $\psi(z)$  are in  $L^1(\mathbb{R})$ , the space of absolutely integrable functions, as well as  $L^2(\mathbb{R})$  (see Gaspari et Cohn, 1999, p. 733). The FT ( $\mathfrak{F}$ ) of a function  $w(z) \in L^1(\mathbb{R})$  is defined as

$$\hat{w}(\hat{z}) = \mathfrak{F}[w(z)] \equiv \int_{-\infty}^{\infty} w(z) e^{-i\hat{z}z} dz \quad (4.23)$$

where  $i$  is a complex number ( $i^2 = -1$ ). The function  $w(z)$  can be recovered from the inverse FT ( $\mathfrak{F}^{-1}$ ) of  $\widehat{w}(\hat{z}) \in L^1(\mathbb{R})$  :

$$w(z) = \mathfrak{F}^{-1}[\widehat{w}(\hat{z})] \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{w}(\hat{z}) e^{iz\hat{z}} d\hat{z}. \quad (4.24)$$

Taking the FT of Equation (4.21) and applying the Convolution Theorem gives

$$\widehat{\varphi}(\hat{z}) = \widehat{f}(\hat{z}) \widehat{\psi}(\hat{z})$$

and thus

$$\widehat{\psi}(\hat{z}) = \frac{\widehat{\varphi}(\hat{z})}{\widehat{f}(\hat{z})} = \widehat{d}(\hat{z}) \widehat{\varphi}(\hat{z}) \quad (4.25)$$

where

$$\widehat{d}(\hat{z}) = \frac{1}{\widehat{f}(\hat{z})}. \quad (4.26)$$

Applying the inverse FT to Equation (4.25) yields

$$\psi(z) = (d * \varphi)(z) = \int_{-\infty}^{\infty} d(z-z') \varphi(z') dz', \quad (4.27)$$

from which we can identify

$$f^{-1}(r) = d(r) \quad (4.28)$$

by comparing with Equation (4.22). This classical result will be applied in Section 4.3.2 to relate the class of AR functions to the solutions of an implicitly-solved diffusion equation.

## 4.3 Correlation operators and the diffusion equation

The diffusion equation is presented here as a mathematical tool for simulating the action of a correlation operator ; i.e., for evaluating convolution integrals of the form Equation (4.21). The basic algorithm is conceptually simple and provides a useful ‘physical’ interpretation to the spatial and/or temporal smoothing property of a correlation operator. In this section we focus on the constant-coefficient diffusion equation and issues surrounding the choice of integration scheme. Some generalizations are also presented at the end of this section.

### 4.3.1 Explicit and implicit formulations of the 1D diffusion equation

Consider the 1D diffusion equation for the scalar function  $\eta = \eta(z, t)$ ,

$$\frac{\partial \eta}{\partial t} - \kappa \frac{\partial^2 \eta}{\partial z^2} = 0 \quad (4.29)$$

where  $\kappa$  is a positive (diffusion) constant. Here  $t$  is to be interpreted as a pseudo-time coordinate which will later be related to the parameters of the correlation model, while the coordinate  $z$  is analogous to the domain variable  $z$  in the covariance integrals in Section 4.2<sup>62</sup>. We consider the general solution to Equation (4.29) subject to the ‘initial’ condition  $\eta(z, 0) = \eta_0(z)$ . To begin with, we assume that the domain is infinite and that the initial condition is bounded at infinity,  $\int_{-\infty}^{\infty} |\eta_0(z')| dz' < \infty$ . In a finite domain, Equation (4.29) must be supplied with boundary conditions. The choice of boundary conditions influences the correlation functions represented by Equation (4.29) and will be discussed in Section 4.4.

The solution of Equation (4.29) is easily found using the FT. Applying the FT to Equation (4.29) yields the first-order differential equation

$$\frac{\partial \widehat{\eta}}{\partial t} + \kappa \hat{z}^2 \widehat{\eta} = 0,$$

<sup>62</sup>In the case of a temporal correlation model,  $z$  should be interpreted as a (real) time coordinate.

whose solution is

$$\widehat{\eta}(\hat{z}, t) = \widehat{\eta}_0(\hat{z}) e^{-\kappa \hat{z}^2 t}, \quad (4.30)$$

where  $\widehat{\eta}_0(\hat{z}) = \widehat{\eta}(\hat{z}, 0)$  is the FT of the initial condition. Now taking the inverse FT (Equation (4.24)) of Equation (4.30) and applying the Convolution Theorem, the solution  $\eta(z, t)$  can be written as

$$\eta(z, t) = \frac{1}{\sqrt{4\pi\kappa t}} \int_{-\infty}^{\infty} e^{-(z-z')^2/4\kappa t} \eta_0(z') dz', \quad (4.31)$$

where we have used the result that

$$h(z) = \mathfrak{F}^{-1}[e^{-\kappa \hat{z}^2 t}] = \frac{1}{\sqrt{4\pi\kappa t}} e^{-z^2/4\kappa t}. \quad (4.32)$$

In Equation (4.31) we recognize the convolution of the initial condition  $\eta_0(z)$  with the Gaussian function Equation (4.32). As the Gaussian function is positive definite, Equation (4.31) defines a covariance operator :

$$\mathcal{H}[\eta_0(z)] \equiv \eta(z, t).$$

Comparing with the notation used in Section 4.2 (cf. Equation (4.21)), we can identify  $\eta_0(z)$  with  $\psi(z)$  and  $\eta(z, t)$  with  $\varphi(z)$ , where  $t$  is a parameter controlling the length scale of the covariance function. Using the conventional definition of the covariance length scale (Daley, 1991, p. 110), we have for the Gaussian that<sup>63</sup>

$$L_g = \sqrt{-\frac{h(0)}{h^{(2)}(0)}} = \sqrt{2\kappa t} \quad (4.33)$$

where  $h^{(2)}(z) = d^2h/dz^2$ . Equation (4.31) can be transformed into a Gaussian *correlation* operator  $\mathcal{G}$  by post-multiplying  $\eta(z, t)$  by the normalization factor

$$\lambda_g = \sqrt{4\pi\kappa t} = \sqrt{2\pi} L_g. \quad (4.34)$$

The correlation operator can thus be represented as

$$\mathcal{G}[\eta_0(z)] \equiv \lambda_g \mathcal{H}[\eta_0(z)],$$

where the kernel of  $\mathcal{G}$  is the unit-amplitude Gaussian function

$$g(r) = e^{-r^2/2L_g^2}. \quad (4.35)$$

The key idea is that, on a numerical grid, we can evaluate the action of a Gaussian correlation operator on a scalar field  $\eta_0(z)$  by ‘time’-stepping a discretized version of the differential equation (4.29) over a pseudo-time interval  $0 \leq t \leq T = L_g^2/2\kappa$  from an initial condition  $\eta(z, 0) = \eta_0(z)$ , and normalizing the result as described above. This is a computationally efficient way of evaluating the convolution integral in Equation (4.31) and is the essence of the diffusion algorithm for defining correlation operators described in WC01, and in the earlier works of Derber et Rosati (1989) and Egbert *et al.* (1994).

WC01 considered a semi-discrete version of Equation (4.29) in which the ‘temporal’ derivative is discretized using an *explicit* forward scheme. Other discretization schemes are of course possible. As shown below, the choice of scheme is important for determining the class of correlation functions that

---

<sup>63</sup>(Gelb, 1974, p. 43) and (Bennett, 2002, p. 54) prefer an alternative definition of length scale based on the distance required to reduce the amplitude of the correlation function to  $1/e$ . For the Gaussian, the  $1/e$  scale is a factor  $\sqrt{2}$  times larger than  $L_g$  defined by Equation (4.33).

can be represented by the diffusion equation. Let  $t_m = m\Delta t$  where  $m$  is a positive integer and  $\Delta t$  a pseudo-time step. At any step  $m$ , the semi-discrete explicit version of (4.29) reads

$$\begin{aligned}\eta(z, t_m) &= \eta(z, t_{m-1}) + \kappa\Delta t \frac{\partial^2}{\partial z^2} \eta(z, t_{m-1}) \\ &= \left(1 + \kappa\Delta t \frac{\partial^2}{\partial z^2}\right) \eta(z, t_{m-1}) \\ &\equiv \mathcal{L}_e[\eta(z, t_{m-1})],\end{aligned}\tag{4.36}$$

where

$$\mathcal{L}_e \equiv 1 + \kappa\Delta t \frac{\partial^2}{\partial z^2}$$

is the one-step explicit diffusion operator. Therefore, to apply a Gaussian correlation operator with length scale  $L_g$ , we can set

$$\kappa\Delta t = \frac{L_g^2}{2M}\tag{4.37}$$

and perform  $M$  steps of Equation (4.36), from  $t_0 = 0$  to  $t_M = M\Delta t = T$  :

$$\eta(z, T) = \mathcal{L}_e^M[\eta_0(z)]$$

where  $\mathcal{L}_e^M = \mathcal{L}_e \cdots \mathcal{L}_e$  ( $M$  times). As the fundamental parameter controlling the length scale is the *product* of the diffusion constant and pseudo-time step, we can arbitrarily choose to set either one of these two parameters to one. It is useful to retain both parameters, however, to avoid confusion about the physical dimensions of  $\kappa\Delta t$  ( $(\text{length})^2/\text{time} \times \text{time} = (\text{length})^2$ ).

The explicit scheme is appealing since it is easy to implement. The drawback is that it is only conditionally stable : for a fixed  $L_g$  and assuming a uniform spatial grid with resolution  $\Delta z$ , the stability condition requires that  $M \geq (L_g/\Delta z)^2$  (e.g., see Haberman, 1987, p. 492). For some applications, the stability criterion can lead to a significant increase in the cost of the algorithm. One possibility to reduce the cost is to replace the explicit scheme with an implicit scheme, which is *unconditionally* stable. In the implicit formulation, the second-order derivative in Equation (4.29) is evaluated at time  $t_m$  instead of  $t_{m-1}$  :

$$\eta(z, t_m) = \eta(z, t_{m-1}) + \kappa\Delta t \frac{\partial^2}{\partial z^2} \eta(z, t_m),$$

which can be written in reverse time order as

$$\begin{aligned}\eta(z, t_{m-1}) &= \eta(z, t_m) - \kappa\Delta t \frac{\partial^2}{\partial z^2} \eta(z, t_m) \\ &= \left(1 - \kappa\Delta t \frac{\partial^2}{\partial z^2}\right) \eta(z, t_m) \\ &\equiv \mathcal{L}_i^{-1}[\eta(z, t_m)],\end{aligned}\tag{4.38}$$

where

$$\mathcal{L}_i^{-1} \equiv 1 - \kappa\Delta t \frac{\partial^2}{\partial z^2}\tag{4.39}$$

is the *inverse* of a one-step implicit diffusion operator.

To iterate the implicit diffusion equation, Equation (4.38) must be inverted on each step. Numerically, this can be done efficiently using, for example, an LU- or Cholesky-decomposition method applied to a spatially discretized matrix representation of the  $\mathcal{L}_i^{-1}$  operator (see Section 4.5). The  $M$ -step implicit diffusion operator can be represented as

$$\eta(z, T) = \mathcal{L}_i^M[\eta_0(z)].\tag{4.40}$$

In general, to achieve an approximate Gaussian response from the diffusion equation, the number of implicit iterations can be chosen to be much less than the number of iterations required by the stability criterion of an explicit scheme. For small values of  $M$ , however, numerical errors will result in a distortion of the Gaussian shape. In the next section, we quantify the actual functional form of the correlation functions generated by an implicit diffusion operator with an arbitrary value of  $M$ . In particular, we show that the implicit scheme can be used to represent the family of AR functions, with the order of the AR function being determined by  $M$ . In this respect,  $M$  can be viewed as a free parameter of the implicit diffusion model which can be used to provide some flexibility in the shape of the correlation function.

### 4.3.2 The kernel of the implicit diffusion operator

In this section we show that the integral (correlation) operator ( $\mathcal{F}_M$ ) whose kernel is given by the  $M$ -th order AR function

$$f_M(z-z') = \left[ 1 + \sum_{j=1}^{M-1} \beta_j \left( \frac{|z-z'|}{L} \right)^j \right] e^{-|z-z'|/L} \quad (4.41)$$

where  $L$  is a scale parameter and

$$\beta_j = \frac{2^j (M-1)! (2M-j-2)!}{j! (M-j-1)! (2M-2)!}, \quad (4.42)$$

is equivalent to the application of an appropriately normalized  $M$ -step implicit diffusion operator. Specific examples of the AR functions include the exponential or first-order AR (FOAR) function ( $M=1$ ), and the second- and third-order AR (SOAR and TOAR) functions ( $M=2$  and  $M=3$ ) as discussed, for example, in Gaspari et Cohn (1999) and Gaspari *et al.* (2006). The AR functions are in fact nested in a more general class of correlation functions known as the *Matérn* family (Stein, 1999; Guttorp et Gneiting, 2006). Indeed, many of the results in this section can be obtained as special cases of more general results in  $\mathbb{R}^n$  for the Matérn family (see Appendix 4.A).

The relationship between the AR correlation operator and the implicit diffusion operator is easily deduced from the expression for the *inverse* of the AR correlation operator. As outlined in Section 4.2, the *inverse* of a homogeneous and isotropic correlation function, which defines the kernel of the inverse of a correlation operator, is defined by the inverse FT of the reciprocal of the FT of the correlation function itself. We now apply this result to derive the inverse correlation function and associated inverse correlation operator for the class of (homogeneous and isotropic) AR correlation functions  $f_M(z-z')$ .

In Appendix 4.A it is shown that the FT of  $f_M(z)$  is

$$\widehat{f}_M(\hat{z}) = \frac{\lambda_M}{[1 + (L\hat{z})^2]^M} \quad (4.43)$$

where

$$\lambda_M = v_M L \quad (4.44)$$

and  $v_M$  is a nondimensional coefficient given by

$$v_M = \frac{2^{2M-1} [(M-1)!]^2}{(2M-2)!}. \quad (4.45)$$

From Equations (4.26) and (4.43), we have

$$\widehat{d}_M(\hat{z}) = \frac{1}{\widehat{f}_M(\hat{z})} = \frac{[1 + (L\hat{z})^2]^M}{\lambda_M},$$

which we can expand as a binomial series

$$\begin{aligned}\widehat{d}_M(\hat{z}) &= \frac{1}{\lambda_M} \sum_{j=0}^M \gamma_j (L\hat{z})^{2j} \\ &= \frac{1}{\lambda_M} (1 - \gamma_1 L^2 (i\hat{z})^2 + \gamma_2 L^4 (i\hat{z})^4 - \dots)\end{aligned}\quad (4.46)$$

where the binomial coefficients

$$\gamma_j = \frac{M!}{j!(M-j)!}.\quad (4.47)$$

Taking the inverse FT of Equation (4.46) and applying the following property from the theory of generalized functions (Jones, 1982, p. 74),

$$\mathfrak{F}^{-1} [(i\hat{z})^n] = \frac{1}{2\pi} \int_{-\infty}^{\infty} (i\hat{z})^n e^{i\hat{z}z} d\hat{z} = \delta^{(n)}(z)$$

where  $\delta^{(0)}(z) = \delta(z)$  and  $\delta^{(n)}(z)$ ,  $n = 1, \dots$ , denotes the  $n$ -th order generalized derivative of the Dirac delta function, yields

$$\begin{aligned}d_M(z) &= \frac{1}{\lambda_M} (\delta(z) - \gamma_1 L^2 \delta^{(2)}(z) + \gamma_2 L^4 \delta^{(4)}(z) - \dots) \\ &= \frac{1}{\lambda_M} \sum_{j=0}^M (-1)^j \gamma_j L^{2j} \delta^{(2j)}(z) = f_M^{-1}(z),\end{aligned}\quad (4.48)$$

where the last relation follows from Equation (4.28). To derive the associated inverse correlation operator, we substitute Equation (4.48) in Equation (4.27),

$$\begin{aligned}\psi(z) &= \int_{-\infty}^{\infty} f_M^{-1}(z-z') \varphi(z') dz' \\ &= \frac{1}{\lambda_M} \sum_{j=0}^M (-1)^j \gamma_j L^{2j} \int_{-\infty}^{\infty} \delta^{(2j)}(z-z') \varphi(z') dz' \\ &= \frac{1}{\lambda_M} \sum_{j=0}^M (-1)^j \gamma_j L^{2j} \varphi^{(2j)}(z)\end{aligned}\quad (4.49)$$

where we have used the property (Jones, 1982, p. 178)

$$\int_{-\infty}^{\infty} \delta^{(2j)}(z-z') \varphi(z') dz' = \varphi^{(2j)}(z).$$

Alternatively, we can rewrite Equation (4.49) as

$$\begin{aligned}\psi(z) &= \frac{1}{\lambda_M} \sum_{j=0}^M \gamma_j \left( -L^2 \frac{\partial^2}{\partial z^2} \right)^j \varphi(z) \\ &= \frac{1}{\lambda_M} \left[ 1 - L^2 \frac{\partial^2}{\partial z^2} \right]^M \varphi(z) \equiv \mathcal{F}_M^{-1} [\varphi(z)].\end{aligned}\quad (4.50)$$

Note that the inverse correlation operator has physical dimensions of inverse length, consistent with the property that an inverse correlation operator is a mapping from the primal space into its dual (Tarantola, 2005, Section 3.1.3).

Comparing Equations (4.39) and (4.50), we see that an  $M$ -step application of the inverse implicit diffusion operator can be matched to the inverse correlation operator  $\mathcal{F}_M^{-1}$  by identifying

$$L = \sqrt{\kappa \Delta t} \quad (4.51)$$

and by post-multiplying the inverse diffusion operator by the factor  $1/\lambda_M$  :

$$\mathcal{F}_M^{-1} [\varphi(z)] = \lambda_M^{-1} (\mathcal{L}_i^{-1})^M [\varphi(z)]. \quad (4.52)$$

Referring to the implicit diffusion operator in Equation (4.40), the correlation operator  $\mathcal{F}_M$  is then

$$\mathcal{F}_M [\eta_0(z)] = \lambda_M \mathcal{L}_i^M [\eta_0(z)]. \quad (4.53)$$

In other words, integrating the implicit diffusion equation for  $M$  steps with  $\kappa \Delta t$  defined according to Equation (4.51), and multiplying the solution by the normalization factor  $\lambda_M$ , is equivalent to applying a correlation operator whose kernel is an  $M$ -th order AR function with scale parameter  $L$ .

In his classical paper, Whittle (1963) derives the family of Matérn functions by computing the correlation function of a random variable obtained as the general solution of a fractional differential equation of multi-dimensional Laplacian type forced by stochastic white noise (Guttorp et Gneiting, 2006; Kelbert *et al.*, 2005, see also) who consider more general diffusion-like stochastic differential operators). In 1D, Whittle's fractional differential operator corresponds to the "square root" of the inverse implicit diffusion operator  $(\mathcal{L}_i^{-1})^{M/2}$  with  $M$  in this case allowed to take on both integer and non-integer values. For non-integer values of  $M$ , the inverse operator  $(\mathcal{L}_i^{-1})^M$  is represented as an infinite binomial series. In numerical applications, the infinite series must be truncated at some finite order although, as remarked by Oliver (1998), there is no guarantee that this results in a useful approximation of the inverse operator. In the implicit diffusion approach, the binomial series is finite as  $M$  is a positive integer. Furthermore, it is convenient to restrict  $M$  to be even to simplify the derivation of a square-root operator which may be required for preconditioning or randomization applications (WC01). The resulting class of even-order AR functions is less general than the full 1D Matérn family but still possesses reasonable flexibility as discussed in the next section.

### 4.3.3 Characteristics of the AR functions

In this section, we review some basic characteristics of the AR functions. The AR functions are valid (positive definite) correlation functions. They are everywhere positive and at least twice differentiable, except the exponential function  $e^{-|z|/L}$  ( $M = 1$ ) which is singular at  $z = 0$ . The correlation function of some variables is often determined through differentiation of the correlation function of some other variable. A classical example is the use of the geostrophic relation to derive covariance functions for the horizontal velocity variables from (derivatives of) the covariance function of the mass field (Buell, 1972; Daley, 1991). The exponential function is clearly unsuitable for this purpose.

Focusing on the AR functions with  $M > 1$ , consider a second-order Taylor series expansion of Equation (4.41) (with  $z' = 0$ ) about  $z = 0$  :

$$\begin{aligned} f_M(z) &\approx f_M(0) + f_M^{(1)}(0)z + \frac{1}{2}f_M^{(2)}(0)z^2 \\ &= 1 - \frac{1}{2(2M-3)} \frac{z^2}{L^2}, \end{aligned}$$

where  $f_M^{(1)}(0) = 0$  and  $f_M^{(2)}(0) = -(2M-3)^{-1}L^{-2}$ . Here  $f^{(n)}$  denotes the  $n$ -th order derivative of  $f$ . The correlation length scale can be deduced directly from the second term in the expansion :

$$L_{\text{AR}} = \sqrt{-\frac{1}{f_M^{(2)}(0)}} = \sqrt{2M-3} L, \quad (4.54)$$

which is equal to the scale parameter  $L$  only for the special case when  $M = 2$  (SOAR). Note that it is not possible to define the correlation length scale of the exponential function via Equation (4.54) since  $f_1(z)$  is not differentiable at  $z = 0$ . For the exponential function it is common to define the correlation length scale by the scale parameter itself (the  $1/e$  scale).

For a fixed length scale, the Gaussian function is a limiting case of the AR functions as  $M \rightarrow \infty$  (e.g., see Stein, 1999, p. 33). To establish the relationship between the normalization factors and length scales for the explicit and implicit correlation operators, we notice from Equations (4.37) and (4.51) that

$$L_g = \sqrt{2M} L \quad (4.55)$$

and hence from Equation (4.54) that

$$L_{AR} = \sqrt{1 - \frac{3}{2M}} L_g. \quad (4.56)$$

For large  $M$ ,  $M! \approx \sqrt{2\pi} M^{M+1/2} e^{-M}$  (Jeffrey, 1995, p. 219), and hence from Equation (4.45)

$$v_M \approx \sqrt{4\pi(M-1)}. \quad (4.57)$$

Combining this approximation with Equations (4.34), (4.44), (4.55) and (4.57) gives

$$\lambda_M \approx \sqrt{1 - \frac{1}{M}} \lambda_g. \quad (4.58)$$

An expression for the inverse of the Gaussian function,  $g^{-1}(z)$ , is easily deduced as a limiting case of  $f_M^{-1}(z)$  as  $M \rightarrow \infty$ . From Equations (4.47) and (4.55) we note that

$$\gamma_j L^{2j} = \left( \frac{1}{j!} + O(M^{-1}) \right) \left( \frac{L_g^2}{2} \right)^j.$$

Substituting this expression in Equation (4.48) and letting  $M \rightarrow \infty$  gives

$$f_\infty^{-1}(z) = \frac{1}{\lambda_g} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left( \frac{L_g^2}{2} \right)^j \delta^{(2j)}(z).$$

The corresponding inverse Gaussian correlation operator is then (cf. Equation (4.50))

$$\begin{aligned} \mathcal{F}_\infty^{-1}[\varphi(z)] &= \frac{1}{\lambda_g} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left( \frac{L_g^2}{2} \frac{\partial^2}{\partial z^2} \right)^j \phi(z) \\ &= \frac{1}{\lambda_g} \exp \left( \frac{-L_g^2}{2} \frac{\partial^2}{\partial z^2} \right) \phi(z) \end{aligned} \quad (4.59)$$

where the last expression follows from the Taylor series expansion of the exponential function. For practical applications,  $\mathcal{F}_\infty^{-1}$  can only be approximated by truncating the summation in Equation (4.59) at some finite order  $M$ ; i.e., by approximating the inverse Gaussian correlation operator as the inverse of an  $M$ -th order AR correlation operator. The accuracy of finite-term expansions of the inverse Gaussian operator is discussed in Oliver (1998) and Xu (2005). A summary of the fundamental equations presented in this section is given Tables 4.1 and 4.2.

Figure 4.1 shows examples of AR functions with different values of  $M$  and a fixed length scale of  $L_{AR} = 6$  units. Given  $M$  and  $L_{AR}$ ,  $L$  has been computed according to Equation (4.54). Compared to the Gaussian (solid curve), the AR functions display fatter tails particularly for small values of  $M$ .

$f_M(z)$	$\left[1 + \sum_{j=1}^{M-1} \beta_j \left(\frac{ z }{L}\right)^j\right] e^{- z /L}$
$\beta_j$	$\frac{2^j (M-1)! (2M-j-2)!}{j! (M-j-1)! (2M-2)!}$
$L$	$(2M-3)^{-\frac{1}{2}} L_{\text{AR}} \quad (M > 1)$
$\mathcal{F}_M[\psi(z)]$	$\int f_M(z-z') \psi(z') dz'$
$f_M^{-1}(z)$	$\lambda_M^{-1} \left[1 - L^2 \frac{\partial^2}{\partial z^2}\right]^M \delta(z)$
$\lambda_M$	$\frac{2^{2M-1} [(M-1)!]^2 L}{(2M-2)!}$
$\mathcal{F}_M^{-1}[\varphi(z)]$	$\lambda_M^{-1} \left[1 - L^2 \frac{\partial^2}{\partial z^2}\right]^M \varphi(z)$

TAB. 4.1 – A summary of expressions associated with the  $M$ -th order AR functions  $f_M(z)$  with scale parameter  $L$ : the AR polynomial coefficients  $\beta_j$ ; the relationship between  $L$  and the Daley length-scale  $L_{\text{AR}}$ ; the AR correlation operator  $\mathcal{F}_M$ ; the inverse AR function  $f_M^{-1}(z)$ ; the normalization coefficient  $\lambda_M$ ; and the inverse AR correlation operator  $\mathcal{F}_M^{-1}$ .

$f_\infty(z)$	$\exp(-z^2/2L_g^2)$
$L_g$	$L_{\text{AR}}$
$\mathcal{F}_\infty[\psi(z)]$	$\int f_\infty(z-z') \psi(z') dz'$
$f_\infty^{-1}(z)$	$\lambda_\infty^{-1} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left(\frac{L_g^2}{2}\right)^j \delta^{(2j)}(z)$
$\lambda_\infty$	$\sqrt{2\pi} L_g$
$\mathcal{F}_\infty^{-1}[\varphi(z)]$	$\lambda_\infty^{-1} \exp\left(-\frac{1}{2} L_g^2 \frac{\partial^2}{\partial z^2}\right) \varphi(z)$

TAB. 4.2 – As in Table I but for the Gaussian function  $g(z)$  with Daley length-scale  $L_g$ , represented here as an “ $\infty$ -order” AR function  $f_\infty(z)$ .

The “fatness” of a correlation function is often expressed in terms of the *kurtosis* coefficient (e.g., von Storch et Zwiers, 1999, p. 32), which is defined as the ratio of the fourth central moment and the square of the second central moment of the function, where the function is normalized to have unit integral. For the AR functions, the kurtosis coefficient is bounded above by 6 for the exponential and decreases rapidly with increasing  $M$  to its lower limit of 3 for the Gaussian.

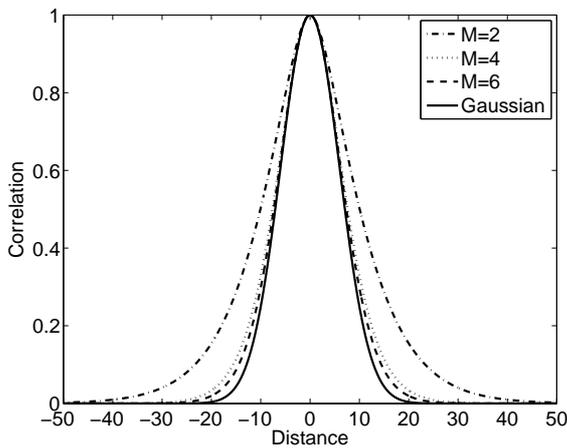


FIG. 4.1 – The AR functions  $f_M(z)$  as a function of  $z$  for different  $M$  and a fixed value of  $L_{\text{AR}} = 6$ . The Gaussian ( $M \rightarrow \infty$ ) is also displayed for comparison.

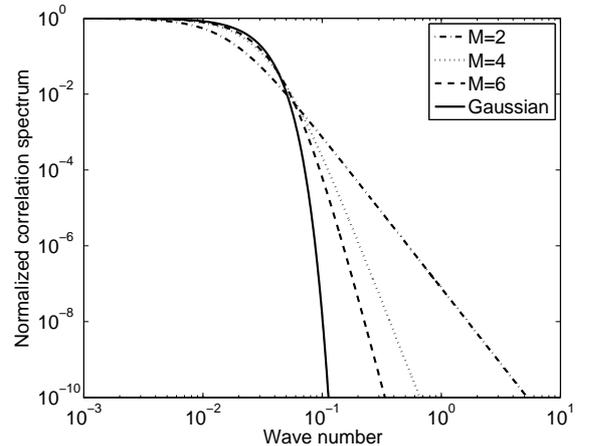


FIG. 4.2 – The normalized power spectrum  $\hat{f}_M(\hat{z})/\hat{f}_M(0)$  where  $\hat{f}_M(0) = \lambda_M$  as a function of  $\hat{z}$  for the same values of  $M$  and  $L_{\text{AR}}$  as in Figure 4.1.

In spectral space, the fatness property translates into a slower drop-off rate of the power spectrum (Fourier transform) at high wavenumbers for small  $M$  as illustrated in Figure 4.2. The drop-off rate sharpens with increasing  $M$ , following a  $\hat{z}^{-2M}$  dependence for high wavenumbers (Equation (4.43)). Correlation functions having more power at high wavenumbers can be desirable in data assimilation to avoid excessive damping of small-scale information in the analysis (Julian et Thiébaux, 1975, P03b).

Figure 4.3 shows the variation of the normalization coefficient  $\lambda_M$  as a function of  $M$ . The values have been normalized by  $\lambda_g$  to indicate the percentage error that would result in the background-error variances if the AR normalization coefficient  $\lambda_M$  was approximated by the Gaussian coefficient  $\lambda_g$ . The case  $M = 1$  has been excluded since it is not possible to use Equation (4.54) to relate  $L_{AR}$  (for  $M = 1$ ) to  $L_g$ . For  $M = 2$ , the error is up to 60% but drops off rapidly for increasing values of  $M$ , with the error being less than 5% for values of  $M > 8$ .

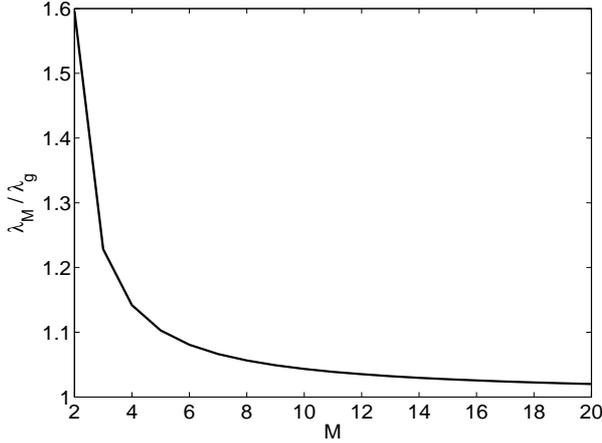


FIG. 4.3 – The ratio of the AR normalization coefficient  $\lambda_M$  (Equations (4.44) and (4.45)) and the Gaussian coefficient  $\lambda_g$  (Equations (4.34) and (4.55)) as a function of  $M$  for  $M \geq 2$ .

#### 4.3.4 Combinations of AR functions

WC01 suggested the following generalization of the diffusion equation

$$\frac{\partial \eta}{\partial t} + \sum_{p=1}^P \kappa_p \left( -\frac{\partial^2}{\partial z^2} \right)^p \eta = 0 \quad (4.60)$$

in order to provide a model with greater flexibility in the choice of shape of the correlation functions. WC01 referred to Equation (4.60) as a generalized diffusion (GD) equation. The general solution of Equation (4.60) is given by the convolution integral

$$\eta(z, t) = \int_{-\infty}^{\infty} h_{GD}(z-z') \eta_0(z') dz' \quad (4.61)$$

where

$$h_{GD}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{h}_{GD}(\hat{z}) e^{iz\hat{z}} d\hat{z} \quad (4.62)$$

is the inverse FT of

$$\hat{h}_{GD}(\hat{z}) = \exp \left( -\sum_{p=1}^P \kappa_p t \hat{z}^{2p} \right). \quad (4.63)$$

Expressed in terms of an  $M$ -step explicit GD operator ( $T = M\Delta t$ ), the correlation operator associated with Equation (4.61) is

$$\mathcal{F}_{GD}[\eta_0(z)] \equiv \lambda_{GD} \mathcal{L}_{GD}^M[\eta_0(z)] \quad (4.64)$$

where

$$\mathcal{L}_{GD} \equiv 1 + \sum_{p=1}^P \kappa_p \Delta t \left( -\frac{\partial^2}{\partial z^2} \right)^p$$

and  $\lambda_{GD} = 1/h_{GD}(0)$ . For a given  $M$ , chosen large enough to keep Equation (4.64) stable, the parameters  $\kappa_p \Delta t$ ,  $p = 1, \dots, P$ , can be tuned from Equations (4.62) and (4.63) to obtain a correlation

function with particular characteristics. Apart from the Gaussian case ( $P = 1$ ), Equation (4.62) must be solved numerically for a given set of parameters. This, as well as the increased cost and complexity of using higher order powers of the second-derivative operator, makes the GD equation less practical than the standard diffusion equation.

Correlation functions can be combined linearly to produce a more general set of correlation functions, as pointed out by P03a. The method is practical if the source functions are relatively simple and inexpensive to manipulate. In particular, consider the following general function  $f_{\text{LC}}(z-z')$ , defined as a linear combination of weighted AR functions,

$$f_{\text{LC}}(z-z') = \frac{1}{\bar{\gamma}} \sum_{p=1}^P \gamma_p f_{M_p}(z-z') \quad (4.65)$$

where  $M_p$  denotes the order of the  $p$ -th AR function in the summation,  $\gamma_p$  are constant weights, and  $\bar{\gamma} = \sum_{p=1}^P \gamma_p$ . The multiplicative factor  $1/\bar{\gamma}$  ensures that the resulting function has unit amplitude. Since the  $f_{M_p}(z-z')$  are valid correlation functions, their linear combination is necessarily a valid correlation function if  $\gamma_p > 0$ . More generally, we can choose some of the  $\gamma_p$  to be negative, although special care must be taken to ensure that the resulting combination is positive definite (Ma, 2005; Gregori *et al.*, 2008). Negative amplitude coefficients are required to produce negative lobes in the correlation function as illustrated in the example in this section.

The correlation operator associated with Equation (4.65) can be modelled using a linear combination of implicit diffusion operators  $\mathcal{L}_i^{M_p}$  :

$$\mathcal{F}_{\text{LC}}[\eta_0(z)] \equiv \frac{1}{\bar{\gamma}} \sum_{p=1}^P \gamma_p \lambda_{M_p} \mathcal{L}_i^{M_p}[\eta_0(z)]$$

where, for each  $\mathcal{L}_i^{M_p}$ , the diffusion parameter  $\kappa_p \Delta t$  can be defined in terms of the scale parameter  $L_p$  using Equation (4.51) and the normalization factor  $\lambda_{M_p}$  can be defined from Equations (4.44) and (4.45). The example below illustrates how the scale parameters and weights can be tuned to produce a desired correlation shape and length scale.

### Example : combining two SOAR functions

As an example, consider the following linear combination of two SOAR functions,

$$f_{\text{LC}}(z) = \frac{1}{\gamma_1 + \gamma_2} \left[ \gamma_1 \left( 1 + \frac{|z|}{L_1} \right) e^{-|z|/L_1} + \gamma_2 \left( 1 + \frac{|z|}{L_2} \right) e^{-|z|/L_2} \right], \quad (4.66)$$

where  $L_1$  and  $L_2$  ( $L_1 \neq L_2$ ) are scale parameters. Martin *et al.* (2007) used a function of the form (4.66) (without the normalization term) to define the background-error covariances in an ocean data assimilation system. In their system, each SOAR function was modelled using a recursive filter (see Section 4.5.3) and their parameters were chosen to represent distinct sources of background error : one SOAR function was used to represent ocean ('synoptic') state errors resulting from errors in the atmospheric forcing fields, while the other SOAR function was used to represent ('mesoscale') state errors arising from errors in the internal dynamics, with the SOAR parameters chosen such that  $L_1 > L_2$  and  $\gamma_1 < \gamma_2$  to reflect the larger scale and smaller variance of the 'synoptic' errors compared to the 'mesoscale' errors.

In this section, we present a different example in which we employ Equation (4.66) to model a correlation function with negative lobes, a feature that may be desirable for certain geophysical fields

but which cannot be represented with a standard AR function. To reduce the number of tunable parameters, we consider the special case where  $\gamma_1 = L_2$  and  $\gamma_2 = -\tau L_1$ , with  $\tau$  a strictly positive (tuning) coefficient that we will use to control the length scale of the resulting correlation function. The amplitude coefficients have physical dimensions of length; they are not error variances as in the previous example taken from Martin *et al.* (2007). With the above parameter settings, Equation (4.66) becomes

$$f_{\text{LC}}(z) = \frac{1}{L_2 - \tau L_1} \left[ L_2 \left( 1 + \frac{|z|}{L_1} \right) e^{-|z|/L_1} - \tau L_1 \left( 1 + \frac{|z|}{L_2} \right) e^{-|z|/L_2} \right]. \quad (4.67)$$

The positive definiteness of Equation (4.67) is guaranteed by ensuring the positiveness of its FT (Bochner's theorem). General conditions for producing positive definite functions from combinations of Matérn functions have been derived by Gregori *et al.* (2008). Applying these results, which are summarized in Table 1 of Gregori *et al.* (2008), to the example here leads to the following conditions on  $L_1$ ,  $L_2$  and  $\tau$  :

$$L_1 < L_2 \quad \text{and} \quad 0 < \tau < 1, \quad \text{or} \quad (4.68)$$

$$L_1 > L_2 \quad \text{and} \quad \tau > 1. \quad (4.69)$$

The length scale of  $f_{\text{LC}}(z)$  is defined in the usual way :

$$L_{\text{LC}} = \sqrt{-\frac{1}{f_{\text{LC}}^{(2)}(0)}} = L_1 L_2 \sqrt{\frac{L_2 - \tau L_1}{L_2^3 - \tau L_1^3}}. \quad (4.70)$$

In terms of  $L_{\text{LC}}$ , conditions (4.68) and (4.69) read

$$L_1 < L_2 \quad \text{and} \quad L_1 < L_{\text{LC}} < \bar{L}_{\text{LC}}, \quad \text{or} \quad (4.71)$$

$$L_1 > L_2 \quad \text{and} \quad \bar{L}_{\text{LC}} < L_{\text{LC}} < L_2 \quad (4.72)$$

where

$$\bar{L}_{\text{LC}} = L_1 L_2 \sqrt{\frac{L_2 - L_1}{L_2^3 - L_1^3}}$$

corresponds to the value of  $L_{\text{LC}}$  for  $\tau = 1$ . For a given  $L_{\text{LC}}$ , scale parameters  $L_1$  and  $L_2$  can be adjusted under conditions (4.71) and (4.72) to produce a desired shape. Then, from the values of  $L_1$  and  $L_2$ , the parameter  $\tau$  can be determined from the inverse of Equation (4.70),

$$\tau = \frac{L_2^3(L_1^2 - L_{\text{LC}}^2)}{L_1^3(L_2^2 - L_{\text{LC}}^2)},$$

in order to maintain the desired value of  $L_{\text{LC}}$ .

Figures 4.4 and 4.5 show examples of the shape and corresponding power spectrum of correlation functions that can be obtained from Equation (4.67) for different choices of  $L_1$  and  $L_2$ . In all the examples, the length scale is fixed at  $L_{\text{LC}} = 2$ . For comparison, a single SOAR function with the same length scale is shown (dash-dotted curve). The negative lobes in the shape (Fig. 4.4) are associated with the increase in power at intermediate wavenumbers (Fig. 4.5). The spectral decay rate at high wavenumbers is similar for all curves and corresponds to the  $\hat{z}^{-4}$  dependence of the standard SOAR function.

Finally, it is worth mentioning that the GD equation (4.60) can also be used to generate correlation functions with negative lobes. For example, negative lobes are supported by the class of so-called

*Spartan Spatial Random Field* correlation functions (Hristopulos, 2003; Hristopulos et Elogne, 2007, 2009) which can be obtained as a special case of the family of implicit GD-operator kernels in which  $M = 1$ ,  $P = 2$ ,  $\kappa_1 \Delta t = \gamma_1 L^2$  and  $\kappa_2 \Delta t = L^4$ , where  $\gamma_1$  and  $L$  are shape and scale parameters analogous to the parameters  $\eta_1$  and  $\xi$  in Equation (27) of Hristopulos (2003). Setting  $\gamma_1 = 2$  recovers the SOAR function of the Matérn correlation family while negative values of  $\gamma_1$  satisfying the permissibility condition  $\gamma_1^2 < 4$  give rise to oscillatory functions with negative lobes as illustrated in Fig. 4 of Hristopulos (2003).

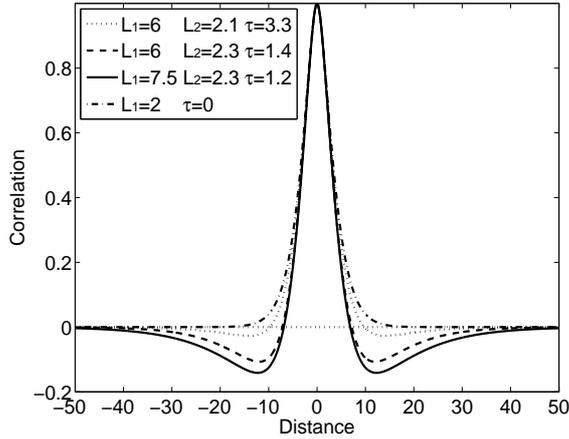


FIG. 4.4 – The correlation functions  $f_{LC}(z)$  constructed from the difference of two SOAR functions for various values of the parameters  $L_1$ ,  $L_2$  and  $\tau$  (Equation (4.67)). The length scale is fixed at  $L_{LC} = 2$  (Equation (4.70)) in all curves. The dash-dotted curve corresponds to a single SOAR function with a length-scale  $L_{LC}$ .

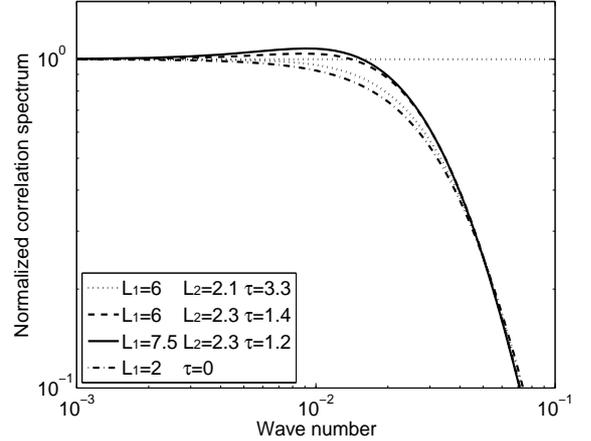


FIG. 4.5 – The normalized power spectrum  $\hat{f}_{LC}(\hat{z})/\hat{f}_{LC}(0)$  of the correlation functions in Fig. 4.4.

## 4.4 Inhomogeneous and anisotropic extensions

### 4.4.1 Location-dependent length scales

The assumption that correlation length scales are constant is very restrictive in geophysical data assimilation applications. The following modification to Equation (4.41) can be made to enable the AR functions to account for spatially varying scale parameters  $L(z)$  :

$$f_M(z, z') = A(z, z') \left[ 1 + \sum_{j=1}^{M-1} \beta_j \left( \frac{|z - z'|}{\sqrt{\overline{L^2(z, z')}}} \right)^j \right] \times \exp \left( \frac{-|z - z'|}{\sqrt{\overline{L^2(z, z')}}} \right), \quad (4.73)$$

where  $\overline{L^2(z, z')}$  denotes the mean of the square of the scale parameters at points  $z$  and  $z'$ , and

$$A(z, z') = \frac{\sqrt{L(z)}\sqrt{L(z')}}{\sqrt{\overline{L^2(z, z')}}}$$

with  $A(z, z) = 1$ . Cummings (2005) used the geometric mean  $\overline{L^2(z, z')} = L(z)L(z')$  with a SOAR function to represent geographically dependent background-error correlations in an operational ocean

data assimilation system based on optimal interpolation. The geometric mean, however, does not guarantee positive-definite AR-like functions for arbitrary  $L(z)$ , although this may not be a problem in practice if  $L(z)$  varies on scales that are sufficiently large compared to  $L(z)$  itself (Daley et Barker, 2001). A better choice is the arithmetic mean  $\overline{L^2(z, z')} = \frac{1}{2}(L^2(z) + L^2(z'))$  which does guarantee positive-definite AR-like functions according to Theorem 1 of Paciorek et Schervish (2006) (see also Theorem 1 in Stein (2005) for a more general result).

An intuitive generalization of the results from Section 4.3 suggests that the correlation operator  $\mathcal{F}_M$  associated with Equation (4.73) can be modelled approximately by

$$\begin{aligned}\mathcal{F}_M[\eta_0(z)] &\approx \sqrt{\tilde{\lambda}_M(z)} \mathcal{L}_i^M \left[ \sqrt{\tilde{\lambda}_M(z)} \eta_0(z) \right] \\ &\equiv \tilde{\mathcal{F}}_M[\eta_0(z)]\end{aligned}\quad (4.74)$$

where  $\mathcal{L}_i^M$  is an  $M$ -step implicit diffusion operator with variable diffusion and normalization coefficients defined such that

$$\begin{aligned}L(z) &= \sqrt{\kappa(z)\Delta t} \\ \text{and } \tilde{\lambda}_M(z) &= v_M L(z).\end{aligned}\quad (4.75)$$

The normalization coefficients have been split into square-root factors to preserve the symmetry of  $\tilde{\mathcal{F}}_M$  (cf. Equation (4.53), (4.51) and (4.44), respectively). The inverse of the one-step implicit diffusion operator with variable diffusion coefficients is given by

$$\mathcal{L}_i^{-1} \equiv 1 - \Delta t \frac{\partial}{\partial z} \left( \kappa(z) \frac{\partial}{\partial z} \right),$$

where by placing the diffusion coefficients within the second derivative we ensure that  $\mathcal{L}_i^{-1}$  remains self-adjoint,  $\langle \mathcal{L}_i^{-1} \eta_1, \eta_2 \rangle = \langle \eta_1, \mathcal{L}_i^{-1} \eta_2 \rangle$ , with respect to the inner product  $\langle \eta_1, \eta_2 \rangle = \int \eta_1(z) \eta_2(z) dz$ . The approximate unit-amplitude functions generated by Equation (4.74) will be denoted by  $\tilde{f}_M(z, z')$ . Note that, by construction, the implicit diffusion operator is positive definite for arbitrary (positive)  $\kappa(z)$  (see Section 4.5).

Equation (4.75) is only an approximation of the normalization factors required to transform  $\mathcal{L}_i^M$  into a correlation operator. A similar approximation was suggested by Pannekoucke et Massart (2008) for modelling the Gaussian function in  $\mathbb{R}^2$  with an explicit version of the 2D diffusion equation. To evaluate the accuracy of this approximation, we need to compute the exact normalization factors  $\lambda_M(z)$  which can be done as follows. First, let  $s_M(z, z')$  denote the kernel (covariance function) of the integral operator implied by  $\mathcal{L}_i^M$ . From Equation (4.19), we can write

$$s_M(z, z') = \sqrt{s_M(z, z)} \sqrt{s_M(z', z')} \bar{f}_M(z, z'),$$

where  $s_M(z, z)$  is the variance (with physical dimensions equal to length) and  $\bar{f}_M(z, z') \approx f_M(z, z')$  is a correlation function ( $\bar{f}_M(z, z) = 1$ ). The associated correlation operator  $\bar{\mathcal{F}}_M$  is then given by

$$\begin{aligned}\bar{\mathcal{F}}_M[\eta_0(z)] &= \int_{-\infty}^{\infty} \bar{f}_M(z, z') \eta_0(z') dz' \\ &= \frac{1}{\sqrt{s_M(z, z)}} \int_{-\infty}^{\infty} \frac{s_M(z, z')}{\sqrt{s_M(z', z')}} \eta_0(z') dz' \\ &= \sqrt{\lambda_M(z)} \mathcal{L}_i^M \left[ \sqrt{\lambda_M(z)} \eta_0(z) \right].\end{aligned}\quad (4.76)$$

From the last two expressions, it is clear that the normalization factors required to transform  $s_M(z, z')$  into  $\bar{f}_M(z, z')$  are equal to the inverse of the variances :

$$\lambda_M(z) = \frac{1}{s_M(z, z)}.$$

At a given point  $z = z_k$ , the covariance function  $s_M(z, z_k)$  and variance  $s_M(z_k, z_k)$  satisfy the identities

$$s_M(z, z_k) = \int_{-\infty}^{\infty} s_M(z, z') \delta(z_k - z') dz' \quad (4.77)$$

$$\text{and } s_M(z_k, z_k) = \int_{-\infty}^{\infty} s_M(z, z_k) \delta(z_k - z) dz. \quad (4.78)$$

Equations (4.77) and (4.78) provide the basis of a numerical algorithm for computing the variance (and hence normalization factor) at each grid point  $z_k$ . In the diffusion algorithm, the integral in Equation (4.77) is evaluated implicitly using a discretized version of the  $\mathcal{L}_i^M$  operator, and the Dirac delta function is represented by a vector with a value at  $z_k$  equal to the inverse of the local grid size and a value of zero at all other grid points (WC01; Bennett, 2002). From Equation (4.78), the variance at  $z_k$  is just the value of  $s_M(z, z_k)$  at  $z = z_k$ .

The implicit diffusion equation can be solved using the numerical techniques detailed in Section 4.5. There, the equation is discretized with respect to general orthogonal curvilinear coordinates to make it easier to see how the algorithm could be applied in a more general application than considered in this paper. For the simple examples presented here, the equation has been discretized on a uniform grid,  $z_k = k\Delta z$ ,  $k = 1, \dots, K$ , where the grid size  $\Delta z$  has been set to one unit and  $K = 100$ . The time step  $\Delta t$  has also been set to one unit. The length scale  $L_{AR}(z_k)$  is assumed to vary with  $z_k$  following a cosine relationship,

$$L_{AR}(z_k) = A \cos\left(\frac{2\pi}{Z}z_k + \theta\right) + B \quad (4.79)$$

where  $A$ ,  $B$ ,  $Z$  and  $\theta$  are constants. From Equation (4.54), the scale parameter  $L(z_k)$  is then set to

$$L(z_k) = \frac{1}{\sqrt{2M-3}} L_{AR}(z_k)$$

The amplitude constants  $A$  and  $B$  are defined in terms of the maximum- and minimum-allowed values of  $L_{AR}(z_k)$ :  $A = \frac{1}{2}(L_{\max} - L_{\min})$  and  $B = \frac{1}{2}(L_{\max} + L_{\min})$ .

Figures 4.6a-c show examples of the 2-step implicit diffusion response to a delta function for three different parameter settings in  $L_{AR}(z)$  (see figure caption). The delta function has been placed far from the domain end-points to avoid being influenced by the boundary conditions. Illustrated are the functions obtained using the approximate as well as the exact normalization factors. For comparison, the analytical functions generated by Equation (4.73), with  $\overline{L^2(z, z')}$  defined by the arithmetic mean of  $L^2(z)$  and  $L^2(z')$ , are also shown for the same parameter settings. The maximum local variation in the scales occurs at  $z_k = ((2p+1)/4 - \theta/2\pi)Z$ ,  $p = 0, 1, \dots$ , where  $|dL_{AR}/dz_k|_{\max} = A2\pi/Z$ . With the parameter settings in panels (a)-(c), the values of  $|dL_{AR}/dz_k|_{\max}$  are  $\pi/200$ ,  $\pi/25$  and  $3\pi/25$ , respectively. In Fig. 4.6a, the variation of  $L_{AR}(z_k)$  is sufficiently slow that the three curves are virtually indistinguishable. Increasing the frequency of the variation of  $L_{AR}(z_k)$  from  $1/Z = 1/200$  to  $1/25$  leads to differences between the diffusion responses and the analytical function mainly near the tail of the function (Fig. 4.6b). Increasing the amplitude of  $L_{AR}(z_k)$  from  $A = 1$  to 3 leads to further differences between the diffusion responses and the analytical function, as well as differences between the two diffusion responses themselves, with the solution obtained using the approximate normalization now producing a 10% error in the amplitude of the correlation function (Fig. 4.6c).

The algorithm for computing the exact normalization factors based on Equations (4.77) and (4.78) requires as many applications of the  $\mathcal{L}_i^M$  operator as there are model grid points. As such, the algorithm is expensive and of limited practical interest for large dimensional problems. WC01 proposed an alternative algorithm that uses a sample of randomized vectors to approximate the normalization

factors. The randomization algorithm is generally much cheaper than the exact method based on Equations (4.77) and (4.78), and with enough randomized vectors, can be made more accurate than the simple approximation based on Equation (4.75), but still requires multiple applications of the  $\mathcal{L}_i^M$  operator and thus remains expensive for certain applications.

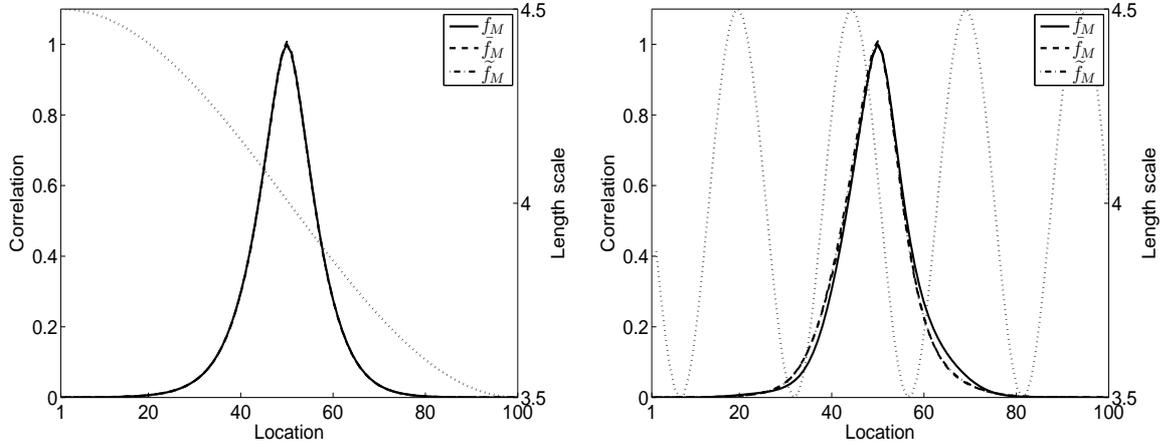
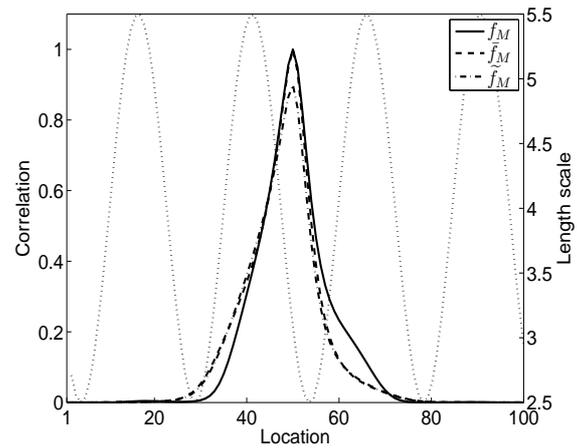
(a) Length scale with  $A = 1/2$ ,  $Z = 200$ ,  $\theta = 0$ (b) Length scale with  $A = 1/2$ ,  $Z = 25$ ,  $\theta = \pi/2$ 

FIG. 4.6 – The result of applying a 2-step implicit diffusion operator with variable coefficients to a Dirac delta function located at  $z_{50}$ . The dashed-dotted and dashed curves have been produced using approximate and exact normalization factors, respectively (Equations (4.74) and (4.76) with  $M = 2$ ). For comparison, the SOAR function with variable scales is also shown (Equation (4.73) with  $M = 2$ ; solid curves). The right axis is associated with the dotted curves which show the length scale variation as a function of  $z_k$  (Equation (4.79)) : (a)  $Z = 200$ ,  $\theta = 0$ ,  $A = 1/2$ ,  $B = 4$ ; (b)  $Z = 25$ ,  $\theta = \pi/2$ ,  $A = 1/2$ ,  $B = 4$ ; and (c)  $Z = 25$ ,  $\theta = 3\pi/4$ ,  $A = 3/2$ ,  $B = 4$ .

(a) Length scale with  $A = 3/2$ ,  $Z = 25$ ,  $\theta = 3\pi/4$ 

By considering an asymptotic expansion of the solution to the diffusion equation for the case of slowly and smoothly varying coefficients  $\kappa(z)$ , P03b showed that a better approximation to the Gaussian normalization factor can be obtained by replacing  $\kappa$  in the constant Gaussian formula (Equation (4.34)) with the variable coefficients  $\kappa(z)$  filtered using the “square-root” of the diffusion operator. Extending this result to the implicit diffusion operator, we have thus the approximation

$$\lambda_M(z) \approx v_M \sqrt{\overline{\kappa(z)} \Delta t} = v_M \sqrt{\overline{(L(z))^2}} \equiv \bar{\lambda}_M(z) \quad (4.80)$$

where  $\overline{\kappa(z)} = \mathcal{L}_i^{M/2}[\kappa(z)]$ . The approximation (4.80) differs from the approximation (4.75) only in the use of  $\overline{\kappa(z)}$  instead of  $\kappa(z)$ . Although this approximation is slightly more expensive than that of (4.75), it is still much cheaper than the randomization algorithm.

The accuracy of the two approximations are compared in Figures 4.7a and b which show their errors relative to the exact values for a low-order ( $M = 2$ ) and higher order ( $M = 10$ ) implicit diffusion operator. The same parameter settings as in Fig. 4.6b have been used. In both examples, the spatial variability in the errors is much reduced using the method of P03b, particularly for the low-order diffusion operator where the amplitude of the error variability is reduced from 6% to less

than 1% (Fig. 4.7a). Both errors exhibit a constant bias between 1% and 2%. This bias is associated with truncation errors in the discretization of the diffusion equation and can be reduced by increasing the grid resolution (not shown).

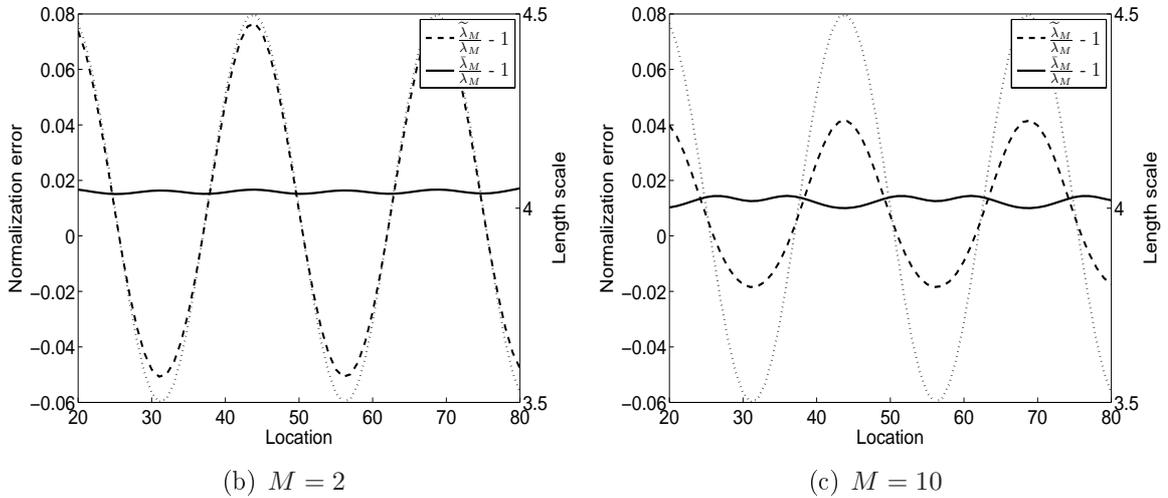


FIG. 4.7 – The relative error between the approximate and exact normalization factors for (a)  $M = 2$ , and (b)  $M = 10$ . The dashed curves are the errors associated with the approximation  $\tilde{\lambda}_M(z)$  based on the analytical relationship (4.75); the solid curves are the errors associated with the approximation  $\bar{\lambda}_M(z)$  based on the method of P03b. The right axis is associated with the dotted curves which show the length scale variation as a function of  $z_k$  (Equation (4.79)). The parameters are the same as in Fig. 4.7b.

#### 4.4.2 Boundary conditions

In a finite domain, the diffusion equation must be supplied with boundary conditions (BCs). Here we study the effect of (non-periodic) BCs on the correlation functions represented by the diffusion equation. For example, for representing horizontal and vertical background-error correlations in an ocean model, BCs must be specified at solid boundaries associated with coastlines and bottom topography, as well as at the ocean-atmosphere or ocean-sea ice interfaces. In the temporal domain, boundaries appear at the end-points of an assimilation window. These artificial boundaries must be taken into account if the diffusion equation is used to model temporally correlated model or observation error over the assimilation window.

WC01 discussed BCs in the context of an explicit version of the 2D diffusion equation. They concluded that, providing the functions are correctly normalized, the difference in the correlation functions resulting from the use of Neumann BCs (setting the normal derivative of the field to be zero at the boundary) and Dirichlet BCs (setting the field to be zero at the boundary) is generally small compared to the uncertainty in our knowledge of actual background-error correlations. Since an explicit scheme with Neumann BCs requires fewer iterations than an explicit scheme with Dirichlet BCs, they argued that, for purely computational reasons, Neumann BCs are preferable.

Given estimates of correlated background error, the role of the diffusion equation is to model these estimates as accurately as possible. This can be done by tuning the diffusion model parameters  $M$  and  $\kappa(z)$  to fit estimates of the correlation spectra and length scales obtained, for example, from statistics of observation-minus-forecast or ensemble-forecast differences. From this perspective, the most appropriate BCs for the diffusion equation are *transparent* BCs. This way the correlation functions defined through parameter tuning are indeed those effectively used by the assimilation method, at all

points in the domain. One possibility to achieve this in a bounded domain is to extend the grid beyond the boundary by adding fictitious active points, and then to remove these points after application of the diffusion equation. The number of fictitious points necessary to avoid contamination by the BC would be a function of the correlation length scale near the boundary. Dobricic et Pinardi (2008) and Liu *et al.* (2009) and applied this technique with a recursive filter to model background-error correlations near coastlines in an ocean 3D-Var scheme. In this section, we explore alternative methods for simulating transparent boundaries which do not require modifications to the analysis grid.

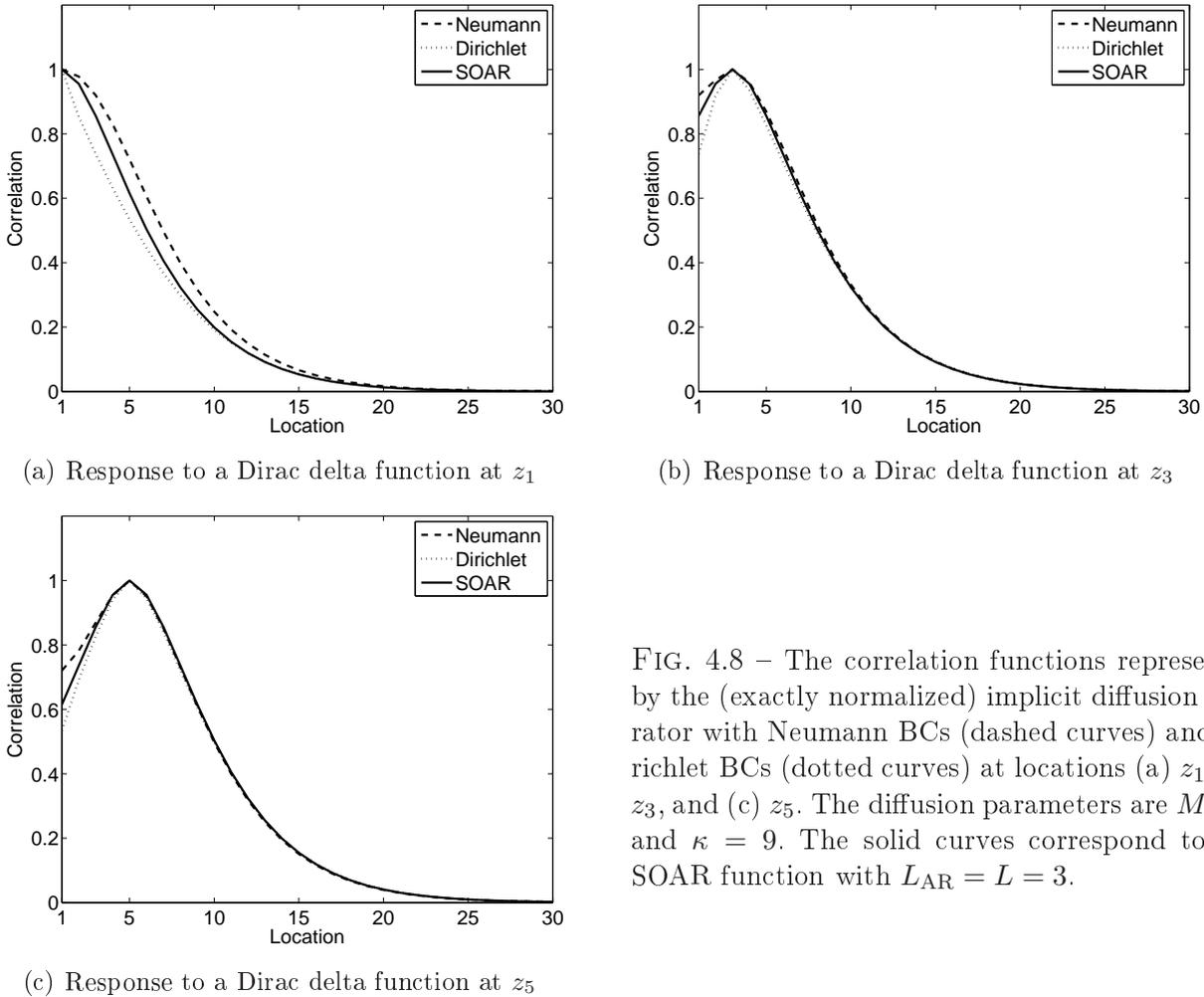


FIG. 4.8 – The correlation functions represented by the (exactly normalized) implicit diffusion operator with Neumann BCs (dashed curves) and Dirichlet BCs (dotted curves) at locations (a)  $z_1$ , (b)  $z_3$ , and (c)  $z_5$ . The diffusion parameters are  $M = 2$  and  $\kappa = 9$ . The solid curves correspond to the SOAR function with  $L_{AR} = L = 3$ .

Consider the 1D diffusion equation on a finite domain  $0 \leq z \leq l$  where the BCs located at the end-points  $z = 0$  and  $z = l$  are assumed to be of the general form,

$$\mu\eta - (1 - \mu)\bar{e}\frac{\partial\eta}{\partial z} = 0 \quad \text{at } z = 0, \quad (4.81)$$

$$\mu\eta + (1 - \mu)\bar{e}\frac{\partial\eta}{\partial z} = 0 \quad \text{at } z = l. \quad (4.82)$$

$\mu$  is a parameter such that  $0 \leq \mu \leq 1$ , and  $\bar{e}$  is a scale parameter to make the terms in Equations (4.81) and (4.82) dimensionally consistent. Equations (4.81) and (4.82) are known as Robin BCs or BCs of the third type. The parameter  $\mu$  controls the relative strength of the two terms, where the limiting cases  $\mu = 0$  and  $\mu = 1$  correspond, respectively, to the Neumann and Dirichlet BCs considered by WC01. Details of the numerical implementation of the Robin BCs in the implicit diffusion equation are given in Section 4.5. As in the previous section, the diffusion equation has been discretized here on a uniform grid with grid size and time step set to one unit.

Figure 4.8 shows the solutions to the implicit diffusion equation obtained near the left end-point when Neumann and Dirichlet BCs are employed (dashed and dotted curves, respectively). Panels a, b and c correspond to the solutions obtained when the initial condition is taken to be a Dirac delta function at grid-point locations  $z_1$ ,  $z_3$  and  $z_5$ , respectively. In this example,  $M = 2$  and the diffusion coefficient  $\kappa$  is defined from Equation (4.51) to give a constant length scale of 3 units. All solutions have been normalized exactly using the procedure outlined in Section 4.4.1. For comparison, the SOAR function with  $L_{AR} = 3$  units is also displayed (solid curves). The SOAR function corresponds to the response of the implicit diffusion equation, with  $M = 2$ , to a Dirac delta function located far from the boundaries.

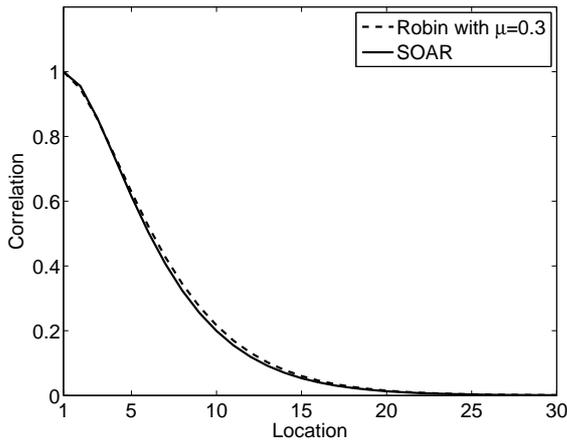


FIG. 4.9 – As in Figure 4.8a but with Robin BCs (dashed curve).

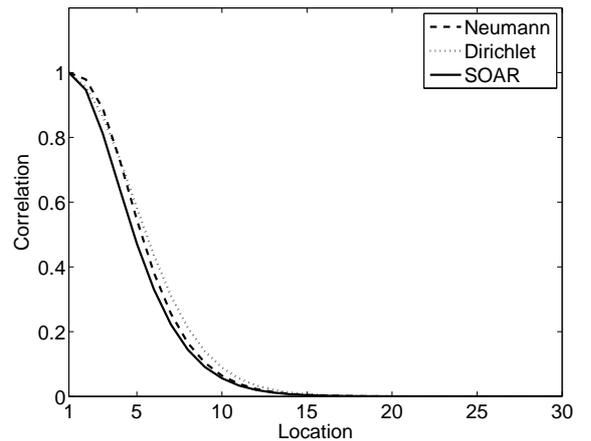


FIG. 4.10 – As in Figure 4.8a but with  $M = 10$ . The solid curve corresponds to a 10-th order AR function.

Figure 4.8 illustrates that, while the amplitudes of the correlation functions are correct, the shape of the functions produced with the diffusion equation are slightly corrupted with respect to the SOAR function. At  $z_3$  and  $z_5$ , the diffusion equation produces an asymmetric response. At  $z_1$ , the length scales of the correlation function produced with Neumann (Dirichlet) BCs are effectively increased (decreased) by approximately 1 unit relative to the length scale of the SOAR function. This is also the case at  $z_3$  and  $z_5$  in the direction of the boundary.

The correlation functions obtained using Neumann and Dirichlet BCs lie on either side of the SOAR function. This suggests that a better approximation of the SOAR function might be obtained using Robin BCs with a value of  $\mu$  between 0 and 1. Figure 4.9 shows that it is possible to obtain a good fit to the SOAR function at all locations by setting  $\mu = 0.3$ . (The scale parameter  $\bar{e}$  has been set equal to  $\Delta z = 1$  unit; see Section 4.5). For higher order AR functions, however, it is not possible to obtain a similar good fit using Robin BCs. For example, Figure 4.10 shows the result of applying a 10-step implicit diffusion operator with Neumann BCs (dashed curve) and Dirichlet BCs (dotted curve). In contrast to the previous example, both curves are now displaced to the right of the corresponding (10-th order) AR function. In this case, the best fit to the AR function is obtained using Neumann BCs ( $\mu = 0$ ).

While Robin BCs can be tuned to provide everywhere a good fit between a 2-step implicit diffusion operator and a SOAR function, they are inadequate for producing a good fit between the implicit diffusion operator and higher order AR functions. Even for the 2-step implicit diffusion operator, a costly computation of the normalization factors is required since the simple methods described in Section 4.4.1 for approximating these factors are inadequate near boundaries. An elegant and robust method for simulating transparent boundaries, which can be used effectively in combination with the approximate normalization factors, is to define the correlation operator as an average of the solutions

from two diffusion problems, one employing Neumann BCs and the other employing Dirichlet BCs. For the case of a constant length scale, the  $M$ -th order AR correlation operator then becomes

$$\mathcal{F}_M[\eta_0(z)] = \frac{1}{2}\lambda_M(\mathcal{L}_N^M[\eta_0(z)] + \mathcal{L}_D^M[\eta_0(z)]) \quad (4.83)$$

where  $\mathcal{L}_N^M$  and  $\mathcal{L}_D^M$  denote the  $M$ -step implicit diffusion operators with Neumann and Dirichlet BCs, respectively. A mathematical proof of this result is given in Appendix 4.B for the Gaussian case ( $M \rightarrow \infty$ ) by examining the analytical solution of the diffusion equation near boundaries. Equation (4.83) follows from an intuitive generalization of this result to account for the (finite  $M$ ) AR functions modelled by the implicit diffusion equation. The validity of this hypothesis is illustrated numerically below for the SOAR function. The result also appears to hold well for the case of spatially varying length scales (not shown).

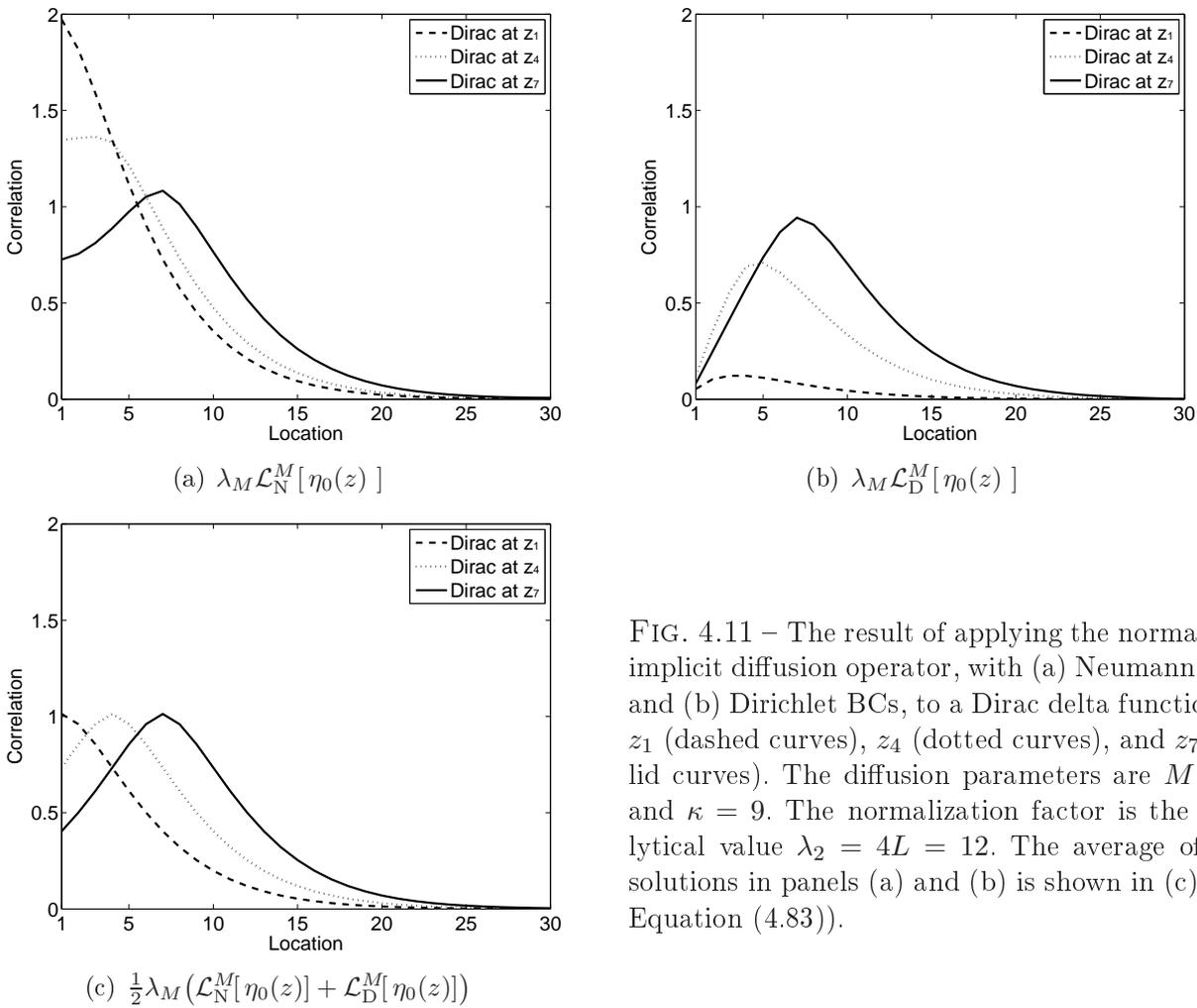


FIG. 4.11 – The result of applying the normalized implicit diffusion operator, with (a) Neumann BCs and (b) Dirichlet BCs, to a Dirac delta function at  $z_1$  (dashed curves),  $z_4$  (dotted curves), and  $z_7$  (solid curves). The diffusion parameters are  $M = 2$  and  $\kappa = 9$ . The normalization factor is the analytical value  $\lambda_2 = 4L = 12$ . The average of the solutions in panels (a) and (b) is shown in (c) (see Equation (4.83)).

Figure 4.11a shows the response of a 2-step implicit diffusion equation to a Dirac delta function at different locations from the boundary when Neumann BCs are employed. The length scale is  $L_{AR} = 3$  as in the previous examples. The solutions have been multiplied by the analytical estimate of the normalization factor  $\lambda_2 = 4L = 12$ . With the Neumann BCs, the boundaries are insulated (no flux is exchanged across the boundary), so the field being diffused effectively accumulates near the boundary giving rise to an overestimation of the amplitude there. Mathematically, the solution with Neumann BCs can be expressed as a  $2L$ -periodic function built from the *sum* of two AR functions (here SOAR) : one centred at  $z_k$  and the other centred at  $-z_k$  (see Equation (4.106) in Appendix 4.B. When  $z_k$  is close to the boundary, the solution is the sum of the AR function centred at  $z_k$  and that part of the

same function centered at  $-z_k$  which appears in the active part of the domain. On the other hand, when Dirichlet BCs are employed, information is effectively lost through the boundary and the result is an underestimation of the solution amplitude near the boundaries (Figure 4.11b). Mathematically, the solution with Dirichlet BCs can be expressed as a  $2l$ -periodic function built from the *difference* of two AR functions : one centered at  $z_k$  and the other centred at  $-z_k$ . When  $z_k$  is close to the boundary, the solution is the difference of the AR function centred at  $z_k$  and that part of the same function centered at  $-z_k$  which appears in the active part of the domain (see Equation (4.107)).

The sum of the two solutions with Neumann and Dirichlet BCs effectively cancels out the contaminating part of the solution arising from the AR functions centred at  $-z_k$ , but increases the amplitude of the solution by a factor of two. Figure 4.11c illustrates the result of averaging the two solutions in panels a and b. The SOAR function is perfectly recovered with correct amplitude at all points. While Equation (4.83) requires an additional application of the diffusion operator, this extra expense is small in comparison to the cost of computing the normalization factors via the exact or randomization methods. This will be especially true for applications requiring regular updating of the normalization factors in response to flow-dependent parameterizations of the length scales.

## 4.5 Numerical aspects

### 4.5.1 Curvilinear coordinates

We consider the diffusion equation defined with respect to general orthogonal curvilinear coordinates  $(i, j, k)$ . In a numerical model, the coordinates  $(i, j, k)$  would correspond to the orthogonal lines that define the three-dimensional model grid, with  $(i, j)$  being the horizontal coordinates and  $k$  the vertical coordinate. Let  $e_1$ ,  $e_2$  and  $e_3$  denote the metric coefficients (partial derivatives with respect to  $i$ ,  $j$  and  $k$ , respectively) that define the curvilinear distance elements  $(ds_1, ds_2, ds_3) = (e_1 di, e_2 dj, e_3 dk)$  in the  $(i, j, k)$  coordinate system (e.g., see Strang, 1986). Here we consider the 1D diffusion equation in which the second-order spatial derivative is constructed from one of the three components of the 3D Laplacian operator defined with respect to  $(i, j, k)$ .

For the 1D problem, it is natural, as an example, to consider the component associated with the vertical coordinate  $k$  at a fixed horizontal position  $(i, j)$ . Assuming that  $k$  is related to vertical distance  $z$  via a continuous and differentiable function  $z = z(k)$ , the 1D diffusion equation reads

$$\frac{\partial \eta}{\partial t} - \frac{1}{e_1 e_2 e_3} \frac{\partial}{\partial k} \left( \kappa \frac{e_1 e_2}{e_3} \frac{\partial \eta}{\partial k} \right) = 0 \quad (4.84)$$

where  $\eta = \eta(z(k), t)$  at a fixed point  $(i, j)$ . The diffusion coefficient  $\kappa$  and metric coefficients  $e_i$ ,  $i = 1, \dots, 3$ , are, in general, functions of  $k$  so must be kept within the second-derivative.

### 4.5.2 Discretization and solution algorithm

The  $M$ -step implicit diffusion operator (Equation (4.40)) requires solving the linear system of equations

$$\begin{aligned} \mathcal{L}_i^{-1}[\eta(z(k), t_1)] &= \eta(z(k), t_0), \\ \mathcal{L}_i^{-1}[\eta(z(k), t_2)] &= \eta(z(k), t_1), \\ &\vdots \\ \mathcal{L}_i^{-1}[\eta(z(k), t_M)] &= \eta(z(k), t_{M-1}) \end{aligned}$$

where, from Equation (4.84),

$$\mathcal{L}_i^{-1} = 1 - \frac{\Delta t}{e_1 e_2 e_3} \frac{\partial}{\partial k} \left( \kappa \frac{e_1 e_2}{e_3} \frac{\partial}{\partial k} \right). \quad (4.85)$$

The  $\mathcal{L}_i^{-1}$  operator is self-adjoint with respect to the inner product  $\iiint \eta_1 \eta_2 e_1 e_2 e_3 di dj dk$ .

Now consider a centred finite-difference representation of the one-step equation

$$\mathcal{L}_i^{-1}[\eta_k^m] = \eta_k^{m-1} \quad (4.86)$$

where  $\eta_k^m = \eta(z_k, t_m)$  and  $z_k = z(k)$ . The  $\eta_k^m$  are assumed to be defined at the mid-points of the spatial grid cells where  $k$  is taken to have an integer value. First-order (centred) derivatives of  $\eta_k^m$  are defined at half-integer points ( $k + \frac{1}{2}$ ). The curvilinear distance  $\Delta s_3$  is the distance from one grid-point or one half grid-point to the next. Since  $\Delta k = 1$ ,  $\Delta s_3$  is simply equal to the metric coefficient :  $\Delta s_3 = e_3 \Delta k = \partial z / \partial k$ .

Solutions to Equation (4.86) are more conveniently computed numerically using a symmetric, rather than self-adjoint, form of the matrix-operator on the left-hand side. The left-hand side is readily made symmetric by multiplying both sides of the equation by  $w_k = e_{1,k} e_{2,k} e_{3,k}$ . The finite-difference representation of Equation (4.86) can then be written as

$$\begin{aligned} w_k \eta_k^{m-1} &= -\alpha_{k-\frac{1}{2}} \eta_{k-1}^m - \alpha_{k+\frac{1}{2}} \eta_{k+1}^m \\ &+ \left( w_k + \alpha_{k-\frac{1}{2}} + \alpha_{k+\frac{1}{2}} \right) \eta_k^m \end{aligned} \quad (4.87)$$

where

$$\alpha_{k+\frac{1}{2}} = \kappa_{k+\frac{1}{2}} \Delta t \left( \frac{e_{1,k+\frac{1}{2}} e_{2,k+\frac{1}{2}}}{e_{3,k+\frac{1}{2}}} \right), \quad (4.88)$$

with  $L_{k+\frac{1}{2}}^2 = \kappa_{k+\frac{1}{2}} \Delta t$ . The subscripts  $k$  and  $k \pm \frac{1}{2}$  indicate that parameters are evaluated at either  $z_k$  or  $z_{k \pm \frac{1}{2}}$ .

Consider solutions to Equation (4.87) on the finite interval  $k = 1, \dots, K-1$ , with solid boundaries assumed at both end-points. The left and right edges of the boundary are defined at the half points  $k = \frac{1}{2}$  and  $k = K - \frac{1}{2}$ . For the BCs that need to be supplied to this problem, it is sufficient to consider the general Robin BCs (Equations (4.81) and (4.82)) which include the Neumann and Dirichlet BCs as limiting cases. At the boundary points  $k = \frac{1}{2}$  and  $k = K - \frac{1}{2}$ , the finite difference form of Equations (4.81) and (4.82) reads

$$\mu \left( \frac{\eta_1^m + \eta_0^m}{2} \right) - (1 - \mu) \bar{e}_{\frac{1}{2}} \left( \frac{\eta_1^m - \eta_0^m}{e_{3,\frac{1}{2}}} \right) = 0, \quad (4.89)$$

$$\mu \left( \frac{\eta_K^m + \eta_{K-1}^m}{2} \right) + (1 - \mu) \bar{e}_{K-\frac{1}{2}} \left( \frac{\eta_K^m - \eta_{K-1}^m}{e_{3,K-\frac{1}{2}}} \right) = 0. \quad (4.90)$$

For convenience, we set  $\bar{e}_{\frac{1}{2}} = e_{3,\frac{1}{2}}$  and  $\bar{e}_{K-\frac{1}{2}} = e_{3,K-\frac{1}{2}}$ . Equations (4.89) and (4.90) can then be rearranged to give the following expressions for the fictitious points  $\eta_0^m$  and  $\eta_K^m$  within the boundaries :

$$\eta_0^m = (1 - \bar{\epsilon}) \eta_1^m, \quad (4.91)$$

$$\eta_K^m = (1 - \bar{\epsilon}) \eta_{K-1}^m, \quad (4.92)$$

where

$$\bar{\epsilon} = \frac{2\mu}{\mu + 2(1 - \mu)}.$$

Since  $0 \leq \mu \leq 1$ , the parameter  $\bar{\epsilon}$  is bounded below by 0 (Neumann) and above by 2 (Dirichlet).

From Equation (4.87), we have

$$\begin{aligned} w_1 \eta_1^{m-1} &= -\alpha_{\frac{1}{2}} \eta_0^m - \alpha_{\frac{3}{2}} \eta_2^m \\ &\quad + \left( w_1 + \alpha_{\frac{1}{2}} + \alpha_{\frac{3}{2}} \right) \eta_1^m, \\ w_{K-1} \eta_{K-1}^{m-1} &= -\alpha_{K-\frac{3}{2}} \eta_{K-2}^m - \alpha_{K-\frac{1}{2}} \eta_K^m \\ &\quad + \left( w_{K-1} + \alpha_{K-\frac{3}{2}} + \alpha_{K-\frac{1}{2}} \right) \eta_{K-1}^m, \end{aligned}$$

which after inserting Equations (4.91) and (4.92) become

$$\begin{aligned} w_1 \eta_1^{m-1} &= -\alpha_{\frac{3}{2}} \eta_2^m \\ &\quad + \left( w_1 + \bar{\varepsilon} \alpha_{\frac{1}{2}} + \alpha_{\frac{3}{2}} \right) \eta_1^m, \\ w_{K-1} \eta_{K-1}^{m-1} &= -\alpha_{K-\frac{3}{2}} \eta_{K-2}^m \\ &\quad + \left( w_{K-1} + \alpha_{K-\frac{3}{2}} + \bar{\varepsilon} \alpha_{K-\frac{1}{2}} \right) \eta_{K-1}^m. \end{aligned}$$

In practice, it is convenient to introduce masking parameters at the integer and half-integer points :

$$\varepsilon_k = \begin{cases} 1 & \text{at active points : } k = 1, \dots, K-1, \\ 0 & \text{within boundaries : } k = 0 \text{ and } k = K, \end{cases}$$

and

$$\varepsilon_{k+\frac{1}{2}} = \begin{cases} 1 & \text{at active points : } k = 1, \dots, K-2, \\ \bar{\varepsilon} & \text{on boundary edges : } k = 0 \text{ and } k = K-1, \end{cases}$$

and to redefine  $\eta_k^m$  and  $\alpha_{k+\frac{1}{2}}$  according to

$$\begin{aligned} \eta_k^m &\longrightarrow \varepsilon_k \eta_k^m, \\ \alpha_{k+\frac{1}{2}} &\longrightarrow \varepsilon_{k+\frac{1}{2}} \alpha_{k+\frac{1}{2}}, \end{aligned}$$

so that the BCs can be enforced directly in the symmetric equation (4.87).

Equation (4.87) forms a symmetric tridiagonal matrix  $\mathbf{A}$  with diagonal elements  $a_{k,k} = w_k + \alpha_{k-\frac{1}{2}} + \alpha_{k+\frac{1}{2}}$  and non-zero off-diagonal elements  $a_{k,k+1} = a_{k+1,k} = -\alpha_{k+\frac{1}{2}}$ . The diagonal is strictly dominant,

$$|a_{k,k}| > \sum_{l=1, l \neq k}^{K-1} |a_{k,l}| \quad \text{for all } k,$$

thus guaranteeing the positive definiteness and invertibility of  $\mathbf{A}$  (Golub et van Loan, 1996, Section 4.2). Equation (4.86) can be written as

$$\mathbf{A} \boldsymbol{\eta}^m = \mathbf{W} \boldsymbol{\eta}^{m-1} \quad (4.93)$$

where  $\boldsymbol{\eta}^m = (\eta_1^m, \dots, \eta_{K-1}^m)^\top$  and  $\mathbf{W} = \text{diag}(w_1, \dots, w_{K-1})$ . For the 1D problem, the size of  $\mathbf{A}$  is generally small enough that it can be inverted using a direct method that requires storing the (non-zero) elements of  $\mathbf{A}$ . An algorithm based on Cholesky factorization is a natural choice for a symmetric and positive definite matrix (Golub et van Loan, 1996, Section 4.2.3). From the Cholesky factorization of  $\mathbf{A}$ , we can write  $\mathbf{A} = \mathbf{G} \mathbf{G}^\top$  where  $\mathbf{G}$  ( $\mathbf{G}^\top$ ) is a lower (upper) triangular matrix. The matrix can then be inverted efficiently in two steps : first using forward elimination

$$\mathbf{G} \boldsymbol{\xi} = \mathbf{W} \boldsymbol{\eta}^{m-1} \quad (4.94)$$

to solve for the intermediate variable  $\xi$ , and then using backward substitution

$$\mathbf{G}^T \boldsymbol{\eta}^m = \xi \quad (4.95)$$

to solve for  $\boldsymbol{\eta}^m$ .

The Cholesky algorithm has been used for the examples presented in this paper. Alternatively, Equation (4.93) can be solved approximately using iterative methods such as conjugate gradient (Golub et van Loan, 1996, Section 10.2) or multi-grid (Briggs *et al.*, 2000). Iterative methods would be preferable for large-matrix applications that arise, for example, in 2D or 3D formulations of the implicit diffusion equation (Weaver et Ricci, 2004). Even for the 1D problem, iterative methods may be better suited for massively parallel implementations with certain models or computer architectures.

### 4.5.3 Link with the recursive filter

The implicit diffusion algorithm presented here is closely related to the first-order recursive filter (Raymond et Garder, 1991; Lorenc, 1992; Hayden et Purser, 1995) which has been developed extensively for modelling correlation functions in atmospheric data assimilation (see P03a and P03b and references therein).

To illustrate the relationship, we consider the simplified case where the diffusion coefficients  $\kappa$  are constant, the underlying grid is uniform with spacing  $\Delta s_3 = \Delta z$ , and the domain is infinite to avoid the complicating effects of BCs. Under these conditions, the one-step implicit diffusion filter described by Equations (4.87) and (4.88) reduces to

$$\eta_k^{m-1} = \eta_k^m - \frac{\kappa \Delta t}{(\Delta z)^2} (\eta_{k-1}^m - 2\eta_k^m + \eta_{k+1}^m). \quad (4.96)$$

The first-order recursive filter is described by the pair of forward and backward equations

$$q_k = (1 - \tilde{\alpha}) p_k + \tilde{\alpha} q_{k-1} \quad \text{for } k = 1, \dots, K-1, \quad (4.97)$$

$$s_k = (1 - \tilde{\alpha}) q_k + \tilde{\alpha} s_{k+1} \quad \text{for } k = K-1, \dots, 1, \quad (4.98)$$

where the filter coefficient  $\tilde{\alpha}$  satisfies  $0 < \tilde{\alpha} < 1$  (cf. Equations (3.17) and (3.18) in P03a with  $n = 1$ ). The quantity  $p_k$  represents the initial input to the filter at grid point  $k$  (cf.  $\eta_k^{m-1}$  in Equation (4.86)) and  $s_k$  is the value at the same point after the filter has been applied in both forward and backward directions (cf.  $\eta_k^m$  in Equation (4.86)). The forward and backward filters are analogous to the forward elimination and backward substitution steps of the Cholesky algorithm (Equations (4.94) and (4.95)).

To determine the relationship between the parameters  $\kappa$  and  $\tilde{\alpha}$ , we can eliminate the intermediate variable  $q_k$  in Equations (4.97) and (4.98) to derive an expression for the inverse of the recursive filter,

$$p_k = s_k - \frac{\tilde{\alpha}}{(1 - \tilde{\alpha})^2} (s_{k-1} - 2s_k + s_{k+1}),$$

which is equivalent to the one-step implicit diffusion equation (4.96) if

$$\frac{\kappa \Delta t}{(\Delta z)^2} = \frac{\tilde{\alpha}}{(1 - \tilde{\alpha})^2}. \quad (4.99)$$

Applying the recursive filter  $M$  times (in both directions) is then clearly equivalent to  $M$  iterations of an implicit diffusion operator, i.e., to convolving a field with an  $M$ -th order AR function. From Equations (4.51) and (4.99),

$$\tilde{\alpha} = 1 + \frac{1}{2} \left( \frac{\Delta z}{L} \right)^2 - \frac{\Delta z}{L} \sqrt{1 + \frac{1}{4} \left( \frac{\Delta z}{L} \right)^2}$$

which can be used to match the scale response of the recursive filter to the scale parameter  $L$  of the AR functions. In terms of  $\tilde{\alpha}$ , the normalization factor is

$$\lambda_M = v_M \left( \frac{\sqrt{\tilde{\alpha}}}{1 - \tilde{\alpha}} \right) \Delta z,$$

where  $v_M$  is given by Equation (4.45).

Lorenz (1992) and P03a provide illustrations of the unnormalized response of the recursive filter for different values of  $M$ . The main differences between the implicit diffusion algorithm presented in this paper and the first-order recursive filter are in the treatment of the BCs in a finite domain. BCs for the recursive filter are discussed in detail in Hayden et Purser (1995) and P03a.

## 4.6 Summary and discussion

Diffusion-based correlation operators are efficient mathematical tools for evaluating correlation matrix-vector products (integral equations) such as those required in the iterative minimization algorithms of variational assimilation. This paper has discussed correlation operators derived from an *implicit* formulation of the 1D diffusion equation. By analyzing the correlation operators formed from kernels belonging to the well-known class of homogeneous and isotropic AR functions, it was shown that the *inverse* of the AR correlation operators could be matched to a diffusion equation integrated in reverse “time” with an implicit scheme. This result was obtained by considering only the temporal discretization of the diffusion equation. A forward integration of the diffusion equation with an implicit scheme requires the inversion of a matrix resulting from the discretization of the spatial derivative term. The solution of this forward problem can be used to model the AR correlation operator itself. The number of pseudo-time steps ( $M$ ) of the implicit scheme corresponds to the order of the AR function. The square-root of the product of the diffusion coefficient ( $\kappa$ ) and the pseudo-time step ( $\Delta t$ ) sets the scale parameter of the AR function. The normalization factor required to produce a unit-amplitude function was shown to have a more complicated relationship involving both  $M$  and  $\kappa\Delta t$ .

The classical solution of the constant-coefficient diffusion equation can be expressed, after normalization, as a correlation operator with a Gaussian kernel. For a fixed-length scale, the AR correlation operator modelled with the implicit scheme converges to the Gaussian correlation operator as  $M \rightarrow \infty$ . The Gaussian correlation operator is the only correlation operator that can be modelled with an explicit scheme. The choice of implicit scheme is thus important for enlarging the class of correlation functions that can be represented by the diffusion equation. Contrary to the explicit scheme, the implicit scheme is unconditionally stable. This is reflected by the additional free parameter  $M$  in the implicit model; in the explicit model  $M$  must be fixed at a value large enough to respect a stability criterion that depends on the square of the ratio of the length scale and grid size. The property of unconditional stability is thus particularly important to keep the computational cost of the diffusion model acceptable when local length scales are large compared to the local grid resolution.

With variable diffusion coefficients, the length scales can be made location dependent. An important consequence of varying the length scales is that the normalization factors are no longer constant and must be estimated using a separate algorithm. As an alternative to the randomization method proposed by WC01, two simpler and less costly methods were tested for approximating these factors when the length scales vary slowly and smoothly. In one approach, the diffusion coefficient appearing in the formula for the constant normalization factor is simply replaced by the spatially varying coefficient (Pannekoucke et Massart, 2008). In the other approach, the spatially varying diffusion coefficient is smoothed using the square-root of the implicit diffusion filter before being used in the same formula (P03b). For the examples considered in this paper, the method of P03b was superior, giving a very

good estimation ( $< 2\%$  error) with only a small cost overhead.

The effects of BCs on the correlation functions induced by the implicit diffusion equation were studied. Both Neumann and Dirichlet BCs result in a distortion of the AR functions near the boundaries. Robin BCs could be used with the implicit diffusion equation to achieve a good fit to a SOAR function near boundaries, but were inadequate for fitting higher order AR functions. A more robust method was presented for simulating transparent boundaries which could be applied at the cost of an additional integration of the diffusion equation. By defining the correlation operator as the average of the solution of two diffusion problems, one employing Neumann BCs and the other employing Dirichlet BCs, the AR response of the implicit diffusion equation could be recovered everywhere in the domain. Importantly, this method can be used in combination with the P03b technique to provide accurate estimates of the normalization factors with location-dependent length scales as well as in the presence of boundaries.

General correlation operators in higher dimensions can be constructed from products of 1D correlation operators as discussed by (Stein, 1999, p. 54–55) in the context of geostatistics, and by Wu *et al.* (2002) and P03a,b in the context of atmospheric data assimilation with the recursive filter. In this respect, the results developed in this paper for the 1D implicit diffusion operator have much wider applications. With few iterations, however, the product of 1D correlation operators leads to a spurious anisotropic response as illustrated by P03a with the recursive filter. Only by increasing the number of iterations to produce a more Gaussian-like response can these numerical artifacts be suppressed. For example, about  $M = 10$  iterations are required with the implicit diffusion operator to obtain an adequate 2D (or 3D) isotropic response. The only satisfactory way to model isotropic, fat-tailed (low-order AR-like) correlation functions with a grid-point filter is to solve a 2D implicit diffusion equation directly, as illustrated by Weaver et Ricci (2004). In doing so, however, we must sacrifice the computational efficiency of the 1D implicit diffusion operator in favour of a generally less efficient and more cumbersome 2D implicit diffusion operator.

The choice of an appropriate correlation model for data assimilation is determined by flexibility considerations, on the one hand, and computational efficiency, on the other. A quasi-Gaussian model built from products of 1D implicit diffusion operators (high-order AR functions) may have limited flexibility in its spectral response at high wavenumbers but can still support location- and flow-dependent length scales in a computationally efficient way. Capturing geographical variations in the correlation length scale is as important as capturing the spectral characteristics of the correlations. To achieve fully anisotropic correlations requires the use of a non-diagonal diffusion tensor (WC01; P03b; Purser, 2005) to allow for preferential stretching or shrinking of local length scales in directions other than those corresponding to the computational coordinates. P03b and Purser (2005) outline the theoretical framework for generating anisotropic correlations with a quasi-Gaussian model formulated using 1D smoothing operators. Applications of the method in meteorology and oceanography are presented in Liu *et al.* (2007) and Liu *et al.* (2009), respectively.

The methods outlined in this paper are being used in the construction of a 3D background-error correlation model for a variational ocean data assimilation system (Weaver *et al.*, 2005; Mogensen *et al.*, 2009) for the NEMO (Nucleus for European Modelling of the Ocean; Madec, 2008) model. Those developments will be reported in a future paper.

**Acknowledgment :** This work was supported by the French National Research Agency (ANR) through the COSINUS programme (project VODA n° ANR-08-COSI-016). Additional support was obtained from the LEFE-ASSIM programme. We would like to thank two anonymous reviewers for pointing out several important references in geostatistics which helped us present our work in much broader context.

## 4.A Matérn and AR functions

A widely-used class of correlation functions in spatial statistical modelling is the so-called Matérn family (Stein, 1999; Guttorp et Gneiting, 2006) :

$$\zeta_\nu(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{L}\right)^\nu K_\nu\left(\frac{r}{L}\right), \quad (4.100)$$

where  $\Gamma(\nu)$  is the Gamma function,  $K_\nu(x)$  is the modified Bessel function defined for  $x > 0$ ,  $r$  is the Euclidean distance between two points in  $\mathbb{R}^n$ , and  $L$  and  $\nu$  are length scale and smoothness parameters. While this correlation family is defined for arbitrary dimension  $n$ , we focus here on the 1D case for which  $r = |z - z'|$ .

Consider the special case  $\nu = M - 1/2$  where  $M$  is a strictly positive integer. The Gamma and modified Bessel functions are then (Jeffrey, 1995, p. 219 and p. 274)

$$\begin{aligned} \Gamma(M-1/2) &= \frac{(2M-3)!!}{2^{M-1}} \sqrt{\pi} \\ &= \frac{1 \times 3 \times \cdots \times (2M-3)}{2^{M-1}} \sqrt{\pi} \end{aligned} \quad (4.101)$$

and

$$K_{M-1/2}(x) = \sqrt{\frac{\pi}{2x}} e^{-x} \sum_{k=0}^{M-1} \frac{(M-1+k)!}{2^k k! (M-1-k)! x^k}. \quad (4.102)$$

Substituting Equations (4.101) and (4.102) in (4.100) gives

$$\begin{aligned} \zeta_{M-1/2}(r) &= \frac{1}{(2M-3)!!} \\ &\times \sum_{k=0}^{M-1} \frac{(M-1+k)!}{2^k k! (M-1-k)!} \left(\frac{r}{L}\right)^{M-1-k} e^{-\frac{r}{L}}. \end{aligned}$$

Making the change of variable  $j = M - k - 1$  yields

$$\begin{aligned} \zeta_{M-1/2}(r) &= \frac{1}{(2M-3)!!} \\ &\times \sum_{j=0}^{M-1} \frac{(2M-j-2)!}{2^{M-j-1} (M-j-1)! j!} \left(\frac{r}{L}\right)^j e^{-\frac{r}{L}}. \end{aligned} \quad (4.103)$$

Since

$$\begin{aligned} &\frac{1}{(2M-3)!!} \\ &= \frac{1}{1 \times 3 \times \cdots \times (2M-3)} \\ &= \frac{(2 \times 1)(2 \times 2) \cdots (2 \times (M-2))(2 \times (M-1))}{(2M-2)!} \\ &= \frac{2^{M-1} (M-1)!}{(2M-2)!}, \end{aligned} \quad (4.104)$$

Equation (4.103) reduces to

$$\zeta_{M-1/2}(r) = f_M(z - z')$$

as defined by Equations (4.41) and (4.42).

The FT of Equation (4.100) is ((Yaglom, 1987, p. 139, Eq. 2.140); (Gneiting *et al.*, 2009, p. 16, Eq. (20)))

$$\begin{aligned}\widehat{\zeta}_\nu(\hat{z}) &= 2\pi \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{2^{\nu-1} \Gamma(\nu+1/2) L^{-2\nu}}{\sqrt{\pi} [L^{-2} + \hat{z}^2]^{\nu+1/2}} \right) \\ &= \frac{2\sqrt{\pi} \Gamma(\nu+1/2)}{\Gamma(\nu)} \left( \frac{L}{[1 + (L\hat{z})^2]^{\nu+1/2}} \right)\end{aligned}$$

where the  $2\pi$  normalization of the FT pair has been defined in accordance with Equations (4.23) and (4.24). Setting  $\nu = M - 1/2$  for the AR functions gives

$$\widehat{\zeta}_{M-1/2}(\hat{z}) = \frac{2\sqrt{\pi} \Gamma(M)}{\Gamma(M-1/2)} \left( \frac{L}{[1 + (L\hat{z})^2]^M} \right).$$

From Equations (4.101) and (4.104), and noting that  $\Gamma(M) = (M - 1)!$ , the coefficient becomes

$$\frac{2\sqrt{\pi} \Gamma(M)}{\Gamma(M-1/2)} = \frac{2^{2M-1} [(M-1)!]^2}{(2M-2)!}$$

as in Equation (4.45). Hence,

$$\widehat{\zeta}_{M-1/2}(\hat{z}) = \widehat{f}_M(\hat{z})$$

as defined by Equations (4.43) and (4.44).

## 4.B Recovering the Gaussian solution near boundaries

In this appendix we show how the Gaussian response of the diffusion equation in an infinite domain can be recovered everywhere in a finite domain by averaging the solutions of two diffusion problems with different boundary conditions. Recall that the solution of the 1D diffusion equation in an unbounded domain is given by the convolution integral in Equation (4.31). The response of this convolution integral to a Dirac delta function,  $\eta(z, 0) = \delta(z - z_0)$ , at  $z = z_0$  is the Gaussian function

$$\eta_g(z, t) = \frac{1}{\sqrt{4\pi\kappa t}} e^{-(z-z_0)^2/4\kappa t}. \quad (4.105)$$

We examine now the response of the diffusion equation to a Dirac delta function in a bounded domain,  $0 \leq z \leq l$ . From the method of separation of variables, the general solution is (e.g., see Section 2.3 in Haberman, 1987)

$$\eta(z, t) = (\nu \cos(\xi z) + \mu \sin(\xi z)) e^{-\xi^2 \kappa t},$$

where the constants  $\nu$ ,  $\mu$  and  $\xi$  are determined from the initial and boundary conditions. First, consider the solution when the boundary conditions are of Neumann type,

$$\frac{\partial \eta}{\partial z} = 0 \quad \text{at } z = 0 \text{ and } z = l,$$

from which it is easy to verify that  $\mu = 0$  and  $\xi = n\pi/l$  with  $n = 0, 1, 2, \dots$ . The solution for the Neumann problem thus has the form

$$\eta_N(z, t) = \sum_{n=0}^{\infty} c_n \cos\left(\frac{n\pi}{l} z\right) e^{-(n\pi/l)^2 \kappa t}.$$

Since the initial condition must satisfy

$$\eta_{\text{N}}(z, 0) = \sum_{n=0}^{\infty} c_n \cos\left(\frac{n\pi}{l}z\right),$$

we can determine the Fourier coefficients  $c_n$  from the orthogonality of the cosines. This yields

$$\begin{aligned} c_0 &= \frac{1}{l} \int_0^l \eta_{\text{N}}(z, 0) dz, \\ c_n &= \frac{2}{l} \int_0^l \eta_{\text{N}}(z, 0) \cos\left(\frac{n\pi}{l}z\right) dz, \quad n = 1, 2, \dots \end{aligned}$$

Taking  $\eta_{\text{N}}(z, 0) = \delta(z - z_0)$ , where  $0 < z_0 < l$ , gives

$$\begin{aligned} c_0 &= \frac{1}{l}, \\ c_n &= \frac{2}{l} \cos\left(\frac{n\pi}{l}z_0\right), \quad n = 1, 2, \dots, \end{aligned}$$

and the solution becomes

$$\begin{aligned} \eta_{\text{N}}(z, t) &= \frac{1}{l} + \sum_{n=1}^{\infty} \frac{2}{l} \cos\left(\frac{n\pi}{l}z_0\right) \cos\left(\frac{n\pi}{l}z\right) \\ &\quad \times e^{-(n\pi/l)^2 \kappa t}, \end{aligned}$$

which using a standard trigonometric identity can be rewritten as

$$\begin{aligned} \eta_{\text{N}}(z, t) &= \frac{1}{2l} + \sum_{n=1}^{\infty} \frac{1}{l} e^{-(n\pi/l)^2 \kappa t} \cos\left(\frac{n\pi}{l}(z - z_0)\right) \\ &\quad + \frac{1}{2l} + \sum_{n=1}^{\infty} \frac{1}{l} e^{-(n\pi/l)^2 \kappa t} \cos\left(\frac{n\pi}{l}(z + z_0)\right). \end{aligned} \quad (4.106)$$

Now consider the solution to the problem when the boundary conditions are of Dirichlet type,

$$\eta = 0 \quad \text{at } z = 0 \text{ and } z = l,$$

which leads to  $\nu = 0$  and the same condition on  $\xi$  as in the Neumann problem. The solution is thus

$$\eta_{\text{D}}(z, t) = \sum_{n=0}^{\infty} d_n \sin\left(\frac{n\pi}{l}z\right) e^{-(n\pi/l)^2 \kappa t}$$

where

$$\eta_{\text{D}}(z, 0) = \sum_{n=0}^{\infty} d_n \sin\left(\frac{n\pi}{l}z\right),$$

and the Fourier coefficients  $d_n$  are determined from the orthogonality of the sines. This yields

$$\begin{aligned} d_0 &= 0, \\ d_n &= \frac{2}{l} \int_0^l \eta_{\text{D}}(z, 0) \sin\left(\frac{n\pi}{l}z\right) dz, \quad n = 1, 2, \dots \end{aligned}$$

Taking  $\eta_{\text{D}}(z, 0) = \delta(z - z_0)$ , where  $0 < z_0 < l$ , gives

$$d_n = \frac{2}{l} \sin\left(\frac{n\pi}{l}z_0\right),$$

and the solution becomes

$$\eta_{\text{D}}(z, t) = \sum_{n=1}^{\infty} \frac{2}{l} \sin\left(\frac{n\pi}{l} z_0\right) \sin\left(\frac{n\pi}{l} z\right) e^{-(n\pi/l)^2 \kappa t},$$

which using a standard trigonometric identity can be rewritten as

$$\begin{aligned} \eta_{\text{D}}(z, t) &= \frac{1}{2l} + \sum_{n=1}^{\infty} \frac{1}{l} e^{-(n\pi/l)^2 \kappa t} \cos\left(\frac{n\pi}{l}(z-z_0)\right) \\ &\quad - \frac{1}{2l} - \sum_{n=1}^{\infty} \frac{1}{l} e^{-(n\pi/l)^2 \kappa t} \cos\left(\frac{n\pi}{l}(z+z_0)\right). \end{aligned} \quad (4.107)$$

Consider again the Gaussian expression (4.105), but centered at  $z_0 = 0$ ,

$$h(z, t) = \frac{1}{\sqrt{4\pi\kappa t}} e^{-z^2/4\kappa t}.$$

Assuming the function is  $2l$ -periodic and that its length scale  $\sqrt{2\kappa t} \ll l$  so that the function effectively vanishes at  $z = \pm l$ , we can expand it as a Fourier series,

$$h(z, t) = a_0 + \sum_{n=1}^{\infty} \left[ a_n \cos\left(\frac{n\pi}{l} z\right) + b_n \sin\left(\frac{n\pi}{l} z\right) \right],$$

where the Fourier coefficients are given by

$$\begin{aligned} a_0 &= \frac{1}{2l} \int_{-l}^l h(z, t) dz, \\ a_n &= \frac{1}{l} \int_{-l}^l h(z, t) \cos\left(\frac{n\pi}{l} z\right) dz, \quad n = 1, 2, \dots, \\ b_n &= \frac{1}{l} \int_{-l}^l h(z, t) \sin\left(\frac{n\pi}{l} z\right) dz, \quad n = 1, 2, \dots \end{aligned}$$

As the Gaussian is an even function, the coefficients  $b_n = 0$  for all  $n$ . The coefficients  $a_0$  and  $a_n$  can be evaluated using Equation 3.894-4 on p. 480 in Gradshteyn et Ryzhik (1980) :

$$\begin{aligned} a_0 &= \frac{1}{2l}, \\ a_n &= \frac{1}{l} e^{-(n\pi/l)^2 \kappa t}, \quad n = 1, 2, \dots, \end{aligned}$$

which leads to

$$h(z, t) = \frac{1}{2l} + \sum_{n=1}^{\infty} \frac{1}{l} e^{-(n\pi/l)^2 \kappa t} \cos\left(\frac{n\pi}{l} z\right). \quad (4.108)$$

Comparing Equation (4.108) with Equations (4.106) and (4.107), we have

$$\begin{aligned} \eta_{\text{N}}(z, t) &= h(z-z_0, t) + h(z+z_0, t), \\ \eta_{\text{D}}(z, t) &= h(z-z_0, t) - h(z+z_0, t), \end{aligned}$$

and thus

$$h(z-z_0, t) = \frac{1}{2} (\eta_{\text{N}}(z, t) + \eta_{\text{D}}(z, t)) = \eta_{\text{g}}(z, t)$$

as in Equation (4.105). In other words, the Gaussian solution can be recovered everywhere in a finite domain by averaging the solutions obtained separately using Neumann and Dirichlet boundary conditions.

## Chapitre 5

# Constructions 2D et 3D à partir de l'opérateur de diffusion implicite 1D

A partir de l'opérateur de diffusion implicite 1D détaillé au chapitre 4, on construit ici des opérateurs 2 ou 3×1D qu'il est possible d'utiliser pour représenter les corrélations spatiales nécessaires à la modélisation des covariances d'erreur (ébauche et/ou modèle) dans un système d'assimilation de données variationnelle. Dans ce chapitre, ces constructions sont appliquées aux corrélations univariées de l'erreur d'ébauche dans le cadre réaliste du système océanique NEMOVAR (voir paragraphe 5.1).

Après en avoir détaillé la construction (paragraphe 5.2), l'opérateur de diffusion implicite 3×1D est comparé à l'opérateur explicite combinant une équation de diffusion 2D à l'horizontale et 1D à la verticale (paragraphe 5.3). La normalisation associée est donnée par la méthode par vecteur aléatoire. Au paragraphe 5.4, on propose d'examiner d'autres possibilités de normalisation.

### 5.1 Le système NEMOVAR

NEMOVAR (Mogensen *et al.*, 2009) est un système d'assimilation variationnelle pour le modèle océanique NEMO<sup>64</sup> (Madec, 2008) du LOCEAN<sup>65</sup>. Développé par le CERFACS<sup>66</sup> en collaboration avec le CEPMMT<sup>67</sup>, le UK Meteorological Office<sup>68</sup>, et l'INRIA<sup>69</sup>/LJK<sup>70</sup>, il fait suite au système OPAVAR (Weaver *et al.*, 2003; Vialard *et al.*, 2003; Weaver *et al.*, 2005; Daget *et al.*, 2009).

La méthode de discrétisation utilisée dans le modèle NEMO est basée sur les différences finies centrées de deuxième ordre, et les variables sont réparties selon la grille "C" de la classification d'Arakawa (Arakawa, 1972). La température, la salinité et la hauteur d'eau sont données au point T, au centre de la maille (voir figure 5.1). Les composantes "zonale" et "méridienne" du courant sont respectivement définies aux points U et V, soit au centre des faces droite et du fond. Le point W au centre de la face du haut est utile au traitements verticaux et le point F au milieu de l'arrête droite du fond, définit l'intersection des prolongements de U et V.

---

<sup>64</sup>Nucleus for European Modelling of the Ocean : <http://www.nemo-ocean.eu>

<sup>65</sup>Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques : <https://www.locean-ipsl.upmc.fr/index.php>

<sup>66</sup>Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique : <http://www.cerfacs.fr/>

<sup>67</sup>Centre Européen pour les Prévisions Météorologiques à Moyen Terme, ECMWF : European Centre for Medium-Range Weather Forecasts : <http://www.ecmwf.int>

<sup>68</sup><http://www.metoffice.gov.uk/>

<sup>69</sup>Institut National de recherche en Informatique et en Automatique : <http://www.inria.fr>

<sup>70</sup>Laboratoire Jean Kuntzmann : <http://www-ljk.imag.fr/>

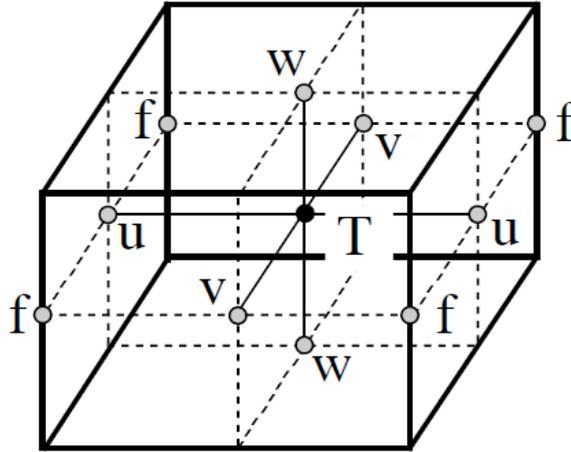


FIG. 5.1 – La grille “C” selon Arakawa définit pour une maille donnée : les points T au centre de la maille, U au centre de la face droite, V au centre de la face du fond, W au centre de la face du haut et F au milieu de l’arrête droite du fond (intersection des prolongements de U et V). Figure 3.1 de Madec (2008).

Le modèle NEMO offre plusieurs configurations possibles, que ce soit en terme de résolution, de régions (global, bassin, ...), de choix de paramètres physiques, de méthodes numériques ou de conditions aux frontières. Pour notre application, nous nous limiterons aux grilles ORCA2 (configuration globale à  $2^\circ$ , soit  $182 \times 149$  mailles horizontales sur 31 niveaux verticaux) et ORCA1 (configuration globale à  $1^\circ$ , soit  $362 \times 292$  mailles horizontales sur 42 niveaux verticaux), avec des conditions aux frontières est/ouest cyclique et recouvrement nord, qui permettent d’associer les côtés droit et gauche puis haut de la grille afin de modéliser une sphère. La singularité due au Pôle Nord est éliminée par construction de deux pôles, l’un sur l’Amérique du Nord, l’autre sur le continent Asiatique. Le pôle Sud étant défini sur l’Antarctique, aucun traitement spécifique n’est nécessaire. On notera que pour ces grilles, un raffinement méridien à l’équateur permet de mieux appréhender cette zone hautement énergétique. Le code est parallélisé en décomposant le domaine global en sous-domaines horizontaux (la verticale n’est pas décomposée) grâce à la librairie MPI<sup>71</sup>. Ces deux configurations à plutôt basse résolution sont principalement employées dans des applications relatives au climat. Elles permettent de fournir des conditions initiales pour des prévisions climatiques (mensuelles, saisonnières ou multi-annuelles), et océaniques, ou encore de reconstruire l’histoire de l’océan par des réanalyses multi-décennales.

NEMOVAR est conçu pour permettre une assimilation incrémentale 3D-Var ou 4D-Var. Pour cette dernière option, les routines adjointes du modèle sont actuellement développées sous l’appellation NEMOTAM. NEMOVAR comprend également un système de contrôle de qualité des observations (NEMOQC). Plusieurs études sont actuellement en cours pour permettre l’amélioration du système, comme la prise en compte de nouveaux types de données à assimiler (altimétrie, température de surface, courants, concentration de glace de mer), la correction des biais du modèle et des observations, les algorithmes de minimisation, ou enfin les opérateurs de covariances qui nous intéressent ici.

Les corrélations univariées de l’erreur d’ébauche  $y$  sont construites par les applications successives d’un opérateur de diffusion explicite 2D sur l’horizontale et 1D sur la verticale, suivies d’une normalisation issue de la méthode par vecteur aléatoire (Weaver et Courtier, 2001). Pour l’horizontale, l’échelle de corrélation est constante ( $4^\circ$  pour ORCA2,  $2^\circ$  pour ORCA1) sauf autour de l’équateur ( $\pm 15^\circ$ ) où elle est étirée dans le sens zonal et rétrécie dans le sens méridien. Cette configuration entraîne un nombre d’itérations de 200 pour respecter le critère de stabilité du schéma explicite (voir paragraphe 4.3.1). Pour la verticale, l’échelle est proportionnelle à la taille de la maille, et le nombre d’itérations est de

<sup>71</sup>Message Passing Interface.

10. Ces paramètres constituent la configuration “standard” mais peuvent bien sûr être modifiés. L'algorithme de minimisation (CONGRAD) est un gradient conjugué (Lanczos) avec un changement de variable préalable du type  $\mathbf{v} = \mathbf{U}^{-1}\delta\mathbf{x}$  (où  $\mathbf{B} = \mathbf{U}\mathbf{U}^T$  ; voir paragraphe 3.1.2).

Au cours de cette thèse, des expériences d'assimilation 3D-Var sont réalisées pour des données de température, salinité et hauteur d'eau (les données de courants ne sont pas assimilés dans NEMOVAR à l'heure actuelle). Pour modéliser les corrélations de la matrice  $\mathbf{B}$ , on construit des opérateurs de diffusion implicite 2D (hauteur d'eau) et 3D (température et salinité) à partir de l'opérateur de diffusion implicite 1D, tout en respectant les contraintes imposées par la parallélisation. Leurs performances sont comparées aux opérateurs explicites mentionnés plus haut. De plus, la définition des échelles de corrélation nécessitant d'être approfondie, on implémente la possibilité de prendre en compte les échelles déterminées pour OPAVAR (ORCA2 uniquement) à partir d'une méthode d'ensemble par Daget (2008). Enfin, on tente d'évaluer les différentes possibilités de normalisation, qu'il s'agisse de la méthode par vecteur aléatoire ou de l'approximation du facteur théorique. C'est dans ce cadre (entre autres), que le nouvel algorithme de minimisation CGMOD a été mis en place (S. Gratton, A. Piacentini, A. Weaver). Moins gourmand en mémoire suivant les options choisies, il utilise un préconditionnement par  $\mathbf{B}$  directement (voir paragraphe 3.1.2) et non plus le changement de variable impliquant  $\mathbf{U}$  nécessaire à CONGRAD. C'est cette dernière caractéristique qui nous intéressera ici pour des raisons que l'on détaillera au paragraphe 5.4.

## 5.2 Construction d'opérateurs de corrélation 2 et 3×1D

La construction en 2D ou 3D de fonctions de covariance (ou corrélation) à partir d'une fonction 1D a été largement étudiée en mathématiques comme en géostatistiques (voir par exemple Stein, 1999, p. 54-55 et les références citées). En assimilation de données, Wu *et al.* (2002) et Purser *et al.* (2003a,b) appliquent un filtre récursif 1D sur chaque axe pour construire des fonctions de corrélations 2D et 3D pour l'erreur d'ébauche. De telles constructions sont généralement plus simples à étudier et à mettre en œuvre que de véritables constructions multi-dimensionnelles. Néanmoins, leur forte dépendance aux axes présente des inconvénients dont il convient de tenir compte. Dans ce paragraphe, des opérateurs de corrélation 2D pour l'horizontale et 3D sont construits suivant le même principe mais à partir de l'opérateur de diffusion implicite 1D étudié au chapitre 4.

### 5.2.1 Fonctions de corrélation séparables

On suppose ici que la fonction de corrélation 3D que l'on cherche à modéliser est séparable suivant les axes  $x$ ,  $y$  et  $z$ , *i.e.* elle peut s'écrire sous la forme d'un produit de fonctions de corrélation 1D (Stein, 1999, p. 54) :

$$c(x, x'; y, y'; z, z') = c_x(x, x')c_y(y, y')c_z(z, z').$$

Chaque fonction de corrélation 1D peut alors être modélisée à l'aide de l'opérateur de diffusion implicite étudié au chapitre 4. Si on suppose de plus que le champ 3D que l'on cherche à corréler est également séparable (bien qu'il ne le soit pas en général)

$$\eta(x, y, z) = \eta_x(x)\eta_y(y)\eta_z(z),$$

l'opérateur de corrélation

$$\begin{aligned} \mathcal{C}[\eta(x, y, z)] &= \int \int \int c(x, x'; y, y'; z, z') \eta(x', y', z') dz' dy' dx' \\ &= \int c_x(x, x')\eta_x(x') dx' \int c_y(y, y')\eta_y(y') dy' \int c_z(z, z')\eta_z(z') dz', \end{aligned}$$

peut alors s'écrire

$$\begin{aligned}\mathcal{C}[\eta(x, y, z)] &= \sqrt{\lambda_x(x)} \mathcal{L}_x^M \left[ \sqrt{\lambda_x(x)} \eta_x(x) \right] \sqrt{\lambda_y(y)} \mathcal{L}_y^M \left[ \sqrt{\lambda_y(y)} \eta_y(y) \right] \sqrt{\lambda_z(z)} \mathcal{L}_z^M \left[ \sqrt{\lambda_z(z)} \eta_z(z) \right] \\ &= \sqrt{\lambda(x, y, z)} \mathcal{L}^M \left[ \sqrt{\lambda(x, y, z)} \eta(x, y, z) \right],\end{aligned}$$

avec

$$\mathcal{L}^M = \mathcal{L}_x^M \mathcal{L}_y^M \mathcal{L}_z^M,$$

et

$$\lambda(x, y, z) = \lambda_x(x) \lambda_y(y) \lambda_z(z).$$

$\mathcal{L}^M$  est l'opérateur de diffusion implicite 3×1D construit à partir des opérateurs de diffusion implicite 1D, avec  $M$  le nombre d'itérations. On notera qu'il est tout à fait possible de choisir un nombre d'itérations différent pour chaque opérateur 1D. Cependant, pour simplifier les notations, on suppose ici qu'ils utilisent tous le même nombre  $M$ .

On discrétise maintenant cet opérateur suivant les coordonnées curvilinéaires orthogonales  $(i, j, k)$  du modèle NEMO. Les opérateurs inverses 1D sont alors donnés par (voir l'exemple du paragraphe 4.5)

$$\mathcal{L}_x^{-1} = 1 - \frac{\Delta t}{e_1 e_2 e_3} \frac{\partial}{\partial i} \left( \kappa \frac{e_2 e_3}{e_1} \frac{\partial}{\partial i} \right) \quad \text{à } (j, k) \text{ fixé,} \quad (5.1)$$

$$\mathcal{L}_y^{-1} = 1 - \frac{\Delta t}{e_1 e_2 e_3} \frac{\partial}{\partial j} \left( \kappa \frac{e_1 e_3}{e_2} \frac{\partial}{\partial j} \right) \quad \text{à } (i, k) \text{ fixé,} \quad (5.2)$$

$$\mathcal{L}_z^{-1} = 1 - \frac{\Delta t}{e_1 e_2 e_3} \frac{\partial}{\partial k} \left( \kappa \frac{e_1 e_2}{e_3} \frac{\partial}{\partial k} \right) \quad \text{à } (i, j) \text{ fixé,} \quad (5.3)$$

où le produit  $e_1 e_2 e_3$  représente les éléments de volume locaux que l'on notera sous la forme de la matrice diagonale  $\mathbf{W} = \text{diag}(\dots, e_1(i, j, k) e_2(i, j, k) e_3(i, j, k), \dots)$ . On rappelle pour mémoire la définition des opérateurs explicite 2D horizontal et 1D vertical discrétisés suivant ces mêmes coordonnées curvilinéaires orthogonales :

$$\mathcal{L}_{exh} = 1 + \frac{\Delta t}{e_1 e_2 e_3} \left[ \frac{\partial}{\partial i} \left( \kappa \frac{e_2 e_3}{e_1} \frac{\partial}{\partial i} \right) + \frac{\partial}{\partial j} \left( \kappa \frac{e_1 e_3}{e_2} \frac{\partial}{\partial j} \right) \right] \quad \text{à } k \text{ fixé,} \quad (5.4)$$

$$\mathcal{L}_{exz} = 1 + \frac{\Delta t}{e_1 e_2 e_3} \frac{\partial}{\partial k} \left( \kappa \frac{e_1 e_2}{e_3} \frac{\partial}{\partial k} \right) \quad \text{à } (i, j) \text{ fixé.} \quad (5.5)$$

Dans cet espace, on définit le produit scalaire par  $\langle \eta_1, \eta_2 \rangle_{\mathbf{W}} = \int \eta_1 \eta_2 e_1 e_2 e_3 di dj dk$ . Pour la fonction coût discrétisée suivant ces mêmes coordonnées, la norme  $L^2$  associée au terme  $J_b$  et découlant de ce produit scalaire (pondéré par  $\mathbf{B}^{-1}$ ) doit également tenir compte des éléments de volume, et par conséquence, la modélisation de  $\mathbf{B}$  doit comprendre une application de  $\mathbf{W}^{-1}$ . Les covariances d'erreur d'ébauche discrétisées selon cette grille curvilinéaire s'écrivent alors

$$\mathbf{B}_{W^{-1}} = \mathbf{K} \mathbf{D}^{1/2} \mathbf{C}_{W^{-1}} \mathbf{D}^{1/2} \mathbf{K}^T,$$

avec  $\mathbf{K}$  l'opérateur d'équilibre,  $\mathbf{D}$  la matrice diagonale des variances, et  $\mathbf{C}_{W^{-1}}$  la matrice bloc diagonale des corrélations univariées, où chaque bloc  $\mathbf{C}$  est donné pour une variable particulière par

$$\mathbf{C} = \mathbf{\Lambda}^{1/2} \mathbf{L}^M \mathbf{W}^{-1} \mathbf{\Lambda}^{1/2}, \quad (5.6)$$

avec

$$\mathbf{L}^M = \mathbf{L}_x^M \mathbf{L}_y^M \mathbf{L}_z^M, \quad (5.7)$$

l'opérateur linéaire résultant de la discrétisation spatiale de chaque opérateur 1D (voir paragraphe 4.5.2).

Les opérateurs inverses donnés par les équations (5.1-5.3) (ainsi que les opérateurs explicites donnés par les équations (5.4) et (5.5)) sont autoadjoints par rapport au produit scalaire  $\langle \eta_1, \eta_2 \rangle_{\mathbf{W}} = \int \eta_1 \eta_2 e_1 e_2 e_3 di dj dk$ . On en rappelle ici la preuve pour l'exemple de  $\mathcal{L}_x^{-1}$ .

**Preuve:**

$$\begin{aligned} \langle \eta_1, \mathcal{L}_x^{-1}[\eta_2] \rangle_{\mathbf{W}} &= \int_{\mathbb{R}^3} \eta_1 \left[ \eta_2 - \frac{\Delta t}{e_1 e_2 e_3} \frac{\partial}{\partial i} \left( \kappa \frac{e_2 e_3}{e_1} \frac{\partial \eta_2}{\partial i} \right) \right] e_1 e_2 e_3 di dj dk \\ &= \int_{\mathbb{R}^3} \eta_1 \eta_2 e_1 e_2 e_3 di dj dk - \int_{\mathbb{R}^3} \eta_1 \Delta t \left[ \frac{\partial}{\partial i} \left( \kappa \frac{e_2 e_3}{e_1} \frac{\partial \eta_2}{\partial i} \right) \right] di dj dk. \end{aligned}$$

En intégrant par partie la seconde intégrale, on obtient :

$$\begin{aligned} &\int \eta_1 \Delta t \left[ \frac{\partial}{\partial i} \left( \kappa \frac{e_2 e_3}{e_1} \frac{\partial \eta_2}{\partial i} \right) \right] di dj dk \\ &= \left[ \eta_1 \Delta t \kappa \frac{e_2 e_3}{e_1} \frac{\partial \eta_2}{\partial i} \right]^\Gamma - \int \Delta t \kappa \frac{\partial \eta_1}{\partial i} \frac{e_2 e_3}{e_1} \frac{\partial \eta_2}{\partial i} di dj dk \\ &= \left[ \Delta t \kappa \frac{e_2 e_3}{e_1} \eta_1 \frac{\partial \eta_2}{\partial i} \right]^\Gamma - \left[ \Delta t \kappa \frac{e_2 e_3}{e_1} \frac{\partial \eta_1}{\partial i} \eta_2 \right]^\Gamma + \int \left[ \frac{\partial}{\partial i} \left( \Delta t \kappa \frac{e_2 e_3}{e_1} \frac{\partial \eta_1}{\partial i} \right) \right] \eta_2 di dj dk \\ &= \left[ \Delta t \kappa \frac{e_2 e_3}{e_1} \left( \eta_1 \frac{\partial \eta_2}{\partial i} - \eta_2 \frac{\partial \eta_1}{\partial i} \right) \right]^\Gamma + \int \Delta t \left[ \frac{\partial}{\partial i} \left( \kappa \frac{e_2 e_3}{e_1} \frac{\partial \eta_1}{\partial i} \right) \right] \eta_2 di dj dk. \end{aligned}$$

Lorsque les conditions sur la frontière  $\Gamma$  du domaine sont de type Neumann ( $\partial \eta_1 / \partial i = \partial \eta_2 / \partial i = 0$ ) ou Dirichlet ( $\eta_1 = \eta_2 = 0$ ), on a alors

$$\eta_1 \frac{\partial \eta_2}{\partial i} - \eta_2 \frac{\partial \eta_1}{\partial i} = 0. \quad (5.8)$$

Lorsque les conditions aux frontières sont de type Robin, on a

$$\begin{aligned} \mu \eta_1 - (1 - \mu) \bar{e} \frac{\partial \eta_1}{\partial i} &= 0, \\ \mu \eta_2 - (1 - \mu) \bar{e} \frac{\partial \eta_2}{\partial i} &= 0, \end{aligned}$$

où  $\mu$  est tel que  $0 < \mu < 1$ , et  $\bar{e}$  un paramètre d'échelle permettant de rendre les dimensions cohérentes dans les expressions précédentes. En multipliant la première expression par  $\eta_2$  et la seconde par  $\eta_1$ , puis en les soustrayant, on obtient

$$\begin{aligned} \mu \eta_1 \eta_2 - (1 - \mu) \bar{e} \eta_2 \frac{\partial \eta_1}{\partial i} - \mu \eta_1 \eta_2 + (1 - \mu) \bar{e} \eta_1 \frac{\partial \eta_2}{\partial i} &= 0 \\ (1 - \mu) \bar{e} \left( \eta_1 \frac{\partial \eta_2}{\partial i} - \eta_2 \frac{\partial \eta_1}{\partial i} \right) &= 0, \end{aligned}$$

et l'on retrouve l'équation (5.8). Au final, on a donc

$$\begin{aligned} \langle \eta_1, \mathcal{L}_x^{-1}[\eta_2] \rangle_{\mathbf{W}} &= \int \eta_1 \eta_2 e_1 e_2 e_3 di dj dk - \int \Delta t \left[ \frac{\partial}{\partial i} \left( \kappa \frac{e_2 e_3}{e_1} \frac{\partial \eta_1}{\partial i} \right) \right] \eta_2 di dj dk \\ &= \int \left[ \eta_1 - \frac{\Delta t}{e_1 e_2 e_3} \frac{\partial}{\partial i} \left( \kappa \frac{e_2 e_3}{e_1} \frac{\partial \eta_1}{\partial i} \right) \right] \eta_2 e_1 e_2 e_3 di dj dk \\ &= \langle \mathcal{L}_x^{-1}[\eta_1], \eta_2 \rangle_{\mathbf{W}}. \end{aligned}$$

La formulation donnée ici étant séparable, les opérateurs 1D commutent et l'opérateur 3×1D est alors

lui-même autoadjoint. On a donc, d'après l'équation (A.2) :

$$\mathbf{L} = \mathbf{L}^* = \mathbf{W}^{-1} \mathbf{L}^T \mathbf{W}.$$

On remarque également que

$$(\mathbf{L}^*)^M = \underbrace{\mathbf{L}^* \times \cdots \times \mathbf{L}^*}_M = \underbrace{(\mathbf{L} \times \cdots \times \mathbf{L})^*}_M = (\mathbf{L}^M)^*,$$

et on a alors

$$\mathbf{L}^M = (\mathbf{L}^*)^M = \mathbf{W}^{-1} (\mathbf{L}^T)^M \mathbf{W} = \mathbf{W}^{-1} (\mathbf{L}^M)^T \mathbf{W}. \quad (5.9)$$

L'opérateur de diffusion implicite horizontale  $2 \times 1D$  est construit sur les mêmes bases mais en retirant la dimension verticale. Dans la suite du paragraphe 5.2, on se limite à cet opérateur  $2 \times 1D$  pour illustrer nos propos.

### 5.2.2 Sensibilité au nombre d'itérations

Dans le cadre du filtre récursif, Purser *et al.* (2003a) montrent qu'un faible nombre d'applications conduit à une réponse altérée par une anisotropie parasite. Le même genre de phénomène est observé avec l'équation de diffusion. La figure 5.2 montre un exemple de l'application d'un opérateur  $2 \times 1D$  sur une grille régulière de résolution 1 point, et avec une échelle de corrélation constante de  $L_{AR} = 20$  points pour chaque axe (le paramètre d'échelle  $L = \sqrt{\kappa \Delta t}$  est ajusté en fonction de  $M$  pour maintenir constante cette échelle de corrélation à l'aide de l'équation (4.54)). La normalisation est donnée par un algorithme de normalisation exacte. Lorsque le nombre d'itérations est faible ( $M = 2$ ; a), une altération apparaît et s'atténue lorsque l'on itère davantage ( $M = 10$ ; b).

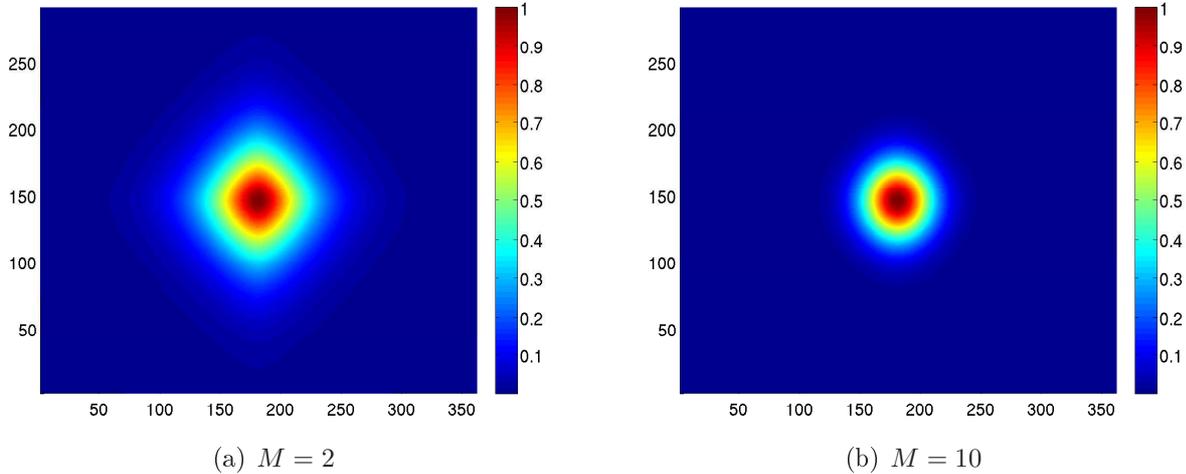


FIG. 5.2 – Exemple de fonctions de corrélation obtenues par l'application d'un opérateur de diffusion implicite  $2 \times 1D$  pour une échelle constante  $L_{AR} = 20$  sur chaque axe. La normalisation est donnée par la méthode exacte. Lorsque le nombre d'itérations  $M$  est faible (a), la fonction présente une anisotropie parasite qui s'atténue en itérant davantage (b).

En raison de cette anisotropie parasite qui apparaît lorsque  $M$  est faible, il n'est pas possible de modéliser des fonctions dont le coefficient d'aplatissement est important (fonctions AR d'ordre faible). En revanche, lorsque  $M$  est grand, chaque fonction de corrélation issue de l'application d'un opérateur de diffusion implicite  $1D$  tend vers la gaussienne, et ainsi, la fonction de corrélation correspondant à l'opérateur  $2 \times 1D$  tend elle-même vers une fonction gaussienne  $2D$  :

$$g(x, y) = e^{-x^2/2L_g^2} \times e^{-y^2/2L_g^2} = e^{-(x^2+y^2)/2L_g^2}.$$

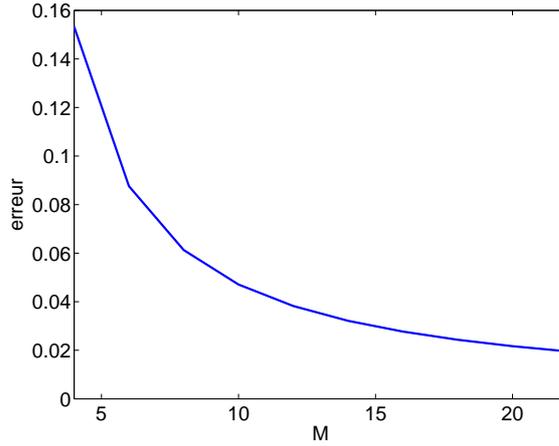


FIG. 5.3 – Erreur suivant l'équation (5.10) en fonction du nombre d'itérations du même opérateur 2×1D que pour la figure (5.2). Pour  $M = 10$ , l'erreur est inférieure à 5%.

Malgré la possibilité qu'offre un opérateur de diffusion implicite 1D de modéliser des fonctions AR, nous ne chercherons à représenter que des fonctions gaussiennes avec les opérateurs de diffusion implicite 2 et 3×1D. La figure 5.3 montre alors dans les mêmes conditions que pour la figure 5.2, l'erreur de modélisation en fonction de  $M$ . Cette erreur est donnée par

$$err = \frac{\|\mathcal{C}[\delta(x - x_0, y - y_0)] - g(x - x_0, y - y_0)\|_2}{\|g(x - x_0, y - y_0)\|_2}, \quad (5.10)$$

où  $g(x - x_0, y - y_0)$  est la fonction gaussienne attendue comme résultat de l'application de l'opérateur 2×1D à  $\delta(x - x_0, y - y_0)$ , la fonction de Dirac au point  $(x_0, y_0)$ . Elevée pour un faible nombre d'itérations, l'erreur décroît rapidement et devient inférieure à 5% à partir de  $M = 10$ .

### 5.2.3 Non séparabilité près des frontières

En océanographie, il est important de prendre en compte les différentes frontières, qu'il s'agisse des côtes ou de la bathymétrie. Or, à l'approche de ces frontières, l'hypothèse de séparabilité des fonctions de corrélation prise dans le paragraphe précédent n'est plus valide. On cherche ici à en appréhender les conséquences à travers différentes formulations garantissant les propriétés de symétrie<sup>72</sup> et de définie positivité des opérateurs. Pour simplifier les écritures et les illustrations, on se limite à l'opérateur de corrélation horizontal 2×1D :

$$\mathbf{C} = \mathbf{\Lambda}^{1/2} \mathbf{L}^M \mathbf{W}^{-1} \mathbf{\Lambda}^{1/2} \quad \text{avec} \quad \mathbf{L}^M = (\mathbf{L}_x \mathbf{L}_y)^M = \mathbf{L}_x^M \mathbf{L}_y^M, \quad (5.11)$$

ou

$$\tilde{\mathbf{C}} = \mathbf{\Lambda}^{1/2} \tilde{\mathbf{L}}^M \mathbf{W}^{-1} \mathbf{\Lambda}^{1/2} \quad \text{avec} \quad \tilde{\mathbf{L}}^M = (\mathbf{L}_y \mathbf{L}_x)^M = \mathbf{L}_y^M \mathbf{L}_x^M. \quad (5.12)$$

Lorsque les opérateurs 1D commutent, les opérateurs de diffusion  $\mathbf{L}$  et  $\tilde{\mathbf{L}}$  sont égaux et par conséquent les opérateurs de corrélations  $\mathbf{C}$  et  $\tilde{\mathbf{C}}$  sont également égaux. De plus,  $\mathbf{L}$  et  $\tilde{\mathbf{L}}$  sont autoadjoints, ce qui permet aux formulations des équations (5.11) et (5.12) d'être symétriques.

En revanche, si les opérateurs 1D ne commutent pas, on a

$$\mathbf{L}^* = (\mathbf{L}_x \mathbf{L}_y)^* = \mathbf{L}_y^* \mathbf{L}_x^* = \mathbf{L}_y \mathbf{L}_x \neq \mathbf{L}_x \mathbf{L}_y = \mathbf{L}.$$

<sup>72</sup>On parle ici de la symétrie de l'opérateur de corrélation dans le sens où l'application de l'opérateur  $\mathbf{C}$  donne le même résultat que l'application de l'opérateur  $\mathbf{C}^T$ , et non pas de la symétrie de la fonction de corrélation résultante telle que définie en Annexe B.

On a alors  $\mathbf{L}^M \mathbf{W}^{-1} \neq (\mathbf{L}^M \mathbf{W}^{-1})^T$  et ainsi  $\mathbf{C} \neq \mathbf{C}^T$ . Les opérateurs de corrélations  $\mathbf{C}$  et  $\tilde{\mathbf{C}}$  (la même démonstration s'y applique) ne sont donc pas symétriques lorsque les opérateurs 1D ne commutent pas. Néanmoins, il est possible de forcer la symétrie de l'opérateur  $\mathbf{C}$  (ou  $\tilde{\mathbf{C}}$ ) en redéfinissant l'opérateur de diffusion par

$$\mathbf{L}_1^M = \frac{1}{2} \left[ \mathbf{L}^M + (\mathbf{L}^M)^* \right] = \frac{1}{2} \left[ \mathbf{L}_x^M \mathbf{L}_y^M + (\mathbf{L}_y^M)^* (\mathbf{L}_x^M)^* \right]. \quad (5.13)$$

Une première formulation de l'opérateur de corrélation dans le cas de non-séparabilité est donc :

$$\begin{aligned} \mathbf{C}_1 &= \mathbf{\Lambda}^{1/2} \frac{1}{2} \left[ \mathbf{L}_x^M \mathbf{L}_y^M + \mathbf{W}^{-1} (\mathbf{L}_y^M)^T \mathbf{W} \mathbf{W}^{-1} (\mathbf{L}_x^M)^T \mathbf{W} \right] \mathbf{W}^{-1} \mathbf{\Lambda}^{1/2} \\ &= \frac{1}{2} \mathbf{\Lambda}^{1/2} \left[ \mathbf{L}_x^M \mathbf{L}_y^M \mathbf{W}^{-1} + \mathbf{W}^{-1} (\mathbf{L}_y^M)^T (\mathbf{L}_x^M)^T \right] \mathbf{\Lambda}^{1/2} \\ &= \frac{1}{2} (\mathbf{C} + \mathbf{C}^T), \end{aligned} \quad (5.14)$$

où on a bien  $\mathbf{C}_1 = \mathbf{C}_1^T$ . On suppose de plus que cet opérateur est défini positif (cette hypothèse n'est pas discutée ici car cette formulation ne sera pas retenue).

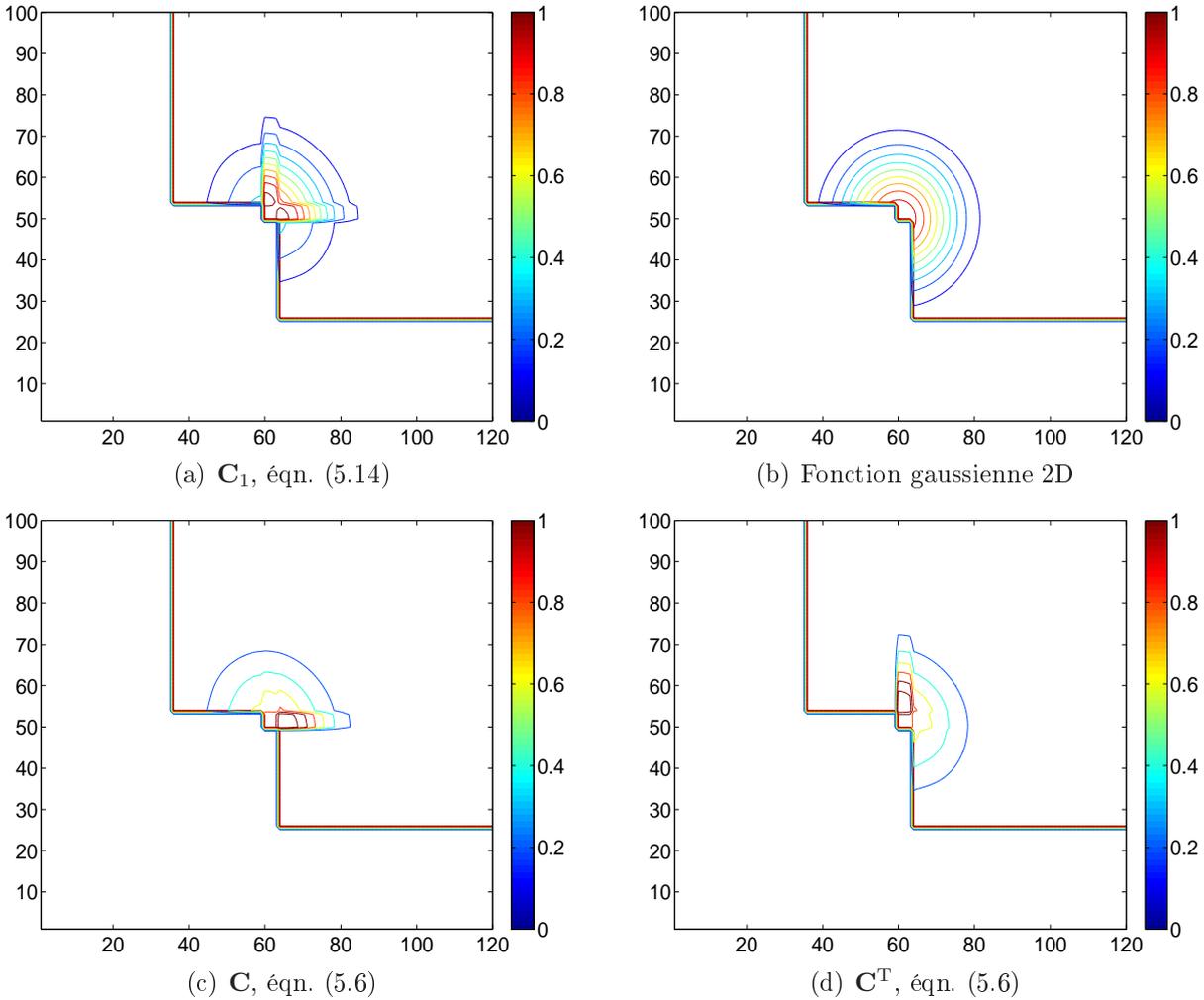


FIG. 5.4 – Solution de l'application de l'opérateur de diffusion implicite  $2 \times 1D$  suivant l'équation (5.14), à une fonction de Dirac située à l'angle de la frontière complexe, sur une grille régulière de résolution 1 point, avec une échelle de corrélation de  $L_{AR} = 10$  points (panneau a). Pour comparaison, la fonction gaussienne attendue est représentée sur le panneau b. Le détail des termes de la moyenne est donné sur les panneaux c et d.

La figure 5.4a montre le résultat de l'application de l'opérateur 2×1D défini par l'équation (5.14) à une fonction de Dirac située à l'angle d'une frontière complexe. On choisit ici délibérément une frontière difficile afin d'exposer plus clairement les difficultés auxquelles sont confrontées les différentes formulations. Le fait que la frontière soit symétrique suivant  $x$  et  $y$  n'a pas d'influence sur les résultats. La grille est régulière et de résolution 1 point, et l'échelle de corrélation est définie sur les deux axes par  $L_{AR} = 10$  points. Pour comparaison, la fonction gaussienne attendue est donnée par la figure 5.4b. On constate l'apparition d'un artefact important qui s'explique par le fait que pour chaque terme de la moyenne, la condition initiale est entièrement diffusée sur un axe avant d'être diffusée sur l'autre. Sur la figure 5.4c qui représente le premier terme de cette moyenne, la condition initiale est d'abord entièrement diffusée suivant l'axe  $y$  avant d'être diffusée suivant l'axe  $x$ , ce qui prolonge artificiellement la frontière zonale. Sur la figure 5.4d qui représente le deuxième terme de la moyenne, la condition initiale est dégagée du recoin de la frontière par la diffusion suivant  $x$  et peut donc ensuite être étalée en haut et en bas. Mais cela entraîne une prolongation artificielle de la première frontière méridienne. On remarque également sur ces deux figures, l'effet d'étirement de l'échelle de corrélation dû à l'application des conditions aux frontières de Neumann avec une normalisation exacte (voir paragraphe 4.4.2).

Les formulations données par les équations (5.11) et (5.14) ne sont pas satisfaisantes. En effet, la définition positive n'est pas garantie et même si la symétrie est assurée par la définition (5.14), le résultat, deux fois plus coûteux, reste insuffisant près des frontières. On propose alors de factoriser l'opérateur de diffusion par  $\mathbf{L}_2^M = \mathbf{L}^{M/2} \mathbf{L}^{M/2}$ , où les termes en  $M/2$  représentent une intégration des équations de diffusion sur le pseudo-temps  $[0, T/2]$  au lieu de  $[0, T]$ . Par convention, on choisira donc obligatoirement un nombre d'itérations  $M$  pair<sup>73</sup>. Si les opérateurs 1D commutent, l'opérateur de diffusion 2×1D est alors autoadjoint et on peut écrire

$$\mathbf{L}_2^M = \mathbf{L}^{M/2} \mathbf{L}^{M/2} = \mathbf{L}^{M/2} \left( \mathbf{L}^{M/2} \right)^* = \mathbf{L}_x^{M/2} \mathbf{L}_y^{M/2} \left( \mathbf{L}_y^{M/2} \right)^* \left( \mathbf{L}_x^{M/2} \right)^*. \quad (5.15)$$

Une nouvelle formulation de l'opérateur de corrélation est alors donnée par

$$\begin{aligned} \mathbf{C}_2 &= \mathbf{\Lambda}^{1/2} \mathbf{L}_x^{M/2} \mathbf{L}_y^{M/2} \mathbf{W}^{-1} \left( \mathbf{L}_y^{M/2} \right)^T \mathbf{W} \mathbf{W}^{-1} \left( \mathbf{L}_x^{M/2} \right)^T \mathbf{W} \mathbf{W}^{-1} \mathbf{\Lambda}^{1/2} \\ &= \mathbf{\Lambda}^{1/2} \mathbf{L}_x^{M/2} \mathbf{L}_y^{M/2} \mathbf{W}^{-1} \left( \mathbf{L}_y^{M/2} \right)^T \left( \mathbf{L}_x^{M/2} \right)^T \mathbf{\Lambda}^{1/2} \\ &= \mathbf{C}^{1/2} \left( \mathbf{C}^{1/2} \right)^T. \end{aligned} \quad (5.16)$$

L'avantage principal de l'équation (5.16) par rapport aux équations (5.11) et (5.14) est que la formulation même garantit un résultat symétrique ( $\mathbf{C}_2 = \mathbf{C}_2^T$ ) et défini positif. De la même manière, on peut définir

$$\begin{aligned} \tilde{\mathbf{C}}_2 &= \mathbf{\Lambda}^{1/2} \mathbf{L}_y^{M/2} \mathbf{L}_x^{M/2} \mathbf{W}^{-1} \left( \mathbf{L}_x^{M/2} \right)^T \left( \mathbf{L}_y^{M/2} \right)^T \mathbf{\Lambda}^{1/2} \\ &= \tilde{\mathbf{C}}^{1/2} \left( \tilde{\mathbf{C}}^{1/2} \right)^T. \end{aligned} \quad (5.17)$$

Néanmoins, si les opérateurs  $\mathbf{L}_x$  et  $\mathbf{L}_y$  ne commutent pas, on obtiendra des résultats différents suivant que l'on applique l'un ou l'autre en premier, puisque

$$\mathbf{C}_2 = \mathbf{C}^{1/2} \left( \mathbf{C}^{1/2} \right)^T \neq \tilde{\mathbf{C}}^{1/2} \left( \tilde{\mathbf{C}}^{1/2} \right)^T = \tilde{\mathbf{C}}_2.$$

Pour palier ce problème, Dobricic et Pinardi (2008) proposent dans le cadre du filtre récursif, de modifier la définition (5.15) en prenant en compte les deux définitions possibles des opérateurs de

<sup>73</sup>Il est en fait possible d'utiliser un nombre d'itérations impair mais au prix de difficultés supplémentaires que l'on n'a pas jugé nécessaires d'introduire ici.

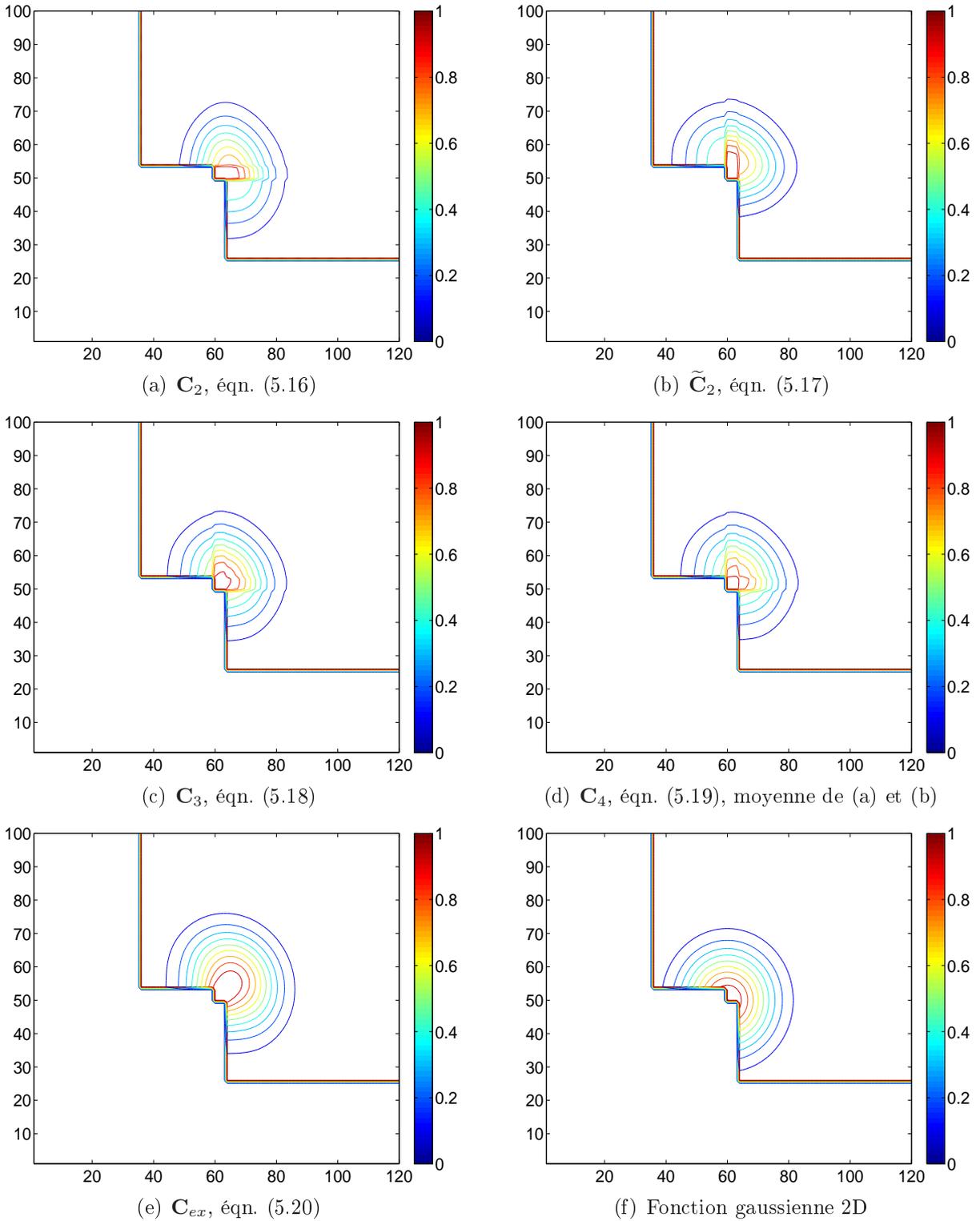


FIG. 5.5 – Solutions de l'application de l'opérateur de diffusion implicite  $2 \times 1D$  à une fonction de Dirac située à l'angle de la frontière complexe, sur une grille régulière de résolution 1 point, avec une échelle de corrélation de  $L_{AR} = 10$  points. L'opérateur  $2 \times 1D$  est donné par l'équation (5.16) (a), l'équation (5.17) (b), l'équation (5.18) (c) ou l'équation (5.19) (d). Pour comparaison, la solution de l'application de l'opérateur explicite 2D suivant l'équation (5.20) (e) et la fonction gaussienne attendue (f) sont également représentés.

diffusion :

$$\begin{aligned}\mathbf{L}_3^M &= \frac{1}{2} \left[ \mathbf{L}^{M/2} + \tilde{\mathbf{L}}^{M/2} \right] \frac{1}{2} \left[ \mathbf{L}^{M/2} + \tilde{\mathbf{L}}^{M/2} \right]^* \\ &= \frac{1}{2} \left[ \mathbf{L}_x^{M/2} \mathbf{L}_y^{M/2} + \mathbf{L}_y^{M/2} \mathbf{L}_x^{M/2} \right] \frac{1}{2} \left[ \left( \mathbf{L}_y^{M/2} \right)^* \left( \mathbf{L}_x^{M/2} \right)^* + \left( \mathbf{L}_x^{M/2} \right)^* \left( \mathbf{L}_y^{M/2} \right)^* \right],\end{aligned}$$

ce qui conduit à la formulation

$$\mathbf{C}_3 = \frac{1}{4} \mathbf{\Lambda}^{1/2} \left[ \mathbf{L}_x^{M/2} \mathbf{L}_y^{M/2} + \mathbf{L}_y^{M/2} \mathbf{L}_x^{M/2} \right] \mathbf{W}^{-1} \left[ \left( \mathbf{L}_y^{M/2} \right)^T \left( \mathbf{L}_x^{M/2} \right)^T + \left( \mathbf{L}_x^{M/2} \right)^T \left( \mathbf{L}_y^{M/2} \right)^T \right] \mathbf{\Lambda}^{1/2}. \quad (5.18)$$

Une autre possibilité est de prendre en compte les deux définitions possibles des opérateurs de corrélation données par les équations (5.16) et (5.17) :

$$\begin{aligned}\mathbf{C}_4 &= \frac{1}{2} \mathbf{\Lambda}^{1/2} \left[ \mathbf{L}_x^{M/2} \mathbf{L}_y^{M/2} \mathbf{W}^{-1} \left( \mathbf{L}_y^{M/2} \right)^T \left( \mathbf{L}_x^{M/2} \right)^T + \mathbf{L}_y^{M/2} \mathbf{L}_x^{M/2} \mathbf{W}^{-1} \left( \mathbf{L}_x^{M/2} \right)^T \left( \mathbf{L}_y^{M/2} \right)^T \right] \mathbf{\Lambda}^{1/2} \\ &= \frac{1}{2} \left( \mathbf{C}_2 + \tilde{\mathbf{C}}_2 \right).\end{aligned} \quad (5.19)$$

Les formulations données par les équations (5.18) et (5.19) diffèrent par les produits croisés qui apparaissent dans la première. Garantissant la symétrie et la définition positive, elles offrent l'avantage sur l'équation (5.16) de s'affranchir des problèmes de non-commutation des opérateurs 1D. Cependant, elles nécessitent deux fois plus d'application des opérateurs de diffusion et sont donc plus coûteuses.

Pour comparaison, l'opérateur de corrélation basé sur l'opérateur de diffusion explicite 2D s'écrit

$$\mathbf{C}_{ex} = \mathbf{\Lambda}_{ex}^{1/2} \mathbf{L}_{exh}^{M/2} \mathbf{W}^{-1} \left( \mathbf{L}_{exh}^{M/2} \right)^T \mathbf{\Lambda}_{ex}^{1/2}, \quad (5.20)$$

où  $\mathbf{L}_{exh}$  représente la discrétisation spatiale de l'équation (5.4).

Les figures 5.5a-d présentent les résultats de l'application d'un opérateur 2×1D suivant les différentes formulations énoncées, à une fonction de Dirac dans les mêmes conditions que pour la figure 5.4a. Pour comparaison, les figures 5.5e-f montrent respectivement le résultat de l'application de l'opérateur explicite 2D suivant l'équation (5.20), et la fonction gaussienne attendue. Dans tous les cas la normalisation associée à l'opérateur de diffusion est donnée par une méthode exacte. Pour la figure 5.5a, les opérateurs sont appliqués suivant l'équation (5.16), où l'application adjointe suivant  $x$  intervient en premier. On voit ici clairement que la première diffusion est bien marquée et permet ensuite, par diffusion suivant  $y$ , d'obtenir un résultat correct le long de la frontière méridienne, tandis qu'il est moins satisfaisant le long de la frontière zonale. Lorsque le sens d'application est inversé (figure 5.5b), les mêmes commentaires s'appliquent mais sur les axes opposés et le résultat devient bien plus satisfaisant sur la frontière zonale que sur la frontière méridienne. Les figures 5.5c-d montrent les solutions obtenues respectivement par l'application des équations (5.18) et (5.19). L'effet du traitement séparé de chaque axe reste visible mais la forme obtenue est bien plus proche de la gaussienne attendue que pour les figures 5.5a-b. On remarque enfin, que l'étirement artificiel de l'échelle de corrélation dû à la normalisation exacte semble plus marqué pour l'opérateur explicite 2D que pour les différentes formulations des opérateurs implicites 2×1D. Mais le traitement étant effectué sur les deux axes simultanément, il n'y a pas d'effet parasite du type de celui qui apparaît sur les figures 5.5a-d.

## 5.2.4 Non séparabilité due aux variations des échelles de corrélation

Dans beaucoup d'applications d'assimilation de données, et en particulier pour l'océanographie, supposer que les échelles de corrélation sont constantes est peu réaliste. La prise en compte de l'in-homogénéité et de l'anisotropie pour une fonction de corrélation gaussienne s'effectue par l'utilisation

d'un tenseur (Weaver et Courtier, 2001; Purser *et al.*, 2003b; Purser, 2005). En se limitant au cas 2D, la fonction gaussienne s'écrit

$$g(\mathbf{x}, \mathbf{y}) = \exp \left[ \frac{1}{2} (\mathbf{x} \ \mathbf{y}) \begin{pmatrix} \kappa_{xx} & \kappa_{xy} \\ \kappa_{yx} & \kappa_{yy} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \right]. \quad (5.21)$$

Les termes diagonaux  $\kappa_{xx}$  et  $\kappa_{yy}$  du tenseur permettent de prendre en compte des échelles de corrélation qui varient suivant les axes  $x$  et  $y$  respectivement. Les figures 5.6a-b représentent le résultat de l'application d'un opérateur de diffusion implicite  $2 \times 1D$  avec  $M = 10$  à une condition initiale composée de deux fonctions de Dirac, associée à une normalisation exacte. Sur le panneau (a), l'échelle de corrélation est constante  $L_{AR} = 7$  pour chaque axe, tandis que sur le panneau (b), l'échelle varie sur son axe autour de cette valeur suivant une sinusoïde. Pour la fonction en bas à gauche, les échelles sont raccourcies tandis qu'elles sont allongées pour la fonction en haut à droite.

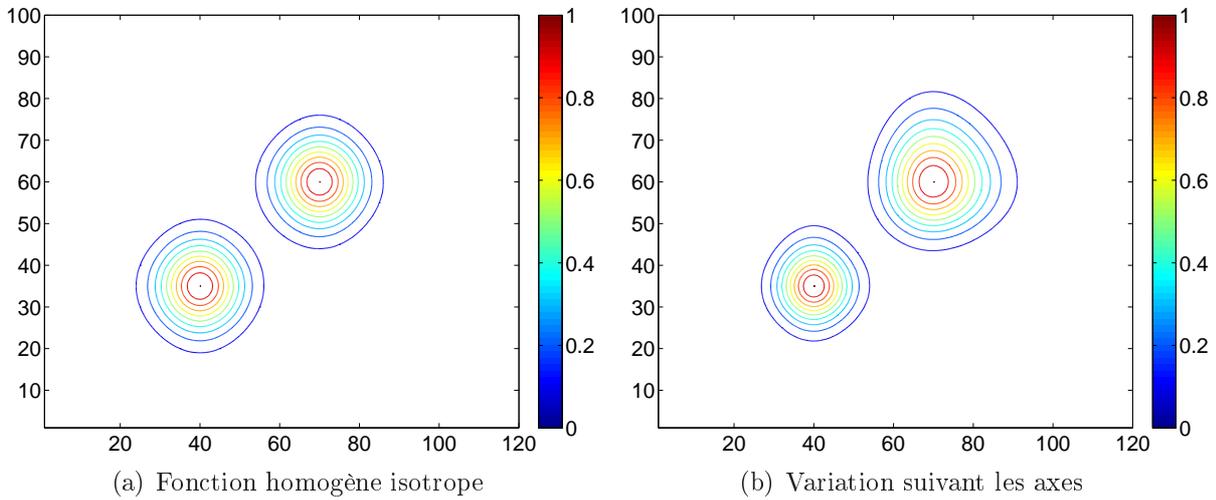


FIG. 5.6 – Application d'un opérateur  $2 \times 1D$  avec  $M = 10$  à une condition initiale composée de deux fonctions de Dirac, lorsque l'échelle de corrélation est constante  $L_{AR} = 7$  points pour chaque axe (a), et lorsqu'elle varie sur son axe autour de cette valeur suivant une sinusoïde (b).

Les termes extra-diagonaux  $\kappa_{xy}$  et  $\kappa_{yx}$  introduisent quant à eux l'anisotropie, mais ne peuvent pas être pris en compte directement par l'opérateur  $2 \times 1D$  à moins d'effectuer un changement de base permettant de ramener ces termes sur la diagonale du tenseur inhérent. Dans le cadre du filtre récursif, Purser *et al.* (2003b) et Purser (2005) proposent une méthode permettant de générer des axes artificiels suivant lesquels il est possible d'allonger ou d'étrécir les échelles. Cette méthode est également applicable à l'équation de diffusion mais ne sera pas testée ici.

Les échelles de corrélation sont souvent estimées à partir de connaissances physiques ou statistiques. Mais lorsque des contraintes numériques existent (critère de stabilité pour l'opérateur de diffusion explicite), elles sont plutôt choisies de manière à satisfaire des conditions de faisabilité. Lorsqu'aucune contrainte numérique n'existe (opérateur de diffusion implicite), une façon de donner davantage de cohérence physique aux échelles de corrélation et même, de les faire dépendre de l'écoulement, est de les estimer à partir d'un ensemble (Daget, 2008), suivant les formules de Pannekoucke *et al.* (2008) par exemple. Néanmoins, il est important de noter que l'hypothèse de séparabilité des fonctions de corrélation n'est plus valide dès que des variations de ces échelles de corrélation sont prises en compte (sauf si pour chaque opérateur 1D, la variation ne dépend que de l'axe afférent).

## 5.3 Application au système NEMOVAR

Les résultats obtenus dans le paragraphe précédent sont maintenant appliqués au système d'assimilation de données océaniques variationnelle NEMOVAR, pour modéliser les corrélations spatiales de l'erreur d'ébauche pour la température (3D) et les composantes non équilibrées de salinité (3D) et hauteur d'eau (2D). Quelque soit la formulation, le nombre d'itérations à appliquer pour les différents opérateurs 1D doit être choisi de manière à ce que l'anisotropie parasite résultant d'un  $M$  petit soit atténuée de façon satisfaisante. Bien qu'il n'y ait aucune obligation de choisir le même nombre d'itérations pour tous les opérateurs, c'est ce que nous faisons dans tout ce qui suit en prenant  $M = 10$ .

L'opérateur explicite précédemment mis en œuvre dans NEMOVAR est donné en 2D par l'équation (5.20) et en 2D+1D par l'extension

$$\mathbf{C}_{ex} = \mathbf{\Lambda}_{ex}^{1/2} \mathbf{L}_{exz}^{M/2} \mathbf{L}_{exh}^{M/2} \mathbf{W}^{-1} \left( \mathbf{L}_{exh}^{M/2} \right)^T \left( \mathbf{L}_{exz}^{M/2} \right)^T \mathbf{\Lambda}_{ex}^{1/2}. \quad (5.22)$$

On note que les opérateurs  $\mathbf{L}_{exz}$  et  $\mathbf{L}_{exh}$  ne commutent pas près des côtes ou de la bathymétrie, et donneraient donc des résultats différents s'ils étaient intervertis dans l'équation (5.22). De la même manière, on choisit pour l'opérateur implicite, d'utiliser la formulation donnée par l'équation (5.16) pour l'opérateur horizontal  $2 \times 1D$  et son extension à trois dimensions

$$\mathbf{C}_2 = \mathbf{\Lambda}^{1/2} \mathbf{L}_x^{M/2} \mathbf{L}_y^{M/2} \mathbf{L}_z^{M/2} \mathbf{W}^{-1} \left( \mathbf{L}_z^{M/2} \right)^T \left( \mathbf{L}_y^{M/2} \right)^T \left( \mathbf{L}_x^{M/2} \right)^T \mathbf{\Lambda}^{1/2},$$

pour l'opérateur  $3 \times 1D$ . L'ordre d'application des opérateurs 1D est choisi de façon arbitraire (mais prendra plus de sens dans le paragraphe 5.4. Comme pour l'explicite, les conditions aux frontières sont de type Neumann et appliquées grâce à l'utilisation de masques comme indiqué au paragraphe 4.5.2. Enfin, la normalisation est réalisée par la méthode par vecteur aléatoire de 1000 membres.

### 5.3.1 Aspects numériques

La résolution numérique du schéma doit tenir compte de la parallélisation adoptée pour le modèle. Or, celle de NEMO (et de NEMOVAR par conséquence) consiste à décomposer le domaine global en sous-domaines horizontaux. La verticale n'étant pas décomposée, une méthode directe basée sur la factorisation de Cholesky peut y être appliquée. Pour les opérateurs zonal et méridien cependant, chaque domaine 1D peut être distribué sur des processeurs différents et n'est donc pas accessible de manière contigue. Un premier choix consiste alors à résoudre le système linéaire de chaque pas de temps par une méthode itérative adaptée à cette parallélisation. Cette approche nécessite d'étudier sa convergence et le nombre d'itérations requis, ainsi que la bonne compatibilité des opérateurs directs et adjoints. Dans cette thèse, on choisit de conserver la résolution de l'horizontale par la même méthode directe que pour la verticale, et donc d'utiliser une factorisation de Cholesky (voir paragraphe 4.5.2). Pour ce faire, les processeurs doivent être réorganisés de manière adéquate avant et après chaque application d'un opérateur horizontal. Le détail de cette réorganisation est donné en Annexe F.

L'opérateur "zonal"  $\mathbf{L}_x$  est appliqué sur chaque ligne complète ( $(j, k)$  fixé) de la grille. Dans certains cas, des côtes sont présentes en début et en fin de ligne (découpage est/ouest de la sphère au milieu de l'Asie, suivant la longitude  $\approx 80E$ ), mais dans d'autres, le raccordement de la ligne en cercle doit être géré. De plus, dans la région du courant circumpolaire Antarctique, aucune terre ne permet de délimiter un domaine. C'est pourquoi il est nécessaire ici d'introduire des conditions aux frontières

périodiques. La matrice tribande  $\mathbf{A}$  du paragraphe 4.5.2 à laquelle on ajoute la périodicité devient :

$$\mathbf{A} = \begin{pmatrix} w_i + \alpha_{I+\frac{1}{2}} + \alpha_{1+\frac{1}{2}} & & -\alpha_{1+\frac{1}{2}} & & & & -\alpha_{I+\frac{1}{2}} \\ & -\alpha_{1+\frac{1}{2}} & w_i + \alpha_{1+\frac{1}{2}} + \alpha_{2+\frac{1}{2}} & & & & \\ & & & \ddots & & & \\ & & & & \ddots & & \\ & & & & & \ddots & \\ & & & -\alpha_{I-\frac{3}{2}} & w_i + \alpha_{I-\frac{3}{2}} + \alpha_{I-\frac{1}{2}} & & -\alpha_{I-\frac{1}{2}} \\ & -\alpha_{I+\frac{1}{2}} & & & -\alpha_{I-\frac{1}{2}} & w_i + \alpha_{I-\frac{1}{2}} + \alpha_{I+\frac{1}{2}} & \end{pmatrix}.$$

Compte tenu de la particularité de  $\mathbf{A}$ , la matrice triangulaire inférieure  $\mathbf{G}$  représentant la décomposition de Cholesky telle que  $\mathbf{A} = \mathbf{G}\mathbf{G}^T$  est alors simplifiée en :

$$\mathbf{G} = \begin{pmatrix} d_1 & & & & & & \\ b_2 & d_2 & & & & & \\ & \ddots & \ddots & & & & \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ p_1 & \cdots & p_{I-2} & b_I + p_{I-1} & d_I & & \end{pmatrix}.$$

Elle nécessite de stocker trois tableaux de taille  $I \times J \times K$ , où  $I$ ,  $J$  et  $K$  représentent le nombre de points du domaine global suivant les coordonnées  $i$ ,  $j$  et  $k$  respectivement. Pour une ligne donnée, *i.e.* à  $(j, k)$  fixé, les composantes  $g_{p,q}$  de la matrice triangulaire inférieure de Cholesky sont données par la diagonale  $g_{p,p} = d_p$ , la bande  $g_{p,p-1} = b_p$  et la ligne de périodicité  $g_{I,q} = p_q$  calculées suivant l'algorithme ci-après, où  $a_{p,q}$  sont les composantes de la matrice  $\mathbf{A}$ . A chaque itération  $m$ , le système est ensuite résolu en introduisant le vecteur  $\mathbf{y}$  tel que  $\mathbf{y} = \mathbf{G}^T \mathbf{x}^{m-1}$  et  $\mathbf{x}^m = \mathbf{G}\mathbf{y}$ , avec  $\mathbf{x}^0$  la condition initiale.

Pour une ligne à $(j, k)$ fixé	
Calcul de $\mathbf{G}$	Résolution du système
<ul style="list-style-type: none"> <li>• <math>b_1 = 0</math></li> <li>• <math>d_1 = \sqrt{a_{1,1}}</math></li> <li>• <math>p_1 = a_{I,1}/d_1</math></li> <li>• Pour <math>i = 2</math> à <math>I - 1</math> <div style="border-left: 1px solid black; padding-left: 10px; margin-left: 20px;"> <math>b_i = a_{i,i-1}/d_{i-1}</math>  <math>d_i = \sqrt{a_{i,i} - b_i^2}</math>  <math>p_i = -p_{i-1}b_i/d_i</math> </div> </li> <li>• Fin pour</li> <li>• <math>b_I = a_{I,I-1}/d_{I-1,I-1}</math></li> <li>• <math>d_I = \sqrt{a_{I,I} - \sum_{n=1}^{I-2} p_n^2 - (b_I + p_{I-1})^2}</math></li> <li>• <math>p_I = 0</math></li> </ul>	<ul style="list-style-type: none"> <li>• <i>Résolution avant</i></li> <li>• <math>y_1 = x_1/d_1</math></li> <li>• Pour <math>i = 2</math> à <math>I - 1</math> <div style="border-left: 1px solid black; padding-left: 10px; margin-left: 20px;"> <math>y_i = (x_i - b_i y_{i-1})/d_i</math> </div> </li> <li>• Fin pour</li> <li>• <math>y_I = (x_I - b_I y_{I-1} - \sum_{n=1}^{I-1} p_n y_n)/d_I</math></li> <li>• <i>Résolution arrière</i></li> <li>• <math>x_I = y_I/d_I</math></li> <li>• Pour <math>i = I - 1</math> à <math>1</math> <div style="border-left: 1px solid black; padding-left: 10px; margin-left: 20px;"> <math>x_i = (y_i - b_{i+1} x_{i+1} - p_i x_I)/d_i</math> </div> </li> <li>• Fin pour</li> </ul>

Pour l'opérateur "méridien"  $\mathbf{L}_y$ , on découpe le globe en une moitié ouest et une moitié est. On considère alors une colonne commençant en Antarctique côté ouest, montant jusqu'au nord et le passant à travers la zone de recouvrement nord, puis redescendant jusqu'en Antarctique côté est. Le domaine 1D étant ainsi borné, des conditions aux frontières périodiques ne sont pas nécessaires et seuls deux tableaux de taille  $I \times J \times K$  doivent être calculés. Le calcul des composantes de la matrice de Cholesky

ainsi que la résolution du problème sont identiques aux algorithmes précédents avec  $p_i = 0$  pour tout  $i$ , et en remplaçant  $I$  par  $2J - r$ , où  $r$  représente le nombre de lignes incluses dans la zone de recouvrement nord. Suivant les options choisies et les variables traitées, cette zone n'est pas définie de la même manière, mais les détails techniques de son traitement ne seront pas exposés ici.

Enfin, l'opérateur vertical  $\mathbf{L}_z$  est appliqué sur chaque colonne d'eau. Le domaine 1D est borné en début par la surface et en fin par la bathymétrie et la décomposition de Cholesky nécessite donc deux tableaux de taille  $I \times J \times K$ . Leurs calculs ainsi que la résolution du problème sont identiques aux algorithmes précédents avec  $p_i = 0$  et en remplaçant  $I$  par  $K$ . Un contrôle permet cependant d'abrégier la boucle dès que le fond océanique est atteint.

La figure 5.7 montre des exemples de corrélations obtenues pour la température sur la grille ORCA1, à partir de l'application de l'opérateur de diffusion implicite  $3 \times 1D$  décrit ici sur une condition initiale constituée de fonctions de Dirac en différents points géographiques. La normalisation est assurée par une méthode par vecteur aléatoire de 1000 membres. Les échelles de corrélation sont paramétrées à  $2^\circ$  pour l'horizontale avec un étirement zonal et rétrécissement méridien autour de l'équateur ( $\pm 15^\circ$ ) sous une forme sinusoïdale. Suivant l'axe zonal par exemple, l'échelle est doublée à  $4^\circ$  sur l'équateur puis raccourcit jusqu'à revenir à  $2^\circ$  dès que l'on atteint la latitude  $15^\circ$  Nord ou Sud. Pour la verticale, l'échelle est paramétrée à deux fois la taille de la maille, sachant que cette taille évolue de 10 mètres sur les 100 premiers mètres de colonne d'eau jusqu'à environ 500 mètres près du fond océanique.

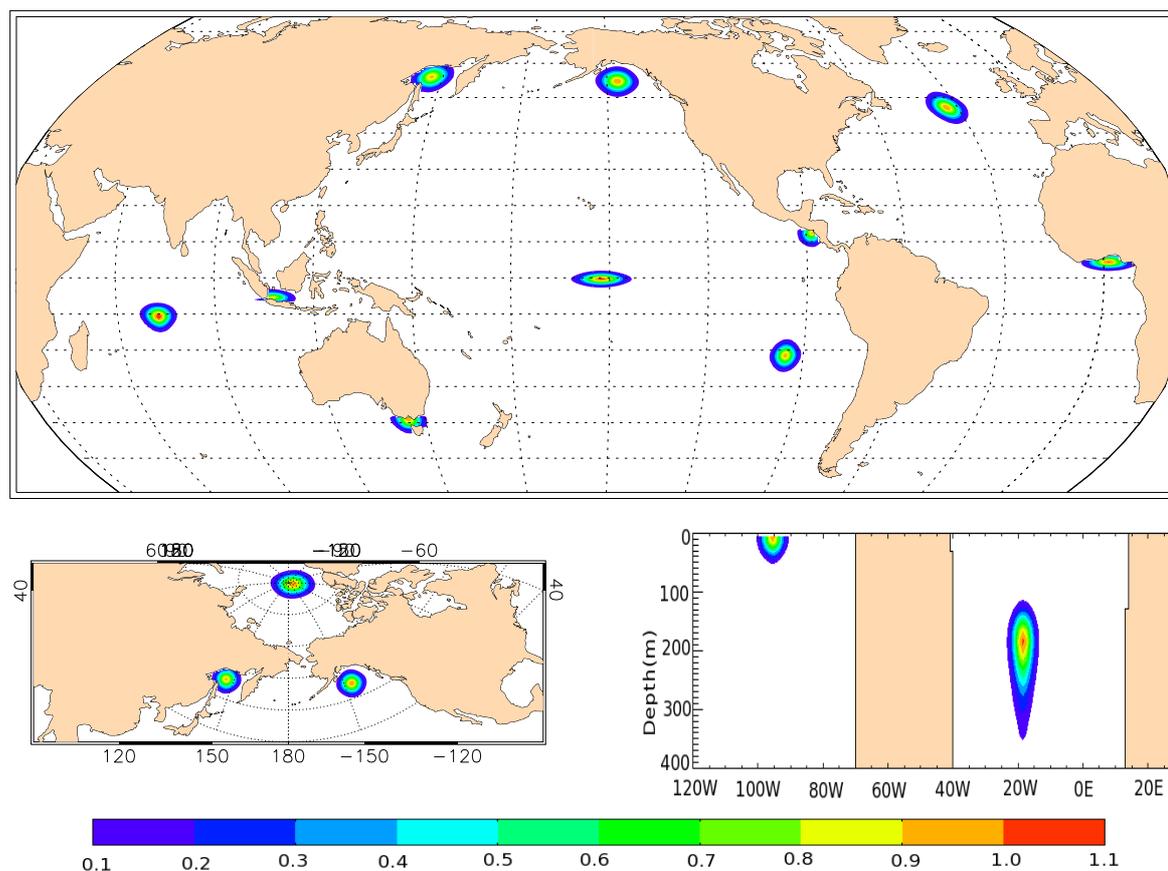


FIG. 5.7 – Exemple de corrélations de température issues de l'application d'un opérateur de diffusion implicite  $3 \times 1D$  avec normalisation par vecteur aléatoire de 1000 membres. Les résultats sont donnés en surface pour le globe (en haut) et au Pôle Nord (en bas à gauche). Les latitudes  $y$  sont représentées tous les  $10^\circ$  et les longitude tous les  $30^\circ$ . En bas à droite, une coupe est réalisée à la latitude  $21^\circ S$  pour le Pacifique Est et l'Atlantique, où une corrélation à été calculée vers 185 mètres de profondeur.

La figure du haut montre les corrélations en surface sur le globe, où la latitude est indiquée tous les  $10^\circ$  et la longitude tous les  $30^\circ$ . La corrélation au sud de l'Inde correspond au raccordement est/ouest. Positionnée aux environs de  $10^\circ\text{S}$ , on peut voir dans la forme légèrement triangulaire de sa moitié basse l'effet du retour de l'échelle zonale à  $2^\circ$ . La variation des échelles est également bien visible au milieu du Pacifique Equatorial ainsi que dans le golfe de Guinée. Au niveau de l'Indonésie, on note que si la propagation de la corrélation est bien arrêtée par les frontières, elle s'infiltré dans l'ouverture présentée par le détroit de la Sonde (entre Sumatra et Java). La figure en bas à gauche montre une corrélation en surface positionnée à travers le recouvrement nord. Enfin, la figure en bas à droite constitue une coupe à la latitude  $21^\circ\text{S}$  pour le Pacifique Est et l'Atlantique. A gauche, la corrélation correspond à celle située à l'ouest de la frontière Pérou/Chili. A droite, la corrélation est issue d'une fonction de Dirac située à 185 mètres de profondeur. On peut voir l'allongement de l'échelle de corrélation verticale en fonction de la profondeur.

Pour évaluer l'impact des frontières sur les corrélations, la figure 5.8 montre un zoom des corrélations obtenues en Indonésie (a), près du détroit de Bass entre Australie et Tasmanie (c) et sur l'Amérique Latine (e). Ces figures sont tracées en fonctions des coordonnées curvilinéaires  $(i, j, 1)$  pour être directement comparées à la figure 5.5a. On remarque alors que les effets parasites visibles sur cette dernière figure sont nettement moins marqués sur les figures 5.8a,c,e, les frontières y étant moins sévères par rapport à l'échelle de corrélation paramétrée. On donne de plus sur les panneaux b, d et f, les corrélations obtenues dans les mêmes conditions mais avec un opérateur explicite 2D ( $M = 200$ ) + 1D ( $M = 14$  pour satisfaire le critère de stabilité avec une échelle de corrélation correspondant à 2 fois la maille verticale).

De façon générale, les corrélations implicites et explicites sont assez similaires, bien qu'il soit possible de distinguer l'effet du traitement des axes séparés dans la forme "un peu moins circulaire" des résultats de l'application de l'opérateur  $3 \times 1\text{D}$ . On note par exemple sur la figure 5.8c, l'arrêt des contours au niveau de la frontière horizontale de l'Australie alors qu'elles remontent légèrement sur la figure 5.8d. De même, on note que la remontée vers les terres des courbes de niveaux inférieures (bleues) est un peu plus brutale sur la figure 5.8e que sur la figure 5.8f. Mais la plus grande différence concerne la propagation de la corrélation à travers le détroit de la Sonde qui reste cantonnée à l'axe  $i$  sur la figure 5.8a alors qu'elle s'étend circulairement sur la figure 5.8b. Globalement, les différences présentées sont légères et semblent acceptables, en particulier dans le cadre des configurations globales à plutôt basse résolution utilisées ici, où l'assimilation de données près des côtes n'est généralement pas privilégiée. On notera que la formulation donnée par l'équation (5.19) étendue au trois dimensions (soit 6 termes au lieu de 2 dans la moyenne) donne des résultats plus proches de ceux de l'explicite (non montrés) bien que la différence dans le détroit de la Sonde soit toujours importante.

Les corrélations résultant de l'application d'un opérateur de diffusion implicite  $3 \times 1\text{D}$  sont globalement similaires à celles résultant de l'application d'un opérateur de diffusion explicite  $2\text{D}+1\text{D}$ . On s'intéresse donc maintenant au coût théorique des deux opérateurs, tant en terme de mémoire que de temps de calcul. On donne ci-après une estimation théorique de ce coût pour le calcul des corrélations d'une variable donnée.

Pour l'opérateur de diffusion explicite  $2\text{D}+1\text{D}$ , les divers paramètres nécessaires à son application sont calculés et sauvegardés en mémoire lors de l'initialisation du code, soit en tout 4 tableaux 3D et 1 tableau 2D (dans NEMO, les éléments de volume de l'horizontale ne dépendent pas du niveau vertical). La résolution proprement dite du schéma se fait en combinant les éléments du champ 3D (5 points sur l'horizontale, 3 points sur la verticale). Pour l'horizontale, le calcul des deux dérivées spatiales est effectué successivement, soit deux boucles sur le champ 3D, tandis que pour la verticale, la matrice explicite est directement appliquée, soit une seule boucle sur le champ 3D. Cette matrice se résume à une diagonale et une bande, soit 2 des 4 tableaux 3D stockés au préalable. Après chaque itération de l'horizontale, les processeurs remettent à jour leur halo par échange d'information avec les processeurs

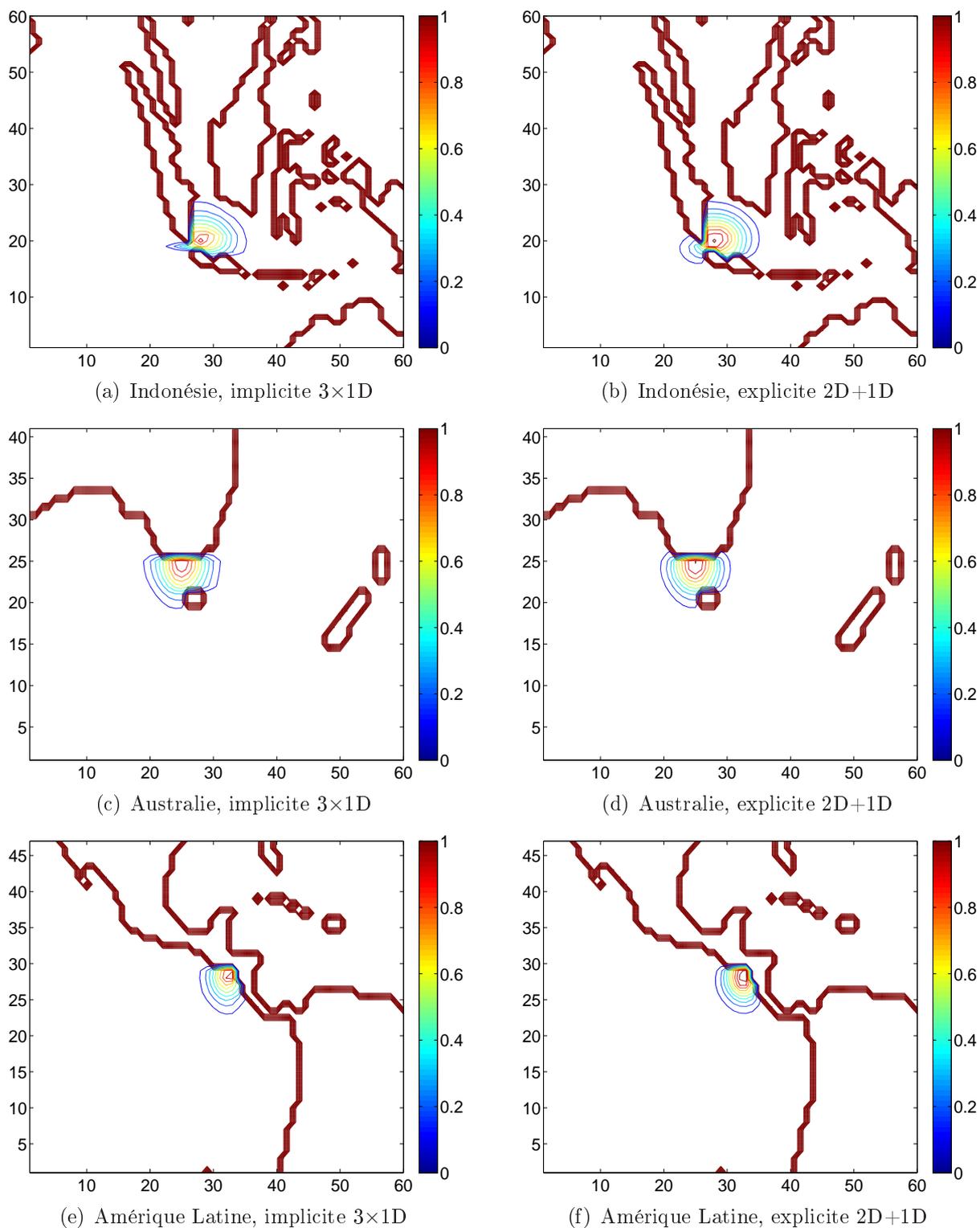


FIG. 5.8 – Zoom près des frontières de la figure 5.7 pour l’Indonésie (a), l’Australie (c) et l’Amérique Latine (e) tracés en surface suivant les coordonnées curvilinéaires  $(i, j, 1)$ . Pour comparaison, les corrélations obtenues dans les mêmes conditions mais pour un opérateur explicite  $2D$  horizontal +  $1D$  vertical sont données sur les panneaux b, d et f.

voisins, l'envoi et la réception de données nécessitant deux communications distinctes. Les échelles de corrélation sont adaptées de façon ad hoc, pour que le nombre d'itérations ne dépasse pas 200 pour l'horizontale et 10 pour la verticale.

Pour l'opérateur de diffusion implicite 3×1D, la décomposition de Cholesky de chacun des opérateurs 1D est également calculée et sauvegardée en mémoire lors de l'initialisation, soit en tout 7 tableaux 3D. Pour résoudre chaque schéma 1D, on effectue deux boucles sur le champ 3D où une combinaison 3 points est calculée à chaque fois. Cependant, une réorganisation des processeurs est effectuée avant et après l'application des opérateurs 1D zonal et méridien, et aucun échange inter-processeur n'est alors nécessaire entre les différentes itérations. 8 réorganisations (2 en  $i$  et 2 en  $j$  pour l'application des opérateurs directs, et 2 en  $i$  et 2 en  $j$  pour l'application des opérateurs adjoints) sont nécessaires.

On peut donc estimer que la résolution de l'opérateur implicite sera plus chère que celle de l'explicite, ce qui semble logique puisque la première nécessite de résoudre un système linéaire à chaque pas de temps. Mais l'algorithme de résolution explicite étant appelé nettement plus souvent, la diffusion implicite devrait au total être plus économe en temps bien que plus coûteuse en mémoire pour le stockage des paramètres et la définition de davantage de variables locales pour la réorganisation des processeurs. Concernant le nombre de communications par processeur, il dépend pour l'explicite du nombre d'itérations horizontales et n'évolue pas (ou peu) avec le nombre de processeurs, contrairement à l'implicite qui dépend en outre de la distribution adoptée. En effet, si  $N_x$  est le nombre de processeurs suivant  $x$ , le nombre de communications nécessaires à la réorganisation de cet axe est proportionnel à  $N_x - 1$ . Si  $N_y$  est le nombre de processeurs suivant  $y$ , le nombre de communications nécessaires à la réorganisation de cet axe est proportionnel à  $2N_y - 1$ . Pour les configurations globales ORCA2 et ORCA1 que nous testons ici, le nombre et la distribution des processeurs employés permet d'avoir un nombre de communications très inférieur pour l'implicite par rapport à l'explicite. Néanmoins, le volume de données échangées à chaque communication est bien plus important pour l'implicite que pour l'explicite. Mais généralement, c'est le temps de latence plutôt que le temps de transfert des données proprement dit qui crée le coût d'une communication, et on peut donc en déduire que le schéma implicite sera moins coûteux de ce point de vue.

En conclusion, si l'opérateur de diffusion implicite 3×1D semble donner un résultat globalement similaire à celui de l'opérateur de diffusion explicite 2D+1D, il est théoriquement moins coûteux en temps bien qu'exigeant plus de mémoire pour les configurations considérées. Mais le nombre de communications pour l'implicite dépend du nombre de processeurs et de leur distribution et peut donc générer des problèmes de scalabilité. Cependant, les configurations sur lesquelles nous travaillons ici sont à plutôt basse résolution, et ne nécessitent donc pas un nombre très important de processeurs. Les problèmes de scalabilité, équilibrage de la charge, ... ne seront donc pas abordés ici, mais devraient l'être si ces opérateurs sont utilisés dans des configurations à plus haute résolution et avec davantage de processeurs.

### 5.3.2 Tests de la boucle interne

Pour comparer le coût de la formulation implicite avec celle de l'explicite, nous effectuons ici une minimisation 3D-Var limitée à la boucle interne de NEMOVAR. Dans l'expérience choisie, les données assimilées correspondent à des profils de température et de salinité *in situ* et d'anomalies altimétriques le long de traces satellite pour la hauteur d'eau. La fenêtre temporelle est de 10 jours à partir du 1er janvier 2006. La minimisation est assurée soit par l'algorithme CONGRAD soit par celui de CGMOD en imposant 40 itérations (soit 42 applications des opérateurs de diffusion en tout). Ces expériences permettent ici d'évaluer les performances de l'opérateur de diffusion implicite 3×1D avec  $M = 10$  pour chaque axe et de les comparer à celles de l'opérateur de diffusion explicite 2D+1D avec  $M = 200$  pour l'horizontale et  $M = 10$  pour la verticale. Les chiffres donnés tout au long de ce paragraphe sont indicatifs, les temps pouvant varier légèrement en fonction de la charge de la machine.

### Profilage pour la configuration ORCA2 sur le *DELL Precision T5500*

Les deux tests suivants<sup>74</sup> permettent d'évaluer et de comparer les performances des opérateurs de diffusion implicite et explicite, et sont effectués pour la configuration ORCA2 sur le PC de bureau *DELL Precision T5500* (2 processeurs Intel Xeon de 4 cœurs chacun). L'opérateur de diffusion est appliqué à deux variables 3D (température et composante non-équilibrée de salinité) et une variable 2D (composante non-équilibrée de hauteur d'eau). Un appel à chaque "racine carrée" des opérateurs direct et adjoint est demandé à chaque itération de la minimisation (CONGRAD avec réorthogonalisation) et pour sa finalisation, soit  $42 \times M/2 \times$  nombre de variables appels des résolution directe et adjointe. Un appel supplémentaire à l'opérateur direct horizontal est également demandé pour le filtrage des variances de la variable température, soit  $M$  résolutions horizontales directes supplémentaires.

	Implicite 2 et 3×1D			Explicite 2D+1D		
	Nb appels	T exec. s	T/appe l ms	Nb appels	T exec. s	T/appe l ms
Paramètre $x$	3	0,070	23,3			
Paramètre $y$	3	0,066	22,0			
Paramètre $h$	3	0,136	45,3	3	0,075	25,0
Paramètre $z$	2	0,084	42,0	2	0,056	28,0
$\mathbf{L}_h^{M/2}$	127	<b>2,47</b>	19,4	127	<i>0,009</i>	0,0
$(\mathbf{L}_h^{M/2})^T$	126	<b>2,36</b>	18,7	126	<i>0,021</i>	0,2
$\mathbf{L}_z^{M/2}$	84	0,001	0,0	84	0,001	0,0
$(\mathbf{L}_z^{M/2})^T$	84	0,001	0,0	84	0,001	0,0
Résol. dir. $x$	640	6,51	10,2			
Résol. adj. $x$	630	6,88	10,9			
Résol. dir. $y$	640	3,51	5,5			
Résol. adj. $y$	630	3,63	5,8			
Résol. dir. $h$	1280	<i>10,02</i>	15,7	12800	<b>46,87</b>	3,7
Résol. adj. $h$	1260	<i>10,51</i>	16,7	12600	<b>69,71</b>	5,5
Résol. dir. $z$	420	3,55	8,5	420	6,76	16,1
Résol. adj. $z$	420	4,14	9,9	420	7,93	18,9

TAB. 5.1 – Comparaison des temps d'exécution des routines explicites et implicites sur le PC *DELL Precision T5500* avec  $1 \times 1$  processeurs et la configuration ORCA2. Le temps d'exécution est le temps nécessaire au code de la routine pour s'exécuter, mais ne comprend pas les temps passés dans d'éventuelles routines appelées. Les temps les plus indicateurs sont comparés en rouge gras et bleu italique.

Le tableau 5.1 compare les temps CPU des opérateurs implicite 2 et 3×1D et explicite 2D+1D lors d'un test sur  $1 \times 1$  processeur<sup>75</sup>. On donne tout d'abord les temps nécessaires au calcul et au stockage des paramètres lors de l'initialisation, puis aux procédures nécessaires à l'application des opérateurs (pour l'implicite on a  $\mathbf{L}_h = \mathbf{L}_x \mathbf{L}_y$ ), et enfin à la résolution du schéma de diffusion (pour l'implicite la résolution suivant  $h$  est en fait la somme des résolutions suivant  $x$  et  $y$ ). Pour chacun, on trouve d'abord le nombre d'appels de la routine (par exemple pour la résolution directe suivant  $x$ , on a 42 itérations de minimisation pour 3 variables avec  $M/2 = 5$ , soit 630 appels auquel on ajoute  $M = 10$  appels pour le filtrage des variances), et leur temps CPU d'exécution en secondes. On en estime alors le temps CPU d'exécution par appel en millisecondes. Ces valeurs correspondent au temps nécessaire

<sup>74</sup>Ces tests sont réalisés grâce au logiciel de profilage *pgprof* : <http://www.pgroup.com/products/pgprof.htm>

<sup>75</sup>Le test est réalisé en multi-processeurs et non pas en monoprocesseur.

à l'exécution de la routine hors appel d'autres routines.

Le calcul des paramètres qui s'effectue à l'initialisation pour chaque variable est plus coûteux pour l'implicite puisqu'il concerne une décomposition de Cholesky, mais reste d'un abord raisonnable. On voit également que, si le temps d'exécution par appel des opérateurs est négligeable pour l'explicite, il est notable pour l'implicite à l'horizontal ( $\approx 20$  ms) qui comprend des calculs supplémentaires (déportés de la résolution) permettant d'anticiper la réorganisation des processeurs. Enfin, concernant la résolution, on peut constater que si le coût d'un appel de la routine explicite horizontale est nettement inférieur à celui de l'implicite ( $\approx 5$  ms contre  $\approx 17$  ms), le nombre d'appels est si élevé que le temps total passé dans la résolution explicite horizontale est entre 4 et 7 fois supérieur à celui de l'implicite. On note également pour l'explicite que le coût de l'adjoint horizontal est bien plus élevé que celui de l'opérateur direct (mais pourrait être optimisé en exploitant le fait que l'opérateur est autoadjoint). Pour la verticale, le coût de l'explicite est plus élevé (mais pourrait être optimisé davantage en stockant un nouveau tableau 3D lors du calcul des paramètres).

Globalement, un appel aux routines implicites est plus coûteux qu'un appel aux routines explicites, ce qui s'explique par les opérations supplémentaires pour le calcul lui-même ainsi que par la prise en compte de la réorganisation des processeurs (une configuration monoprocesseur serait probablement un peu plus rapide). Néanmoins, le nombre d'appels de ces routines est bien plus important pour l'explicite que pour l'implicite. Au final, le temps d'exécution total est largement inférieur pour l'implicite que pour l'explicite.

Le tableau 5.2 donne les temps CPU des mêmes routines que précédemment mais pour une configuration  $4 \times 2$  processeurs. On s'intéresse donc également aux procédures de communication entre les différents processeurs. Dans ce tableau, on donne comme précédemment le nombre d'appels et le temps CPU d'exécution (hors appel d'autres routines) mais également le coût total d'une routine en secondes, *i.e.* le temps total passé dans la routine y compris dans celles qui y sont appelées.

En comparant tout d'abord les temps d'exécution des résolutions avec celles du tableau 5.1, on constate que le fait de travailler avec 8 processeurs réduit le temps d'exécution de l'implicite horizontale d'un facteur 5. En détaillant suivant les axes, le facteur est de 6 pour  $x$  et 4 pour  $y$ , puisque l'on a choisi d'affecter davantage de processeurs suivant l'axe  $x$ . Pour l'explicite, le facteur de réduction est à peu près équivalent pour l'opérateur adjoint mais n'est que de 3,5 pour l'opérateur direct, rétablissant l'équilibre entre les deux opérateurs. Pour la résolution verticale, le facteur de réduction est plus important pour l'explicite que pour l'implicite, et équilibre ainsi les coûts. Le fait de distribuer le calcul sur plusieurs processeurs entraîne la nécessité d'un échange de données entre eux. On sépare ici le coût lié à la réorganisation des processeurs nécessaire aux opérateurs implicites horizontaux (3,74 s au total), du coût des communications liées au voisinage. Ces dernières, également utilisées dans d'autres parties du code, sont intensivement utilisées dans les opérateurs explicites tandis qu'elles sont réduites pour les opérateurs implicites. Au total, les coûts liés à l'échange de données sont 3,6 fois plus importants pour l'explicite que pour l'implicite.

Le coût total de l'application des opérateurs de diffusion explicite et implicite (résolution, réorganisation et communication compris) est également donné pour comparaison. Comparé à la configuration  $1 \times 1$  processeur, le coût des opérateurs implicites est réduit d'un facteur 4 (il était respectivement de 12,50 et 12,89 pour l'horizontal, 3,55 et 4,14 pour la verticale avec la configuration  $1 \times 1$  processeur) tandis que pour l'explicite le facteur de réduction est d'environ 3 pour l'horizontale et 5 pour la verticale (il était respectivement de 47,26 et 70,34 pour l'horizontale, 7,08 et 8,29 pour la verticale avec la configuration  $1 \times 1$  processeur). Dans cette configuration  $4 \times 2$  processeurs, si les opérateurs verticaux sont à peu près équivalents, les opérateurs explicites horizontaux sont au moins 5 fois plus chers que les opérateurs implicites.

	Implicite 2 et 3×1D			Explicite 2D+1D		
	Nb appels	T exec. s	Coût s	Nb appels	T exec. s	Coût s
Paramètre $x$	3	0,013				
Paramètre $y$	3	0,011				
Paramètre $h$	3	0,024	0,045	3	0,017	0,020
Paramètre $z$	2	0,013	0,013	2	0,012	0,012
$\mathbf{L}_h^{M/2}$	127	0,096	<i>3,85</i>	127	0,010	<b>19,49</b>
$(\mathbf{L}_h^{M/2})^T$	126	0,17	<i>3,58</i>	126	0,19	<b>20,53</b>
$\mathbf{L}_z^{M/2}$	84	0,001	1,04	84	0,001	1,39
$(\mathbf{L}_z^{M/2})^T$	84	0,001	1,08	84	0,001	1,49
Résol. dir. $x$	640	1,11				
Résol. adj. $x$	630	1,18				
Résol. dir. $y$	640	0,84				
Résol. adj. $y$	630	0,85				
Résol. dir. $h$	1280	<i>1,95</i>		12800	<b>13,26</b>	
Résol. adj. $h$	1260	<i>2,03</i>		12600	<b>14,89</b>	
Résol. dir. $z$	420	1,04		420	1,22	
Résol. adj. $z$	420	1,08		420	1,33	
Réorg. $x$	515		0,80	0		0,00
Retour réorg. $x$	253		0,63	0		0,00
Réorg. $y$	515		1,59	0		0,00
Retour réorg. $y$	253		0,72	0		0,00
Comm. dir.	1018		<i>1,12</i>	13986		<b>11,92</b>
Comm. adj.	1134		<i>1,21</i>	13902		<b>10,15</b>
Total comm.			<i>6,07</i>			<b>22,07</b>

TAB. 5.2 – Comparaison des temps d’exécution des routines explicites et implicites sur le PC *DELL Precision T5500* avec  $4 \times 2$  processeurs et la configuration ORCA2. Le temps d’exécution est le temps nécessaire au code de la routine pour s’exécuter, mais ne comprend pas les temps passés dans d’éventuelles routines appelées, alors que le coût les comprend. Les temps les plus indicateurs sont comparés en rouge gras et bleu italique.

### Temps d’exécution et consommation mémoire pour ORCA1 sur l’*IBM POWER6*

On réalise maintenant un test de comparaison du temps total d’exécution de la boucle interne et de sa consommation mémoire. Il s’agit ici de quantifier l’apport des opérateurs de diffusion implicite par rapport à l’explicite dans l’ensemble de la boucle interne. Ce test est effectué dans le même cadre d’expérimentation que le test précédent mais cette fois pour la configuration ORCA1 et sur la plateforme *IBM POWER6 Cluster 1600* du CEPMMT. La minimisation est effectuée maintenant par 40 itérations de l’algorithme CGMOD (au lieu de CONGRAD) avec réorthogonalisation.

Le tableau 5.3 compare les coûts CPU et mémoire pour différentes configurations de processeurs lorsque la normalisation est calculée par une méthode par vecteur aléatoire de 1000 membres lors de l’initialisation (soit 1000 applications supplémentaires des opérateurs de diffusion directs en  $M/2$ ). Globalement, l’utilisation de l’opérateur de diffusion implicite permet de réduire le temps d’un facteur d’environ 2 mais consomme un peu plus de mémoire. La valeur de la fonctionnelle incrémentale  $J$  mini-

		Temps CPU s			Mémoire Mo		
		Implicite	Explicite	Gain	Implicite	Explicite	Diff.
Norm. 1000	$1 \times 1$	4566 s	9611 s	<i>50 %</i>	12467 Mo	12107 Mo	<b>350 Mo</b>
	$4 \times 4$	408 s	772 s	<i>45 %</i>	14505 Mo	13823 Mo	<b>700 Mo</b>
	$8 \times 4$	269 s	504 s	<i>45 %</i>	16037 Mo	15351 Mo	<b>700 Mo</b>
	<i>J</i>	78119, 2471					

		Temps CPU s			Mémoire Mo		
		Implicite	Explicite	Gain	Implicite	Explicite	Diff.
Norm. lue	$4 \times 4$	145 s	217 s	<i>30 %</i>	14348 Mo	13806 Mo	<b>550 Mo</b>

TAB. 5.3 – Comparaison des temps CPU et de la consommation mémoire pour une minimisation quadratique (CGMOD avec réorthogonalisation) utilisant un opérateur de diffusion soit implicite, soit explicite. La normalisation est donnée soit par une méthode de calcul “on-line” par vecteur aléatoire de 1000 membres, soit considérée connue et lue à partir d’un fichier. Les écarts sont notés en bleu italique lorsqu’ils sont à l’avantage de l’implicite, en rouge gras sinon.

		Temps CPU s			Mémoire Mo		
		Implicite	Explicite	Gain	Implicite	Explicite	Diff.
Norm.	CONGRAD	393 s	711 s	<i>45 %</i>	13958 Mo	13416 Mo	<b>550 Mo</b>
	CGMOD -	362 s	678 s	<i>45 %</i>	<i>8918 Mo</i>	<i>8235 Mo</i>	<b>700 Mo</b>
1000		<i>J</i>					
	CONGRAD	78119, 2471					
	CGMOD -	78119, 3837					

		Temps CPU s			Mémoire Mo		
		Implicite	Explicite	Gain	Implicite	Explicite	Diff.
Norm. lue	CONGRAD	133 s	170 s	<i>20 %</i>	13628 Mo	13401 Mo	<b>250 Mo</b>
	CGMOD -	100 s	136 s	<i>25 %</i>	<i>8760 Mo</i>	<i>8218 Mo</i>	<b>550 Mo</b>

TAB. 5.4 – Comparaison des temps CPU et de la consommation mémoire pour une minimisation quadratique par CONGRAD avec réorthogonalisation ou CGMOD sans réorthogonalisation (indiquée par le -) utilisant un opérateur de diffusion soit implicite, soit explicite. La normalisation est donnée soit par une méthode de calcul “on-line” par vecteur aléatoire de 1000 membres, soit considérée connue et lue à partir d’un fichier. Les écarts sont notés en bleu italique lorsqu’ils sont à l’avantage de l’implicite, en rouge gras sinon.

male est donnée pour l'opérateur implicite (les valeurs pour les différentes configurations de processeurs sont équivalentes à  $10^{-10}$  près). Dans la deuxième partie du tableau, on compare les temps CPU de la configuration  $4 \times 4$  lorsque la normalisation a été calculée au préalable et est lue à partir d'un fichier. La réduction du temps CPU est bien sûr moins importante mais reste cependant intéressante (30%), pour une consommation de mémoire toujours supérieure à l'explicite.

Pour connaître l'influence de l'algorithme de minimisation choisi, le tableau 5.4 donne les mêmes informations que le tableau 5.3 mais limitées à la configuration  $4 \times 4$ , et en minimisant la fonctionnelle soit par la méthode CONGRAD avec réorthogonalisation, soit à nouveau par CGMOD mais sans réorthogonalisation (indiquée par le -). Globalement les résultats de comparaison entre explicite et implicite sont du même ordre de grandeur, bien que l'on note que l'amélioration apportée par l'implicite est un peu moindre lorsque la minimisation est effectuée par CONGRAD et que la normalisation est supposée connue. Néanmoins, le résultat le plus remarquable est donné par l'algorithme CGMOD sans réorthogonalisation (explicite et implicite), qui réduit la consommation mémoire de plus de 35 % pour un résultat de minimisation similaire (valeur de  $J$ ) aux algorithmes CGMOD ou CONGRAD avec réorthogonalisation. En fait, pour les configurations utilisées ici, cinq itérations supplémentaires sont nécessaires pour atteindre la même valeur de  $J$  bien que l'on remarque d'avantage d'oscillations autour de la trajectoire de descente. Une étude complémentaire reste donc à mener sur les performances de l'algorithme CGMOD sans réorthogonalisation dans un cadre plus général.

### 5.3.3 Expérience d'assimilation

Un dernier test de comparaison est effectué pour la configuration ORCA1 globale sur l'*IBM POWER6* avec une distribution des processeurs  $4 \times 4$ , correspondant à la configuration pré-opérationnelle de NEMOVAR pour la prévision saisonnière du CEPMMT. Il s'agit de l'assimilation 3D-Var de profils de température et de salinité issus du jeu de données contrôlées ENSEMBLES EN3<sup>76</sup> sur une période de 5 ans, de janvier 2004 à décembre 2008, avec une fenêtre d'assimilation de 10 jours. A moins de 300 kilomètres des côtes ( $\approx 2,7^\circ$ ) les variances des erreurs d'observation sont augmentées d'un facteur 2 afin de donner moins de poids à ces observations dans l'assimilation. De plus, les observations situées dans des zones de moins de 500 mètres de profondeur sont rejetées. La boucle externe ne comporte qu'une seule itération, et un cycle d'assimilation correspond donc à une seule minimisation quadratique.

Dans une première expérience, la matrice des covariances d'erreur d'ébauche est construite autour de la modélisation des corrélations par un opérateur de diffusion explicite 2D+1D ( $M = 400$  pour l'horizontale,  $M = 10$  pour la verticale), et la minimisation quadratique est assurée par 40 itérations de l'algorithme CONGRAD avec réorthogonalisation. Cette configuration représente la "**référence**", *i.e.* la configuration habituellement utilisée avant cette thèse (bien que  $M$  soit généralement donné à 200). Dans une deuxième expérience, la matrice des covariances d'erreur d'ébauche est construite autour de la modélisation des corrélations par un opérateur de diffusion implicite 3×1D ( $M = 10$  pour chaque axe), et la minimisation quadratique est assurée par 40 itérations de l'algorithme CGMOD sans réorthogonalisation. Cette configuration représente la "**contribution**" des algorithmes développés dans cette thèse, associés à la nouvelle minimisation mise en place par S. Gratton, A. Piacentini et A. Weaver. Pour les deux expériences, la normalisation est calculée au préalable à partir d'une méthode par vecteur aléatoire de 1000 membres.

Pour comparer ces deux expériences, on utilise les diagnostics classiques permettant d'analyser les performances d'assimilation (les expériences et leurs diagnostics ont été réalisés par T. Pangaud et A. Weaver, on se contente ici d'en rapporter les résultats). Pour chaque fenêtre d'assimilation, on détermine l'écart entre l'ébauche et les observations à assimiler ( $\mathcal{H}[\mathbf{x}^b] - \mathbf{y}^o$ ) d'une part, et l'écart entre l'analyse et les observations assimilées ( $\mathcal{H}[\mathbf{x}^b] + \mathbf{H}\delta\mathbf{x}^a - \mathbf{y}^o$ ) d'autre part. Des statistiques de moyenne,

---

<sup>76</sup><http://hadobs.metoffice.com/en3/>

d'erreur rms et d'écart-type sont ensuite calculées à partir de l'ensemble des fenêtres d'assimilation. Un exemple pour la température moyenne globale est donné par la figure 5.9. Ces diagnostics ont également été calculés pour des zooms sur l'Atlantique Nord-Ouest où le Pacifique Tropical (Nino3.4), ainsi que pour la salinité.

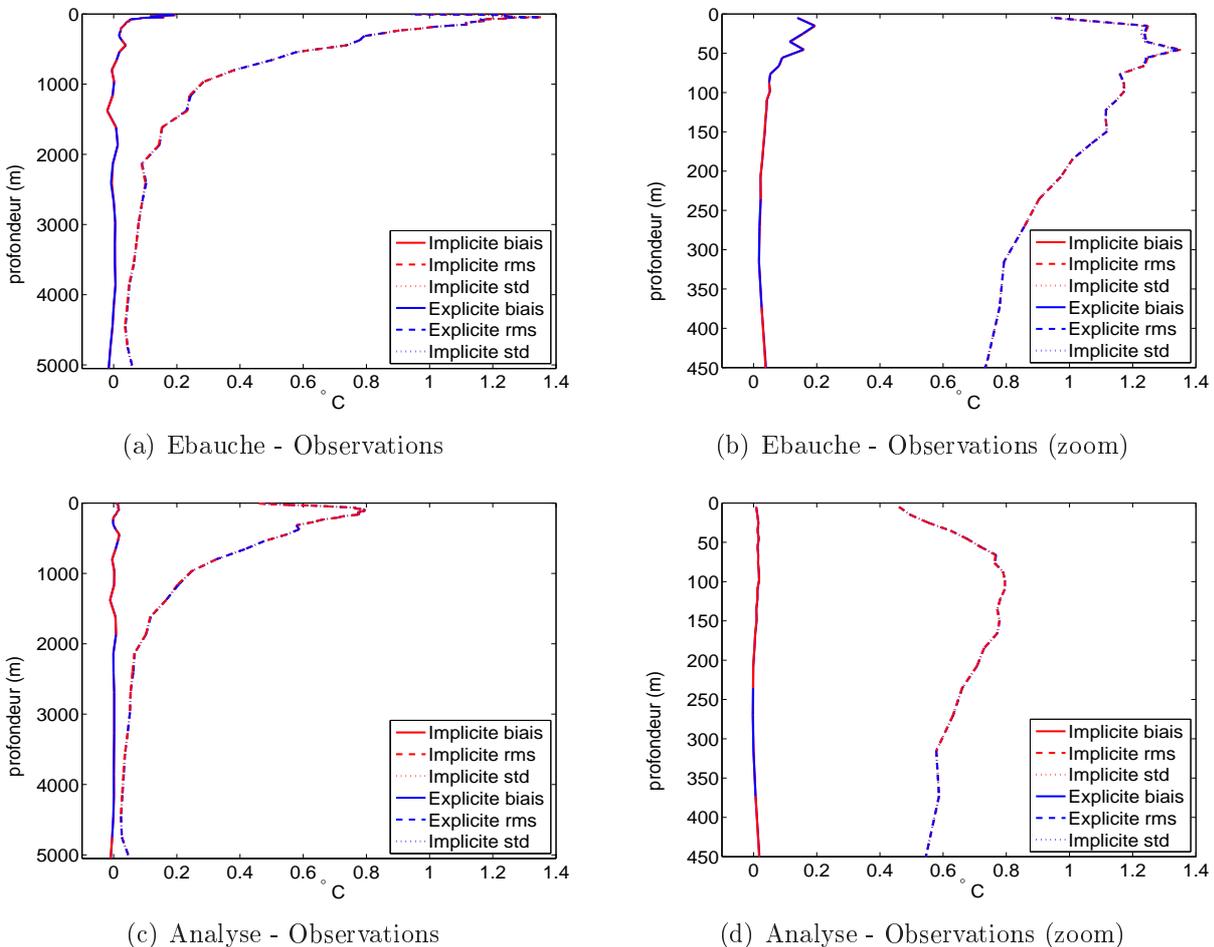


FIG. 5.9 – Diagnostics de performance de l'assimilation de données pour la température moyenne globale. L'écart moyen (trait plein), l'erreur rms (tirets) et l'écart-type (points) sont calculés pour l'écart entre l'ébauche et les observations à assimiler d'une part (a, b) et l'écart entre l'analyse et les observations assimilées d'autre part (c, d). Les résultats sont donnés sur toute la colonne d'eau (a, c) et sur le détail des 450 premiers mètres (b, d). Les courbes bleues concernent l'expérience de "référence" (explicite, CONGRAD avec réorthogonalisation) et les rouges l'expérience de "contribution" (implicite, CGMOD sans réorthogonalisation). Communication personnelle T. Pangaud (2010).

Globalement, les statistiques obtenues pour l'expérience de "contribution" (courbes rouges) sont similaires à celles obtenues pour l'expérience de "référence" (courbes bleues). On peut donc en conclure, que l'utilisation d'un opérateur de diffusion implicite  $3 \times 1D$  est statistiquement équivalente à celle d'un opérateur de diffusion explicite  $2D+1D$ . Néanmoins, comme le montrent les résultats précédents, la différence de coût entre les deux expériences est substantielle. En effet, pour chaque cycle d'assimilation, le temps CPU d'une expérience identique à celle de référence mais avec  $M = 200$  itérations pour l'horizontale (soit 2 fois moins) est réduite d'environ 30% (principalement dû à l'opérateur implicite) tandis que la consommation mémoire chute d'un facteur 1,5 (dû à l'absence de réorthogonalisation de CGMOD).

### 5.3.4 Variation géographique des échelles

L'inconvénient majeur de l'opérateur de diffusion explicite est que son nombre d'itérations est directement lié à un critère de stabilité qui dépend du carré de l'échelle de corrélation. La configuration "standard" pour ORCA2 (ORCA1) consiste à paramétrer les échelles horizontales à 4° (2°) avec un étirement zonal et un rétrécissement méridien autour de l'équateur ( $\pm 15^\circ$ ). Ces variations sont obtenues sous la forme d'un facteur multiplicatif sinusoïdal dont l'extremum situé sur l'équateur donne une échelle de 8° (4°) en zonal et 2° (1°) en méridien, avant de revenir au 4° (2°) paramétré aux alentours des  $\pm 15^\circ$  de latitude. Avec ce paramétrage, le nombre d'itérations requis est de 200 (200 également puisque les échelles ont été divisées par 2). Pour la verticale, le paramétrage s'obtient en imposant un facteur multiplicatif à la taille de la maille. Lorsque ce facteur est de 1, le nombre d'itérations requis est de 10. Mais ce paramétrage repose plus sur des propriétés numériques (pas spatial au moins équivalent à la maille pour les différences finies) que statistiques ou physiques. Pourtant, estimer objectivement ces échelles de corrélation devrait bénéficier aux performances de l'assimilation, à condition bien sûr de pouvoir les prendre en compte, ce que permet l'opérateur de diffusion implicite sans aucun coût supplémentaire, puisqu'il n'est soumis à aucun critère de stabilité.

Daget (2008) construit une méthode d'ensemble pour OPAVAR (prédécesseur de NEMOVAR) permettant d'estimer les autocorrélations de chaque variable en chaque point de la grille ORCA2 à partir des différences entre les éléments issus de 8 membres perturbés et d'1 membre non perturbé pour chaque cycle d'assimilation de 10 jours sur une période de 45 ans (de janvier 1960 à décembre 2005), soit un total de 13448 éléments. Les échelles de corrélation locales sont ensuite estimées en utilisant la formule donnée par Pannekoucke *et al.* (2008) :

$$L_g = \frac{\delta x}{\sqrt{-2 \ln \rho(\delta x)}}, \quad (5.23)$$

où  $\rho(\delta x)$  est la corrélation du point de grille pour la distance  $\delta x$ . L'équation (5.23) est obtenue en supposant que les fonctions de corrélation sont gaussiennes, *i.e.*  $\rho(\delta x) = e^{-\delta x^2 / 2L_g^2}$ , et en inversant cette dernière équation. Grâce à un calcul directionnel, Daget (2008) estime pour chaque variable traitée, des échelles de corrélation dans chaque direction. En se restreignant à la variable de température, nous étudions ici la possibilité de prendre en compte ces échelles par un opérateur de diffusion implicite.

Les échelles de corrélation estimées par l'étude de Daget (2008) sont des valeurs brutes. On ajoute donc un simple contrôle permettant de les borner entre un minimum correspondant à la taille de la maille et un maximum de 1000 km à l'horizontale (soit 9°) et 1000 m à la verticale. Aucun contrôle supplémentaire n'est effectué, puisqu'il s'agit ici de montrer la possibilité de prendre en compte ces échelles, et non pas de déterminer leur validité. La figure 5.10 présente l'exemple des échelles zonales de surface en degré. De manière générale, elles sont inférieures aux 4° précédemment paramétrés, en particulier dans l'hémisphère nord et près des côtes et on ne note pas d'augmentation franche de ces valeurs dans la zone équatoriale. Néanmoins, de fortes valeurs (limitées à 1000 kilomètres par le contrôle) apparaissent dans les zones subtropicales et dans l'hémisphère sud. Un schéma similaire est visible pour les échelles méridiennes (non montré), excepté pour la zone équatoriale où les valeurs sont effectivement réduites à moins de 2°. Quant aux échelles verticales, on constate dans la couche de mélange où les mailles sont d'environ de 10 mètres de profondeur, que certaines échelles du Pacifique Tropical ou de la zone circumpolaire présentent des coefficients multiplicatifs de l'ordre de 5 mais pouvant localement aller jusqu'à 40 (non montré). Globalement, les échelles montrent des caractéristiques liées à la densité des observations (plus courtes dans les zones densément observées) et aux dynamiques océaniques (Pacifique Tropical, courants de bord ouest comme le Gulf Stream ou le Kuroshio).

Pour prendre en compte ces échelles avec un opérateur implicite 2D+1D, il est nécessaire de calculer le nombre d'itérations nécessaire permettant de respecter le critère de stabilité. Or, les rapports échelle locale sur taille de la maille montrent des maxima de 21 pour les axes  $x$  et  $y$ , et de 44 pour l'axe  $z$ .

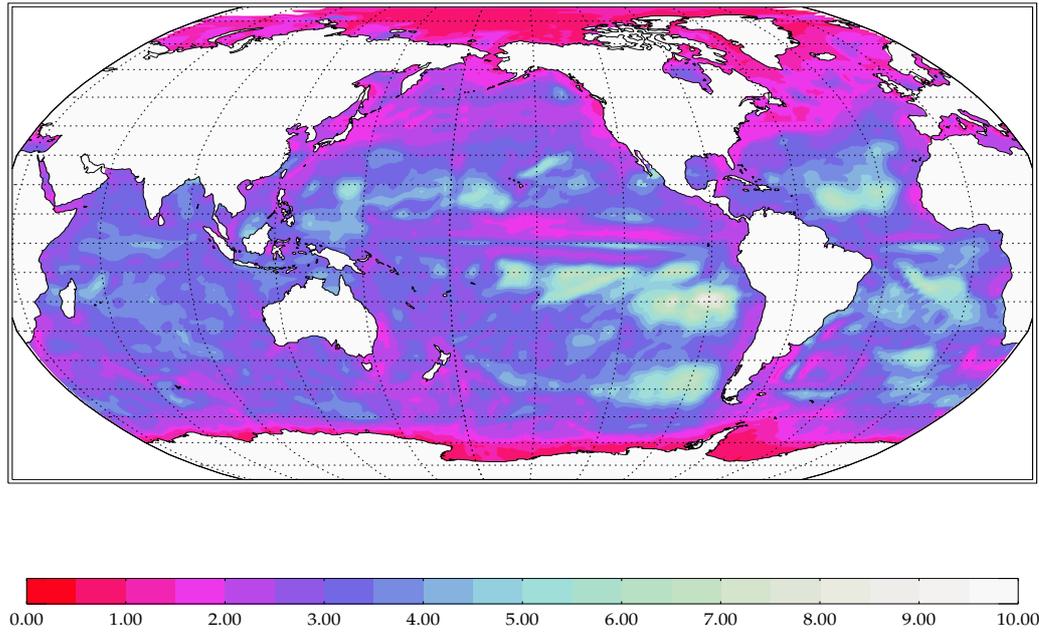


FIG. 5.10 – Echelles de corrélation zonales de surface en degré (Daget, 2008).

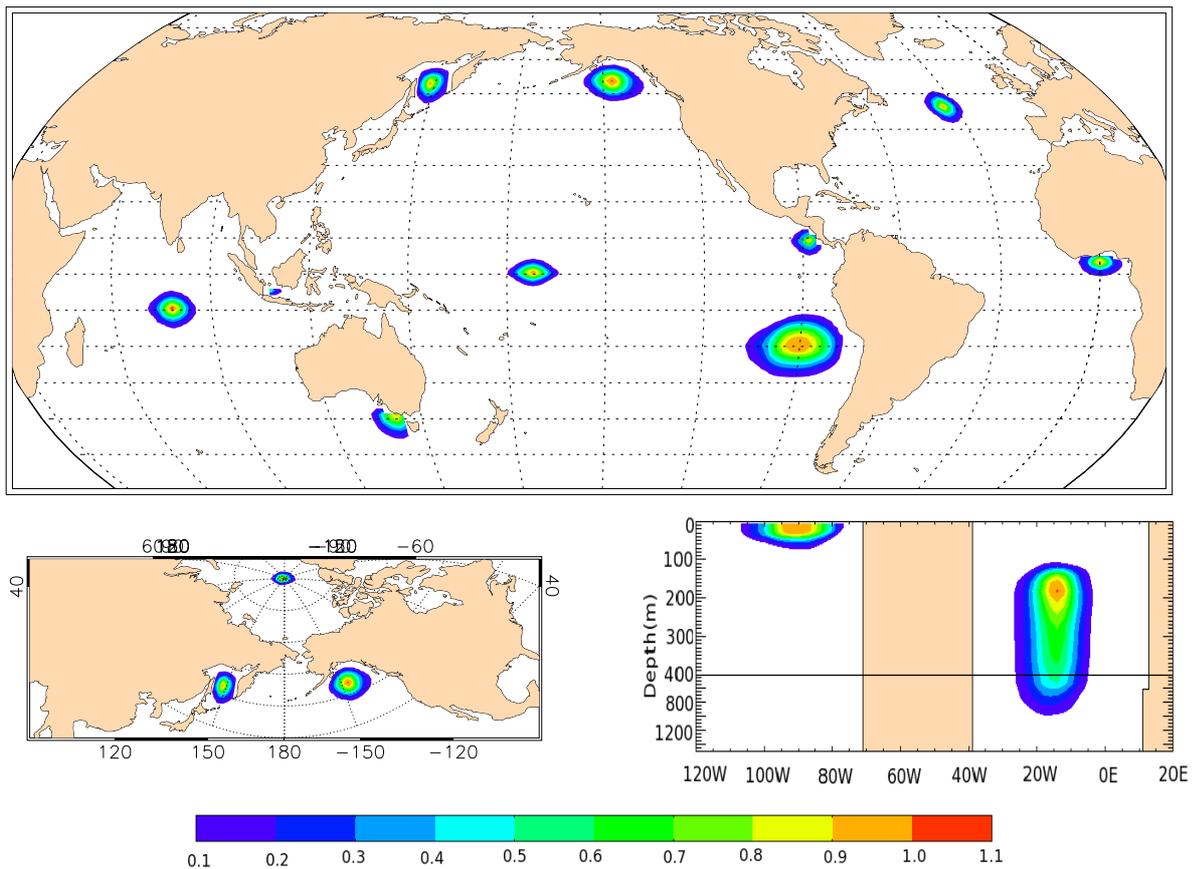


FIG. 5.11 – Exemple de corrélations de température issues de l'application d'un opérateur de diffusion implicite  $3 \times 1D$  avec normalisation par vecteur aléatoire de 1000 membres. Les échelles de corrélations sont issues d'une méthode d'ensemble (Daget, 2008). Les résultats sont donnés en surface pour le globe (en haut) et au Pôle Nord (en bas à gauche). Les latitudes  $y$  sont représentées tous les  $10^\circ$  et les longitude tous les  $30^\circ$ . En bas à droite, une coupe est réalisée à la latitude  $20^\circ S$  dans le Pacifique Est et l'Atlantique, où une corrélation a été calculée par rapport à un point situé à 182 mètres de profondeur. La profondeur est représentée de 0 à 1500 mètres avec un zoom sur les 400 premiers mètres.

Soit, un nombre d'itérations minimal pour l'horizontale et la verticale de

$$M_h > 2 \times 21^2 \approx 900, \quad M_v > 44^2 \approx 1900.$$

Prendre en compte ces nombres d'itérations peut être extrêmement coûteux. En revanche, l'opérateur implicite  $3 \times 1D$  peut être appliqué sans aucun coût supplémentaire. La figure 5.11 montre un exemple de corrélations obtenues avec ces échelles, pour une condition initiale constituée de fonctions de Dirac situées en différents points géographiques. L'opérateur de diffusion implicite  $3 \times 1D$  est associé à une normalisation par vecteur aléatoire de 1000 membres. Le nombre d'itérations est fixé à  $M = 10$  sur chaque axe. On voit en particulier l'effet des échelles allongées sur tous les axes dans le Pacifique subtropical et l'Atlantique subtropical (coupe verticale), et notamment sur la verticale, l'effet d'un facteur multiplicatif important sur des mailles qui s'étirent rapidement avec la profondeur (de 20 à 115 mètres en 5 mailles). A contrario, on remarque les échelles courtes sur le Pôle Nord. On note que contrairement à la configuration ORCA1 utilisée pour la figure 5.7, le détroit de la Sonde (Indonésie) est fermé pour la configuration ORCA2. On note également quelques artefacts (détroits de Bass par exemple) strictement liés à la représentation graphique sur une grille grossière.

Si la prise en compte d'échelles de corrélation issues d'un ensemble est techniquement faisable avec un opérateur de diffusion explicite, le critère de stabilité de ce schéma peut imposer des temps de calcul prohibitifs. L'utilisation d'un opérateur de diffusion implicite permet en revanche de prendre en compte ces échelles sans aucun surcoût par rapport aux échelles paramétrées.

## 5.4 Précision de la normalisation

Dans les paragraphes précédents, nous avons exploré les performances des opérateurs de diffusion implicite  $2$  et  $3 \times 1D$  pour représenter des opérateurs de covariances. Ici, nous étudions la procédure de normalisation qui permet de transformer ces opérateurs de covariance en opérateurs de corrélation.

### 5.4.1 Les différentes méthodes de normalisation

Une première technique pour normaliser les opérateurs de diffusion est la méthode par vecteur aléatoire, alternative moins coûteuse de la méthode exacte. Lorsque les paramètres ne varient pas dans le temps (*i.e.* de cycle à cycle), il est possible d'effectuer cette procédure au préalable, et d'en stocker le résultat dans un fichier qu'il sera possible de lire lors de l'initialisation de la boucle interne. Mais lorsque ces paramètres varient, comme c'est le cas par exemple lorsque l'on introduit une dépendance à l'écoulement, la procédure de normalisation doit être appliquée en début de chaque cycle. Or, chaque membre nécessite l'application de l'opérateur direct de diffusion en  $M/2$  si sa formule est factorisée en "racines carrées", et en  $M$  sinon. On cherche donc naturellement à réduire le plus possible le nombre de membres. Dans le paragraphe 5.3, nous avons choisi d'utiliser 1000 membres pour nos tests. Souvent, lorsque la méthode est utilisée "on-line", le nombre de membres est réduit à 100 pour des raisons de coût. Lorsqu'elle est appliquée au préalable, le nombre de membres peut alors être significativement augmenté (typiquement 10000 membres sont utilisés par la configuration pré-opérationnelle pour les prévisions saisonnières du CEPMMT).

S'affranchir de la méthode de normalisation par vecteur aléatoire nécessite de connaître le facteur théorique à appliquer. Au paragraphe 4.3.2, nous avons pu établir, que le coefficient de normalisation de l'opérateur de diffusion implicite  $1D$  était donné pour un paramètre d'échelle  $L$  constant, par :

$$\lambda_M = \nu_M L,$$

avec

$$\nu_M = \frac{2^{2M-1} [(M-1)!]^2}{(2M-2)!}.$$

Le calcul de factorielles pouvant s'avérer périlleux, on remarque en partant de  $\nu_1 = 2$ , qu'il existe une relation de récurrence, en écrivant pour  $i = 1, \dots, M - 1$  :

$$\nu_{i+1} = \frac{2^{2i+2-1} [(i+1-1)!]^2}{(2i+2-2)!} = \frac{2^2 \times 2^{2i-1} [i \times (i-1)!]^2}{2i(2i-1) \times (2i-2)!} = \frac{2i}{2i-1} \nu_i.$$

Prendre en compte des échelles qui varient géographiquement est possible, et dans l'hypothèse où les fonctions de corrélations sont séparables, le facteur de normalisation de l'opérateur 2 ou 3×1D correspond au produit des facteurs de normalisation 1D. Lorsque la variation est suffisamment "lente", une approximation du facteur de normalisation théorique est donc donnée par

$$\tilde{\lambda}_M(x, y, z) = \nu_M \tilde{L}_x(x, y, z) \nu_M \tilde{L}_y(x, y, z) \nu_M \tilde{L}_z(x, y, z) = \nu_M^3 \tilde{L}(x, y, z), \quad (5.24)$$

où

$$\tilde{L}(x, y, z) = \tilde{L}_x(x, y, z) \tilde{L}_y(x, y, z) \tilde{L}_z(x, y, z),$$

et  $\tilde{L}_x(x, y, z)$ ,  $\tilde{L}_y(x, y, z)$  et  $\tilde{L}_z(x, y, z)$  correspondent aux paramètres d'échelle locaux interpolés au point de la maille adéquat. Prenons en effet le cas de la variable température. Sur la gille "C" selon Arakawa, cette variable est définie au point T, *i.e.* au centre de la maille. Mais le coefficient de diffusion nécessaire à la résolution de l'équation de diffusion est appliqué sur la première dérivée spatiale du champ de température. Cette dérivée étant calculée par des différences finies centrées, elle est donc définie sur la face de la maille. Le coefficient de diffusion, et par conséquent l'échelle de corrélation doit donc également être définie sur la face de la maille. La seconde dérivée spatiale du champ de température, et donc du champ filtré, est à nouveau définie au centre de la maille. Ainsi, si les paramètres d'échelle  $L_x(x, y, z)$ ,  $L_y(x, y, z)$  et  $L_z(x, y, z)$  utilisés pour définir les coefficients de diffusion sont respectivement définis aux points U, V et W pour la température, les paramètres d'échelle  $\tilde{L}_x(x, y, z)$ ,  $\tilde{L}_y(x, y, z)$  et  $\tilde{L}_z(x, y, z)$  utilisés dans le calcul de la normalisation doivent être définis au point T. Dans le système de coordonnées curvilinéaires, l'interpolation bilinéaire suivante est donc appliquée en fonction de la longitude (*lon*), de la latitude (*lat*) et de la profondeur (*pro*) :

$$\begin{aligned} \tilde{L}_{x,i}^2 &= \left[ L_{x,i-\frac{1}{2}}^2 \left( lon_{i+\frac{1}{2}} - lon_i \right) + L_{x,i+\frac{1}{2}}^2 \left( lon_i - lon_{i-\frac{1}{2}} \right) \right] / \left( lon_{i+\frac{1}{2}} - lon_{i-\frac{1}{2}} \right), \\ \tilde{L}_{y,j}^2 &= \left[ L_{y,j-\frac{1}{2}}^2 \left( lat_{j+\frac{1}{2}} - lat_j \right) + L_{y,j+\frac{1}{2}}^2 \left( lat_j - lat_{j-\frac{1}{2}} \right) \right] / \left( lat_{j+\frac{1}{2}} - lat_{j-\frac{1}{2}} \right), \\ \tilde{L}_{z,k}^2 &= \left[ L_{z,k-\frac{1}{2}}^2 \left( pro_k - pro_{k+\frac{1}{2}} \right) + L_{z,k+\frac{1}{2}}^2 \left( pro_{k-\frac{1}{2}} - pro_k \right) \right] / \left( pro_{k-\frac{1}{2}} - pro_{k+\frac{1}{2}} \right). \end{aligned}$$

On rappelle que le système de coordonnées curvilinéaires de NEMO est défini tel que  $i$  s'incrémente de l'ouest vers l'est,  $j$  du Pôle Sud vers le Pôle Nord et  $k$  de la surface vers les profondeurs. Aux hautes latitudes nord, cette approximation bilinéaire peut cependant devenir insatisfaisante du fait de l'incurvation de la grille.

La méthode proposée par Purser *et al.* (2003a) (voir Annexe E) semble donner de bons résultats dans un cas 1D académique (voir paragraphe 4.4.1). On tente donc d'étendre cette méthode au cas 3D en proposant l'approximation suivante :

$$\bar{\lambda}_M(x, y, z) = \nu_M \bar{L}_x(x, y, z) \nu_M \bar{L}_y(x, y, z) \nu_M \bar{L}_z(x, y, z) = \nu_M^3 \bar{L}(x, y, z), \quad (5.25)$$

avec

$$\begin{aligned} \bar{L}(x, y, z) &= \bar{L}_x(x, y, z) \bar{L}_y(x, y, z) \bar{L}_z(x, y, z) \\ &= \sqrt{\mathcal{L}_x^{M/2} \left[ \tilde{L}_x^2(x, y, z) \right] \mathcal{L}_y^{M/2} \left[ \tilde{L}_y^2(x, y, z) \right] \mathcal{L}_z^{M/2} \left[ \tilde{L}_z^2(x, y, z) \right]}. \end{aligned}$$

### 5.4.2 Simulation de frontières transparentes

L'opérateur de diffusion utilisé avec des conditions aux bords de Neumann (ou de Dirichlet) génère un opérateur de corrélation dont le noyau est corrompu près des frontières (voir paragraphe 4.4.2). Lorsque la normalisation est effectuée par la méthode par vecteur aléatoire, l'altération consiste en une modification de l'échelle de corrélation. Mais lorsque la normalisation est donnée par le facteur de normalisation théorique ou son approximation, la déformation peut être très importante. Pour palier ce problème, il est possible de simuler des frontières transparentes en définissant l'opérateur de diffusion comme la moyenne de deux problèmes de diffusion, l'un avec des conditions aux frontières de Neumann, et l'autre avec des conditions aux frontières de Dirichlet :

$$\mathbf{L}^M = \frac{1}{2} [\mathbf{L}_N^M + \mathbf{L}_D^M],$$

où les indices N et D donnent les conditions aux frontières utilisées, soit Neumann et Dirichlet respectivement. Ecrire cet opérateur sous une forme factorisée de termes  $\mathbf{L}^{M/2}$  demande néanmoins quelques manipulations. En effet, si  $\mathbf{L}^{M/2}$  est construit comme une somme, des termes croisés apparaissent :

$$\begin{aligned} & \frac{1}{2} (\mathbf{L}_N^{M/2} + \mathbf{L}_D^{M/2}) \mathbf{W}^{-1} (\mathbf{L}_N^{M/2} + \mathbf{L}_D^{M/2})^T \\ &= \mathbf{L}^M \mathbf{W}^{-1} + \frac{1}{2} \mathbf{L}_N^{M/2} \mathbf{W}^{-1} (\mathbf{L}_D^{M/2})^T + \frac{1}{2} \mathbf{L}_D^{M/2} \mathbf{W}^{-1} (\mathbf{L}_N^{M/2})^T. \end{aligned}$$

Pour éviter ce problème, l'opérateur  $\mathbf{L}^{M/2}$  doit être construit sous la forme d'une matrice rectangulaire. Le nombre de termes dans les matrices rectangulaires dépend de la dimension. Pour l'opérateur  $2 \times 1D$ , il est de  $2^2 = 4$  tandis que pour l'opérateur  $3 \times 1D$ , il est de  $2^3 = 8$  :

$$\mathbf{L}^M \mathbf{W}^{-1} = \frac{1}{8} \mathbf{L}^{M/2} \begin{pmatrix} \mathbf{W}^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W}^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{W}^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{W}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{W}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{W}^{-1} \end{pmatrix} (\mathbf{L}^{M/2})^T,$$

avec

$$\begin{aligned} \mathbf{L}^{M/2} &= \begin{pmatrix} \mathbf{L}_{N_x}^{M/2} \mathbf{L}_{N_y}^{M/2} \mathbf{L}_{N_z}^{M/2} & \mathbf{L}_{N_x}^{M/2} \mathbf{L}_{N_y}^{M/2} \mathbf{L}_{D_z}^{M/2} & \mathbf{L}_{N_x}^{M/2} \mathbf{L}_{D_y}^{M/2} \mathbf{L}_{N_z}^{M/2} & \mathbf{L}_{N_x}^{M/2} \mathbf{L}_{D_y}^{M/2} \mathbf{L}_{D_z}^{M/2} \\ \mathbf{L}_{D_x}^{M/2} \mathbf{L}_{N_y}^{M/2} \mathbf{L}_{N_z}^{M/2} & \mathbf{L}_{D_x}^{M/2} \mathbf{L}_{N_y}^{M/2} \mathbf{L}_{D_z}^{M/2} & \mathbf{L}_{D_x}^{M/2} \mathbf{L}_{D_y}^{M/2} \mathbf{L}_{N_z}^{M/2} & \mathbf{L}_{D_x}^{M/2} \mathbf{L}_{D_y}^{M/2} \mathbf{L}_{D_z}^{M/2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{L}_{N_x}^{M/2} \left( \mathbf{L}_{N_y}^{M/2} \left( \mathbf{L}_{N_z}^{M/2} & \mathbf{L}_{D_z}^{M/2} \right) \right) & \mathbf{L}_{D_y}^{M/2} \left( \mathbf{L}_{N_z}^{M/2} & \mathbf{L}_{D_z}^{M/2} \right) \\ \mathbf{L}_{D_x}^{M/2} \left( \mathbf{L}_{N_y}^{M/2} \left( \mathbf{L}_{N_z}^{M/2} & \mathbf{L}_{D_z}^{M/2} \right) \right) & \mathbf{L}_{D_y}^{M/2} \left( \mathbf{L}_{N_z}^{M/2} & \mathbf{L}_{D_z}^{M/2} \right) \end{pmatrix}. \end{aligned} \quad (5.26)$$

L'opérateur direct donné par l'équation (5.26) nécessite en entrée un vecteur de 8 fois la taille du champ auquel est appliqué l'opérateur de diffusion. Chaque partie est ensuite associée aux 8 parties de la matrice rectangulaire avant d'en effectuer la somme pour finir. Ainsi, appliquer l'opérateur direct nécessite 8 applications de  $\mathbf{L}_z$ , 8 applications de  $\mathbf{L}_y$  et de la réorganisation des processeurs inhérente, et 8 applications de  $\mathbf{L}_x$  et de la réorganisation des processeurs inhérente. Pour l'opérateur adjoint, c'est le vecteur résultant de son application qui possède une taille de 8 fois celle du champ auquel est appliqué l'opérateur de diffusion. Ainsi, l'utiliser nécessite toujours 8 applications de  $\mathbf{L}_z$ , mais seulement 4 applications de  $\mathbf{L}_y$  et de la réorganisation des processeurs inhérente, et 2 applications de  $\mathbf{L}_x$  et de la réorganisation des processeurs inhérente. Ainsi, l'opérateur adjoint est moins coûteux que l'opérateur direct, et le choix de l'ordre des opérateurs 1D influence ce coût. Un test de profilage sur le PC de bureau *DELL Precision T5500* de la boucle interne pour la configurations ORCA2 donne un coût de 15,7 secondes pour 42 applications de la "racine carrée" de l'opérateur de corrélation direct et 9,0

secondes pour 42 applications de la “racine carrée” de l'opérateur de corrélation adjoint. On notera que l'opérateur de corrélation donné par l'équation (5.19) et étendu à 3 dimensions comprend 6 termes dans sa moyenne. Dans ce cas, utiliser la combinaison de deux problèmes de diffusion, l'un avec des conditions aux frontières de Neumann et l'autre avec des conditions aux frontières de Dirichlet, conduit à  $6 \times 8 = 48$  termes dans la moyenne.

On rappelle d'autre part, que l'utilisation de l'algorithme de minimisation CONGRAD requiert un changement de variable, tel que le problème de minimisation à résoudre devient

$$(\mathbf{I} + \mathbf{U}^T \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \mathbf{U}) \mathbf{v} = \mathbf{U}^T \mathbf{G}^T \mathbf{R}^{-1} \mathbf{d},$$

où  $\mathbf{B} = \mathbf{U} \mathbf{U}^T$ . En définissant  $\mathbf{U}$  et  $\mathbf{U}^T$  à partir de la matrice rectangulaire de l'équation (5.26), on voit bien que la minimisation devra se dérouler sur des vecteurs de 8 fois la taille du vecteur de contrôle, augmentant d'autant le coût d'exécution et surtout le besoin en mémoire. L'algorithme CGMOD offre alors une alternative attrayante puisqu'il ne nécessite pas ce changement de variable, et ne requiert que l'application du  $\mathbf{B}$  complet. Même si ce dernier est appliqué par l'intermédiaire de ses “racines carrées” pour garantir les propriétés de symétrie et de définition positive, celles-ci sont appliquées consécutivement et la taille du vecteur n'est alors augmentée que de manière intermédiaire.

Les configurations que nous utilisons ici sont des configurations globales à plutôt basse résolution, intéressantes par exemple, pour étudier la circulation océanique à grande échelle. Les régions côtières y ont donc moins d'importance et l'assimilation de données près des côtes n'est généralement pas privilégiée. La surface en revanche reste une frontière primordiale. En effet, une grande partie des observations pour l'océanographie se trouvent près de la surface. De plus, c'est à travers la surface que les forçages atmosphériques se transmettent à tout l'océan. Dans ce cadre, on peut souhaiter diminuer le coût de l'opérateur défini par l'équation (5.26) en limitant la simulation de frontières transparentes à la verticale. L'opérateur de diffusion est alors défini par

$$\mathbf{L}^M \mathbf{W}^{-1} = \frac{1}{2} \mathbf{L}_x^{M/2} \mathbf{L}_y^{M/2} \begin{pmatrix} \mathbf{L}_{Nz}^{M/2} & \mathbf{L}_{Dz}^{M/2} \\ \mathbf{L}_{Nz}^{M/2} & \mathbf{L}_{Dz}^{M/2} \end{pmatrix} \begin{pmatrix} \mathbf{W}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^{-1} \end{pmatrix} \begin{pmatrix} \left( \mathbf{L}_{Nz}^{M/2} \right)^T \\ \left( \mathbf{L}_{Dz}^{M/2} \right)^T \end{pmatrix} \begin{pmatrix} \left( \mathbf{L}_y^{M/2} \right)^T \\ \left( \mathbf{L}_x^{M/2} \right)^T \end{pmatrix}.$$

### 5.4.3 Méthode d'évaluation

Générer des corrélations en des points géographiques particuliers et en relever le maximum (qui doit être à 1), permet d'obtenir une première idée de la performance de la normalisation appliquée. Pour mieux évaluer l'erreur engendrée, on se sert du principe de la méthode par vecteur aléatoire. Mais on va ici l'appliquer à l'opérateur de corrélation total plutôt qu'à l'opérateur de diffusion seul, et le nombre de membres sera porté à  $N = 10000$ . A partir du vecteur aléatoire  $\mathbf{v}$ , tel que  $E[\mathbf{v}] = \mathbf{0}$  et  $E[\mathbf{v}\mathbf{v}^T] = \mathbf{I}$ , on estime

$$\text{diag}(\mathbf{C}) = \text{diag} \left( \frac{1}{N-1} \sum_{n=1}^N E \left[ \mathbf{C}^{1/2} \mathbf{v} \left( \mathbf{C}^{1/2} \mathbf{v} \right)^T \right] \right), \quad (5.27)$$

puisque

$$E \left[ \mathbf{C}^{1/2} \mathbf{v} \left( \mathbf{C}^{1/2} \mathbf{v} \right)^T \right] = \mathbf{C}^{1/2} E[\mathbf{v}\mathbf{v}^T] \left( \mathbf{C}^{1/2} \right)^T = \mathbf{C}.$$

Les valeurs de la diagonale ainsi estimée doivent être proches de 1.

Lorsque des frontières transparentes sont simulées pour les trois axes, la procédure de l'équation (5.27) doit être modifiée pour prendre en compte la matrice rectangulaire définie par l'équa-

tion (5.26). A partir du vecteur aléatoire  $\mathbf{v}$ , tel que  $E[\mathbf{v}] = 0$  et  $E[\mathbf{v}\mathbf{v}^T] = \mathbf{I}$ , on définit les vecteurs

$$\mathbf{V}_1 = \begin{pmatrix} \mathbf{v} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{V}_2 = \begin{pmatrix} \mathbf{0} \\ \mathbf{v} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad \dots \quad \mathbf{V}_8 = \begin{pmatrix} \mathbf{0} \\ \mathbf{v} \end{pmatrix}.$$

On a alors pour  $N = 10000$  membres

$$\text{diag}(\mathbf{C}) = \text{diag} \left( \frac{1}{N-1} \sum_{n=1}^N \frac{1}{8} \sum_{k=1}^8 E \left[ \mathbf{C}^{1/2} \mathbf{V}_k (\mathbf{C}^{1/2} \mathbf{V}_k)^T \right] \right). \quad (5.28)$$

Normalisation		$\mathbf{L}^{M/2}$	Evaluation	Désignation
Méthode par vecteur aléatoire	$P = 100$	Eq. (5.29)	Eq. (5.27)	N° 1
	$P = 1000$			N° 2
Méthode par approximation	$\tilde{\lambda}_M$ , Eq. (5.24)	Eq. (5.26)	Eq. (5.28)	N° 3
	$\tilde{\lambda}_M$ , Eq. (5.24)			N° 4
	$\bar{\lambda}_M$ , Eq. (5.25)			N° 5

TAB. 5.5 – Détail des cinq configurations de normalisation évaluées.

Le tableau 5.5 résume les cinq configurations évaluées, et dont les résultats sont rapportés dans les paragraphes suivants. Deux types de méthodes de normalisation sont évaluées ici : les méthodes par vecteur aléatoire avec  $P = 100$  ou  $P = 1000$  membres, et les méthodes par approximation du facteur théorique. Pour les premières, la “racine carrée” de l’opérateur de diffusion est donnée par l’application d’un seul problème de diffusion avec des conditions aux frontières de Neumann :

$$\mathbf{L}^{M/2} = \mathbf{L}_{N_x}^{M/2} \mathbf{L}_{N_y}^{M/2} \mathbf{L}_{N_z}^{M/2}. \quad (5.29)$$

Pour les secondes, on teste également leurs performances lorsque des frontières transparentes sont simulées.

#### 5.4.4 Méthodes par vecteurs aléatoires

En théorie (voir paragraphe 3.6.4), l’erreur commise par la méthode par vecteur aléatoire est donnée par  $1/\sqrt{2P}$ , où  $P$  est le nombre de membres. Pour  $P = 1000$  membres, l’erreur maximale devrait donc être d’un peu plus de 2% et atteindre les 7% pour  $P = 100$  membres. Le tableau 5.6 donne les valeurs maximales obtenues pour chaque corrélation dans le cas de test de la figure 5.11 lorsque  $P = 100$  et  $P = 1000$ . Les valeurs sont notées en rouge gras lorsque leur erreur dépasse la valeur théorique, et en bleu italique sinon. Pour les deux configurations, certaines valeurs sont au-delà de l’erreur théorique (en particulier pour  $P = 100$ ) allant jusqu’à 26% et 8% respectivement.

Dans un premier temps, les configurations N°1 ( $P = 100$ ) et N°2 ( $P = 1000$ ) sont testées sur ORCA1 avec des échelles paramétrées (voir paragraphe 5.3.4). La figure 5.12 montre le pourcentage d’erreur relevé au niveau de la surface. Pour la figure 5.12a (configuration N°1), de nombreuses zones

Zone géographique	Coordonnées		Valeur maximale	
	$(i, j, k)$	(lon,lat,prof)	$P = 100$	$P = 1000$
Raccordement Est/Ouest	(181, 62, 1)	(78°E,9°S,5 m)	<b>0,88</b>	<b>1,08</b>
Indonésie	(15, 66, 1)	(106°E,5°S,5 m)	<b>0,82</b>	<i>1,01</i>
Australie, Dét. de Bass	(33, 45, 1)	(142°E,39°S,5 m)	<b>0,87</b>	<b>0,94</b>
Mer d'Okhotsk	(37, 114, 1)	(148°E,53°N,5 m)	<i>0,94</i>	<i>0,99</i>
Pacifique Equatorial	(56, 75, 1)	(172°W,1°N,5 m)	<b>0,83</b>	<i>0,99</i>
Golfe d'Alaska	(68, 114, 1)	(147°W,54°N,5 m)	<b>1,26</b>	<b>1,03</b>
Pacifique Sud Est	(97, 56, 1)	(90°W,20°S,5 m)	<b>1,18</b>	<i>1,00</i>
Amérique latine	(98, 86, 1)	(88°W,9°N,5 m)	<i>1,02</i>	<b>0,97</b>
Atlantique Nord	(125, 108, 1)	(34°W,46°N,5 m)	<i>1,02</i>	<i>1,02</i>
Golfe de Guinée	(142, 80, 1)	(0°E,3°N,5 m)	<b>1,10</b>	<i>0,98</i>
Océan Arctique	(50, 148, 1)	(16°E,90°N,5 m)	<b>0,85</b>	<i>1,00</i>
Atlantique Sud	(135, 56, 16)	(14°W,20°S,182 m)	<i>0,95</i>	<i>1,01</i>

TAB. 5.6 – Maximum des corrélations de la figure 5.11 pour une normalisation par vecteurs aléatoires de  $P = 100$  et  $P = 1000$  membres. Les valeurs sont notées en rouge gras lorsqu'elles dépassent l'erreur théorique (respectivement 7% et 2%), en bleu italique sinon.

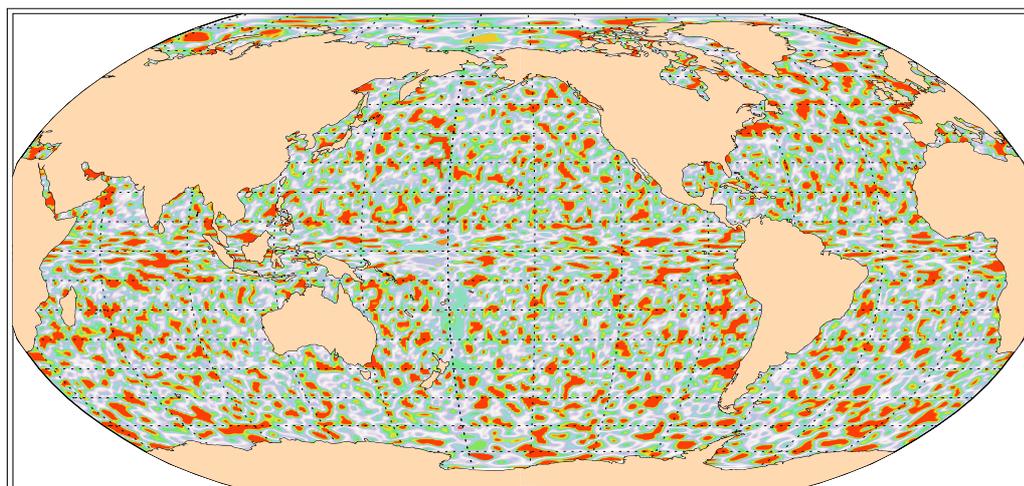
montrent une erreur supérieure à 20%, le maximum pouvant atteindre plus de 100%. Si la figure 5.12b (configuration N° 2) montre une nette amélioration de la normalisation par rapport à la figure 5.12a, on relève néanmoins un certain nombre de zones où l'erreur est aux alentours de 10 à 12%, l'erreur maximale atteignant très localement 25%. Tous les niveaux verticaux présentent le même motif d'erreur dans les deux cas. L'évaluation de la configuration N°2 est ensuite réalisée pour ORCA2 avec des échelles estimées par une méthode d'ensemble (voir paragraphe 5.3.4). Les résultats sont similaires à la figure 5.12a et ne sont donc pas montrés.

Les structures d'erreur qui semblent apparaître ne dépendent pas de la variation des échelles. Elles pourraient provenir en fait du générateur aléatoire utilisé (RAN1 de *Numerical Recipes*, voir Press *et al.*, 1992). La normalisation par vecteurs aléatoires peut générer des erreurs importantes lorsque le nombre de membres utilisé est réduit. C'est pourquoi cette méthode devrait plutôt être utilisée lors d'une procédure préalable à l'assimilation lorsque les paramètres ne varient pas dans le temps, et avec un nombre de membres suffisant.

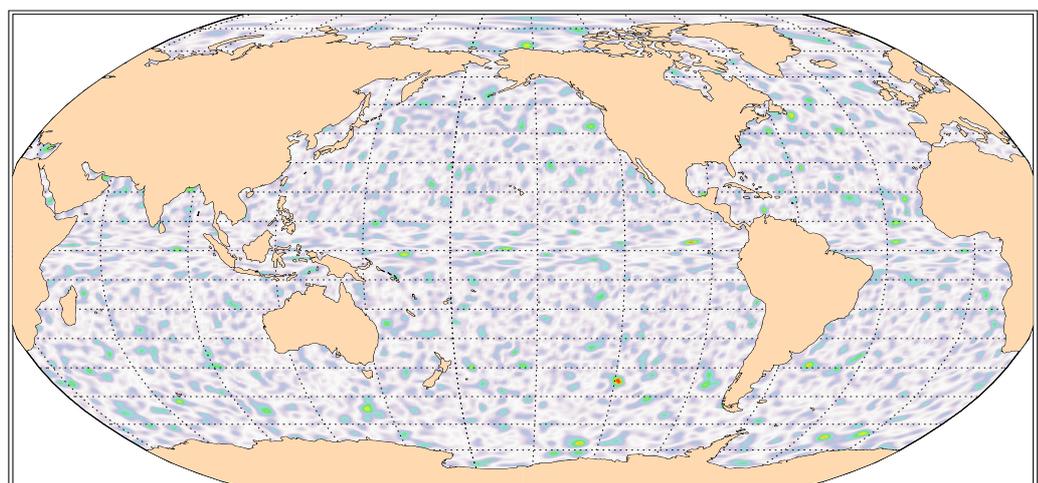
### 5.4.5 Méthodes par approximation du facteur théorique

La configuration N°3 est maintenant testée sur ORCA1 avec des échelles paramétrées. La figure 5.13 montre le pourcentage d'erreur relevé en surface. Loin des côtes, l'erreur de la normalisation se situe autour des 100% (soit un maximum à 2) mais s'atténue rapidement avec la profondeur pour revenir à des valeurs acceptables avant de s'élever à nouveau près de la bathymétrie (non montré). Près des côtes néanmoins, l'erreur peut atteindre les 400 ou 600%, voir plus dans certaines zones particulières comme la Mer Rouge, le nord du Canada ou encore la Mer Baltique. Ceci s'explique par le fait que les conditions aux frontières de Neumann doublent la valeur d'une corrélation calculée pour une fonction de Dirac contre une frontière (voir paragraphe 4.4.2). En plein océan, l'erreur est provoquée par la frontière verticale, mais près des côtes, l'effet des frontières latérales s'y ajoute.

Pour palier ce problème, il est donc nécessaire de simuler des frontières transparentes en définissant l'opérateur de diffusion comme la moyenne de deux problèmes de diffusion, l'un avec des conditions aux frontières de Neumann et l'autre avec des conditions aux frontières de Dirichlet. La figure 5.14 montre le pourcentage d'erreur relevé en surface pour la configuration N° 4. L'amélioration apportée



(a) Configuration N°1, ORCA1, échelles paramétrées



(b) Configuration N°2, ORCA1, échelles paramétrées

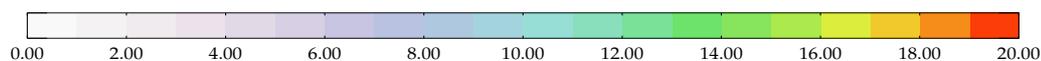


FIG. 5.12 – Pourcentage d’erreur donné par l’évaluation des configurations N° 1 (a) et N° 2 (b). L’exemple est donné en surface pour la ORCA1 avec des échelles paramétrées. La palette est saturée à une erreur de 20%.

par la simulation de frontières transparentes est évidente. On remarque néanmoins (comme pour la figure 5.13) que l’erreur est plus importante aux latitudes comprises entre 10 et 35 degrés (jusqu’à  $\approx 14\%$ ) par rapport à l’équateur et aux hautes latitudes (jusqu’à  $\approx 6\%$ ). En effet, l’erreur numérique due à la discrétisation spatiale de l’équation de diffusion n’est pas prise en compte par ce type de normalisation alors qu’elle l’était implicitement pour la normalisation par vecteurs aléatoires. Le raffinement à l’équateur et le rétrécissement naturel des mailles aux hautes latitudes permet au rapport échelle sur maille de s’accroître et donc de faire baisser cette erreur numérique. Enfin, certaines zones très fermées comme la Mer Rouge ou le nord du Canada, présentent ici aussi une erreur très élevée.

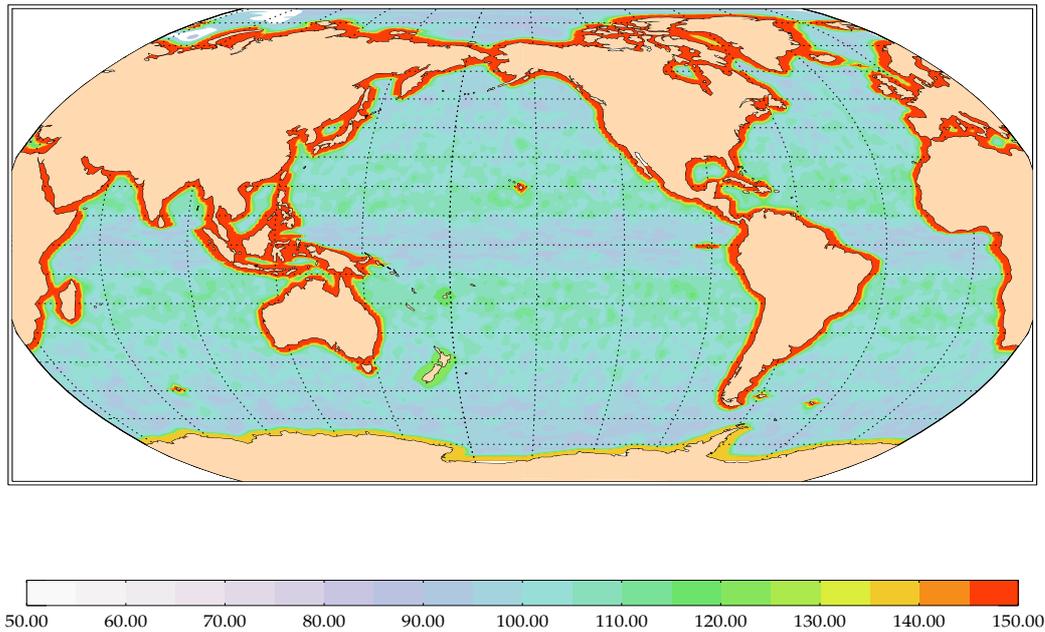


FIG. 5.13 – Pourcentage d'erreur donné par l'évaluation de la configuration N° 3. L'exemple est donné en surface pour ORCA1 avec des échelles paramétrées. La palette est saturée à une erreur de 150%.

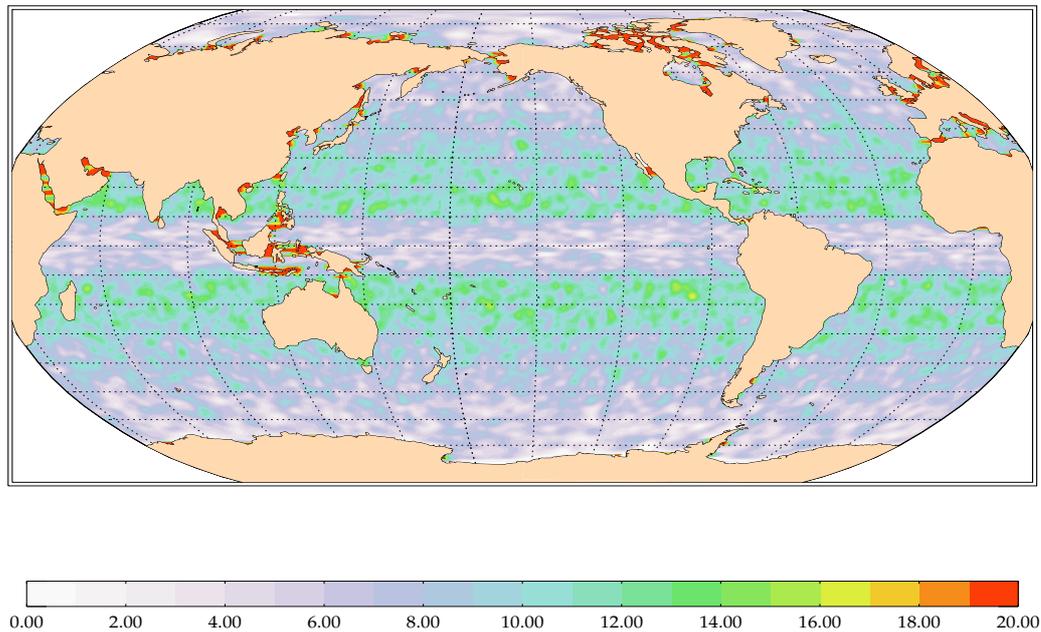


FIG. 5.14 – Pourcentage d'erreur donné par l'évaluation de la configuration N° 4. L'exemple est donné en surface pour ORCA1 avec des échelles paramétrées. La palette est saturée à une erreur de 20%.

En comparant la figure 5.14 avec le résultat donné par la figure 5.12a pour une normalisation par vecteurs aléatoires de  $P = 1000$  membres, on constate que les erreurs les plus importantes (hors zones particulières) sont concentrées entre 10 et 35 degrés de latitude dans le premier cas alors que, moins nombreuses dans le deuxième cas, elles y sont distribuées de façon aléatoire. La précision de la normalisation selon l'équation (5.24) dépend donc directement des erreurs numériques dues à la discrétisation spatiale, et donc du paramétrage des échelles par rapport à la maille. En modifiant par

exemple le facteur de l'échelle de corrélation verticale de telle sorte que cette échelle soit de la même longueur de la maille (au lieu de deux fois la maille ici), l'erreur numérique sera plus importante et par conséquent, l'erreur due à la normalisation sera également plus importante.

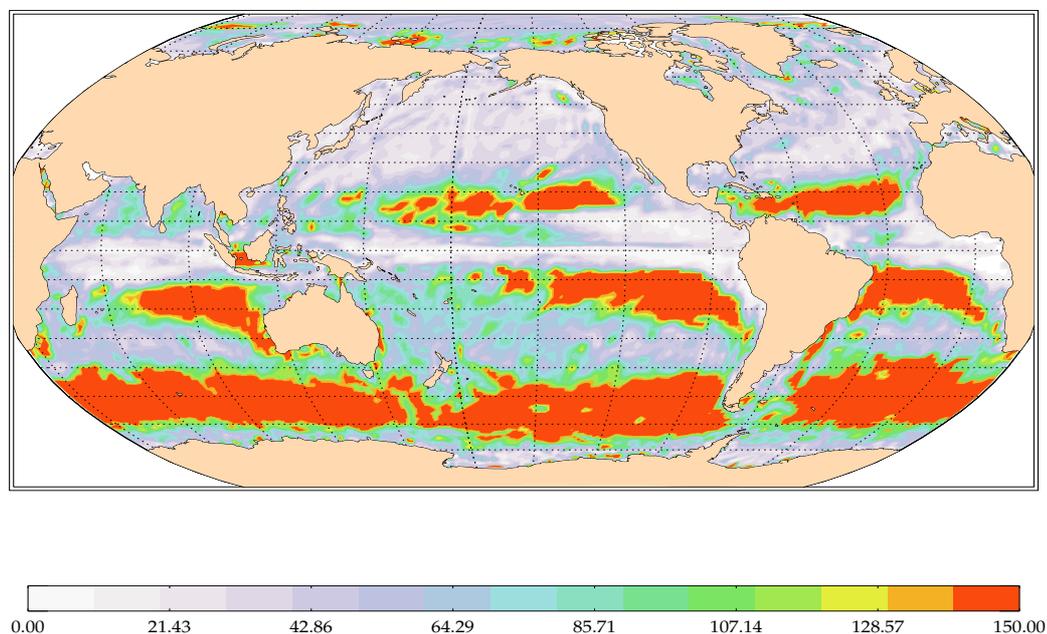


FIG. 5.15 – Pourcentage d'erreur donné par l'évaluation de la configuration N° 4. L'exemple est donné en surface pour ORCA2 avec des échelles estimées par une méthode d'ensemble (Daget, 2008). La palette est saturée à une erreur de 150%.

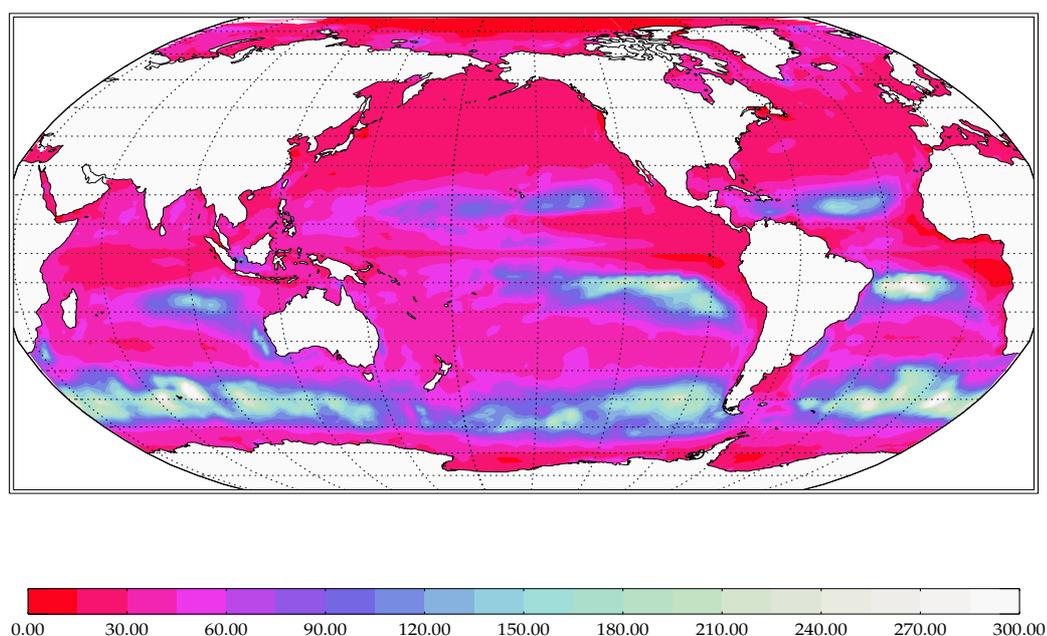


FIG. 5.16 – Echelles de corrélation verticales en mètres pour la surface (Daget, 2008).

On évalue maintenant la configuration N° 4 sur ORCA2 avec des échelles estimées par une méthode d'ensemble (Daget, 2008, voir paragraphe 5.3.4). La figure 5.15 montre le pourcentage d'erreur relevé

en surface. Bien que les latitudes au delà de  $30^\circ$  nord et la bande équatoriale montrent une erreur acceptable, on décerne néanmoins de larges zones dont l'erreur peut dépasser localement les 400%. Or, en rapprochant ce résultat de la figure 5.16, on constate que ces zones d'erreur importante correspondent en fait à des zones où l'échelle de corrélation verticale (et dans une moindre mesure les échelles de corrélation horizontales, voir figure 5.10) est la plus élevée. En descendant dans les niveaux verticaux, ces erreurs s'atténuent rapidement.

Dans ces zones, les échelles verticales se raccourcissent très rapidement avant de se rallonger à partir de 200 mètres de profondeur. Néanmoins, les erreurs à ces niveaux restent tout à fait convenables. Or, si dans les premiers 100 mètres de colonne d'eau les mailles sont approximativement de 10 mètres de profondeur, elles sont d'environ 40 mètres à 200 mètres de profondeur et s'élargissent rapidement. Pour une échelle de corrélation verticale de 250 mètres, le facteur multiplicatif de la maille est donc de 25 sur les premiers niveaux, et de moins de 6 à partir de 200 mètres de profondeur. On pourrait donc penser qu'un facteur multiplicatif trop élevé entraîne une erreur importante. On effectue alors un test de corrélation dans les mêmes conditions que la figure 5.11 mais selon la configuration N° 4. Dans un premier cas, les échelles verticales sont les échelles estimées par la méthode d'ensemble, et dans un deuxième cas, les échelles verticales sont paramétrées à 25 fois la longueur de la maille sur les 15 premiers niveaux, puis à 2 fois la maille pour le reste des niveaux. Dans le premier cas, le maximum de la corrélation située à l'ouest du Pérou est de 6,44 alors qu'elle n'est plus que de 2,57 dans le deuxième cas. L'importance de l'erreur semble donc liée aux brusques variations des échelles plutôt qu'à leur longueur proprement dite.

La figure 5.17 représente des corrélations de température (courbes rouges) dans le Pacifique Equatorial (a, c, e) et dans le détroit de Bass en Australie (b, d, f) pour la configuration N° 4. Dans le premier cas, on obtient une corrélation maximale de 1,03, soit une erreur sur le maximum de 3%. Dans le deuxième cas en revanche, le maximum est à 1,69, soit près de 70% d'erreur. On s'intéresse alors aux échelles de corrélation utilisées (courbes bleues). En comparant les échelles zonales (a et b), on s'aperçoit que dans le premier cas, il n'y a pas beaucoup de variation autour de la zone de corrélation alors que dans le deuxième cas l'échelle passe d'environ 350 à 200 mètres. Concernant les échelles méridiennes (c et d), la variation est importante pour le Pacifique Equatorial mais le raffinement de la grille ( $0,5^\circ$  au lieu de  $2^\circ$ ) semble permettre de mieux la gérer. Enfin, les échelles verticales (e et f) montrent également une variation plus importante pour la corrélation près de l'Australie. Ainsi, l'erreur de normalisation est plus importante lorsque les échelles varient brusquement.

La configuration N° 5 est enfin testée sur ORCA2 avec des échelles estimées par une méthode d'ensemble (non montrée). Elle présente une nette amélioration de l'erreur en surface par rapport à la figure 5.15. Cependant, vers 200 mètres de profondeur alors que l'erreur pour la configuration N° 4 était tout à fait acceptable, la configuration N° 5 provoque des erreurs très élevées. L'adaptation de la méthode de Purser *et al.* (2003b) au cas 3D telle que proposée par l'équation (5.25) ne semble pas satisfaisante dans tous les cas et demande donc à être étudiée plus avant.

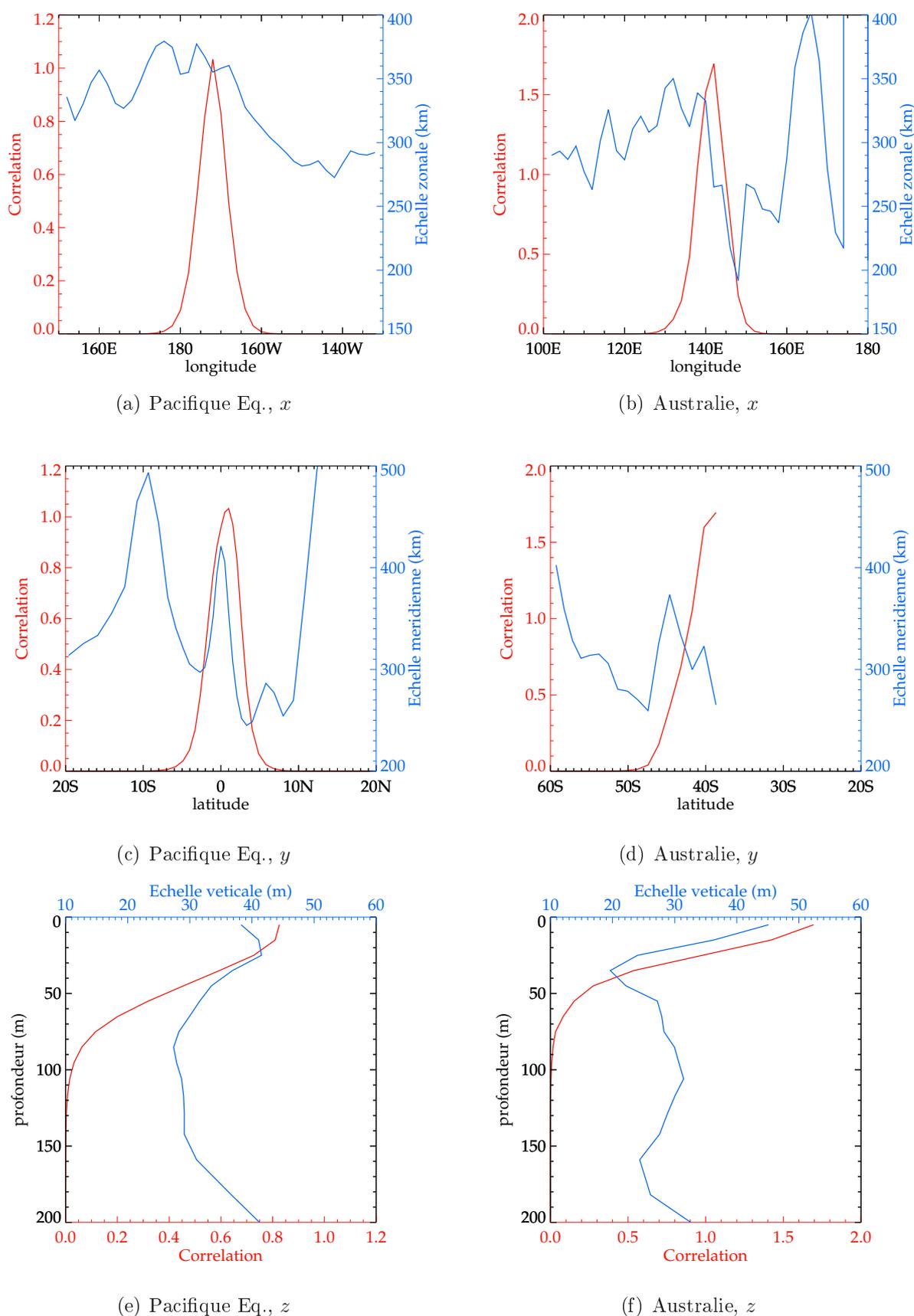


FIG. 5.17 – Exemple de corrélations de température (courbes rouges) dans le Pacifique Equatorial (a, c, e) et dans le détroit de Bass en Australie (b, d, f), et des échelles de corrélation associées (courbes bleues). Les coupes sont réalisées suivant l'axe zonal  $x$  (a, b), l'axe méridien  $y$  (c, d) et l'axe vertical  $z$  (e, f). Pour le panneau (d), l'arrêt des courbes à 38°S marque le début de la terre australienne.



---

## Chapitre 6

# Conclusions et perspectives

Dans cette thèse, nous avons poursuivi les travaux de Weaver et Courtier (2001) et de Weaver et Ricci (2004) sur la modélisation de fonctions de corrélation pour l'assimilation de données variationnelle, à partir de l'équation de diffusion. Le but était de définir des opérateurs de corrélation efficaces à partir d'opérateurs de diffusion *implicite*. En effet, bien que le schéma de diffusion *explicite* soit très simple à mettre en œuvre, il est soumis à un critère de stabilité qui peut le rendre très coûteux. A contrario, le schéma implicite est inconditionnellement stable, mais son efficacité dépend de celle des algorithmes utilisés pour inverser sa matrice. Les différents développements réalisés devaient ensuite être appliqués au système NEMOVAR pour modéliser les corrélations univariées de l'erreur d'ébauche, et se substituer à terme aux opérateurs de diffusion explicite.

Une étude théorique de l'équation de diffusion 1D et de sa discrétisation temporelle implicite a tout d'abord été réalisée. Ces travaux ont permis de mettre en évidence le lien direct entre la solution d'une équation de diffusion implicite et les fonctions AR (famille de Matérn), dans le cas homogène et isotrope. Nous avons en particulier pu préciser les relations entre, d'une part, le nombre d'itérations du filtre de diffusion et l'ordre de la fonction AR, et d'autre part, le coefficient de diffusion et l'échelle de corrélation. De plus, le facteur de normalisation qui, une fois multiplié au résultat de l'opérateur de diffusion, permet de générer l'opérateur de corrélation, a pu être déterminé théoriquement. Nous avons ensuite établi des approximations permettant d'étendre ces résultats à la modélisation de fonctions de corrélation dont les échelles varient en fonction de leur position. Nous nous sommes également intéressés à l'influence que pouvaient avoir les frontières sur le noyau de l'opérateur de corrélation. Ainsi, il a été possible de simuler des frontières transparentes, *i.e.* qui n'affectent pas le noyau de corrélation, en associant deux problèmes de diffusion, l'un avec des conditions aux frontières de Neumann, et l'autre avec des conditions aux frontières de Dirichlet.

A partir des résultats de cette étude, des opérateurs de diffusion 2 et 3×1D ont été construits, sous l'hypothèse de fonctions de corrélation séparables, en appliquant un opérateur de diffusion implicite 1D dans chaque direction. Parce que les axes sont traités séparément, un nombre faible d'itérations génère une anisotropie parasite, et on limite donc ces opérateurs à la représentation de fonctions de corrélation gaussienne (typiquement  $\geq 10$  itérations dans chaque direction pour obtenir une approximation satisfaisante). Plusieurs formulations de l'opérateur de diffusion 2×1D ont été testées. En particulier, en supposant que l'opérateur 2×1D est autoadjoint, il peut s'écrire sous la forme d'une factorisation en "racines carrées" et garantir ainsi un résultat symétrique et défini positif. Mais près des frontières ou lorsque les échelles de corrélation varient suivant leur position géographique et pas seulement suivant leur position sur l'axe traité, l'hypothèse de séparabilité n'est plus valide. Palier ce problème est néanmoins possible en moyennant les différentes combinaisons d'opérateurs 1D. Mais cela est bien sûr au prix d'applications d'opérateurs supplémentaires et doit donc être réservé aux expériences le nécessitant.

Ces opérateurs de diffusion implicite  $2$  et  $3 \times 1D$  ont été mis en œuvre dans le système NEMOVAR dans le cadre de configurations globales plutôt basse résolution. L’assimilation d’observations près des côtes n’y étant pas privilégiée, et compte tenu du paramétrage simplifié des échelles de corrélation et de la formulation explicite actuelle, nous avons retenu la simple formulation en “racines carrées” et non pas la moyenne des différentes combinaisons d’opérateurs  $1D$ . Dans cette thèse, nous avons choisi d’inverser la matrice de diffusion implicite par une décomposition de Cholesky. Pour cela, nous avons dû mettre en place des procédures de réorganisation des processeurs pour nous conformer à la gestion multi-processeurs du modèle NEMO. La comparaison de ces opérateurs de diffusion implicite à leurs équivalents explicites montre que si les premiers nécessitent plus de temps pour leur exécution puisqu’une matrice doit y être inversée, ils nécessitent moins d’itérations que les seconds qui doivent respecter le critère de stabilité. De plus, alors que des communications inter-processeurs sont nécessaires à chaque itération de la résolution de l’équation de diffusion explicite, elles sont à peu près limitées aux réorganisations des processeurs dans le cas implicite. Ainsi, pour les différentes dispositions des processeurs que nous avons utilisées, les opérateurs de corrélation générés par les opérateurs de diffusion implicite étaient environ 4 fois moins chers que ceux générés par les opérateurs de diffusion explicite, bien que nécessitant un peu plus de mémoire. En fonction des paramètres utilisés (algorithme de minimisation, type de normalisation, diagnostics), l’utilisation d’opérateurs de diffusion implicite dans la boucle interne de nos configurations NEMOVAR a pu réduire le temps d’exécution de celle-ci de 20 à 50%, pour des résultats d’assimilation statistiquement équivalents. Enfin, le fait que les opérateurs de diffusion implicite ne soient soumis à aucun critère de stabilité permet également de prendre en compte de grandes échelles de corrélation sans coût supplémentaire. En particulier, nous avons pu montrer qu’il était possible d’utiliser des échelles de corrélation issues d’une méthode d’ensemble sans surcoût pour l’implicite, tandis que les mêmes échelles utilisées dans le cadre d’un opérateur de diffusion explicite auraient rendu le coût de l’opérateur de corrélation prohibitif.

Après avoir mis en œuvre les opérateurs de diffusion implicite, nous nous sommes intéressés à la normalisation permettant de les transformer en opérateurs de corrélation. L’étude  $1D$  préalable a permis de mettre en évidence que la méthode de normalisation par vecteur aléatoire couramment utilisée, associée à un opérateur de diffusion avec des conditions aux frontières soit de Neumann, soit de Dirichlet, modifiait les échelles de corrélation près des frontières. De plus, le nombre de membres à utiliser dépend de la précision attendue, et la procédure peut donc s’avérer très coûteuse lorsqu’elle est effectuée en début de cycle d’assimilation. Cette méthode reste néanmoins bien adaptée aux configurations dont les paramètres n’évoluent pas dans le temps et où l’assimilation près des côtes n’est pas privilégiée. Au besoin, l’altération des échelles de corrélation près de la surface peut être corrigée en définissant l’opérateur de diffusion vertical comme la moyenne de deux problèmes de diffusion, l’un avec des conditions aux frontières de Neumann, et l’autre avec des conditions aux frontières de Dirichlet. La procédure de normalisation peut alors être effectuée au préalable avec un nombre de membres suffisant, et le résultat stocké pour être utilisé dans l’expérience d’assimilation. Néanmoins, lorsque l’on souhaite introduire une dépendance à l’écoulement en redéfinissant par exemple les échelles de corrélation à chaque cycle, une méthode de normalisation moins coûteuse doit être utilisée. Nous avons testé ici la possibilité de définir une approximation des facteurs de normalisation théoriques. Attractive par son coût négligeable, cette méthode nécessite néanmoins la simulation de frontières transparentes et augmente donc le coût de l’opérateur de diffusion. De plus, l’erreur numérique de discrétisation spatiale de l’équation de diffusion ainsi que les erreurs générées par des variations d’échelles de corrélation trop brusques doivent être corrigées expressément. Une étude complémentaire est donc nécessaire pour appréhender les possibilités que pourrait offrir cette méthode.

Pour la configuration pré-opérationnelle de NEMOVAR utilisée pour la prévision saisonnière du CEPMMT (ORCA1, normalisation préalable par la méthode par vecteur aléatoire de 10000 membres, diagnostics nécessitant des applications supplémentaires des opérateurs de diffusion), les opérateurs de diffusion implicite développés au cours de cette thèse permettent de réduire le temps d’exécution de la boucle interne d’environ 30% pour un besoin en mémoire légèrement plus élevé mais pouvant être

totalemment compensé par le nouvel algorithme de minimisation CGMOD qui offre la possibilité de ne pas stocker certaines informations. Ils permettent en outre de prendre en compte des échelles de corrélation paramétrées en fonction de connaissances physiques ou statistiques sans restriction numérique. Enfin, ils facilitent l'accès à plusieurs études dont on donne quelques exemples ci-dessous.

Nous nous sommes en effet limités ici, à une assimilation 3D-Var sur des configurations plutôt basse résolution. Cependant, l'algorithme du 4D-Var, bien plus coûteux que le 3D-Var, est actuellement en cours de développement pour NEMOVAR, et nécessite des opérateurs les moins chers possibles. D'autre part, NEMOVAR doit être appliqué à des configurations à plus haute résolution telle que le  $1/4^\circ$  (ou le  $1/12^\circ$ ), actuellement opérationnelle à Mercator-Océan et au MetOffice pour les prévisions océaniques à courte échéance, et pressentie au CEPMMT pour la prévision saisonnière. Dans tous les cas, l'utilisation des opérateurs de diffusion explicite et de leur nombre d'itérations important (malgré la limitation des échelles de corrélation) peut alourdir considérablement les calculs (suivant le poids relatif de leur application par rapport à l'intégration du modèle). Les opérateurs de diffusion implicite développés ici peuvent permettre de pallier ce problème, sous réserve d'un comportement acceptable près des frontières. Toutefois, il est probable qu'un plus grand nombre de processeurs soit utilisé pour ces expériences par rapport à celles déroulées dans cette thèse. Or, si le nombre de communications dépend essentiellement du nombre d'itérations pour les opérateurs de diffusion explicite, il dépend surtout du nombre de processeurs et de leur distribution pour les opérateurs de diffusion implicite, compte tenu de la méthode de résolution adoptée ici. Il conviendrait donc de s'assurer que la scalabilité de ces derniers n'atteint pas les limites qui les rendraient alors inefficaces.

Le critère de stabilité des opérateurs de diffusion explicite entraîne également une augmentation du nombre d'itérations lorsque les échelles de corrélation s'allongent, et ces dernières doivent donc éventuellement être limitées artificiellement pour ne pas rendre le coût de l'algorithme prohibitif. Or, ces échelles devraient être définies par rapport à des connaissances physiques ou statistiques et non pas pour des raisons numériques. De plus, une perspective attrayante consiste à introduire une dépendance à l'écoulement dans le schéma d'assimilation, en redéfinissant ces échelles de corrélation à chaque cycle. Une première approche peut être mise en œuvre en déterminant les échelles verticales en fonction du gradient de densité comme suggéré par Cummings (2005). Ainsi par exemple, elles pourraient être allongées dans la couche de mélange où ce gradient est faible, et rétrécies dans la thermocline où ce gradient est plus fort. Une seconde approche consiste à caractériser les échelles de corrélation à partir d'une méthode d'ensemble. Dans tous les cas, il est tout d'abord nécessaire de pouvoir prendre en compte facilement n'importe quelle échelle de corrélation, possibilité qu'offrent les opérateurs de diffusion implicite développés ici.

Ces opérateurs ne se limitent d'ailleurs pas à la représentation des corrélations de l'erreur d'ébauche, même s'ils ont été mis en œuvre dans ce cadre au cours de cette thèse. Ils peuvent également servir pour modéliser les corrélations spatiales ou temporelles de l'erreur modèle ou de l'erreur des forçages atmosphériques dans le cadre d'un 4D-Var à contrainte faible. Pour l'erreur d'observation, ils peuvent aussi être utilisés pour représenter les corrélations existant lorsque les observations sont issues d'un produit établi sur une grille. C'est par exemple le cas pour les produits standard de température de surface. On note cependant, que c'est l'inverse de la matrice de corrélation d'erreur d'observation qui est nécessaire à l'évaluation de la fonction coût et de son gradient. Or, si l'opérateur de diffusion implicite demande l'inversion d'une matrice, son opérateur inverse utilise donc cette matrice directement et devrait être encore plus simple à appliquer lorsqu'il n'est pas nécessaire de simuler des frontières transparentes. Plus complexes, les erreurs d'observations de hauteur de mer sont souvent données sous la forme d'une anomalie par rapport à une topographie dynamique moyenne (MDT<sup>77</sup>). Cette MDT est généralement donnée sur une grille (celle du modèle ou d'un autre) et est corrélée spatialement par construction. Une approche possible pour tenir compte de ces corrélations est de les modéliser grâce

---

<sup>77</sup>Mean Dynamic Topography.

aux opérateurs de diffusion implicite développés ici. La somme des covariances d’erreur de la MDT et des anomalies de hauteur d’eau est ensuite inversée pour le calcul de la fonction coût et de son gradient.

Dans un autre registre, l’estimation des covariances de l’erreur d’ébauche à partir d’une méthode d’ensemble nécessite généralement d’appliquer une procédure de localisation. En effet, la méthode d’ensemble étant très coûteuse, elle ne peut être utilisée qu’avec un faible nombre de membres, ce qui génère un certain bruit d’échantillonnage ainsi que des corrélations à longue distance parasites. On cherche alors à localiser les covariances, *i.e.* à s’assurer qu’elles s’annulent à partir d’une certaine distance. Une approche possible est d’effectuer cette localisation grâce au filtre récursif (Wang *et al.*, 2008) ou de la même manière grâce aux opérateurs de diffusion implicite.

Utilisables actuellement dans NEMOVAR, les opérateurs développés au cours de cette thèse permettent de réduire les coûts sur les configurations globales actuellement utilisées pour les applications sur le climat (réanalyses océaniques, initialisation pour la prévision saisonnière, ...) et simplifient l’accès à de nouvelles études. Néanmoins, ces opérateurs restent soumis à plusieurs limitations. En effet, près des frontières ou lorsque les échelles varient géographiquement, l’hypothèse de séparabilité sur laquelle ils sont construits n’est plus valable. L’opérateur multi-dimensionnel n’est alors pas autoadjoint et des artefacts liés à l’ordre d’application des opérateurs 1D apparaissent. De plus, bien qu’il soit possible de modéliser des fonctions de corrélation inhomogènes en faisant varier les échelles dans chaque direction, le traitement séparé des axes ne permet pas de modéliser des fonctions dont les variations d’échelles ne suivent pas ces axes, et donc la plupart des fonctions anisotropes. Purser *et al.* (2003b) proposent des techniques pour pallier ces problèmes dans le cadre du filtre récursif. Pour l’approche diffusion, une autre possibilité est de construire des opérateurs de diffusion implicite 2D et 3D complets avec un tenseur de diffusion non diagonal (Weaver et Courtier, 2001). Pour ce faire, une première étude théorique basée en particulier sur les travaux de Whittle (1954, 1963) et Weaver et Ricci (2004) permettrait de redéfinir de manière plus directe les liens entre l’équation de diffusion implicite et les fonctions de corrélations qu’elles permettent de représenter en 2D (plan ou sphère) ou 3D. Les efforts devraient ensuite porter sur la résolution du système linéaire à chaque pas de temps. En effet, l’efficacité des opérateurs de diffusion implicite repose sur celle des algorithmes assurant cette résolution. Une étude préliminaire dans le cadre 1D a permis de montrer qu’un algorithme itératif multi-grille pourrait être satisfaisant. Néanmoins, mettre en œuvre un tel algorithme requiert la définition d’opérateurs de simplification (restriction) et d’interpolation (prolongation) pour NEMOVAR, tout en respectant la parallélisation adoptée par NEMO.

Générer des opérateurs de corrélation à partir des opérateurs de diffusion nécessite de définir les facteurs de normalisation à appliquer. Ceux-ci doivent être suffisamment précis, et leur méthode de définition efficace, en particulier lorsque l’on envisage d’introduire une dépendance à l’écoulement contraignant à les recalculer à chaque cycle. La précision de la méthode par vecteur aléatoire, généralement utilisée, dépend du nombre de membres dont on dispose. Utiliser alors des techniques de filtrage spatial (Berre *et al.*, 2007; Raynaud *et al.*, 2008, 2009) pourrait réduire le nombre de membres nécessaires pour une même précision. La méthode par approximation du facteur théorique nécessite quant à elle des études complémentaires pour atténuer de manière satisfaisante et fiable les erreurs numériques dues à la discrétisation spatiale de l’équation de diffusion et aux variations des échelles de corrélation. Enfin, on peut se demander comment chacune de ces méthodes pourrait bénéficier à l’autre.

Les travaux réalisés au cours de cette thèse représentent une première étape dans la direction de l’utilisation de la diffusion implicite pour la modélisation de fonctions de corrélation. Disponibles dans NEMOVAR, les opérateurs de diffusion implicite proposés ici peuvent permettre de faire progresser plus facilement certaines études et développements tels que le 4D-Var et sa formulation contrainte faible, l’introduction d’une dépendance à l’écoulement des corrélations, ou encore l’utilisation de matrices de covariances non diagonales pour certaines observations. De plus, ils donnent des directions à explorer pour mettre en place des opérateurs de corrélation encore plus performants.

---

# Bibliographie

- Akhiezer NI, Glazman IM. 1961. *Theory of Linear Operators in Hilbert Space. Volume I*. Frederick Ungar Publishing Co., Inc, New York.
- Amodei L. 1995. Solution approchée pour un problème d'assimilation de données météorologiques avec prise en compte de l'erreur modèle. *C. R. Acad. Sci.* **321 II a**, 1087–1094.
- Arakawa A. 1972. *Design of the UCLA general circulation model*. Technical Report 7, Department of Meteorology, University of California, Los Angeles.
- Arfken G. 1966. *Mathematical Methods for Physicists*. Academic Press, New York.
- Aster RC, Borchers B, Thurber CH. 2005. *Parameter Estimation and Inverse Problems*. Elsevier Academic Press, London.
- Balmaseda MA, Dee D, Vidard A, Anderson DLT. 2007. A multivariate treatment of bias for sequential data assimilation : Application to the Tropical oceans. *Q. J. R. Meteorol. Soc.* **133**, 167–179.
- Balmaseda MA, Vidard A, Anderson DLT. 2008. The ECMWF ocean analysis system : ORA-S3. *Mon. Weather Rev.* **136**, 3018–3034.
- Bannister RN. 2007. Can wavelets improve the representation of forecast error covariances in variational data assimilation ? *Mon. Weather Rev.* **135**, 387–408.
- Bannister RN. 2008a. A review of forecast error covariance statistics in atmospheric variational data assimilation. I : Characteristics and measurements of forecast error covariances. *Q. J. R. Meteorol. Soc.* **134**, 1951–1970.
- Bannister RN. 2008b. A review of forecast error covariance statistics in atmospheric variational data assimilation. II : Modelling the forecast error covariance statistics. *Q. J. R. Meteorol. Soc.* **134**, 1971–1996.
- Belo Pereira M, Berre L. 2006. The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Weather Rev.* **134**, 2466–2489.
- Bennett AF. 2002. *Inverse Modeling of the Ocean and Atmosphere*. Cambridge University Press, Cambridge.
- Bennett AF, Chua BS, Harrison DE, McPhaden MJ. 1998. Generalized inversion of tropical atmosphere-ocean data and a coupled model of the Tropical Pacific. *J. Climate.* **11**, 1768–1792.
- Bennett AF, Chua BS, Leslie LM. 1996. Generalized inversion of a global numerical weather prediction model. *Meteorol. Atmos. Phys.* **60**, 165–178.
- Bennett AF, Chua BS, Leslie LM. 1997. Generalized inversion of a global numerical weather prediction model. II : Analysis and implementation. *Meteorol. Atmos. Phys.* **62**, 129–140.
- Bergthórsson P, Döös BR. 1955. Numerical weather map analysis. *Tellus.* **7**, 329–340.

- Berre L. 2000. Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Mon. Weather Rev.* **128**, 644–667.
- Berre L, Pannekoucke O, Desroziers G, Stefănescu SE, Chapnik B, Raynaud L. 2007. ‘A variational assimilation ensemble and the spatial filtering of its error covariances : Increase of sample size by local spatial averaging’. Dans Proceedings of the workshop on flow-dependent aspects of data assimilation, p. 151–168. ECMWF, Reading, UK.
- Berre L, Stefănescu SE, Belo Pereira M. 2006. The representation of the analysis effect in three error simulation techniques. *Tellus*. **58A**, 196–209.
- Bloom SC, Takacs LL, Da Silva AM, Ledvina D. 1996. Data assimilation using incremental analysis updates. *Mon. Weather Rev.* **124**, 1256–1271.
- Bouttier F, Courtier P. 1999. Data assimilation concepts and methods, lecture notes. URL [http://www.ecmwf.int/newsevents/training/lecture\\_notes/pdf\\_files/ASSIM/Ass\\_cons.pdf](http://www.ecmwf.int/newsevents/training/lecture_notes/pdf_files/ASSIM/Ass_cons.pdf).
- Briggs WL, Henson VE, McCormick SF. 2000. *A Multigrid Tutorial*. 2<sup>e</sup> éd. Siam, Philadelphia.
- Buehner M. 2005. Ensemble-derived stationary and flow-dependent background-error covariances : evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.* **131**, 1013–1043.
- Buehner M, Charron M. 2007. Spectral and spatial localization of background-error correlations for data assimilation. *Q. J. R. Meteorol. Soc.* **133**, 615–630.
- Buehner M, Houtekamer PL, Charette C, Mitchell HL, He B. 2009a. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I : Description and single-observation experiments. *Mon. Weather Rev.* DOI 10.1175/2009MWR3157.1.
- Buehner M, Houtekamer PL, Charette C, Mitchell HL, He B. 2009b. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II : One-month experiments with real observations. *Mon. Weather Rev.* DOI 10.1175/2009MWR3158.1.
- Buehner M, Malanotte-Rizzoli P. 2003. Reduced-rank Kalman filters applied to an idealized model of the wind-driven ocean circulation. *J. Geophys. Res.* **108 (C6)**, 23.
- Buell CE. 1972. Correlation functions for wind and geopotential on isobaric surfaces. *J. Appl. Meteorol.* **11**, 51–59.
- Busalacchi AJ. 1997. Oceanic observations. *J. Meteorol. Soc. Japan* **75**, 131–154.
- Chua BS, Bennett AF. 2001. An inverse ocean modeling system. *Ocean Model.* **3**, 137–165.
- Cohn S. 1997. An introduction to estimation theory. *J. Meteorol. Soc. Japan* **75**, 257–288.
- Cohn S, Da Silva A, Guo J, Sienkiewicz M, Lamich D. 1998. Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Weather Rev.* **126**, 2913–2926.
- Courtier P. 1997. Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.* **123**, 2449–2462.
- Courtier P, Anderson E, Heckley W, Pailleux J, Vasiljevic D, Hamrud M, Hollingsworth A, Rabier F, Fisher M. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). part I : Formulation. *Q. J. R. Meteorol. Soc.* **124**, 1783–1808.
- Courtier P, Thépaut JN, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1388.

- Cressman GP. 1959. An operational objective analysis system. *Mon. Weather Rev.* **87**, 367–374.
- Cummings JA. 2005. Operational multivariate ocean data assimilation. *Q. J. R. Meteorol. Soc.* **131**, 3583–3604.
- Daget N. 2008. *Estimation d'ensemble des paramètres des covariances d'erreur d'ébauche dans un système d'assimilation variationnelle de données océaniques*. Thèse de doctorat, Université Paul Sabatier - Toulouse III, France.
- Daget N, Weaver AT, Balmaseda MA. 2009. Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Q. J. R. Meteorol. Soc.* **135**, 1071–1094.
- Daley R. 1991. *Atmospheric Data Analysis*. Cambridge University Press, Cambridge.
- Daley R. 1992. Estimating model-error covariances for application to atmospheric data assimilation. *Mon. Weather Rev.* **120**, 1735–1746.
- Daley R, Barker E. 2001. *NRL atmospheric variational data assimilation system*. NAVDAS Source Book NRL/PU/7530-01-441, Naval Research Laboratory, Monterey, CA.
- Deckmyn A, Berre L. 2005. A wavelet approach to representing background error covariances in a limited-area model. *Mon. Weather Rev.* **133**, 1279–1294.
- Dee D. 2005. Bias and data assimilation. *Q. J. R. Meteorol. Soc.* **131**, 3323–3343.
- Derber J, Bouttier F. 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus.* **51A**, 195–221.
- Derber JC, Purser RJ, Wu WS, Treadon R, Pondevca M, Parrish D, Kleist D. 2004. 'Flow dependent jb in a global grid-point 3D-Var'. Dans ECMWF proceedings of the seminar on recent developments in data assimilation for atmosphere and ocean, p. 125–134. ECMWF, Reading, UK.
- Derber JC, Rosati A. 1989. A global oceanic data assimilation system. *J. Phys. Oceanogr.* **19**, 1333–1347.
- Desroziers G. 1997. A coordinate change for data assimilation in spherical geometry of frontal structures. *Mon. Weather Rev.* **125**, 3030–3038.
- Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **131**, 3385–3396.
- Desroziers G, Brachemi O, Hamadache B. 2001. Estimation of the representativeness error caused by the incremental formulation of variational data assimilation. *Q. J. R. Meteorol. Soc.* **127**, 1775–1794.
- Di Lorenzo E, Moore AM, Arango HG, Cornuelle BD, Miller AJ, Powell B, Chua BS, Bennett AF. 2007. Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS) : Development and application for a baroclinic coastal upwelling system. *Ocean Model.* **16**, 160–187.
- Dobricic S, Pinardi N. 2008. An oceanographic three-dimensional variational data assimilation scheme. *Ocean Model.* **22**, 89–105.
- Dowd K, Severance C. 1998. *High Performance Computing*. O'Reilly, Sebastopol.
- Egbert G, Bennett A, Foreman M. 1994. Topex/Poseidon tides estimated using a global inverse model. *J. Geophys. Res.* **99**, 24821–24852.

- El Akkraoui A, Gauthier P, Pellerin S, Buis S. 2008. Intercomparison of the primal and dual formulations of variational data assimilation. *Q. J. R. Meteorol. Soc.* **134**, 1015–1025.
- Elbern H, Schwinger J, Botchorishvili R. 2010. Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation : System configuration. *J. Geophys. Res.* **115**, 1–23.
- Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10143–10162.
- Evensen G. 2003. The Ensemble Kalman Filter : Theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.
- Evensen G. 2007. *Data Assimilation. The Ensemble Kalman Filter*. Springer-Verlag, Berlin.
- Fisher M. 1998. ‘Minimization algorithms for variational data assimilation’. Dans Proceedings of seminar on recent developments in numerical methods for atmospheric modelling, p. 364–385. ECMWF, Reading, UK.
- Fisher M. 2004. ‘Background error covariance modelling’. Dans ECMWF proceedings of the seminar on recent developments in data assimilation for atmosphere and ocean, p. 45–63. ECMWF, Reading, UK.
- Fisher M, Andersson E. 2001. *Developments in 4D-Var and Kalman filtering*. Tech. Memo. 347, ECMWF, Reading.
- Fukumori I, Malanotte-Rizzoli P. 1995. An approximate Kalman filter for ocean data assimilation : An example with an idealized Gulf Stream model. *J. Geophys. Res.* **100** ( ), 6777–6793.
- Gandin L. 1963. *Objective Analysis of Meteorological Fields*. English translation by israel program for scientific translation, jerusalem, 1965 éd. Gridromet, Leningrad.
- Gaspari G, Cohn S. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**, 723–757.
- Gaspari G, Cohn SE, Guo J, Pawson S. 2006. Construction and application of covariance functions with variable length-fields. *Q. J. R. Meteorol. Soc.* **132**, 1815–1838.
- Gauthier P, Tanguay M, Laroche S, Pellerin S, Morneau J. 2007. Extension of 3D-Var to 4D-Var : Implementation of 4D-Var at the Meteorological Service of Canada. *Mon. Weather Rev.* **135**, 2339–2354.
- Gelb A. 1974. *Applied Optimal Estimation*. The MIT Press, Cambridge.
- Ghil M, cohn S, Tavantzis J, Bube K, Isaacson E. 1981. Applications of estimation theory to numerical weather prediction. *Appl. Math. Sci.* **36**, 139–224.
- Gneiting T. 1999. Correlation functions for atmospheric data analysis. *Q. J. R. Meteorol. Soc.* **125**, 2449–2464.
- Gneiting T. 2002. Compactly supported correlation functions. *J. Multivariate Anal.* **83**, 493–508.
- Gneiting T, Kleiber W, Schlather M. 2009. *Matèrn cross-covariance functions for multivariate random fields*. Technical Report No. 549, University of Washington, Department of Statistics, Washington, USA.
- Golub GH, van Loan CF. 1996. *Matrix Computations*. 3<sup>e</sup> éd. Johns Hopkins University Press, Baltimore.

- Gradshteyn IS, Ryzhik IM. 1980. *Table of Integrals, Series, and Products*. Corrected and enlarged éd. Academic Press, San Diego.
- Gratton S, Tshimanga J. 2009. An observation-space formulation of variational assimilation using a restricted preconditioned conjugate gradient algorithm. *Q. J. R. Meteorol. Soc.* **135**, 1573–1585.
- Gregori P, Porcu E, Mateu J, Sasvári Z. 2008. On potentially negative space time covariances obtained as sum of products of marginal ones. *Ann. Inst. Stat. Math.* **60**, 865–882.
- Gustafsson N. 2007. Discussion on ‘4D-Var or EnKF ?’. *Tellus.* **59A**, 774–777.
- Guttorp P, Gneiting T. 2006. Miscellanea studies in the history of probability and statistics XLIX : On the Matérn correlation family. *Biometrika.* **93**, 989–995.
- Haberman R. 1987. *Elementary Applied Partial Differential Equations*. 2<sup>e</sup> éd. Prentice-Hall, Englewood Cliffs.
- Haidvogel DB, Beckmann A. 1999. *Numerical ocean circulation modeling*. Imperial College Press, London.
- Hamill TM, Snyder C. 2000. A Hybrid Ensemble Kalman Filter-3D variational analysis scheme. *Mon. Weather Rev.* **128**, 2905–2919.
- Hayden CM, Purser RJ. 1995. Recursive filter objective analysis of meteorological fields : Applications to NESDIS operational processing. *J. Appl. Meteorol.* **34**, 3–15.
- Hollingsworth A, Lönnberg P. 1986. The statistical structure of short range forecast errors as determined from radiosonde data. part I : The wind field. *Tellus.* **38A**, 111–136.
- Houtekamer PL, Mitchell HL. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **126**, 796–811.
- Houtekamer PL, Mitchell HL. 2001. A sequential Ensemble Kalman Filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**, 123–137.
- Hristopulos DT. 2003. Spartan Gibbs random field models for geostatistical applications. *SIAM J. Sci. Comput.* **24**, 2125–2162.
- Hristopulos DT, Elogne SN. 2007. Analytic properties and covariance functions of a new class of generalized Gibbs random fields. *IEEE Trans. Inform. Theory.* **53**, 4467–4679.
- Hristopulos DT, Elogne SN. 2009. Computationally efficient spatial interpolators based on Spartan spatial random fields. *IEEE Trans. Signal Process.* **57**, 3475–3487.
- Janjić T, Cohn S. 2006. Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Mon. Weather Rev.* **134**, 2900–2915.
- Järvinen H. 2001. Temporal evolution of innovation and residual statistics in the ECMWF variational data assimilation systems. *Tellus.* **53A**, 333–347.
- Jazwinski AH. 1970. *Stochastic Processes and filtering Theory*. Academic Press, Inc., New York.
- Jeffrey A. 1995. *Handbook of Mathematical Formulas and Integrals*. Academic Press, San Diego.
- Jones DS. 1982. *The Theory of Generalised Functions*. 2<sup>e</sup> éd. Cambridge University Press, Cambridge.
- Julian PR, Thiébaux HJ. 1975. On some properties of correlation functions used in optimum interpolation schemes. *Mon. Weather Rev.* **103**, 605–616.

- Kalman RE. 1960. A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Eng.* **82**, 35–45.
- Kalman RE, Bucy R. 1961. New results in linear filtering and prediction theory. *Trans. ASME, J. Basic Eng.* **83**, 95–108.
- Kalnay E. 2003. *Atmospheric Modelling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge.
- Kalnay E, Li H, Miyoshi T, Yang SC, Ballabrera-Poy J. 2007a. 4D-Var or ensemble Kalman filter? *Tellus.* **59A**, 758–773.
- Kalnay E, Li H, Miyoshi T, Yang SC, Ballabrera-Poy J. 2007b. Response to the discussion on "4D-Var or EnKF?" by Nils Gustafsson. *Tellus.* **59A**, 778–78.
- Kelbert M, Leonenko N, Ruiz-Medina MD. 2005. Fractional random fields associated with stochastic fractional heat equations. *Adv. in Appl. Prob.* **37**, 108–133.
- Kepernt JD. 2009. Covariance localisation and balance in an Ensemble Kalman Filter. *Q. J. R. Meteorol. Soc.* **135**, 1157–1176.
- Lawless AS, Gratton S, Nichols NK. 2005. An investigation of incremental 4D-Var using non-tangent linear models. *Q. J. R. Meteorol. Soc.* **131**, 459–476.
- Le Dimet FX, Talagrand O. 1986. Variational algorithms for analysis and assimilation of meteorological observations : Theoretical aspects. *Tellus.* **38A**, 97–110.
- Li Z, Chao Y, McWilliams JC, Ide K. 2008. A three-dimensional variational data assimilation scheme for the Regional Ocean Modeling System. *J. Atmos. Oceanic Technol.* **25**, 2074–2090.
- Liu H, Xue M, Purser RJ, Parrish DF. 2007. Retrieval of moisture from simulated GPS slant-path water vapor observations using 3DVAR with anisotropic recursive filter. *Mon. Weather Rev.* **135**, 1506–1521.
- Liu Y, Zhu J, She J, Zhuang S, Fu W, Gao J. 2009. Assimilating temperature and salinity profile observations using an anisotropic recursive filter in a coastal ocean model. *Ocean Model.* **30**, 75–87.
- Lönnerberg P, Hollingsworth A. 1986. The statistical structure of short range forecast errors as determined from radiosonde data. part II : The covariance of height and wind errors. *Tellus.* **38A**, 137–161.
- Lorenc A. 2003. The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **129**, 3183–3203.
- Lorenc AC. 1981. A global three-dimensional multivariate statistical interpolation scheme. *Mon. Weather Rev.* **109**, 701–721.
- Lorenc AC. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**, 1177–1194.
- Lorenc AC. 1992. Iterative analysis using covariance functions and filters. *Q. J. R. Meteorol. Soc.* **118**, 569–591.
- Ma C. 2005. Linear combinations of space-time covariance functions and variograms. *IEEE Trans. Signal Process.* **53**, 857–864.
- Madec G. 2008. *NEMO ocean engine*. Note du Pole de modélisation No 27 ISSN No 1288-1619, Institut Pierre-Simon Laplace, Paris.

- Mallat S. 2009. *A Wavelet Tour of Signal Processing. The Sparse Way*. Academic Press, Burlington.
- Martin MJ, Hines A, Bell MJ. 2007. Data assimilation in the FOAM operational short-range ocean forecasting system : A description of the scheme and its impact. *Q. J. R. Meteorol. Soc.* **133**, 981–995.
- Ménard R, Daley R. 1996. The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus*. **48A**, 221–237.
- Menke W. 1989. *Geophysical Data Analysis : Discrete Inverse Theory*. Academic Press, San Diego.
- Mirouze I, Weaver AT. 2010. Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Q. J. R. Meteorol. Soc.* **136**, 1421–1443.
- Mogensen KS, Balmaseda MA, Weaver AT, Martin M, Vidard A. 2009. ‘NEMOVAR : A variational data assimilation system for the NEMO ocean model’. Dans ECMWF Newsletter 120 - Summer 2009, p. 17–21. ECMWF, Reading, UK.
- Ne-Zheng S. 1994. *Inverse Problems in Groundwater Modeling*. Kluwer Academic Publishers, Dordrecht.
- Ngodock HE. 2005. Efficient implementation of covariance multiplication for data assimilation with the representer method. *Ocean Model.* **8**, 237–251.
- Oke PR. 2007. Impacts of localisation in the EnKF and EnOI : Experiments with a small model. *Ocean Dyn.* **57**, 32–45.
- Oliver DS. 1998. Calculation of the inverse of the covariance. *Math. Geology.* **30**, 911–933.
- Paciorek CJ, Schervish MJ. 2006. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics.* **17**, 483–506.
- Pannekoucke O. 2008. *Modélisation des structures locales de covariance des erreurs de prévision à l’aide des ondelettes*. Thèse de doctorat, Université Paul Sabatier - Toulouse III, France.
- Pannekoucke O. 2009. Heterogeneous correlation modeling based on the wavelet diagonal assumption and on the diffusion operator. *Mon. Weather Rev.* **137**, 2995–3012.
- Pannekoucke O, Berre L, Desroziers G. 2007. Filtering properties of wavelets for local background-error correlations. *Q. J. R. Meteorol. Soc.* **133**, 363–379.
- Pannekoucke O, Berre L, Desroziers G. 2008. Background-error correlation length-scale estimates and their sampling statistics. *Q. J. R. Meteorol. Soc.* **134**, 497–508.
- Pannekoucke O, Massart S. 2008. Estimation of the local diffusion tensor and normalization for heterogeneous correlation modelling using a diffusion equation. *Q. J. R. Meteorol. Soc.* **134**, 1425–1438.
- Parrish DF, Derber JC. 1992. The National Meteorological Center’s spectral statistical interpolation analysis system. *Mon. Weather Rev.* **120**, 1747–1763.
- Pham DT, Verron J, Roubaud C. 1998. A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Sys.* **16**, 323–340.
- Powell BS, Arango HG, Moore AM, Di Lorenzo E, Milliff RF, D F. 2008. 4DVAR data assimilation in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). *Ocean Model.* **23**, 130–145.

- Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 1992. *Numerical Recipes. The Art of Scientific Computing*. 3<sup>e</sup> éd. Cambridge University Press, Cambridge.
- Priestley MB. 1981. *Spectral Analysis and Time Series*. Elsevier Academic Press, London.
- Purser RJ. 2005. *A geometrical approach to the synthesis of smooth anisotropic covariance operators for data assimilation*. Office Note 447, National Centers for Environmental Prediction, Camp Springs.
- Purser RJ, Wu WS, Parrish DF, Roberts NM. 2003a. Numerical aspects of the application of recursive filters to variational statistical analysis. part I : Spatially homogeneous and isotropic Gaussian covariances. *Mon. Weather Rev.* **131**, 1524–1535.
- Purser RJ, Wu WS, Parrish DF, Roberts NM. 2003b. Numerical aspects of the application of recursive filters to variational statistical analysis. part II : Spatially inhomogeneous and anisotropic general covariances. *Mon. Weather Rev.* **131**, 1536–1548.
- Rabier F. 2005. Overview of global data assimilation developments in numerical weather-prediction centres. *Q. J. R. Meteorol. Soc.* **131**, 3215–3233.
- Rabier F, Järvinen H, Klinker E, Mahfouf JF, Simmons A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. part I : Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.* **126**, 1143–1170.
- Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, J PT. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society* **133**, 347–362.
- Raymond WH, Garder A. 1991. A review of recursive and implicit filters. *Mon. Weather Rev.* **119**, 477–495.
- Raynaud L, Berre L, Desroziers G. 2008. Spatial averaging of ensemble-based background-error variances. *Q. J. R. Meteorol. Soc.* **134**, 1003–1014.
- Raynaud L, Berre L, Desroziers G. 2009. Objective filtering of ensemble-based background-error variances. *Q. J. R. Meteorol. Soc.* **135**, 1177–1199.
- Ricci S. 2004. *Assimilation variationnelle océanique : Modélisation multivariée de la matrice de covariance d'erreur d'ébauche*. Thèse de doctorat, Université Paul Sabatier - Toulouse III, France.
- Ricci S, Weaver AT, Vialard J, Rogel P. 2005. Incorporating state-dependent temperature-salinity constraints in the background error covariance of variational ocean data assimilation. *Mon. Weather Rev.* **133**, 317–338.
- Robert C, Blayo E, Verron J. 2006a. Comparison of reduced-order sequential, variational and hybrid data assimilation in the context of a Tropical Pacific ocean model. *Ocean Dyn.* **56**, 624–633.
- Robert C, Blayo E, Verron J. 2006b. Reduced-order 4D-Var : a preconditioner for the full 4D-Var data assimilation method. *Geophys. Res. Lett.* **33**, L18609.
- Saporta G. 2006. *Probabilités, Analyse des Données et Statistique*. 2<sup>e</sup> éd. Technip, Paris.
- Sasaki Y. 1958. An objective analysis based on the variational method. *J. Meteorol. Soc. Japan.* **II-36**, 77–88.
- Sasaki Y. 1970. Some basic formalisms in numerical variational analysis. *Mon. Weather Rev.* **98**, 875–883.

- Shewchuk JR. 1994. An introduction to the conjugate gradient method without the agonizing pain.
- Snyder C, Bengtsson T, Bickel P, Anderson J. 2008. Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136**, 4629–4640.
- Stammer D, Wunsch C, Giering R, Eckert C, Heimbach P, Marotzke J, Adcroft A, Hill CN, Marshall J. 2002. Global ocean circulation during 1992-1997, estimated from ocean observations and a general circulation model. *J. Geophys. Res.* **107 (C9)**, 3118.
- Stein M. 1999. *Interpolation of Spatial Data. Some Theory for Kriging*. Springer, New-York.
- Stein M. 2005. *Nonstationary spatial covariance functions*. Technical report No. 21, University of Chicago, Center for Integrating Statistical and Environmental Science, Chicago, USA. URL <http://www.stat.uchicago.edu/cises/research>.
- von Storch H, Zwiers FW. 1999. *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge.
- Strang G. 1986. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley.
- Talagrand O. 1997. Assimilation of observations, an introduction. *J. Meteorol. Soc. Japan* **75**, 191–209.
- Talagrand O, Courtier P. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I : Theory. *Q. J. R. Meteorol. Soc.* **113**, 1311–1328.
- Tarantola A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia.
- Thacker WC, Long RB. 1988. Fitting dynamics to data. *J. Geophys. Res.* **93**, 1227–1240.
- Trémolet Y. 2006. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.* **132**, 2483–2504.
- Trémolet Y. 2007a. Incremental 4D-Var convergence study. *Tellus*. **59A**, 706–718.
- Trémolet Y. 2007b. Model-error estimation in 4D-Var. *Q. J. R. Meteorol. Soc.* **133**, 1267–1280.
- Tshimanga J, Gratton S, Weaver AT, Sartenaer A. 2008. Limited-memory preconditioners with application to incremental four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.* **134**, 751–769.
- Van Leeuwen PJ. 2003. A variance-minimizing filter for large-scale applications. *Mon. Weather Rev.* **131**, 2071–2084.
- Vialard J, Weaver AT, Anderson DLT, Delecluse P. 2003. Three- and Four-Dimensional variational assimilation with a general circulation model of the Tropical Pacific Ocean. Part II : Physical validation. *Mon. Weather Rev.* **131**, 1379–1395.
- Wang X, Barker DM, Snyder C, Hamill TM. 2008. A Hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I : Observing system simulation experiment. *Mon. Weather Rev.* **136**, 5116–5131.
- Weaver AT, Courtier P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.* **127**, 1815–1846.
- Weaver AT, Deltel C, Machu E, Ricci S, Daget N. 2005. A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc.* **131**, 3605–3625.

- Weaver AT, Ricci S. 2004. 'Constructing a background-error correlation model using generalized diffusion operators'. Dans ECMWF proceedings of the seminar on recent developments in data assimilation for atmosphere and ocean, p. 327–339. ECMWF, Reading, UK.
- Weaver AT, Vialard J, Anderson DLT. 2003. Three- and Four-Dimensional variational assimilation with a general circulation model of the Tropical Pacific Ocean. Part I : Formulation, internal diagnostics, and consistency checks. *Mon. Weather Rev.* **131**, 1360–1378.
- Whittle P. 1954. On stationary processes in the plane. *Biometrika.* **41**, 434–449.
- Whittle P. 1963. Stochastic processes in several dimensions. *Bull. Inst. Internat. Statist.* **40**, 974–994.
- Wikle CK, Berliner LM. 2007. A Bayesian tutorial for data assimilation. *Physica D.* **230**, 1–16.
- Wu WS, Purser RJ, Parrish DF. 2002. Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Weather Rev.* **130**, 2905–2916.
- Wunsch C. 1996. *The Ocean Circulation Inverse Problem*. Cambridge University Press, Cambridge.
- Xu Q. 2005. Representations of inverse covariances by differential operators. *Adv. Atmos. Sci.* **22**, 181–198.
- Yaglom AM. 1987. *Correlation Theory of Stationary and Related Random Functions. I : Basic Results*. Springer-Verlag, New York.

# Annexe A

## Analyse fonctionnelle

Par souci de simplification, tous les outils mathématiques sont volontairement restreints au corps des réels  $\mathbb{R}$ .

On note  $\mathbf{x}$  le vecteur  $(x_1, \dots, x_n)^T \in \mathbb{R}^n$ .

$L^1(\Omega)$  est l'espace des fonctions de valeur absolue sommable sur l'espace  $\Omega$

$$\int_{\Omega} |f(\mathbf{x})| d\mathbf{x} < +\infty.$$

$L^2(\Omega)$  est l'espace des fonctions de carré sommable sur l'espace  $\Omega$

$$\int_{\Omega} |f(\mathbf{x})|^2 d\mathbf{x} < +\infty.$$

### A.1 Rappel sur les espaces

Une suite  $(\mathbf{x}_n)_n$  est dite de Cauchy dans un espace vectoriel normé, si ses termes se rapprochent à partir d'un certain rang :  $\forall \epsilon > 0 \quad \exists N_{\epsilon} \in \mathbb{N} \quad / \quad \forall p, q > N_{\epsilon} \quad \|\mathbf{x}_p - \mathbf{x}_q\| < \epsilon$ .

Un **espace de Banach** est un espace vectoriel muni d'une norme  $\|\cdot\|$  dérivant d'une distance  $d(\cdot, \cdot)$  ( $\|\mathbf{x}\| = \sqrt{d(\mathbf{x}, \mathbf{x})}$ ), et complet pour cette distance, *i.e.* que toute suite de Cauchy de cet espace converge dans cet espace.

Un **espace préhilbertien** est un espace vectoriel muni d'une norme  $\|\cdot\|$  dérivant d'un produit scalaire  $\langle \cdot, \cdot \rangle$  ( $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ ).

Un **espace de Hilbert** est un espace préhilbertien complet pour la norme associée à son produit scalaire, *i.e.* que toute suite de Cauchy de cet espace converge dans cet espace.

Lorsqu'un espace préhilbertien est réel et de dimension finie, on parle alors d'**espace euclidien**. Comme tous les espaces normés de dimension finie sont complets, un espace euclidien est donc un cas particulier d'espace de Hilbert.

Un espace euclidien de dimension  $n$  est dit **canonique**, lorsqu'il est muni du produit scalaire canonique  $\langle \mathbf{x}, \mathbf{y} \rangle = \langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle = \sum_{i=1}^n x_i y_i$ .

On définit la norme euclidienne du vecteur  $\mathbf{x}$  par  $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$   
et la distance euclidienne entre deux vecteurs  $\mathbf{x}$  et  $\mathbf{y}$  par  $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ .

$\mathbb{R}^n$  est un espace euclidien (et donc de Hilbert) muni du produit scalaire canonique

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle = \sum_{i=1}^n x_i y_i.$$

## A.2 Formes linéaires et espace dual

Une **forme linéaire** sur un espace vectoriel  $E$  est une application linéaire  $f$  de  $E$  dans  $\mathbb{R}$  :

$$f : E \longrightarrow \mathbb{R} \quad / \quad \forall \mathbf{x}, \mathbf{y} \in E \text{ et } \lambda, \mu \in \mathbb{R} \quad \exists f(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda f(\mathbf{x}) + \mu f(\mathbf{y}) .$$

Si  $E$  est un espace normé de dimension finie, alors toutes les formes linéaires sur  $E$  sont continues.

En dimension finie, il est possible de mettre une forme linéaire sous une forme matricielle. En effet, si on se donne  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$  pour base de  $E$ , tout élément  $\mathbf{x} \in E$  s'écrit  $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$ . Par linéarité, on a  $f(\mathbf{x}) = \sum_{i=1}^n x_i f(\mathbf{e}_i)$ , doù

$$f(\mathbf{x}) \equiv \underbrace{\left( f(\mathbf{e}_1) \quad \dots \quad f(\mathbf{e}_n) \right)}_{\mathbf{F}} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

où  $\mathbf{F}$  est la forme matricielle (matrice  $1 \times n$ ) de  $f$ .

### Exemple:

Soit la forme linéaire  $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$  telle que  $(x_1, x_2) \longmapsto 2x_1 + 3x_2$ . Soient  $(\mathbf{e}_1, \mathbf{e}_2)$  la base canonique de  $\mathbb{R}^2$ . On a  $f(\mathbf{e}_1) = 2$  et  $f(\mathbf{e}_2) = 3$  et on peut écrire

$$f(\mathbf{x}) \equiv ( 2 \quad 3 ) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

On appelle **espace dual** de  $E$ , et on note  $E'$ , l'ensemble des formes linéaires sur  $E$  :

$$E' = \{ f : E \longrightarrow \mathbb{R} \quad / \quad \forall \mathbf{x}, \mathbf{y} \in E \text{ et } \lambda, \mu \in \mathbb{R} \quad \exists f(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda f(\mathbf{x}) + \mu f(\mathbf{y}) \} .$$

$E'$  est un espace vectoriel dont les dimensions physiques correspondent aux réciproques de celles de l'espace  $E$ . Si  $E$  est de dimension finie  $n$ ,  $E'$  est aussi de dimension finie  $n$ .

Si  $E$  est un espace préhilbertien (muni d'un produit scalaire), on peut définir une application linéaire  $\varphi$  telle que

$$\begin{aligned} \varphi : E &\longrightarrow E' \\ \mathbf{x} &\longmapsto f_{\mathbf{x}} : E \longrightarrow \mathbb{R} \\ &\quad \mathbf{y} \longmapsto \langle \mathbf{x}, \mathbf{y} \rangle \end{aligned} .$$

Si  $E$  est un espace de Hilbert, une conséquence du théorème de représentation de Riesz-Fréchet (voir Akhiezer et Glazman, 1961, pp. 33 et 42) est qu'il existe un isomorphisme entre cet espace et son dual, permettant alors d'identifier les deux espaces, *i.e.*  $E \approx E'$ . En d'autres termes, il est possible de représenter les formes linéaires continues sur  $E$  par des éléments de  $E$ .

L'**espace dual** d'un espace vectoriel  $E$  est l'ensemble des applications linéaires continues de  $E$  dans  $\mathbb{R}$ . Il est noté  $E'$ .

Les éléments du dual de  $\mathbb{R}^n$  peuvent être représentés par des éléments de  $\mathbb{R}^n$ .

## A.3 Opérateurs linéaires et adjoints

Un **opérateur linéaire**  $\mathcal{K} : E \longrightarrow F$  est une application qui transforme toute combinaison linéaire de  $E$  en une combinaison linéaire de  $F$  :

$$\forall \mathbf{x}, \mathbf{y} \in E \text{ et } \lambda, \mu \in \mathbb{R} \quad \mathcal{K}[\lambda \mathbf{x} + \mu \mathbf{y}] = \lambda \mathcal{K}[\mathbf{x}] + \mu \mathcal{K}[\mathbf{y}] \in F.$$

Si  $E$  et  $F$  sont deux espaces de Hilbert<sup>78</sup> identifiés à leurs espaces duals respectifs, il est alors possible de définir l'**opérateur adjoint**  $\mathcal{K}^* : F \longrightarrow E$  tel que :

$$\forall \mathbf{x} \in E \text{ et } \mathbf{y} \in F \quad \langle \mathbf{x}, \mathcal{K}^*[\mathbf{y}] \rangle_E = \langle \mathcal{K}[\mathbf{x}], \mathbf{y} \rangle_F. \quad (\text{A.1})$$

Si  $\mathcal{K}$  et  $\mathcal{G}$  sont deux opérateurs linéaires continus, alors  $(\mathcal{K}\mathcal{G})^* = \mathcal{G}^*\mathcal{K}^*$ . Si  $E = F$  et si  $\mathcal{K} = \mathcal{K}^*$ , l'opérateur est dit **auto-adjoint**. S'il commute avec son adjoint,  $\mathcal{K}^*\mathcal{K} = \mathcal{K}\mathcal{K}^*$ , il est dit **normal**. Le produit de deux opérateurs auto-adjoints est auto-adjoint si et seulement si les opérateurs commutent.

D'après le théorème de représentation de Riesz-Fréchet (voir Akhiezer et Glazman, 1961, pp. 33 et 42), si  $\mathcal{K}$  est un opérateur linéaire continu défini d'un espace de Hilbert dans lui-même, alors l'opérateur adjoint  $\mathcal{K}^*$  existe et est unique.

En dimension finie  $n$ , le produit scalaire étant une forme bilinéaire, il peut être exprimé sous une forme matricielle  $\mathbf{P}$  (matrice  $n \times n$ ) telle que  $\langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^T \mathbf{P} \mathbf{x}$ . Si de plus  $\mathbf{K}$  est la forme matricielle de l'opérateur  $\mathcal{K}$  et  $\mathbf{K}^*$  celle de  $\mathcal{K}^*$ , on a

$$\left. \begin{aligned} \langle \mathcal{K}[\mathbf{x}], \mathbf{y} \rangle_F &= \mathbf{x}^T \mathbf{K}^T \mathbf{P}_F \mathbf{y} \\ \langle \mathbf{x}, \mathcal{K}^*[\mathbf{y}] \rangle_E &= \mathbf{x}^T \mathbf{P}_E \mathbf{K}^* \mathbf{y} \end{aligned} \right\} \implies \mathbf{K}^* = \mathbf{P}_E^{-1} \mathbf{K}^T \mathbf{P}_F. \quad (\text{A.2})$$

Si le produit scalaire utilisé est canonique,  $\mathbf{P} = \mathbf{I}$  la matrice identité, l'opérateur adjoint correspond à l'opérateur transposé  $\mathcal{K}^* = \mathcal{K}^T$ .

### Exemple:

$$\left| \begin{array}{l} \text{Soit l'opérateur linéaire } \mathcal{K} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2 \text{ tel que } (x_1, x_2) \longmapsto (2x_1 + 3x_2, x_1). \\ \text{On a } \mathbf{K} = \begin{pmatrix} 2 & 3 \\ 1 & 0 \end{pmatrix} \text{ et donc } \mathbf{K}^* = \begin{pmatrix} 2 & 1 \\ 3 & 0 \end{pmatrix}. \\ \text{L'opérateur adjoint } \mathcal{K}^* : \mathbb{R}^2 \longrightarrow \mathbb{R}^2 \text{ est donc tel que } (x_1, x_2) \longmapsto (2x_1 + x_2, 3x_1). \end{array} \right.$$

Pour tout opérateur linéaire continu sur  $\mathbb{R}^n$  il existe un unique opérateur adjoint. Il correspond à l'opérateur transposé si le produit scalaire utilisé est canonique.

<sup>78</sup>On peut également définir un opérateur adjoint pour  $E$  et  $F$  deux espaces de Banach. Cette définition prend alors une autre expression puisque d'une part, le produit scalaire n'y est pas défini et d'autre part, il n'existe pas d'isomorphisme entre un espace de Banach et son dual permettant de les identifier.



## Annexe B

# Probabilités et statistiques

Cette annexe constitue un rappel des principaux résultats de probabilités. Seules les variables aléatoires réelles sont traitées ici. Pour plus de détail, en particulier concernant les hypothèses inhérentes à ces définitions, on se reportera par exemple à Priestley (1981), Tarantola (2005) ou Saporta (2006).

## B.1 Variables aléatoires

### B.1.1 Définitions

Une **variable aléatoire** représente une grandeur qui varie selon le résultat d'une expérience aléatoire. Par exemple, si l'on considère le lancé de deux dés parfaitement équilibrés, elle représente les faces sorties pour chacun des dés et prend donc ses valeurs dans l'ensemble  $\Omega = \{(1, 1); (1, 2); \dots; (6, 6)\}$ . La loi de probabilité associée est donnée par

$$P(\omega) = \frac{1}{36}, \quad \forall \omega \in \Omega,$$

et représente la probabilité qu'un élément de  $\Omega$  particulier sorte lors d'un jet. On définit maintenant une application  $X$  de  $\Omega$  dans  $E = X(\Omega)$ .

La **loi de probabilité** de  $X$  est définie par

$$P_X(\alpha) = \frac{\text{Nombre de cas permettant de réaliser } X = \alpha}{\text{Nombre de cas possibles}}, \quad \forall \alpha \in E.$$

Si on définit par exemple comme application la somme des deux dés, on a  $E = \{2, 3, \dots, 12\}$ .  $X = 4$  est réalisée par les couples  $(1, 3), (2, 2), (3, 1)$ , soit  $P_X(X = 4) = 3/36$ . Lorsque la variable aléatoire est **discrète**, la loi de probabilité  $P_X$  est constituée de masses ponctuelles.

Lorsqu'une variable aléatoire est **continue**, elle admet une **densité de probabilité**  $f$

$$P_X(I) = \int_I f(x) dx, \quad \text{et on a} \quad \int_{\mathbb{R}} f(x) dx = 1.$$

La **fonction de répartition** est l'application monotone croissante

$$\begin{aligned} F : \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto F(x) = P(X < x). \end{aligned}$$

Dans le cas continu, on a

$$F(x) = \int_{-\infty}^x f(x') dx'.$$

Dans notre exemple,  $X < 4$  est réalisé par les couples  $(1, 1)$ ,  $(1, 2)$ ,  $(2, 1)$  et on a donc  $P(X < 4) = 3/36$ . De même  $X < 10$  est réalisé par tous les couples sauf  $(4, 6)$ ,  $(6, 4)$ ,  $(5, 5)$ ,  $(5, 6)$ ,  $(6, 5)$ ,  $(6, 6)$  et on a donc  $P(X < 10) = 30/36$ . La fonction de répartition permet également de calculer la probabilité d'un intervalle sur  $\mathbb{R}$  :  $P(a \leq X < b) = F(b) - F(a)$ . Par exemple  $P(4 \leq X < 10) = 30/36 - 3/36 = 27/36$ .

La **fonction caractéristique** d'une variable aléatoire  $X$  est la transformée de Fourier de sa loi de probabilité. Dans le cas continu, on a

$$\varphi_X(\hat{x}) = \int_{\mathbb{R}} e^{i\hat{x}x} f(x) dx.$$

Cette fonction existe toujours et est continue.

## B.1.2 Moments

### Moment centré d'ordre 1

L'**espérance** est définie par

$$E[X] = \sum_j x_j P(X = x_j), \quad \text{pour une variable discrète,}$$

$$E[X] = \int_{\mathbb{R}} x f(x) dx, \quad \text{pour une variable continue.}$$

Lorsqu'elle existe, l'espérance représente la moyenne arithmétique des différentes valeurs de  $X$  pondérées par leurs probabilités. Ses propriétés principales sont la linéarité

$$E[aX] = aE[X],$$

où  $a$  est un réel, et l'additivité

$$E[X_1 + X_2] = E[X_1] + E[X_2].$$

Si  $X_1$  et  $X_2$  sont deux **variables indépendantes**, alors l'espérance de leur produit est le produit de leurs espérances

$$E[X_1 X_2] = E[X_1] E[X_2].$$

### Moment centré d'ordre 2

La **variance** est définie par

$$V[X] = \sigma^2 = E[(X - E[X])^2] = E[X^2] - (E[X])^2,$$

où  $\sigma$  est l'**écart-type**.

La variance représente la dispersion de  $X$  autour de sa moyenne. Sa principale propriété est donnée par

$$V[aX] = a^2 V[X].$$

L'espérance et l'écart-type sont reliés par l'**inégalité de Bienaymé-Tchebyshev** :

$$P(|X - E[X]| > p\sigma) \leq \frac{1}{p^2}.$$

On appelle **covariance** des variables aléatoires  $X_1$  et  $X_2$ , la quantité

$$\text{cov}(X_1, X_2) = E[X_1 X_2] - E[X_1] E[X_2].$$

Si  $X_1$  et  $X_2$  sont **indépendantes**, alors  $\text{cov}(X_1, X_2) = 0$ .

La variance de la somme des variables aléatoires  $X_1$  et  $X_2$  s'écrit

$$V[X_1 + X_2] = V[X_1] + V[X_2] + 2\text{cov}(X_1, X_2).$$

On appelle **corrélacion** des variables aléatoires  $X_1$  et  $X_2$  d'écart-type respectifs  $\sigma_1$  et  $\sigma_2$ , la quantité normalisée

$$\text{cor}(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sigma_1 \sigma_2}$$

La corrélation est adimensionnelle et prend ses valeurs dans l'intervalle  $[-1, 1]$ .

La corrélation entre deux variables définit l'intensité du **lien linéaire** qui peut exister entre ces variables. Une corrélation de 1 signifie qu'une des variables est une fonction affine croissante de l'autre. Si la corrélation est à  $-1$ , la fonction affine est décroissante. Une corrélation à 0 signifie que les variables sont indépendantes.

### Moment centré d'ordre k

S'il existe, le **moment centré d'ordre k** est défini par

$$\mu_k = E[(X - E[X])^k].$$

Si la distribution de la variable aléatoire est symétrique, on a  $\mu_{2k+1} = 0 \forall k$ .

Pour caractériser la forme de la distribution, on définit  $\gamma_1$  le **coefficient d'asymétrie** ("skewness") et  $\gamma_2$  le **coefficient d'aplatissement** ("kurtosis") qui mesure en particulier les "queues" de distribution :

$$\gamma_1 = \frac{\mu_3}{\sigma^3}, \quad \gamma_2 = \frac{\mu_4}{\sigma^4}.$$

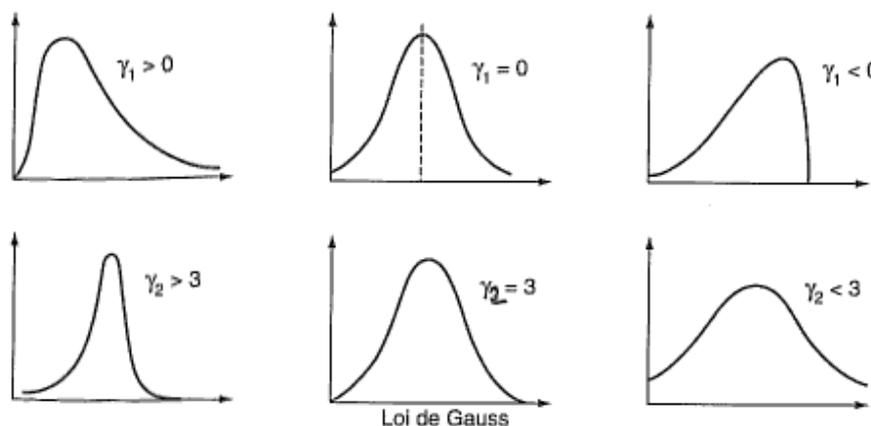


FIG. B.1 – Distributions avec différents coefficients d'asymétrie en haut et d'aplatissement en bas (Figure 2.9 Saporta, 2006).

### B.1.3 Loi de Gauss

Une variable aléatoire  $X$  suit une loi **gaussienne** (ou **normale**) d'espérance  $m$  et d'écart-type  $\sigma$ , si sa densité est donnée par

$$X \sim \mathcal{N}(m, \sigma) \Rightarrow f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right).$$

Les variables gaussiennes possèdent la propriété d'additivité

$$\left. \begin{array}{l} X_1 \sim \mathcal{N}(m_1, \sigma_1) \\ X_2 \sim \mathcal{N}(m_2, \sigma_2) \end{array} \right\} \Rightarrow X_1 + X_2 \sim \mathcal{N}\left(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right).$$

Les moments de la loi gaussienne existent pour tout ordre, mais en raison de la symétrie de la distribution, les moments d'ordre impairs sont nuls. Sa fonction caractéristique est donnée par

$$\varphi_X(\hat{x}) = \exp\left(-\frac{1}{2}\sigma^2\hat{x} + im\hat{x}\right).$$

On note enfin le changement de variable aléatoire

$$\text{si } U = \frac{X - m}{\sigma}, \quad \text{alors } U \sim \mathcal{N}(0, 1),$$

où  $\mathcal{N}(0, 1)$  est dite **loi normale centrée réduite**.

### B.1.4 Probabilités conditionnelles

Soient  $X$  et  $Y$  deux variables aléatoires continues. La **fonction de répartition du couple**  $(X, Y)$  est définie par

$$H(x, y) = P(X < x \cap Y < y).$$

La densité de probabilité du couple  $(X, Y)$  ou **loi jointe** est alors donnée par

$$h(x, y) = \frac{\partial^2 H}{\partial x \partial y}.$$

Les densités marginales ou **lois marginales** s'écrivent

$$\begin{aligned} f(x) &= \int_{\mathbb{R}} h(x, y) dy, \\ g(y) &= \int_{\mathbb{R}} h(x, y) dx. \end{aligned}$$

Si  $X$  et  $Y$  sont **indépendantes**, on a

$$h(x, y) = f(x)g(y).$$

On peut également définir quelle est la probabilité qu'une des variables ait une certaine valeur lorsque l'on connaît la valeur de l'autre.

On définit la **densité de probabilité conditionnelle de X sachant Y** par

$$P(X = x | Y = y) = f(x | y) = \frac{h(x, y)}{g(y)},$$

et la **densité de probabilité conditionnelle de Y sachant X** par

$$P(Y = y | X = x) = g(y | x) = \frac{h(x, y)}{f(x)}.$$

Les **formules de Bayes** permettent d'exprimer une probabilité conditionnelle en fonction de l'autre :

$$f(x | y) = \frac{g(y | x)f(x)}{\int_{\mathbb{R}} g(y | x)f(x) dx}, \quad g(y | x) = \frac{f(x | y)g(y)}{\int_{\mathbb{R}} f(x | y)g(y) dy}.$$

L'intégrale du dénominateur permet de normaliser la probabilité conditionnelle ( $\int_{\mathbb{R}} f(x | y) dx = 1$ ) et représente une simple valeur numérique. On écrit souvent les probabilités conditionnelles sous les relations proportionnelles

$$f(x | y) \propto g(y | x)f(x), \quad g(y | x) \propto f(x | y)g(y).$$

## B.2 Champs aléatoires et opérateurs statistiques

Un **champ aléatoire** (ou **vecteur aléatoire**) représente un ensemble de variables aléatoires. On parle également de **fonctions aléatoires** ou de **processus stochastique**. De la même manière que pour un couple de variables aléatoires, on définit la **fonction de répartition du champ** par

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n),$$

et sa **densité de probabilité** par

$$f(x_1, \dots, x_n) = \frac{\partial^n F}{\partial x_1 \cdots \partial x_n}.$$

Si les variables aléatoires sont **mutuellement indépendantes**, alors la loi jointe est le produit des lois marginales

$$f(x_1, \dots, x_n) = \prod_{j=1}^n f(x_j).$$

Pour notre application, on considère un champ aléatoire  $X$  constitué de variables aléatoires en chaque point de grille  $X(z)$ .

La **fonction de covariance** du champ aléatoire  $X$  est définie par

$$b(z, z') = E [(X(z) - E[X(z)]) (X(z') - E[X(z')])],$$

et sa **fonction de corrélation** par

$$c(z, z') = \frac{b(z, z')}{\sqrt{b(z, z)} \sqrt{b(z', z')}}.$$

Suivant les définitions données par Gaspari et Cohn (1999) (voir figure B.2) :

- une fonction de corrélation définie sur  $\mathbb{R}^n \times \mathbb{R}^n$  est **homogène** si elle est invariante par toute translation  $\mathbf{T}$  de  $\mathbb{R}^n$  :  $c(\mathbf{T}(z), \mathbf{T}(z')) = c(z, z')$ .
- une fonction de corrélation définie sur  $\mathbb{R}^n \times \mathbb{R}^n$  ( $S^{n-1} \times S^{n-1}$ ) est **isotrope** si elle est invariante par toute transformation orthogonale  $\mathbf{G}$  de  $\mathbb{R}^n$  ( $S^{n-1}$ ) :  $c(\mathbf{G}(z), \mathbf{G}(z')) = c(z, z')$ .

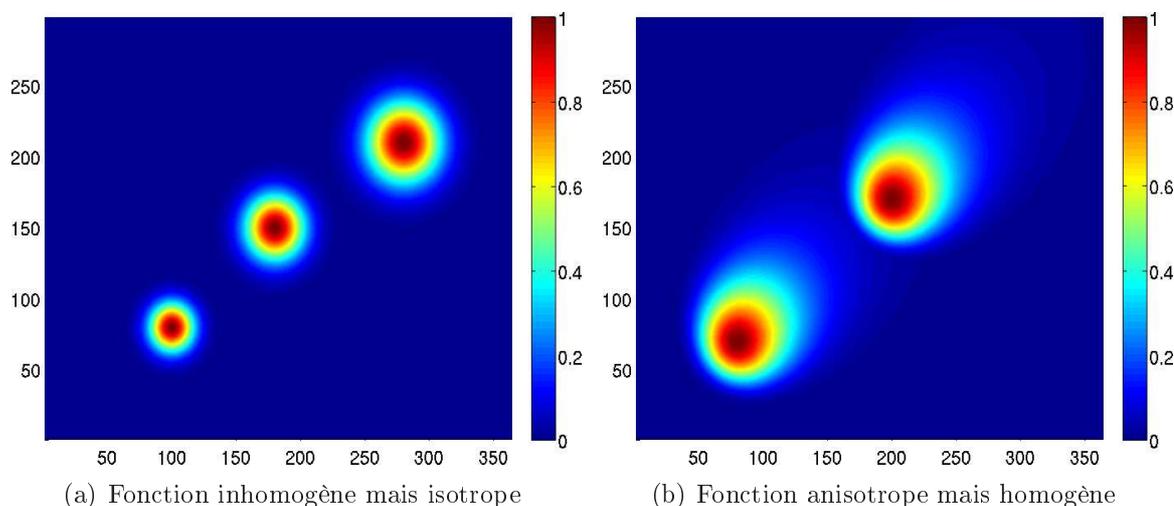


FIG. B.2 – Exemples d’une fonction de corrélation inhomogène (a) et d’une fonction de corrélation anisotrope (b) en 2D.

La symétrie d'une fonction de corrélation inhomogène ou anisotrope ne concerne pas la forme de la fonction mais le fait que centrée au point  $z$ , sa valeur par rapport à  $z'$  est la même que celle obtenue par rapport à  $z$  lorsque la fonction est centrée en  $z'$  (voir figure B.3).

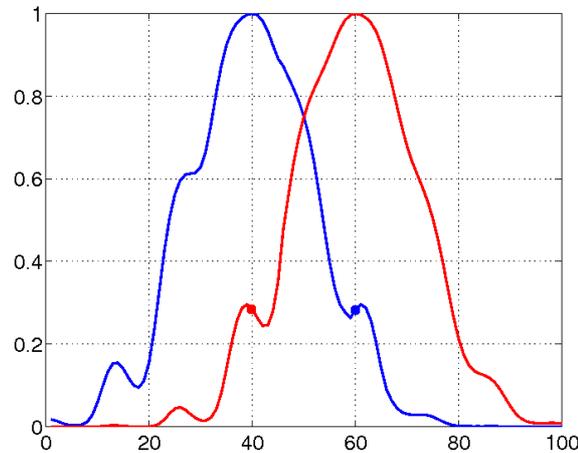


FIG. B.3 – Exemple d'une fonction de corrélation gaussienne 1D inhomogène. L'échelle varie en sinus suivant le même principe que celui décrit au paragraphe 4.4.1. La courbe bleue représente la fonction de corrélation lorsqu'elle est centrée au point  $z = 40$  tandis que la courbe rouge la représente centrée au point  $z' = 60$ . La valeur de corrélation de la courbe bleue au point  $z' = 60$  est la même que celle de la courbe rouge au point  $z = 40$ . On a donc bien  $f(z, z') = f(z', z)$  et la fonction de corrélation est donc symétrique malgré sa forme asymétrique.

Si  $E$  est espace linéaire et  $E'$  son dual, un **opérateur de covariance** est une application linéaire, symétrique (semi-)définie positive telle que

$$\begin{aligned} \mathcal{B} : E' &\rightarrow E \\ d'(z) &\mapsto d(z) = \int_E b(z, z') d'(z') dz', \end{aligned}$$

où  $b(z, z')$  est une fonction de covariance appelée **noyau** de l'opérateur.

On définit maintenant les variables normalisées  $e'(z) = d'(z)\sqrt{b(z, z)}$  et  $e(z) = d(z)/\sqrt{b(z, z)}$ .

Si  $E$  est espace linéaire et  $E'$  son dual, un **opérateur de corrélation** est une application linéaire, symétrique (semi-)définie positive telle que

$$\begin{aligned} \mathcal{C} : E' &\rightarrow E \\ e'(z) &\mapsto e(z) = \int_E c(z, z') d'(z') dz', \end{aligned}$$

où  $c(z, z')$  est une fonction de corrélation (amplitude maximale à 1) appelée **noyau** de l'opérateur.

Lorsque le noyau de l'opérateur est une fonction homogène et isotrope, les intégrales de l'opérateur sont des intégrales de convolution.

## Annexe C

# Estimateur de la moyenne conditionnelle

### Exemple simple

Nous reprenons tout d'abord l'exemple simple de la section 2.1.3 où on cherche à estimer la quantité  $u$  à partir de l'ébauche  $x$  et d'une mesure  $y$

$$\begin{aligned}x &= u + \varepsilon_x \\ y &= u + \varepsilon_y,\end{aligned}$$

où  $\varepsilon_x$  et  $\varepsilon_y$  sont des erreurs non biaisées

$$E[\varepsilon_x] = E[\varepsilon_y] = 0,$$

de variance d'erreur  $\sigma_x^2$  et  $\sigma_y^2$  respectivement, et décorrélées

$$E[\varepsilon_x \varepsilon_y] = 0.$$

On cherche ici à combiner linéairement l'ébauche et l'observation pour trouver  $\tilde{u}$ , un estimateur de  $u$

$$\tilde{u} = \alpha_1 x + \alpha_2 y, \tag{C.1}$$

où  $\alpha_1$  et  $\alpha_2$  sont deux coefficients à déterminer. Pour se faire, nous prenons tout d'abord comme hypothèse que l'estimateur  $\tilde{u}$  est non biaisé par rapport à  $u$

$$\begin{aligned}E[\tilde{u} - u] &= E[\alpha_1 (u + \varepsilon_x) + \alpha_2 (u + \varepsilon_y) - u] \\ &= (\alpha_1 + \alpha_2 - 1)E[u] = 0,\end{aligned}$$

ce qui impose la contrainte

$$\alpha_1 + \alpha_2 = 1.$$

On peut donc reformuler la combinaison linéaire donnée par l'équation (C.1) en

$$\tilde{u} = \alpha_1 x + (1 - \alpha_1) y.$$

Parmi tous les estimateurs qui respectent cette contrainte, on choisit celui qui minimise la variance de l'erreur d'estimation

$$\begin{aligned}\tilde{\sigma}^2 = E[(\tilde{u} - u)^2] &= E[(\alpha_1 (u + \varepsilon_x) + (1 - \alpha_1) (u + \varepsilon_y) - u)^2] \\ &= E[(\alpha_1 \varepsilon_x + (1 - \alpha_1) \varepsilon_y)^2] \\ &= \alpha_1^2 \sigma_x^2 + (1 - \alpha_1)^2 \sigma_y^2.\end{aligned}$$

Il faut donc trouver la valeur minimale de  $\alpha_1$ , et donc chercher les solutions qui annulent la dérivée par rapport à  $\alpha_1$ , soit

$$\begin{aligned}\frac{d\tilde{\sigma}^2}{d\alpha_1} &= 2\alpha_1 \sigma_x^2 - 2(1 - \alpha_1) \sigma_y^2 \\ &= 2\alpha_1 (\sigma_x^2 + \sigma_y^2) - 2\sigma_y^2 = 0,\end{aligned}$$

ce qui donne les solutions

$$\alpha_1 = \frac{\sigma_y^2}{\sigma_x^2 + \sigma_y^2} \quad \alpha_2 = 1 - \alpha_1 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2}. \quad (\text{C.2})$$

La seconde dérivée

$$\frac{d^2 \tilde{\sigma}^2}{d\alpha_1^2} = 2(\sigma_x^2 + \sigma_y^2)$$

étant positive, la valeur de  $\alpha_1$  donnée par l'équation (C.2) est donc bien un minimum et l'estimateur que l'on cherche est donc donné par

$$\tilde{u} = \frac{\sigma_y^2}{\sigma_x^2 + \sigma_y^2} x + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2} y. \quad (\text{C.3})$$

On retrouve donc l'expression (2.6) donnée par l'approche suivie en section 2.1 .

On notera que si on reporte la valeur obtenue pour  $\alpha_1$  dans le calcul de la variance de l'erreur d'estimation, on obtient

$$\begin{aligned} \tilde{\sigma}^2 &= \alpha_1^2 \sigma_x^2 + (1 - \alpha_1)^2 \sigma_y^2 \\ &= \frac{\sigma_y^4}{(\sigma_x^2 + \sigma_y^2)^2} \sigma_x^2 + \frac{\sigma_x^4}{(\sigma_x^2 + \sigma_y^2)^2} \sigma_y^2 \\ &= \frac{\sigma_x^2 \sigma_y^2 (\sigma_x^2 + \sigma_y^2)}{(\sigma_x^2 + \sigma_y^2)^2} \\ &= \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 + \sigma_y^2}, \end{aligned}$$

que l'on peut réécrire sous la forme

$$\frac{1}{\tilde{\sigma}^2} = \frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2}.$$

En appelant **précision** l'inverse de la variance d'une erreur, on remarque que la précision de l'estimateur  $\tilde{u}$  correspond à la somme des précisions des observations.

## Généralisation

On rappelle que la pdf de l'analyse est donnée par (voir section 2.1.3)

$$P^a(\mathbf{x}) \propto P(\mathbf{x} - \mathbf{x}^b) P_o(\mathcal{H}[\mathbf{x}] - \mathbf{y}^o),$$

avec  $\mathbf{x}$  l'estimateur du vecteur état vrai  $\mathbf{x}^t$  et  $\mathbf{y}^o$  le vecteur des observations.  $\mathbf{x}^b$  représente l'information *a priori* et  $\mathcal{H}$  l'opérateur d'observation permettant de ramener l'espace du modèle dans l'espace des observations (on ne considère pas la propagation du modèle ici, *i.e.*  $\mathcal{G} = \mathcal{H}$ ).

Si l'erreur d'estimation est donnée par

$$\boldsymbol{\varepsilon}^a = \mathbf{x} - \mathbf{x}^t,$$

on peut définir la **fonction perte** quadratique

$$L(\boldsymbol{\varepsilon}^a) = (\boldsymbol{\varepsilon}^a)^T \mathbf{S} \boldsymbol{\varepsilon}^a, \quad (\text{C.4})$$

où  $\mathbf{S}$  est une matrice symétrique définie positive arbitraire<sup>79</sup>. Une façon de choisir l'estimateur optimal est de prendre celui qui minimise la moyenne de cette fonction perte  $E[L(\boldsymbol{\varepsilon}^a)]$ . D'après le théorème

<sup>79</sup>Jazwinski (1970) p.147 définit de manière plus large les fonctions pertes admissibles. Ici, on se contente de prendre le cas particulier qui nous intéresse.

5.2 de Jazwinski (1970) (p. 148), cet estimateur est également celui qui minimise la **moyenne conditionnelle**

$$E[L(P^a(\mathbf{x}))] = \int \mathbf{x}P^a(\mathbf{x}) d\mathbf{x}. \quad (\text{C.5})$$

Le choix particulier donné par l'équation (C.4) que l'on a fait pour la fonction perte conduit à l'équivalence entre l'estimateur de la moyenne conditionnelle et de l'estimateur du minimum de variance (Jazwinski, 1970, théorème 5.3 p.149).

On cherche ici l'estimateur optimal sous la forme d'une correction à apporter à l'ébauche

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b), \quad (\text{C.6})$$

où  $\mathbf{K}$  est une matrice à déterminer, et  $\mathbf{H}$  est l'opérateur d'observation linéaire. On définit l'erreur d'ébauche par

$$\boldsymbol{\varepsilon}^b = \mathbf{x}^b - \mathbf{x}^t,$$

et l'erreur d'observation par

$$\boldsymbol{\varepsilon}^o = \mathbf{y}^o - \mathbf{H}\mathbf{x}^t.$$

En reprenant alors l'équation (C.6), et en soustrayant  $\mathbf{x}^t$  de chaque côté, on obtient

$$\begin{aligned} \boldsymbol{\varepsilon}^a &= \boldsymbol{\varepsilon}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \\ &= \boldsymbol{\varepsilon}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}(\boldsymbol{\varepsilon}^b + \mathbf{x}^t)) \\ &= \boldsymbol{\varepsilon}^b + \mathbf{K}(\boldsymbol{\varepsilon}^o - \mathbf{H}\boldsymbol{\varepsilon}^b) \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})\boldsymbol{\varepsilon}^b + \mathbf{K}\boldsymbol{\varepsilon}^o, \end{aligned}$$

et ainsi

$$E[\boldsymbol{\varepsilon}^a] = (\mathbf{I} - \mathbf{K}\mathbf{H})E[\boldsymbol{\varepsilon}^b] + \mathbf{K}E[\boldsymbol{\varepsilon}^o].$$

En prenant pour hypothèse que les erreurs d'ébauche et d'observations sont décorréliées (voir la note 40 à ce sujet), on a alors

$$\begin{aligned} \mathbf{A} = E[\boldsymbol{\varepsilon}^a(\boldsymbol{\varepsilon}^a)^T] &= (\mathbf{I} - \mathbf{K}\mathbf{H})E[\boldsymbol{\varepsilon}^b(\boldsymbol{\varepsilon}^b)^T](\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}E[\boldsymbol{\varepsilon}^o(\boldsymbol{\varepsilon}^o)^T]\mathbf{K}^T \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T, \end{aligned}$$

où  $\mathbf{A}$ ,  $\mathbf{B}$  et  $\mathbf{R}$  sont les matrices de covariances d'erreur d'analyse, d'ébauche et d'observation respectivement.

Minimiser la variance de l'erreur d'estimation, c'est minimiser la trace de  $\mathbf{A}$ , *i.e.* minimiser la somme des éléments de sa diagonale. On a

$$\text{Tr}(\mathbf{A}) = \text{Tr}(\mathbf{B}) + \text{Tr}(\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K}^T) - 2\text{Tr}(\mathbf{B}\mathbf{H}^T\mathbf{K}^T) + \text{Tr}(\mathbf{K}\mathbf{R}\mathbf{K}^T).$$

En considérant  $\text{Tr}(\mathbf{A})$  comme une fonction dépendant de la variable  $\mathbf{K}$ , on peut donc en calculer la différentielle et l'annuler pour trouver le minimum que l'on cherche. On obtient

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}. \quad (\text{C.7})$$

L'estimateur donné par l'équation (C.6) où  $\mathbf{K}$  est défini d'après (C.7) est appelé "Meilleur Estimateur Linéaire Non Biaisé" (**BLUE**<sup>80</sup>).

---

<sup>80</sup>Best Linear Unbiased Estimator

## Filtre de Kalman

La moyenne conditionnelle est à la base des méthodes d'assimilation séquentielles comme l'EnKF et l'OI. Elle permet d'obtenir un estimateur non biaisé mais les opérateurs d'observation non linéaires doivent être pris en compte par des techniques appropriées. Un des avantages de cette formulation est qu'aucune hypothèse n'est faite sur la forme des distributions conditionnelles. Cependant une distribution multimodale (plusieurs pics) peut conduire à des difficultés.

L'algorithme du filtre de Kalman est donné ci-après entre les instants d'observations  $t_i$  et  $t_{i+1}$ . Le vecteur état est noté  $\mathbf{x}^f$ , la matrice des covariances d'erreur de prévision et notée  $\mathbf{P}^f$ , la matrice des covariances d'erreur d'estimation  $\mathbf{P}^a$ , et la matrice des covariances d'erreur du modèle  $\mathbf{Q}$ . L'indice  $i$  indique le temps  $t_i$  et la propagation du modèle de l'instant  $t_i$  à l'instant  $t_{i+1}$  est notée  $\mathcal{M}_{i \rightarrow i+1}$  lorsqu'il s'agit du modèle non linéaire et  $\mathbf{M}_{i+1}$  lorsqu'il s'agit du modèle linéarisé.

- Calcul du vecteur des innovations :

$$\mathbf{d}_i = \mathbf{y}_i^o - \mathcal{H}_i \left[ \mathbf{x}_i^f \right];$$

- Calcul de la matrice de gain :

$$\mathbf{K}_i = \mathbf{P}_i^f \mathbf{H}_i^T \left( \mathbf{H}_i \mathbf{P}_i^f \mathbf{H}_i^T + \mathbf{R}_i \right)^{-1};$$

- Détermination de l'analyse :

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

- Mise à jour des covariances d'erreur d'estimation :

$$\mathbf{P}_i^a = \left( \mathbf{I} - \mathbf{K}_i \mathbf{H}_i^T \right) \mathbf{P}_i^f;$$

- Evolution du vecteur d'état :

$$\mathbf{x}_{i+1}^f = \mathcal{M}_{i \rightarrow i+1} \left[ \mathbf{x}_i^f \right];$$

- Evolution des covariances d'erreur d'estimation :

$$\mathbf{P}_{i+1}^f = \mathbf{M}_{i+1} \mathbf{P}_i^a \mathbf{M}_{i+1}^T + \mathbf{Q}_i.$$

## Annexe D

# Minimisation par la méthode du gradient conjugué

Shewchuk (1994) donne une explication très compréhensible du gradient conjugué. Dans cette annexe, on se contente d'en reprendre les grandes lignes. On peut également se référer au chapitre 10.2 de Golub et van Loan (1996) pour les théorèmes et définitions plus formels.

Résoudre le problème

$$\mathbf{Ax} = \mathbf{b}, \quad (\text{D.1})$$

avec  $\mathbf{A}$  une matrice symétrique définie positive et  $\mathbf{b}$  un vecteur connu, est équivalent à considérer que l'on cherche à minimiser la fonction quadratique

$$\varphi(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{Ax} - \mathbf{x}^T \mathbf{b} + c,$$

où  $c = \varphi(\mathbf{0})$  est une constante, puisque la valeur optimale  $\mathbf{x}^*$  de ce problème est donnée lorsque le gradient s'annule

$$\varphi'(\mathbf{x}^*) = \mathbf{Ax}^* - \mathbf{b} = 0.$$

Le fait que la matrice  $\mathbf{A}$ , Hessienne de la fonction  $\varphi(\mathbf{x})$ , soit définie positive implique que la fonction  $\varphi(\mathbf{x})$  à minimiser a une forme de bol parabolicoïde (voir Figure D.1). Son gradient en un point quelconque pointe vers la direction du plus grand accroissement de  $\varphi(\mathbf{x})$ .

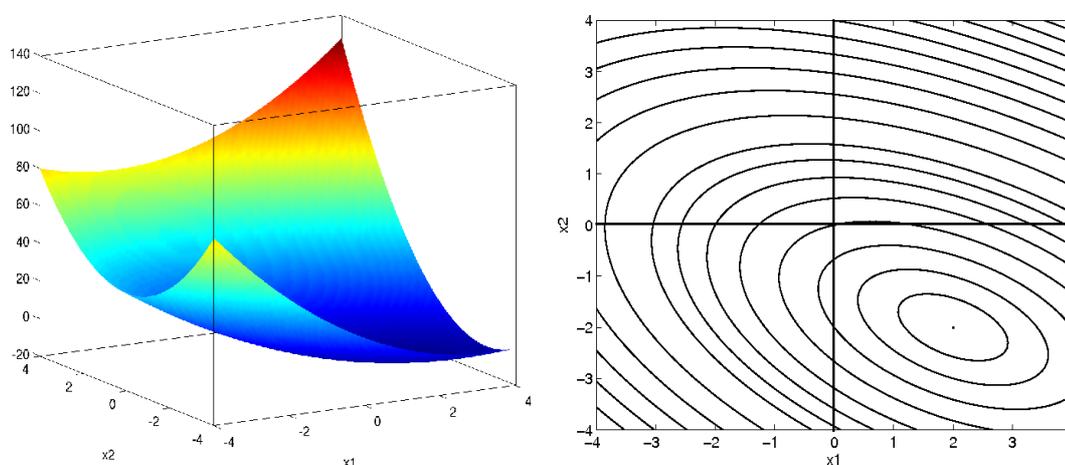


FIG. D.1 –  $\mathbf{A}$  étant définie positive, la fonction  $\varphi(\mathbf{x})$  a une forme de bol parabolicoïde (gauche). Le graphe de droite représente ses contours de niveaux. Dans cet exemple, le minimum est le point  $\mathbf{x}^* = (2, -2)^T$ .

## D.1 Méthode de la plus forte descente

A partir d'un point  $\mathbf{x}_0$  ("initial guess") choisi arbitrairement ou proche de la solution possible, il s'agit de définir itérativement une série de points  $\mathbf{x}_1, \mathbf{x}_2, \dots$ , en suivant des directions de descente jusqu'à approcher au plus près le minimum  $\mathbf{x}^*$ .

A chaque itération, le vecteur **erreur**  $\mathbf{e}_i$  représente l'écart entre le point courant et la solution, tandis que le vecteur **résidu**  $\mathbf{r}_i$  indique l'écart que le point courant induit par rapport à la valeur de  $\mathbf{b}$

$$\mathbf{e}_i = \mathbf{x}_i - \mathbf{x}^*, \quad \mathbf{r}_i = \mathbf{b} - \mathbf{A}\mathbf{x}_i.$$

Il est facile de voir que

$$\mathbf{r}_i = \mathbf{b} - \mathbf{A}(\mathbf{e}_i + \mathbf{x}^*) = -\mathbf{A}\mathbf{e}_i. \quad (\text{D.2})$$

On remarque également que

$$\mathbf{r}_i = -\varphi'(\mathbf{x}_i).$$

Le résidu représente donc la direction de la plus forte descente. A chaque itération, on définit donc un nouveau point le long de cette direction

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha \mathbf{r}_i. \quad (\text{D.3})$$

Il nous reste maintenant à définir le pas  $\alpha$ . Pour cela, on cherche celui qui minimise la fonction  $\varphi(\mathbf{x}_{i+1})$  et donc qui annule la dérivée directionnelle

$$\frac{d}{d\alpha}\varphi(\mathbf{x}_{i+1}) = \frac{d}{d\mathbf{x}}\varphi(\mathbf{x}_{i+1})\frac{d\mathbf{x}}{d\alpha} = \varphi'(\mathbf{x}_{i+1})^T \mathbf{r}_i = -\mathbf{r}_{i+1}^T \mathbf{r}_i = 0.$$

$\alpha$  est donc défini de manière à ce que les directions de descente soient orthogonales. On a alors

$$\mathbf{r}_{i+1}^T \mathbf{r}_i = (\mathbf{b} - \mathbf{A}\mathbf{x}_{i+1})^T \mathbf{r}_i = (\mathbf{b} - \mathbf{A}(\mathbf{x}_i + \alpha \mathbf{r}_i))^T \mathbf{r}_i = (\mathbf{b} - \mathbf{A}\mathbf{x}_i)^T \mathbf{r}_i - \alpha (\mathbf{A}\mathbf{r}_i)^T \mathbf{r}_i = 0,$$

et ainsi

$$\begin{aligned} \alpha (\mathbf{A}\mathbf{r}_i)^T \mathbf{r}_i &= (\mathbf{b} - \mathbf{A}\mathbf{x}_i)^T \mathbf{r}_i \\ \alpha \mathbf{r}_i^T \mathbf{A}\mathbf{r}_i &= \mathbf{r}_i^T \mathbf{r}_i \\ \alpha &= \frac{\mathbf{r}_i^T \mathbf{r}_i}{\mathbf{r}_i^T \mathbf{A}\mathbf{r}_i}. \end{aligned} \quad (\text{D.4})$$

A chaque itération, deux produits matrice-vecteurs différents doivent être effectués : un pour calculer  $\alpha$  et un pour calculer le résidu. Cependant, il est possible d'éliminer ce dernier en remarquant que le résidu peut être calculé récursivement. Pour cela, on multiplie par  $-\mathbf{A}$  et on ajoute  $\mathbf{b}$  aux deux membres de l'équation (D.3) :

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{x}_i + \alpha \mathbf{r}_i \\ -\mathbf{A}\mathbf{x}_{i+1} + \mathbf{b} &= -\mathbf{A}\mathbf{x}_i + \mathbf{b} - \alpha \mathbf{A}\mathbf{r}_i \\ \mathbf{r}_{i+1} &= \mathbf{r}_i - \alpha \mathbf{A}\mathbf{r}_i. \end{aligned}$$

Cette astuce nécessite néanmoins la définition (et donc le stockage) d'un nouveau vecteur  $\mathbf{q}_i$  tel que

$$\mathbf{q}_i = \mathbf{A}\mathbf{r}_i.$$

D'où l'algorithme

## Méthode de la plus forte descente

- 
- $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$
  - $i = 0$
  - Tant que le critère d'arrêt n'est pas satisfait
 

•	$\mathbf{q}_i = \mathbf{A}\mathbf{r}_i$
•	$\alpha_i = \mathbf{r}_i^T \mathbf{r}_i / \mathbf{q}_i^T \mathbf{r}_i,$
•	$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{r}_i$
•	$\mathbf{r}_{i+1} = \mathbf{r}_i - \alpha \mathbf{q}_i$
•	$i = i + 1$
  - Fin tant que
- 

## D.2 Méthode du gradient conjugué

Lorsque toutes les valeurs propres  $\lambda_k$  de la matrice  $\mathbf{A}$  sont égales, les contours de niveaux de la fonction  $\varphi(\mathbf{x})$  sont circulaires. La méthode de la plus forte descente converge alors en une seule itération quelque soit le point de départ. Mais ce n'est généralement pas le cas, et on définit le **conditionnement** d'une matrice par la formule

$$\kappa(\mathbf{A}) = \frac{\lambda_{\max}}{\lambda_{\min}}. \quad (\text{D.5})$$

Plus ce nombre est grand et plus la matrice est mal conditionnée. Cela se traduit par l'allongement en ellipse des contours et la convergence de la méthode peut alors être très lente, les directions de descente rebondissant d'une pente à l'autre. Les directions données par les résidus n'étant pas forcément optimales, on considère un nouveau jeu de directions  $\{\mathbf{p}_1, \mathbf{p}_2, \dots\}$   $\mathbf{A}$ -orthogonales. Deux directions  $\mathbf{p}_i$  et  $\mathbf{p}_j$  sont dites  **$\mathbf{A}$ -orthogonales** ou  **$\mathbf{A}$ -conjuguées** si

$$\mathbf{p}_i^T \mathbf{A} \mathbf{p}_j = 0.$$

La Figure D.2 illustre cette définition. A gauche, les vecteurs représentés sont  $\mathbf{A}$ -orthogonaux. Si

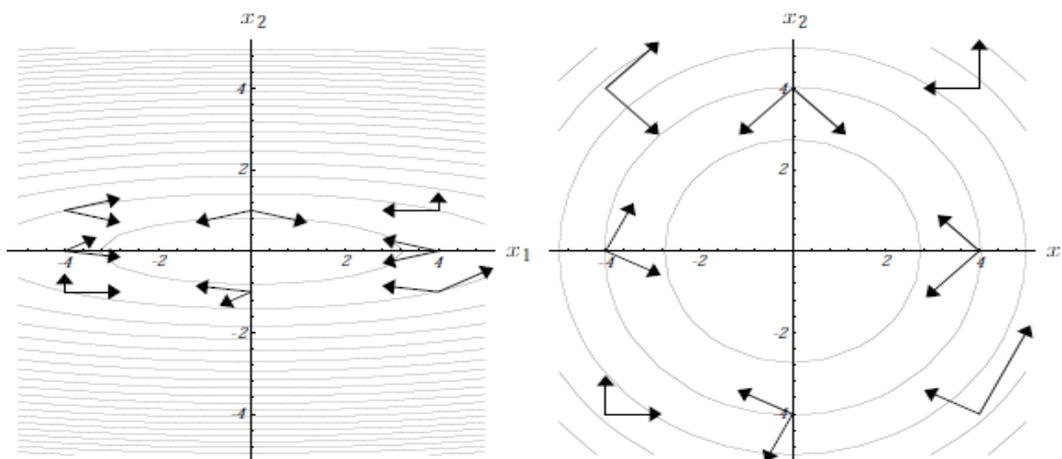


FIG. D.2 – Si les contours ellipsoïdaux de la figure de gauche étaient étirés jusqu'à les rendre circulaires, on obtiendrait la figure de droite. Les vecteurs représentés à gauche sont  $\mathbf{A}$ -orthogonaux parce qu'orthogonaux à droite (Figure 22 de Shewchuk, 1994).

on pouvait étirer les contours ellipsoïdaux jusqu'à les rendre circulaires, les vecteurs seraient alors orthogonaux (droite). On a alors

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i.$$

En soustrayant  $\mathbf{x}^*$  aux deux membres de l'équation précédente, on obtient

$$\mathbf{e}_{i+1} = \mathbf{e}_i + \alpha_i \mathbf{p}_i.$$

Pour calculer le pas  $\alpha_i$ , on utilise le fait que la nouvelle erreur  $\mathbf{e}_{i+1}$  doit être  $\mathbf{A}$ -orthogonale à la direction  $\mathbf{p}_i$  :

$$\mathbf{p}_i^T \mathbf{A} \mathbf{e}_{i+1} = \mathbf{p}_i^T \mathbf{A} (\mathbf{e}_i + \alpha_i \mathbf{p}_i) = -\mathbf{p}_i^T \mathbf{r}_i + \alpha_i \mathbf{p}_i^T \mathbf{A} \mathbf{p}_i = 0,$$

où on a utilisé le résultat de l'équation (D.2), d'où

$$\begin{aligned} \alpha_i \mathbf{p}_i^T \mathbf{A} \mathbf{p}_i &= \mathbf{p}_i^T \mathbf{r}_i \\ \alpha_i &= \frac{\mathbf{p}_i^T \mathbf{r}_i}{\mathbf{p}_i^T \mathbf{A} \mathbf{p}_i}. \end{aligned} \quad (\text{D.6})$$

Pour la première direction de descente, on impose de prendre le résidu  $\mathbf{p}_0 = \mathbf{r}_0$ . Les directions suivantes sont choisies pour être  $\mathbf{A}$ -orthogonales mais proches des résidus. On pose alors

$$\mathbf{p}_i = \mathbf{r}_i + \beta_i \mathbf{p}_{i-1}, \quad (\text{D.7})$$

et on déduit  $\beta_i$  de l' $\mathbf{A}$ -orthogonalité des directions :

$$\mathbf{p}_i^T \mathbf{A} \mathbf{p}_{i-1} = (\mathbf{r}_i + \beta_i \mathbf{p}_{i-1})^T \mathbf{A} \mathbf{p}_{i-1} = \mathbf{r}_i^T \mathbf{A} \mathbf{p}_{i-1} + \beta_i \mathbf{p}_{i-1}^T \mathbf{A} \mathbf{p}_{i-1} = 0,$$

d'où

$$\begin{aligned} \beta_i \mathbf{p}_{i-1}^T \mathbf{A} \mathbf{p}_{i-1} &= -\mathbf{r}_i^T \mathbf{A} \mathbf{p}_{i-1} \\ \beta_i &= -\frac{\mathbf{r}_i^T \mathbf{A} \mathbf{p}_{i-1}}{\mathbf{p}_{i-1}^T \mathbf{A} \mathbf{p}_{i-1}}. \end{aligned} \quad (\text{D.8})$$

D'après l'équation (D.2), on peut écrire

$$\mathbf{r}_i = -\mathbf{A} \mathbf{e}_i = -\mathbf{A} (\mathbf{e}_{i-1} + \alpha_{i-1} \mathbf{p}_{i-1}) = \mathbf{r}_{i-1} - \alpha_{i-1} \mathbf{A} \mathbf{p}_{i-1}.$$

En multipliant les deux membres de cette équation par  $\mathbf{r}_i^T$ , on a alors

$$\mathbf{r}_i^T \mathbf{r}_i = \mathbf{r}_i^T \mathbf{r}_{i-1} - \alpha_{i-1} \mathbf{r}_i^T \mathbf{A} \mathbf{p}_{i-1} = -\alpha_{i-1} \mathbf{r}_i^T \mathbf{A} \mathbf{p}_{i-1},$$

car les résidus sont orthogonaux (puisque les directions sont construites à partir de ces résidus), d'où

$$\mathbf{r}_i^T \mathbf{A} \mathbf{p}_{i-1} = -\frac{1}{\alpha_{i-1}} \times \mathbf{r}_i^T \mathbf{r}_i.$$

A l'aide de l'équation (D.6), l'équation (D.8) devient alors

$$\begin{aligned} \beta_i &= \frac{1}{\alpha_{i-1}} \times \frac{\mathbf{r}_i^T \mathbf{r}_i}{\mathbf{p}_{i-1}^T \mathbf{A} \mathbf{p}_{i-1}} \\ &= \frac{\mathbf{p}_{i-1}^T \mathbf{A} \mathbf{p}_{i-1}}{\mathbf{p}_{i-1}^T \mathbf{r}_{i-1}} \times \frac{\mathbf{r}_i^T \mathbf{r}_i}{\mathbf{p}_{i-1}^T \mathbf{A} \mathbf{p}_{i-1}} \\ &= \frac{\mathbf{r}_i^T \mathbf{r}_i}{\mathbf{p}_{i-1}^T \mathbf{r}_{i-1}}. \end{aligned} \quad (\text{D.9})$$

D'autre part, en multipliant les deux membres de l'équation (D.7) par  $\mathbf{r}_i$ , on montre que

$$\mathbf{p}_i^T \mathbf{r}_i = (\mathbf{r}_i + \beta_i \mathbf{p}_{i-1})^T \mathbf{r}_i = \mathbf{r}_i^T \mathbf{r}_i + \beta_i \mathbf{p}_{i-1}^T \mathbf{r}_i = \mathbf{r}_i^T \mathbf{r}_i,$$

car le résidu est orthogonal à la direction précédente. L'équation (D.9) devient alors

$$\beta_i = \frac{\mathbf{r}_i^T \mathbf{r}_i}{\mathbf{r}_{i-1}^T \mathbf{r}_{i-1}},$$

et l'équation (D.6)

$$\alpha_i = \frac{\mathbf{r}_i^T \mathbf{r}_i}{\mathbf{p}_i^T \mathbf{A} \mathbf{p}_i}.$$

Comme précédemment pour la méthode de la plus forte descente, on peut écrire le résidu de façon récursive :

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{x}_i + \alpha \mathbf{p}_i \\ -\mathbf{A} \mathbf{x}_{i+1} + \mathbf{b} &= -\mathbf{A} \mathbf{x}_i + \mathbf{b} - \alpha \mathbf{A} \mathbf{p}_i \\ \mathbf{r}_{i+1} &= \mathbf{r}_i - \alpha \mathbf{A} \mathbf{p}_i, \end{aligned}$$

et on définit  $\mathbf{q}_i$  tel que

$$\mathbf{q}_i = \mathbf{A} \mathbf{p}_i.$$

On obtient donc l'algorithme CG<sup>81</sup>

---

Gradient Conjugué (CG)
<ul style="list-style-type: none"> <li>• <math>\mathbf{r}_0 = \mathbf{b} - \mathbf{A} \mathbf{x}_0</math></li> <li>• <math>\mathbf{p}_0 = \mathbf{r}_0</math></li> <li>• <math>i = 0</math></li> <li>• Tant que le critère d'arrêt n'est pas satisfait           <div style="border-left: 1px solid black; padding-left: 10px; margin-left: 20px;"> <ul style="list-style-type: none"> <li>• <math>\mathbf{q}_i = \mathbf{A} \mathbf{p}_i</math></li> <li>• <math>\alpha_i = \mathbf{r}_i^T \mathbf{r}_i / \mathbf{q}_i^T \mathbf{p}_i</math></li> <li>• <math>\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i</math></li> <li>• <math>\mathbf{r}_{i+1} = \mathbf{r}_i - \alpha_i \mathbf{q}_i</math></li> <li>• <math>i = i + 1</math></li> <li>• <math>\beta_i = \mathbf{r}_i^T \mathbf{r}_i / \mathbf{r}_{i-1}^T \mathbf{r}_{i-1}</math></li> <li>• <math>\mathbf{p}_i = \mathbf{r}_i + \beta_i \mathbf{p}_{i-1}</math></li> </ul> </div> </li> </ul>
<ul style="list-style-type: none"> <li>• Fin tant que</li> </ul>

Lorsque l'on se place dans le cadre d'une arithmétique exacte (pas d'erreur d'arrondi), pour une matrice  $\mathbf{A}$  de dimension  $n \times n$ , le nombre d'itérations nécessaires pour trouver la solution est inférieur à  $r$ , le nombre de valeurs propres distinctes de la matrices ( $r \leq n$ ). Le taux de convergence est alors

$$\frac{\|\mathbf{e}_i\|_{\mathbf{A}}}{\|\mathbf{e}_0\|_{\mathbf{A}}} \leq 2 \left( \frac{\sqrt{\kappa(\mathbf{A})} - 1}{\sqrt{\kappa(\mathbf{A})} + 1} \right)^i,$$

avec  $\|\mathbf{e}_i\|_{\mathbf{A}} = (\mathbf{e}_i^T \mathbf{A} \mathbf{e}_i)^{\frac{1}{2}}$ . Cependant, les erreurs d'arrondi de la matrice nécessitent de définir un critère d'arrêt fonction de l'erreur en norme. On peut également définir un nombre d'itérations maximal.

---

<sup>81</sup>Conjugate Gradient

### D.3 Gradient conjugué préconditionné

Lorsque le nombre  $\kappa(\mathbf{A})$  défini par l'équation (D.5) est grand, la matrice  $\mathbf{A}$  est mal conditionnée et la convergence peut être lente. On cherche alors à préconditionner le problème, *i.e.* à trouver une matrice  $\mathbf{M}$  symétrique, définie positive telle que  $\kappa(\mathbf{MA}) < \kappa(\mathbf{A})$ . En factorisant cette matrice sous la forme  $\mathbf{M} = \mathbf{PP}^T$ , le problème donné par l'équation (D.1) devient

$$\mathbf{P}^T \mathbf{A} \mathbf{P} \mathbf{v} = \mathbf{P}^T \mathbf{b} \quad \text{avec} \quad \mathbf{v} = \mathbf{P}^{-1} \mathbf{x}.$$

Le résidu de ce nouveau problème peut s'écrire

$$\tilde{\mathbf{r}}_i = \mathbf{P}^T \mathbf{b} - \mathbf{P}^T \mathbf{A} \mathbf{P} \mathbf{v} = \mathbf{P}^T (\mathbf{b} - \mathbf{A} \mathbf{x}) = \mathbf{P}^T \mathbf{r}_i,$$

et la direction est réécrite en

$$\tilde{\mathbf{p}}_i = \mathbf{P}^{-1} \mathbf{p}_i.$$

On a alors

$$\alpha_i = \frac{\tilde{\mathbf{r}}_i^T \tilde{\mathbf{r}}_i}{\tilde{\mathbf{p}}_i^T \mathbf{P}^T \mathbf{A} \mathbf{P} \tilde{\mathbf{p}}_i} = \frac{\mathbf{r}_i^T \mathbf{P} \mathbf{P}^T \mathbf{r}_i}{\mathbf{p}_i^T (\mathbf{P}^T)^{-1} \mathbf{P}^T \mathbf{A} \mathbf{P} \mathbf{P}^{-1} \mathbf{p}_i} = \frac{\mathbf{r}_i^T \mathbf{M} \mathbf{r}_i}{\mathbf{p}_i^T \mathbf{A} \mathbf{p}_i},$$

et

$$\beta_i = \frac{\tilde{\mathbf{r}}_i^T \tilde{\mathbf{r}}_i}{\tilde{\mathbf{r}}_{i-1}^T \tilde{\mathbf{r}}_{i-1}} = \frac{\mathbf{r}_i^T \mathbf{P} \mathbf{P}^T \mathbf{r}_i}{\mathbf{r}_{i-1}^T \mathbf{P} \mathbf{P}^T \mathbf{r}_{i-1}} = \frac{\mathbf{r}_i^T \mathbf{M} \mathbf{r}_i}{\mathbf{r}_{i-1}^T \mathbf{M} \mathbf{r}_{i-1}}.$$

A chaque itération, outre le produit  $\mathbf{q}_i = \mathbf{A} \mathbf{p}_i$ , il faut également calculer le produit  $\mathbf{M} \mathbf{r}_i$ . Pour que l'algorithme du PCG soit moins coûteux que l'algorithme du CG, ce dernier produit doit être facile à calculer ou évaluer. Il faut cependant définir (et stocker) un nouveau vecteur  $\mathbf{z}_i$  tel que

$$\mathbf{z}_i = \mathbf{M} \mathbf{r}_i.$$

L'algorithme PCG<sup>82</sup> devient

---

#### Gradient Conjugué Préconditionné (PCG)

---

- $\mathbf{r}_0 = \mathbf{b} - \mathbf{A} \mathbf{x}_0$
  - $\mathbf{z}_0 = \mathbf{M} \mathbf{r}_0$
  - $\mathbf{p}_0 = \mathbf{z}_0$
  - $i = 0$
  - Tant que le critère d'arrêt n'est pas satisfait
    - $\mathbf{q}_i = \mathbf{A} \mathbf{p}_i$
    - $\alpha_i = \mathbf{r}_i^T \mathbf{z}_i / \mathbf{q}_i^T \mathbf{p}_i$
    - $\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$
    - $\mathbf{r}_{i+1} = \mathbf{r}_i - \alpha_i \mathbf{q}_i$
    - $i = i + 1$
    - $\mathbf{z}_i = \mathbf{M} \mathbf{r}_i$
    - $\beta_i = \mathbf{r}_i^T \mathbf{z}_i / \mathbf{r}_{i-1}^T \mathbf{z}_{i-1}$
    - $\mathbf{p}_i = \mathbf{z}_i + \beta_i \mathbf{p}_{i-1}$
  - Fin tant que
- 

Lorsque l'on utilise la matrice des covariances d'erreur d'ébauche comme préconditionneur ( $\mathbf{M} = \mathbf{B}$ ), on retrouve l'approche proposée par Derber et Rosati (1989).

<sup>82</sup>Preconditioned Conjugate Gradient

## Annexe E

# Méthode de normalisation proposée par Purser *et al.* (2003b)

La normalisation proposée par Purser *et al.* (2003b) est donnée par

$$\bar{\lambda}(z) = \sqrt{4\pi\overline{\kappa(z)}t} = \sqrt{2\pi(\overline{L(z)})^2},$$

avec

$$\overline{\kappa(z)} = \mathbf{F}^{1/2}\kappa(z),$$

où  $\mathbf{F}^{1/2}$  représente l'application de la "racine carrée" de l'opérateur de diffusion et  $\kappa(z)$  est le coefficient de diffusion inhomogène. Bien que la justification de cette approche, détaillée ci-dessous, soit basée sur l'équation de diffusion, elle n'en reste pas moins valide pour le filtre récursif.

Soit l'équation de diffusion inhomogène appliquée à un champ  $\varphi = \varphi(z, t)$  où  $\kappa(z)$  représente le coefficient de diffusion

$$\frac{\partial\varphi}{\partial t} - \frac{\partial}{\partial z} \left( \kappa \frac{\partial\varphi}{\partial z} \right) = 0 \iff \dot{\varphi} - (\kappa\varphi)' = 0, \quad (\text{E.1})$$

où  $\dot{\varphi}$  représente la dérivée de  $\varphi$  par rapport à  $t$  et  $\varphi'$  sa dérivée par rapport à  $z$ . Lorsque la condition initiale est une fonction de Dirac au point  $z_0$ , la solution de cette équation représente la fonction de Green, *i.e.* à un facteur prêt, la fonction de corrélation que l'on souhaite modéliser centrée en ce point  $z_0$ . Ici, nous supposons que cette solution, ou fonction de covariance, peut s'écrire

$$\varphi(z, t) = e^{g(r, t)}, \quad (\text{E.2})$$

où  $r = z - z_0$ . Dans le cas d'un coefficient de diffusion constant on a  $g(r, t) = -r^2/4\kappa t$ . On cherche ici à évaluer la variance de cette solution, *i.e.* son amplitude lorsque  $r = 0$ , et donc  $z = z_0$ . Pour cela, on suppose qu'il est possible de développer  $g(r, t)$  et  $\kappa(z)$  en séries entières au voisinage de 0 et  $z_0$ , respectivement

$$g(r, t) = g_0(t) + g_1(t)r + g_2(t)r^2 + \dots, \quad (\text{E.3})$$

$$\begin{aligned} \kappa(z) &= \kappa_0 + \kappa_1(z - z_0) + \kappa_2(z - z_0)^2 + \dots \\ &= \kappa_0 + \kappa_1 r + \kappa_2 r^2 + \dots \end{aligned} \quad (\text{E.4})$$

Si  $\kappa(z)$  est constant, on a :

$$\begin{cases} \kappa(z) = \kappa_0, \\ g_0(t) = -\frac{1}{2} \ln t - \frac{1}{2} \ln(4\pi\kappa_0), \\ g_2(t) = -\frac{1}{4\kappa_0 t}, \\ g_k(t) = 0, \text{ pour } k = 1 \text{ ou } k > 2. \end{cases} \quad (\text{E.5})$$

**Preuve:**

$$\begin{aligned}
 \varphi(z, t) &= e^{g(z-z_0, t)} = \exp\left(-\frac{1}{2} \ln t - \frac{1}{2} \ln(4\pi\kappa_0) - \frac{(z-z_0)^2}{4\kappa_0 t}\right) \\
 &= \exp\left(\ln t^{-1/2}\right) \exp\left(\ln(4\pi\kappa_0)^{-1/2}\right) \exp\left(-\frac{(z-z_0)^2}{4\kappa_0 t}\right) \\
 &= \frac{1}{\sqrt{4\pi\kappa_0 t}} e^{-(z-z_0)^2/4\kappa_0 t},
 \end{aligned}$$

solution de l'équation de diffusion homogène lorsque la condition initiale  $\varphi(z, 0)$  est une fonction de Dirac en  $z_0$ .

**Question :** Quel est l'effet de perturbations de premier ordre de  $\kappa(z)$  sur le terme d'amplitude  $g_0(t)$  ?

Compte tenu de l'équation (E.2), les termes de l'équation (E.1) peuvent s'écrire

$$\begin{aligned}
 \dot{\varphi} &= (\dot{e}^g) = \dot{g}e^g = \dot{g}\varphi, \\
 (\kappa\varphi)' &= (\kappa(e^g)')' = (\kappa g' e^g)' = (\kappa g')' e^g + \kappa g' (e^g)' \\
 &= (\kappa g')' e^g + \kappa g' g' e^g = (\kappa g')' \varphi + \kappa (g')^2 \varphi.
 \end{aligned}$$

L'équation de diffusion (E.1) donne alors l'égalité

$$\dot{g} = (\kappa g')' + \kappa (g')^2. \quad (\text{E.6})$$

La solution (E.2) étant symétrique, les termes  $g_k(t)$  avec  $k$  impaires sont nuls. De plus, les termes  $\kappa_k$  avec  $k$  impaires ne peuvent pas affecter l'amplitude et nous ne considérerons que les termes en  $k$  paires. À partir de l'équation (E.3) on peut alors dériver

$$\begin{aligned}
 g &= g_0 + g_2 r^2 + g_4 r^4 + g_6 r^6 + \dots, \\
 \dot{g} &= \dot{g}_0 + \dot{g}_2 r^2 + \dot{g}_4 r^4 + \dots, \\
 g' &= 2g_2 r + 4g_4 r^3 + 6g_6 r^5 + \dots, \\
 (g')^2 &= 4g_2^2 r^2 + 16g_2 g_4 r^4 + (24g_2 g_6 + 16g_4^2) r^6 + \dots,
 \end{aligned} \quad (\text{E.7})$$

où la dépendance à  $t$  des coefficients est sous-entendue pour simplifier l'écriture. En combinant ces résultats avec l'équation (E.4), on obtient

$$\begin{aligned}
 \kappa(z) &= \kappa_0 + \kappa_2 r^2 + \kappa_4 r^4 + \dots, \\
 \kappa g' &= 2\kappa_0 g_2 r + (4\kappa_0 g_4 + 2\kappa_2 g_2) r^3 + (6\kappa_0 g_6 + 4\kappa_2 g_4 + 2\kappa_4 g_2) r^5 \dots, \\
 (\kappa g')' &= 2\kappa_0 g_2 + (12\kappa_0 g_4 + 6\kappa_2 g_2) r^2 + (30\kappa_0 g_6 + 20\kappa_2 g_4 + 10\kappa_4 g_2) r^4 \dots,
 \end{aligned} \quad (\text{E.8})$$

$$\kappa (g')^2 = 4\kappa_0 g_2^2 r^2 + (16\kappa_0 g_2 g_4 + 4\kappa_2 g_2^2) r^4 + \dots. \quad (\text{E.9})$$

L'équation (E.6) permet grâce aux expressions des équations (E.7), (E.8) et (E.9) d'identifier les termes en fonction de la puissance de  $r$

$$\dot{g}_0 = 2\kappa_0 g_2, \quad (\text{E.10})$$

$$\dot{g}_2 = 4\kappa_0 g_2^2 + 12\kappa_0 g_4 + 6\kappa_2 g_2, \quad (\text{E.11})$$

$$\dot{g}_4 = 30\kappa_0 g_6 + 20\kappa_2 g_4 + 10\kappa_4 g_2 + 16\kappa_0 g_2 g_4 + 4\kappa_2 g_2^2, \quad (\text{E.12})$$

⋮

En se rappelant que les coefficients  $g_k$  sont fonction de  $t$ , on suppose qu'il est possible de les développer en puissance de  $t$  en commençant par  $t^{-1}$  :

$$\begin{aligned} g_2 &= g_{2,-1} t^{-1} + g_{2,0} + g_{2,1} t + \dots, \\ g_4 &= g_{4,-1} t^{-1} + g_{4,0} + g_{4,1} t + \dots. \end{aligned}$$

L'équation (E.11) devient

$$\begin{aligned} -g_{2,-1} t^{-2} + g_{2,1} + \dots &= 4\kappa_0 (g_{2,-1} t^{-1} + g_{2,0} + g_{2,1} t + \dots)^2 \\ &\quad + 12\kappa_0 (g_{4,-1} t^{-1} + g_{4,0} + g_{4,1} t + \dots) \\ &\quad + 6\kappa_2 (g_{2,-1} t^{-1} + g_{2,0} + g_{2,1} t + \dots) \\ &= 4\kappa_0 g_{2,-1}^2 t^{-2} \\ &\quad + (8\kappa_0 g_{2,-1} g_{2,0} + 12\kappa_0 g_{4,-1} + 6\kappa_2 g_{2,-1}) t^{-1} + \dots, \end{aligned} \quad (\text{E.13})$$

soit, par identification des termes en  $t^{-2}$

$$-g_{2,-1} = 4\kappa_0 g_{2,-1}^2 \Rightarrow g_{2,-1} = -\frac{1}{4\kappa_0}. \quad (\text{E.14})$$

De la même manière, l'équation (E.12) devient

$$\begin{aligned} -g_{4,-1} t^{-2} + g_{4,1} + \dots &= 30\kappa_0 g_6 + 20\kappa_2 (g_{4,-1} t^{-1} + g_{4,0} g_{4,1} t + \dots) \\ &\quad + 10\kappa_4 (g_{2,-1} t^{-1} + g_{2,0} + g_{2,1} t + \dots) \\ &\quad + 16\kappa_0 (g_{2,-1} t^{-1} + g_{2,0} + g_{2,1} t + \dots) (g_{4,-1} t^{-1} + g_{4,0} g_{4,1} t + \dots) \\ &\quad + 4\kappa_2 (g_{2,-1} t^{-1} + g_{2,0} + g_{2,1} t + \dots)^2 \\ &= (16\kappa_0 g_{2,-1} g_{4,-1} + g_{2,-1}^2) t^{-2} + \dots, \end{aligned}$$

soit, par identification des termes en  $t^{-2}$  et en appliquant l'équation (E.14)

$$\begin{aligned} -g_{4,-1} &= 16\kappa_0 g_{2,-1} g_{4,-1} + 4\kappa_2 g_{2,-1}^2 = -\frac{16\kappa_0}{4\kappa_0} g_{4,-1} + \frac{4\kappa_2}{16\kappa_0^2} \\ &= -4g_{4,-1} + \frac{\kappa_2}{4\kappa_0^2} \Rightarrow g_{4,-1} = \frac{\kappa_2}{12\kappa_0^2}. \end{aligned} \quad (\text{E.15})$$

En identifiant maintenant les termes en  $t^{-1}$  de l'équation (E.13) et en appliquant les équations (E.14) et (E.15), on obtient

$$\begin{aligned} 0 &= 8\kappa_0 g_{2,-1} g_{2,0} + 12\kappa_0 g_{4,-1} + 6\kappa_2 g_{2,-1} \\ &= -\frac{8\kappa_0}{4\kappa_0} g_{2,0} + \frac{12\kappa_0 \kappa_2}{12\kappa_0^2} - \frac{6\kappa_2}{4\kappa_0} = -2g_{2,0} - \frac{\kappa_2}{2\kappa_0} \Rightarrow g_{2,0} = -\frac{\kappa_2}{4\kappa_0}. \end{aligned} \quad (\text{E.16})$$

Enfin, l'équation (E.10) devient

$$\begin{aligned} \dot{g}_0 &= 2\kappa_0 (g_{2,-1} t^{-1} + g_{2,0} + g_{2,1} t + \dots) \\ &= -\frac{2\kappa_0}{4\kappa_0} t^{-1} - \frac{2\kappa_0 \kappa_2}{4\kappa_0} + 2\kappa_0 g_{2,1} t + \dots \\ &= -\frac{1}{2} t^{-1} - \frac{\kappa_2}{2} + 2\kappa_0 g_{2,1} t + \dots, \end{aligned}$$

que l'on peut intégrer par rapport à  $t$  pour obtenir

$$g_0(t) = -\frac{1}{2} \ln t + g_{0,0} - \frac{1}{2} \kappa_2 t + O(t^2),$$

où  $g_{0,0}$  est la constante d'intégration que l'on définit par

$$g_{0,0} = -\frac{1}{2} \ln(4\pi\kappa_0),$$

par cohérence avec le cas du coefficient de diffusion constant donné par le système (E.5). Pour absorber la perturbation de premier ordre dans le coefficient d'amplitude  $g_0$ , on écrit

$$g_0(t) = -\frac{1}{2} \ln t - \frac{1}{2} \ln(4\pi\overline{\kappa_{z_0}}),$$

avec

$$\begin{aligned} -\frac{1}{2} \ln(4\pi\overline{\kappa_{z_0}}) &= -\frac{1}{2} \ln(4\pi\kappa_0) - \frac{1}{2} \kappa_2 t + O(t^2) \\ \Rightarrow 4\pi\overline{\kappa_{z_0}} &= 4\pi\kappa_0 \exp(\kappa_2 t + O(t^2)). \end{aligned}$$

En utilisant le développement en série de l'exponentielle (Jeffrey, 1995, p.12), on définit

$$\overline{\kappa_{z_0}} = \kappa_0 + \kappa_0\kappa_2 t + O(t^2). \quad (\text{E.17})$$

Ainsi on a

Si  $\kappa(z)$  varie suivant une perturbation de premier ordre, on a

$$\begin{cases} \kappa(z) = \kappa_0 + \kappa_2(z - z_0)^2 + \kappa_4(z - z_0)^4 + \dots, \\ \overline{\kappa_{z_0}} = \kappa_0 + \kappa_0\kappa_2 t + O(t^2), \\ g_0(t) = -\frac{1}{2} \ln t - \frac{1}{2} \ln(4\pi\overline{\kappa_{z_0}}), \\ \vdots \end{cases} \quad (\text{E.18})$$

Soit maintenant  $\varphi(z, t/2)$ , la solution de l'équation de diffusion intégrée jusqu'à  $t/2$  avec un coefficient de diffusion constant  $\kappa(z) = \kappa_0$  et une fonction de Dirac en  $z_0$  comme condition initiale :

$$\varphi(z, t/2) = \frac{1}{\sqrt{2\pi\kappa_0 t}} e^{-(z-z_0)^2/2\kappa_0 t}.$$

L'intégrale d'une fonction gaussienne est donnée par (Jeffrey, 1995, équation (30), p. 250)

$$\int_{-\infty}^{\infty} \varphi(z, t/2) dz = \frac{1}{\sqrt{2\pi\kappa_0 t}} \int_{-\infty}^{\infty} e^{-(z-z_0)^2/2\kappa_0 t} dz = \frac{1}{\sqrt{2\pi\kappa_0 t}} \times \sqrt{2\pi\kappa_0 t} = 1.$$

On a également (Jeffrey, 1995, équation (26), p. 249)

$$\begin{aligned} \int_{-\infty}^{\infty} (z - z_0)^2 \varphi(z, t/2) dz &= \frac{2}{\sqrt{2\pi\kappa_0 t}} \int_0^{\infty} (z - z_0)^2 e^{-(z-z_0)^2/2\kappa_0 t} dz \\ &= \frac{2}{\sqrt{2\pi\kappa_0 t}} \times \frac{2\kappa_0 t}{4} \sqrt{2\pi\kappa_0 t} = \kappa_0 t. \end{aligned}$$

L'équation (E.17) peut alors s'écrire

$$\begin{aligned} \overline{\kappa_{z_0}} &= \kappa_0 + \kappa_0\kappa_2 t + O(t^2) \\ &\approx \kappa_0 \int_{-\infty}^{\infty} \varphi(z, t/2) dz + \kappa_2 \int_{-\infty}^{\infty} (z - z_0)^2 \varphi(z, t/2) dz \\ &\approx \int_{-\infty}^{\infty} \kappa(z) \varphi(z, t/2) dz. \end{aligned}$$

En généralisant cette approche à tous les points  $z_0$  de la grille, on a

$$\overline{\kappa(z)} \approx \frac{1}{\sqrt{2\pi\kappa_0 t}} \int_{-\infty}^{\infty} e^{-(z-z')^2/2\kappa_0 t} \kappa(z') dz'.$$

On retrouve ici un produit de convolution équivalent à l'application de la "racine carrée" de l'opérateur de diffusion sur le champ  $\kappa(z)$

$$\overline{\kappa(z)} \approx \mathbf{F}^{1/2} \kappa(z).$$

## Annexe F

# Réorganisation des processeurs

On expose dans cette annexe le principe de réorganisation des processeurs utilisé lors de la résolution du schéma numérique de l'équation de diffusion 1D. Bien que plusieurs configurations soient disponibles dans le modèle NEMO, on se limite aux configurations globales ORCA1 et ORCA2 sans entrer trop avant dans les détails techniques. Les exemples donnés dans cette annexe, le sont pour la grille ORCA2 avec 4 processeurs suivant l'axe  $i$  et 2 processeurs suivant l'axe  $j$ , soit 8 processeurs en tout.

A l'instar de NEMO (Madec, 2008), le code NEMOVAR est défini suivant le système de coordonnées curvilinéaires  $(i, j, k)$ , où  $(i, j)$  représente l'horizontale et  $k$  la verticale. Le principe de parallélisation qui lui est associé, consiste à décomposer le domaine global en autant de sous-domaines horizontaux qu'il y a de processeurs. Tous les sous-domaines horizontaux comprennent la dimension verticale entière. Chaque processeur effectuera alors les tâches du code sur le sous-domaine qui lui est affecté. La mémoire n'étant pas partagée, les informations utiles sont échangées entre les processeurs grâce à la librairie MPI<sup>83</sup>.

On appelle **zone de recouvrement** ou **halo** d'un processeur des lignes ou colonnes qui sont nécessaires au traitement des données de ce processeur mais qui appartiennent au domaine d'un autre processeur et sont donc traitées par ce dernier. Ce sont ces zones de recouvrement qu'il convient de mettre à jour chaque fois que nécessaire. Les zones de recouvrement inter-processeurs sont constituées par des lignes ou colonnes appartenant aux processeurs voisins. Les zones de recouvrement est/ouest des processeurs aux extrêmes gauche ou droit sont constituées de colonnes appartenant aux processeurs à l'autre extrême. Les zones de recouvrement nord dépendent de la configuration choisie et de la variable traitée.

On appelle  $jpiglo$  et  $jpglo$  le nombre de points du domaine global suivant l'axe  $i$  et l'axe  $j$  respectivement. Les colonnes 1 et  $jpglo$  constituent des zones de recouvrement permettant de raccorder les côtés gauche et droit de la grille pour simuler la sphéricité de la Terre. Les lignes  $jpglo$  et éventuellement  $jpglo - 1$  définissent le recouvrement utilisé pour relier les zones nord afin de simuler la continuité du pôle Nord. Aucun recouvrement n'est nécessaire pour les zones sud couvertes par les terres de l'Antarctique, mais en raison des autres options disponibles, la ligne 1 est également considérée comme zone de recouvrement. En tenant compte des zones de recouvrement inter-processeur (que l'on prend ici de taille 1), le nombre total de points pour chaque axe est divisé par le nombre de processeurs suivant cet axe.

On appelle alors  $jpi$  et  $ppj$  le nombre de points des sous-domaines suivant les axes  $i$  et  $j$  respectivement. Lorsque le nombre total de points n'est pas divisible par le nombre de processeurs, les tailles des sous-domaines peuvent ne pas être les mêmes. Dans ce cas, on définit des halos supplémentaires

---

<sup>83</sup>Message Passing Interface.

permettant de réajuster les sous-domaines les plus petits.  $jpgk$  représente le nombre de points suivant la verticale.

### Exemple ORCA2:

$$\begin{aligned} jpiglo &= 182 & jjpglo &= 149 \\ jpi &= 47 & jpj &= 76 & jpgk &= 31 \end{aligned}$$

La figure F.1 montre le détail des sous-domaines pour les processeurs 1, 2 et 5. Pour chacun, les zones de recouvrement sont hachurées en oblique. On a, par exemple, la colonne de recouvrement ouest du processeur 2 qui correspond à la colonne 46, colonne appartenant au processeur 1. Pour le processeur 5 (et tous les processeurs de la ligne du haut), une zone supplémentaire (hachures verticales) est définie pour ajuster la taille du domaine. Les zones blanches sont les zones à traiter par le processeur. Les numéros des lignes et des colonnes sont indiqués en noir pour la numérotation locale et en bleu pour la numérotation globale.

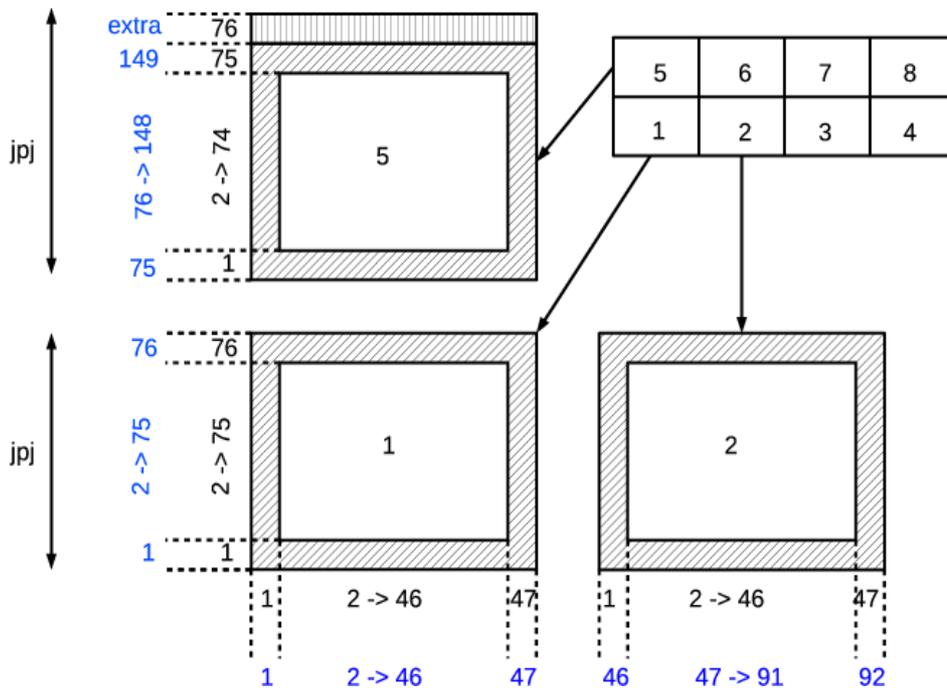


FIG. F.1 – Détail de la décomposition par domaines horizontaux. On identifie les zones utiles (blanc), les zones de recouvrement (hachures obliques) et les zones permettant d’ajuster la taille du domaine (hachures verticales). La numérotation locale est donnée en noire et la numérotation globale en bleu.

## F.1 Réorganisation zonale

La réorganisation zonale consiste à redéfinir les domaines pour chaque ligne de processeurs. Les échanges de données ne s’effectuent donc qu’entre les processeurs d’une même ligne, ce qui permet de restreindre les communications. De plus, pour limiter le nombre de données à échanger, seules les données utiles sont prises en compte, et pas les zones de recouvrement.

Pour chaque processeur, on définit  $njpgloc$ , le nombre de lignes du nouveau domaine. Si  $jpj$  moins les lignes de recouvrement n’est pas divisible par le nombre de processeurs suivant l’axe  $i$ , les différents  $njpgloc$  ne seront pas égaux (différence de 1 ligne maximum). Le processeur le plus à gauche récupère les  $njpgloc$  premières lignes de chacun des processeurs à sa droite et définit ainsi son domaine par  $jpiglo$

moins les zones de recouvrement est/ouest colonnes sur  $n_{jpploc}$  lignes sur  $jpk$  niveaux. Le deuxième processeur couvrira les lignes au-dessus et ainsi de suite. Lorsque la configuration ne comporte qu'un seul processeur suivant l'axe  $i$ , aucune réorganisation n'est nécessaire et les zones de recouvrement est/ouest sont directement prises en comptes.

### Exemple ORCA2:

La figure (F.2) schématise la réorganisation zonale des 8 processeurs.

1 :  $n_{jpploc} = 19$     2 :  $n_{jpploc} = 19$     3 :  $n_{jpploc} = 18$     4 :  $n_{jpploc} = 18$   
 5 :  $n_{jpploc} = 19$     6 :  $n_{jpploc} = 18$     7 :  $n_{jpploc} = 18$     8 :  $n_{jpploc} = 18$

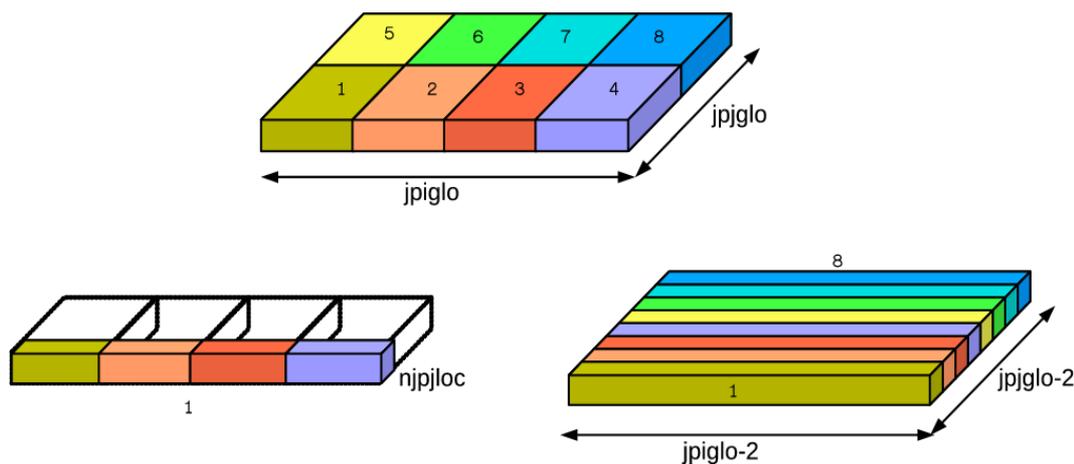


FIG. F.2 – Le processeur 1 construit son domaine en récupérant les lignes 2 à 20 sur les colonnes 2 à 46 de chacun des processeurs 2, 3 et 4 ainsi que les siennes propres.

Les échanges de données nécessaires sont définis lors de l'initialisation du code sous la forme d'un tableau de communication (technique utilisée dans le logiciel PALM<sup>84</sup>, communication personnelle T. Morel, 2009) indiquant le processeur initiant la communication, le type de communication (copie, envoi, réception), le processeur destinataire, les indices d'après la numérotation locale d'origine et les indices d'après la numérotation locale réorganisée.

## F.2 Réorganisation méridienne

Le principe de la réorganisation verticale est de considérer qu'une colonne complète démarre à un point de terre de l'Antarctique, traverse le pôle Nord en suivant le recouvrement demandé en option et revient sur un point de terre de l'Antarctique. Le recouvrement nord consiste à raccorder les colonnes de la moitié gauche du domaine global aux colonnes de la moitié droite, en reliant les extrêmes entre elles et ainsi de suite jusqu'à celles du milieu. L'identification des lignes de recouvrement nord dépend de l'option choisie mais également de la variable. Afin de limiter les données échangées, les zones de recouvrement de la grille ne font pas partie des nouveaux domaines. Pour limiter les communications, les domaines sont réorganisés suivant les colonnes de processeurs : colonnes d'extrême gauche et droite ensemble, et ainsi de suite jusqu'aux colonnes du milieu.

Pour chaque processeur, on définit  $n_{jpiloc}$ , le nombre de colonnes du nouveau domaine. Si  $jpi$  moins les colonnes de recouvrement n'est pas divisible par le double du nombre de processeurs suivant l'axe  $j$ , les différents  $n_{jpiloc}$  ne seront pas égaux (différence de 1 colonne maximum). Le processeur en bas à gauche récupère les  $n_{jpiloc}$  premières colonnes de chacun des processeurs au-dessus de lui ainsi que les  $n_{jpiloc}$  dernières colonnes des processeurs à l'extrême droite du domaine global. Son

<sup>84</sup>[http://www.cerfacs.fr/globc/PALM\\_WEB/](http://www.cerfacs.fr/globc/PALM_WEB/)

domaine est donc défini par  $2 \times j_{pglo}$  moins les zones de recouvrement nord et sud lignes sur  $n_{jpiloc}$  colonnes sur  $j_{pk}$  niveaux. Le processeur du dessus couvrira les colonnes plus à droite et ainsi de suite jusqu'au processeur en haut à gauche. Le relais est ensuite pris par le processeur en haut à droite, puis celui du dessous et ainsi de suite jusqu'au processeur en bas à droite. Lorsque le système tourne en mono-processeur, aucune réorganisation n'est nécessaire, l'opérateur de diffusion prenant en compte directement cette possibilité.

### Exemple ORCA2:

La figure (F.3) schématise la réorganisation méridienne des 8 processeurs.

1 :  $n_{jpiloc} = 12$     5 :  $n_{jpiloc} = 11$     8 :  $n_{jpiloc} = 11$     4 :  $n_{jpiloc} = 11$   
 2 :  $n_{jpiloc} = 12$     6 :  $n_{jpiloc} = 11$     7 :  $n_{jpiloc} = 11$     3 :  $n_{jpiloc} = 11$

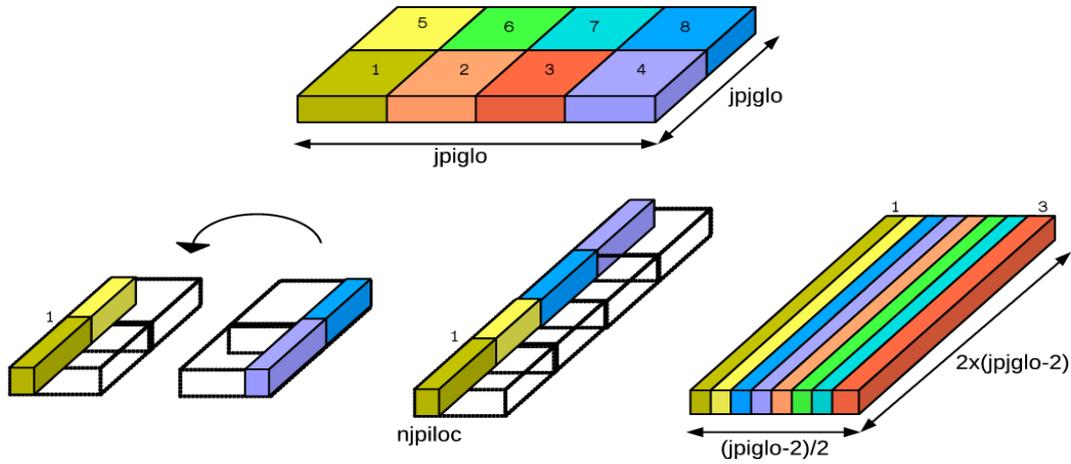


FIG. F.3 – Le processeur 1 construit son domaine en récupérant les colonnes 2 à 13 sur les lignes 2 à 75 du processeur 5 ainsi que les siennes propres. Puis il récupère les colonnes 35 à 46 sur les lignes 2 à 75 des processeurs 8 et 4 qu'il retourne.

Les échanges de données nécessaires sont définis lors de l'initialisation du code sous la forme d'un tableau de communication (technique utilisée dans le logiciel PALM, communication personnelle T. Morel, 2009) indiquant le processeur initiant la communication, le type de communication (copie, envoi, réception), le processeur destinataire, les indices d'après la numérotation locale d'origine, les indices d'après la numérotation locale réorganisée et un drapeau signalant des données à inverser.



# Régularisation de problèmes inverses à l'aide de l'équation de diffusion, avec application à l'assimilation variationnelle de données océaniques

**AUTEUR :** Isabelle MIROUZE  
**DIRECTEURS :** Serge GRATTON - Anthony WEAVER - Luca AMODEI  
**DISCIPLINE :** Mathématiques Appliquées  
**SOUTENANCE :** Le 17 septembre 2010 au CERFACS  
**LABORATOIRE :** CERFACS 42, avenue Gaspard Coriolis 31057 Toulouse Cedex 1  
IMT - Université Paul Sabatier 31 062 Toulouse Cedex 9

---

## Résumé :

En assimilation de données, les modèles de corrélation permettent de caractériser les structures d'erreurs pour les variables définies sur une grille numérique. L'équation de diffusion fournit un cadre flexible et efficace pour représenter des fonctions de corrélation pour des problèmes de grande dimension tels que ceux rencontrés en assimilation variationnelle pour l'atmosphère ou l'océan.

Dans cette thèse, une formulation implicite est d'abord étudiée en détail en dimension un (1D). On montre qu'intégrer une équation de diffusion implicite à coefficient constant sur  $M$  pas de temps est équivalent à convoluer la condition initiale à une fonction autorégressive (AR) d'ordre  $M$ . L'échelle de corrélation de la fonction AR et le facteur de normalisation requis pour générer une amplitude égale à 1 sont donnés en fonction du coefficient de diffusion et de  $M$ . Des extensions du modèle de diffusion permettant aux fonctions de corrélation de ne pas être affectées par les frontières, et tenant compte de variations des échelles sont décrites. Une approximation des facteurs de normalisation est alors proposée.

Des produits d'opérateurs de diffusion implicite 1D sont ensuite utilisés pour construire des modèles de corrélation en dimension deux et trois pour des configurations globales d'un système d'assimilation variationnelle pour le modèle océanique NEMO. Leurs performances sont comparées au modèle de diffusion explicite existant, et des exemples de structures de corrélation où les échelles sont soit paramétrées, soit issues d'une méthode d'ensemble, sont montrés. Enfin, les performances de différentes techniques de normalisation sont comparées.

**Mots clés :** assimilation variationnelle, covariances, corrélations, diffusion, normalisation.

---

## Abstract :

Correlation models are required in data assimilation to characterize the error structures of variables defined on a numerical grid. The diffusion equation provides a flexible and efficient framework for representing correlation functions for problems of large dimension such as those encountered in variational atmospheric or ocean data assimilation.

In this thesis, an implicit formulation of the diffusion equation is first analyzed in detail for the one-dimensional (1D) case. It is shown that integrating a constant-coefficient implicit diffusion equation over  $M$  time steps is equivalent to convolving the initial condition with an  $M$ -th order autoregressive (AR) function. The correlation length scale of the AR function and the normalization factor required for generating a unit amplitude are given in terms of the diffusion coefficient and  $M$ . Extensions of the diffusion model to allow for correlation functions that are not affected by solid boundaries, and that account for varying length scales are described. An approximation of the normalization factors is then proposed.

Products of 1D implicit diffusion operators are then used for constructing two- and three-dimensional correlation models for global configurations of a variational assimilation system for the NEMO ocean model. Their efficiency are compared to the existing explicit diffusion model, and examples of correlation structures are shown, where the length scales are either parametrized or estimated using an ensemble method. Finally, the efficiency of different normalization techniques are compared.

**Keywords :** variational assimilation, covariances, correlations, diffusion, normalization.

---